Modeling mRNA Populations

R. Urquidi Camacho[a], N. Pollesch[b,1], M.A. Gilchrist[a,c,d,1,*]

[a] *Genome Science and Technology Program, University of Tennessee, Knoxville, TN 37996-XXX*
[b] *Department of Mathematics, University of Tennessee, Knoxville, TN 37996-1320*
[c] *Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1610*
[d] *National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996-3410*

Abstract

This paper presents a model to describe the dynamics of protein translation. A system of ordinary differential equations is derived to describe the number of ribosomes bound to a strand of mRNA at a given time. The number of ribosomes bound to an mRNA at a given time is referred to its ribosome load. The mRNA is classified based on its ribosome load and whether or not it's decapped for future degradation. Distribution of ribosome counts is assumed to be related to the translation initiation rate, translation completion rate, degredation marking rate, and length of the mRNA. The length of the mRNA's coding region plays the role of controlling the number of ribosome counts which, in turn, determines the number of ODEs in the system. A goal of this work is to see how the equilibrium distribution between classes as changes with coding region length. A closed form solution to the density in the i^{th} ribosomal class in a system with $i_{\max}$ states is presented for the equilibrium distribution of the decapped classes in terms of the capped classes. The equilibrium solutions in the capped classes are shown to be related to the full determinant of the tri-diagonal matrix used to describe the system, as well as all the determinants of the minors associated to it. In general, there is no closed form for the determinant of a tri-diagonal matrix, only a recurrence relation that can be used to find determinants. However, in this model a closed form exists for the full determinant as it changes with changing values of $i_{\max}$ and its formula is presented . This closed form for the determinant provides a method to efficiently find equilibrium solutions for the entire system. Additionally, a continuous approximation using PDE is derived and also used to find equilibrium solutions to the system. Both of these methods for determining equilibrium solutions are utilized in an effort to find the set of parameters that maximizes the likelihood of a given data set. A process for mapping the equilibrium model results to data is also presented and used to begin preliminary estimation of model parameters

*Corresponding author
Email address: mikeg@utk.edu (M.A. Gilchrist)

and to verify model function.

alternate abstract: Modeling Ribosomal Loading of mRNA

A model is presented to describe the dynamics of protein translation related to the ribosomal load of an mRNA. The number of ribosomes bound at a given time is referred to as ribosome load, and using this value a population of mRNA are classified. A system of ordinary differential equations (ODEs) is derived and solved for the equilibrium distribution of a population of mRNA. Distribution of ribosome counts is assumed to be related to the translation initiation rate, translation completion rate, degradation marking rate, and length of the mRNA. Methods are developed to find analytical equilibrium solutions to the system of ODEs and a system of partial differential equations (PDEs) are derived to find numerical approximations to the ODE system at equilibrium as well. Both the PDE continuous approximation and the analytical solutions to the ODE system agree offering two different methods for finding solutions at equilibrium within optimization routines. Additionally, a tool is developed and presented that is used to compare the model results to empirical microarray data measures of ribosome load.

Keywords: bioinformatics, mRNA population, protein translation, ribosome loading, ribosome count, polysome, mathematical model

---

Paper Outline

1. Motivation - **(Mike)**
   (a) Why is this process important?
   (b) What will this model enable researchers to do?
   (c) Other modeling efforts?
2. Derivation and Assumptions
   (a) Physical processes captured (Ideally, have a quick discussion of process and inline definitions of variables used to represent process, followed by a total recap in a table) - **(Nate)**
      i. System described as population model: Dichotomy of decapped and capped mRNA. State variables based on an mRNA's ribosome load.
      ii. Process of mRNA production
      iii. Process of Marking mRNA for degradation supposed
      iv. Three processes of : Initiation, translation, and completion
   (b) Definition/Discussion of system boundaries - (**MIKE**)
      i. Physical boundaries as a cell and relation to parameters
      ii. Discussion of perceived upper and lower limits to state variables and parameters
      iii. Temporal boundaries and relation to steady state
   (c) Assumptions: Such as initial assumptions of specific functional forms, i.e. marking rate constant among classes - **(Nate)**
   (d) Justify consideration of system as two subsystems, decapped and capped. - **(Nate)**
3. Model Formulation: Total model presented and then analysis of capped and decapped systems - **(Nate)**
   (a) ODE/Discrete system
      i. Present system of ODEs (Total, capped, and decapped)
      ii. Matrix Representation of ODE model (Total, capped, and decapped)
      iii. Steady state formulations
   (b) PDE/Continuous system

i. Explain motivation for deriving PDE

ii. Explain framing as 'non-linear birth and death process'

iii. Explain derivation using Taylor expansion

iv. Present PDE for capped class

v. Present non-dimensionalized system

vi. Present 2nd order ODE to be solved for non-dimensionalized PDE at Equilibrium

vii. Motivate and present equation for decapped class at equilibrium

viii. (Make decision to present results for steady state values for PDE here or in a separate section to follow)

4. Results - **(Nate)**

   (a) Present solution strategies/methods

      i. ODE/Discrete system: Matrix inversion technique

      ii. PDE/Continuous system: Numerical solver of 2nd order ODE that arises at equilibrium

      iii. Discussion of alternative solution approaches

   (b) Present actual solutions for a couple sets of parameters: Highlight agreement of ODE and PDE system

   (c) Present solutions for discrete system under further simplifications for translation and initiation

5. Opportunities for Future Research - **(Nate and Mike)**

   (a) Application of model to real data. Can highlight sources of data.

   (b) Alternate functional forms and relaxed assumptions

   (c) Further establish connection (in simplified system) to potential probability distributions

   (d) How to move forward with analytical solutions, specifically connection to solving 2nd order partial difference equation arising from tri-diagonal form of matrix, note here that boundary conditions exist that may be utilized which are not normally present.

1. Introduction

This section addresses such topics as why modeling this process important, what this model will enable researchers to do, and what other modeling efforts exist that seek to achieve the same goals.

1.1. mRNA and Translation

1. Gene expression short overview
   (a) Gene expression is often stated as the central dogma in which genetic information encoded in the DNA is transcribed into mRNA which is subsequently translated into protein.
   (b) Often, a greater amount of attention is focused on explaining gene expression at the transcriptional level and prevailing changes of mRNA transcript levels.
   (c) However, multiple studies across all kingdoms of life have shown that transcript expression level is only moderately predictive of the final protein expression.
   (d) Gene expression at the post transcriptional level is controlled by mRNA transcript stability and degradation, translation and protein maturation/degradation.
   (e) The model presented in this paper encompasses gene expression regulation occuring at the translational and the mature mRNA population level.
2. Biology controlling mRNA stability and translation
   (a) Mature mRNAs in the cytosol are called the free mRNA pool, and are in one of three states.
   (b) They are actively being translated by ribosomes and will continue to initiate new rounds of translation until the transcript is degraded.
   (c) Transcripts are degraded directly from the free mRNA pool.
   (d) Transcripts are protected from degradation by RNA binding protein chaperones or are found in processesing bodies awaiting translation initiation or degradation.
   (e) Degradation of mature mRNAs is controlled by numerous processes depending on whether they are bound to ribosome, in processing bodies or in the free mRNA pool.
   (f) Free mRNAs can be decapped or deadenylated followed by exonuclease digestion.
   (g) Ribosomes can destine transcripts to degradation under multiple conditions.
   (h) The first ribosome to bind to a freshly exported transcript performs the "pioneer round of translation", which is charged with assesing the mRNA's quality.

(i) There are 3 processes which occur in the pioneer round of translation, all of which detect different mRNA defects.

(j) No Go Decay (NGD) detects a stalled ribosome, either due to mRNA structural features, slowly translating sequence or inteference of translation elongation.

(k) No stop decay (NSD) detects a missing stop codon and nonsense mediated decay (NMD) detects potential mis splicing or nonsense mutations.

(l) All three decay mechanisms, NMD, NSD and NGD lead to the eventual degradation of their bound transcripts.

(m) While NSD and NMD are a restricted to the pioneering round of translation, NGD can also uccur during the following rounds of translation.

(n) As transcripts are cleared by the pioneering round of translations more ribosomes can attach to the transcript, once more than one ribosome is on a transcript this ribosome mRNA complex is called a polysome.

(o) Transcripts associated to ribosomes are generally assumed to be protected from degradation and only degraded once ribosomes are off the transcript, however both NGD, (sRNA silencing) and a process called cotranslational decay can degrade actively translated transcripts.

(p) Cotranslational decay involved the decapping of actively translating mRNA transcripts and subsequent 5' to 3' mRNA degradation which follows a 3 nucleotide periodic pattern in step with the Ribosome.

3. Current Models/Research and how our model fits in the current field

1. The basic representation of the central dogma dictates that epression of protein coding genes starts from genes encoded in DNA that are transcribed to mRNA and subsequently translated to Protein.

2. A more careful representation considers that the final protein production is dependent on both the maintenance of an actively translating mRNA population, the association of ribosomes on the population and finally the degradation of the protein itself.

3. The maintenance of mRNA populations relies on the balance of mRNA transcription rates, the translation status of transcripts and numerous mRNA decay pathways.

4. mRNA degradation relies on removing protective and translation enhancing componenets of the mRNA. These include the 5' mG cap and the 3' polyadenosine tail.

5. Additionally mRNA degradation can be promoted through endonucleolytic cleavage by RISC (and siRNAs).

6. mRNA degradation can occur in both a ribosomal associated or a ribosome free manner.

7. Ribosomal association of transcripts can lead to both protection of viable transcripts as well as quality control degradation of faulty transcripts.

8. When a viable transcript is bound by the ribosomal and translational machinery, the 5' cap is bound by translational initation factors and the 3' tail is bound by poly A binding proteins. This protects transcripts from exonucleic attack and degradation.

9. Endonucleic degradation is still possible, but reduced due to a reduced accessibility of the siRNA binding sequence on the transcript through competition with ribosomes.

