NORTHWESTERN UNIVERSITY

Large Scale Complex Protein Synthesis Networks: Modeling and Analysis of Translation Elongation Behavior

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Chemical Engineering

By

Hermioni Despina Zouridis

EVANSTON, ILLINOIS

December 2008

ABSTRACT

Large Scale Complex Protein Synthesis Networks: Modeling and Analysis of Translation Elongation Behavior

Hermioni Despina Zouridis

The genetic information in DNA is transcribed to mRNA and then translated to proteins, which form the building blocks of life. Translation, or protein synthesis, is hence a central cellular process. Decades of experimentation have elucidated a vast wealth of molecular information about discrete translation steps, but the sheer complexity of the translation mechanism necessitates that these results be integrated in systematic mathematical frameworks to better understand the system properties of translation and make quantitative predictions.

In this work we have developed a deterministic, sequence specific kinetic model of the translation mechanism that accounts for all its elementary steps. Specifically, our model includes all the elementary steps involved in the elongation cycle at every codon along the length of the mRNA. We performed a sensitivity analysis in order to determine the effects of the kinetic parameters and concentrations of the translational components on the protein synthesis rate. Utilizing our mechanistic framework and sensitivity analysis, we investigate the steady state protein synthesis properties of mRNAs.

In this thesis we first introduce the mechanistic framework and sensitivity analysis, and we utilize them to investigate the protein synthesis properties of a single mRNA species. We then expand our mechanistic framework and sensitivity analysis to account for ternary complex

competitive binding to the ribosomal A site to study effects of codon specific elongation cycle properties on the protein synthesis properties of mRNAs. We finally apply our expanded mechanistic framework and sensitivity analysis to all the genes in the *E. coli* genome. We determine (i) the interplay between ribosomal occupancy of elongation cycle intermediate states and ribosome distributions with respect to codon position along the length of the mRNA leads to polysome self-organization that drives translation rate to maximum levels (ii) the relative position of codons along the mRNA determines the optimal protein synthesis rate, and the translation rates of mRNAs are controlled by segments of sequence specific rate limiting codons, and (iii) minor codons play a role in optimizing translation rate, and their usage is important to the optimized, systemic allocation of ribosomes in the translation of mRNAs throughout the cell.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Vassily Hatzimanikatis, and co-advisor, Olke Uhlenbeck, for their guidance and support throughout my doctoral studies. I would like to thank my thesis committee members, Linda Broadbelt, Mary Silber, and Terry Papoutsakis, for sharing their insight with me. I would like to thank my family for their love and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1 Motivation

The genetic information in DNA is transcribed to mRNA and then translated to proteins, which form the building blocks of life. Translation, or protein synthesis, is a process that is central to cellular function and well conserved among all living organisms. It is one of processes in the "central dogma" of molecular biology and the last step in information transfer from DNA to protein. Decades of experimentation have elucidated a vast wealth of molecular information about discrete translation steps, but the sheer complexity of the translation mechanism necessitates that these results be integrated in systematic mathematical frameworks to better understand the system properties of translation and make quantitative predictions. A better understanding of protein synthesis is of great importance in many areas of medicine and biotechnology. For instance, many antibiotics function by inhibiting bacterial translational machinery. With antibiotic resistance an increasing problem (1) understanding translation will help elucidate mechanisms of antibiotic action and resistance. Also, translational malfunction has been implicated in cancer cell proliferation (2). Moreover, understanding protein synthesis will aid in optimization of pharmaceutical protein production. A large number of important pharmaceuticals, such as insulin and anti-anemia drugs, are proteins produced by recombinant DNA technology from *E. coli*. Quantifying in a systematic manner how growth and environmental conditions, along with genetic perturbations, manifest themselves at the level of proteins will serve as a tool for expressing needed proteins.

Several studies have been conducted involving investigating the kinetics of protein synthesis that take into account the ribosome movement on mRNAs (3-5), and other studies (6, 7) have involved the effects of competition for ribosomes between mRNAs on cell wide mapping

between mRNA and protein levels. An assumption in these studies is that the elongation kinetics at each codon depends on a single rate constant that is the same for all codon species at all positions along the length of the mRNA. In reality, codons have varying elongation kinetics due to different tRNA availabilities (8) and codon – anticodon compatibilities (9-11), and the multiple elementary steps and translational components involved in the elongation cycle at every codon. Therefore, a better understanding of the properties of translation requires the consideration of the translation elongation phase, accounting for all elongation cycle intermediate steps. Hence, in this work we have developed a deterministic, sequence specific kinetic model of the translational machinery that accounts for all the elementary steps of the translation mechanism. Specifically, our model includes all the elementary steps involved in the elongation cycle at every codon along the length of the mRNA. We performed a sensitivity analysis in order to determine the effects of the kinetic parameters and concentrations of the translational components on the protein synthesis rate. Utilizing our mechanistic framework and sensitivity analysis, we investigate the steady state protein synthesis properties, i.e., translation rate, rate limiting intermediate translation steps, polysome size, and ribosome distributions, of mRNAs.

1.2 Research outline

In this thesis we present a deterministic, sequence specific mechanistic framework for protein synthesis that accounts for all of its elementary steps and perform a sensitivity analysis in order to investigate the steady state protein synthesis properties of mRNAs. In Chapter 2 we introduce the mechanistic framework and sensitivity analysis. Applying the mechanistic framework and sensitivity analysis to investigate the protein synthesis properties of a single mRNA species, we determine that translation rate of at a given polysome size depends on the

complex interplay between ribosomal occupancy of elongation cycle intermediate states and ribosome distributions with respect to codon position along the length of the mRNA, and this interplay leads to polysome self-organization that drives translation rate to maximum levels. This work has been published in (12). In Chapter 3 we expand the mechanistic framework from Chapter 2 to account for ternary complex competitive binding to the ribosomal A site. We also expand our sensitivity analysis to make it codon specific, meaning that we account for the contribution of kinetic parameters and translational component concentrations of each codon on the overall protein synthesis rate. We applied our expanded mechanistic framework and sensitivity analysis to investigate the protein synthesis properties of mRNAs having codon frequencies representative of those of the *E. coli* genome, observe that the relative position of codons along the mRNA determines the optimal protein synthesis rate, and find that the translation rates of mRNAs are controlled by segments of rate limiting codons that are sequence specific. This work has been accepted for publication by an academic journal. In Chapter 4 we apply the expanded mechanistic framework and sensitivity analysis from Chapter 3 to all the genes in the *E. coli* genome. By calculating the relative translational efficiency of each gene, which we define as the measure of the tradeoff between the overall translation rate of the gene and the number of ribosomes needed to achieve that rate, we determine that less abundant, or minor, codons play a role in optimizing translation rate. We hypothesize that the usage of minor codons are important to the optimized, systemic allocation of ribosomes in the translation of mRNAs throughout the cell. This work is being prepared for submission to an academic journal. In Chapter 5 we summarize the novel hypotheses emerging from this work, collaborative research efforts, and recommendations for future work.

Chapter 2: Polysome self-organization and protein synthesis properties

2.1 Introduction

In this chapter we present a deterministic, sequence specific kinetic model of the translational machinery that accounts for all the elementary steps of the translation mechanism. Specifically, our model includes all the elementary steps involved in the elongation cycle at every codon along the length of the mRNA. We performed a sensitivity analysis in order to determine the effects of the kinetic parameters and concentrations of the translational components on the protein synthesis rate. Utilizing our mechanistic framework and sensitivity analysis, we investigate the steady state protein synthesis properties of a single mRNA species. We determine the protein synthesis rate as a function of polysome size and then identify ranges of polysome sizes in which the translation kinetics are initiation, elongation, and termination limited. Additionally, we investigate how ribosomes are distributed with respect to elongation cycle intermediate state and sequence position under initiation, elongation, and termination limited regimes. To understand how each elongation cycle elementary step contributes to the kinetics of a given elongation cycle, we introduce a reduced version of our model. We propose that translation rate at a given polysome size depends on the complex interplay between ribosomal occupancy of elongation cycle intermediate states and ribosome distributions with respect to codon position along the length of the mRNA, and this interplay leads to polysome self-organization that drives translation rate to maximum levels.

2.2 Methods

2.2.1 The overall translation mechanism

Translation is essentially a template polymerization process (13) consisting of initiation, elongation, and termination phases. Messenger RNA (mRNA), composed of a sequence of

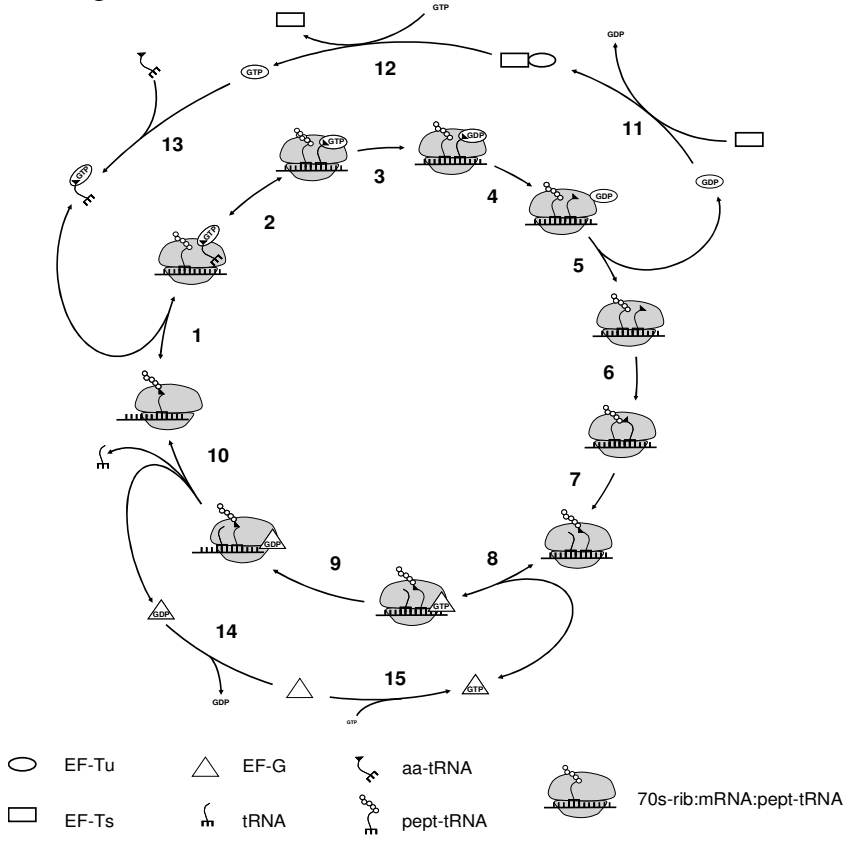
codons coding for amino acids, carries genetic information. Initiation occurs with binding of the ribosome to the ribosomal binding site near the 5' end of the mRNA. During the elongation phase the ribosome facilitates assembly of the polypeptide chain with one amino acid (aa) added per elongation cycle at each codon. Amino acids are delivered to the ribosome by transfer RNAs (tRNAs) in the form of ternary complexes that serve as adapter molecules between the amino acid and the codon present in the ribosomal A site. Termination involves release of the completed peptide from the ribosome near the 3' end of the mRNA. Multiple proteins can be synthesized simultaneously on a single mRNA molecule, forming a structure called the polysome (or polyribosome) consisting of several ribosomes simultaneously translating the same mRNA. Polysome size is the number of ribosomes bound to a single mRNA molecule. Hence, the higher the polysome size, the greater the coverage of the mRNA due to ribosomes translating it is. Polysomes have been observed experimentally (14), and modern techniques have allowed the quantification of polysome size for almost every mRNA in yeast cells (15).

2.2.2 Elementary steps of the elongation cycle

The translation elongation phase is a cyclic process that involves codons, ribosomes, amino acids, tRNAs, elongation factors Tu, Ts, and G, and leads to the assembly of polypeptide chains (Figure 2.1). Each aminoacyl-tRNA (aa-tRNA) binds to Ef-Tu:GTP, forming a ternary complex (step 13). The ternary complex then binds reversibly to the ribosomal A site in a codon independent manner (step 1). After finding the correct codon match and reversible codon dependent binding (step 2), GTP is hydrolyzed (step 3), Ef-Tu:GDP changes position on the ribosome (step 4) and is released (step 5). In a two step process, Ef-Ts catalyzes regeneration of Ef-Tu:GTP (steps 11 and 12). During accommodation the aa-tRNA undergoes a conformation change and enters the A site (step 6). Transpeptidation then occurs (step 7), where the peptide

chain is transferred from the peptidyl-tRNA to the aa-tRNA, resulting in the elongation of the polypeptide chain by one amino acid. Reversible binding of Ef-G:GTP (step 8) facilitates translocation (step 9). During translocation the P site tRNA and codon move to the E site of the ribosome and the A site tRNA and codon move to the P site, resulting in the complex moving toward the 3' end of the mRNA by one codon. The tRNA in the E site is released along with Ef-G:GDP (step 10), and Ef-G:GTP is recycled in a two step process (steps 14 and 15).

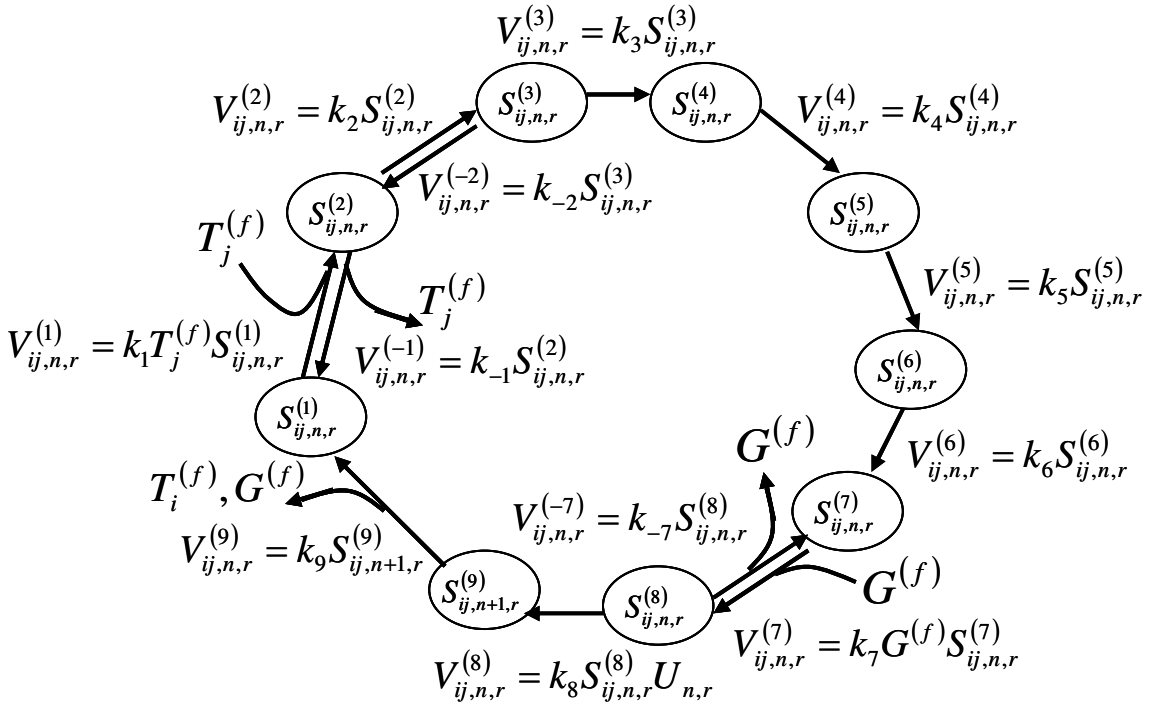
Figure 2.1: The elementary mechanistic steps of the translation elongation process. Ribosomal A, P, and E sites indicated on the intermediates between steps 1 and 2 and steps 9 and 10. Step 1: Reversible, codon independent binding of the ternary complex to the ribosomal A site. Step 2: Reversible, codon dependent binding of the ternary complex to the ribosomal A site. Step 3: GTP hydrolysis. Step 4: Ef-Tu:GDP position change on the ribosome. Step 5: Ef-Tu:GDP release. Step 6: aa-tRNA accommodation. Step 7: Transpeptidation. Step 8: Reversible binding of Ef-G:GTP. Step 9: Translocation. Step 10: E site tRNA release. Steps 11 and 12: Ef-Ts catalyzed regeneration of Ef-Tu:GTP. Step 13: Ef-Tu:GTP binding to the aa-tRNA. Steps 14 and 15: Regeneration of Ef-G:GTP.



2.2.3 Model formulation

We have employed the following assumptions in the formulation of the mathematical model for the elongation cycle. A graphical representation of the elementary steps of the elongation cycle with nomenclature from the model formulation is included in Figure 2.2.

Figure 2.2: A graphical representation of the elementary elongation steps of the translation elongation process with nomenclature from the model formulation as explained in the text. Fluxes $V_{ij,n,r}^{(1)}$ and $V_{ij,n,r}^{(-1)}$ represent reversible, codon independent binding of the ternary complex to the ribosomal A site. Fluxes $V_{ij,n,r}^{(2)}$ and $V_{ij,n,r}^{(-2)}$ represent reversible, codon dependent binding of the ternary complex to the ribosomal A site. Flux $V_{ij,n,r}^{(3)}$ represents GTP hydrolysis. Fluxes $V_{ij,n,r}^{(4)}$ and $V_{ij,n,r}^{(5)}$ represent Ef-Tu:GDP position change on the ribosome and Ef-Tu:GDP release, respectively. Flux $V_{ij,n,r}^{(6)}$ represents aa-tRNA accommodation. Fluxes $V_{ij,n,r}^{(7)}$ and $V_{ij,n,r}^{(-7)}$ represent reversible binding of Ef-G:GTP. Flux $V_{ij,n,r}^{(8)}$ represents ribosomal translocation. Flux $V_{ij,n,r}^{(9)}$ represents E site tRNA release. The intermediate elongation cycle states that occur before and after transpeptidation (step 7, Figure 2.1) are considered to be one state in our model $\left(S_{ij,n,r}^{(7)}\right)$. After release of the tRNA in the ribosomal E site $\left(V_{ij,n,r}^{(9)}\right)$, the subsequent elongation cycle is ready to begin with the ribosomal P site positioned at codon *n+1*.



Assumption 1: The elongation cycle is modeled in terms of the states of the ribosome.

Ribosomes at different stages of the elongation cycle are considered to be separate states, $S_{ij,n,r}^{(\sigma)}$. Each state σ is of type *ij*, where *i* is the codon species occupying the P site and *j* is the codon species occupying the A site, and *n* denotes the position of the ribosomal P site codon. Ribosomes are bound to mRNA species *r*, with $r \in M$, and *M* is the set of mRNA species. Codon positions along mRNA sequences are numbered from *1* to N_r starting at the 5' end of the protein coding region, N_r denotes the number of codons (length) of mRNA species *r*.

Assumption 2: Ribosomes cover 12 codons on the mRNA.

We assume that a ribosome covers $L = 12$ codons on an mRNA (16-18), where *L* is the length of the ribosome. The front and back of the ribosome are defined to be on the sides closest to the 3' and 5' ends of the mRNA, respectively, with the A and P sites covering the sixth and seventh codons relative to the front of the ribosome. Hence, in addition to the codons occupying the ribosomal A and P sites in an elongation cycle state, the five codons preceding and following the A and P site codons are also covered by the ribosome.

Assumption 3: All free tRNAs are in the form of ternary complexes

Because most tRNAs are charged (19), Ef-Tu is present in a one-to-one ratio with tRNA (13), and the association rate constant of Ef-Tu:GTP to charged tRNAs is very high (20), we consider all free tRNAs to be in the form of ternary complexes. Free ternary complex concentrations $\left(T_k^{(f)}\right)$ are of species *k*, with $k \in K$, where *K* is the set of ternary complex species. This assumption can be relaxed by including flux expressions corresponding to ternary complex formation.

Assumption 4: The set of reaction rate constants governing the kinetics of the intermediate elongation cycle steps are the same for all codons.

We assume that every elongation cycle, regardless of the ternary complex species that binds to the ribosomal A site, the tRNA occupying the ribosomal P site, and the codon species occupying the A and P sites, have the same elementary steps and the same rate constants for each elementary step. Although experimental evidence suggests that synonymous codons translated by the same tRNA are not necessarily translated at the same rate (9, 10), rate constants specific to each codon species have yet to be determined. Hence, in the absence of this information, the same set of reaction rate constants (Table 2.1) adapted from the literature (21-24) were applied to the flux expressions (Table 2.2) of the intermediate steps of all elongation cycles. We assume the temperature and Mg concentration to be 37 ^{o}C and 7 mM, respectively, so the reaction rate constants not determined experimentally at these conditions were adjusted accordingly.

Table 2.1: Kinetic parameters

Parameter	Definition	Characteristic value
k_1	Rate constant of ternary complex codon independent binding[*†]	$100\,\mu M^{-1}s^{-1}$
k_{-1}	Rate constant of ternary complex codon independent binding reverse reaction[*†]	$79\,s^{-1}$
k_2	Rate constant of ternary complex codon dependent binding[*†]	$207\,s^{-1}$
k_{-2}	Rate constant of ternary complex codon dependent binding reverse reaction[*†]	$3.45\,s^{-1}$
k_3	Rate constant of GTP hydrolysis[‡]	$100\,s^{-1}$
k_4	Rate constant of Ef-Tu:GDP position change on the ribosome[*]	$638\,s^{-1}$
k_5	Rate constant of Ef-Tu:GDP release[*]	$15\,s^{-1}$
k_6	Rate constant of A site tRNA accommodation[‡]	$20\,s^{-1}$
k_7	Rate constant of Ef-G:GTP binding[¶]	$150\,\mu M^{-1}s^{-1}$
k_{-7}	Rate constant of Ef-G:GTP binding reverse reaction[¶]	$140\,s^{-1}$
k_8	Rate constant of ribosome translocation[¶]	$250\,s^{-1}$
k_9	Rate constant of E site tRNA release[¶]	$20\,s^{-1}$
$k_{I,r}$	Translation initiation rate constant	Allowed to vary
$k_{T,r}$	Translation termination rate constant	Allowed to vary

*from (23), †from (22) (activation energies were used to adjust rate constants for temperature), ‡from (21), ¶from (24)

Table 2.2: Flux expressions

Flux	Expression	Description
$V_{ij,n,r}^{(1)}$	$k_1 T_j^{(f)} S_{ij,n,r}^{(1)}$	Ternary complex codon independent binding
$V_{ij,n,r}^{(-1)}$	$k_{-1} S_{ij,n,r}^{(2)}$	Ternary complex codon independent binding reverse reaction
$V_{ij,n,r}^{(2)}$	$k_2 S_{ij,n,r}^{(2)}$	Ternary complex codon dependent binding
$V_{ij,n,r}^{(-2)}$	$k_{-2} S_{ij,n,r}^{(3)}$	Ternary complex codon dependent binding reverse reaction
$V_{ij,n,r}^{(3)}$	$k_3 S_{ij,n,r}^{(3)}$	GTP hydrolysis
$V_{ij,n,r}^{(4)}$	$k_4 S_{ij,n,r}^{(4)}$	Ef-Tu:GDP position change on the ribosome
$V_{ij,n,r}^{(5)}$	$k_5 S_{ij,n,r}^{(5)}$	Ef-Tu:GDP release
$V_{ij,n,r}^{(6)}$	$k_6 S_{ij,n,r}^{(6)}$	A site tRNA accommodation
$V_{ij,n,r}^{(7)}$	$k_7 G^{(f)} S_{ij,n,r}^{(7)}$	Ef-G:GTP binding
$V_{ij,n,r}^{(-7)}$	$k_{-7} S_{ij,n,r}^{(8)}$	Ef-G:GTP binding reverse reaction
$V_{ij,n,r}^{(8)}$	$k_8 S_{ij,n,r}^{(8)} U_{n,r}$	Ribosome translocation
$V_{ij,n,r}^{(9)}$	$k_9 S_{ij,n+1,r}^{(9)}$	E site tRNA release
$V_{I,r}$	$k_{I,r} R^{(f)} C_{n+6,r}^{(f)}, n=1$	Translation initiation
$V_{T,r}$	$k_{T,r} S_r^T$	Translation termination

Assumption 5: Ternary complex binding kinetics is first order with respect to free ternary complex concentrations and the ribosomal state having the A site empty.

The elongation cycle begins with binding of the ternary complex $\left(T_k^{(f)}\right)$ to state 1, which is the ribosomal state having the A site empty $\left(S_{ij,n,r}^{(1)}\right)$, to form state 2 $\left(S_{ij,n,r}^{(2)}\right)$. This step corresponds to flux $V_{ij,n,r}^{(1)}$. State 2 represents non specific binding of ternary complexes to the A site.

Assumption 6: Only ternary complexes cognate to the A site codon bind to the ribosome during non specific binding.

Although ternary complexes can be incorrectly bound to non-cognate codons at this point in the elongation cycle, for simplicity we assume that the concentration of incorrectly bound ternary complexes to the A site is comparatively small and consider state 2 to consist only of correctly bound ternary complexes to the A site. However, this assumption can be relaxed by adding additional states to the model.

Assumption 7: Elongation cycle kinetics after ternary complex binding and before Ef-G:GTP binding are first order with respect to ribosomal states.

After non specific binding of the ternary complex to the ribosomal A site the correct codon – anticodon match is verified, which corresponds to flux $V_{ij,n,r}^{(2)}$, and leads to the formation of state 3 $\left(S_{ij,n,r}^{(3)}\right)$. State 3 participates in GTP hydrolysis, which corresponds to flux $V_{ij,n,r}^{(3)}$, forming state 4 $\left(S_{ij,n,r}^{(4)}\right)$. Ef-Tu:GDP changes position on the ribosome to form state 5 $\left(S_{ij,n,r}^{(5)}\right)$ and then dissociates from the ribosome to form state 6 $\left(S_{ij,n,r}^{(6)}\right)$, corresponding to fluxes $V_{ij,n,r}^{(4)}$ and $V_{ij,n,r}^{(5)}$, respectively. Accommodation, corresponding to flux $V_{ij,n,r}^{(6)}$, occurs when the aa-tRNA enters the A site of the ribosome and leads to formation of state 7 $\left(S_{ij,n,r}^{(7)}\right)$.

Assumption 8: Ef-G:GTP binding kinetics is first order with respect to free Ef-G:GTP concentration and the ribosomal state that exists before Ef-G:GTP binding.

Ef-G:GTP $\left(G^{(f)}\right)$ binds to state 7 $\left(S_{ij,n,r}^{(7)}\right)$, corresponding to flux $V_{ij,n,r}^{(7)}$, leading to the formation of state 8 $\left(S_{ij,n,r}^{(8)}\right)$ which participates in translocation. In addition, we assume that all

free Ef-G, before Ef-G:GTP binding to and after Ef-G:GDP release from the ribosome, is in the form of Ef-G:GTP

Assumption 9: Ribosome translocation also depends on the conditional probability that the codon adjacent to the codon occupied by the front of the ribosome is free given that the previous codon is occupied by the front of the ribosome.

