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 decapping 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 and to verify model function.

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

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 decapping 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. decapping 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

Intro Outline 3.1.1. Gene regulation, translation and mRNA stability 3.1.1.1. Short introduction to Gene expression, transcription, translation, and the regulation of mRNA populations both dependent and independent of translation 3.1.2. Ever increasing methods of measuring mRNA decay and Translation provide ample grounds for testing and knitting together hypothesis underlying the mechanism of translation. 3.1.2.1. Ribo-seq, microarrays, polysome profiling, proteomics and live imaging. 3.1.2.1.1. But most of these approaches are not measurements of single transcripts, but ensemble measurements of populations 3.1.3. Mathematical modeling as a tool to interpret and generate hypotheses to better understand translation 3.1.3.1. TASEP 3.1.3.2. Riboflow 3.1.3.3. Other bulk "cell-wide" approaches shah 2013 3.1.4. Our model acts as an intermediate between cell wide approaches and single transcript models such as TASEP and Riboflow. Our coarse-grained model of translation focuses on the behavior of transcript populations. This includes effects originating from transcription and mRNA decay as well as translation initiation and elongation/termination. By modeling translation at the population level, we can also use the model in the future to better understand the information held in ribo-seq and proteomics experiments.

1. Gene expression relies on transfer of information encoded in DNA into a final functional form, often protein.
    (a) Protein production begins when DNA is transcribed into mobile messenger RNAs (mRNA).
    (b) Subsequently, the nucleotide code in the mRNA is translated by the Ribosome into the final protein sequence.
    (c) While this process represents the basic flow of genetic information, each step has multiple regulatory mechanisms adjusting gene expression.
2. The process of translation
    (a) ss

The interplay between the translational machinery, mRNA degradation machinery and mRNA properties such as codon usage, secondary structure or modifications all have been reported to play a role in mRNA stability (Wu 2019, Medina-Munoz 2021, Bae and Collier 2022).

3. mRNA populations are regulated by a series of degradation mechanism
    (a) mRNA degradation
    (b) 5’ decapping
    (c) 3’ deg
    (d) Ribsome quality control mechanisms and degadation
4. modelling intro
    (a) TASEP
    (b) Ribo flow
    (c) Shah
    (d) Degradation papers
    (e) Translation and mRNA degradation have both received ample attention in the literature, however few have explored the interaction between mRNA degradation and translation.
5. Paper summary
    (a) stuff

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

   Stuff for intro

4. mRNA degradation mechanisms (Cao and Parker 2001, Cao and Parker 2003, Wu 2013, Wu 2016, Zupanic 2016, reviewed in Ashworth 2019) and translation (Reuveni 2011, Nanikashvili 2019, Raveh 2016, Shaw 2003, Shah 2013) have been modeled separately in the past, but only rarely together (Reuveni 2011, Valleriani 2011).

5. in our model we are going to be focusing on the process of mRNA 5' decapping.

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.

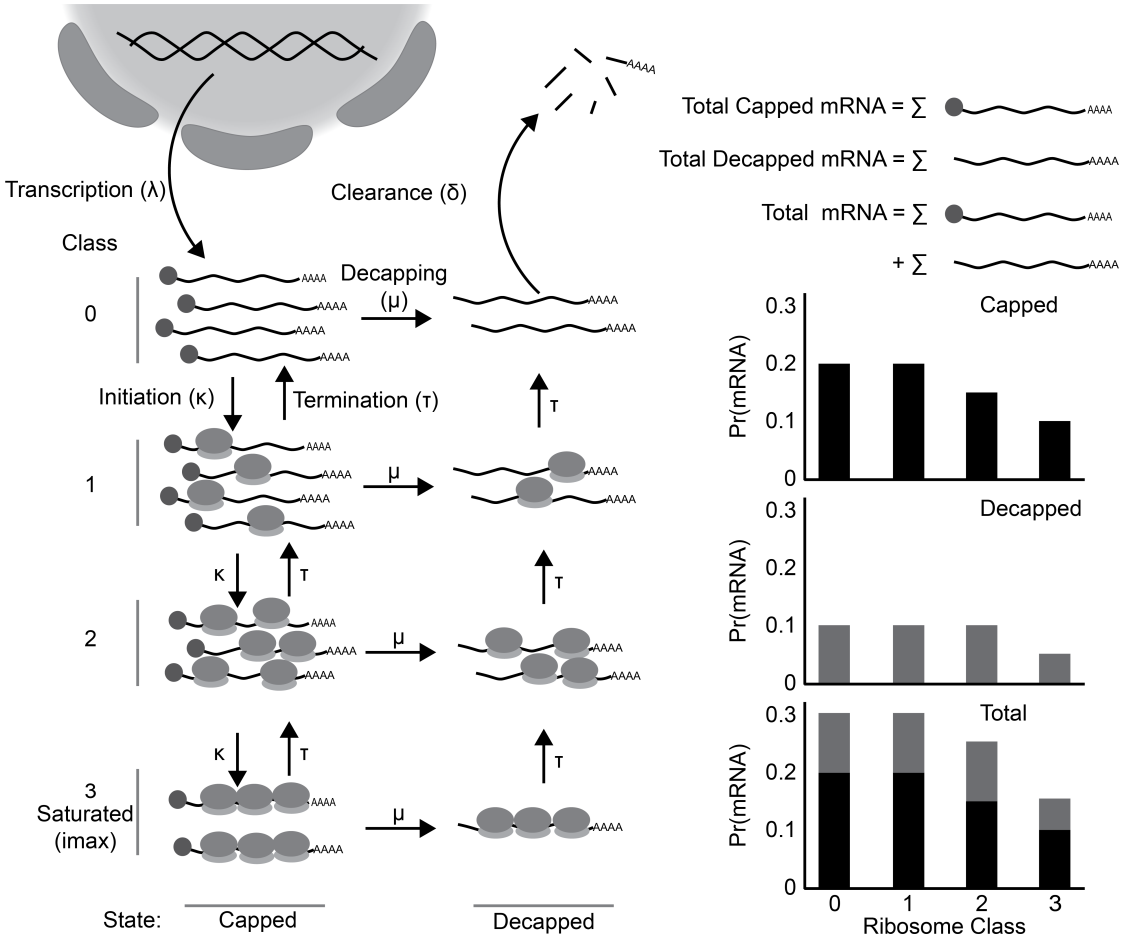
7. Ribosomal association of transcripts can lead to both protection of viable transcripts as well as quality control degradation of faulty transcripts.
8. When a viable transcript is bound by the ribosomal and translational machinery, the 5' cap is bound by translational initation factors and the 3' 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.

many models seek to understand separate aspects of mRNA biology. Some focus on the mechanistic aspects of decay, separate from the interaction with the transaltional machinery. Others model translation directly, but not decay (TASEP and RFM). Others focus on a bulk measure of all processes, but with no particular allusion to specific types of degradation. here we present a model focused on integrating from mRNA procution to final degradation, the maintenance of mRNA populations with regards to translation and co transaltaional decay.

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 on the mRNAs in the capped, ribosome free polysome class 0 at rate κ and be loaded onto the transcript and move it into capped class 1. Because only one ribosomes can occupy a particular location on the mRNA at any given time and our model does not track ribosomal positions, , we model translation intiation across polysome classes $i = 0$ to i_{max}, more generally as

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

where i is the mRNA polysome class, i_{max} is the maximal ribosomal occupancy on the transcript. We note i/i_{max} represents, under the assumptions of a uniform distribution, the probability a randomly chosen codon position is occupied by a ribosome. Correspondingly, $(1-i/i_{\text{max}})$ represents probability a randomly chose codon position, such as the initiation site is unoccupied. Because, the natural units for mRNA coding sequence length in our model is the amount of space the bound ribosome takes when translating it follows that $i_{\text{max}} = n_c/9$ where n_c is the length of the mRNA's coding sequence in codons and 9 represents the length in codons of space a single ribosome occupies. This attempts to account for the ribosomal density dependent effects on initiation and is called the density dependent initiation (DDI) model.

A ribosome on the transcript elongates the peptide and, in turn, terminates at a rate of τ, Given that the length scale of our model is formulated in terms of ribosome widths but most estimates of elongation are at the scale of codons, if τ_c is the average elongation rate of an mRNA in codons, $\tau = \tau_c/9$. As the number of ribosomes on a transcript increase, the probability of a ribosome being at the end of the transcript also increases. Again assuming ribosomes are distributed across a transcript

according to a uniform distribution, the expected ribosome termination rate on a mRNA in polysome class i is simply,

$$\tau_i = \tau \frac{i}{i_{\text{max}}}. \tag{2}$$

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 mRNA sub population is:

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

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.	$1/s$
$\mu(i)$	decapping 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 mRNA sub population is:

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

In closing, we note the parameters $i_{\text{max}}, \kappa, \mu$, and τ likely vary between genes.