10. However multiple mechanisms of mRNA decay are carried out in association with the ribosome. Nonsense mediated decay, no go decay and no stop decay all rely on ribosomes detecting faults in the transcript and susbsequently interacting with degradation machinery to remove the faulty transcript.

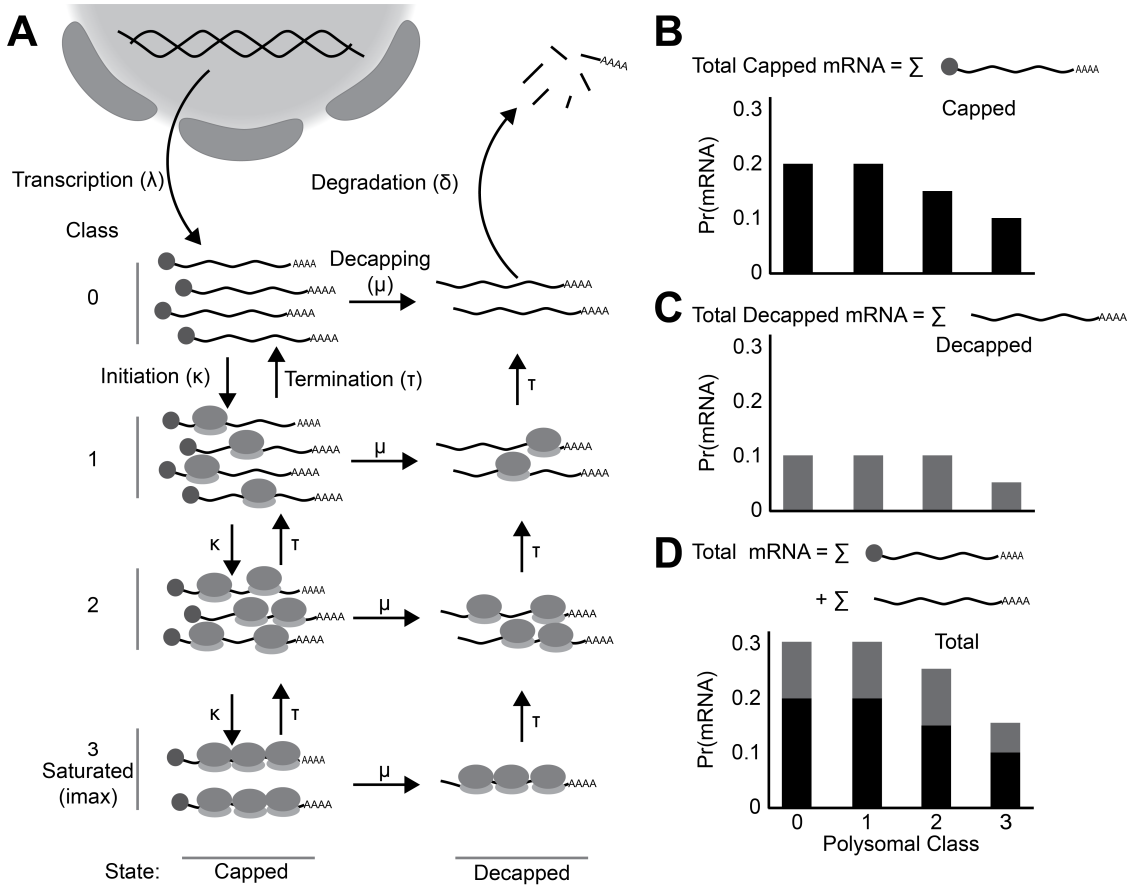
11. With some mechanisms of mRNA decay, decay can occur co-translationally. This is mainly seen in 5' decapping. When a translating transcript is decapped the 5' to 3' exonucleic degradation machinery trails the most upstream ribosome. As the ribosome translates the mRNA is degraded.

12.

2. Methods

2.1. Model Overview

Figure 1: Cartoon Representation of model in biological context. A) Model overview. Transcripts enter the sytem into the capped state at class 0 (no ribosomes bound). They enter the state at rate λ through transcription. Transcripts are free to move up and down ribosomal classes at rates κ for translation initiation and τ for elongation/termination. Transcripts can also be decapped and enter the decapped state at rate μ. Finally, upon reaching class 0 in the decapped state transcripts are fully degraded at rate δ. B) Probability of finding an mRNA in each class in the capped state. C) probability of finding an mRNA in each class in the decapped state. D) Joint probabilty of finding an mRNA in each class across each state. This reflect the total protein production potential.



The model captures some of the basic processes governing mRNA populations: transcript production, degradation and the process of translation (Figure 1A). Transcripts can exist in one of two states: capped and decapped which captures the the role of the 5' cap in mRNA protection and translation initiation. Capped transcripts are translationally competent, meaning that new ribosome can be loaded onto the transcript. Individual transcripts in the cell will be found with a set number of ribosomes (none, 1, 2, etc). The number of ribosomes on a transcript determines that transcripts polysomal class. The model seeks to determine how the population of transcripts of a single gene are distributed between ribosomal classes and capped and decapped states.

Transcripts enter into the model as defined by the transcription rate λ into the capped state with no ribosomes (polysome class 0) From capped class 0 a transcript can have two fates. The transcript can be decapped, thus marked for degradation at rate μ and move into the decapped class 0. Alternatively, a ribosome can initiate translation at rate $\kappa(1 - i/i_{\max})$ and be loaded onto the transcript and move it into capped class 1. Where i is the current transcript class and $i_{\max}$ is the maximal ribosomal occupancy on the transcript. The first term, κ is the average initiation rate on an empty transcript. For ribosomes initiating on transcripts already harboring ribosomes, this initiation rate is scaled to reflect the probability of a ribosome being present at or near the start codon. This attempts to account for the ribosomal density dependent effects on initiation and is called the density dependent initiation (DDI) model. As our model does not track ribosomal positions, we assume a uniform distribution of ribosomes across a transcript.

A ribosome on transcript can then elongate and terminate at a rate of $\tau' \times i$. After a ribosome fully elongates and terminates it leaves the transcript and the transcript falls to a lower class. The term τ' is calculated by using the average elongation rate τ_0 on a particular transcript divided by the length of the transcript $\tau' = \tau_0/\text{protein length in aa}$. As the number of ribosomes on a transcript increase, the probability of a ribosome being at the end of the transcript also increases, again following a uniform distribution. To better represent this mathematically, τ' can also be written as:

$$\tau' = \frac{\tau_0}{protein\ length} = \frac{\tau_0}{9 \times i_{\max}} \tag{1}$$

Where the saturated state of a transcript is denoted as $i_{\max}$, meaning that that transcript can no longer accept any more ribosomes. The factor nine arises from the average number of codons that is occupied by a ribosome. We can now formulate $\tau' \times i$ in the same way we formulated κ.

$$\tau' \times i = \frac{\tau_0}{9} \times \frac{i}{i_{\max}} \tag{2}$$

Where $\tau_0/9$ is the scaled elongation/termination rate, and $i/i_{\max}$ is the probability that a ribosome is at the end of a transcript.

Capped transcripts move through rounds of translation initiation and elongation-termination and distribute along the different polysomal classes. From any ribosomal class in the capped state the transcript can be decapped at rate μ and move into the decapped state while maintaining the same polysomal class. Decapped transcripts can no longer initiate new rounds of translation, but allow for currently loaded ribosomes to complete translation. This process represent co-translational decay, a common method of mRNA decay in eukaryotes (Hu 2009, Pelechano 2015, Collart 2020) After all ribosome complete translation, the mRNA is in decapped class 0 and completely degraded at a rate δ. The model produces two outputs. First, the total mRNA in each state and therefore the system (Figure 1B-D). Second, The distribution of the mRNAs in each mRNA in each ribosomal class. (Figure 1 B-D). The total protein output at steady state from our model can be obtained by calculating the average ribosomal class in the system by the total mRNA in the system (Figure 1D).

2.2. Formal Model Definition

We formalize the DDI model presented in Figure 1 by converting each state in to a series of ordinary differential equations (ODEs) representing the mRNA population for each polysomal class. The functional form of the capped subsystem is:

$$
\begin{aligned}
\frac{dm_0}{dt} &= \lambda + \frac{\tau_0}{9}\frac{1}{i_{\max}}m_1 - \left(\kappa_0\left(1-\frac{0}{i_{\max}}\right)+\mu\right)m_0 \\
\frac{dm_1}{dt} &= \kappa(0)m_0 + \frac{\tau_0}{9}\frac{2}{i_{\max}}m_2 - \left(\frac{\tau_0}{9}\frac{1}{i_{\max}} + \kappa_0\left(1-\frac{1}{i_{\max}}\right)+\mu\right)m_1 \\
&\vdots \\
\frac{dm_i}{dt} &= \kappa(i-1)m_{i-1} + \frac{\tau_0}{9}\frac{i+1}{i_{\max}}m_{i+1} - \left(\frac{\tau_0}{9}\frac{i}{i_{\max}} + \kappa_0\left(1-\frac{i-1}{i_{\max}}\right)+\mu\right)m_i \\
&\vdots \\
\frac{dm_{i_{\max}}}{dt} &= \kappa_0\left(1-\frac{i_{\max}-1}{i_{\max}}\right)m_{i_{\max}-1} - \left(\frac{\tau_0}{9}\frac{i_{\max}}{i_{\max}}+\mu\right)m_{i_{\max}}
\end{aligned}
\tag{3}
$$

Table 1: State variables and model parameters for ODE model of mRNA populations. Variable $i_{\max}$ is in the domain of non-negative integers; all other variables are non-negative real numbers.

