Modeling mRNA Populations

R.A. 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

Understanding the underlying mechanisms of protein production is essential to understanding life and for biotech applications. Translation and mRNA degradation are key coupled processes in the cell, which regulate protein output. While both translation and mRNA degradation have been extensively studied through modeling, there is little work exploring their interaction. Here we introduce a novel coupled ODE model which integrates mRNA transcription, 5' mRNA degradation and translation. Using empirically derived parameter estimates, our model predicts the protein productions from two populations of mRNA: Stable and actively translating 5' capped mRNA and decapped mRNA undergoing cotranslational decay. We are able to predict the capped and decapped mRNA population abudances, distribution of ribosomes loaded onto transcripts and the mean number of ribosomes associated to a transcript. Surprisingly, we find that genes with a high decapping rate (short half life) will produce just under half their protein from decapped transcripts. Our model proves useful toward understanding a more complete view of RNA biology and protein production and acts as a stepping stone for future development.

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

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

1. Introduction

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 encompasses three steps: initiation, elongation and termination and is reviewed in (Browning and Bailey-Serres 2015, Urquidi Camacho 2020).
   (a) In eukaryotes, translation initiation consists of a ribosome binding to the capped 5' end of an mRNA molecule.
   (b) Following binding to the mRNA a Ribosome proceeds to scan the mRNA until it encounters a start codon and initiates protein synthesis.
   (c) Protein is synthesized through the process of elongations, as the ribosome adds one amino acid at a time to the nascent peptide chain.
   (d) Finally, in the termination step, the ribosome and protein are released from the transcript.
   (e) Multiple ribosomes can be on a transcript at once.
   (f) This polyribosomal mRNAs, or polysomes are common and often measured as a proxy for protein production by ribosome footprinting, polysome profiling and single molecule imaging of translation.
3. To produce protein Translation relies a pool of capped mRNAs being available.
   (a) The maintenance of this mRNA pool relies on a broad range of ribsome dependent and independent mechanisms (reviewed in Collart 2020).
   (b) 5' decapping removes the protective 5' cap from a transcript allowing for the 5'-3' exonuclease XNR1 to degrade the transcript (Collart 2020).
   (c) If a ribosome is already present on the transcript, XRN1 trails behind performing cotranslational decay (Pelechano 2015).

(d) This allows for protein to be produced from decaying transcripts.

(e) Another mechanism of decay relies on 3'-5' degradation of transcripts, which does not permit for any more protein to be made.

(f) Decay can also occur following ribosomal collisions.

(g) Collided ribsomes can also initiate endonucleolytic decay pathways such as nonsense mediated decay or no go decay or ribosomal quality control.

(h) The resulting 3' fragments may complete translation.

(i) 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).

and translation (Reuveni 2011, , Raveh 2016, Shaw 2003, Shah 2013) have been modeled separately in the past, but only rarely together (Reuveni 2011, Valleriani 2011).

4. Both mRNA decay and translation have been explored in the literature (Yadav 2021, von der Haar 2012.)

   (a) Some early translation modeling papers used the totally assymetric exclusion process (TASEP) framework to study codon by codon movement of ribosomes (Duc and Song 2018, Shaw 2003)

   (b) While powerful, TASEP is computationally expesive to run.

   (c) The Ribosome Flow Model (RFM) introduced by (Reuveni 2011), make a course grain approximation to translation by splitting an mRNA into regions reducing computational cost at little accuracy (Reuveni 2011).

   (d) Extension on the RFM look at the combined translational behavior of pools of different transcript species (Nanikashvili 2019).

   (e) Other protein synthesis models take a cell wide approach to model transaltion (Shah 2013) providng a systems wide overview of protein production.

   (f) mRNA degradation mechanisms have also been explored, either mechanistically (Cao and Parker 2001, Cao and Parker 2003, Wu 2013, Wu 2016) or using a basic protein produciton model (Zupanic 2016) and are reviewed in (Ashworth 2019).

(g) A combined 5' mRNA decay model has previously been explored briefly by (Valleriani 2010), finding a moderate effect of decay on ribosomal load.

(h) Translation and mRNA degradation have both received ample attention in the literature, however few have explored the interaction between mRNA degradation and translation.

5. Here we introduce a novel coupled ODE model of mRNA polysome classes which integrates mRNA transcription, 5' mRNA degradatiaon and translation.

(a) Our model despite not being fit to data can demonstrate biological behavior using empirically derived parametres.

(b) The structure of the model allows for the exploration of protein production from both capped and uncapped mRNAs undergoing cotranslational decay.