2.2.1. Analytical steady state solutions of the capped transcript population

Analytical exploration of the model's capped system presents no closed form solution for the capped system. However, the model solution can be represented in the following form,

$$\vec{m} = \frac{\lambda}{\mu}\vec{p}_m \tag{5}$$

Where $\vec{m}$ is a vector of the steady state mRNA abundances in each polysomal class. $\vec{m}$ is calculated from by scaling the vector $\vec{p}$, which represents the distribution of the mRNA across the polysomal classes, by transcript production rate λ and the decapping rate μ scale s. While the individual components of $\vec{p}$ are functions of i, i_{max}, the translation initiation rate κ, the elongation rate τ_0 and μ and have no closed form solution, it is worth noting that because the probability distribution of the capped population must sum to 1, by definition, it follows that

$$\sum_{i=0}^{i_{\text{max}}} m_i = \lambda/\mu. \tag{6}$$

2.2.2. Analytical steady state solutions of the decapped transcript population

The solution for the decapped system is dependent on the underlying distribution of the capped system and can be represented as:

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

We can simplify the model by converting the mRNA quantity m_j to the probability p_j by 5. Additionally, for any $i = j$ where S_j is cumulative probability from i = class j to $i = i_{\text{max}}$.

$$S_j = \sum_{i=j}^{i_{\text{max}}} \vec{p_i} \tag{7}$$

Now the solution becomes,

$$\begin{aligned} m_0^* &= \frac{\lambda}{\delta} S_0 = \frac{\lambda}{\delta} \\ m_1^* &= \frac{\lambda}{\tau} S_1 \\ &\vdots \\ m_i^* &= \frac{\lambda}{i\,\tau} S_i \\ &\vdots \\ m_{i_{\text{max}}}^* &= \frac{\lambda}{i_{\text{max}}\,\tau} S_{i_{\text{max}}} \end{aligned} \tag{8}$$

Note that $S_0 = 1$ and $S_0 \geq S_1 \geq ... \geq S_i \geq ... \geq S_{i_{\text{max}}}$ dependant on the distribution of $\hat{\vec{m}}$ of the capped state.

2.3. Calculation of the decapped mRNA population

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

$$m_{tot}^* = \sum_{i=0}^{i_{\text{max}}} m_i^* = \frac{\lambda}{\delta} + \frac{\lambda}{\tau} S_1 + \ldots + \frac{\lambda}{i\tau} S_i + \ldots + \frac{\lambda}{i_{\text{max}}\tau} S_{i_{\text{max}}}$$

This can be further shortened to:

$$m_{tot}^* = \lambda(\frac{1}{\delta} + \frac{1}{\tau}\vec{S} \cdot \vec{l}) \tag{9}$$

Where $\vec{S}$ is a vector of all the cumulative sums and $\vec{l}$ is a vector of $1, 1/2, ..., 1/i, ..., 1/i_{\text{max}}$.

2.4. Probability distribution in the decapped state

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

$$p^*_0 = \frac{1}{1 + \frac{\delta}{\tau}\vec{S} \cdot \vec{l}} \tag{10}$$

$$p^*_j = \frac{S_j}{j(\frac{\tau}{\delta} + \vec{S} \cdot \vec{l})} \quad \text{for } j = 1, 2, ..., i, ..., i_{\text{max}} \tag{11}$$

2.5. Calculation of the total mRNA population and its distribution between capped and decapped states

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

$$M_{tot} = \frac{\lambda}{\mu} + \lambda(\frac{1}{\delta} + \frac{1}{\tau}\vec{S} \cdot \vec{l}) \tag{12}$$

To understand how mRNA is divided between we start with the probability of finding an mRNA in the capped state.

$$p_{mtot} = \frac{1}{(1 + \frac{\mu}{\delta} + \frac{\mu}{\tau}\vec{S} \cdot \vec{l})}$$

Then you calculate the odds,

$$odds_{\hat{m}} = \frac{1}{\mu(\frac{1}{\delta} + \frac{1}{\tau}\vec{S} \cdot \vec{l}})} \tag{13}$$

2.6. Calculating expected ribosomal load and protein production

The expected ribosomal load for either the capped or decapped state is calculated by:

$$E(ribosome) = \sum_{i=0}^{i_{\text{max}}} j \times p_i \tag{14}$$

Where $\vec{p_m}$ is the distribution in either state and i is the polysome class.

To find the global mean ribosomal load we obtain,

$$\text{Total Ribosomal Load} = p_{mtot} \times E(ribosome)_{mtot} + (1 - p_{mtot}) \times E(ribosome)_{mtot^*} \tag{15}$$

2.7. Numerical solution implementation in R

Code to solve the model was written in the R package Ribosome (`https://github.com/rurquidi/Ribosome`). To solve the capped subsystem of the model, the solve.tridiag algorithm from limSolve package (V 1.5.6) (Soetaert,K 2009). The decapped solution was obtained by using the capped solutions into 8. Utility functions, plots and statistics were created using R (v 3.6) (R core team), and data.table (v1.14.0) (Dowle 2021).

2.8. Data Sources

In order to biologically contextualize and illustrate our model's behavior, 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 (*saccharomyces cerevisiae*) and the plant *Arabidopsis thaliana*. To determine i_{max}, protein lengths are divided by the average number of codons covered by a ribosome, which is 9 codons (Figure 2A and C). The range of i_{max} is 48 $\pm$ 36 for yeast and 47 $\pm$ 30 for Arabidopsis. Protein lengths were extracted from the Ensembl (version 109) and Ensembl plants (version 56) respectively (Cunningham 2022, Yates 2022, Kinsella 2011). The decapping 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). We approximated gene specific μ from the half lives with the following:

$$\mu_i = \frac{ln(2)}{t_{1/2_i}}$$

Where $t_{1/2}$ is the half-life. The resulting range of μ is from $1.3 \times 10^{-}3 \pm 1.8 \times 10^{-}3$ for yeast and $1.7 \times 10^{-}4 \pm 2 \times 10^{-}4$ for Arabidopsis.

Translation initiation and average elongation rates (κ and τ_0) were obtained for Yeast from Duc and Song 2018. In Duc and Song 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.

$$\text{initiation to elongation ratio} = \kappa' = \frac{\kappa}{\tau_0} \tag{16}$$

This simplifies the model behavior to one generalized parameter with a unique response (Figure 2E). The initiation to elongation ratio 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 expression ranges just under 5 orders of magnitude.

The mRNA clearance rate δ only determines the accumulation of transcripts in the m_0^* class, which

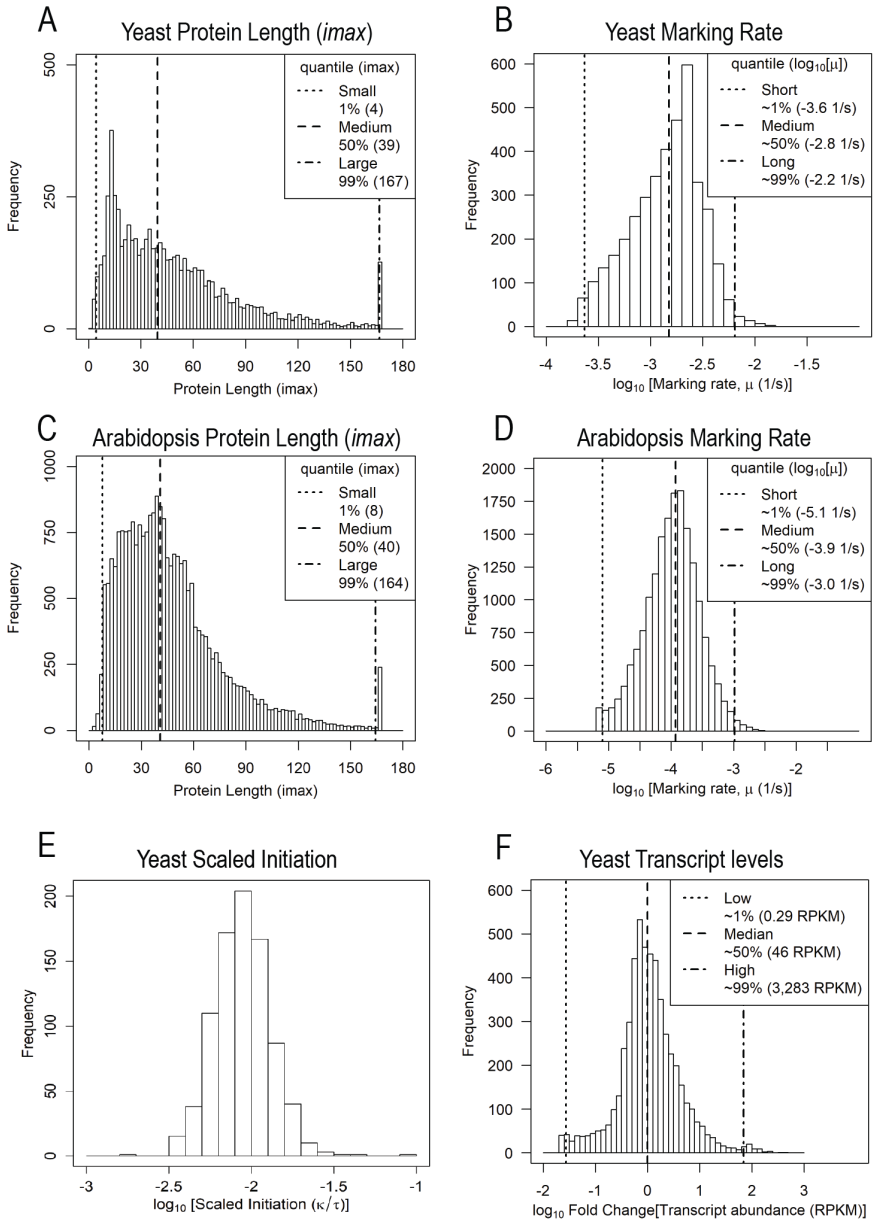
for simplicity of interpreting results has been set to be $>> \tau$ and thus will not accumulate transcripts in m_0^*.