Symbol	Description	Unit
State Variables		
m_i	Abundance of mRNAs with a ribosome load of i in capped state.	$mRNA$
m_i^*	Abundance of mRNAs with a ribosome load of i in decapped state.	$mRNA$
Model Parameters		
i	ribosomal load index	Ribosome
$i_{\max}$	Maximum number of ribosomes able to bind to mRNA; defines number of state variables and is a function of gene length.	Ribosome
$\kappa(i)$	Translation initiation rate for unmarked mRNAs with a ribosome load of i.	$1/s$
$\tau(i)$	Translation completion rate for the marked and unmarked mRNAs with a ribosome load of i.	$1/s$
$\mu(i)$	Marking rate for unmarked mRNAs with a ribosome load of i.	$1/s$
λ	Production rate of newly produced, ribosome free, and unmarked mRNA to the m_0 class.	$mRNA/s$
δ	Removal rate of marked mRNA with a ribosome load of 0 from the m_0^* class.	$1/s$

Similarly, the functional form of the decapped subsystem is:

$$\begin{aligned}
\frac{dm_0^*}{dt} &= \mu m_0 + \frac{\tau_0}{9}\frac{1}{i_{\max}} m_1^* - \delta m_0^* \\
\frac{dm_1^*}{dt} &= \mu m_1 + \frac{\tau_0}{9}\frac{2}{i_{\max}} m_2^* - \tau(1) m_1^* \\
&\vdots \\
\frac{dm_i^*}{dt} &= \mu m_i + \frac{\tau_0}{9}\frac{i+1}{i_{\max}} m_{i+1}^* - \tau(i) m_i^* \\
&\vdots \\
\frac{dm_{i_{\max}}^*}{dt} &= \mu m_{i_{\max}}^* - \frac{\tau_0}{9}\frac{i_{\max}}{i_{\max}} m_{i_{\max}}
\end{aligned} \tag{4}$$

A less constrained version of the model does not account for the DDI effects and doesn't scale κ by $(1 - i/i_{\max})$. This is the density independent initiation (DII) version of the model. Parameters and their units are fully defined in Table 1.

At equilibrium every $\tau(i)$ must equal every $\kappa(i)$ by definition. Only mRNAs decapped for degrada-

tion and with a ribosome load of 0 are removed from the system. Specifically, decapped mRNAs are removed from the 0^{th} class at a rate of $\delta\, m_0^*(t)$. All parameters are assumed to fixed for a given gene, but may vary between genes.

2.3. The density dependent initiation model

The DII model above is insensitive to the current ribosomal load on a transcript. But this doesn't mimic what happens in reality. Steric hindrance of recently initiated ribosomes prevent subsequent initiation. This might be one explanation for the sometimes observed "translational ramp", a short strecth of amino acids right after the start codon which translate slower than the remaining transcript (Verma 2019). Specifically, we assume the start codon must be unoccupied by a ribosome in order for translation initiation to be successful. As a consequence of this assumption, the probability of a ribosome occupying a given position on an mRNA with a ribosome load of i is simply $i/i_{\max}$. Thus, the probability the start codon is unoccupied is $1 - i/i_{\max}$ and, in turn, our translation initiation rate function can be defined as,

$$\kappa(i) = \kappa_0 \left(1 - \frac{i}{i_{\max}}\right), \tag{5}$$

where κ_0 is a gene specific parameter that describes the rate at which capped mRNAs encounter and are bound by ribosomes within the cytosol (i.e. it is an implicit function of the abundance of free ribosomes which we assume is constant).

Incorporating Equation 1 into the DII system yeild the density dependent intiation (DDI) model:

$$\begin{aligned}
\frac{dm_0}{dt} &= \lambda + \tau(1)m_1 - \left(\kappa_0 \left(1 - \frac{0}{i_{\max}}\right) + \mu(0)\right) m_0 \\
\frac{dm_1}{dt} &= \kappa(0)m_0 + \tau(2)m_2 - \left(\tau(1) + \kappa_0 \left(1 - \frac{1}{i_{\max}}\right) + \mu(1)\right) m_1 \\
&\vdots \\
\frac{dm_i}{dt} &= \kappa(i-1)m_{i-1} + \tau(i+1)m_{i+1} - \left(\tau(i) + \kappa_0 \left(1 - \frac{i-1}{i_{\max}}\right) + \mu(i)\right) m_i \\
&\vdots \\
\frac{dm_{i_{\max}}}{dt} &= \kappa_0 \left(1 - \frac{i_{\max} - 1}{i_{\max}}\right) m_{i_{\max}-1} - \left(\tau(i_{\max}) + \mu(i_{\max})\right) m_{i_{\max}}
\end{aligned}$$

and the decapped subsystem is unchanged.

2.4. Matrix-vector Formulation of ODE System

It is frequently useful to work with the matrix-vector formulation for a system of ODE. In this model, the dynamics of the decapped and capped mRNAs can be represented as,

$$\vec{M}' = \boldsymbol{F}\vec{M} + \vec{B}, \tag{6}$$

where $\vec{M} \in \mathbb{R}^{2(i_{\max}+1)}$ is a vector of all state variables, ordered here as m_0, m_1, ..., $m_{i_{\max}}$, m_0^*, m_1^*, ..., $m_{i_{\max}}^*$, $\vec{M}'$ is the vector containing the first derivatives of $\vec{M}$ with respect to time, $\boldsymbol{F} \in \mathbb{R}^{2(i_{\max}+1)\times 2(i_{\max}+1)}$ is the matrix representing the full system (Equation 7), and $\vec{B} \in \mathbb{R}^{2(i_{\max}+1)}$ is the vector of λ as the first component and 0s else. Using the functional forms presented above, matrix formulations are provided next.

As opposed to explicitly listing elements of the full system matrix-vector representation we found that it is more convenient to utilize the block structure that emerges in this system and explicitly provide the block components. The matrix $\boldsymbol{F}$ is block lower-diagonal and is given in Equation 7.

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{U} & \boldsymbol{0} \\ \\ \boldsymbol{\mu} & \boldsymbol{R} \end{pmatrix}. \tag{7}$$

The upper-left block, $\boldsymbol{U}$, corresponds to the capped state variables, where $\boldsymbol{U}$'s general form is provided in Equation 8. The upper-right block is a matrix of all zeros, $\boldsymbol{0} \in \mathbb{R}^{i_{\max}+1\times i_{\max}+1}$. Using $\boldsymbol{I}$ to represent the $i_{\max}+1 \times i_{\max}+1$ identity matrix, the lower-left block is $\boldsymbol{\mu} = \mu_0\boldsymbol{I}$, a diagonal matrix with the constant μ_0 on the diagonal and 0s else. The lower-right block, $\boldsymbol{R}$, corresponds to the decapped state variables and its form is provided in Equation 9.

The matrix $\boldsymbol{U}$ is $(i_{\max}+1 \times i_{\max}+1)$ dimensional and is tri-diagonal with non-zero entries on the

diagonal, super-, and sub-diagonals,

$$\boldsymbol{U} = \begin{pmatrix} -(\kappa_0+\mu_0) & \tau_0\frac{1}{i_{\max}} & & & & & & \\ \kappa_0 & \left(1-\frac{1}{i_{\max}}\kappa_0+\mu_0+\tau_0\frac{1}{i_{\max}}\right) & \tau_0\frac{2}{i_{\max}} & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & & 1-\frac{(i-1)}{i_{\max}}\kappa_0 & -\left(1-\frac{i}{i_{\max}}\kappa_0+\mu_0+\tau_0\frac{i}{i_{\max}}\right) & \tau_0\frac{i+1}{i_{\max}} & & \\ & & & & \ddots & \ddots & \ddots & \\ & & & & & \frac{1}{i_{\max}}\kappa_0 & -\left(\mu_0+\tau_0\frac{i_{\max}}{i_{\max}}\right) & \end{pmatrix} \tag{8}$$

In the representation given in Equation 8, all blank entries are 0. The $(i_{\max}-1)^{\text{th}}$ row has been suppressed in Equation 8, but it can be generated using the formula included for the i^{th} row.

The matrix $\boldsymbol{R}$ is the lower-right block in the block lower-diagonal matrix $\boldsymbol{F}$ (Equation 7),

$$\boldsymbol{R} = \begin{pmatrix} -\delta & \tau_0\frac{\text{`}}{i_{\max}} & & & & & \\ & -\tau_0\frac{1}{i_{\max}} & \tau_0\frac{2}{i_{\max}} & & & & \\ & & \ddots & \ddots & & & \\ & & & -\tau_0\frac{i-1}{i_{\max}} & \tau_0\frac{(i+1)}{i_{\max}} & & \\ & & & & \ddots & \ddots & \\ & & & & & -\tau_0\frac{(i_{\max}-2)}{i_{\max}} & \tau_0\frac{i_{\max}}{i_{\max}} \\ & & & & & & -\tau_0\frac{i_{\max}}{i_{\max}} \end{pmatrix}, \tag{9}$$

$\boldsymbol{R}$ is upper-diagonal with only non-zero entries on the diagonal and the super-diagonal.

2.4.1. Capped Subsystem Matrix-vector Representation

As a group the capped subsystem decouples from the decapped subsystem, as such the capped subsystem can be solved independently of the decapped subsystem. The matrix-vector formula representing the capped subsystem is

$$\vec{m}' = \boldsymbol{U}\vec{m} + \vec{b}, \tag{10}$$

where $\vec{m} \in \mathbb{R}^{i_{\max}+1}$ is the vector of capped state variables ordered m_0, ..., $m_{i_{\max}}$, $\vec{m}'$ is the vector containing the first derivatives of $\vec{m}$ with respect to time, $\boldsymbol{U} \in \mathbb{R}^{i_{\max}+1 \times i_{\max}+1}$ is the matrix representing the capped subsystem (Figure 8), and $\vec{b} \in \mathbb{R}^{i_{\max}+1}$ is the vector of λ as the first component and 0s else. With all equations defined for the full ODE system, include matrix-vector representations, the next section outlines methods for finding steady-state solutions to the system.