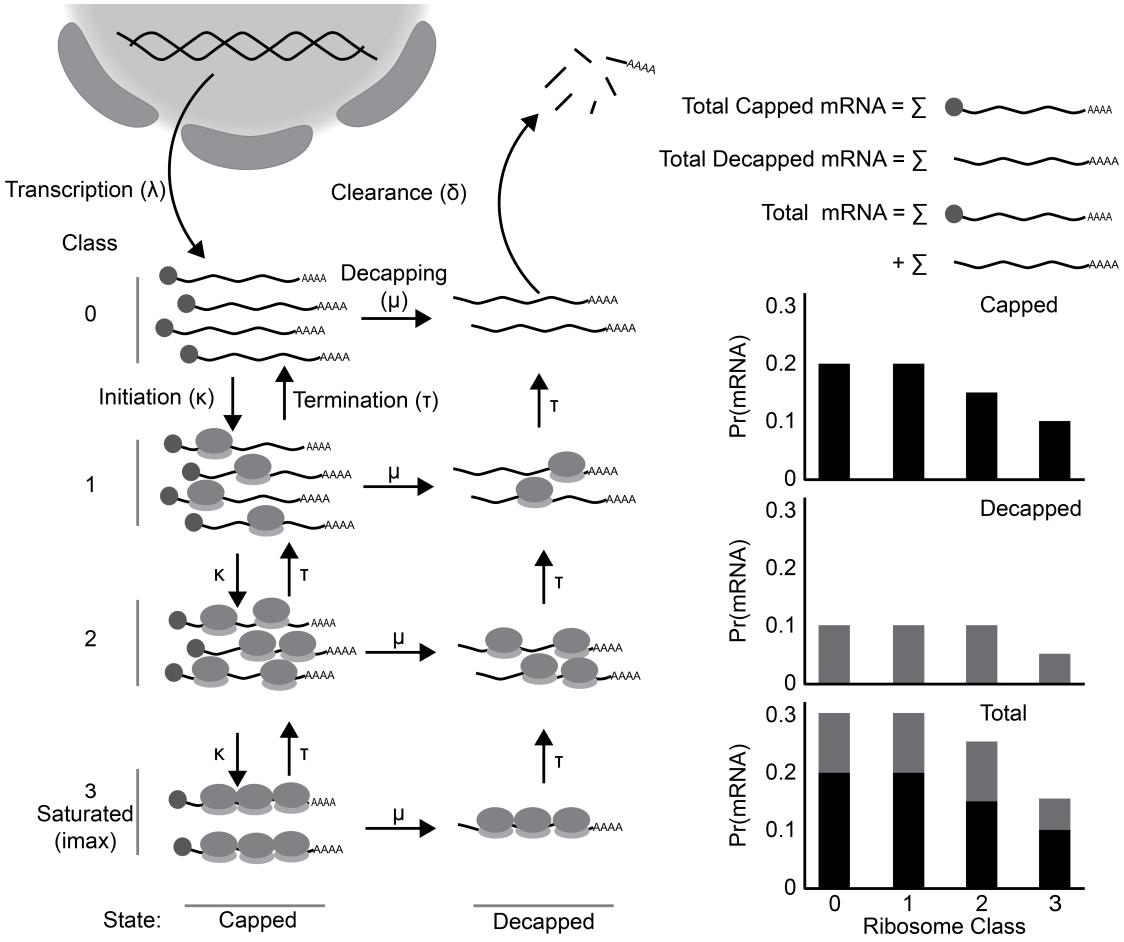
(c) Moreover, we find that genes with a high decapping rate (short half life) will produce just under half their protein from decapped transcripts.

(d) Our model proves useful toward understanding a more complete view of RNA biology and protein production and acts as a stepping stone for future development.

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 polysome 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 role of the 5' cap in mRNA protection and translation initiation. Capped transcripts are translation initiation competent, while decapped transcripts are not. Individual transcripts in the cell are categorized by an integer number of ribosomes (0, 1, 2, ..., $i_{\max}$). The number of ribosomes on a transcript determines that transcripts polysomal class. The model seeks to determine how the population of transcripts of a given gene are distributed between polysome 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, ie. polysome class 0 (m_0). From the m_0 class a transcript can have two fates. The transcript can be decapped, thus marked for degradation at rate μ and move into the decapped class m_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 m_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

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

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, $(1 - i/i_{\max})$ 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 $i_{\max} = n_c/9$ 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,

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

Capped transcripts move through rounds of translation initiation and elongation-termination and distribute along the different polysomal classes. From any polysome class in the capped state m_i the transcript can be decapped at rate μ and move into the decapped state while maintaining the same polysomal class m_i^*. Decapped transcripts can no longer initiate new rounds of translation, but allow for currently loaded ribosomes to complete translation. The behavior of the m_i^* classes represent co-translational decay, a common method of mRNA decay in eukaryotes (Hu 2009, Pelechano 2015, Collart 2020) After all ribosomes on a decapped transcript complete translation, the mRNA is in decapped class 0 and completely degraded at the mRNA clearance rate δ.