The empirical mean ribosomal load (MRL) for the 850 genes in Duc and song 2018 was calculated from the mRNA-seq read per kilbase million (mRNA RPKM) and the ribo-seq footprints (RPF RPKM) from Weinberg 2016. The following equation was used.

$$MRL_i = \frac{RPF\ RPKM_i}{mRNA\ RPKM_i \times \frac{200}{lengthmRNA_i}} \tag{17}$$

Where the gene specific scaling factor $\frac{200}{lengthmRNA_i}$ corrects for the bias in read counts due to longer transcripts producing more fragments. The value 200 arises from the average fragment size of a library prep and can be adjusted according the experimental method used.

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

3.1. Model provides a unique distribution of mRNAs across polysome classes for each initiation to elongation ratio

1. The model predicts the abundances of the mRNA in the capped and decapped states as well as the mRNA's distribution across polysome classes.

   (a) Steady state solution of the capped class

      i. The analytical steady state solution 5 is composed of a vector of probabilities $\vec{p_m}$ that an mRNA is in polysome class i and a scaling term (the transcription rate λ divided by the decapping rate μ).

      ii. This solution highlights two separate roles of μ.

         A. First, the scaling term λ/μ determines mRNA abundance in the capped state.

         B. Second, the vector of probabilities is a function of the initiation rate κ, the elongation/termination rate τ and μ and is independent of λ.

      iii. Figure 3A shows the mRNA distribution in the capped state for four different initiation to elongation ratios κ', for a protein of median length (i_{max} of 39) with a low decapping rate (2×10^{-4}).

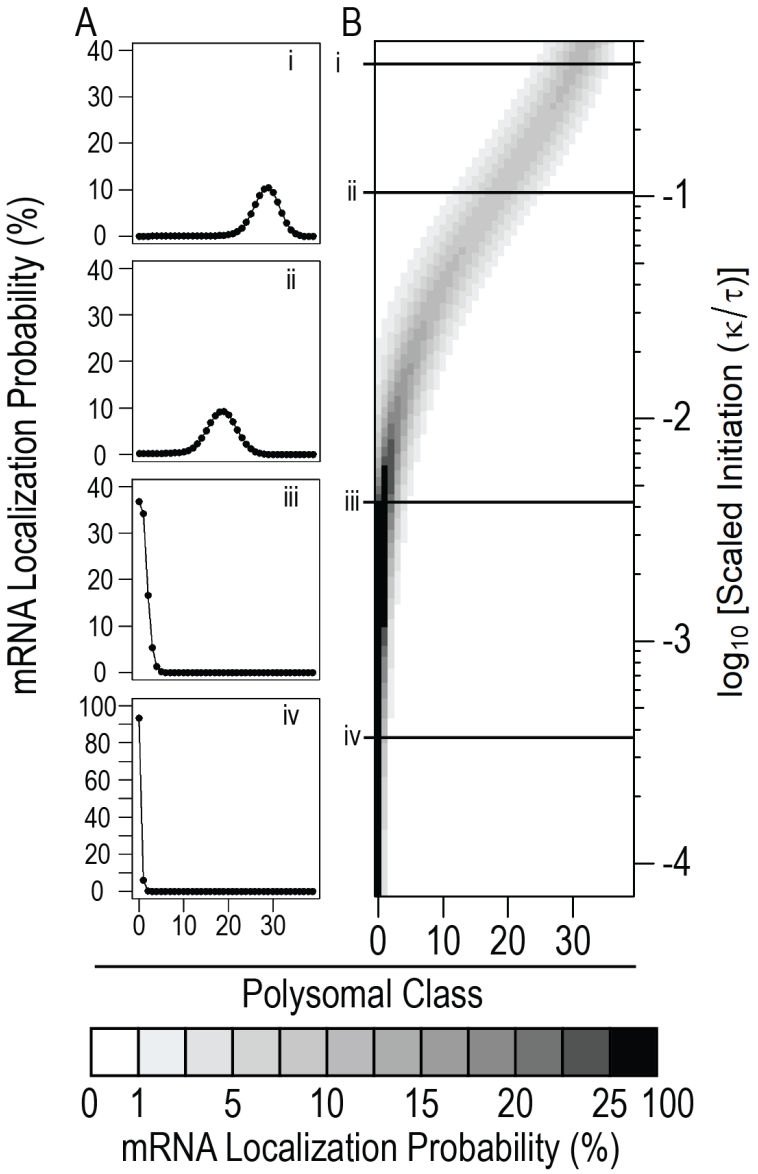
      iv. To summarize the model results across a range of parameters a heatmap where each row is the steady state distribution of mRNA at a particular κ' is shown (Figure 3B).

      v. The steady state density in the capped system is bounded at class 0 and class i_{max} and can be roughly approximated by a truncated gaussian.

      vi. When $\kappa' << \tau$, the distribution concentrates at low i near the $i = 0$ boundary.

      vii. As κ' increases, the steady state distribution moves towards higher i.

Figure 3: mRNA distribution in capped state. A) Distribution profiles for four scaled initiation values i) 2×10^{-1} ii) 1.03×10^{-1} iii) 3×10^{-3} iv) 2×10^{-4} 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 low decapping rate of 2×10^{-4} (99$^{\text{th}}$ percentile). Color bar shows probability of finding mRNA in particular ribosomal class.



(b) Steady state solution of the decapped class

i. The whole system is again scaled by the transcription rate λ.

ii. The analytical steady state solution eq. 8 can be understood in two parts.

iii. m_0^* and the remaining m_i^* for $i > 0$.

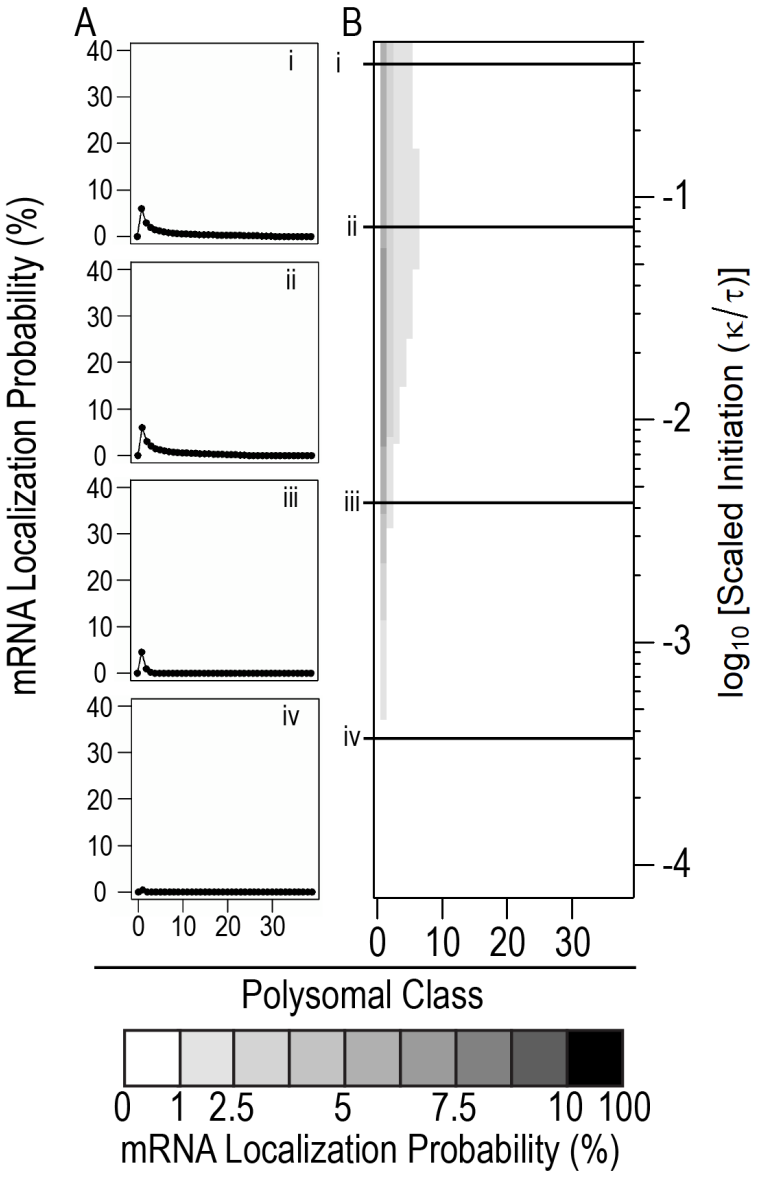
iv. m_0^* is solely determined by the ratio of the mRNA clearance rate 1/δ.

v. The remaining decapped polysome classes depend on the elongation/termination rate τ and the distribution of $\hat{\vec{m}}$.

vi. Figure 4A shows the mRNA distribution in the decapped state for four different initiation to elongation ratios κ', for a median length protein with a low decapping rate (2×10^{-4}) and the full range of κ' are shown in the heatmap in Figure 4B.

vii. The mRNA distributions are centered around low i and are monotonically decreasing.