2.4.2. Capped state steady state solution

The capped system can be split into two components: Total transcripts in the capped state and how the transcripts are distributed across ribosomal classes. From manual exploration of model solutions of the capped state at low $i_{\max}$ values. We discovered that the capped class transcript number is determined by λ/μ If you take the simplest version of the model consisting of only the zeroth capped class.

$$\frac{dm_0}{dt} = \lambda + \mu m_0 \tag{11}$$

which, at equilibrium results in,

$$m_0 = \lambda/\mu \tag{12}$$

When the number of classes increases we find the the m_0 solution always has λ/μ factored out. As the m_0 solution propagates to higher classes all classes gain a λ/μ out front. This means you can factor out λ/μ from the whole system. This result makes logical sense as the overall transcript production rate into the capped state has to equal the marking rate out of it. For only one class $\lambda = \mu$. For mulitple classes, as the transcripts get distrbuted, each class contribute a weighted port of the total μ. Therefore, adding all the contributions togeter equals:

$$\frac{\lambda}{\mu} = \sum_{i=0}^{i_{\max}} m_i, \tag{13}$$

Where λ is only a scaling factor for the system as a whole. I.e. the distribution of transcripts across all classes is determined by κ, τ, μ and δ. μ affects both the total transcript abundance and the distribution of ribosomal classes across a particular species of transcript. First μ controls the rate of outflow from capped unto decapped, and second it shifts mRNAs to lower ribosomal classes. The solution to the system, as presented previously, can be expressed in the determinant-adjoint form:

$$labeleq : eq_general_adjugate_determinant_solution \vec{m} = -\frac{1}{\det[\boldsymbol{U}]} Adj[\boldsymbol{U}]\vec{b}.$$

As $\vec{b}$ is [λ 0 0 0 ... 0]. Only the first column of the adjoint matrix contributes to the result.

$$Adj[\boldsymbol{U}]\vec{b} = \lambda\vec{a}$$

and

$$\sum_{j=0}^{i_{\max}} \vec{a}_j = a_{tot}$$

With this we can factor our solution into two parts: 1) the total transcript abundance and 2) The distribution of transcript across the ribosomal classes.

$$\vec{m} = -\frac{\lambda a_{tot}}{\det[\boldsymbol{U}]}\frac{\vec{a}}{a_{tot}}$$

Where:

$$\frac{\vec{a}}{a_{tot}} = \vec{p}_m$$

The vector $\vec{p}_m$ sums to one and contains the probabilities of finding and mRNA in each class in the capped state. Now we are left with

$$\vec{m} = -\frac{a_{tot}}{\det[\boldsymbol{U}]}\,\lambda\vec{p}_m$$

If we sum across all classes to get the total mRNA population we find,

$$\sum_{i=0}^{i_{\max}} m_i = -\sum_{i=0}^{i_{\max}} \frac{a_{tot}}{\det[\boldsymbol{U}]}\,\lambda\vec{p}_m = -\frac{a_{tot}}{\det[\boldsymbol{U}]}\,\lambda = \frac{\lambda}{\mu}$$

$$-\frac{a_{tot}}{\det[\boldsymbol{U}]} = \frac{1}{\mu}$$

We finally arrive at,

$$\vec{m} = \frac{\lambda}{\mu}\vec{p}_m \tag{14}$$

The terms on the left hand side of the equation represent the total transcript population. The right hand side is the vector of probabilities, one entry for each class and is a function of κ, τ, and μ. This formulation has three interesting properties

First it gives a determinant free solution to our system. Now, to obatin a full solution of the capped solution to our model we only need the first column of the Adjugate matrix. Second it splits the two funtions of μ; Its effect on transcript number and its effect on transcript distribution. And allows for their separate analysis. Third, it permits analysis of the underlying transcript distribution even under conditions where the model has no solution. For example, when $\mu = 0$, both solutions are indeterminate. However, the determinant free solution allows for us to explore what the transcript

distribution would be when μ=0.

2.4.3. Decapped Subsystem steady state solution

Starting with the decapped subsystem of equations:

$$\begin{aligned}
\frac{dm_0^*}{dt} &= \mu(0)m_0 + \tau(1)m_1^* - \delta m_0^* \\
\frac{dm_1^*}{dt} &= \mu(1)m_1 + \tau(2)m_2^* - \tau(1)m_1^* \\
&\vdots \\
\frac{dm_i^*}{dt} &= \mu(i)m_i + \tau(i+1)m_{i+1}^* - \tau(i)m_i^* \\
&\vdots \\
\frac{dm_{i_{\max}}^*}{dt} &= \mu(i_{\max})m_{i_{\max}}^* - \tau(i_{\max})m_{i_{\max}}
\end{aligned}$$

We get the following solutions at steady state:

$$\begin{aligned}
m_0^* &= \frac{\mu\, m_0 + \tau(1)m_1^*}{\delta} \\
m_1^* &= \frac{\mu\, m_1 + \tau(2)m_2^*}{\tau(1)} \\
&\vdots \\
m_i^* &= \frac{\mu\, m_i + \tau(2)m_{i+1}^*}{\tau(i)} \\
&\vdots \\
m_{imax}^* &= \frac{\mu\, m_{imax}}{\tau(imax)}
\end{aligned}$$

We can rearrange the solutions and simplify to find,

$$
\begin{aligned}
m_0^* &= \frac{\mu}{\delta} \sum_{j=0}^{i_{\max}} m_j \\
m_1^* &= \frac{\mu}{\tau} \sum_{j=1}^{i_{\max}} m_j \\
&\vdots \\
m_i^* &= \frac{\mu}{i\,\tau} \sum_{j=i}^{i_{\max}} m_j \\
&\vdots \\
m_{imax}^* &= \frac{\mu}{i_{\max}\,\tau} \sum_{j=imax}^{i_{\max}} m_j
\end{aligned}
$$

We can simplify the model by converting the mRNA quantity m_j to the probability p_j by the following.

$$\frac{\lambda}{\mu} = \sum_{i=0}^{i_{\max}} m_i \tag{15}$$

Therefore,

$$1 = \frac{\mu}{\lambda} \sum_{i=0}^{i_{\max}} m_i \tag{16}$$

For any $i = j$ where S_j is cummulative probability from $i = classj$ to $i = i_{\max}$.

$$S_j = \frac{\mu}{\lambda} \sum_{i=j}^{i_{\max}} m_i \tag{17}$$

Now the solution becomes,

$$
\begin{aligned}
m_0^* &= \frac{\lambda}{\delta} S_0 = \frac{\lambda}{\delta} \\
m_1^* &= \frac{\lambda}{\tau} S_1 \\
&\vdots \\
m_i^* &= \frac{\lambda}{i\,\tau} S_i \\
&\vdots \\
m_{imax}^* &= \frac{\lambda}{i_{\max}\,\tau} S_{i_{\max}}
\end{aligned} \tag{18}
$$

The total transcript population in the decapped state does not have a closed form solution. However it can be summarized as follows,

$$m^*_{tot} = \sum_{i=0}^{i_{\max}} m^*_i = \frac{\lambda}{\delta} + \frac{\lambda}{\tau} S_1 + \ldots + \frac{\lambda}{i\tau} S_i + \ldots + \frac{\lambda}{i_{\max}\tau} S_{i_{\max}} \tag{19}$$

This can be further shortened using element wise multiplication denoted by the hadamard product ($\odot$).

$$m^*_{tot} = \lambda(\frac{1}{\delta} + \frac{1}{\tau}\vec{S} \odot \vec{l}) \tag{20}$$

Where $\vec{S}$ is a vector of all the cummulative sums and $\vec{l}$ is a vector of $1, 1/2, ..., 1/i, ..., 1/i_{\max}$. The S_i have the following arrangement $S_0 = 1$ and $S_0 \geq S_1 \geq ... \geq S_i \geq ... \geq S_{imax}$. This depends on the distribution of $\vec{m}$ of the capped state. Exploring the result we find a few properties of our system. Transcription rate (λ) again serves only to scale the entire system. The first decapped class's population m^*_0 is only dependent on the degradation rate (δ). The total mRNA in the decapped state can wildly vary according to the value of degradation. In this work we shall set delta to be large and focus on the effects of the marking rate and elongation/termination rate. This result will be explored further in the results.

To get the probability distribution of transcripts across the decapped state we can divide $\vec{m^*}/m^*_{tot}$ which results in,

$$p^*_0 = \frac{1}{1 + \frac{\delta}{\tau}\vec{S} \odot \vec{l}}$$
$$p^*_j = \frac{S_j}{j(\frac{\tau}{\delta} + \vec{S} \odot \vec{l})} \quad , for j = 1, 2, ..., i, ..., i_{\max}$$

2.5. Complete system mRNA population

The total mRNA (M_{tot}) in the system is defined by,

$$M_{tot} = \lambda(\frac{1}{\mu} + \frac{1}{\delta} + \frac{1}{\tau}\vec{S} \odot \vec{l}) \tag{21}$$

To understand how mRNA is divided between the two subsystem we can calculate the log odd of finding an mRNA in the decapped class. Again we will set δ to very large.

$$p_{mtot} = m_{tot}/M_{tot} = \frac{\frac{\lambda}{\mu}}{\lambda(\frac{1}{\mu} + \frac{1}{\tau}\vec{S} \odot \vec{l})}$$
$$p_{mtot} = \frac{1}{(1 + \frac{\mu}{\tau}\vec{S} \odot \vec{l})}$$

Then you calculate the odds,

$$odds_m = \frac{p_{mtot}}{1 - p_{mtot}} \tag{22}$$

$$odds_m = \frac{\frac{1}{(1+\frac{\mu}{\tau}\vec{S}\odot\vec{l})}}{1 - \frac{1}{(1+\frac{\mu}{\tau}\vec{S}\odot\vec{l})}} \tag{23}$$

It simplifies to,

$$odds_m = \frac{1}{\frac{\mu}{\tau}\vec{S}\odot\vec{l}} \tag{24}$$

$$log_{10}(odds_m) = -log_{10}(\frac{\mu}{\tau}\vec{S}\odot\vec{l}) \tag{25}$$

2.6. model implementation in R

Code to solve the model was written in the R package Ribosome. To solve the capped subsystem of the model, the Solve.tridiag routing from limSolve (Soetaert,K 2009). The model solutions presented previously serve their purpose to make the mathematics more interpretable by humans. However, solving for just the first column of the adjugate matrix is far slower than solving the whole tridiagonal system using limSolve. To solve for the deapped subsytem, the values of the capped state were plugged in to the solutions above. Utility functions, plots and statistics were created using R (v 3.6), data.table (v1.14.0), and limSolve (v 1.5.6).