Translocation kinetics are dependent on the conditional probability that the codon adjacent to the codon occupied by the front of the ribosome is free given that the previous codon is occupied by the front of the ribosome (equation 2.1).

$$U_{n,r} = \begin{cases} \dfrac{C^{(f)}_{n+7,r}}{C^{(f)}_{n+6,r} + \sum_{\sigma} S^{(\sigma)}_{ij,n,r}} & , n \in \left[1, N_r - (L+1)\right] \\ 1 & , n \in \left[N_r - L, N_r - 1\right] \end{cases} \tag{2.1}$$

This relationship is adapted from (4). Instead of all free codons at position *n+7* $\left(C^{(f)}_{n+7,r}\right)$ being available to ribosomes participating in elongation cycles at position *n*, only the fraction of free codons at position *n+7* preceded by codons at position *n+6* occupied by the front of ribosomes are available for ribosome occupancy after translocation. We assume that the flux corresponding to translocation, $V^{(8)}_{ij,n,r}$, is first order with respect to $U_{n,r}$ and state 8 $\left(S^{(8)}_{ij,n,r}\right)$.

Assumption 10: The ribosomal state that exists before E site tRNA release does not have a tRNA present in the A site.

During translocation the codon and tRNA in the P site move to the E site, the codon and tRNA in the A site move to the P site, and the downstream codon in the sequence moves to the A site to form state 9 $\left(S^{(9)}_{ij,n+1,r}\right)$. The tRNA in the E site dissociates from state 9 to form state 1 of

the following elongation cycle, and we assume that this step, corresponding to flux $V_{ij,n,r}^{(9)}$, is first order with respect to state 9.

Assumption 11: Ribosomes, ternary complexes, Ef-G, and codons are conserved species with constant total concentrations.

The total ribosome $\left(R^{(t)}\right)$, ternary complex $\left(T_k^{(t)}\right)$, Ef-G:GT(D)P $\left(G^{(t)}\right)$, and mRNA species $\left(M_r\right)$ concentrations are assumed to be constant (time-invariant) and are described by the following conservation equations. Free ribosomes, ternary complexes, Ef-G complexes, and codons are denoted by $R^{(f)}$, $T_k^{(f)}$, $G^{(f)}$, and $C_{n,r}^{(f)}$, respectively.

$$R^{(t)} = \sum_r \sum_{n=1}^{N_r-1} \sum_{\sigma} \left(S_{ij,n,r}^{(\sigma)} + S_r^T\right) + R^{(f)} \tag{2.2}$$

$$T_k^{(t)} = \sum_r \sum_{n=1}^{N_r-1} \left\{ \left(S_{kk,n,r}^{(\sigma)} + 2\sum_{\sigma=2}^{9} S_{kk,n,r}^{(\sigma)} \right) + \sum_{\sigma=2}^{9} S_{ik,n,r}^{(\sigma)} + \sum_{\sigma} S_{kj,n,r}^{(\sigma)} \right\} + T_k^{(f)} \qquad , k \in K \tag{2.3}$$

$$G^{(t)} = \sum_r \sum_{n=1}^{N_r-1} \sum_{\sigma=8}^{9} S_{ij,n,r}^{(\sigma)} + G^{(f)} \tag{2.4}$$

Ribosomes participate in all states at every position on the mRNA (equation 2.2), with the state S_r^T described in following assumptions. Ternary complexes *k*, with $k \in K$ and *K* the set of ternary complex species, are bound to states where they are cognate to either or both of the A and P site codons (equation 2.3), and Ef-G:GT(D)P is bound to states 7, 8, and 9 of every elongation cycle (equation 2.4). Codons participate in all states in which they occupy either the ribosomal A sites or P sites. Additionally, because ribosomes cover twelve codons on the mRNA, along with the codons occupying the A and P sites the five preceding and following codons are also covered during an elongation cycle (***Assumption 2***) and are unavailable for participation in other elongation cycles. The total concentration of a codon at a specific position

on mRNA r is equal to the concentration of mRNA r $\left(M_r\right)$, which is why the free codon concentrations are dependent on M_r. Below are the conservation equations for codons:

$$M_r = \sum_{n}^{n+6} \sum_{\sigma=1}^{8} S_{ij,n,r}^{(\sigma)} + C_{n,r}^{(f)} \qquad , n = 1 \tag{2.5}$$

$$M_r = \sum_{n}^{n+6} \left(S_{ij,n-1,r}^{(9)} + \sum_{\sigma=1}^{8} S_{ij,n,r}^{(\sigma)} \right) + C_{n,r}^{(f)} \quad , n \in \left[2, N_r - \left(L+1\right)\right] \tag{2.6}$$

Assumption 12: Translation initiation is a bimolecular reaction between the ribosome and the initiation site of the mRNA.

Translation initiation, corresponding to flux $V_{I,r}$, is considered to be first order with respect to free ribosomes $\left(R^{(f)}\right)$ and the free mRNA initiation sites. The initiation site is defined here as the first seven codons of the protein coding region of the mRNA, with the first codon of the protein coding region as the start codon, and the adjacent noncoding five codons upstream of the start codon. Hence, we assume that the concentration of free mRNA initiation sites is equal to $C_{7,r}^{(f)}$. Translation initiation results in positioning of the ribosomal P site over the start codon.

Assumption 13: Translation termination is a single step process in which the ribosome dissociates from the mRNA.

We introduce the state S_r^T, which corresponds to the state after the completion of the final elongation cycle and prior to termination. We assume translation termination kinetics to be first order with respect to this state, and corresponds to flux $V_{T,r}$.

2.2.4 Mathematical model

The equations that describe the dynamics of the transition between the nine states of the elongation cycle, along with initiation and termination, are as follows:

$$\frac{dS^{(1)}_{ij,n,r}}{dt} = V_{I,r} + V^{(-1)}_{ij,n,r} - V^{(1)}_{ij,n,r} \qquad , n = 1 \qquad (2.7)$$

$$\frac{dS^{(1)}_{ij,n,r}}{dt} = V^{(9)}_{ij,n-1,r} + V^{(-1)}_{ij,n,r} - V^{(1)}_{ij,n,r} \qquad , n \in [2, N_r - 1] \qquad (2.8)$$

$$\frac{dS^{(2)}_{ij,n,r}}{dt} = V^{(1)}_{ij,n,r} + V^{(-2)}_{ij,n,r} - V^{(2)}_{ij,n,r} - V^{(-1)}_{ij,n,r} \qquad , n \in [1, N_r - 1] \qquad (2.9)$$

$$\frac{dS^{(3)}_{ij,n,r}}{dt} = V^{(2)}_{ij,n,r} - V^{(-2)}_{ij,n,r} - V^{(3)}_{ij,n,r} \qquad , n \in [1, N_r - 1] \qquad (2.10)$$

$$\frac{dS^{(4)}_{ij,n,r}}{dt} = V^{(3)}_{ij,n,r} - V^{(4)}_{ij,n,r} \qquad , n \in [1, N_r - 1] \qquad (2.11)$$

$$\frac{dS^{(5)}_{ij,n,r}}{dt} = V^{(4)}_{ij,n,r} - V^{(5)}_{ij,n,r} \qquad , n \in [1, N_r - 1] \qquad (2.12)$$

$$\frac{dS^{(6)}_{ij,n,r}}{dt} = V^{(5)}_{ij,n,r} - V^{(6)}_{ij,n,r} \qquad , n \in [1, N_r - 1] \qquad (2.13)$$

$$\frac{dS^{(7)}_{ij,n,r}}{dt} = V^{(6)}_{ij,n,r} + V^{(-7)}_{ij,n,r} - V^{(7)}_{ij,n,r} \qquad , n \in [1, N_r - 1] \qquad (2.14)$$

$$\frac{dS^{(8)}_{ij,n,r}}{dt} = V^{(7)}_{ij,n,r} - V^{(-7)}_{ij,n,r} - V^{(8)}_{ij,n,r} \qquad , n \in [1, N_r - 1] \qquad (2.15)$$

$$\frac{dS^{(9)}_{ij,n+1,r}}{dt} = V^{(8)}_{ij,n,r} - V^{(9)}_{ij,n+1,r} \qquad , n \in [1, N_r - 3] \qquad (2.16)$$

$$\frac{dS^{T}_{r}}{dt} = V^{(9)}_{ij,n+1,r} - V_{T,r} \qquad , n = N_r - 2 \qquad (2.17)$$

Equations (2.7) – (2.17) above, together with the conservation equations (2.2) – (2.6), comprise our model that we call the ZH model. In our computational studies presented in following sections we also consider the lattice model of protein synthesis first proposed by MacDonald and others (4) and MacDonald and Gibbs (3) and later extended by Heinrich and Rapoport (5), or what we call the

MG-HR model. A description of the MG-HR model is included in APPENDIX A. We performed a sensitivity analysis based on the metabolic control analysis framework (25-30) to determine the sensitivity of steady state concentrations and fluxes with respect to input parameters for our model. Details of the sensitivity analysis are included in APPENDIX B.

2.3 Computational studies

We applied our mathematical model of translation elongation to investigate the steady state properties of translation of the *trpR* gene in *E. coli*. The kinetic data available on the intermediate steps of the *E. coli* elongation cycle and the data available on the intracellular concentrations of the translational machinery make it possible to readily study protein synthesis properties of *E. coli* genes. However, our mechanistic framework is applicable to other organisms. Estimates for the concentration of a single mRNA species r $\left(M_r\right)$, the total ribosome concentration available to participate in translation $\left(R^{(t)}\right)$, the total Ef-G concentration available to participate in translation $\left(G^{(t)}\right)$, and the total concentrations of ternary complexes available for ribosomal A site binding $\left(T_k^{(t)}\right)$ used in the computational studies are included in APPENDIX C.

2.3.1 Protein synthesis properties and polysome size

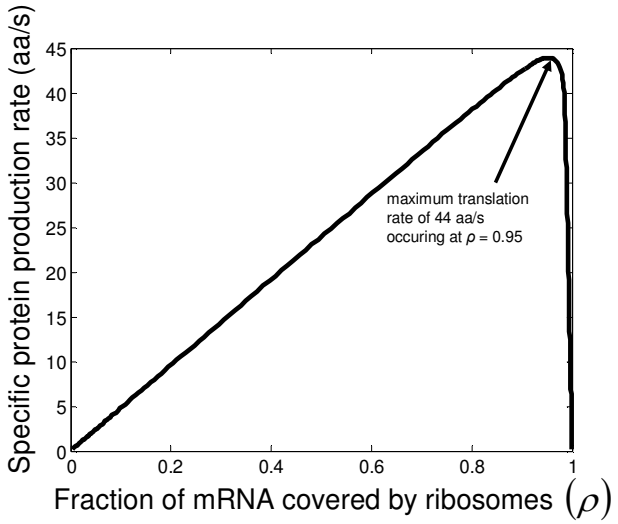
We investigated how translation rate and control relate to polysome size. We define ribosomal fractional coverage, i.e., ribosome density, as the fraction of mRNAs covered by bound ribosomes, where:

$$\rho = \frac{L\sum_{n}\sum_{\sigma} S_{ij,n,r}^{(\sigma)}}{M_r N_r} \tag{2.18}$$

The ribosomal fractional coverage is proportionate to polysome size. At steady state, for a given ribosomal fractional coverage, a set of pairs of initiation and termination rate constants can be

determined. Each of these pairs corresponds to a unique protein synthesis rate. We first determined the pairs of initiation and termination rate constants corresponding to each ribosomal fractional coverage for $0 < \rho < 1$. We hypothesized that at any given growth condition the cell maximizes the protein production rates from each of its mRNAs. Therefore, to determine the relationship between translation rate and polysome size we considered the pair of initiation and termination rate constants corresponding to the maximum specific protein synthesis rate, i.e., the protein synthesis rate per mRNA molecule, for each ribosomal fractional coverage. Figure 2.3 shows the specific protein production rate as a function of the fraction of the mRNA covered by ribosomes, ρ. We observe that as ribosomal fractional coverage increases the protein synthesis rate increases, reaches a maximum, and then decreases. Our model predicts that the maximum translation rate of 44 amino acids/s occurs at $\rho = 0.95$. Moreover, the observed range of protein synthesis rates is consistent with experimental reports (31, 32).

Figure 2.3: Relationships between translation properties and polysome size: Specific protein production rate as a function of polysome size.

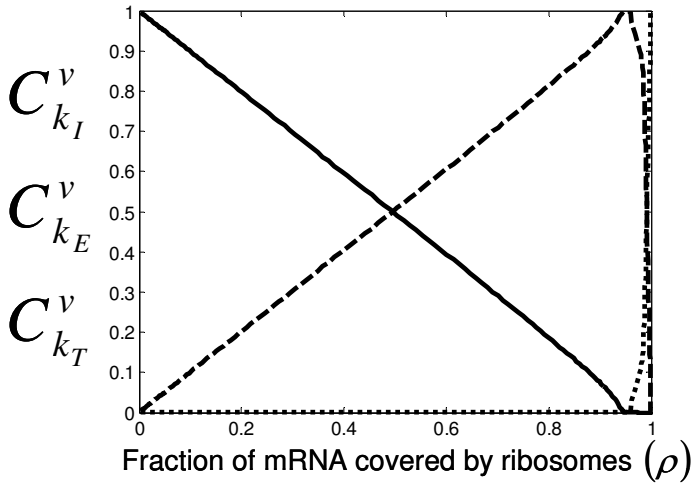


Rate limiting steps and polysome size

We applied the control analysis framework to the model in order to determine if translation is initiation, elongation, or termination limited under different polysome sizes.

We observe that the initiation control coefficients are maximal for low ribosome density. As the ribosome density increases, the elongation control coefficients increase, reach a maximum, and then decrease, and the termination control coefficients are maximum at high polysome sizes (Figure 2.4). We define translation kinetics at a single polysome size to be initiation limited if $C_{k_I}^v > C_{k_E}^v + C_{k_T}^v$, elongation limited if $C_{k_E}^v > C_{k_I}^v + C_{k_T}^v$, and termination limited if $C_{k_T}^v > C_{k_I}^v + C_{k_E}^v$. We observe that translation is initiation limited for $\rho < 0.5$, elongation limited for $0.5 < \rho < 0.99$, with elongation control maximal at the same ribosomal fractional coverage that specific protein production rate is maximal, and termination limited for $\rho > 0.99$.

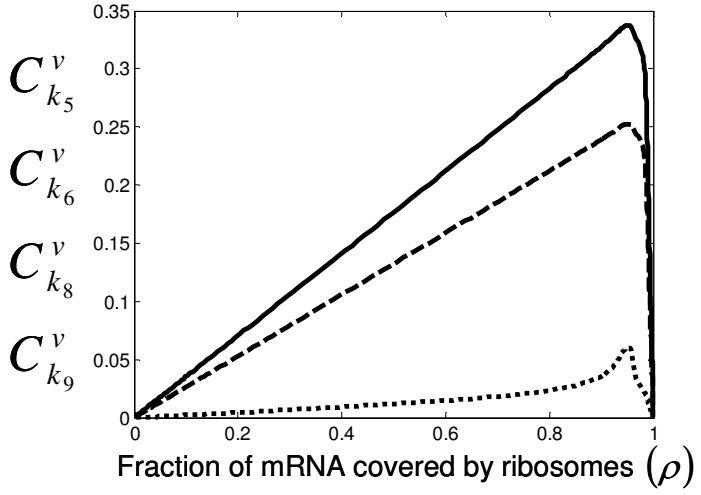
Figure 2.4: Relationships between translation properties and polysome size: Initiation control coefficients, $C_{k_I}^v$ (solid line), elongation control coefficients, $C_{k_E}^v$ (dashed line), and termination control coefficients, $C_{k_T}^v$ (dotted line), as functions of polysome size.



Rate limiting steps in the elongation cycle

The flux control coefficients with respect to the rate constants corresponding to the elementary elongation cycle steps were also investigated as functions of polysome size. We observe that the control coefficient with respect to the Ef-Tu:GDP release rate constant, $C_{k_5}^{v}$, is maximum (Figure 2.5). This result is consistent with experimental reports which identify Ef-Tu:GDP release as the rate limiting step of the elongation cycle (23). Control coefficients for A site tRNA accommodation $\left(C_{k_6}^{v}\right)$ and E site tRNA release $\left(C_{k_9}^{v}\right)$ are equal to each other and also high (Figure 2.5). The control coefficient with the third highest magnitude, $C_{k_8}^{v}$, corresponds to the translocation step (Figure 2.5). The remaining elongation cycle intermediate steps have low control coefficients.

Figure 2.5: Relationships between translation properties and polysome size: Elongation cycle intermediate control coefficients with respect to k_5, $C_{k_5}^{v}$ (solid line), k_6, $C_{k_6}^{v}$, and k_9, $C_{k_9}^{v}$ (dashed line), and k_8, $C_{k_8}^{v}$ (dotted line), as functions of polysome size.



2.3.2 Ribosome distributions

To analyze steady state ribosomal position and state occupancies the following quantities are introduced:

$$S_{ij,n,r} = \sum_{\sigma} S_{ij,n,r}^{(\sigma)} \Big/ M_r \tag{2.19}$$

$$\delta_{n,r}^{(\sigma)} = S_{ij,n,r}^{(\sigma)} \Big/ \sum_{\sigma} S_{ij,n,r}^{(\sigma)} \tag{2.20}$$

where $S_{ij,n,r}$ is the total dimensionless concentration of ribosomes with P sites occupying state *n* regardless of state, and $\delta_{n,r}^{(\sigma)}$ is the fraction of ribosomes at codon *n* occupying state σ.

Initiation and elongation limited kinetics

Under initiation and elongation limited kinetics ribosomes are uniformly distributed with respect to sequence position throughout the ensemble of mRNAs, and as the polysome size increases the concentration of ribosomal P sites at each codon increases (Figure 2.6). Most ribosomes at each position occupy the state existing prior to Ef-Tu:GDP release, state 5 $\left(S_{ij,n,r}^{(5)}\right)$, and the distribution of states is identical for all ribosomes at each codon (Figure 2.7) for every polysome size. Uniform ribosome distributions are expected under these conditions because once the ribosome binds to the initiation site, movement along the length of the mRNA is relatively unrestricted. Hence, the progress of each elongation cycle is restricted only by the relative magnitudes of the reaction rate constants of the elementary steps. Consequently most bound ribosomes at each codon occupy state 5 as expected because the control coefficient corresponding to the reaction rate constant for Ef-Tu:GDP release $\left(C_{k_5}^{v}\right)$ is the highest of all the control coefficients corresponding to the elongation cycle intermediate step rate constants in the initiation and elongation limited regimes. In addition, this result demonstrates that our

assumption about the total Ef-G concentration free to participate in translation being approximately equal to the total cellular Ef-G concentration is reasonable (this assumption is described in detail in APPENDIX C). Ef-G is not bound to the ribosome at state 5. Therefore, because most ribosomes participating in translation throughout the initiation and elongation limited regimes occupy state 5, and because the initiation and elongation limited regimes comprise almost the entire range of polysome sizes, the cellular Ef-G concentration bound to translating ribosmes is negligibly small.

Figure 2.6: Ribosome distributions under initiation and elongation limited kinetics: Ribosome distributions with respect to codon sequence position for ρ=0.0033 (solid line), ρ=0.50 (dashed line), and ρ=0.95 (dotted line).

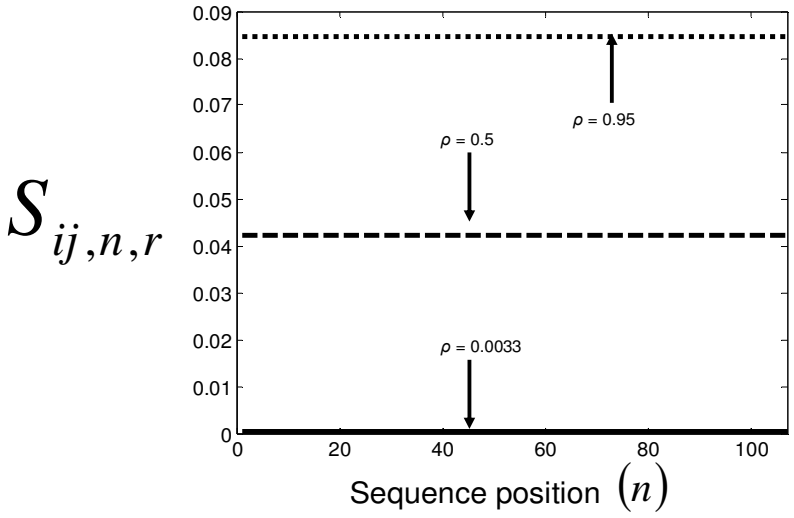
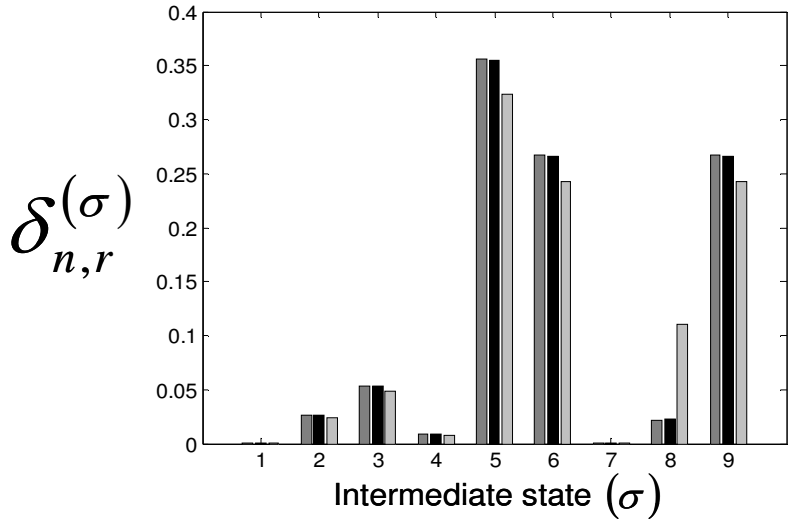


Figure 2.7: Ribosome distributions under initiation and elongation limited kinetics: Ribosome distributions with respect to intermediate elongation cycle state for ρ=0.0033 (dark grey), ρ=0.50 (black), and ρ=0.95 (light grey) for all codons along the length of the sequence.



As polysome size increases, ribosomal crowding develops. With ribosomal crowding it is more likely that the progress of an elongation cycle at position *n* is limited by the presence of the tail end of a downstream ribosome occupying the *n+7* codon position. Presence of the downstream ribosome prevents translocation of the ribosome at position *n*. Hence, as polysome size increases more ribosomes at each position occupy state 8 $\left(S_{ij,n,r}^{(8)}\right)$, the intermediate that exists prior to translocation (Figure 2.7). We observe that $\rho = 0.95$, the ribosomal fractional coverage corresponding to maximum translation rate, is the maximum ribosomal fractional coverage at which the ribosome distribution with respect to codon position is uniform (Figure 2.6). As the kinetics transition from being elongation limited to termination limited in the ribosomal fractional coverage range $0.95 < \rho < 1$, we observe that the ribosome distribution with respect to codon position is not uniform, and that the distribution of states at each codon is not the same for all codons along the length of the sequence (Figures 2.8 and 2.9). As ribosomal fractional

coverage increases from $\rho = 0.95$ ribosome movement becomes restricted at the 3' end of the mRNA, causing the ribosomes to queue along the length of the chain (Figure 2.8). However, the ribosomal crowding is not high enough in the range of polysome sizes where kinetics transition from being elongation to termination limited to cause ribosomes to queue along the entire length of the mRNA, so the ribosome distribution near the 5' end of the mRNA resembles a uniform distribution (Figure 2.8). Consequently, near the 5' end of the mRNA the distribution of states at each codon is similar to that in Figure 2.7, with state 5 ribosomal occupancy being the highest (Figure 2.9). However, near the 3' end of the mRNA ribosomal queuing occurs along the length of the sequence, causing state 8 ribosomal occupancy to increase sharply at positions spaced one ribosome length apart (Figure 2.9). As a result the state 5 ribosomal occupancy is similar to the state 5 occupancy under initiation and elongation limited conditions, but decreases sharply at these positions (Figure 2.9). The remaining state occupancies (not shown) are also similar to their respective occupancies under initiation and elongation limited conditions and decrease sharply at positions where state 8 occupancy is maximal.

Figure 2.8: Ribosome distributions under initiation and elongation limited kinetics: Ribosome distribution with respect to codon sequence position for ρ=0.976.

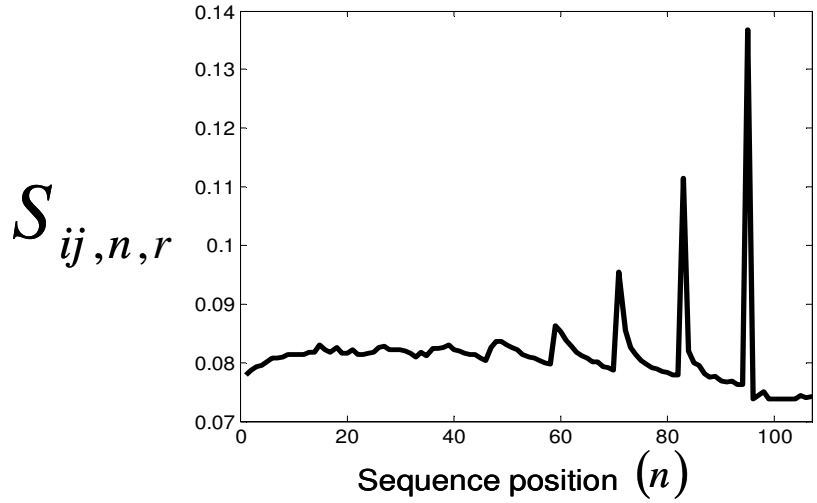
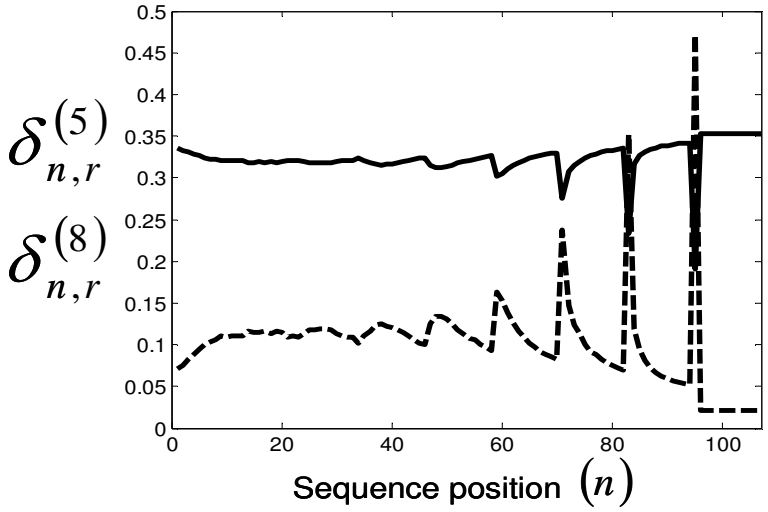


Figure 2.9: Ribosome distributions under initiation and elongation limited kinetics: Fraction of ribosomes at each codon position occupying state 5 (solid line), and the fraction of ribosomes at each codon position occupying state 8 (dashed line) for ρ=0.976.