Figure 4: mRNA distribution in decapped state. A) Distribution profiles for four scaled initiation values i) 2×10^{-1} ii) 1.03×10^{-1} iii) 3×10^{-3} iv) 2×10^{-4} 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 low decapping rate of 2×10^{-4} (99$^{\text{th}}$ percentile). Color bar shows probability of finding mRNA in particular ribosomal class.



(c) Steady states solution for the full system

i. The full system combines the mRNA distributions from the capped and decapped states.

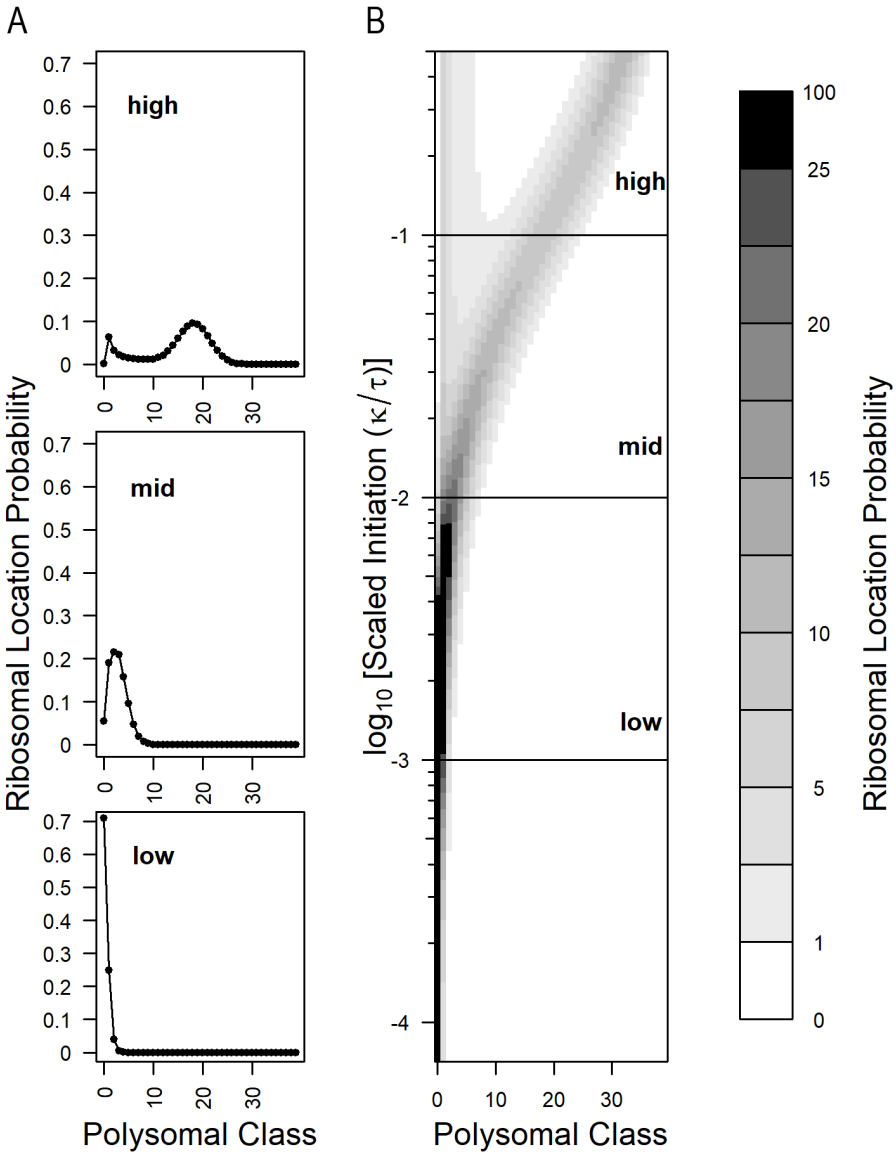
A. Figure 5A shows the mRNA distribution in the full system for three values of κ', for a median length protein with a low decapping rate (2×10^{-4}) and the full range of κ' are shown in the heatmap in Figure 5B.

B. The system is unimodal at low $\kappa' < 0.01$ when the capped and decapped distributions overlap around low i (Figure 5A mid and low).

C. As $\kappa' > 0.01$ increases, the full distribution becomes bimodal (Figure 5A high).

D. The peak at low i representing the decapped distribution, and the higher gaussian peak representing the capped distribution.

Figure 5: mRNA distribution in the full system. A) Distribution profiles for three scaled initiation values low) 1×10^{-3} mid) 2×10^{-2} and high) 1×10^{-1} 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 low decapping rate of 2×10^{-4} (99$^{\text{th}}$ percentile). Color bar shows probability of finding mRNA in particular ribosomal class.



3.2. Higher decapping rates reduce capped state ribosomal loads

1. To explore the role of mRNA stability on mRNA populations we varied the decapping rate μ from the 1st,

   50th and 99th percentile values as determined from Presnyak 2015.

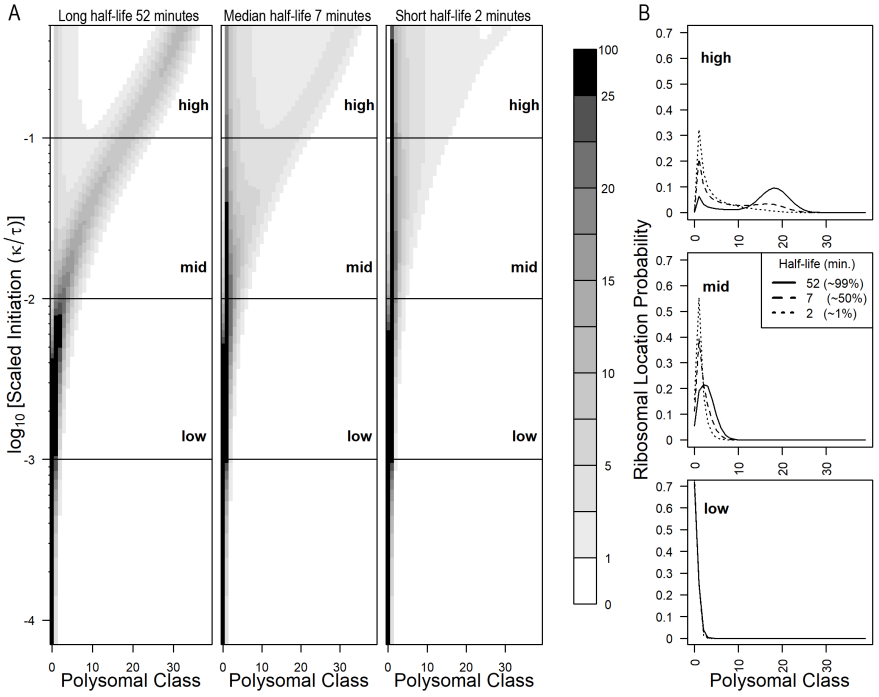
   (a) As μ increases the distribution of mRNAs changes in two ways.

      i. First there is shift to lower ribosomal classes in the capped state (Figure 6).

      ii. This is likely due to the mRNAs leaving the capped state at a higher rate and driving the equilibrium towards lower ribosomal loads.

      iii. Secondly, as half-life decreases, a larger proportion of the mRNA is found in the decapped state. This is further explored later.

Figure 6: Higher decapping rates μ reduce ribosome load in the capped system in yeast. A) Heatmaps for the full system. Left) low μ (2×10^{-4} /s) Center) median μ (1.7×10^{-3} /s) Right) μ (5.7×10^{-3} /s) B) individual density profiles for low (0.001), mid (0.01) and high (0.1) scaled initiation values for each μ All results calculate with $i_{\max}$ = 39.