2.7. Data Sources

In order to interpret the model within biological context, we will focus on parameter ranges derived from the literature. The range of $i_{\max}$ is determined from the distribution of protein lengths obtained from yeast (Figure 2A) or Arabidopsis (Figure 2C) which is then divided by the average number of codons covered by a ribosome (9 codons). Both were extracted from the Ensembl (version 109) and Ensembl plants (version 56) respectively (Cunningham 2022, Yates 2022, Kinsella 2011). The marking rate between the Capped and uncapped system was approximated from the protein half-lives from Presnyak 2015 for yeast (Figure 2B) and Sorenson 2018 for arabidopsis (Figure 2D). Presnyak utilized the temperature sensitive *rbp-1* RNA polymerase mutant in yeast. This mutant can not undergo transcription at non-optimal temperatures, thus allowing for the measurement of mRNA decay over time. Sorenson (2018) used the transcriptional inhibitor, cordycepin, to treat *Arabidopsis thaliana* seedlings and measured their decay using RNA-Seq. To approximate μ transcript half-lives were converted to rates with the following:

$$\mu = \frac{ln(2)}{t_{1/2}} \tag{26}$$

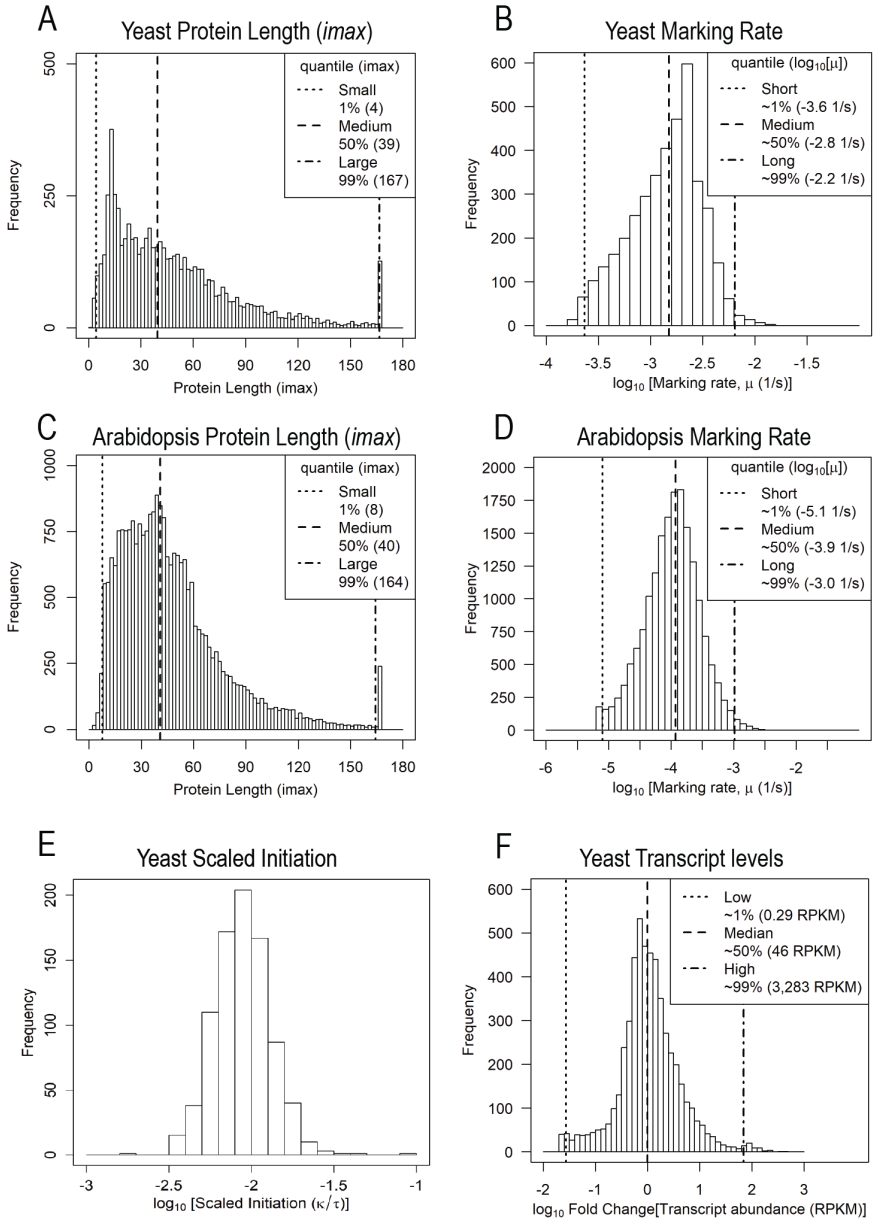
Where $t_{1/2}$ is the half-life. The resulting range of μ is from 10^-2 to 10^-4 for yeast and 10^-3 to 10^-6 for arabidopsis.

Translation initiation and elongation rates (κ and $\tau*9*i_{\max}$) were obtained for Yeast from Duc and Song 2018. Remember, that τ is the overall rate of translating a transcript, not the average elongation rate (see section 2.2). In Duc and So stng 2018, the authors used 850 highly translated transcripts from the ribo-seq dataset from Weinberg 2016. They employed a TASEP model to estimate the initiation rates and correct the empirical elongation rates from the footprint distributions. We calculated an average gene specific elongation rate from the corrected elongations rates. We scale the each gene specific initiation rate by dividing it by the gene specific elongation rate.This is due to the fact that many combinations of κ and τ can yield the same scaled initiation rate (e.g. κ =0.02 and $\tau * 9 * i_{\max}$ = 2, and κ=0.04 and $\tau * 9 * i_{\max}$=4, both yield κ/τ=0.01). This simplifies the model behavior to one generalized parameter with a unique response (Figure 2E). The scaled initiation rate ranges from $0.1s^{-1}$ to $0.001s^{-1}$.

The transcription rate, λ only acts as a scaling factor throughout the model and does not affect the distribution of the ribosomes. For solutions provided in this work λ has been set to one. However, as a point of reference, the transcriptomic results from Weinberg 2016 are included in Figure 2F. In short, reads per kilobase million from Weinberg were further converted into a log10 fold change based on the median expression level. Figure 2F shows that the absolute range of transcriptional control ranges just under 5 orders of magnitude.

Finally, δ only determines the accumulation of transcripts in the m_0^* class. In our model the remaining transcripts in class m_0^* are either only the 3' end of co-translationally degraded transcripts or full transcripts from class m_0. Recently, the rate of degradation for the 5' - 3' exonuclease XRN1 was determined to be 26 nt/s (Atthapattu 2021). XRN1 is the primary exonuclease involved in co-translational degradation and 5' degradation pathways (Sorenson 2018,Yu 2016, Collart 2019, Pelechano 2015). For an average 3' UTR of 121 nts (Kebaara 2009) this would take 4.6s, and an average transcript of 1400nt would take 54s to degrade. This means the average degradation rate δ would take between 1/54s = 0.019/s or 1/4.6s = 0.22/s . The total population of mRNA in m_tot^* is determined by $1/\delta$, $1/\tau$ and μ (as part of $\vec{S}$) as shown Equation 20. τ ranges from 0.03 to 10^-4. This makes $1/\delta \leq 1/\tau$. It is reasonable to explore the model with large δ since decapped transcripts are translationally incompetent.

Figure 2: Histograms of empirical values of model parameters. A) Yeast protein lengths. B) Yeast half-life C) Arabidopsis Protein Lengths. D) Arabidopsis Half-Life. E) Yeast Scaled elongation rates (Translational initiation rate/average translation elongation rate) on a per gene basis. F) Log 10 Fold Changes between all transcripts compared ot the median transcript expression in yeast.



3. Results

The model outputs two sets of results. First it provides the total mRNA population and how this population is split between capped and decapped states. Secondly, it describes how transcripts are distributed among ribosomal classes within each state.