Termination limited kinetics

Under termination limited kinetics ribosome movement is strongly restricted at the 3' end of the mRNA, causing the ribosomes to queue along the entire length of the chain. Consequently, almost all bound ribosomes have P sites spaced one ribosome length apart (Figure 2.10). Also,

under these conditions ribosomal occupancy of state 8 $\left(S_{ij,n,r}^{(8)}\right)$ is maximal, with almost all bound ribosomes occupying state 8 at each codon where ribosomal P site occupancy is high (Figure 2.11). The fraction of ribosomes at each codon occupying state 5 is slightly lower under termination limited conditions than the state 5 occupancy under initiation and elongation limited conditions, and approaches zero at each position where ribosomal state 8 occupancy is maximal (Figure 2.11). Ribosomal occupancies of the remaining states (not shown) are also slightly lower under these conditions than their respective occupancies under initiation and elongation limited conditions, and also approach zero at each position where ribosomal state 8 occupancy is maximal. The progress of an elongation cycle at position *n* is more strongly limited by the presence of the tail end of the proceeding ribosome occupying the *n+7* codon position under termination limited conditions than under elongation limited conditions, resulting in most bound ribosomes occupying state 8.

Figure 2.10: Ribosome distributions under termination limited kinetics for ρ=1: Ribosome distribution with respect to codon sequence position.

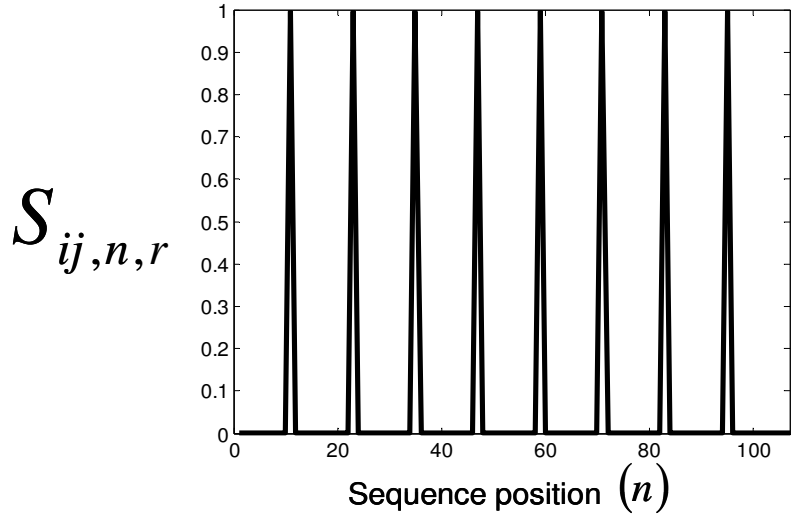
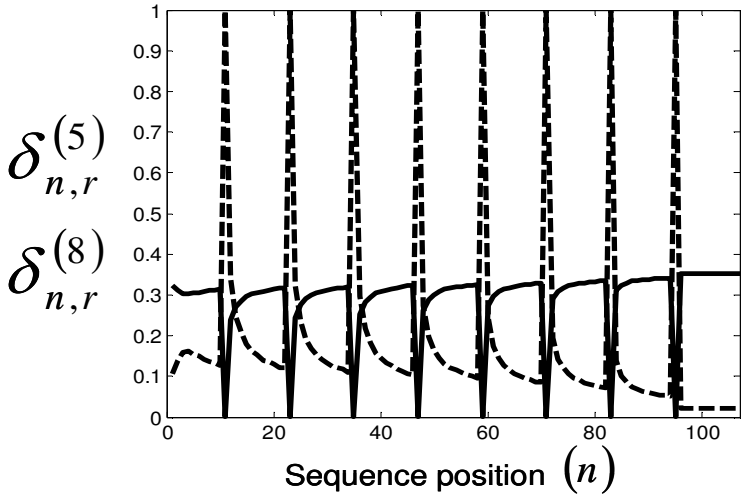


Figure 2.11: Ribosome distributions under termination limited kinetics for ρ=1: Fraction of ribosomes at each codon position occupying state 5 (solid line), and the fraction of ribosomes occupying state 8 (dashed line).



Effects of ribosome, ternary complex, and Ef-G concentrations on translation rate

We observe that total free ribosome, ternary complex, and Ef-G:GTP concentrations do not limit translation rate of a single gene by examining respective conservation equations (equations 2.2 – 2.4) and control coefficients. Free ribosome $\left(R^{(f)}\right)$, ternary complex $\left(T_k^{(f)}\right)$, Ef-G:GTP $\left(G^{(f)}\right)$, and ribosomal state concentrations $\left(S_{ij,n,r}^{(\sigma)}\right)$ are made dimensionless by scaling with total ribosome $\left(R^{(t)}\right)$, ternary complex $\left(T_k^{(t)}\right)$, Ef-G:GTP $\left(G^{(t)}\right)$, and mRNA $\left(M_r\right)$ concentrations, respectively, allowing the conservation equations for a single mRNA species r to be rewritten as:

$$1 = \mu_r \sum_{n=1}^{N_r-1} \sum_{\sigma} \left(\tilde{S}_{ij,n,r}^{(\sigma)} + \tilde{S}_r^T \right) + \tilde{R}^{(f)} \tag{2.21}$$

$$1 = \lambda_{k,r} \sum_{n=1}^{N_r-1} \left\{ \left(\tilde{S}_{kk,n,r}^{(\sigma)} + 2\sum_{\sigma=2}^{9} \tilde{S}_{kk,n,r}^{(\sigma)} \right) + \sum_{\sigma=2}^{9} \tilde{S}_{ik,n,r}^{(\sigma)} + \sum_{\sigma} \tilde{S}_{kj,n,r}^{(\sigma)} \right\} + \tilde{T}_k^{(f)} \qquad , k \in K \tag{2.22}$$

$$1 = \varphi_r \sum_{n=1}^{N_r-1} \sum_{\sigma=8}^{9} \tilde{S}_{ij,n,r}^{(\sigma)} + \tilde{G}^{(f)} \tag{2.23}$$

where $\tilde{R}^{(f)}$, $\tilde{T}_k^{(f)}$, $\tilde{G}^{(f)}$, and $\tilde{S}_{ij,n,r}^{(\sigma)}$, are the dimensionless free ribosome, ternary complex, Ef-G:GTP, and ribosome state concentrations, respectively, and μ_r, $\lambda_{k,r}$, and φ_r are dimensionless quantities with:

$$\mu_r = M_r / R^{(t)} \tag{2.24}$$

$$\lambda_{k,r} = M_r / T_k^{(t)} \tag{2.25}$$

$$\varphi_r = M_r / G^{(t)} \tag{2.26}$$

Based on the cellular concentrations calculated previously for $M_r, R^{(t)}$, $T_k^{(t)}$, and $G^{(t)}$ we determine that $\mu_r = 2.7 \times 10^{-4}$, $\lambda_{k,r} = \left[6.9 \times 10^{-5}, 4.3 \times 10^{-3}\right]$, and $\varphi_r = 5.3 \times 10^{-5}$. Because the concentration of a single mRNA species is low relative to the concentrations of the other available translational components, the ribosome, ternary complex, and Ef-G concentrations sequestered in the ribosomal states on a single mRNA species are low relative to their respective total available concentrations. This finding demonstrates that the coupling that exists between ribosomal states on a single mRNA species due to shared translational resources is low. Hence, when considering the dimensional conservation equations for ribosomes, ternary complexes, and Ef-G:GTP (equations 2.2 – 2.4) $R^{(f)}$, $T_k^{(f)}$, and $G^{(f)}$ are not strong functions of ribosomal state concentrations.

To determine if the total concentrations of the available translational machinery limit protein synthesis rate we consider the flux control coefficients with respect to total ribosome, ternary complex, and Ef-G:GTP concentrations. We observe that the flux control coefficients with respect to total ternary complex and Ef-G:GTP concentrations are approximately zero under

initiation, elongation, and termination limited kinetics, so the total concentrations of these translational components do not limit protein synthesis rate. The flux control coefficient with respect to total ribosome concentration is equal to one at low polysome size, and decreases and approaches zero with increasing polysome size. Hence, the total ribosome concentration can significantly impact protein synthesis rate at low polysome size.

2.3.3 Comparison between the ZH and MG-HR models

To simplify the ZH model and further understand the contributions of each elongation cycle intermediate step to the overall elongation cycle kinetics, we developed a formulation that is an equivalent description of steady state translation kinetics to the ZH model description of steady state translation kinetics. By setting the time derivatives equal to zero in equations (2.7) – (2.17) and solving for states 1 to 8 in terms of state 9, the flux at position *n* can be written as:

$$V_{ij,n,r} = k_{E,n,r}^{eff} S_{ij,n,r} U_{n,r} M_r \tag{2.27}$$

where

$$k_{E,n,r}^{eff} = \frac{1}{U_{n,r}\left(\alpha_{1,j} + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7 + \alpha_9\right) + \alpha_8} \tag{2.28}$$

with descriptions of the terms included in Table 2.3.

Table 2.3: Dimensionless parameters and reduced model terms

Parameter	Expression	Elongation cycle intermediate step	Magnitude
$\alpha_{1,j}$	$\left[\frac{(k_{-1}+k_2)(k_{-2}+k_3)}{k_2k_3}-\frac{k_{-2}}{k_3}\right]\frac{1}{k_1T_j^{(f)}}$	Codon independent binding of the ternary complex	6×10^{-4} – 0.04
α_2	$\frac{k_{-2}+k_3}{k_2k_3}$	Codon dependent binding	0.005
α_3	$1/k_3$	GTP hydrolysis	0.01
α_4	$1/k_4$	Ef-G:GDP position change on ribosome	0.0015
α_5	$1/k_5$	Ef-G:GDP release	0.067
α_6	$1/k_6$	A site tRNA accommodation	0.05
α_7	$\frac{k_{-7}+k_8}{k_7k_8G^{(f)}}$	Ef-G:GTP binding	3.5×10^{-4}
α_8	$1/k_8$	Translocation	0.004
α_9	$1/k_9$	E site tRNA release	0.05

The magnitudes of the terms in the expression for the effective elongation rate constant (equation 2.28) reflect the influence each elongation cycle intermediate step has over the overall kinetics of the elongation cycle. Similar to the results from the control analysis, the reduced model identifies

Ef-Tu:GDP release, accommodation, and E site tRNA release as the elongation cycle intermediate steps that have the greatest influence over elongation cycle kinetics.

The conservation equations for ribosomes and codons are expressed by equations (2.2), (2.5), and (2.6), and conditional probability of ribosome translocation is expressed by equation (2.1), similar to the full model. We have previously identified low flux control coefficients with respect to total ternary complex and Ef-G:GTP concentrations, and therefore we assume in the reduced model that free ternary complex and Ef-G:GTP concentrations are fixed to their respective total available concentrations, leading to $T_k^{(f)} = T_k^{(t)}$, and $G^{(f)} = G^{(t)}$.

With the reduced version of our model the elongation cycle at a given codon is expressed in terms of a single flux (equation 2.27) whose terms map exactly to the MG-HR elongation flux expression (equation A4). Both flux expressions at a given codon position depend on the elongation rate constant $k_{E,n,r}^{eff}$ (ZH model) and k_E (MG-HR model), the probability that the codon is occupied either by the P site (ZH model), $S_{ij,n,r}$, or front of the ribosome (MG-HR model), $x_{n,r}$, and the conditional probability governing ribosome movement $U_{n,r}$ (ZH model) and $W_{n+1,r}$ (MG-HR model).

We used the MG-HR mechanistic framework to determine the specific protein production rate and the initiation, elongation, and termination control coefficients as functions of polysome size (Figures 2.12 and 2.13). For the elongation rate constant $\left(k_E\right)$ we used a value of $5.26 s^{-1}$, which we refer to as the characteristic effective elongation rate constant. This value is equal to the effective elongation rate constant discussed previously (equation 2.28) evaluated at $U_{n,r} = 1$ and at the average free ternary complex concentration of $6.31 \mu M$. The MG-HR model predicts

that the maximum translation rate of 29 amino acids/s occurs at $\rho = 0.77$ (Figure 2.12). Also, the MG-HR model predicts initiation limited kinetics for $\rho < 0.45$, elongation limited kinetics for $0.45 < \rho < 0.95$, and termination limited kinetics for $\rho > 0.95$ (Figure 2.13).

Figure 2.12: Relationships between translation properties and polysome size using the MG-HR model: Specific protein production rate as a function of polysome size.

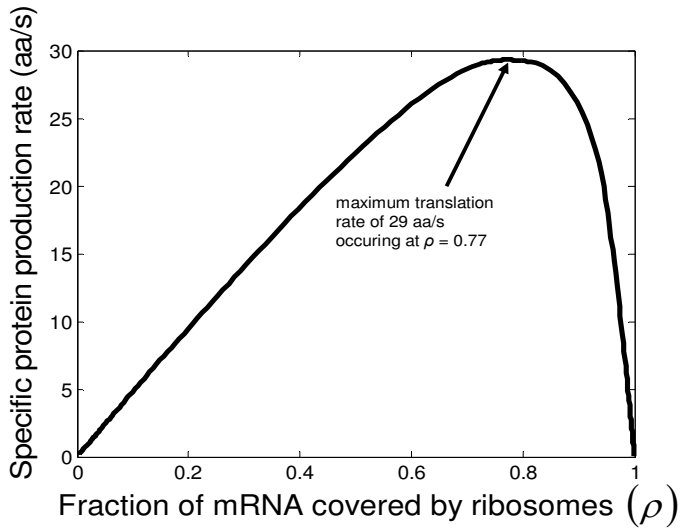
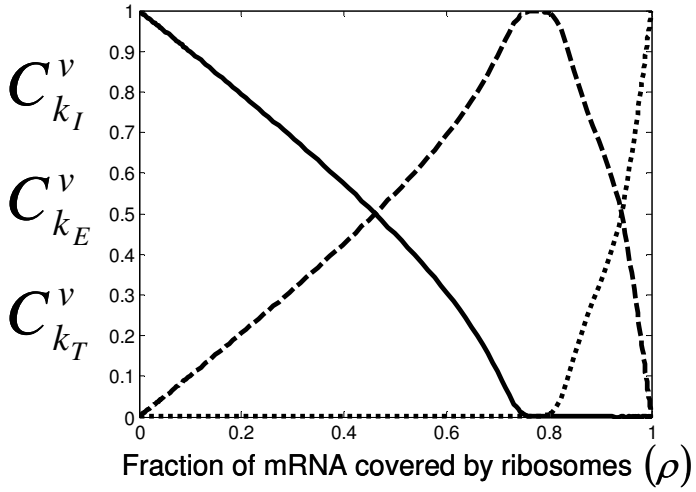


Figure 2.13: Relationships between translation properties and polysome size using the MG-HR model: Initiation (solid line), elongation (dashed line), and termination (dotted line) control coefficients as functions of polysome size.



2.3.4 Polysome self-organization

We observe two main differences between the ZH model and the MG-HR model in predicting translational behavior with respect to polysome size:

(i) The ZH model predicts a higher maximum translation rate than the MG-HR model predicts.

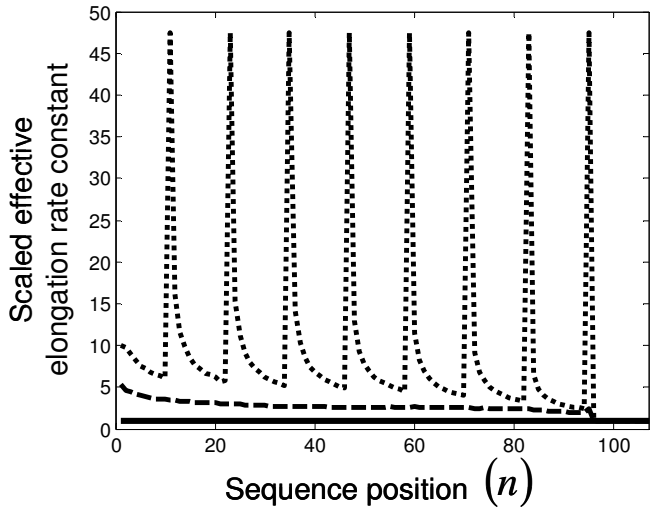
(ii) The ZH model predicts that the maximum translation rate occurs at a ribosomal fractional coverage that is higher than what the MG-HR model predicts.

In this section these differences are addressed by discussing how ribosome occupancies with respect to state and sequence position lead to varying configurations of effective elongation rate constant magnitudes that are specific to different polysome sizes. These results are used to investigate how self organization of bound ribosomes with respect to the elongation cycle state and position occupancy affects the relationship between translational behavior and polysome size.

To investigate differences between the results of the ZH model and the MG-HR model we scale the effective elongation rate constants by dividing them by the characteristic effective elongation rate constant, and we compare the scaled effective elongation rate constants derived from our reduced model (equation 2.28) under initiation, elongation, and termination limited kinetics (Figure 2.14). We observe that under initiation limited conditions the effective elongation rate constants along the length of the sequence are approximately equal to the characteristic effective elongation rate constant. Under elongation limited kinetics the reduced model predicts effective elongation constants approximately equal to five times the characteristic effective elongation rate constant and protein synthesis rates are driven to maximum levels. Under termination limited conditions the effective elongation constants are approximately equal

to 48 times the characteristic effective elongation rate constant at positions spaced one ribosome length apart due to ribosomal queuing along the length of the mRNA, while the rest of the effective elongation rate constants vary between ten and five times as much as the characteristic effective elongation rate constants.

Figure 2.14: Scaled effective elongation rate constants with respect to codon sequence position under initiation (ρ=0.0033, solid line), elongation (ρ=0.95, dashed line), and termination (ρ=1, dotted line) limited conditions.

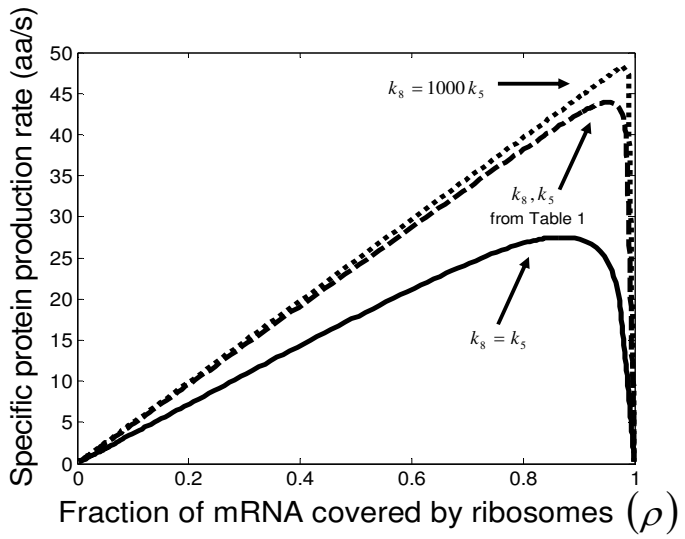


In the reduced model the conditional probability $(U_{n,r})$ is in the denominator of the expression for the effective elongation rate constant, $k_{E,n,r}^{eff}$ (equation 2.28), suggesting that $k_{E,n,r}^{eff}$ increases as $U_{n,r}$ decreases due to crowding. At low polysome size, ribosomal crowding on the mRNA is minimal, so $U_{n,r} \approx 1$ and $k_{E,n,r}^{eff}$ is approximately equal to the characteristic effective elongation rate constant. As polysome size increases, ribosomal crowding on the mRNA increases; $U_{n,r}$ decreases causing $k_{E,n,r}^{eff}$ to increase. At high polysome size ribosomal crowding on the mRNA is maximal and $U_{n,r} \approx 0$, causing $k_{E,n,r}^{eff}$ to approach the magnitude of the translocation rate

constant, k_8. Hence, at high polysome size the effective elongation rate constants at positions spaced one ribosome length apart are approximately equal to k_8. The maximum protein synthesis rate occurs at the polysome size corresponding to the set of effective elongation rate constants that are maximal at each sequence position, while still uniformly distributed along the length of the mRNA.

We observe that the effective elongation rate constants along the length of the mRNA transition as polysome size increases from having magnitudes driving high translation rates to magnitudes that decrease translation rates. To understand how this relationship develops, we investigated effects of relative values of the elongation cycle intermediate rate constants on the magnitudes of the effective elongation rate constants. Furthermore, we investigated how the altered effective elongation rate constant magnitudes impact the relationship between translation rate and polysome size (Figure 2.15). Because Ef-Tu:GDP release was found to be the elongation cycle intermediate step to contribute the most to the control of the elongation phase over translation rate (Figure 2.5) and to have the highest contribution in the expression for the effective elongation rate constant (equation 2.28), the rate constant corresponding to translocation, k_8, was manipulated relative to the rate constant corresponding to Ef-Tu:GDP release, k_5. As k_8 increases relative to k_5, the maximum specific protein production rate increases and the ribosomal fractional coverage where the maximum specific protein production rate occurs at also increases (Figure 2.15). Also, the relation between translation rate and polysome size does not change when $k_8 > 1000k_5$, so the highest possible maximum protein synthesis rate is equal to 48 amino acids/s and corresponds to a ribosomal fractional coverage of 0.98.

Figure 2.15: Relationships between translation rate and polysome size with k_8 manipulated relative to k_5. Specific protein production rate as a function of polysome size for $k_8 = k_5$ (solid line), the k_8 and k_5 values from Table 1 (dashed line), and $k_8 = 1000k_5$ (dotted line).



Hence, we observe that the magnitudes of the effective elongation rate constants depend both on the level of crowding in the sequence and on how fast the ribosome can be transferred to the next codon in the sequence. As polysome size increases, the number of bound ribosomes to the mRNA increases, and the concentration of the state existing prior to translocation, state 8, increases (Figure 2.7). This behavior is due to the increased likelihood that the progress of an elongation cycle at position *n* is limited by the presence of the tail of a downstream ribosome occupying the *n+7* codon position as previously discussed. The higher the translocation rate constant (k_8), the more ribosomes can be bound to the mRNA without this limitation occurring. As a result, as k_8 increases, the maximum translation rate increases and occurs at increasing polysome sizes because the polysome size can be higher before ribosomal steric effects become significant and limit translation rate. These results demonstrate that it is the configuration of effective elongation rate constant magnitudes with respect to sequence position that lead to optimum translation rate at a specific polysome size. As expected, when we substitute the set of

effective elongation rate constants at each polysome size for the elongation rate constants in the MG-HR model, the relationship between protein synthesis rate and ribosomal fractional coverage that is observed is the same as that observed with the ZH model (Figure 2.3).

2.4 Discussion

We presented a theoretical analysis of the translation mechanism accounting for the initiation, elongation, and termination phases. Our model of the elongation phase is sequence specific and includes all the intermediate steps of the elongation cycles taking place at every codon along the length of the mRNA. Consideration of protein synthesis kinetics in the context of polysome size provides insights into quantifying the systemic contributions of the translational components and kinetic parameters to the translational output of genes. As polysome size increases, the ribosomal occupancy with respect to both elongation cycle intermediate and position on the mRNA changes (Figures 2.6, 2.7, 2.8, 2.9, 2.10, and 2.11). These changes affect the protein synthesis rate (Figure 2.3) and the extent to which the initiation, elongation, and termination kinetic parameters limit translation rate (Figure 2.4).

These results suggest that polysomes self organize with respect to ribosomal state and sequence position occupancies to achieve maximum translation rates. The relative values of the kinetic parameters corresponding to the intermediate steps of the elongation cycle are such that the polysome size can become very high before ribosomal crowding on the mRNA limits translation rate. We observe that the maximum protein synthesis rate of 44 amino acids/s occurs at a ribosomal fractional coverage of 0.95 (Figure 2.3). Also, by changing the relative magnitudes of the elongation cycle kinetic parameters we observe that the increases in maximum possible translation rate and corresponding ribosomal fractional coverage are not dramatic

(Figure 2.15), which suggests that the relative reported values for the elongation kinetic parameters are near optimal.

It is important to note that the same set of reaction rate constants were used for the elongation cycle intermediate steps at every codon along the length of the sequence. However, due to the interplay between the level of ribosomal crowding on the mRNA and the contributions of the intermediate elongation cycle steps to the kinetics of the overall elongation cycle, the effective elongation rate constants change with polysome size and sequence position (Figure 2.14). The magnitudes of the effective elongation rate constants at a given polysome size determine the translation rate at that polysome size. Additionally, similar translational behavior with respect to polysome size was observed when the relative values of the elongation cycle kinetic parameters were altered (Figure 2.15). Hence, our findings suggest that the behavior we observe is intrinsic to the translation mechanism. In future studies it will be important to incorporate codon specific kinetic parameters into our model to investigate how translational behavior is affected.

The elongation cycle reaction rate constants used in our model were determined experimentally *in vitro*, so although we hypothesize that our findings are intrinsic to the translation mechanism, we do not expect that the numerical results we observe are exactly what would be observed *in vivo*. However, the translation rates predicted by our model compare well with those of *in vivo* results. According to (31), *E. coli* has a bulk translation rate of 18 amino acids/s. In recent work (33) the bulk initiation rate constant corresponding to this translation rate was determined to be equal to 6×10^{4} M^{-1} s^{-1}. Our model predicts that this initiation rate constant corresponds to a translation rate of 22 amino acids/s, which is a commonly reported value (31). Also, protein synthesis rates as high as 40 amino acids/s have been observed (32), and our model

predicts a maximum protein synthesis rate of 44 amino acids/s (Figure 2.3). Hence, this study appears to predict extremely well the translation properties *in vivo*.

Because the translational behavior for a given ribosomal fractional coverage corresponds to a unique pair of initiation and termination rate constants, we do not necessarily expect a single mRNA species to be able to take on the full ribosomal fractional coverage range. However, a recent study in *S. cerevisiae* (15) demonstrates that different mRNA species take on different polysome sizes, and the full ribosomal fractional coverage range is observed. Hence, our mechanistic framework can be used to provide insight into the translation of proteins of high and low expression levels and, subsequently, how cellular resources are allocated for the synthesis of different proteins.

2.5 Conclusions

In this chapter we developed a deterministic, sequence specific kinetic model of the translational machinery that accounts for all the elementary steps of the translation mechanism. We also performed a sensitivity analysis in order to determine the effects of the kinetic parameters and concentrations of the translational components on the protein synthesis rate. Moreover, we developed a reduced formulation utilizing effective elongation rate constants that is an equivalent description of steady state translation kinetics to the ZH model description of steady state translation kinetics. We determined the following:

(i) As polysome size increases translation rate increases, reaches a maximum, and then decreases.

(ii) Translation kinetics are either initiation or elongation limited for almost the entire range of polysome sizes, and are termination limited at very high polysome sizes.

(iii) The elongation cycle intermediate step with the most control over a given elongation cycle is Ef-Tu:GDP release, i.e., for the elongation cycle at a given codon, Ef-Tu:GDP release is the most rate limiting intermediate step.

(iv) Ribosome distributions with respect to codon position in the initiation and elongation limited regimes are uniformly distributed. Ribosome distributions in the termination limited regime are composed of ribosomal P sites spaced one ribosome length apart due to queuing that occurs along the length of the mRNA.