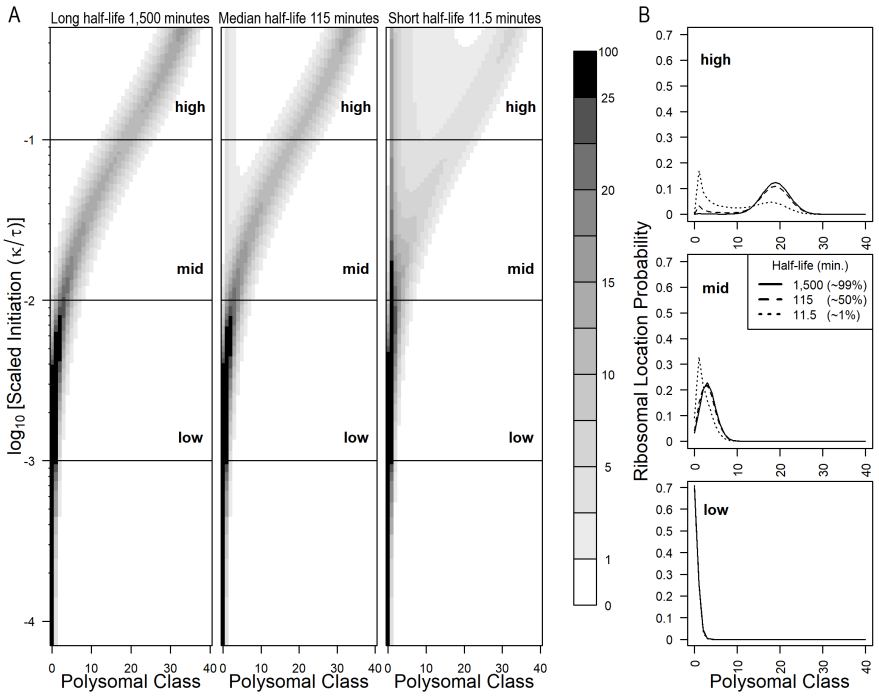
2. Plants and other multicellular eukaryotes tend to have slower translation initiation and elongation rates as well as slower cell division when compared to single celled organisms such as yeast.

   (a) This is highlighted by the current gold standard study of mRNA half-lives in the model organism *Arabidopsis thaliana*, where the decapping rate μ measured are ten to one hundred times lower than those in yeast.

   (b) To explore the effect of lower μ in Arabidopsis, we ran the model using the same initiation to elongation ratios as in yeast, the median Arabidopsis i_{max} of 41.

   (c) As expected, the longer half-lives have a higher mRNA distribution (Figure 7) and are mostly in the capped state (Figure 10).

Figure 7: Low decapping rates μ in Arabidopsis result in a smaller effect on mRNA distributions in the capped system. A) Heatmaps for the full system. Left) low μ (7.7×10^{-6} /s) Center) median μ (1×10^{-4} /s) Right) μ (1×10^{-3} /s) B) individual density profiles for low (0.001), mid (0.01) and high (0.1) scaled initiation values for each μ. All results calculate with i_{max} = 41.



3.3. *decapping rate and ribosomal load determine mRNA distribution between capped and decapped states*

1. As shown in previous results, higher decapping rates μ lead to lower MRL in the capped state and increase mRNA abundance in the decapped state.

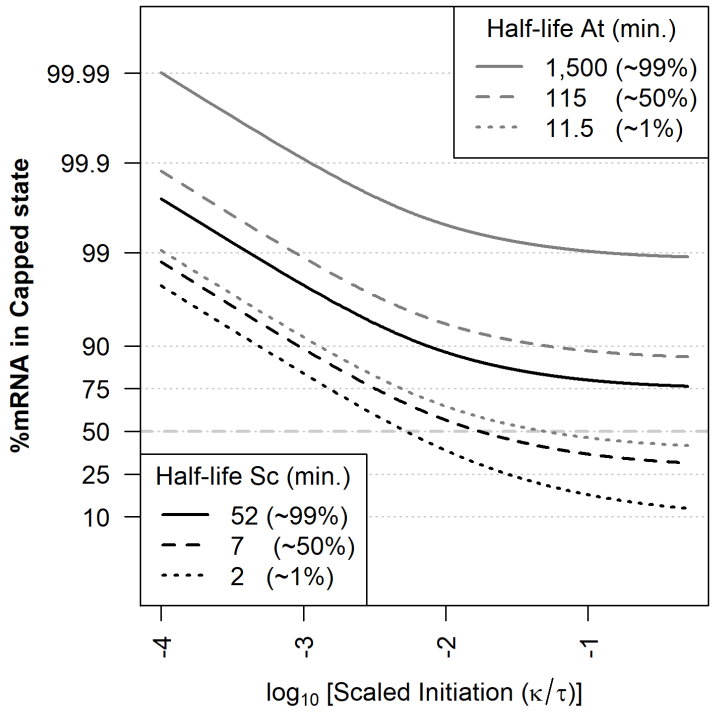
   (a) Using 13, we can determine how much of the mRNA population is in the capped state.

   (b) We produced output across all scaled initiation values κ' and under the 1st, 50th and 99th percentiles for decapping rates in both yeast and Arabidopsis (Figure 8).

   (c) We note two patterns. First as the κ' increases, so does the amount of mRNA in the decapped class $\hat{\vec{m}}^*$ increases.

   (d) Secondly and similarly, higher μ shifts mRNA population from the capped state to the decapped state as previously seen in Figures 6 and 7.

Figure 8: Percentage of mRNA in the capped state for a range of decapping rates in yeast (low μ (2×10^{-4} /s), median μ (1.7×10^{-3} /s), high μ (5.7×10^{-3} /s)) and Arabidopsis(low μ (7.7×10^{-6} /s, median μ (1×10^{-4} /s), high μ (1×10^{-3} /s)).



3.4. At steady state protein production is scales with coding sequence length i_{max}

1. At steady state the MRL increases with coding sequence length and begins to assymptote at high initiation to elongation ratio κ' (Figure 9).

   (a) While the capped state MRL is always greater than the decapped MRL, the MRL of the whole system is defined by both capped and decapped MRLs as well as the transcript abundance in each state as shown in eq 17.

   (b) As protein length i_{max} increases, mRNAs enter the decapped state at higher polysome classes.

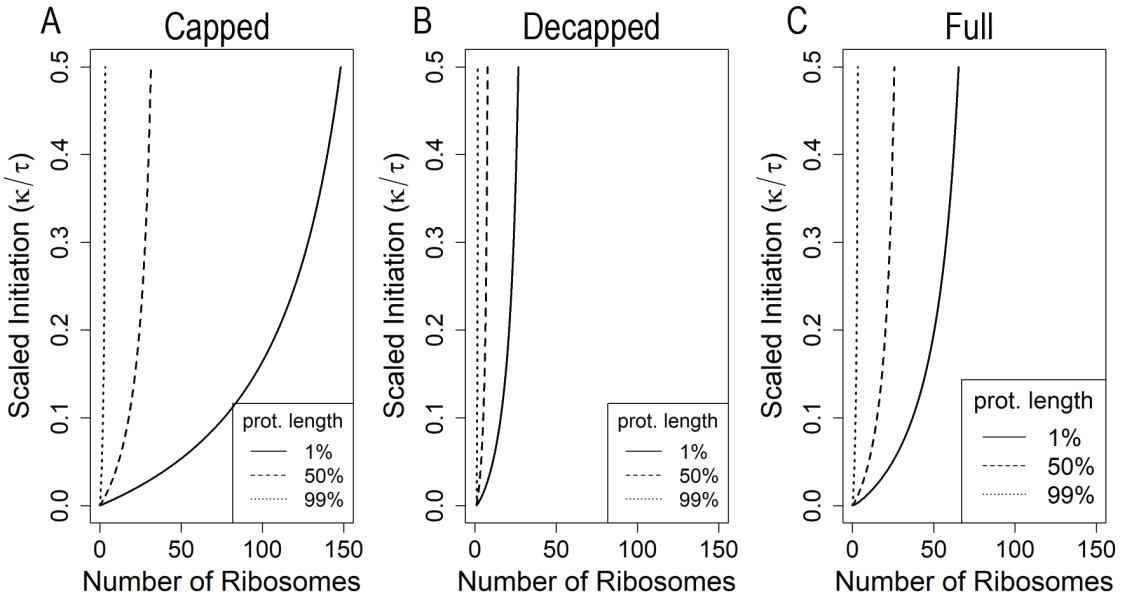
   (c) Therefore ribosomes take longer to clear the mRNA, and thus increase the contribution from the decapped state.

   (d) For κ'= 0.1, the percentage of the mRNA in the capped class is 99% 78% and 35% for i_{max} of 4, 39 and 194 respectively.

   (e) The i_{max} dependence is captured in the 1/τ term in eq 13, remembering that $\tau = \tau_c/9 * i_{\text{max}}$.

R: Mike, you were right. There is a DDI effect on mean ribosomal density. After looking into it deeper I found the old findings completely wrong. Created this new set of results to properly explain the MRL behavior

Figure 9: The mean ribosomal density on a transcript is dependent on coding sequence length. MRL per transcript is higher for longer transcripts. A) Capped state B) Decapped state C) Full system. Yeast parameters were used i_{max} = 4 (1st percentile), 39 (50th percentile), 194 (99th percentile), low decapping rate (2.2×10^{-4} /s), over the full scaled initiation range 0.0001- 0.5.