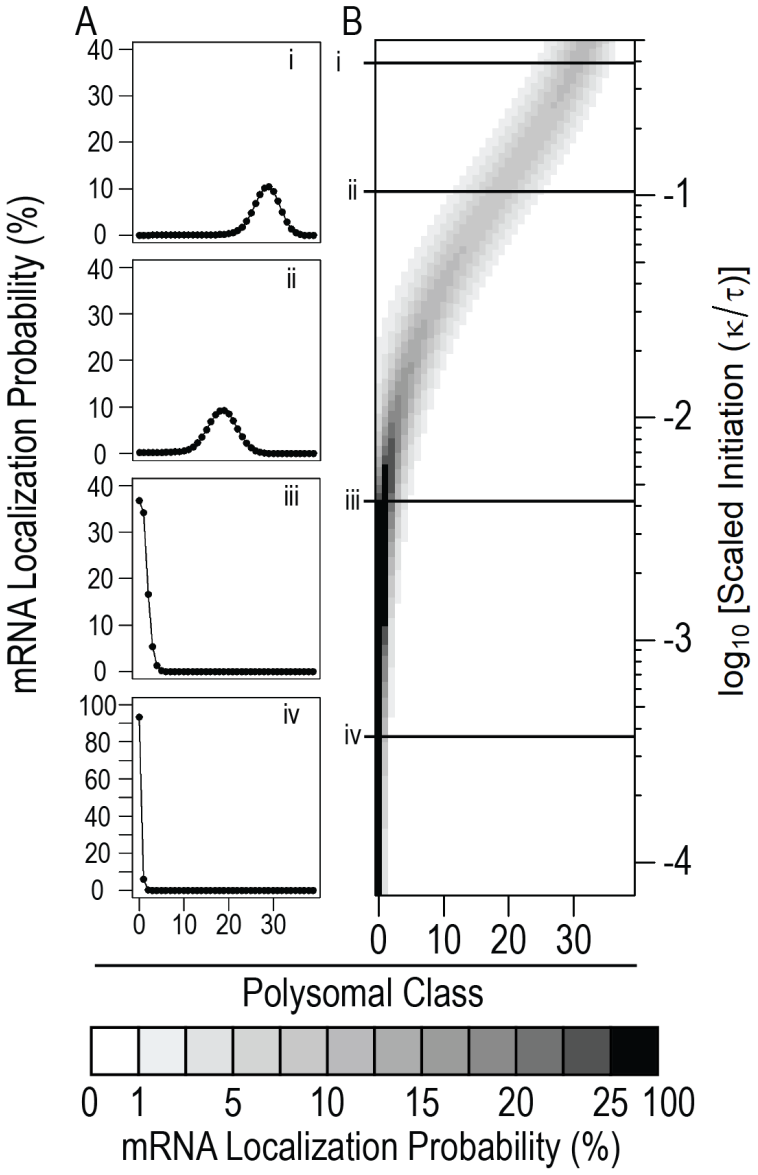
3.1. Model provides a unique distribution of mRNAs across ribosomal classes for each scaled initiation rate

The model outputs the probability of finding a transcript in each ribosomal class for a particular parameter set. In figure 3A the mRNA distribution in the capped state is presented for four different scaled initiation rates (as shown in Figure 3B) for a median length protein with a long (52 minute) half life. To summarize the model results across a range of parameters a heatmap where each row is the distribution of mRNA at a particular scaled initiation rate is shown. As the scaled initiation rate increases the density moves to the right and spreads out in the capped system. The same heatmap can be produced for the marked class (Figure 4) and for the whole system (Figure 5). The distribution in the capped system is bounded at class 0 and class $i_{\max}$ and roughly symmetrical away from the boundaries (Figure 3B). The decapped system is centered around the lowest classes (Figure 4). This is a result of higher classes having a higher chance of terminating, therefore skewing the distribution left. As scaled initiation increases, the distribution in the capped class shifts to higher ribosomal loads which results in an increase in the ribosomal class of the decapped system. This is due to transcripts entering the decapped system at a higher ribosomal class. The whole system incorporates the transcript distributions from both capped and decapped systems. This is apparent in the bimodal peaks at higher scaled initiation values, with a peak at lower ribosomal load representing the decapped system and a high ribosomal load from the capped system (Figure 5).

3.2. Initiation interference due to increasing ribosome density is noticeable beginning at moderate to high ribosomal loads

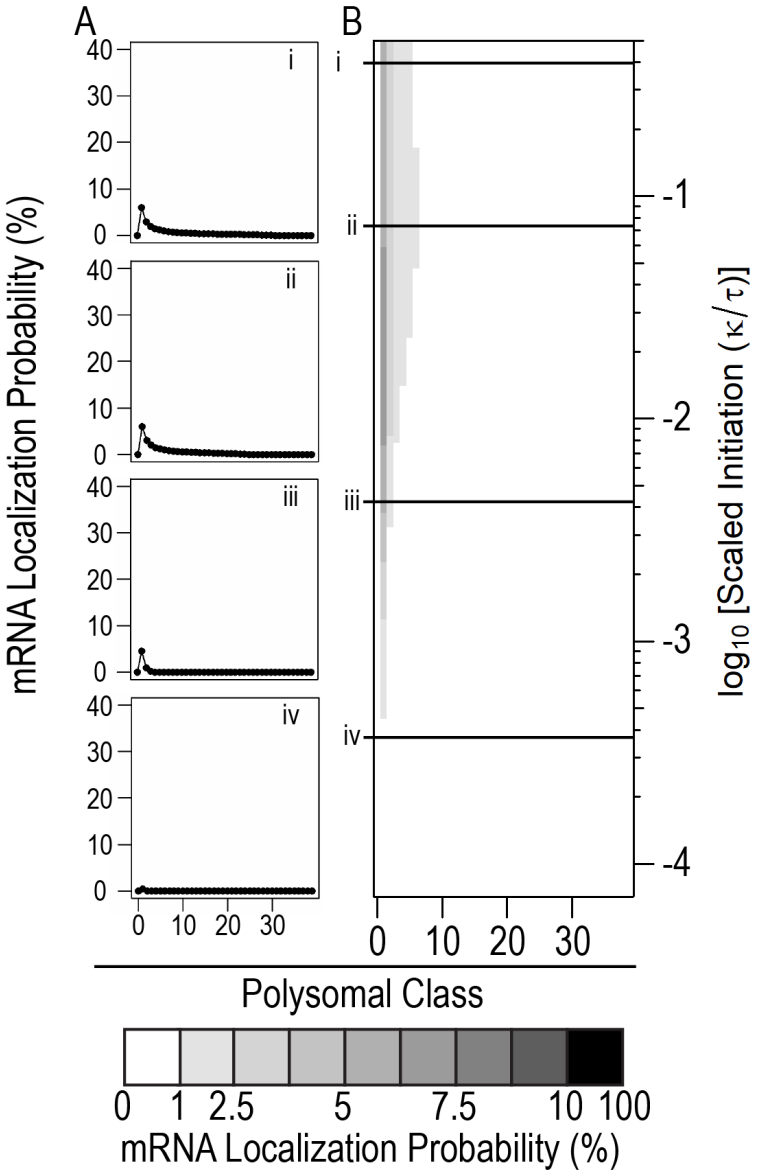
On a physical level, ribosomes initiating on a transcript depend on having sufficient space around the start codon. Transcripts at with higher ribosomal loads therefore have a higher probability of having an obstructed start codon. In order to explore the effect of ribosomal load on initiation we created two versions of the model. The first is independent on the ribosomal load of a transcript, the density independent model (DII). A second dependent on the ribosomal load of a transcript, the density dependent initiation (DDI) model. The DII model is presented in figure 5 A-B, with individual probability profiles presented in 5A and the summary heatmap in 5B. The DDI model is presented in Figure 5C-D. Individual probability profiles are generated at low, mid and high scaled initiation rates as determined from Duc and Song 2018. Note that in the DII model the system saturates just above the high scaled initiation rate, while the DDI model doesn't. Additionally, model profiles are very similar at low to mid scaled initiation values for both models. Figures 5 B and D suggest that the density dependent effects start appearing between mid and high scaled initiation rates. Polysome

Figure 3: mRNA distribution in Capped state. A) Individual distribution profiles for four scaled initiation values B) Heatmap of model output across a range of scaled initiation values. Lines represent slice represented in A). Results produced with $i_{\max}$ of 39 and a long half life of 52 minutes (99th percentile). Color bar shows probability of finding mRNA in particular ribosomal class.



profiling experiments usually ony resolve 8-10 ribosomal peaks, with the majority of the signal arising from polysomes 2-5. This generally agrees with single molecule imaging of nascent peptides where ribosomal densities are on the range of 0.5% - 30% (Morisaki 2016, Wang 2016, Wu 2016, Yan 2016). While both the DII and DDI models show similar behavior and low density at low to mid scaled initiation values, empirical evidence and physical reality would indicate that a transcript is unlikely to ever reach saturation. This supports the DDI model. From here on out all results will be solely based

Figure 4: mRNA distribution in decapped state. A) Individual distribution profiles for four scaled initiation values B) Heatmap of model output across a range of scaled initiation values. Lines represent slice represented in A). Results produced with $i_{\max}$ of 39 and a long half life of 52 minutes (99th percentile). Color bar shows probability of finding mRNA in particular ribosomal class.

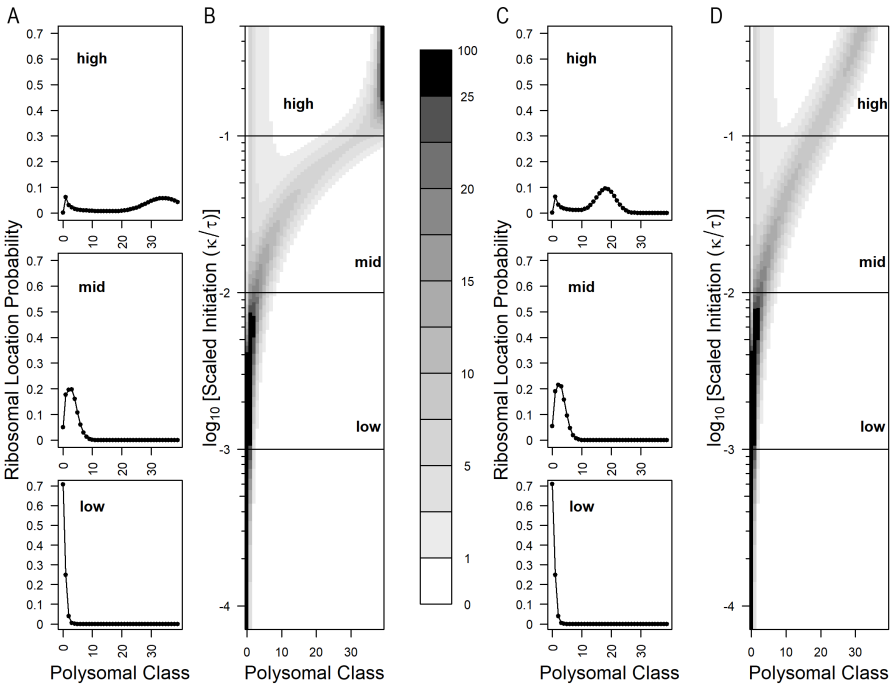


on the DDI model.