(v) In the initiation and elongation limited regimes ribosomes primarily occupy the state existing prior to Ef-Tu:GDP release. In the termination limited regime ribosomes primarily occupy the state existing prior to translocation.

(vi) The maximum protein synthesis rate occurs at the polysome size corresponding to the set of effective elongation rate constants that are maximal at each sequence position, while still uniformly distributed along the length of the mRNA. The configuration of effective elongation rate constants depends on the complex interplay between ribosomal occupancy of elongation cycle intermediate states and ribosome distributions with respect to codon position along the length of the mRNA.

Chapter 3: Effects of codon distributions and tRNA competition on protein translation

3.1 Introduction

A finding from our model in Chapter 2 was that tRNA concentrations have almost no impact on protein synthesis rate. However, experimental evidence suggests tRNA concentrations are significant to translation kinetics. The work by Ikemura (34) shows a correlation between tRNA abundances and codon frequencies. Other work demonstrates that synonymous codons (different codons coding for the same amino acid) are not translated at the same rate (10), with higher translation rates for more abundant or major codons (35). Given the difference between experimental results and those determined from our computational studies in Chapter 2, it is important to note that a simplifying assumption made in our model in Chapter 2 is that only ternary complexes that recognize the A site codon can bind to the ribosome. In reality, ternary complexes initially bind nonspecifically to the ribosomal A site, which means that both ternary complexes recognizing and not recognizing the ribosomal A site codon can bind to the ribosome in the first intermediate step of the elongation cycle of each codon. The experimentally observed importance of tRNA concentration to protein synthesis kinetics, coupled with our observation that tRNA concentrations are not scarce enough to modulate translation rate, motivates questions about the role the competition between ternary complexes for ribosomal A site binding plays in protein synthesis kinetics.

Hence, in this chapter we expand our mechanistic framework to account for ternary complex competitive binding to the ribosomal A site. We also expand our sensitivity analysis to make it codon specific, meaning that we account for the contribution of kinetic parameters and translational component concentrations of each codon on the overall protein synthesis rate. We find that our expanded mechanistic framework predicts lower protein synthesis rates than our

framework in Chapter 2. Our sensitivity analysis predicts that at low polysome sizes the codons near the 5' end of the mRNA control protein synthesis rate, at intermediate polysome sizes different configurations of codons along the length of the mRNA control protein synthesis rate, and at high polysome sizes the codons near the 3' end of the mRNA control protein synthesis rate. Moreover, our sensitivity analysis identifies the competitive, non-specific binding of the tRNAs to the ribosomal A site as rate limiting to the elongation cycle for every codon. By introducing our model from Chapter 2 and our model from this chapter in terms of the Michaelis – Menten kinetic framework, we determine that these results are due to the tRNAs that do not recognize the ribosomal A site codon acting as competitive inhibitors to the tRNAs that do recognize the ribosomal A site codon. We also observe that the relative position of codons along the mRNA determines the optimal protein synthesis rate, and that the translation rates of mRNAs are controlled by segments of rate limiting codons that are sequence specific.

3.2 Methods

3.2.1 Mathematical model

In this section we introduce a mechanistic framework that incorporates the kinetics of all the intermediate steps of the translation elongation cycle occurring at a given codon in a single expression that is applied to the studies in this chapter. A summary of the assumptions made in this formulation, along with descriptions of the variables and parameters can be found in APPENDIX D. A detailed description of this model can be found in Chapter 2.

The initiation rate is described as follows:

$$V_{I,r} = k_{I,r} R^{(f)} C^{(f)}_{n+6,r} \qquad ,n = 1 \qquad (3.1)$$

where $k_{I,r}$ is the initiation rate constant of mRNA r, $R^{(f)}$ is the free ribosome concentration, and $C^{(f)}_{n+6,r}$ is the concentration of mRNA *r* having a free ribosomal binding site.

The elongation rate at codon *n* along the length of the mRNA species *r* is described as follows:

$$V_{ij,n,r} = k^{eff}_{E,n,r} S_{ij,n,r} U_{n,r} M_r \qquad , n \in [1, N_r - 1] \qquad (3.2)$$

where the subscript *i* denotes the P site codon species, the subscript *j* denotes the A site codon species, the subscript *n* denotes the position of the ribosomal P site codon, $k^{eff}_{E,n,r}$ is the effective elongation rate constant, and $S_{ij,n,r}$ is the fraction of the mRNA species *r* concentration with codon position *n* occupied by the P site of a translating ribosome. Ribosome movement along the length of the sequence is dependent on the conditional probability that the codon adjacent to the codon occupied by the front of the ribosome is free given that the previous codon is occupied by the front of the ribosome, $U_{n,r}$, and M_r is the concentration of mRNA *r*.

The effective elongation rate constant at codon position *n*, $k^{eff}_{E,n,r}$ (equation 2.28), is comprised of terms representing the kinetics of each of the translation elongation cycle intermediate steps occurring at that codon, and these terms depend on the reaction rate constants corresponding to the elongation cycle intermediate steps (a detailed discussion of the effective elongation rate constant can be found in Chapter 2):

$$k^{eff}_{E,n,r} = \frac{1}{U_{n,r}\left(\alpha_{1,j} + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7 + \alpha_9\right) + \alpha_8} \qquad (2.28)$$

A summary of the effective elongation rate constant terms is included in Table 2.3. In this work we investigate effects of ternary complex competition for ribosomal A site binding, so it is important to note that the first effective elongation rate constant term, $\alpha_{1,j}$, corresponds to the

reversible, codon independent binding of the ternary complex to the ribosomal A site codon species *j*. The expression for $\alpha_{1,j}$ is as follows:

$$\alpha_{1,j} = \left[\frac{(k_{-1} + k_2)(k_{-2} + k_3)}{k_2 k_3} - \frac{k_{-2}}{k_3} \right] \frac{1}{k_1 T_j^{(f)}} \tag{3.3}$$

where k_1, k_{-1}, k_2, k_{-2}, and k_3 are reaction rate constants corresponding to ternary complex binding (Table 2.1). Free ternary complex concentrations $\left(T_k^{(f)}\right)$ are of species *k*, with $k \in K$, where *K* is the set of ternary complex species. Hence, $T_j^{(f)}$ is the free ternary complex concentration of species *j* recognizing A site codon species *j*. Equation (3.3) was derived assuming that only ternary complexes recognizing the ribosomal A site codon bind to the ribosome during nonspecific binding. In reality, all ternary complexes species can bind to the ribosome during the codon independent binding intermediate step, regardless of whether or not they recognize the A site codon. Hence, in this work we relax our original assumption by allowing all ternary complex species to be able to bind to the ribosomal A site at this step, yielding the following expression for the nonspecific ternary complex binding term of the effective elongation rate constant:

$$\alpha_{1,j}^{T} = \left[\frac{(k_{-1} + k_2)(k_{-2} + k_3)}{k_2 k_3} - \frac{k_{-2}}{k_3} \right] \frac{1}{k_1 T_j^{(f)}} \left(1 + K_1 \sum_{k \neq j} T_k^{(f)} \right) \tag{3.4}$$

where the term $\left(1 + K_1 \sum_{k \neq j} T_k^{(f)} \right)$ accounts for ternary complex competitive binding and $K_1 = k_1 / k_{-1}$. By replacing $\alpha_{1,j}$ with $\alpha_{1,j}^{T}$ in the expression for the effective elongation rate constant (equation 2.28), we define $k_{E,n,r}^{eff,T}$ to be the effective elongation rate constant accounting for ternary complex competitive binding:

$$k_{E,n,r}^{eff,T} = \frac{1}{U_{n,r}\left(\alpha_{1,j}^{T} + \alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7 + \alpha_9\right) + \alpha_8} \tag{3.5}$$

The termination rate is described as follows:

$$V_{T,r} = k_{T,r} S_r^T \tag{3.6}$$

where $k_{T,r}$ is the termination rate constant of mRNA *r* and S_r^T is the total concentration of ribosomes on mRNA *r* that have completed the translation elongation phase.

The dynamics describing the transition between the states of the elongation phase are as follows:

$$\frac{dS_{ij,n,r}}{dt} = V_{I,r} - V_{E,n,r} \qquad , n = 1 \tag{3.7}$$

$$\frac{dS_{ij,n,r}}{dt} = V_{E,n-1,r} - V_{E,n,r} \qquad , n \in \left[2, N_r - 1\right] \tag{3.8}$$

$$\frac{dS_r^T}{dt} = V_{E,n,r} - V_{T,r} \qquad , n = N_r - 1 \tag{3.9}$$

The total ribosome and codon concentrations are expressed by equations (3.10) and (3.11), respectively:

$$R^{(t)} = \sum_r \sum_{n=1}^{N_r - 1} \left(S_{ij,n,r} + S_r^T\right) + R^{(f)} \tag{3.10}$$

$$M_r = \begin{cases} \sum_{n}^{n+6} S_{ij,n,r} + C_{n,r}^{(f)} & , n = 1 \\ \sum_{n}^{n+6} \left(S_{ij,n,r}\right) + C_{n,r}^{(f)} & , n \in \left[2, N_r - \left(L+1\right)\right] \end{cases} \tag{3.11}$$

where $C_{n,r}^{(f)}$ is the concentration of free codons at position *n* of mRNA *r*.

3.2.2 Sensitivity analysis

In this section we introduce the sensitivity analysis as it pertains to the studies in this chapter. A detailed discussion of the sensitivity analysis is included in APPENDIX B. We

investigate the effects of elongation cycle kinetics at each codon along the length of the mRNA on the steady state protein synthesis rate by examining the flux control coefficients, C_p^v, which are defined as fractional flux changes with respect to fractional input parameter changes (25). Similar to the Summation Theorem (25), we can show that the sum of the control coefficients with respect to the reaction rate constants for an mRNA species that is not competing for translational resources with other mRNA species is equal to one:

$$C_{k_{I,r}}^v + C_{k_{E,r}}^v + C_{k_{T,r}}^v = 1 \tag{3.12}$$

where $C_{k_{I,r}}^v$ and $C_{k_{T,r}}^v$ are the fractional changes in flux with respect to fractional changes in the initiation and termination rate constants, respectively. The control coefficient $C_{k_{E,r}}^v$ is the fractional change in flux with respect to the simultaneous fractional change in the elongation rate constant, $k_{E,n,r}^{eff}$, of every codon expressed as:

$$C_{k_{E,r}}^v = \sum_{n=1}^{N_r-1} C_{k_{E,n,r}^{eff}}^v \tag{3.13}$$

where $C_{k_{E,n,r}^{eff}}^v$ is the control coefficient corresponding to the elongation step occurring at the codon at position *n* on the mRNA and is the fractional change in flux with respect to the fractional change in the effective elongation rate constant at position *n*. The control coefficient with respect to the effective elongation rate constant at codon position *n*, $C_{k_{E,n,r}^{eff}}^v$, is equal to the sum of the control coefficients with respect to the reaction rate constants of the elongation cycle intermediate steps at codon position *n*, where:

$$\begin{aligned} C_{k_{E,n,r}^{eff}}^v &= C_{k_1,n}^v + C_{k_{-1},n}^v + C_{k_2,n}^v + C_{k_{-2},n}^v + C_{k_3,n}^v + C_{k_4,n}^v + \\ & C_{k_5,n}^v + C_{k_6,n}^v + C_{k_7,n}^v + C_{k_{-7},n}^v + C_{k_8,n}^v + C_{k_9,n}^v \end{aligned} \tag{3.14}$$

Equations (3.13) and (3.14) are also applied to determine control coefficients of the elongation steps along the length of the mRNA with under ternary complex competitive binding conditions, $C^{v}_{k^{eff,T}_{E,n,r}}$.

3.3 Computational studies

We utilize our mathematical model of protein synthesis and the sensitivity analysis to investigate the steady state translation properties of *E. coli* mRNAs as functions of polysome size with and without accounting for ternary complex competitive binding to the ribosomal A site. In this chapter we define ribosomal fractional coverage, ρ, to be the fraction of the mRNA molecule covered by translating ribosomes, where:

$$\rho = \frac{L\sum_{n} S_{ij,n,r}}{M_r N_r} \tag{3.15}$$

where $L = 12$ is the number of codons covered by the ribosome (16-18) and N_r is the number of codons of mRNA r. Further discussion of the ribosomal fractional coverage is included in Chapter 2. The values for the concentration of each mRNA species r, M_r, the free ribosome concentration, $R^{(f)}$, the free ternary complex concentrations, $T_k^{(f)}\left(k \in K\right)$, and the free Ef-G concentration, $G^{(f)}$, applied in these studies, along with the reaction rate constants, k_1, k_{-1}, k_2, k_{-2}, k_3, k_4, k_5, k_6, k_7, k_{-7}, k_8, k_9, are the same as those used in Chapter 2. Also, the method used to calculate steady state translation rate as a function of polysome size is the same as that from Chapter 2. The obtained steady state translation rates are applied to the sensitivity analysis in order to determine the flux control coefficients.

In this work we consider three cases with respect to ternary complex binding to the ribosomal A site codon. These cases differ by how the free ternary complex concentrations are

applied for the quantification of the effective elongation rate constants $k_{E,n,r}^{eff}$ and $k_{E,n,r}^{eff,T}$. In Table 3.1 we list the free ternary complex concentrations and corresponding magnitudes for $k_{E,n,r}^{eff}$ and $k_{E,n,r}^{eff,T}$ for all the ternary complex species. The effective elongation rate constant magnitudes shown correspond to $U_{n,r}=1$. The following assumptions were employed for each case:

Table 3.1: Effective elongation rate constant magnitudes for each *E. coli* ternary complex species*

Species	$T^{(f)}$ (μM)	k_E^{eff} $(s^{-1})^{\dagger}$	$k_E^{eff,T}$ $(s^{-1})^{\dagger}$	Species	$T^{(f)}$ (μM)	k_E^{eff} $(s^{-1})^{\dagger}$	$k_E^{eff,T}$ $(s^{-1})^{\dagger}$
Ala1B	14.7	5.3	2.0	Leu5	3.5	5.1	0.7
Ala2	1.9	4.9	0.4	Lys	6.1	5.2	1.1
Arg2	23.1	5.3	2.7	Met	2.5	5.0	0.5
Arg3	2.3	5.0	0.4	Phe	4.0	5.1	0.7
Arg4	3.4	5.1	0.6	Pro1	2.2	5.0	0.4
Arg5	2.3	5.0	0.4	Pro2	3.8	5.1	0.7
Asn	5.4	5.2	0.9	Pro3	2.0	4.9	0.4
Asp1	9.5	5.2	1.5	Ser1	6.9	5.2	1.2
Cys	7.2	5.2	1.2	Ser2	1.3	4.7	0.3
Gln1	3.0	5.0	0.6	Ser3	5.4	5.2	1.0
Gln2	3.8	5.1	0.7	Ser5	3.8	5.1	0.7
Glu2	20.7	5.3	2.5	Thr1	0.4	3.7	0.1
Gly1	5.6	5.2	1.0	Thr2	2.8	5.0	0.5
Gly2	5.6	5.2	1.0	Thr3	4.3	5.1	0.8
Gly3	18.0	5.3	2.3	Thr4	5.4	5.2	1.0
His	3.6	5.1	0.7	Trp	4.1	5.1	0.7
Ile1	7.2	5.2	1.2	Tyr1	4.4	5.1	0.8
Ile2	10.8	5.3	1.6	Tyr2	5.1	5.2	0.9
Leu1	14.9	5.3	2.0	Val1	17.1	5.3	2.2
Leu2	5.2	5.2	0.9	Val2A	2.6	5.0	0.5
Leu3	2.2	5.0	0.4	Val2B	3.7	5.1	0.69
Leu4	9.4	5.2	1.5				

*Numerical values for the free ternary complex species concentrations, $T^{(f)}$, are estimated in APPENDIX C. Experimental data for the ternary complex concentrations can be found in (36).
†Evaluated at $U_{n,r}=1$.

Case I: Noncompetitive binding

Only the ternary complex species recognizing the ribosomal A site codon are allowed to participate in the nonspecific binding step of the elongation cycle. In studies considering ***Case I*** the ternary complex concentrations recognizing the A site codons, $T_j^{(f)}$, are set equal to the median concentration, $T_{med}^{(f)} = 4.3\,\mu M$, in the effective elongation rate constant expressions $\left(k_{E,n,r}^{eff}\right)$ of every codon. Although variations in ternary complex concentrations cause variations in effective elongation rate constant magnitudes, we have observed that these differences are negligibly small under noncompetitive binding conditions (a detailed discussion is included in Chapter 2).

Case II: Uniform competitive binding

All ternary complex species are allowed to participate in the nonspecific binding step of the elongation cycle. Similar to ***Case I***, in studies considering ***Case II*** the ternary complex concentrations recognizing the A site codons, $T_j^{(f)}$, are set equal to the median concentration, $T_{med}^{(f)} = 4.3\,\mu M$ in the effective elongation rate constant expressions $\left(k_{E,n,r}^{eff,T}\right)$ of every codon. Although we observe in this work that variations in ternary complex concentrations cause significant variations in elongation rate constant magnitudes, ***Case II*** allows us to study the effects of ternary complex competitive binding in a codon independent manner. Moreover, because all codons are treated uniformly in ***Case II***, we study the effects of ternary complex competitive binding in a sequence independent manner.

Case III: Nonuniform competitive binding

Similar to ***Case II***, all ternary complex species are allowed to participate in the nonspecific binding step of the elongation cycle. However, in studies considering ***Case III*** the ternary

complex concentrations recognizing the A site codons, $T_j^{(f)}$, are set equal to their respective physiological levels in the effective elongation rate constant expression $\left(k_{E,n,r}^{eff,T}\right)$. Because codons are not treated uniformly in ***Case III***, we study the effects of ternary complex competitive binding in both a codon and sequence specific manner.

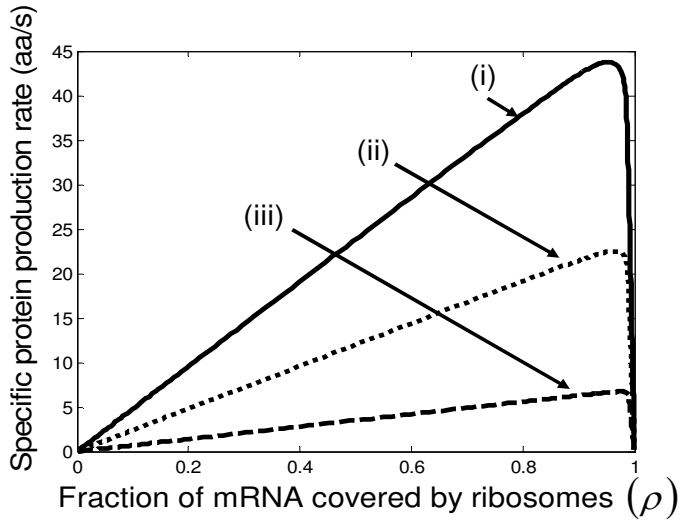
3.3.1 Effects of ternary complex competitive binding on the relationships between protein synthesis properties and polysome size

In these studies we apply ***Cases I*** and ***II*** to investigate the translation properties of the *trpR* gene of *E. coli* in both a codon and sequence independent manner.

Effects of ternary complex competitive binding on the relationship between translation rate and polysome size

We observe that as ribosomal fractional coverage increases, the protein synthesis rate increases, reaches a maximum, and then decreases under both competitive (Figure 3.1, curves ii and iii) and noncompetitive (Figure 3.1, curve i) binding conditions. Included in Figure 3.1 are results for ***Cases I*** and ***II*** (curves i and iii, respectively), and ***Case II*** with all the codons in the sequence recognized by the ternary complex species having the maximum free concentration of $23.1\,\mu M$ (curve ii). The translation rates determined under ***Case I*** are higher at each polysome size than those determined under ***Case II***. This result is due to the large difference in the effective elongation rate constant magnitudes under the two cases. For $U_{n,r}=1$ under ***Case I*** $k_{E,n,r}^{eff}=5.1\,s^{-1}$ (curve i), while under ***Case II*** $k_{E,n,r}^{eff,T}=2.7\,s^{-1}$ (curve ii) and $k_{E,n,r}^{eff,T}=0.8\,s^{-1}$ (curve iii). Because the effective elongation rate constant magnitudes under ***Case I*** are higher than those under ***Case II***, the translation rates observed under ***Case I*** are higher than those observed under ***Case II***.

Figure 3.1: Relationship between translation rate and polysome size for (i) ***Case I***, and ***Case II*** with all codons recognized by the ternary complex species of maximum (ii) and median (iii) concentrations.



Rate limiting steps and polysome size

We applied the control analysis framework to the model in order to determine if translation is initiation, elongation, or termination limited under different polysome sizes. We observe that under both ***Cases I*** and ***II*** translation is initiation limited for $\rho < 0.5$, elongation limited for $0.5 < \rho < 0.99$, with elongation control maximal at the same ribosomal fractional coverage that specific protein production rate is maximal, and termination limited for $\rho > 0.99$.

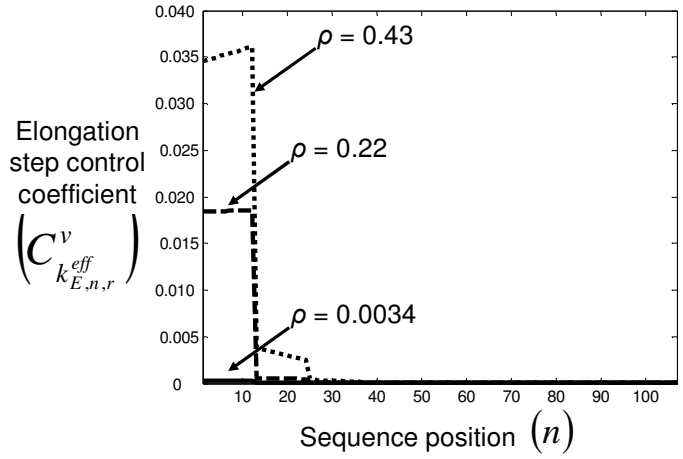
Relationship between codon specific control of protein translation rate and polysome size

We investigated how control of the elongation phase over translation rate $\left(C^{v}_{k_{E,r}}\right)$ is distributed with respect to the codons along the length of the mRNA at different polysome sizes by examining the control coefficients corresponding to the effective elongation rate constants, $C^{v}_{k^{eff}_{E,n,r}}$ (Figure 3.2). We observe that at low polysome sizes the elongation phase control over translation rate lies in the codons near the 5' end of the mRNA. This result is in agreement with early experimental results demonstrating that point mutations near the start codon of the mRNA

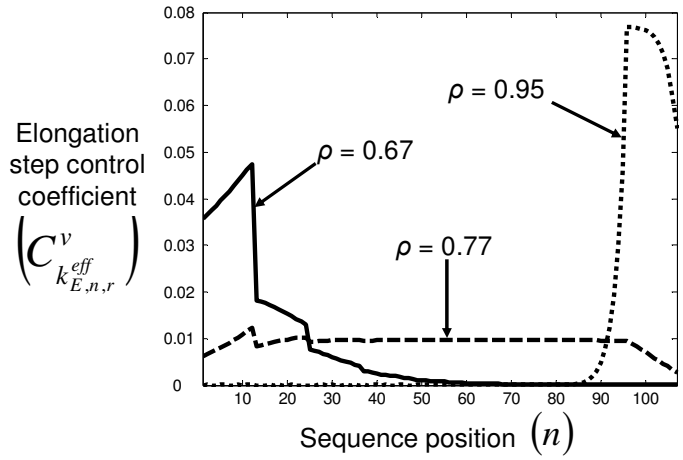
cause dramatic changes in protein expression levels (37, 38). Also, at intermediate polysome sizes the control is distributed along the length of the mRNA in different configurations, and at high polysome sizes the control lies in the codons near the 3' end of the mRNA. We observe the same results under both ***Cases I*** and ***II***. These results are expected because at low polysome sizes kinetics are initiation limited (see previous paragraph for discussion), which means that the initiation process limits the progress of protein translation. Hence, the more efficiently the codons near the 5' end of the mRNA can be translated, the more ribosomes can be transferred to downstream codons along the length of the sequence. Faster transfer of ribosomes due to more efficient translation of these codons elevates protein synthesis rate by increasing the probability of an initiation event occuring without changes to the initiation process being made. The converse is true for termination limited conditions.

Figure 3.2: Elongation step control coefficients, $C^{v}_{k^{eff}_{E,n,r}}$, with respect to sequence position under initiation (A), elongation (B), and termination (C) limited conditions for ***Cases I*** and ***II***.

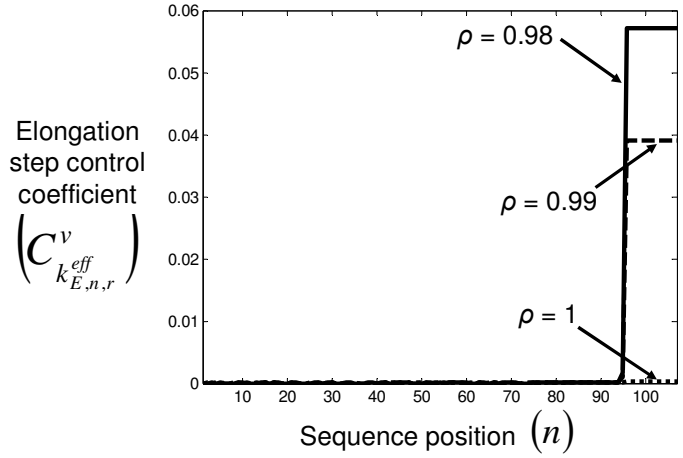
A



B



C



Relationship between codon specific elongation cycle intermediate step control of protein translation rate and polysome size

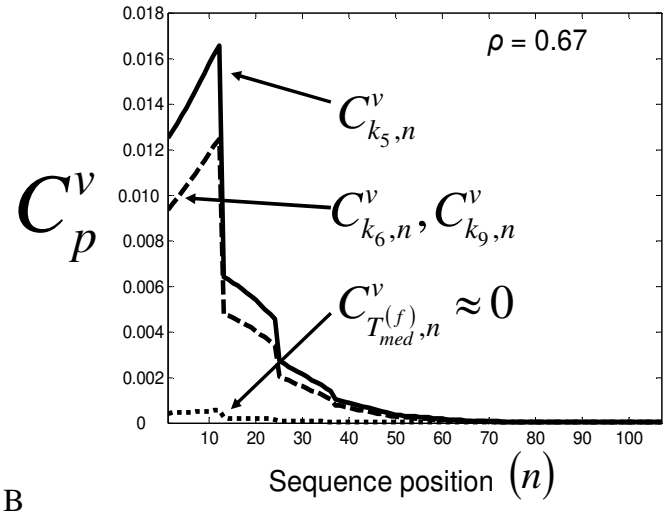
We investigated how the elongation phase control is distributed with respect to the elongation cycle intermediate steps at each codon along the length of the mRNA by examining the control coefficients: $C^{v}_{k_1,n}$, $C^{v}_{k_{-1},n}$, $C^{v}_{k_2,n}$, $C^{v}_{k_{-2},n}$, $C^{v}_{k_3,n}$, $C^{v}_{k_4,n}$, $C^{v}_{k_5,n}$, $C^{v}_{k_6,n}$, $C^{v}_{k_7,n}$, $C^{v}_{k_{-7},n}$, $C^{v}_{k_8,n}$, and $C^{v}_{k_9,n}$, along with the control coefficients corresponding to free ternary complex concentration, $C^{v}_{T_j^{(f)},n}$. We observe that the rate limiting step at each codon along the length of the mRNA is different between ***Cases I*** and ***II***. Under ***Case I*** we observe that the control coefficient with respect to the Ef-Tu:GDP release rate constant, $C^{v}_{k_5,n}$, is the highest of the control coefficients corresponding to elongation cycle intermediate steps at every sequence position and polysome size (Figure 3.3A, results shown only for ρ = 0.67), indicating that this intermediate step is rate limiting to the elongation cycle. This result is consistent with experimental reports which identify Ef-Tu:GDP release as one of the rate limiting steps of the elongation cycle at a given codon (23). Control coefficients for A site tRNA accommodation $\left(C^{v}_{k_6,n}\right)$ and E site tRNA release $\left(C^{v}_{k_9,n}\right)$ are equal to each other and also high (Figure 3.3A, results shown only for ρ = 0.67) at every sequence position and polysome size. The remaining elongation cycle intermediate steps have low control coefficients, including that for the free ternary complex concentration control coefficient $\left(C^{v}_{T_j^{(f)},n}\right)$.