3.5. Decapped state can be a significant source of protein production

1. Protein production rate (PPR) is a function of full system MRL $\times\tau$ and plots normalized to highest protein output are shown in Figure 10. **R:** may now be superfluous

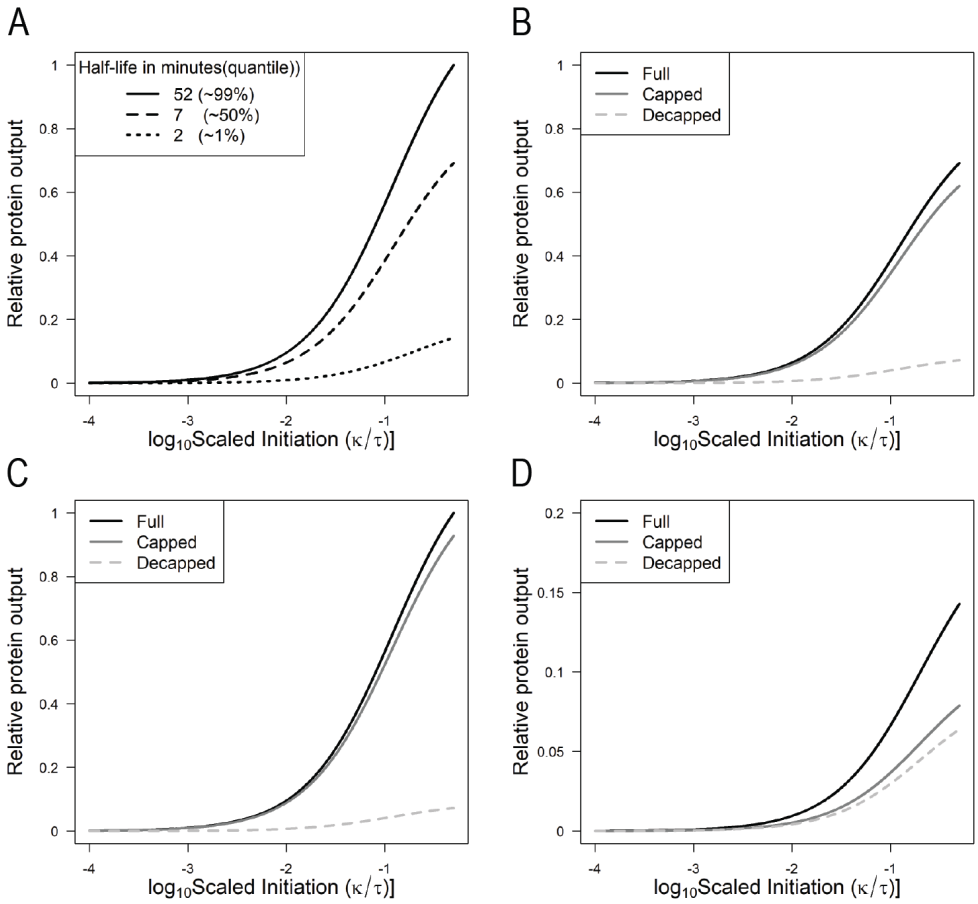
   (a) As the decapping rate μ increases it reduces the capped and uncapped MRL as well as shifting transcript abundance to the decapped state (Figure 10A).

   (b) Each of the three cases in Figure 10 A, has been broken down into the PPR contributions from the capped and decapped states (Figure 10 B-D).

   (c) A surprising finding from out model is that when μ is high (5.7×10^{-3} /s), 41% of all protein production can arise from the decapped state (Figure 10 D).

   (d) The reason behind this is despite the the relative MRL of the decapped state being lower than the capped state as scaled initiation rises, the amount of mRNA in the decapped state rises faster.

Figure 10: Estimated average protein production in yeast. A) Protein production across different decapping rates low μ (2.2×10^{-4} /s), median μ (1.7×10^{-3} /s), high μ (5.7×10^{-3} /s). Total protein production is normalized to the maximal protein production across all parameters. B-D) Contribution of capped and decapped states to total protein production. B) Low decapping C) Medium decapping D) High decapping.



3.6. Model validation

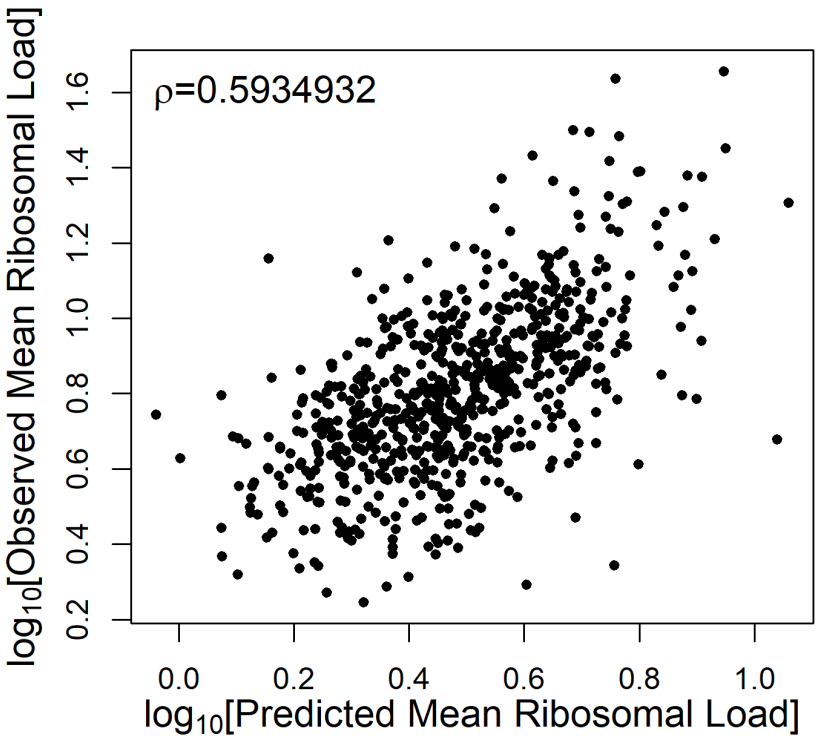
1. Gene specific MRL 17 were calculated for the genes analyzed in Duc and Song 2018 and compared to the empirical MRL calculated from raw data from Weinberg 2016 (Figure 11).

   (a) Model predictions of MRL showed a significant positive correlation to empirical MRLs.

   (b) This result is impressive as the model performs well despite no model fitting being performed.

   (c) Model performance is further corroborated with single molecule imaging analyses. Rescaling ribosome abundances from each single molecule study to an i_{max} of 39 results in loads of 1, 2.4 and 4 ribosomes from (Morisaki 2016), 3 ribosomes (Yan 2016), 4 ribosomes (Wang 2016) and 1.6 (Wu 2016).

Figure 11: Predicted mean ribosomal loads coincide with observed mean ribosomal loads from Weinberg 2016. Using the 850 genes from Duc and song 2018, decapping rates from Presnyak 2015 were mined. Gene specific MRL were calculated and compared to the empircal MRL. Spearman's ρ was calculated and found to be significant, pvalue $< 10^{-16}$



4. Discussion

In this study we develop, analyze, and validate a novel coupled ODE model of mRNA polysome classes which includes the contributions of mRNA transcription, the initiation, elongation (implicitly), and termination of translation as well as mRNA degradation through 5' decapping and cotranslational decay.

4.1. Model Formulation & Structure

Although our model is only a very simplified description of the mRNA polysome population and, in turn, protein translation, it studies the interaction of protein translation with the process of mRNA degradation, a topic which has been underexplored (Yadav 2021). The process of translation is dependent on the underlying population of capped, translationally competent mRNAs. However, empirical measurements suggest that $\sim 12\%$ of transcripts are undergoing co-translational decay (Pelechano 2015). To undergo co-translational decay, the 5' cap has to removed and exonucleases trail behind the last loaded ribosome on a transcript, processing codons as they exit behind the ribosome. 5' decapping is a common pathway in many organisms and accounts for decay for $\sim 68\%$ of Arabadopsis genes (Sorenson 2018). Our model includes 5' mRNA decapping followed by cotranslational decay, permiting the analysis of the decapped mRNA state (called degratome in Ma 2020), changes in MRL for capped and decapped states and the contribution of cotranslational decay to protein production.

In addition to being more biologically realistic, structuring the mRNA population by its polysome classes (ribosome load) and the status of its 5' cap allows us to understand how the the rates mRNA production λ, decapping μ, protein elongation τ, and the clearance rate δ of decapped and ribosome free mRNAs $\hat{m}_0^*$ shape the steady state distribution of a gene's mRNA population across polysome classes and capping state (Figures 3-5). Overall, we find that

Analytical and numerical solutions show transcription rate λ acts as a scaling factor such that the abundances of all of the mRNA polysome classes are proportional to λ. In other words, the total abundance of the capped and decapped mRNA polysome classes $\hat{\vec{m}}$ and $\hat{\vec{m}}^*$ are simply proportional to λ (see (6) where $\sum_{i=0}^{i_{\max}} m_i = \lambda/\mu$ and (9), respectively). The fact that the abundance of the entire capped and decapped mRNA polysome classes are proportional to the transcrption rate λ is consistent with intuition, as λ increases, so does the abundance of both the capped and decapped populations. Similarly, the fact that the abundance of the capped mRNA polysome classes declines as an inverse funcdtion of the decapping rate μ is also consistent with intuition. Because it is the ratio of λ and μ, rather than their individual values, that determine the size of the capped mRNA polysome classes, our

model indicates that there will be an infinite set of transcription λ and decapping rates μ that can result in the same population size of capped mRNA polysomes. All else being equal, this result suggests that these rates could vary greatly between genes with similar abundances.