3.3. Higher marking rates reduce capped state ribosomal loads

The interplay between the translational machinery, mRNA degradation machinery and mRNA properties such as codon optimality, secondary structure or modifications all have been reported to play a role in mRNA stability (Wu 2019, Medina-Munoz 2021, Bae and Collier 2022). To explore the role of mRNA stability on mRNA populations we varied the marking rate from the 1st percentile,

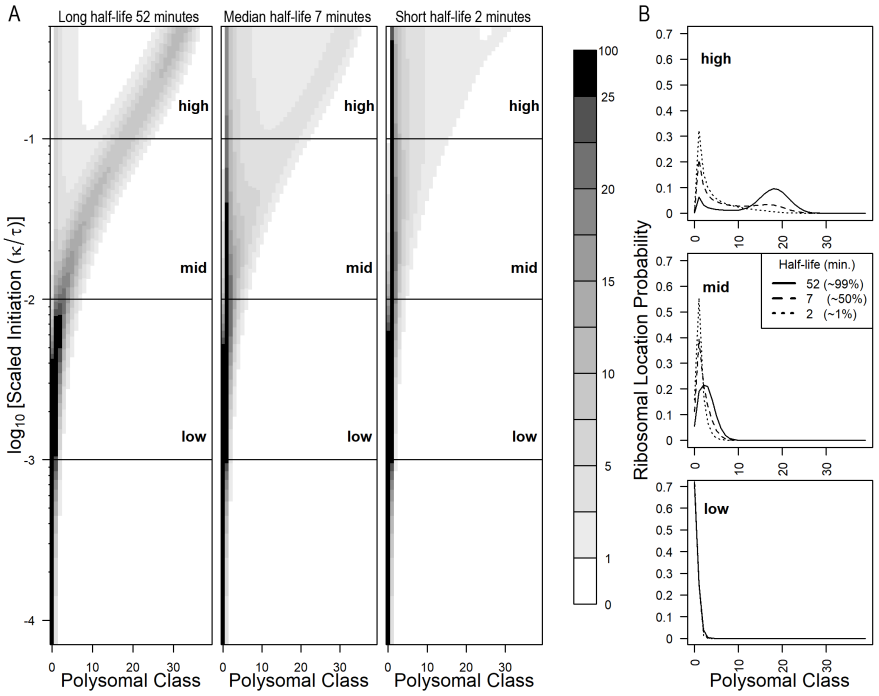
Figure 5: Comparison of density independent initiation (DII) and density dependent initiation (DDI) in full system. A) DII individual density profiles for low (0.001), mid (0.01) and high (0.1) scaled initiation values. B) DII density Heatmap for the full system. C) DDI individual density profiles for low (0.001), mid (0.01) and high (0.1) scaled initiation values. B) DDI density Heatmap for the full system. All results calculate with $i_{\max} = 39$ and long half life of 52 minutes (99th percentile).



median and 99th percentile values as determined from the half-life values in Presnyak 2015. We find that as half-life decreases the distribution of mRNAs change in two ways. First there is shift to lower ribosomal classes in the capped state (Figure 6). This is likely due to the mRNAs leaving the capped state at a higher rate and driving the equilibrium towards lower ribosomal loads. Secondly, as half-life decreases, a larger proportion of the mRNA is found in the decapped state. This is further explored later.

Multicellular eukaryotes, such as plants, face a different set of environmental challenges and tend to have slower translation initiation and elongation rates as well as slower cell division when compared to single celled organisms. This is highlighted by the current gold standard study of mRNA half-lives in the model organism *Arabidopsis thaliana*, where the half-lives measured are one two two orders of magnitude longer than those in yeast. To explore this we ran the model using the same scaled initiation rate as in yeast, the median arabidopsis $i_{\max}$ of 41, and arabidopsis half-lives (long half life (1,500 minutes), median half life (115 minutes), and short half life (11.5 minutes)). As expected, the longer half-lives have higher ribosomal loads and are mostly in the capped state.

Figure 6: Shorter half-lives reduce ribosome load in capped system in yeast. A) Heatmaps for the full system. Left) long half life (52 minutes) Center) median half life (7 minutes) Right) short half life (2 minutes) B) individual density profiles for low (0.001), mid (0.01) and high (0.1) scaled initiation values for each half life value. All results calculate with $i_{\max} = 39$.



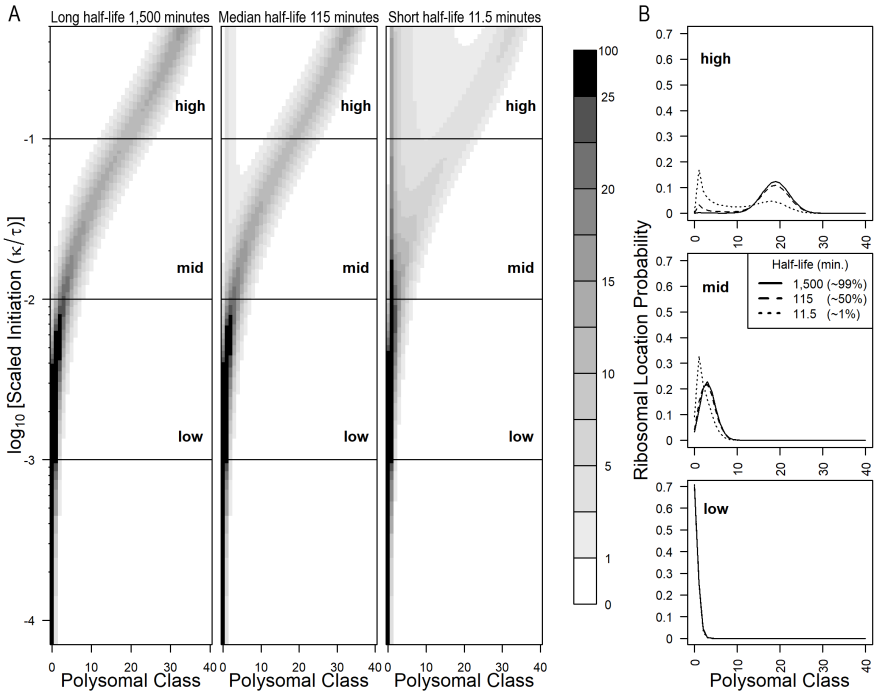
3.4. Under equilibrium, protein length does not affect the ribosomal density on transcripts

One particular aspect of analyzing the model at equilibrium is that while the total number of ribosomes on a particular transcript is dependent on length (Figure 8A), the density per unit length is not (Figure 8B). In other words, the flux of ribosomes through a transcript is independent of the length of the transcript. At equilibrium and under the same parameters, except for $i_{\max}$, the number of ribosomes initiating and terminating is the same regardless of $i_{\max}$. This property means that for any model solution, if the ribosomal classes are converted into densities (i.e. dividing the x axis of figure 5 by 39), you now have a general solution for all transcripts with those parameters. It also implies that the rate of protein production is independent of transcript length. 0

3.5. Marking rate and ribosomal load determine mRNA distribution between states

As shown in previous results, shorter half-lives (faster marking) leads to lower ribosomal load and a shift towards the decapped state. To explore this shift we can use the results in equation 14, which splits the mRNA population in the capped state and the distribution of reads within and equation 20 the total transcript population to derive the log odds in equation 25. Using equation 25, we can see under which parameter regimes mRNA is more abundant in the capped state. We produced output

Figure 7: Longer half-lives in arabidopsis result in a smaller effect on ribosomal load in the capped system. A) Heatmaps for the full system. Left) long half life (1,500 minutes) Center) median half life (115 minutes) Right) short half life (11.5 minutes) B) individual density profiles for low (0.001), mid (0.01) and high (0.1) scaled initiation values for each half life value. All results calculate with $i_{\max} = 41$.



across all scaled initiation values and under the 1%, 50% and 99% percentiles for marking rates in both yeast and arabidopsis (Figure 9). We note two patterns. First as the scaled initiation rate increases the amount of mRNA in the decapped class increases. Secondly, shorter mRNA half-life bias transcripts towards the decapped class as previously seen in Figures 6 and 7.

18

3.6. Protein production is a function of

4. Discussion

Paragraph summary

paragraph in context to other models TASEP, RFM cell wide models and results comparison

paragraph that talks about marking rate effects vis a vis effects on translation

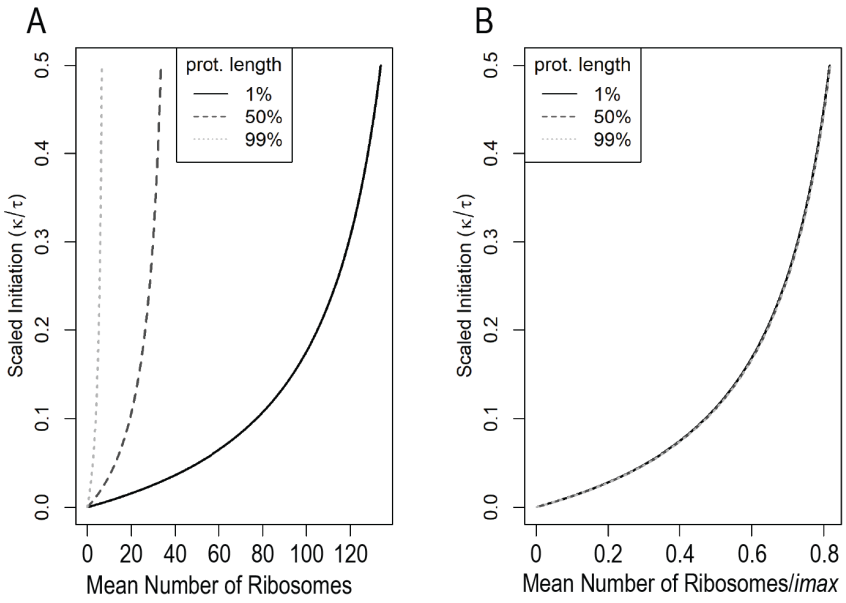
paragraph that talks about limitations and extensions

paragraph that talks about protein production

Text for discussion of figure 1

1. The model presented most closely resembles the co-translational decay pathway from eukaryotes.

Figure 8: The ribosomal density on a transcript is independent of their length. A) Ribosomal load per transcript is higher for longer transcripts. B) When the ribosomal load is corrected for the length of the transcript, the ribsomal density collapses to the same curve for all transcripts under the same parameters. For this plot arabidopsis parameters were used $i_{\max} = 8$ (1% percentile), 40 (1% percentile), 164 (99% percentile), 1500 minute half life, over the full scaled initiation range 0.0001- 0.5.