However, under ***Case II*** we observe that the control coefficient with respect to the free ternary complex concentration, $C^{v}_{T_j^{(f)},n}$, is highest at every sequence position and polysome size (Figure 3.3B, results shown only for ρ = 0.67), indicating that ternary complex nonspecific

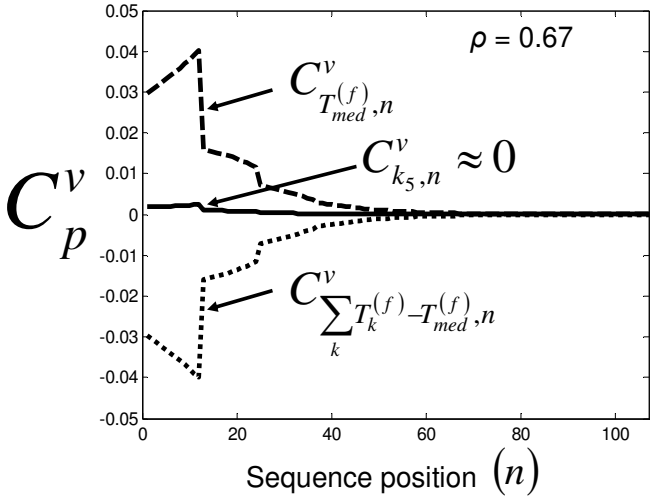
binding is rate limiting to the elongation cycle. The remaining elongation cycle control coefficients are close to zero, indicating that the intermediate steps following ternary complex nonspecific binding have very little influence on elongation cycle kinetics. Moreover, the rate limiting effects of ternary complex nonspecific binding are much higher under ***Case II*** than the rate limiting effects of Ef-Tu:GDP release under ***Case I***, with free ternary complex concentration control coefficients $\left(C^{v}_{T_j^{(f)},n}\right)$ under ***Case II*** more than twice as high as Ef-Tu:GDP release control coefficients $\left(C^{v}_{k_5,n}\right)$ under ***Case I***. We also observe that the concentrations of ternary complexes that do not recognize the A site codon, $T_k^{(f)}\left(k \neq j\right)$, have an inhibitory effect on translation kinetics because the corresponding control coefficients for the combined concentration of the incorrect ternary complexes, $C^{v}_{\sum_k T_k^{(f)} - T_{med}^{(f)},n}$, are negative (Figure 3.3B, results shown only for ρ = 0.67), meaning that an increase in this concentration would cause a decrease in translation rate.

Figure 3.3: Elongation cycle intermediate control coefficients with respect to sequence position under ***Case I*** (A) and ***Case II*** (B) binding conditions. Results shown are for elongation limited conditions ($\rho = 0.67$).

A



B



It is important to note that it is the relative magnitudes of the terms in the effective elongation rate constant, $k_{E,n,r}^{eff,T}$, that play a significant role in the distribution of control with respect to the elongation cycle intermediate steps at each codon. The influence each elongation cycle intermediate step has over the overall kinetics of the elongation cycle at a given codon is proportionate to the magnitude of its corresponding term in the effective elongation rate constant.

Under noncompetitive binding conditions, Ef-Tu:GDP release is rate limiting, with $\alpha_5 = 0.067$ (Table 2.3) being the largest term in $k_{E,n,r}^{eff}$, and nonspecific ternary complex binding has almost no influence over elongation cycle kinetics, with $\alpha_{1,j} = 6\times10^{-4} - 0.04$ (Table 2.3). However, under competitive binding conditions, the magnitude of the nonspecific ternary complex binding term in the effective elongation rate constant is much higher than α_5, with $\alpha_{1,j}^{T} = 0.19 - 12.9$, making nonspecific binding rate limiting.

To further understand the relationship between the magnitudes of the effective elongation rate constant terms and the control the corresponding elongation cycle intermediate steps have over translation rate, we introduce the elasticities of the elongation rate at codon *n* with respect to the free ternary complex concentration, $\varepsilon_{T_j^{(f)}}^{V_{ij,n,r}}$, and the reaction rate constant for Ef-Tu:GDP release, $\varepsilon_{k_5}^{V_{ij,n,r}}$, under competitive binding conditions:

$$\varepsilon_{T_j^{(f)}}^{V_{ij,n,r}} \equiv \frac{\partial \ln V_{ij,n,r}}{\partial \ln T_j^{(f)}} = \frac{T_j^{(f)}}{V_{ij,n,r}} \frac{\partial V_{ij,n,r}}{\partial T_j^{(f)}} = \alpha_{1,j}^{T} k_{E,n,r}^{eff,T} U_{n,r} \tag{3.16}$$

$$\varepsilon_{k_5}^{V_{ij,n,r}} \equiv \frac{\partial \ln V_{ij,n,r}}{\partial \ln k_5} = \frac{k_5}{V_{ij,n,r}} \frac{\partial V_{ij,n,r}}{\partial k_5} = \alpha_5 k_{E,n,r}^{eff,T} U_{n,r} \tag{3.17}$$

Elasticity is defined as the differential change in the rate of a single reaction step, i.e., in this case $V_{ij,n,r}$. Unlike the control coefficients, which pertain to the overall translation rate of the mRNA, elasticity is therefore a property local to that reaction step and not a systemic property. However, due to the compactness of the elasticity expressions, they are useful for obtaining general quantitative insight into the impact of the individual reaction rate constants and translational components on their respective control coefficient magnitudes. It is evident from the above

expressions that the elasticity of $V_{ij,n,r}$ with respect to a given parameter is dependent on the effective elongation rate constant term to which the parameter pertains, and not only on that parameter. Along these lines, the relative magnitudes of the elasticities are proportionate to the relative magnitudes of the corresponding effective elongation rate constant terms, and similar relationships are obtained between the remaining effective elongation rate constant terms and their corresponding elasticities. Consequently, the control the elongation cycle intermediate steps have over the translation rate of each codon is strongly influenced by the magnitudes of their respective effective elongation rate constant terms.

Overall, in these studies we observe that ternary complex competitive binding to the ribosomal A site introduces changes to translation rate (Figure 3.1). However, competitive binding does not cause changes to the distribution of overall initiation, elongation, and termination control with respect to polysome size. Moreover, competitive binding does not affect the codon specific distribution of control with respect to polysome size (Figure 3.2), but instead introduces changes to the distribution of control with respect to elongation cycle intermediate step at each codon (Figure 3.3).

3.3.2 Ternary complexes not recognizing the ribosomal A site codon act as competitive inhibitors to elongation cycle kinetics

To further investigate the inhibitory effects of ternary complexes not recognizing the ribosomal A site codon, we derive our mechanistic framework in the context of Michaelis – Menten enzyme kinetics. Treating all the translating ribosomes in a single *E. coli* cell having codon species *j* occupying the A site as the enzyme and the ternary complex species *j* recognizing the A site codon as the substrate, it can be shown that in the absence of ternary complex competitive binding:

$$v_{MM,j} = V_{\max,j} \frac{T_j^{(f)}}{K_M + T_j^{(f)}} \tag{3.18}$$

where the Michaelis – Menten constant is:

$$K_M = \frac{k_{-1} + k_M}{k_1}, \tag{3.19}$$

and the maximum reaction rate is,

$$V_{\max,j} = k_M R_j, \tag{3.20}$$

In the above expression R_j is the cellular concentration of ribosomes in the cell participating in translation with codon species *j* occupying the A site. By accounting for ternary complex competitive binding, it can be shown that:

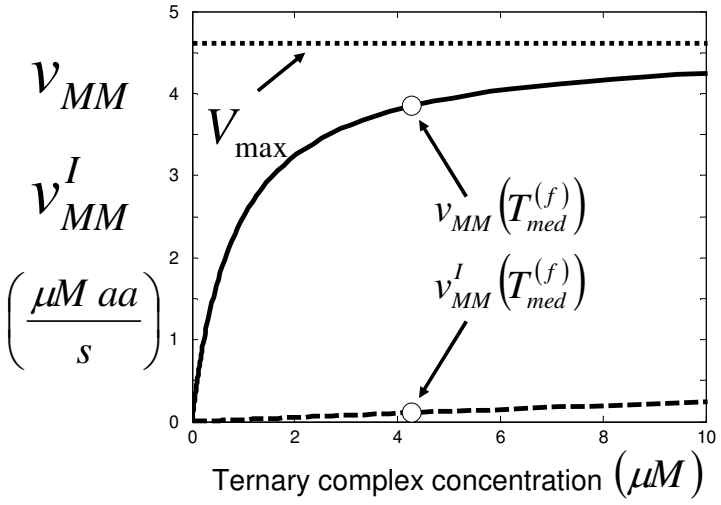
$$v_{MM,j}^{I} = V_{\max,j} \frac{T_j^{(f)}}{K_M \left(1 + \frac{\sum_{k \neq j} T_k^{(f)}}{K_1} \right) + T_j^{(f)}} \tag{3.21}$$

Details of the derivation of the above equations, along with the estimation of R_j, are included in APPENDIX E.

Under competitive binding conditions the ternary complexes that do not recognize the A site codon $\left(T_k^{(f)}\left(k \neq j\right)\right)$ bind to the ribosome as the ternary complexes that do $\left(T_j^{(f)}\right)$ during the nonspecific binding step of the elongation cycle, but do not proceed to the subsequent intermediate steps. The ternary complexes $T_k^{(f)}\left(k \neq j\right)$ occupying the ribosomal A site prevent the ternary complexes $T_j^{(f)}$ from binding to the ribosome, so the apparent affinity the ternary complexes recognizing the A site codon $\left(T_j^{(f)}\right)$ have for the ribosome decreases. This decrease is

due to the term multiplied by the Michaelis – Menten constant (K_M) in the expression for $v^I_{MM,j}$ (equation 3.21), which represents the inhibitory effects of the ternary complexes $T^{(f)}_k (k \neq j)$ on translation rate. However, the maximum reaction rate $(V_{\max,j})$ is the same under both noncompetitive and competitive binding conditions. Figure 3.4 shows the relationship between translation rates $v_{MM,j}$ and $v^I_{MM,j}$ as functions of the ternary complex concentration recognizing the A site codon, $T^{(f)}_j$, for the median ternary complex concentration, $T^{(f)}_{med}$. The maximum reaction rate $(V_{\max,j})$ is proportional to the concentration of translating ribosomes having the codon recognized by the ternary complex species of median concentration present in the A site, and the ternary complex concentration is allowed to vary. When evaluating the translation rate expressions with and without competitive binding at the median ternary complex concentration $\left(v^I_{MM}\left(T^{(f)}_{med}\right), v_{MM}\left(T^{(f)}_{med}\right)\right)$, we observe a much lower translation rate under competitive binding conditions than under noncompetitive binding conditions (Figure 3.4). Similar results are observed for the remaining ternary complex species.

Figure 3.4: Amino acid rate of incorporation as a function of ternary complex concentration for ***Case I*** (solid line) and ***Case II*** (dashed line). Results shown are at the *E. coli* cellular level for the ternary complex species of median concentration. Similar results are observed for all ternary complex species.



We examined the expressions for the elasticities, $\varepsilon_{MM,T_j^{(f)}}$ and $\varepsilon^I_{MM,T_j^{(f)}}$, of the reaction rates $v_{MM,j}$ and $v^I_{MM,j}$ with respect to the ternary complex concentration recognizing the ribosomal A site, i.e. the ratios of the proportional changes in $v_{MM,j}$ and $v^I_{MM,j}$ with respect to the proportional change in $T_j^{(f)}$:

$$\varepsilon_{MM,T_j^{(f)}} \equiv \frac{\partial \ln v_{MM,j}}{\partial \ln T_j^{(f)}} = \frac{T_j^{(f)}}{v_{MM,j}} \frac{\partial v_{MM,j}}{\partial T_j^{(f)}} \tag{3.22}$$

$$\varepsilon^I_{MM,T_j^{(f)}} \equiv \frac{\partial \ln v^I_{MM,j}}{\partial \ln T_j^{(f)}} = \frac{T_j^{(f)}}{v_{MM,j}} \frac{\partial v^I_{MM,j}}{\partial T_j^{(f)}} \tag{3.23}$$

Evaluating equations (3.22) and (3.23) yields:

$$\varepsilon_{MM,T_j^{(f)}} = 1 - \frac{T_j^{(f)}}{K_M + T_j^{(f)}} = \frac{K_M}{K_M + T_j^{(f)}} \tag{3.24}$$

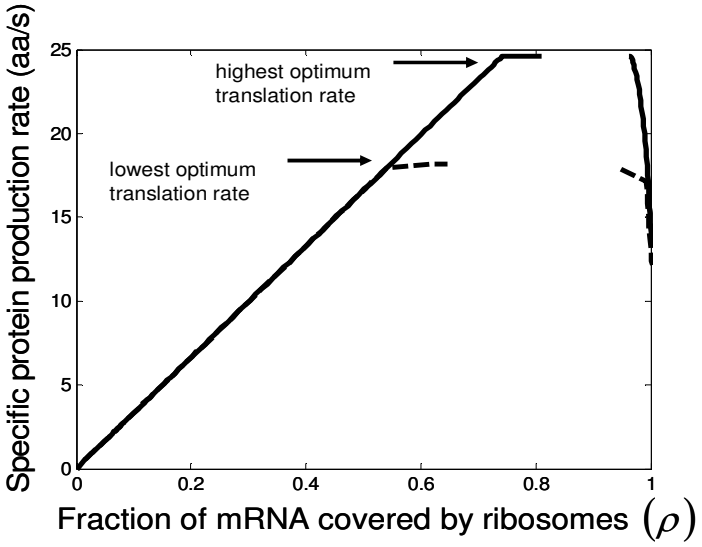
$$\varepsilon^{I}_{MM,T_j^{(f)}} = 1 - \frac{T_j^{(f)}}{K_M\left(1+\frac{\sum_{k\neq j} T_k^{(f)}}{K_1}\right)+T_j^{(f)}} = \frac{K_M\left(1+\frac{\sum_{k\neq j} T_k^{(f)}}{K_1}\right)}{K_M\left(1+\frac{\sum_{k\neq j} T_k^{(f)}}{K_1}\right)+T_j^{(f)}} \tag{3.25}$$

We observe that the elasticities determined under competitive binding conditions $\left(\varepsilon^{I}_{MM,T_j^{(f)}}\right)$ are much greater than those determined under noncompetitive binding conditions $\left(\varepsilon_{MM,T_j^{(f)}}\right)$, with $\varepsilon^{I}_{MM,T_j^{(f)}} = 0.98$ and $\varepsilon_{MM,T_j^{(f)}} = 0.17$ for the median ternary complex concentration, $T_{med}^{(f)}$. Equations (3.24) and (3.25) suggest that the lower the ternary complex concentration recognizing the ribosomal A site codon, $T_j^{(f)}$, the stronger the sensitivity to change under competitive binding conditions $\left(\varepsilon^{I}_{MM,T_j^{(f)}}\right)$ than noncompetitive binding conditions $\left(\varepsilon_{MM,T_j^{(f)}}\right)$. Similar results are observed for the remaining ternary complex species. As we observed in the results above relating to reaction rates $v_{MM,j}$ and $v^{I}_{MM,j}$, the increased elasticities under competitive binding conditions are observed because of the term multiplied by the Michaelis – Menten constant $\left(K_M\right)$ in the expression for $\varepsilon^{I}_{MM,T_j^{(f)}}$ (equation 3.25) that represents the inhibitory effects of the ternary complexes $T_k^{(f)}\left(k \neq j\right)$ on translation rate. The results in this section support our results discussed in previous sections pertaining to ternary complex competitive binding lowering translation rate and causing the nonspecific binding intermediate step to be rate limiting to the elongation cycle at each codon. However, the results presented in this section suggest that the effects of competitive binding are due to the ternary complexes not recognizing the A site codon $\left(T_k^{(f)}\left(k \neq j\right)\right)$ acting as competitive inhibitors to elongation cycle kinetics.

3.3.3 The relative position of codons along the mRNA determines the optimal protein syntheisis rate and the rate limiting effect of the individual codons

In these studies we apply ***Case III*** to investigate the translation properties of mRNAs in both a codon and sequence dependent manner**.** We applied our mechanistic framework to one hundred randomly permuted sequences having identical codon frequencies representative of those of the *E. coli* genome. Each sequence is 361 codons long, approximately the average length of an *E. coli* mRNA (32). Similar to our results in previous sections, we observe that the translation rate increases, reaches a maximum, and then decreases as polysome size increases (Figure 3.5). However, optimum protein synthesis rates vary with sequence (Figure 3.5, results shown only for sequences producing highest and lowest optimum translation rates). We also observe that the optimum rate occurs at multiple polysome sizes for each sequence and that there are regimes of polysome sizes for which translation properties are highly sensitive to the input parameters of our model (Figure 3.5). Because all the sequences in this study have the same codon frequencies, the results presented in this section emphasize that the relative positions of codons along the length of the mRNA can influence protein synthesis properties.

Figure 3.5: Relationship between translation rate and polysome size under ***Case III*** conditions.

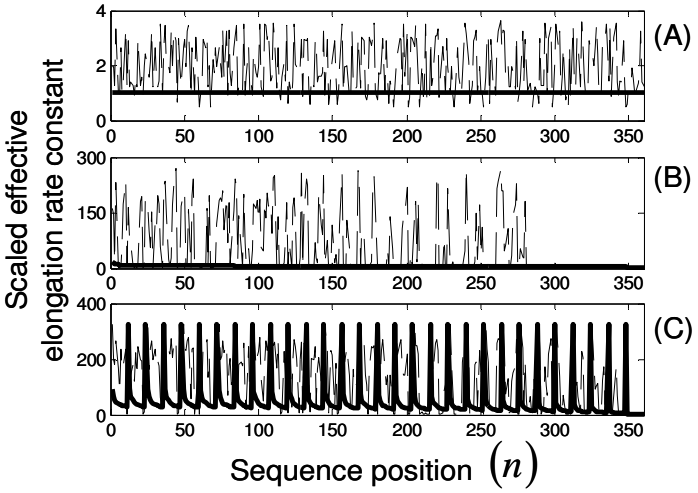


Relationship between effective elongation rate constant magnitudes and polysome size

In order to investigate the overall relationship between translation rate and polysome size with ***Case III*** conditions, we examined changes in the effective elongation rate constant magnitudes with polysome size. We scaled the effective elongation rate constants by dividing them by the effective elongation rate constant, $k_{E,n,r}^{eff,T}$, evaluated at $U_{n,r}=1$ and $T_j^{(f)}=T_{med}^{(f)}$ that has a magnitude of $0.8\,s^{-1}$. In the absence of ribosomal crowding on the mRNA, i.e. when $U_{n,r}=1$, the scaled effective elongation rate constant magnitudes vary between 0.10 and 3.44 due to differences in the nonspecific binding term, $\alpha_{1,j}^{T}$ (equation 3.4). Using effective elongation rate constants determined with ***Case II*** conditions as a reference, we scaled them the same way by dividing them by the effective elongation rate constant, $k_{E,n,r}^{eff}$, evaluated at $U_{n,r}=1$ and $T_j^{(f)}=T_{med}^{(f)}$ that has a magnitude of $5.1\,s^{-1}$. Included in Figure 3.6 are the distributions of scaled effective elongation rate constant magnitudes as functions of sequence position under initiation (A), elongation (B), and termination (C) limited conditions for one of the sequences

used in this section (similar results are observed for the other sequences). The dashed lines represent magnitudes under ***Case III*** conditions, and the solid lines represent magnitudes under ***Case II*** conditions.

Figure 3.6: Scaled effective elongation rate constant magnitudes under initiation (A), elongation (B), and termination (C) limited conditions for one of the sequences used in this section (similar results are observed for the other sequences). The dashed lines represent magnitudes determined with ***Case III*** conditions, while the solid lines represent magnitudes determined with ***Case II*** conditions.



We observe that under initiation limited conditions the scaled effective elongation rate constants for both ***Case II*** and ***Case III*** are approximately equal to the values they take on when $U_{n,r} = 1$, and this result is expected because the polysome size is low and hence the mRNA is not crowded. Under elongation limited conditions the level of crowding on the mRNA is higher as reflected in the conditional probability term, $U_{n,r}$, of the effective elongation rate constant decreasing, which results in the scaled effective elongation rate constant magnitudes and translation rates increasing. The level of ribosomal crowding on the mRNA determines the magnitudes under ***Case II*** conditions (see Chapter 2 for more discussion), while the complex

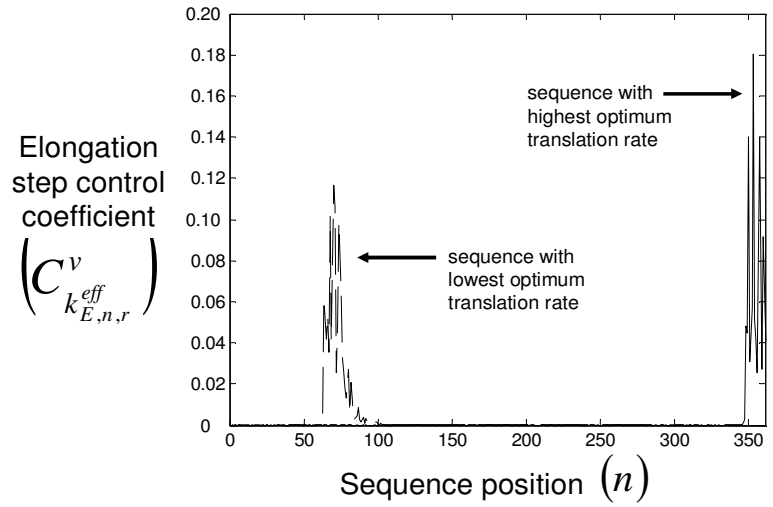
interplay between the level of ribosomal crowding on the mRNA and the level of ternary complex competition for the ribosomal A site at each codon determines the scaled effective elongation rate constant magnitudes under ***Case III*** conditions. Codons that experience a lot of ternary complex competitive inhibition have lower effective elongation rate constants and are hence translated more slowly than codons that do not, causing high ribosome density upstream on the mRNA and large variation in the conditional probability term, $U_{n,r}$, that is not observed under ***Case II*** conditions. Consequently the scaled effective elongation rate constants determined with ***Case III*** are much higher than those determined with ***Case II***. Under termination limited conditions the polysome size is high, so crowding on the mRNA is maximal and $U_{n,r} \approx 0$, regardless of whether the binding conditions are uniform (***Case II***) or nonuniform (***Case III***). Due to the ribosomal queuing that occurs along the length of the mRNA at high polysome size (see Chapter 2 for more discussion), the effective elongation rate constants at positions spaced one ribosome length apart are approximately equal to the translocation rate constant, k_8 (see Chapter 2 for more discussion).

Effects of rate limiting codon segments on the relationship between optimum translation rate and polysome size

In order to investigate translation properties occurring in the regimes of polysome sizes associated with optimum rates, we obtained the elongation step control coefficients $\left(C^{v}_{k^{eff,T}_{E,n,r}}\right)$ of each sequence at its respective optimum translation rate (Figure 3.5). Similar to results in Chapter 2, at the optimum rate the kinetics are completely elongation limited, with $\sum_{n=1}^{N_r-1} C^{v}_{k^{eff,T}_{E,n,r}} = C^{v}_{k_{E,r}} = 1$. The control over rate is dominated by segments of codons that have high

elongation step control coefficients (Figure 3.7, results shown only for sequences producing highest and lowest optimum translation rates). For all of the polysome sizes that the translation kinetics are completely elongation limited, we observe that the configuration of elongation step control coefficients $\left(C^{v}_{k^{eff,T}_{E,n,r}}\right)$ does not change, and therefore the segments of rate limiting codons do not change.

Figure 3.7: Elongation step control coefficients, $C^{v}_{k^{eff}_{E,n,r}}$, with respect to sequence position under ***Case III*** conditions.



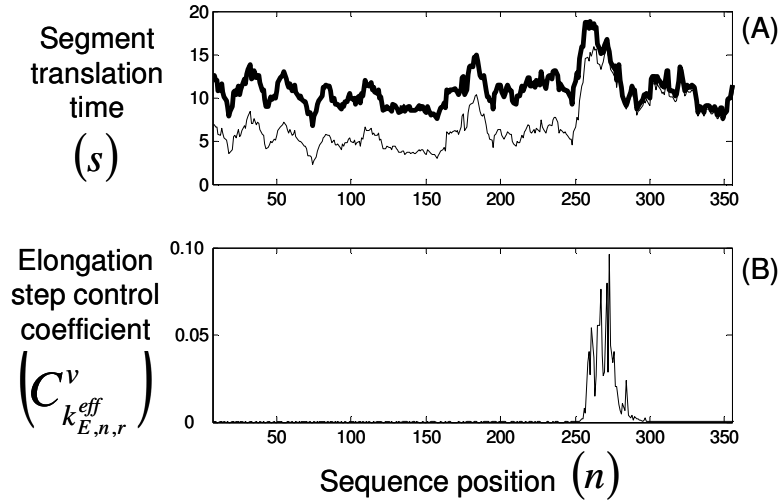
The positions of rate limiting codon segments are expected because they correspond to segments of high translation time (Figure 3.8). We define the translation time of the codon segments to be:

$$t^{seg}_{n} = \sum_{n-5}^{n+6} \frac{1}{k^{eff,T}_{E,n,r} U_{n,r}} \qquad ,n \in [6,355] \tag{3.26}$$

where the codon segments are equal to one ribosome length, and the translation time of the segment corresponding to codon *n* is equal to the combined translation time of that codon along with the five upstream and six downstream codons. We consider the codon segments of one

ribosome length in this way because *n* denotes the position of the ribosomal P site codon, and in this work and in Chapter 2 we assume the front and back ends of the ribosome are on the sides closest to the 3' and 5' ends of the mRNA, respectively, with the P site covering the seventh codon relative to the front of the ribosome. We estimated the segment translation times with $U_{n,r} = 1$ for all of the codons in the sequence and with $U_{n,r}$ values corresponding to the ribosome distribution at the optimum translation rate (Figure 3.8A). The part of the sequence having the highest translation times (Figure 3.8A) corresponds to the rate limiting codon segment (Figure 3.8B). Additionally, the segment translation times with $U_{n,r} = 1$ can be interpreted as the translation times in the absence of ribosomal crowding. At the optimum translation rate the ribosome density is high, resulting in low values for $U_{n,r}$ for all of the codons in the sequence. Consequently, the translation times at high ribosome densities are higher than those at low ribosome densities. However, the part of the sequence having the highest segment translation times remains the same at both high and low ribosome densities, which is consistent with the rate limiting codon segment remaining unchanged for the polysome sizes that the translation kinetics are completely elongation limited.