The fact that changes in the mRNA transcription rate λ only scales, rather than shapes, the relative distribution of mRNA polysome classes allows us to turn our focus to how the remaining model parameters, $\kappa' = \kappa/\tau, \mu,$ andδ alter the *relative* distribution of the capped and decapped mRNA polysome classes $\hat{\vec{m}}$ and $\hat{\vec{m}}^*$, respectively.

For example, focusing on the relative distribution of the capped mRNA polysome classes $\hat{\vec{m}}$, our model indicates that it is the ratio of scaled translation initiation $\kappa' = \kappa/\tau$ to decapping μ which determines the distribution, and thus mean ribosome load, of $\hat{\vec{m}}$ (Figure 3). For example, when $\kappa'/\mu \ll 1$, the distribution of capped mRNA polysome classes $\hat{\vec{m}}$ is greatest in the ribosome free polysome class $i = 0$ and declines rapidly with with ribosome load i. As κ'/μ increases, the distribution of capped mRNA polysome classes shifts away from the lower bound of $i = 0$ appears to follow a truncated gaussian distribution. In contrast, it is only at very high and generally unrealistic values of κ'/μ (i.e. $\kappa'/\mu > 10$) do we see the peak of the distribution of capped mRNA polysome classes approach i_{max}.

Shifting our focus to the relative distribution of the decapped mRNA polysome classes $\hat{\vec{m}}^*$, our model provides a number of important insights . Surprisingly, in the special case of the decapped, ribosome free mRNA class $\hat{m}_0^*$, we find its abundance is decoupled from the dynamics of the rest of the population. This decoupling has a number of important implications. For example, the steady state abundance of $\hat{m}_0^* = \lambda/\delta$ and, thus, depends only on the ratio of the mRNA transcription rate λ to the mRNA clearance rate δ (equation 8). If the transcription rate λ of new, capped, but ribosome free mRNAs $\hat{m}_0$ is substantially lower than the per capita mRNA clearance rate of decapped, ribosome free mRNAs δ, such that $\lambda \ll \delta$, then our model predicts that there will be few mRNAs in the $\hat{m}_0^*$ class $\hat{m}_0^* \ll 1$. Because $\hat{m}_0^*$ has no impact on the rest of the mRNA population, this result allows us to greatly simplify our analysis since we need not consider $\hat{m}_0^*$ nor the parameter δ.

Focusing now on the steady state abundance of the ribosome occupied decapped mRNA polysome classes, i.e. $\hat{m}_i^*$ where $i > 0$, we find that the distribtuion of $\hat{m}_i^*$ depends on the gene specific ribosome elongation rate τ_0 (where 'elongation' includes the ribosome's reading of the mRNA's stop codon) and the distribution of capped mRNA $\hat{m}_i$ with $i > 0$ (Figure 4). This finding implies that because the density of $\hat{m}_i^*$ monotonically decreases with i, the distribution of decapped mRNA polysome classes is skewed and dominated by lower polysomal classes. This is monotonic decline coupled with the fact

that the decapped ribosome free polysome class $\hat{m}_0$ does not contribute to protein production, implies that MRL of the decapped mRNA polysomes is must be less than MRL the of the capped mRNA polysomes. Thus, while the decapped class does contribute to protein production, substantially under particular parameter values ($\mu > 5.7 \times 10^{-3}$ or $i_{\text{max}} >> median i_{\text{max}}$), its contributution to the mRNA population's protein production will always be less than 50%.

Combined distributions of the capped and decapped polysome class

1. Distribution is strictly unimodal when $\sum \hat{\vec{m}} >> \sum \hat{\vec{m}}^*$, effectively when $\kappa'/\mu \ll 1$ (Figure 6 and 7).

2. The distribution is bimodal otherwise. The bimodal peaks of this arised from the distributions from the capped and decapped states.

3. Increasing decapping rate μ and scaled translation initiation rate κ/τ both increase the proportion of transcripts in the decapped state.

   (a) The model formalizes the interplay between mRNA decapping μ and translation initiation κ.

   (b) As expected, increasing μ results in an increase in the proportion of decapped transcripts $\hat{\vec{m}}^*$ compared to capped transcripts $\hat{\vec{m}}$.

   (c) However, increasing ratio of initiation to elongation rates κ/τ also results in an increase of $\hat{\vec{m}}^*$.

   (d) As κ/τ increases the MRL of the capped population $\hat{\vec{m}}$ increases, transcripts enter the decapped state at higher polysomal classes and thus take longer to reach m_0^*.

4. A surprising prediction from our model is that genes with high decapping rates (e.g. $\mu \sim 5 \times 10^{-3}$ or a half-life of ~ 120 sec) has almost (but never more than) half of its protein production coming from the decapped mRNA polysome classes (Figure 11).

   M: The following points be incorporated into the previous discussion of MRL and protien production (4.1).

   (a) As μ increases, a greater proportion of the MRL arises from the decapped states.

   (b) This increase is mainly a function of a shift in transcripts to the decapped state $\hat{\vec{m}}^*$ due to larger μ.

   (c) The decapped MRL $\leq$ the capped MRL.

(d) The high relative protein production from the decapped state suggests that high decapping rate transcripts can produce more protein than expected from their capped MRL alone.

(e) As we are assuming the mRNA clearance rate $\delta >> \lambda$, $\hat{m}_0^*$ will be negligible.

(f) Thus, our current results act as an upper bound of the protein production contribution from the decapped class.

R: Relative protein production here $MRL \times \sum \hat{m}$/total protein production is not a rate, but a comparison of the relative amount of protein produced by a state over a set period of time. This is explained (poorly) in the results. I chose this over $E(i) \times \tau$ because while the off rate for the capped class is simply τ, the off rate for ribosomes in the decapped class is some constant $c \times \tau$ c is determined by the distribution in the capped class. Can I numerically calculate it? Yes. But I dont have an analytical solution for it.

4.2. Model Validation

In addition to studying the general behavior of our model, we validate this behavior using empirically based parameter values from the literature. In general, we find that our model's predictions of mRNA distributions, when parametrized with biologically relevant values, are highly consistent with a wide range of empirical data. For example, we predicted MRL using empirical values for initiation elongation ratio κ' (Duc and Song 2018) and decapping rate μ (Presnyak 2015) and compared them to the empirical MRL from (Weinberg 2016) and found a strong correlation despite having performed no fitting. This supports the idea that the model is a useful representation of the complex processes underlying protein production. The κ' estimates utilized are only for highly translated genes (16% of alll detected genes), and most others would fall in a range of $\kappa' < 0.01$ (Duc and Song 2018). Taking the overall low κ' values, our model predicts that a median length protein of i_{max} =39 (351aa) would have 10 or fewer ribosomes loaded, which agree with the predominance of low polysomes (<10) seen polysome gradient traces (Lokdarshi 2020, Dasgupta 2023). By the same logic, we find that single molecule measurements of translation (Morisaki 2016, Yan 2016, Wang 2016, Wu 2016, Section 3.6) all fall in the same low polysome range. Finally, the fraction of mRNA predicted in the capped and decapped class are consistent with population wide estimates (Pelechano 2015, Figure 10).

4.3. Model limitations, extensions and future work

Our model's assumptions about the process of mRNA decapping, the continued competence of ribosomes present prior to decapping, and degradation of mRNA solely from the decapped and ribosome free class m_0^* closely resembles the biological process of co-translational mRNA decay. While the existence of co-translational mRNA decay is well established (Sorenson 2018, Pelechano 2015), other mechanisms exist with different outcomes for translation. 3' decay results in no ribosomes terminating and would send all transcripts into the m_0^* class . Mechanisms utilizing endonucleolytic decay due to no go decay or nonsense mediated decay would potentially allow ribosomes downstream of the cleavage site to terminate but not those upstream (Urquidi-Camacho 2020, Merchante 2017). Thus,

depending on the ditribution of mRNAs in the capped class, and the site of the endonucleolytic decay a transcript in m_i would end up in m_j, where $j < i$. Developing a more quantitative understanding of how different factors affect a gene's mRNA stability and, in turn, protein expression, relevant to a wide range of applied molecular biology (e.g. the design of efficient heterologous genes expression and mRNA vaccines) (Cheng 2023 viruses, Boo and Kim 2020).