2. There are a few current limitations of the model. First and foremost it doesn't model endonucleolytic decay pathways such as gene silencing by siRNA and the RISC complex or the NMD, NGD and NSD pathways or degradation initiated through ribosomal collisions.

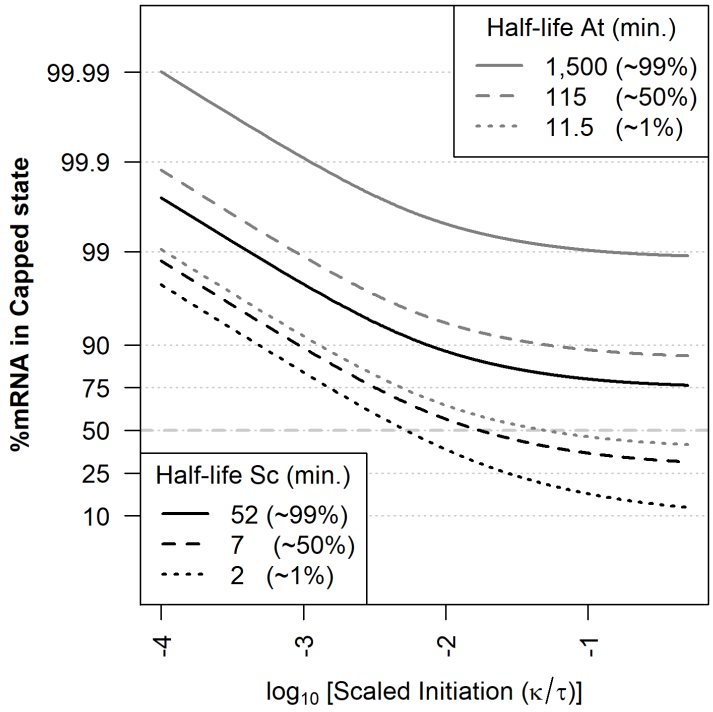
3. Implementation of endonucleolytic decay would require an ability to track ribosomal position, which is not possible in our model.

4. Alternatively a probabilistic model of ribsomal position given a particular load could allow for redistribution of capped transcripts into decapped states of lower ribosomal class.

5. Currently there is a debate whether mRNA stability is regulated primarily through the protective effects of ribosomal association or through the suboptimal codons causing ribosomal stalling and the subsequect ribosome associated decay pathways (Chan elife).

6. We explored the interaction of ribosomal loading and marking rate, noting that increased marking leads to a lower expected ribosomal load. This particularly noticable for genes with short half lives and is less sensitive with transcripts with halflives above 2 hours.

Figure 9: Log odds of finding mRNA in the capped state for a range of marking rates in yeast and arabidopsis.



7. In the current implementation of the model we did not directly explore the protective effects of ribosomal loading. This could be first implemented by including a similar weighting term analogous to the wieght for the intiation rate of (1-i/imax).

8. However, biology suggests a more complex behavior.

9. The protective effects of translation could increase per ribosome, but eventually at high loads could trigger ribosome associated decay pathways through ribosomal collisions.

10. Our model does not model elongation at a single codon level like TASEP, nor at a course grained approximation of groups of codons like the ribo flow models. This means that it is not capable of analyzing the effect of individual non optimal codons, or stretches of mRNA which are non-optimal

11. Our model reflects the average elongation rate over a transcript. However it can be set to reflect the slowest bottle neck across a transcript.

12. Another alternative to modeling bottlenecks is increasing granularity to a very coarse grained

ribo-flow model comprising of only two or three sectors, depending whether the transcript has one or two bottlenecks.

13. This splits transcripts into an intial sections where the bottleneck effectively "shortens" the transcript length to the first bottleneck site. Increasing the effect of ribosome loading on iinitiation interference.

14. However, the transcript physically has more length and can bear a larger ribosomal load on the section section. This results in a transcript having and "effective" imax between the actual coding sequence length and the lenght of the transcript to the first bottleneck.

what is interesting here is that our model does not explicitly capture the protective role of translation on mRNA stability. Nor does it consider local fluctuations in codon optimality. In many ways it simply takes the base behavior of a system, acting as it's null. The fact that at fast marking rates dominate translation is interesting. It might be that in-vivo it is a competition between the decapping machinery and the translational machinery. A more bi modal response might be more realistic if translation has a protective effect.

outside of shah and reuveni how many models have explored the parallel mRNA stability and translation interplay? Not many. Many papers modeling features which may affect translation, but not translation itself

all these methods of translational attenuation have the greatest impact at the higher scaled initiation rate values. Everything is driven towards having low ribosomal densities. DDI, marking, ribsomal collisions. This doesnt mean high translational rates arent possible but much of the things that can be done have detrimental impacts on big translational loads. Implies that translation is robust at low rates and highly variable and unstable at high rates. Literature suggest however that low rates of processivity also result in instability. So there is a goldilocks zone. Why would the cell not want bursty behavior? Energy saving. Also remember that even a low load integrated over time is significant. a 52 minute half life with an average protein is plenty. 352 aa at even 1codon a sec. 352sec = 6 min, at density of 1 ribosome that is almost 8 proteins during a halflife per transcript in a population.

a half life of two minutes is very short and prompts the question on how it can exist? The model suggest that the only a 0.28 average ribosomal load is possible and that most mRNA is in the decapped state. So production of protein from these transcripts seems unlikely. One way this could work is that Translation is stabilizing and creates a bimodal system where there is a low probability of associating

to ribosomes, but once associated you are stable for a while and produce protein.

In terms of protein production TL is a scalar multiplier of TX abundances. But it also acts as general regulator and quality control mechanism for the overall mRNA population. The ability of TL to regulate gene expression is on the face of it only about one and a half orders of magnitude above parity with mRNA expression. While transcription can range from four to five orders of magnitude. So very low translation rates below a one to one ratio with the transcript population can still result in substantial protein production given enough mRNA substrate or given enough time (low marking rate). Translation does have the power to completely abrogate all protein production however, acting as an on off switch under extreme conditions (heat shock and hypoxia). Also selective translation seems to rely on mechanisms that avoid or take advantage of the global translational repression pathways. TL seems to be, at base value, optimal for a gene in question, with little that can happen in-vivo to directly increase translational efficiency. This doesn't mean that there isn't space for sequences to be optimized, secondary structure to be reduced and the base translational behavior of a transcript to be improved over a naturally occurring transcript. However, mRNA abundance and stability seem to be greater drivers of total protein production.

5. Appendix

For simplicity, we begin by defining our model equations using generic functions to describe the transition of mRNAs between different classes or states. We then constrain the model by assuming specific functions to describe the transition of mRNAs between classes.

General model equations of the density independent initiation model

Our model consists of two sets of time dependent and coupled ODEs. Each set of ODEs describes the abundance of mRNAs that are either capped and decapped for degradation. The ODEs within each sets equations are structured by the ribosome load of the mRNA. The coupled ODEs within a set of equations describe how mRNAs are introduced to the set, the transitions in ribosome load via initiation or completion of protein translation, and the transition between sets either via the marking of capped mRNAs or the degradation of decapped mRNAs with a ribosome load of 0.

Specifically, new mRNA enter the 0^{th} capped class $m_0(t)$ at a rate.. Ribosomal bind mRNAs in the i^{th} capped class at a rate $\kappa(i)$, increasing the mRNA's ribsome load to the $i + 1^{th}$ class. By definition, $\kappa(i_{\max}) = 0$, i.e. mRNAs with a ribosome load of $i_{\max}$ cannot accommodate any addtional mRNAs. In the density independent model (DII), we assume that the current ribosomal load has no

effect on the ability of another ribosome to bind to the transcript. An average ribosomal footprint covers 9 codons (27 nucleotides). Therefore for a protein of 270 amino acids in length, the maximal ribsomal load, $i_{\max} = 10$. Capped mRNAs with ribosome load i are decapped at a rate of $\mu(i)$. We assume that capped mRNAs are decapped for degradation rate independent of their ribosome load, i.e. $\mu(i) = \mu_0$. Accordingly, the ribosome load of decapped mRNAs remains unchanged, but they are transitioned from the capped class $m_i(t)$ to the decapped class $m_i^*(t)$. Ribosome movement along an mRNA is assumed to occur independent of whether or not its capped or decapped for degradation. Thus, ribosomes complete translation of both decapped and capped mRNAs with ribosome load i at rate $\tau(i)$, decreasing the mmRNA's ribosome load to the $i - 1^{th}$ class. Where $\tau(i) = i \cdot \tau(1)$ and . This is because we not modeling the explicit movement of ribosomes along an mRNA, we assume that at steady state probability of finding a ribosome at any given codon position within the coding sequence follows a uniform distribution. Thus, the chance that a ribosome on a transcript of class i will complete translation increases as ribosome load increases. Since mRNA's with a ribosome load of 0 have no ribosomes which can complete translation, by definition $\tau(0) = 0$. It is important to note that $\tau(1)$ is not the same as the average elongation rate. $\tau(1) = average\ elongation\ rate/(9 \cdot i_{\max})$. That is, the average elongation rate in aa/s is rescaled to the average rate of total elongation and termination through a transcript in units of $1/s$.