Figure 3.8: (A) Segment translation times, t_n^{seg}, with respect to sequence position for one of the sequences used in this study (similar results are observed for the other sequences). (B) Elongation step control coefficients, $C^{v}_{k^{eff}_{E,n,r}}$, with respect to sequence position under ***Case III*** conditions for the same sequence used in Figure 3.8A.



Because the rate limiting codon segments correspond to regions of high translation time, they also lead to nonuniform ribosome distributions along the length of the mRNA. Defining *p* to be the sequence position of the codon at the 3' end of the rate limiting codon segment, the ribosome density upstream of *p* is as follows:

$$\rho^{u} = \frac{L\sum_{n=1}^{p-1} S_{ij,n,r}}{M_r(p-1)} \tag{3.27}$$

while the ribosome density downstream of *p* is as follows:

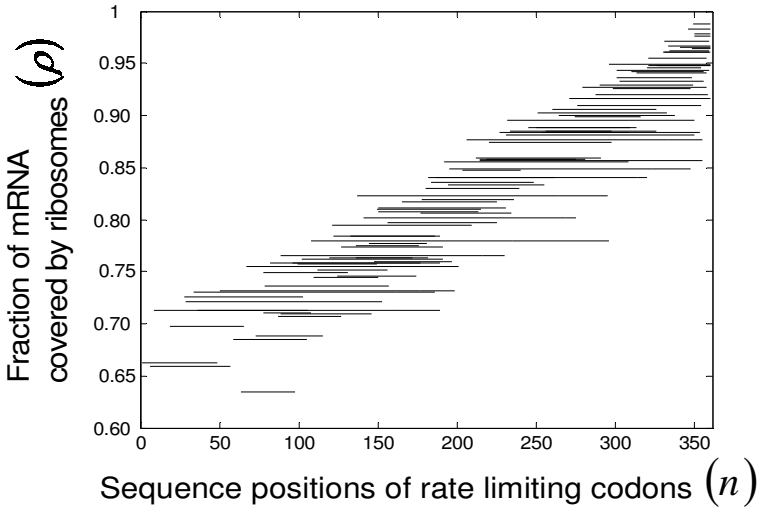
$$\rho^{d} = \frac{L\sum_{n=p}^{361} S_{ij,n,r}}{M_r(N_r - p + 1)} \tag{3.28}$$

The ribosome density upstream of *p*, ρ^{u}, corresponds to the part of the sequence between the 5' end of the mRNA and the 3' end of the rate limiting codon segment, while the ribosome density

downstream of p, ρ^d, corresponds to the remainder of the sequence. For all of the sequences studied at the optimum translation rate $\rho^u > \rho^d$, with $0.61 \le \rho^u \le 0.89$ and $0.0091 \le \rho^d \le 0.76$. Ribosomes translate the codons in the rate limiting segments slower than those in the remainder of the sequence, leading to higher ribosome densities upstream of the rate limiting codon segments than downstream. It is important to note that, at the regime of polysome sizes corresponding to optimum translation rate, the translation kinetics shift from elongation to termination limited. Under elongation limited conditions we observe nonuniform ribosome densities, while under termination limited conditions we observe uniform queuing of the ribosomes along the length of the mRNA (see Chapter 2 for more discussion). In this transitional regime the ribosome density becomes very sensitive to the input parameters of our model, making it difficult to obtain data. Hence, in Figure 3.5 there are regimes of polysome sizes for which we do not show translation rates.

Furthermore, the positions of the rate limiting codon segments determine the minimum polysome size at which the optimum translation rate occurs (Figure 3.9). The closer to the 3' end of the sequence the rate limiting segment is, the more ribosomes are accommodated on the mRNA, the higher the polysome size, and the higher the protein synthesis rate. This result indicates that the positioning of the rate limiting codon segment influences the optimum translation rate. Translation of the sequence with the highest optimum rate is limited by a codon segment near the 3' end of the mRNA (Figure 3.7). Consequently this sequence can accommodate the most ribosomes, maximizing the probability of a translation termination event occurring and hence maximizing the optimum protein synthesis rate. The converse is true for the sequence with the lowest optimum rate, because its translation is limited by a codon segment near the 5' end of the mRNA (Figure 3.7).

Figure 3.9: Relationship between the positions of the rate limiting codon segments and ribosomal fractional coverage.



3.4 Discussion

In this chapter we presented a theoretical analysis of protein synthesis that includes all the elementary steps of the translation mechanism and accounts for ternary complex competitive binding to the ribosomal A site. Considering protein synthesis kinetics in the context of ternary complex competitive binding provides insights into quantifying the systemic contributions of ternary complex concentrations to the translational output of genes. Moreover, our codon specific sensitivity analysis allows us to separately quantify the influence the concentration of the ternary complex recognizing each codon along the length of the mRNA has on the overall protein synthesis rate. We find that the expanded mechanistic framework predicts lower protein synthesis rates than our framework in Chapter 2 (Figure 3.1), the configuration of codons that have the most control over protein synthesis rate changes with polysome size (Figure 3.2), and competitive, nonspecific binding of the ternary complexes to the ribosomal A site is rate limiting to the elongation cycle for every codon (Figure 3.3). These results suggest that the ternary

complexes that do not recognize the ribosomal A site codon act as competitive inhibitors to the ternary complexes recognizing the A site codon (Figure 3.4). Considering this model in the context of a Michaelis – Menten mechanistic framework demonstrates that translation rates are lower and more sensitive to ternary complex concentrations under competitive binding conditions than under noncompetitive binding conditions, which is consistent with what Michaelis - Menten kinetics predicts under competitive inhibition conditions where a substrate and inhibitor are competing for access to the active site of an enzyme.

In these studies the same set of reaction rate constants were used for the elongation cycle intermediate steps at every codon along the length of the sequence. Hence, the results suggest that it is the interplay between the level of ternary complex competition for the ribosomal A site at each codon and the level of ribosomal crowding on the mRNA that determines the effective elongation rate constant magnitudes at each codon and polysome size (Figure 3.6). This configuration at a given polysome size determines the corresponding protein synthesis properties (Figure 3.5). However, given that codon – anticodon compatibilities affect translation rates (9-11, 39), in future studies it will be important to incorporate anticodon specific kinetic and thermodynamic (40) parameters into our model to investigate how translational behavior is affected.

The set of elongation cycle reaction rate constants used in these studies were the same as those used in our studies in Chapter 2, which did not account for ternary complex competitive binding, and they predict higher translation rates, suggesting that Ef-Tu:GDP release is the rate limiting step of the elongation cycle for every codon. The expanded mechanistic framework in this study predicts lower translation rates and indicates that ternary complex nonspecific binding to the ribosome is the rate limiting step of the elongation cycle for every codon. It has been

shown experimentally that the ternary complexes not recognizing the ribosomal A site codon do not inhibit translation rate (39), and that Ef-Tu:GDP release is one of the rate limiting steps of the elongation cycle (23). Although these experimental results are consistent with the results of our previous study in Chapter 2, these experiments were performed *in vitro* and consequently do not reflect *in vivo* conditions. The results from (39) were obtained by examining the competitive binding effects $1.3\,\mu M$ Phe ternary complex experiences from Leu2 and Leu4 varying from $0\,\mu M$ to $16\,\mu M$ during poly(Phe) synthesis. By increasing the Leu2 and Leu4 concentrations from $0\,\mu M$ to $16\,\mu M$, the authors observe that the translation rates per ribosome decrease from $4.0\,s^{-1}$ to $3.0\,s^{-1}$. Hence, they conclude that ternary complex species not recognizing the ribosomal A site codon have almost no inhibitory effects *in vitro*. However, the total concentration of tRNA in *E. coli* is roughly $332\,\mu M$ (32), so *in vivo* a ternary complex species having a concentration of $1.3\,\mu M$ would experience much higher competitive effects than what is predicted in (39). By rearranging equation (3.2) we can express the translation rate per ribosome, $v_{ij,n,r}$, evaluated at $U_{n,r} = 1$ as:

$$v_{ij,n,r} = k_{E,n,r}^{eff,T} \qquad , n \in [1, N_r - 1] \tag{3.29}$$

Applying $1.3\,\mu M$ to $T_j^{(f)}$ and $0\,\mu M - 16\,\mu M$ to $\sum_{k \neq j} T_k^{(f)}$ in the above expression, we observe that the translation rate per ribosome decreases from $5.0\,s^{-1}$ to $2.4\,s^{-1}$, which is close to the range of translation rates per ribosome observed in (39). Similar to (39), our model predicts low inhibitory effects of ternary complexes not recognizing the ribosomal A site codon on translation rate. On the contrary, when we apply $1.3\,\mu M$ to $T_j^{(f)}$ and $332\,\mu M$ to $\sum_{k \neq j} T_k^{(f)}$ in equation (3.29), we obtain a translation rate per ribosome of $0.2\,s^{-1}$, which is much lower than what is observed

in (39) and what is predicted using our model above. This result indicates that ternary complexes have significant competitive effects *in vivo*. Moreover, our model predicts a 2 – 9 fold reduction in optimal translation rate due to ternary complex competitive binding (Figure 3.1), which is consistent with estimates in previous experimental work (22). Hence, the difference in the results from our framework in Chapter 2 and the mechanistic framework presented in this chapter further suggests that ternary complexes have a significant effect on translation kinetics by acting as competitive inhibitors.

3.5 Conclusions

In this chapter we expanded our mechanistic framework from Chapter 2 to account for ternary complex competitive binding to the ribosomal A site. We also performed a sensitivity analysis in order to determine the effects of the kinetic parameters and concentrations of the translational components on the protein synthesis rate. We determined the following:

(i) Translation rates are lower under ternary complex competitive binding conditions than under noncompetitive binding conditions. This result is due to the tRNAs that do not recognize the ribosomal A site codon acting as competitive inhibitors to the tRNAs that do recognize the ribosomal A site codon. Along these lines, the competitive, non-specific binding of the tRNAs to the ribosomal A site is rate limiting to the elongation cycle for every codon.

(ii) At low polysome sizes the codons near the 5’ end of the mRNA control protein synthesis rate, at intermediate polysome sizes different configurations of codons along the length of the mRNA control protein synthesis rate, and at high polysome sizes the codons near the 3’ end of the mRNA control protein synthesis rate.

(iii) The relative position of codons along the mRNA determines the optimal protein synthesis rate. Optimal translation rates of mRNAs are controlled by segments of rate limiting codons that are sequence specific. The segments of rate limiting codons correspond to regions of high translation time that cause nonuniform ribosome distributions on mRNAs.

Chapter 4: A genome scale analysis of the relative translational efficiencies of *E. coli* genes

4.1 Introduction

The genetic code is degenerate, with 61 codons encoding 20 amino acids. Hence, most amino acids are encoded by multiple, or synonymous, codons. Preferential usage of more abundant, or major, synonymous codons over less abundant, or minor, synonymous codons has been observed in the genomes of many species, with codon usage patterns having been correlated with gene expression levels (41), with position within a gene (42-44), and with gene length (45). In *E. coli* the frequencies of synonymous codons are proportionate to the relative abundance of the concentrations of the tRNAs that recognize them (34, 35). Synonymous codons are not translated at the same rate (10), with higher translation rates observed for major codons than for minor codons (35). The relationship between synonymous codon frequencies and tRNA abundances, along with the relationship between translation rate and synonymous codon frequency suggest that the bias toward major synonymous codons is to increase the speed and accuracy of translation (43, 46, 47).

In this chapter we apply our detailed, mechanistic framework of protein translation to all of the genes in the *E. coli* genome. We determine translation rates and the fractions of the mRNAs covered by translating ribosomes. We also perform a sensitivity analysis to determine the codons that are rate limiting to protein synthesis for all the genes. We utilize these data to calculate *relative translational efficiencies*. We define the relative translational efficiency of a gene to be the measure of the tradeoff between the overall translation rate of the gene and the number of ribosomes needed to achieve that rate. We observe that efficient genes achieve high translation rates using ribosomes economically while inefficient genes do not. Additionally,

efficient genes have their rate limiting codons near the start codons (5' ends) of the genes, while inefficient genes have their rate limiting codons near the stop codons (3' ends) of the genes. Along these lines, we find that efficient genes have an overrepresentation of the minor arginine (Arg) AGA and AGG codons near the 5' ends of the genes, and inefficient genes have an overrepresentation of the AGA and AGG codons near the 3' ends of the genes. These results suggest that minor codons play a role in optimizing translation rate. Moreover, these results present an important implication in the usage of minor codons for the optimized, systemic allocation of ribosomes in the translation of mRNAs throughout the cell.

4.2 Computational studies

We utilize our mathematical model of protein synthesis from Chapter 3 and the sensitivity analysis discussed in Chapter 3 and APPENDIX B to investigate the steady state translation properties of all the genes in the *E. coli* genome (48). The values for the concentration of each mRNA species *r*, M_r, the free ribosome concentration, $R^{(f)}$, the free ternary complex concentrations, $T_k^{(f)}\left(k \in K\right)$, and the free Ef-G concentration, $G^{(f)}$, applied in these studies, along with the reaction rate constants, k_1, k_{-1}, k_2, k_{-2}, k_3, k_4, k_5, k_6, k_7, k_{-7}, k_8, k_9, are the same as those used in Chapters 2 and 3. We hypothesized that at any given growth condition the cell maximizes the protein production rates from each of its mRNAs. Hence, we examine the protein synthesis properties of each gene corresponding to its maximum specific protein synthesis rate, i.e., the protein synthesis rate per mRNA molecule. The method used to calculate the maximum specific protein synthesis rates of mRNAs is the same as that from Chapters 2 and 3. We found in Chapters 2 and 3 that the maximum specific protein synthesis rate of a given mRNA species occurs under completely elongation limited conditions. Similarly, in this chapter

the protein synthesis properties determined corresponding to the maximum specific protein synthesis rate of each mRNA also occur under completely elongation limited conditions, such that:

$$C^{v}_{k_{E,r}} = \sum_{n=1}^{N_r - 1} C^{v}_{k^{eff}_{E,n,r}} = 1 \qquad (4.1)$$

Consequently the protein synthesis properties of mRNAs we determine under these conditions are not influenced by the translation initiation and termination phases and are based *only* on the sequences of codons translated in the elongation phase.

4.2.1 Overall protein synthesis properties

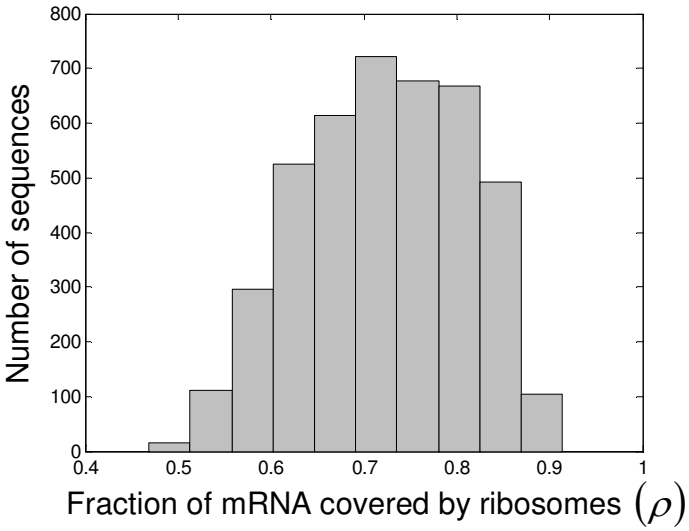
In these studies we apply our mathematical model to determine polysome sizes and translation rates for all the genes in the *E. coli* genome. We apply the sensitivity analysis to find the positions of rate limiting codons in all of the sequences.

Polysome size and translation rate

We consider the polysome sizes of all the mRNAs by introducing the ribosomal fractional coverage, ρ, i.e., the fraction of the mRNA molecule covered by translating ribosomes, as defined mathematically in Chapter 3. Included in Figure 4.1 is the ribosomal fractional coverage for all the mRNAs. We observe large variation in ribosomal fractional coverage, with values between 0.47 and 0.91. The average ribosomal fractional coverage is 0.72, with a standard deviation of 0.088. This result is approximately the same as a result that has been determined experimentally in a recent study with yeast cells (49). The polysome size of almost every mRNA was quantified, and the average ribosomal fractional coverage was determined to be 0.71 with a standard deviation of 0.081. Our result, combined with the result from the yeast study (49),

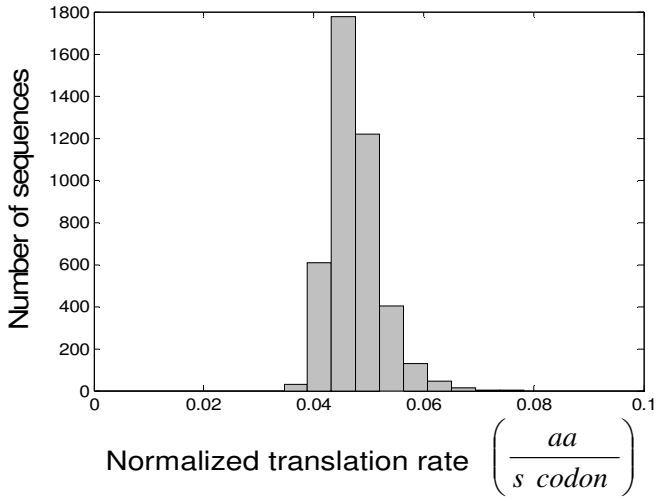
imply that the number of ribosomes engaged in translation may be systemically optimized throughout cells in a way that is well conserved among living organisms.

Figure 4.1: Distribution of the fractions of mRNAs covered by ribosomes



We also consider the translation rates of all the mRNAs. Included in Figure 4.2 is the normalized translation rate, i.e., the translation rate of the sequence divided by the number of codons in the sequence, for each of the genes. Unlike our observation regarding the ribosomal fractional coverages, we observe little variation in the normalized translation rates, with values between $0.03 \ \frac{aa}{s \ codon}$ and $0.08 \ \frac{aa}{s \ codon}$. We also observe no correlation between the normalized translation rates and the ribosomal fractional coverages. These results indicate that protein synthesis rates may not necessarily be directly proportionate to the number of ribosomes bound to the mRNAs, and suggest that the common assumptions that translation rates and ribosomes bound to the mRNAs are directly proportional and that the elongation rates are the same for all codons (49) can lead to erroneous conclusions.

Figure 4.2: Distribution of normalized translation rates



Rate limiting codon segments

To elucidate the rate limiting effects of individual codons on translation, we examined the control coefficient with respect to the elongation step, $C^{v}_{k^{eff}_{E,n,r}}$, occurring at each codon along the length of every mRNA. For each mRNA we observe that all of the control over translation rate is dominated by a segment of rate limiting codons (Figure 4.3A), such that:

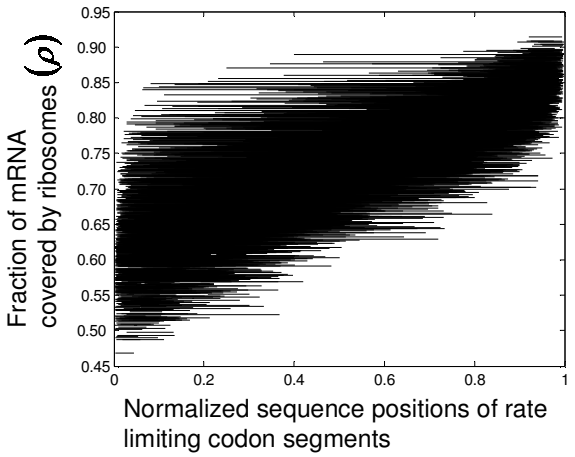
$$C^{v}_{k_{E,r}} = \sum_{n=n^{seg}_{5'}}^{n^{seg}_{3'}} C^{v}_{k^{eff}_{E,n,r}} = 1 \qquad (4.2)$$

where $n^{seg}_{5'}$ is the position of the 5' end of the rate limiting codon segment and $n^{seg}_{3'}$ is the position of the 3' end of the rate limiting codon segment, and this result has also been observed in Chapter 3. The positions of the rate limiting codon segments correlate positively with the fraction of mRNAs covered by ribosomes (Figure 4.3A), which is expected because the closer to the 3' end of the mRNA the rate limiting codon segment is positioned, the more translating ribosomes can be accommodated on the mRNA, and hence the higher the corresponding ribosomal fractional coverage is. The rate limiting codon segments can occur at all sequence positions, with no bias observed with respect to the normalized positions of the centers of the

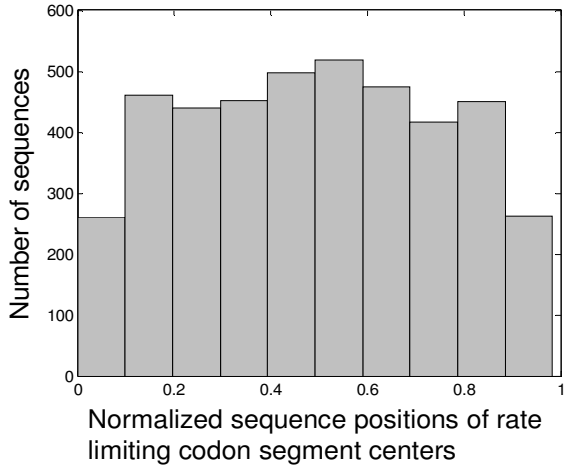
rate limiting codon segments (Figure 4.3B). It is also important to note that most of the rate limiting codon segments are between 40 and 60 codons long (Figure 4.3C), so the rate limiting codon segments that appear to cover entire mRNAs in Figure 4.3A correspond to short mRNAs with lengths approximately equal to the length of a typical rate limiting codon segment.

Figure 4.3: (A) Relationship between fractions of mRNAs covered by ribosomes and normalized positions of rate limiting codon segments (B) Distribution of normalized positions of rate limiting codon segment centers (C) Distribution of rate limiting codon segment lengths

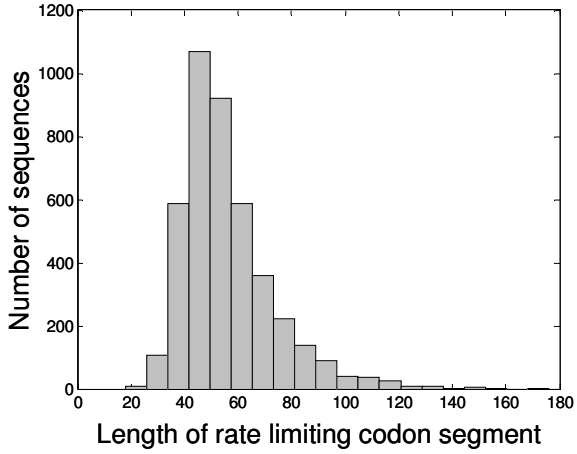
A



B



C



4.2.2 Systemic relationships between relative translational efficiency and sequence properties

In these studies we consider effects of silent mutations, i.e., mutations in the sequence of nucleotides that cause changes in synonymous codons without causing changes in the corresponding amino acids, on all of the genes in the *E. coli* genome. We apply our mathematical model to the mutated genes, and utilize the protein synthesis properties of the mutated sequences obtained from our model to determine their corresponding relative translational efficiencies. Since our studies are performed under steady state conditions, it holds that the initiation rate, the elongation rate of all the codons, and the termination rates are equal to each other:

$$V_{I,r} = V_{E,n,r} = V_{T,r} \qquad ,n \in [1, N_r - 1] \tag{4.3}$$

Taking the termination rate, $V_{T,r}$, to be the steady state protein synthesis rate, we define $\hat{V}_r$ to be the protein yield per ribosome of gene *r*, such that:

$$\hat{V}_r = \frac{V_{T,r}}{\sum_{n=1}^{N_r - 1} S_{ij,n,r}} \tag{4.4}$$

We simulate silent mutations for all the genes by replacing the codon at each position along the length of the mRNA with the most abundant synonymous codon. If the most abundant synonymous codon already occupies a given position we do not make a replacement. Applying our mathematical model to the mutated sequences to obtain the steady state translation rates and ribosome distributions, it follows from equations (4.3) and (4.4) that the protein yield per ribosome of gene *r*, $\hat{V}_r^*$, is expressed as:

$$\hat{V}_r^* = \frac{V_{T,r}^*}{\sum_{n=1}^{N_r - 1} S_{ij,n,r}^*} \tag{4.5}$$

where the superscript * denotes protein synthesis properties determined from the mutated sequences. Dividing the protein synthesis rate per ribosome for the original gene by that for the mutated gene, we obtain the relative translational efficiency of gene *r*:

$$\eta_r = \frac{\hat{V}_r}{\hat{V}_r^*} \tag{4.6}$$

Relative translational efficiency and ribosome economics

We observe a broad distribution of relative translational efficiencies, varying from 0.50 to 1.58, with 600 genes having relative translational efficiencies greater than one (Figure 4.4). We also investigated ribosome utilization by introducing the ribosome density ratio, $\hat{\rho}_r$:

$$\hat{\rho}_r = \frac{\rho_r}{\rho_r^*} \tag{4.7}$$

where ρ_r^* is the fraction of mRNA *r* covered by ribosomes determined after making the silent mutations. The ribosome density ratio varies inversely with relative translational efficiency (Figure 4.5). Hence, the greater the relative translational efficiency, the better the balance between high protein synthesis rate and economical ribosome utilization is. This result is consistent with the positioning of the centers of the rate limiting codon segments in the most and least efficient genes (Figure 4.6). The positioning of the rate limiting codon segment centers is strongly biased near the 5' end of the mRNA in the most efficient genes (Figure 4.6A), and strongly biased near the 3' end of the mRNA in the least efficient genes (Figure 4.6B). Consequently, because the positions of the rate limiting codons correlate positively with ribosomal fractional coverages (Figure 4.3A), efficient genes achieve higher translation rates at lower polysome sizes than inefficient genes. Given that a common hypothesis for synonymous codon usage is that codon choice is to maximize translation rate (43, 46, 47), these results

suggest that codon usage is to optimize other or additional objectives such as economical ribosome utilization. Additionally, by calculating the average relative translational efficiencies of genes belonging to gene function categories as defined by the NCBI (48) (Table 4.1), we do not find any noticeable differences in relative translational efficiency with respect to gene function. This result implies that if codon choice is to optimize either translation rate, economical ribosome usage, or other objectives, then it is not done at the level of gene function.