Current debate is focused on the contributions of the protective effects of ribosome association vs. ribosome stalling to mRNA transcript stability. While our model currently does not include the protective effects of translation or stall prone codons, it should be possible to do so. The protective effects of ribosomal loading which could be modeled by making the decapping rate μ by $(1-i/i_{\mathsf{max}})$, or having one, higher decapping rate for m_0 and a lower decapping rate for the other polysomal classes. The protective effects of translation could increase per ribosome, but eventually at high MRL could trigger ribosome associated decay pathways through ribosomal collisions. This would require analysis on an individual transcript basis. Our model does not consider codon specific effects such as pausing sites, difficult to fold regions of a protein or codon optimality, or protein quality control (Wu and Bazzini 2023). Pausing sites could be addressed by splitting each polysome class into two regions and could approximate a ribosome flow model of only two regions, a 5' and 3', split by the pausing site. Current models of translation focus mainly on the behavior of the average transcripts. However this ignores tha

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 decapping 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, $\kappa(i_{\text{max}}) = 0$, 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 at a 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. $\tau(1) = average\ elongation\ rate/(9 \cdot i_{\text{max}})$. 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 $1/s$.

5.1. 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,

$$\vec{M}' = \boldsymbol{F}\vec{M} + \vec{B}, \tag{18}$$

where $\vec{M} \in \mathbb{R}^{2(i_{\text{max}}+1)}$ is a vector of all state variables, ordered here as m_0, m_1, ..., $m_{i_{\text{max}}}$, m_0^*, m_1^*, ..., $m_{i_{\text{max}}}^*$, $\vec{M}'$ is the vector containing the first derivatives of $\vec{M}$ with respect to time, $\boldsymbol{F} \in \mathbb{R}^{2(i_{\text{max}}+1)\times 2(i_{\text{max}}+1)}$ is the matrix representing the full system (Equation **??**), and $\vec{B} \in \mathbb{R}^{2(i_{\text{max}}+1)}$ 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 **??**.

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{U} & \boldsymbol{0} \\ \\ \boldsymbol{\mu} & \boldsymbol{R} \end{pmatrix}. \tag{19}$$

The upper-left block, $\boldsymbol{U}$, corresponds to the capped state variables, where $\boldsymbol{U}$'s general form is provided in Equation **??**. The upper-right block is a matrix of all zeros, $\boldsymbol{0} \in \mathbb{R}^{i_{\text{max}}+1 \times i_{\text{max}}+1}$. Using $\boldsymbol{I}$ to represent the $i_{\text{max}}+1 \times i_{\text{max}}+1$ 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 **??**.

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

$$\boldsymbol{U} = \begin{pmatrix} -(\kappa_0 + \mu_0) & \tau_0 \frac{1}{i_{\text{max}}} & & & & & \\ \kappa_0 & \left(1 - \frac{1}{i_{\text{max}}}\kappa_0 + \mu_0 + \tau_0 \frac{1}{i_{\text{max}}}\right) & \tau_0 \frac{2}{i_{\text{max}}} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & & 1 - \frac{(i-1)}{i_{\text{max}}}\kappa_0 & -\left(1 - \frac{i}{i_{\text{max}}}\kappa_0 + \mu_0 + \tau_0 \frac{i}{i_{\text{max}}}\right) & \tau_0 \frac{i+1}{i_{\text{max}}} & \\ & & & & \ddots & \ddots & \ddots \\ & & & & & \frac{1}{i_{\text{max}}}\kappa_0 & -\left(\mu_0 + \tau_0 \frac{i_{\text{max}}}{i_{\text{max}}}\right) \end{pmatrix} \tag{20}$$

In the representation given in Equation **??**, all blank entries are 0. The $(i_{\text{max}} - 1)^{\text{th}}$ row has been suppressed in Equation **??**, 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 **??**),

$$\boldsymbol{R} = \begin{pmatrix} -\delta & \tau_0 \frac{1}{i_{\text{max}}} & & & & & & \\ & -\tau_0 \frac{1}{i_{\text{max}}} & \tau_0 \frac{2}{i_{\text{max}}} & & & & & \\ & \ddots & \ddots & & & & & \\ & & -\tau_0 \frac{i-1}{i_{\text{max}}} & \tau_0 \frac{(i+1)}{i_{\text{max}}} & & & & \\ & & & \ddots & \ddots & & & \\ & & & & -\tau_0 \frac{(i_{\text{max}}-2)}{i_{\text{max}}} & \tau_0 \frac{i_{\text{max}}}{i_{\text{max}}} & \\ & & & & & -\tau_0 \frac{i_{\text{max}}}{i_{\text{max}}} & \end{pmatrix}, \tag{21}$$

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

5.1.1. 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

$$\vec{m}' = \boldsymbol{U}\vec{m} + \vec{b}, \tag{22}$$

where $\vec{m} \in \mathbb{R}^{i_{\text{max}}+1}$ is the vector of capped state variables ordered m_0, ..., $m_{i_{\text{max}}}$, $\vec{m}'$ is the vector containing the first derivatives of $\vec{m}$ with respect to time, $\boldsymbol{U} \in \mathbb{R}^{i_{\text{max}}+1 \times i_{\text{max}}+1}$ is the matrix representing the capped subsystem (Figure **??**), and $\vec{b} \in \mathbb{R}^{i_{\text{max}}+1}$ 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.

5.2. The density dependent initiation model

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 $1 - i/i_{\text{max}}$ and, in turn, our translation initiation rate function can be defined as,

$$\kappa(i) = \kappa_0 \left(1 - \frac{i}{i_{\text{max}}}\right), \tag{23}$$

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:

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

and the decapped subsystem is unchanged.

5.2.1. Capped state steady state solution

1. m

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.

$$\frac{dm_0}{dt} = \lambda + \mu m_0 \tag{24}$$

which, at equilibrium results in,

$$m_0 = \lambda/\mu \tag{25}$$

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 decapping 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:

$$\frac{\lambda}{\mu} = \sum_{i=0}^{i_{\max}} m_i, \tag{26}$$

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:

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

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,

$$\vec{m} = \frac{\lambda}{\mu}\vec{p}_m \tag{27}$$

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 μ = 0, both solutions are indeterminate. However, the determinant free solution allows for us to explore what the transcript distribution would be when μ=0.

5.2.2. Decapped Subsystem steady state solution

Starting with the decapped subsystem of equations:

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

We get the following solutions at steady state:

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

We can rearrange the solutions and simplify to find,

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

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

$$
\frac{\lambda}{\mu} = \sum_{i=0}^{i_{\mathsf{max}}} m_i \tag{28}
$$

Therefore,

$$
1 = \frac{\mu}{\lambda} \sum_{i=0}^{i_{\mathsf{max}}} m_i \tag{29}
$$

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

$$S_j = \frac{\mu}{\lambda} \sum_{i=j}^{i_{\max}} m_i \tag{30}$$

Now the solution becomes,

$$\begin{aligned} m_0^* &= \frac{\lambda}{\delta} S_0 = \frac{\lambda}{\delta} \\ m_1^* &= \frac{\lambda}{\tau} S_1 \\ &\vdots \\ m_i^* &= \frac{\lambda}{i\,\tau} S_i \\ &\vdots \\ m_{i_{\max}}^* &= \frac{\lambda}{i_{\max}\,\tau} S_{i_{\max}} \end{aligned} \tag{31}$$

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

$$m_{tot}^* = \sum_{i=0}^{i_{\max}} m_i^* = \frac{\lambda}{\delta} + \frac{\lambda}{\tau} S_1 + \ldots + \frac{\lambda}{i\tau} S_i + \ldots + \frac{\lambda}{i_{\max}\tau} S_{i_{\max}} \tag{32}$$

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

$$m_{tot}^* = \lambda(\frac{1}{\delta} + \frac{1}{\tau} \vec{S} \odot \vec{l}) \tag{33}$$

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 $S_0 \geq S_1 \geq ... \geq S_i \geq ... \geq S_{i_{\max}}$. 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 mRNA clearance 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 decapping 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,

$$p_0^* = \frac{1}{1 + \frac{\delta}{\tau}\vec{S} \odot \vec{l}}$$
$$p_j^* = \frac{S_j}{j(\frac{\tau}{\delta} + \vec{S} \odot \vec{l})} \quad , for j = 1, 2, ..., i, ..., i_{\text{max}}$$

5.3. Complete system mRNA population

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

$$M_{tot} = \lambda(\frac{1}{\mu} + \frac{1}{\delta} + \frac{1}{\tau}\vec{S} \odot \vec{l}) \tag{34}$$

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,

$$odds_m = \frac{p_{mtot}}{1 - p_{mtot}} \tag{35}$$
$$odds_m = \frac{\frac{1}{(1+\frac{\mu}{\tau}\vec{S}\odot\vec{l})}}{1 - \frac{1}{(1+\frac{\mu}{\tau}\vec{S}\odot\vec{l})}} \tag{36}$$

It simplifies to,

$$odds_m = \frac{1}{\frac{\mu}{\tau}\vec{S} \odot \vec{l}} \tag{37}$$
$$log_{10}(odds_m) = -log_{10}(\frac{\mu}{\tau}\vec{S} \odot \vec{l}) \tag{38}$$

5.4. Additional text

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.

.This is due to the fact that many combinations of κ and τ can yield the same scaled translation initiation rate (e.g. κ =0.02 and $\tau * 9 * i_{\text{max}}$ = 2, and κ=0.04 and $\tau * 9 * i_{\text{max}}$=4, both yield κ/τ=0.01).

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 mRNA clearance 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 9. 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.