Figure 4.4: Distribution of relative translational efficiencies

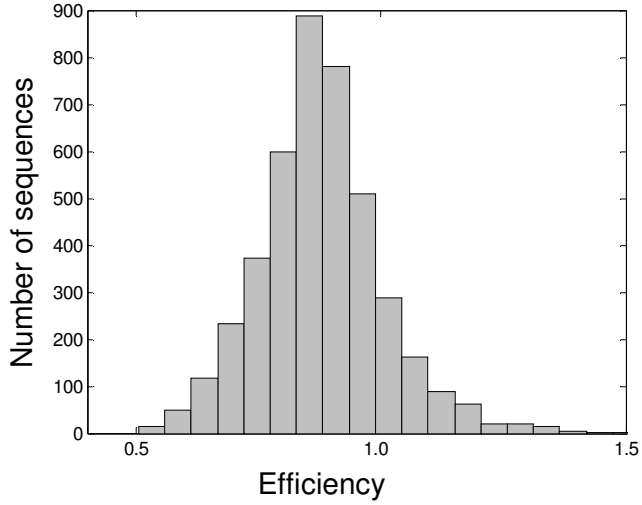


Figure 4.5: Relationship between ribosome density ratio and relative translational efficiency

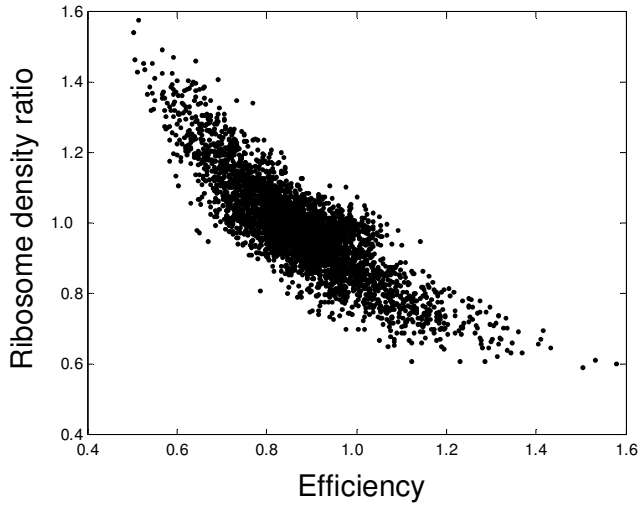
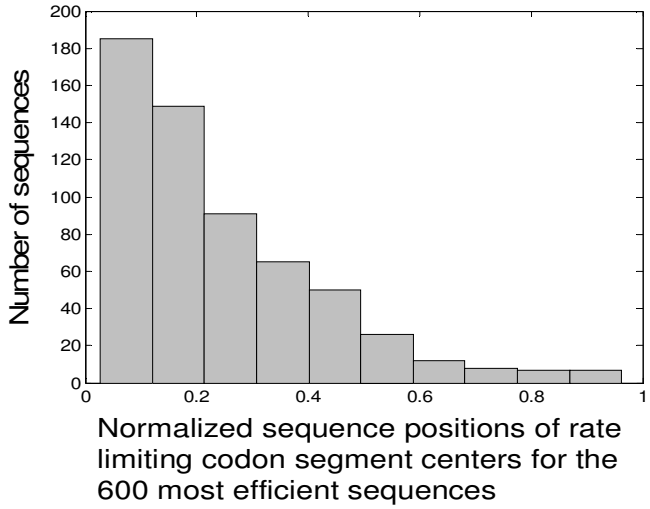


Figure 4.6: (A) Distribution of normalized positions of rate limiting codon segment centers for the 600 most efficient sequences (B) Distribution of normalized positions of rate limiting codon segment centers for the 600 least efficient sequences

A



B

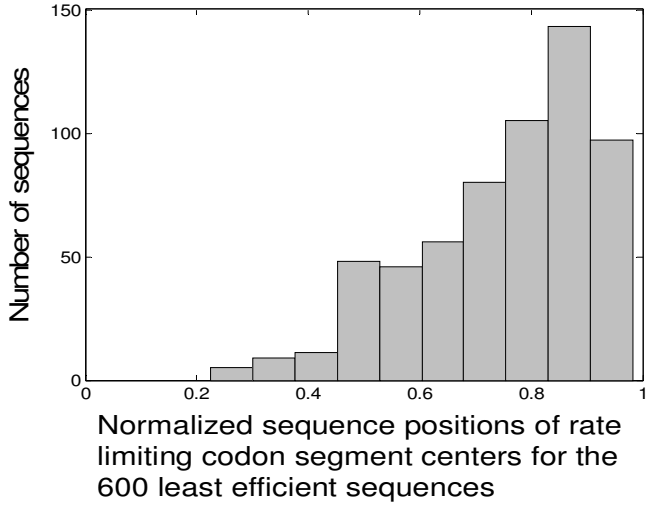


Table 4.1: Relative translational efficiencies of various gene function categories*

Gene function category	Number of genes	Average relative translational efficiency	Standard deviation
RNA processing and modification	2	0.89	0.026
Energy production and conversion	302	0.89	0.14
Cell cycle control, cell division, chromosome partitioning	45	0.91	0.11
Amino acid transport and metabolism	454	0.87	0.13
Nucleotide transport and metabolism	94	0.89	0.13
Carbohydrate transport and metabolism	417	0.88	0.13
Coenzyme transport and metabolism	153	0.86	0.12
Lipid transport and metabolism	107	0.86	0.12
Translation, ribosome structure and biogenesis	187	0.92	0.13
Transcription	337	0.87	0.12
Replication, recombination and repair	234	0.86	0.14
Cell wall/membrane/envelope biogenesis	244	0.89	0.14
Cell motility	110	0.87	0.12
Posttranslational modification, protein turnover, chaperones	147	0.90	0.14
Inorganic ion transport and metabolism	302	0.87	0.13
Secondary metabolites biosynthesis, transport and catabolism	80	0.86	0.14
General function prediction only	538	0.87	0.13
Function unknown	318	0.87	0.11
Signal transduction mechanisms	193	0.88	0.13
Intracellular trafficking, secretion, and vesicular transport	130	0.88	0.14
Defense mechanisms	51	0.87	0.13

*Gene function categories are as defined by the NCBI (48)

Codon usage bias and relative translational efficiency

In order to understand the differences in the positioning of the rate limiting codon segment centers with respect to relative translational efficiency, we investigated variations in the codon usage bias of different regions of the mRNAs. We define the initiation region of the mRNA to be the first 30 codons after the start codon near the 5' end, and the termination region of the mRNA to be the last 30 codons before the stop codon near the 3' end. We quantify codon usage bias in the initiation and termination regions of all the *E. coli* genes by introducing the factors F_j^{ini} and F_j^{ter}, respectively, for each codon species *j*, such that:

$$F_j^{ini} = \log\left(\frac{f_j^{ini}}{f_j^{mRNA \neq ini}} \right) \tag{4.8}$$

where f_j^{ini} is the combined frequency of codon species *j* in the initiation regions of all of the mRNAs and $f_j^{mRNA \neq ini}$ is the combined frequency of codon species *j* in all of the mRNAs excluding the initiation regions, and:

$$F_j^{ter} = \log\left(\frac{f_j^{ter}}{f_j^{mRNA \neq ter}} \right) \tag{4.9}$$

where f_j^{ter} is the combined frequency of codon species *j* in the termination regions of all of the mRNAs and $f_j^{mRNA \neq ter}$ is the combined frequency of codon species *j* in all of the mRNAs excluding the termination regions. We repeat the calculation in equation (4.8) to obtain $F_j^{ini,+}$ and $F_j^{ini,-}$, which are the codon usage biases in the initiation regions of the 600 most and least efficient genes, respectively, and we repeat the calculations in equation (4.9) to obtain $F_j^{ter,+}$ and $F_j^{ter,-}$, which are the codon usage biases in the termination regions of the 600 most and least

efficient genes, respectively. Initiation region codon biases are included in Table 4.2 and termination region codon biases are included in Table 4.3.

Table 4.2: Codon biases in the initiation regions of all the genes

Amino acid	Codon species j	F_j^{ini}	$F_j^{ini,+}$	$F_j^{ini,-}$	Amino acid	Codon species j	F_j^{ini}	$F_j^{ini,+}$	$F_j^{ini,-}$
Ala	GCU	0.20	0.17	0.24	Leu	CUU	0.46	0.64	0.37
	GCC	-0.01	0.54	-0.49		CUC	0.18	0.36	0.09
	GCA	0.15	0.29	0.05		CUA	0.37	0.82	-0.06
	GCG	-0.26	-0.35	-0.26		CUG	-0.18	-0.22	-0.19
Arg	CGU	-0.13	-0.20	-0.11	Lys	AAA	0.14	0.16	0.12
	CGC	-0.11	-0.15	-0.09		AAG	0.32	0.36	0.36
	CGA	0.39	0.30	0.40	Met	AUG	-0.04	-0.06	0.06
	CGG	0.07	0.38	-0.17	Phe	UUU	0.28	0.33	0.36
	AGA	0.97	1.45	0.81		UUC	-0.12	-0.27	0.19
	AGG	1.04	1.62	0.82	Pro	CCU	0.32	0.23	0.32
Asn	AAU	0.15	0.31	0.19		CCC	0.39	0.52	0.30
	AAC	-0.17	-0.30	-0.15		CCA	-0.01	0.19	0.02
Asp	GAU	-0.21	-0.23	-0.17		CCG	-0.47	-0.53	-0.38
	GAC	-0.27	-0.33	-0.20		AGU	0.27	0.15	0.38
Cys	UGU	0.27	0.44	0.15		AGC	-0.01	0.00	0.04
	UGC	0.07	0.05	0.01	Ser	UCU	0.33	0.27	0.45
Gln	CAA	0.18	0.38	0.15		UCC	0.10	0.10	0.09
	CAG	-0.18	-0.32	-0.16		UCA	0.43	0.34	0.46
Glu	GAA	-0.18	-0.22	-0.26		UCG	-0.06	-0.17	-0.07
	GAG	-0.22	-0.24	-0.35	Thr	ACU	0.27	0.21	0.38
Gly	GGU	-0.28	-0.33	-0.21		ACC	-0.08	-0.17	0.02
	GGC	-0.31	-0.40	-0.29		ACA	0.60	0.67	0.65
	GGA	0.23	0.38	0.14		ACG	-0.02	-0.14	0.04
	GGG	-0.14	-0.06	-0.25	Trp	UGG	-0.07	-0.15	0.04
His	CAU	0.02	0.03	0.00	Tyr	UAU	-0.12	-0.28	-0.13
	CAC	0.00	-0.08	0.21		UAC	-0.24	-0.65	-0.22
Ile	AUU	0.12	0.11	0.08	Val	GUU	0.11	-0.07	0.18
	AUC	-0.02	-0.08	0.13		GUC	0.03	0.15	0.09
	AUA	0.82	0.94	0.76		GUA	0.14	0.20	0.10
Leu	UUA	0.52	0.90	0.28		GUG	-0.37	-0.44	-0.42
	UUG	0.14	0.16	0.11					

Table 4.3: Codon biases in the termination regions of all the genes

Amino acid	Codon species j	F_j^{ter}	$F_j^{ter,+}$	$F_j^{ter,-}$	Amino acid	Codon species j	F_j^{ter}	$F_j^{ter,+}$	$F_j^{ter,-}$
Ala	GCU	0.03	0.11	-0.05	Leu	CUU	0.00	-0.18	0.14
	GCC	-0.13	-0.47	0.16		CUC	-0.17	-0.31	-0.02
	GCA	0.11	0.13	0.09		CUA	0.02	-0.13	0.24
	GCG	-0.06	-0.03	-0.07		CUG	-0.12	-0.12	-0.16
Arg	CGU	0.09	0.09	0.14	Lys	AAA	0.21	0.16	0.20
	CGC	0.04	0.04	-0.08		AAG	0.50	0.50	0.31
	CGA	0.42	0.41	0.47	Met	AUG	-0.10	0.03	-0.12
	CGG	0.37	0.16	0.68	Phe	UUU	-0.09	-0.03	-0.05
	AGA	0.84	0.71	1.01		UUC	-0.07	-0.04	-0.13
	AGG	0.85	0.32	1.06	Pro	CCU	0.01	0.02	-0.17
Asn	AAU	0.04	0.04	0.05		CCC	-0.04	0.09	-0.11
	AAC	-0.13	-0.08	-0.04		CCA	0.01	-0.06	0.14
Asp	GAU	-0.03	-0.07	-0.08		CCG	-0.14	-0.05	-0.21
	GAC	-0.03	-0.10	-0.07		AGU	0.02	0.05	0.10
Cys	UGU	0.09	0.18	0.18		AGC	-0.04	0.09	-0.12
	UGC	0.02	-0.11	0.19	Ser	UCU	-0.09	-0.10	-0.09
Gln	CAA	0.17	0.07	0.22		UCC	-0.26	-0.15	-0.21
	CAG	0.05	0.17	-0.01		UCA	0.11	0.06	0.14
Glu	GAA	0.10	0.08	0.04		UCG	-0.14	-0.20	-0.16
	GAG	0.23	0.20	0.16	Thr	ACU	-0.02	0.04	-0.15
Gly	GGU	-0.19	-0.20	-0.24		ACC	-0.23	-0.21	-0.28
	GGC	-0.26	-0.19	-0.27		ACA	0.14	0.16	0.06
	GGA	0.17	0.13	0.18		ACG	0.01	0.16	-0.01
	GGG	0.02	0.19	-0.18	Trp	UGG	0.03	0.07	0.06
His	CAU	0.08	0.18	-0.01	Tyr	UAU	0.08	0.05	0.13
	CAC	0.01	0.04	0.07		UAC	-0.06	-0.01	-0.16
Ile	AUU	-0.05	-0.08	-0.05	Val	GUU	-0.01	0.06	0.00
	AUC	-0.14	-0.11	-0.19		GUC	-0.08	-0.15	-0.06
	AUA	0.31	0.32	0.48		GUA	0.05	-0.04	0.02
Leu	UUA	0.10	-0.07	0.27		GUG	-0.03	-0.04	-0.10
	UUG	0.09	0.15	0.08					

We observe that all the genes have an overrepresentation of the minor Arg AGA and AGG codons in the initiation regions. An overrepresentation of AGA and AGG codons has been observed in the initiation regions of *E. coli* genes in previous work (37), and insertion and removal of AGA and AGG codons in the initiation regions of mRNAs have pronounced effects

on their translation (37, 50). The initiation region bias is highest for the most efficient genes and lowest for the least efficient genes. We also observe that all the genes have an overrepresentation of AGA and AGG in the termination regions of the genes. The termination region bias is highest for the least efficient genes and lowest for the most efficient genes. The codons AGA and AGG are the least used codons in *E. coli*, are recognized by some of the least abundant tRNAs (34, 35), and have pronounced effects on the translation rates of genes (37, 50). These results imply that AGA and AGG codons play a role in the relative translational efficiencies of sequences. Because the positioning of the rate limiting codon segment centers is strongly biased near the 5' end of the mRNA in the most efficient genes (Figure 4.6A), and strongly biased near the 3' end of the mRNA in the least efficient genes (Figure 4.6B), these results also imply that AGA and AGG codons may also play a role in the way ribosomes are allocated for translation of mRNAs throughout cells.

4.3 Discussion

In this work we apply our detailed, mechanistic framework of protein translation summarized in Chapter 3 to all of the genes in the *E. coli* genome. We determined large variation in the fractions of the mRNAs covered by translating ribosomes (Figure 4.1) and small variation in the normalized translation rates (Figure 4.2) of all the genes. These results, together with our observation that ribosomal fractional coverage and normalized translation rate do not correlate, suggest that the overall translation rates of mRNAs are not necessarily proportionate to the number of bound ribosomes and that the varying rate limiting effects of different codons may play a role in the protein synthesis properties of mRNAs. By performing a sensitivity analysis, we determine the codons that are rate limiting to protein synthesis for all the genes. We observe that the positions of the rate limiting codons correlate positively with the ribosomal fractional

coverages (Figure 4.3A) and that there is no bias with respect to the positioning of the rate limiting codons in the genes (Figure 4.3B). We simulate silent mutations to determine the relative translational efficiencies of the genes (Figure 4.4), and determine that efficient genes use ribosomes the most economically (Figure 4.5). We also find that efficient genes have their rate limiting codons near the 5' ends of the sequences (Figure 4.6A) and an overrepresentation of the minor AGA and AGG codons near the 5' ends (Table 4.2), while inefficient genes have their rate limiting codons near the 3' ends of the sequences (Figure 4.6B) and an overrepresentation of AGA and AGG near the 3' end (Table 4.3).

The results from this study present an important implication in the usage of minor codons for the optimized, systemic allocation of ribosomes in the translation of mRNAs throughout the cell. Due to the positioning of the rate limiting codons near the 5' ends of the mRNAs in efficient genes and near the 3' ends of the mRNAs in inefficient genes, efficient genes achieve high translation rates at low polysome sizes, and inefficient genes achieve low translation rates at high polysome sizes. The overrepresentation of AGA and AGG, which are the least used codons in *E. coli* and have pronounced effects on the translation rates of genes (37, 50), in the rate limiting regions of the efficient and inefficient genes suggest that AGA and AGG play an important role in the relationship between the relative translational efficiency and ribosome utilization on mRNAs. Moreover, our finding that the average ribosomal fractional coverage is 0.72, with a standard deviation of 0.088, is approximately the same as the average ribosomal fractional coverage of 0.71 and standard deviation of 0.081 that have been determined experimentally in a recent study with yeast cells (49). These results imply that minor codons may play a role in the systemic optimization of the number of ribosomes engaged in translation throughout cells in a way that may be well conserved among living organisms. Given that the translation mechanism

is well conserved among living organisms and that it is energetically expensive, with 80 % of the cell's energy devoted to it (13), these results also indicate that ribosomes may be allocated for translation in a way that is most energetically and thermodynamically favorable for the cell.

It is important to emphasize that we make no assumptions *a priori* about the translation properties of individual codons. In previous modeling studies of translation (44, 51-53), a common assumption is that the translation rate of each codon depends on a single rate constant that is the same for all codons on the mRNA and is directly proportionate to the tRNA concentration recognizing the codon. In our mechanistic framework we account for all the elementary translation steps that occur at every codon along the length of the mRNA, so the protein synthesis properties we observe for the genes in this study are emergent from all the intermediate steps, along with the concentrations of all the translational components, working together in concert. However, it is important to point out that the tRNA concentrations applied to our mechanistic framework in this study were determined experimentally (36) under optimal cellular growth conditions and at high growth rate, and we have not considered effects of variations in ternary complex concentrations on the protein synthesis properties of genes. A finding in a recent modeling study (54) is that the charged levels of tRNA isoacceptors, i.e., different tRNA species that carry the same amino acid, differ when *E. coli* cells are starved of that amino acid, causing changes in the translation rates of the codons the isoacceptors recognize under amino acid starvation conditions. A finding from an experimental study (37) is that the minor AGA and AGG codons have no negative effects on gene expression in exponentially growing cells, but begin to negatively affect gene expression once cell growth passes mid-log phase. To explain these observations the authors of this study (37) propose the AGA/AGG modulator hypothesis, where the transcriptional control of tRNA recognizing AGA and AGG

becomes tighter after cell growth passes mid-log phase, making its concentration limited. Hence, given that ternary complex concentrations change with growth condition (54) and cell cycle phase (37), and these changes affect translation rates, the relative translational efficiencies of genes determined in this study and the way resources are systemically allocated throughout the cell for translation can potentially also vary with growth condition and cell cycle phase.

4.4 Conclusions

In this chapter we applied the expanded mechanistic framework and sensitivity analysis from Chapter 3 to investigate the protein synthesis properties of all the genes in the *E. coli* genome. We determined the following:

(i) There is large variation in the fractions of the mRNAs covered by translating ribosomes, but small variation in the normalized translation rates of all the genes.

(ii) The positions of the rate limiting codons correlate positively with the ribosomal fractional coverages, and there is no bias with respect to the positioning of the rate limiting codons when all of the genes are considered.

(iii) The greater the relative translational efficiency of a gene, the better its balance between high protein synthesis rate and economical ribosome utilization is.

(iv) Efficient genes have their rate limiting codons near the 5' ends of the sequences and an overrepresentation of the minor AGA and AGG codons near the 5' ends, while inefficient genes have their rate limiting codons near the 3' ends of the sequences and an overrepresentation of AGA and AGG near the 3' end.

Chapter 5: Conclusions

5.1 Novel hypotheses emerging from this thesis research

In this thesis we present a deterministic, sequence specific mechanistic framework for protein synthesis that accounts for all of its elementary steps and perform a sensitivity analysis in order to investigate the steady state protein synthesis properties of mRNAs. The three overarching, novel hypotheses that emerge from this work are as follows:

(i) The translation rate at a given polysome size depends on the complex interplay between ribosomal occupancy of elongation cycle intermediate states and ribosome distributions with respect to codon position along the length of the mRNA, and this interplay leads to polysome self-organization that drives translation rate to maximum levels (Chapter 2).

(ii) Due to the complex interplay between the level of ribosomal crowding on the mRNA and the level of ternary complex competition for the ribosomal A site at each codon, the relative position of codons along the mRNA determines the optimal protein synthesis rate, and the translation rates of mRNAs are controlled by segments of rate limiting codons that are sequence specific (Chapter 3).

(iii) Less abundant, or minor, codons play a role in optimizing translation rate, and the usage of minor codons is important to the optimized, systemic allocation of ribosomes in the translation of mRNAs throughout the cell (Chapter 4).

While some of the conclusions drawn from our studies might be as expected to those experienced with protein synthesis, the computational framework presented in this thesis provides a quantitative verification and allows the formulation of hypotheses for the origins of the observed phenomena that mental simulations alone cannot offer. Throughout this work we

incorporated information about discreet intermediate steps in our framework to make quantitative predictions about the translation mechanism. These mathematical models allow us to consider each part of the complex biological process and to develop a more complete understanding of translation at the systems level.

5.2 Collaborative research efforts

Our collaborators in the Uhlenbeck group have studied the structure and function of tRNA for many years. In ongoing research Kevin Keegan, who is a graduate student jointly advised by Olke Uhlenbeck and Vassily Hatzimanikatis, is developing an *in vitro* translation system of short peptide sequences in order to test major computational results discussed in this thesis experimentally.

5.3 Recommendations for future work

5.3.1 Effects of mRNA competition for common translational resources in cells

Throughout this thesis we assume that the total concentrations of ribosomes, ternary complexes, and Ef-G:GTP free to participate in translation are constant. To further investigate allocation of cellular resources and the translational output of genes, our mechanistic framework can be expanded to account for the simultaneous translation of multiple mRNA species. Previous experimental studies on the relative changes in mRNA and protein levels in response to an environmental and/or genetic perturbation (55, 56) have shown a nonlinear, not *one-to-one*, relation between mRNA and protein expression. Also, previous computational studies suggest that this nonlinear relation is due to system-wide competition for ribosomes and tRNAs (6, 7). By applying our mechanistic framework to the simultaneous translation of multiple mRNA species, we will be able to understand how the coupling between ribosomal states due to shared translational components relates to system-wide properties.

5.3.2 Stochastic studies

The work in this thesis is based on deterministic models. However, translation networks are composed of small numbers of macromolecules, making randomness of molecular encounters and fluctuations in macromolecule levels important factors in protein synthesis (57). Reconciling randomness requires stochastic simulations, and our deterministic mechanistic framework can be adapted to stochastic modeling frameworks. Previous studies demonstrate that proteins are produced in random and sharp bursts and that phenotypic noise is essentially determined at the translational level (58, 59). Hence, with our mechanistic framework modeled in terms of stochastic simulations the effects of every elongation step on stochastic fluctuations in protein levels can be investigated. Such studies will provide insight on control the elongation cycles occurring at different codons have over phenotypic noise at the level of translation. It has also been hypothesized that noise is an evolvable trait that can be optimized to balance fidelity and diversity in gene expression (60). The expansion of the mechanistic framework presented in this thesis to stochastic models will allow for the study of correlations between codon usage and phenotypic noise levels in translation.

5.3.3 Application of the mechanistic framework for protein synthesis to other organisms

The mechanistic framework for protein synthesis presented in this thesis was applied to the translation of *E. coli* mRNAs. Because protein synthesis is well conserved among organisms, our mechanistic model can be readily applied to the translation of mRNAs in other species. The rapid sequencing of genomes of many organisms, along with current high throughput technologies revealing cellular usage of the translational machinery, provide a wealth of information that must be integrated at the systems level to fully characterize genome scale translation networks. The mathematical model of protein synthesis presented in this thesis will

provide a framework to integrate this information and help us understand how individual mechanistic steps and translational machinery allocation contribute to emergent systems properties of many organisms. Such studies will allow us to understand how species compare systemically at the level of translation, providing insight on the evolution of sequences at the level of translation from a systems perspective.

APPENDIX

APPENDIX A: The MG-HR model of the translation mechanism

In this model, the dynamics of protein synthesis from the mRNA of species *r* with length N_r codons described by (N_r+1) differential equations of the form:

$$\frac{d(M_r x_{n,r})}{dt} = V_{I,r} - V_{n,r} \qquad n=1 \tag{A1}$$

$$\frac{d(M_r x_{n,r})}{dt} = V_{n-1,r} - V_{n,r} \qquad n=[2, N_r] \tag{A2}$$

where $M_r x_{n,r}$ is the concentration of codons occupied by the front of the ribosome, $x_{n,r}$ is the probability of each codon *n* on mRNA *r* being occupied by the front of the ribosome, and $V_{n-1,r}$ and $V_{n,r}$ are the rates of ribosome movement from codon *n*-1 to *n* and from *n* to *n*+1, respectively. The initiation rate is described as follows:

$$V_{I,r} = k_{I,r} R^{(f)} W_{I,r} M_r \tag{A3}$$

where $W_{I,r}$ is the probability that the initiation site is free $\left(W_{I,r} = 1 - \sum_{n=1}^{L} x_{n,r}\right)$. The free ribosome concentration $\left(R^{(f)}\right)$is a function of the total ribosome and mRNA species concentrations, along with the occupation probabilities of each codon on every mRNA $\left(R^{(t)} - \sum_r M_r \sum_{n=1}^{N_r+1} x_{n,l}\right)$. The rates of movement of the ribosomes are described by the following equations:

$$V_{n,r} = \begin{cases} k_E x_{n,r} W_{n+1,r} M_r & 1 \le n \le N_r - L + 1 \\ k_E x_{n,r} M_r & N_r - L + 2 \le n \le N_r \end{cases} \tag{A4}$$

where k_E is the elongation constant and $W_{n+1,r}$ denotes the conditional probability that codon

n+1 is free given that codon n is occupied, formulated as $W_{n+1,r} = \left\{ \left(1 - \sum_{n=1}^{L} x_{n+1,r} \right) \Big/ \left(1 - \sum_{n=1}^{L-1} x_{n+1,r} \right) \right\}$. The rate of termination is expressed as:

$$V_{T,r} = k_{T,r} x_{N_r+1,r} M_r \qquad \text{(A5)}$$

where $k_{T,r}$ is the termination rate constant.

The main difference between our model and the MG-HR model is that our model accounts for all intermediate steps of the elongation cycle, while the MG-HR model treats the elongation cycle as a single step that is dependent on a single elongation rate constant (k_E) that is the same for all codons. In our model we use the position of the ribosomal P site on the mRNA as the reference position of the ribosome, while the MG-HR model uses the position of the front of the ribosome on the mRNA as the reference position of the ribosome.

APPENDIX B: Sensitivity analysis

Sensitivities are quantified using the concentration control coefficients, $\boldsymbol{C}_{\boldsymbol{p}}^{\boldsymbol{x}_i}$, and flux control coefficients, $\boldsymbol{C}_{\boldsymbol{p}}^{\boldsymbol{v}}$, which are defined as fractional concentration and flux changes, respectively, with respect to fractional input parameter changes (25). In this model input parameters include total ribosome, ternary complexes, Ef-G complexes, and mRNA concentrations, along with reaction rate constants. Steady state mass balances can be expressed with the following equation:

$$\boldsymbol{N}_{\boldsymbol{R}}\boldsymbol{v}\left(\boldsymbol{x}_i, \boldsymbol{x}_d\left(\boldsymbol{x}_i, \boldsymbol{p}_t\right), \boldsymbol{p}_r\right) = 0 \tag{B1}$$

where $\boldsymbol{N}_{\boldsymbol{R}}$ is the stoichiometric matrix consisting only of rows corresponding to independent species, $\boldsymbol{x}_i$ is the vector of independent species concentrations $\left(S_{ij,n,r}^{(\sigma)}, S_r^T\right)$, and $\boldsymbol{x}_d$ is the vector of dependent species concentrations $\left(R^{(f)}, T_k^{(f)}, G^{(f)}, C_{n,r}^{(f)}\right)$. The vector of fluxes is represented by $\boldsymbol{v}$, the vector of total species concentrations $\left(R^{(t)}, T_k^{(t)}, G^{(t)}, M_r\right)$ is represented by $\boldsymbol{p}_t$, and the vector of reaction rate constants $\left(k_1, k_{-1}, k_2, k_{-2}, k_3, k_4, k_5, k_6, k_7, k_{-7}, k_8, k_9, k_{I,r}, k_{T,r}\right)$ is represented by $\boldsymbol{p}_r$. Using the established (log)linear model formalism (28, 61, 62), we can linearize the system around the steady state and derive expressions for the control coefficients:

$$\boldsymbol{C}_{\boldsymbol{p}}^{\boldsymbol{x}_i} = -\left(\boldsymbol{N}_{\boldsymbol{R}}\boldsymbol{V}\boldsymbol{E}_{\boldsymbol{i}} + \boldsymbol{N}_{\boldsymbol{R}}\boldsymbol{V}\boldsymbol{E}_{\boldsymbol{d}}\boldsymbol{Q}_{\boldsymbol{i}}\right)^{-1}\left[\boldsymbol{N}_{\boldsymbol{R}}\boldsymbol{V}\boldsymbol{E}_{\boldsymbol{d}}\boldsymbol{Q}_{\boldsymbol{t}} \vdots \boldsymbol{N}_{\boldsymbol{R}}\boldsymbol{V}\boldsymbol{\Pi}\right] \tag{B2}$$

$$\boldsymbol{C}_{\boldsymbol{p}}^{\boldsymbol{v}} = \left(\boldsymbol{E}_{\boldsymbol{i}} + \boldsymbol{E}_{\boldsymbol{d}}\boldsymbol{Q}_{\boldsymbol{i}}\right)\boldsymbol{C}_{\boldsymbol{p}}^{\boldsymbol{x}_i} + \left[\boldsymbol{E}_{\boldsymbol{d}}\boldsymbol{Q}_{\boldsymbol{t}} \vdots \boldsymbol{\Pi}\right] \tag{B3}$$

where $\boldsymbol{p}$ is the generalized parameter set $\left(\boldsymbol{p} = \left[\boldsymbol{p}_t \vdots \boldsymbol{p}_r\right]\right)$, $\boldsymbol{V}$ is the diagonal matrix whose elements are the steady state fluxes; $\boldsymbol{E}_{\boldsymbol{i}}$ and $\boldsymbol{E}_{\boldsymbol{d}}$ are the matrices of elasticities with respect to independent and dependent species, respectively, defined as the local sensitivities of fluxes to species concentrations; $\boldsymbol{Q}_{\boldsymbol{i}}$ is the matrix that quantifies the relative abundance of dependent species with

respect to the abundance of the independent ones, and a second matrix; $\boldsymbol{Q}_t$, is the matrix that quantifies the relative abundance of the dependent species with respect to the levels of their corresponding total moieties; $\boldsymbol{\Pi}$ is the matrix of elasticities with respect to reaction rate constants, defined as the local sensitivities of fluxes to these values. Similar to the Summation Theorem (25), we can show that the sum of the control coefficients with respect to the reaction rate constants is equal to one:

$$C_{k_I}^{v} + C_{k_E}^{v} + C_{k_T}^{v} = 1 \tag{B4}$$

A detailed derivation of the Summation Theorem is included in section 5.3.1 of (63). The control coefficient $C_{k_E}^{v}$ in the MG-HR formulation is equal to the fractional change in flux with respect to the fractional change in the elongation reaction rate constant, k_E, of every codon, while $C_{k_E}^{v}$ in our formulation is equal to the sum of the flux control coefficients with respect to the elongation cycle intermediate step reaction rate constants, expressed as:

$$C_{k_E}^{v} = C_{k_1}^{v} + C_{k_{-1}}^{v} + C_{k_2}^{v} + C_{k_{-2}}^{v} + C_{k_3}^{v} + C_{k_4}^{v} + C_{k_5}^{v} + C_{k_6}^{v} + C_{k_7}^{v} + C_{k_{-7}}^{v} + C_{k_8}^{v} + C_{k_9}^{v} \tag{B5}$$

APPENDIX C: Ribosome, ternary complex, Ef-G:GTP, and mRNA abundances

The total mRNA concentration in *E. coli* is roughly $6.64\,\mu M$ (32), and there are 4,237 known protein coding regions in the *E. coli* genome (48). Hence, we estimate the concentration, M_r, of a given mRNA species *r* by assuming equal mRNA levels for all genes, dividing the total mRNA concentration by the number of genes, and obtaining a concentration of $0.0016\,\mu M$. The total concentration of ribosomes in *E. coli* is roughly $30\,\mu M$ (32) and 80 % of the ribosomes are engaged in translation throughout the cell (64). Therefore, the total ribosome concentration engaged in translation is estimated to be $24\,\mu M$, and the total ribosome concentration free to participate in translation $\left(R^{(t)}\right)$ is $6\,\mu M$ $\left(G^{(t)}\right)$ is also present in a one-to-one ratio with ribosomes (13), and therefore the total Ef-G concentration is assumed to be $30\,\mu M$. Because Ef-G is bound to the ribosome only at states 8 and 9 $\left(S_{ij,n,r}^{(8)}, S_{ij,n,r}^{(9)}\right)$, we assume that the cellular concentration of ribosomes occupying these states combined is negligibly small. Results of the computational studies confirm that this assumption is reasonable. Therefore, we assume the total Ef-G:GT(D)P concentration free to participate in translation $\left(G^{(f)}\right)$ to be the total Ef-G concentration $\left(G^{(t)}\right)$.

The total concentration of tRNA in *E. coli* is roughly $332\,\mu M$ (32). Relative concentrations of different tRNA species taken from (36), along with the total tRNA concentration, are used to estimate the concentration of each tRNA species. To determine total concentrations of free tRNAs available for ribosomal A site binding $\left(T_k^{(t)}\right)$, tRNA concentrations participating in elongation cycles throughout the cell must be accounted for and subtracted from total tRNA concentrations. In order to estimate cellular concentrations of tRNAs sequestered in ribosomal A and P sites, we approximate concentrations of ribosomes participating in translation with A and P sites occupying different codon species pairs by multiplying the estimated cellular

concentration of ribosomes participating in translation ($24\,\mu M$) by the fractional codon pair frequencies in the protein coding regions of the *E. coli* genome from (65). We determine the maximum tRNA concentration to be $23.13\,\mu M$, the minimum tRNA concentration to be $0.37\,\mu M$, and the average tRNA concentration to be $6.31\,\mu M$. The free tRNA concentrations determined are equal to the total ternary complex concentrations because we assume all free tRNAs are in the form of ternary complexes (***Assumption 3***).

APPENDIX D: Mechanistic framework assumptions

We have applied the following assumption in our current mechanistic formulation:

All ternary complex species can bind to the ribosomal A site during the codon independent binding intermediate step of the elongation cycle, regardless of the codon species present in the ribosomal A site.

Introducing the ternary complex subscript *k* to the fluxes and state corresponding to nonspecific binding yields: $V^{(1)}_{k,ij,n,r}$, $V^{(-1)}_{k,ij,n,r}$, and $S^{(2)}_{k,ij,n,r}$, and denotes the nonspecific binding between each ternary complex species *k* and A site codon species *j*. Detailed descriptions of ribosomal states and fluxes can be found in Chapter 2. The equations describing the dynamics of the transitions of state 1, the state existing prior to ternary complex binding, are as follows:

$$\frac{dS^{(1)}_{ij,n,r}}{dt} = V_{I,r} + \sum_k \left(V^{(-1)}_{k,ij,n,r} - V^{(1)}_{k,ij,n,r} \right) \qquad , n = 1 \qquad \text{(D1)}$$

$$\frac{dS^{(1)}_{ij,n,r}}{dt} = V^{(9)}_{ij,n-1,r} + \sum_k \left(V^{(-1)}_{k,ij,n,r} - V^{(1)}_{k,ij,n,r} \right) \qquad , n \in \left[2, N_r - 1\right] \qquad \text{(D2)}$$

We assume that the ternary complexes that do not recognize the A site codon cannot proceed past the nonspecific binding intermediate step of the elongation cycle, while ternary complexes recognizing the A site codon can continue on to the remaining steps of the elongation cycle. These assumptions yield the following expressions for the dynamics of the transitions of state 2:

$$\frac{dS^{(2)}_{k,ij,n,r}}{dt} = V^{(-1)}_{k,ij,n,r} - V^{(1)}_{k,ij,n,r} \qquad , n \in \left[1, N_r - 1\right], k \neq j \qquad \text{(D3)}$$

$$\frac{dS^{(2)}_{k,ij,n,r}}{dt} = V^{(1)}_{k,ij,n,r} + V^{(-2)}_{ij,n,r} - V^{(2)}_{ij,n,r} - V^{(-1)}_{k,ij,n,r} \qquad , n \in \left[1, N_r - 1\right], k = j \qquad \text{(D4)}$$

The expressions describing the dynamics of the transitions between the remaining elongation cycle intermediate states are the same as those described in Chapter 2. Equations (D1) – (D4),

together with the expressions for the remaining intermediate states, are used to derive the expression for the effective elongation rate constant accounting for ternary complex competitive binding (equation 3.5) in the same manner the original effective elongation rate constant (equation 2.28) was derived in Chapter 2.

APPENDIX E: Michaelis – Menten reaction rate expression derivation

In the absence of ternary complex competitive binding we consider the following reaction scheme for the elongation cycle occurring at a given codon:

$$T_j^{(f)} + S_{ij,n,r}^{(1)} \underset{k\text{-}1}{\overset{k_1}{\rightleftarrows}} S_{ij,n,r}^{(2)} \xrightarrow{k_2} S_{ij,n,r}^{(3)} \xrightarrow{k_3} S_{ij,n,r}^{(4)} \xrightarrow{k_4} S_{ij,n,r}^{(5)} \xrightarrow{k_5} S_{ij,n,r}^{(6)} \xrightarrow{k_6} S_{ij,n,r}^{(7)} + G^{(f)} \underset{k\text{-}7}{\overset{k_7}{\rightleftarrows}} S_{ij,n,r}^{(8)} \xrightarrow{k_8} S_{ij,n+1,r}^{(9)} \xrightarrow{k_9} \quad \text{(E1)}$$

The states $S_{ij,n,r}^{(1)} - S_{ij,n+1,r}^{(9)}$ represent the intermediate elongation cycle ribosomal states that are described in detail in Chapter 2. The first state, $S_{ij,n,r}^{(1)}$, represents the ribosomal state that exists prior to ternary complex binding with the A site empty, and the remaining states have the A site occupied by the ternary complex. We allow $S_{ij,n,r}^{(M)}$ to be the grouped ribosomal state including all the intermediate elongation cycle states having the ternary complex bound to the ribosomal A site, where:

$$S_{ij,n,r}^{(M)} = S_{ij,n,r}^{(2)} + S_{ij,n,r}^{(3)} + S_{ij,n,r}^{(4)} + S_{ij,n,r}^{(5)} + S_{ij,n,r}^{(6)} + S_{ij,n,r}^{(7)} + S_{ij,n,r}^{(8)} + S_{ij,n+1,r}^{(9)} \quad \text{(E2)}$$

By introducing the grouped state, $S_{ij,n,r}^{(M)}$, the reaction scheme in equation (E1) simplifies to:

$$T_j^{(f)} + S_{ij,n,r}^{(1)} \underset{k\text{-}1}{\overset{k_1}{\rightleftarrows}} S_{ij,n,r}^{(M)} \xrightarrow{k_M} \quad \text{(E3)}$$

Since our studies are performed at steady state, we obtain the expression for k_M as we obtained the expression for the effective elongation rate constant in Chapter 2, yielding:

$$k_M = \frac{1}{U_{n,r}\left(\alpha_2 + \alpha_3 + \alpha_4 + \alpha_5 + \alpha_6 + \alpha_7 + \alpha_9\right) + \alpha_8} \quad \text{(E4)}$$

In this work k_M is evaluated with $U_{n,r} = 1$.

The equation describing the dynamics of the transitions between states is as follows:

$$\frac{dS_{ij,n,r}^{(M)}}{dt} = k_1 T_j^{(f)} S_{ij,n,r}^{(1)} - k_{-1} S_{ij,n,r}^{(1)} - k_M S_{ij,n,r}^{(M)} \quad \text{(E5)}$$

Following from the pseudo steady state approximation, the concentrations of the intermediates are assumed to reach steady state much faster than those of the product and substrate. Hence, we set the time derivative in the above equation equal to zero and rearrange it to obtain an expression for $S_{ij,n,r}^{(M)}$, yielding:

$$S_{ij,n,r}^{(M)} = \frac{k_1 T_j^{(f)} S_{ij,n,r}^{(1)}}{\left(k_{-1} + k_M\right)} \quad \text{(E6)}$$

By allowing $K_M = \frac{k_{-1} + k_M}{k_1}$ (equation 3.19) the above equation becomes:

$$S_{ij,n,r}^{(M)} = \frac{T_j^{(f)} S_{ij,n,r}^{(1)}}{K_M} \quad \text{(E7)}$$

The total concentration of translating ribosomes is equal to the sum of the concentration of ribosomes with the A site empty, $S_{ij,n,r}^{(1)}$, and the concentration of ribosomes with a ternary complex bound to the A site, $S_{ij,n,r}^{(M)}$. We assume that the concentration of translating ribosomes with codon species *j* in the A site, R_j, is equal to the concentration of ribosomes participating in translation in an *E. coli* cell (estimated to be $24\,\mu M$ in APPENDIX C) multiplied by the frequency of codon species *j* in the *E. coli* genome. Also, we assume that R_j is constant and can be expressed as:

$$R_j = S_{ij,n,r}^{(1)} + S_{ij,n,r}^{(M)} \quad \text{(E8)}$$

Rearranging the above equation and applying it to equation (E7) yields:

$$S_{ij,n,r}^{(M)} = R_j \frac{1}{1+\frac{K_M}{T_j^{(f)}}} \tag{E9}$$

Because the amino acid incorporation rate is equal to $k_M S_{ij,n,r}^{(M)}$, it can be shown that:

$$v_{MM,j} = k_M R_j \frac{1}{1+\frac{K_M}{T_j^{(f)}}} \tag{E10}$$

which is equivalent to equation (3.18).

Under ternary complex competitive binding conditions we consider the following reaction scheme:

$$\begin{array}{c} S_{k,ij,n,r}^{(2)} \\ k \neq j \\ k_1 \uparrow \downarrow k_{-1} \end{array}$$

$$T_j^{(f)} + S_{ij,n,r}^{(1)} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} S_{ij,n,r}^{(2)} \xrightarrow{k_2} S_{ij,n,r}^{(3)} \xrightarrow{k_3} S_{ij,n,r}^{(4)} \xrightarrow{k_4} S_{ij,n,r}^{(5)} \xrightarrow{k_5} S_{ij,n,r}^{(6)} \xrightarrow{k_6} S_{ij,n,r}^{(7)} \underset{k_{-7}}{\overset{k_7}{\rightleftarrows}} S_{ij,n,r}^{(8)} \xrightarrow{k_8} S_{ij,n+1,r}^{(9)} \xrightarrow{k_9} \tag{E11}$$

$$+ \; G^{(f)}$$

$$+ \; T_k^{(f)}, \quad k \neq j$$

A similar derivation to the one presented above for noncompetitive binding conditions yields equation (3.21) for competitive binding conditions.

REFERENCES

1. Harms, J. M., H. Bartels, F. Schlunzen, and A. Yonath. 2003. Antibiotics acting on the translational machinery. J Cell Sci 116:1391-1393.

2. Lee, J. M. 2003. The role of protein elongation factor eEF1A2 in ovarian cancer. Reproductive Biology and Endocrinology 1:69.

3. MacDonald, C. T., Gibbs, J.H. 1969. Concerning kinetics of polypeptide synthesis on polyribosomes. Biopolymers 7:707-725.

4. MacDonald, C. T., J. H. Gibbs, and A. C. Pipkin. 1968. Kinetics of biopolymerization on nucleic acid templates. Biopolymers 6:1-5.

5. Heinrich, R., and T. A. Rapoport. 1980. Mathematical modelling of translation of mRNA in eucaryotes; steady state, time-dependent processes and application to reticulocytes. J Theor Biol 86:279-313.

6. Mehra, A., K. H. Lee, and V. Hatzimanikatis. 2003. Insights into the relation between mRNA and protein expression patterns: I. Theoretical considerations. Biotechnol Bioeng 84:822-833.

7. Mehra, A., and V. Hatzimanikatis. 2006. An algorithmic framework for genome-wide modeling and analysis of translation networks. Biophys J 90:1136-1146.

8. Varenne, S., J. Buc, R. Lloubes, and C. Lazdunski. 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.

9. Curran, J. F., and M. Yarus. 1989. Rates of aminoacyl-tRNA selection at 29 sense codons in vivo. J Mol Biol 209:65-77.

10. Soerensen, M. A., and 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. Journal of Molecular Biology 222:265-280.

11. Kruger, M. K., S. Pedersen, T. G. Hagervall, and M. A. Sorensen. 1998. The modification of the wobble base of tRNAGlu modulates the translation rate of glutamic acid codons in vivo. J Mol Biol 284:621-631.

12. Zouridis, H., and V. Hatzimanikatis. 2007. A model for protein translation: polysome self-organization leads to maximum protein synthesis rates. Biophys J 92:717-730.

13. Hershey, J. 1987. Protein Synthesis. In Escherichia coli and Salmonella typhimuriium: Cellular and Molecular Biology. F. C. Neidhardt, Ingraham, J.L., Low, K.B., Magasanik,

B., Schaechter, M., Umbarger, H.E., editor. American Society for Microbiology, Washington, D.C. 613-647.

14. Miller, O. L., Jr., B. A. Hamkalo, and C. A. Thomas, Jr. 1970. Visualization of bacterial genes in action. Science 169:392-395.

15. Arava, Y., Y. Wang, J. D. Storey, C. L. Liu, P. O. Brown, and D. Herschlag. 2003. Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 100:3889-3894.

16. Kazazian, H. H., Jr., and M. L. Freedman. 1968. The characterization of separated alpha and beta-chain polyribosomes in rabbit reticulocytes. J Biol Chem 243:6446-6450.

17. Rose, J. K. 1977. Nucleotide sequences of ribosome recgonition sites in messenger RNAs of vesicular stomatitis virus. Proc Natl Acad Sci U S A 74:3672-3676.

18. Revel, M., and Y. Groner. 1978. Post-transcriptional and translational controls of gene expression in eukaryotes. Annu Rev Biochem 47:1079-1126.

19. Dittmar, K. A., M. A. Sorensen, J. Elf, M. Ehrenberg, and T. Pan. 2005. Selective charging of tRNA isoacceptors induced by amino-acid starvation. EMBO Rep 6:151-157.

20. Louie, A., and F. Jurnak. 1985. Kinetic studies of Escherichia coli elongation factor Tu-guanosine 5'-triphosphate-aminoacyl-tRNA complexes. Biochemistry 24:6433-6439.

21. Bilgin, N., F. Claesens, H. Pahverk, and M. Ehrenberg. 1992. Kinetic properties of Escherichia coli ribosomes with altered forms of S12. J Mol Biol 224:1011-1027.

22. Rodnina, M. V., T. Pape, R. Fricke, L. Kuhn, and W. Wintermeyer. 1996. Initial binding of the elongation factor Tu.GTP.aminoacyl-tRNA complex preceding codon recognition on the ribosome. J Biol Chem 271:646-652.

23. Pape, T., W. Wintermeyer, and M. V. Rodnina. 1998. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome. Embo J 17:7490-7497.

24. Savelsbergh, A., V. I. Katunin, D. Mohr, F. Peske, M. V. Rodnina, and W. Wintermeyer. 2003. An elongation factor G-induced ribosome rearrangement precedes tRNA-mRNA translocation. Mol Cell 11:1517-1523.

25. Kacser, H., and J. A. Burns. 1973. The control of flux. Symp Soc Exp Biol 27:65-104.

26. Heinrich, R., and T. A. Rapoport. 1974. A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. Eur J Biochem 42:89-95.

27. Fell, D. A., and H. M. Sauro. 1985. Metabolic control and its analysis. Additional relationships between elasticities and control coefficients. Eur J Biochem 148:555-561.

28. Reder, C. 1988. Metabolic control theory: a structural approach. J Theor Biol 135:175-201.

29. Kholodenko, B. N., and H. V. Westerhoff. 1993. Metabolic channelling and control of the flux. FEBS Lett 320:71-74.

30. Hatzimanikatis, V., and J. E. Bailey. 1996. MCA has more to say. J Theor Biol 182:233-242.

31. Neidhardt, F. C. 1987. Escherichia coli and Salmonella typhimuriium: Cellular and Molecular Biology. American Society for Microbiology, Washington, D.C.

32. Sundararaj, S., A. Guo, B. Habibi-Nazhad, M. Rouani, P. Stothard, M. Ellison, and D. S. Wishart. 2004. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of Escherichia coli. Nucleic Acids Res 32:D293-295.

33. Underwood, K. A., J. R. Swartz, and J. D. Puglisi. 2005. Quantitative polysome analysis identifies limitations in bacterial cell-free protein synthesis. Biotechnol Bioeng 91:425-435.

34. 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.

35. Ikemura, T. 1981. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the E. coli translational system. J Mol Biol 151:389-409.

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

37. Chen, G. T., and M. Inouye. 1990. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the Escherichia coli genes. Nucleic Acids Res 18:1465 - 1473.

38. Kozak, M. 1984. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. Nature 308:241 - 246.

39. Bilgin, N., M. Ehrenberg, and C. Kurland. 1988. Is translation inhibited by noncognate ternary complexes? FEBS Lett 233:95-99.

40. LaRiviere, F. D., A. D. Wolfson, and O. C. Uhlenbeck. 2001. Uniform Binding of Aminoacyl-tRNAs to Elongation Factor Tu by Thermodynamic Compensation. Science 294:165 - 168.

41. Grantham, R., C. Gautier, M. Gouy, M. Jacobzone, and R. Mercier. 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res 9:r43-74.

42. Bulmer, M. 1988. Codon usage and intragenic position. J Theor Biol 133:67-71.

43. Bulmer, M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics 129:897-907.

44. Liljenstrom, H., and G. von Heijne. 1987. Translation rate modification by preferential codon usage: intragenic position effects. J Theor Biol 124:43-55.

45. Comeron, J. M., M. Kreitman, and M. Aguade. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila. Genetics 151:239-249.

46. Smith, N. G. C., and A. Eyre-Walker. 2001. Why are translationally sub-optimal synonymous codons used in Escherichia coli? J Mol Evol 53:225-236.

47. Akashi, H., and A. Eyre-Walker. 1998. Translational selection and molecular evolution. Current opinion in genetics and development 8:688-693.

48. NCBI. Escherichia coli K12, complete genome.

49. Arava, Y., F. E. Boas, P. O. Brown, and D. Herschlag. 2005. Dissecting eukaryotic translation and its control by ribosome density mapping. Nucleic Acids Res 33:2421-2432.

50. Rosenberg, A. H., E. Goldman, J. J. Dunn, F. W. Studier, and G. Zubay. 1993. Effects of consecutive AGG codons on translation in Escherichia coli, demonstrated with a versatile codon test system. J Bacteriology 175:716-722.

51. Lesnik, T., J. Solomovici, A. Deana, R. Ehrlich, and C. Reiss. 2000. Ribosome traffic in E. coli and regulation of gene expression. J Theor Biol 202:175-185.

52. Solomovici, J., T. Lesnik, and C. Reiss. 1997. Does Escherichia coli optimize the economics of the translation process? J Theor Biol 185:511-521.

53. Zhang, S., E. Goldman, and G. Zubay. 1994. Clustering of low usage codons and ribosome movement. J Theor Biol 170:339-354.

54. Elf, J., D. Nilsson, T. Tenson, and M. Ehrenberg. 2003. Selective charging of tRNA isoacceptors explains patterns of codon usage. Science 300:1718-1722.

55. Ideker, T., V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold, and L. Hood. 2001. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 292:929-934.

56. Lee, P. S., L. B. Shaw, L. H. Choe, A. Mehra, V. Hatzimanikatis, and K. H. Lee. 2003. Insights into the relation between mrna and protein expression patterns: II. Experimental observations in Escherichia coli. Biotechnol Bioeng 84:834-841.

57. Fedoroff, N., and W. Fontana. 2002. Genetic networks: Small numbers of big molecules. Science 297:1129-1131.

58. Ozbudak, E. M., M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. 2002. Regulation of noise in the expression of a single gene. Nat Genet 31:69-73.

59. Thattai, M., and A. van Oudenaarden. 2001. Intrinsic noise in gene regulatory networks. P Natl Acad Sci USA 98:8614-8619.

60. Raser, J. M., and E. K. O'Shea. 2004. Control of stochasticity in eukaryotic gene expression. Science 304:1811-1814.

61. Hatzimanikatis, V., Floudas, C.A., Bailey, J.E. 1996. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. AIChE Journal 42:1277-1292.

62. Wang, L., I. Birol, and V. Hatzimanikatis. 2004. Metabolic control analysis under uncertainty: framework development and case studies. Biophys J 87:3750-3763.

63. Heinrich, R., and S. Schuster. 1996. Metabolic control analysis. In The regulation of cellular systems. Springer. 138-291.

64. Bremer, H., and Dennis, P.P. 1996. Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli and Salmonella. F. C. Neidhardt, editor. American Society for Microbiology, Washington, D.C. 1553-1569.

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