The model produces two outputs. First, the total mRNA in either the capped or decapped state and therefore the system (Figure 1B-D). Second, the distribution of the mRNAs in each mRNA in each polysome class. (Figure 1 B-D).

We represent each state by a series of coupled ordinary differential equations (ODEs), one equation for the mRNA population for each polysome class. The functional form of the capped mRNA sub population is:

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

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.	$mRNA$
p_i	Distribution of mRNAs with a ribosome load of i in capped state.	$mRNA$
p_i^*	Distribution 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	
	Ribosome	
κ_0	Translation initiation rate for capped mRNAs with a ribosome load of $i = 0$.	$1/s$
τ	Translation completion rate for one ribosome	$1/s$
μ	Decapping rate.	$1/s$
λ	Production rate of newly produced, ribosome free, and capped mRNA to the m_0 class.	$mRNA/s$
δ	Removal rate of decapped 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:

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

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

2.1.1. Steady state solutions of the capped transcript population

The model solution for the capped state can be represented in the following form,

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

Where $\hat{\vec{m}}$ is a vector of the steady state mRNA abundances in each polysomal class. $\hat{\vec{m}}$ is calculated from by scaling the vector $\vec{p}$ (where $\sum(\vec{p}) = 1$), which represents the steady state distribution of the mRNA across the polysomal classes, by transcript production rate λ and the decapping rate μ. 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 we could not find a 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

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

2.1.2. 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:

$$\begin{aligned}
\hat{m}_0^* &= \frac{\mu}{\delta}\sum_{j=0}^{i_{\max}} m_j \\
\hat{m}_1^* &= \frac{\mu}{\tau}\sum_{j=1}^{i_{\max}} m_j \\
&\vdots \\
\hat{m}_i^* &= \frac{1}{i}\frac{\mu}{\tau}\sum_{j=i}^{i_{\max}} m_j \\
&\vdots \\
\hat{m}_{i_{\max}}^* &= \frac{1}{i_{\max}}\frac{\mu}{\tau}\hat{\vec{m}}_{i_{\max}}
\end{aligned} \tag{7}$$

We can simplify the model by converting the mRNA quantity m_j to the probability p_j by eq. (5). By defining S_j as the cumulative probability from an mRNA found in class i and above as,

$$S_i = \sum_{i}^{i_{\max}} \vec{p_i} \tag{8}$$

Now the solution to eq. (7) becomes,

$$\begin{aligned}
\hat{m}_0^* &= \frac{\lambda}{\delta} S_0 = \frac{\lambda}{\delta} \\
\hat{m}_1^* &= \frac{\lambda}{\tau} S_1 \\
&\vdots \\
\hat{m}_i^* &= \frac{1}{i} \frac{\lambda}{\tau} S_i \\
&\vdots \\
\hat{m}_{i_{\max}}^* &= \frac{1}{i_{\max}} \frac{\lambda}{\tau} S_{i_{\max}}
\end{aligned} \tag{9}$$

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

The total transcript population in the decapped state $\hat{m}_{tot}^*$ does not have a closed form solution. However it can be summarized as follows,

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

This can be further shortened to:

$$m_{tot}^* = \lambda \left(\frac{1}{\delta} + \frac{1}{\tau} \vec{S} \cdot \vec{l} \right) \tag{10}$$

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

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

$$p_0^* = \frac{1}{1 + \frac{\delta}{\tau} \vec{S} \cdot \vec{l}} \tag{11}$$

$$p_i^* = \frac{S_i}{i(\frac{\tau}{\delta} + \vec{S} \cdot \vec{l})} \quad \text{for } i = 1, 2, ..., i_{\max} \tag{12}$$

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

The total mRNA (M) in the system is defined by,

$$M = \lambda \left(\frac{1}{\mu} + \frac{1}{\delta} + \frac{1}{\tau} \vec{S} \cdot \vec{l} \right) \tag{13}$$

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 we calculate the odds,

$$odds_{\hat{m}} = \frac{1}{\mu(\frac{1}{\delta} + \frac{1}{\tau}\vec{S} \cdot \vec{l})} \tag{14}$$

2.3. Calculating mean ribosomal load and protein production

The mean ribosomal load (MRL) for either the capped polysome classes calculated by:

$$E(\hat{\vec{m}}) = \sum_{i=0}^{i_{\max}} j \times p_i \tag{15}$$

and for the decapped polysome classes by:

$$E(\hat{\vec{m}}^*) = \sum_{i=0}^{i_{\max}} j \times p_i^* \tag{16}$$

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,

$$\text{Total Ribosomal Load} = p_{mtot} \times E(\hat{\vec{m}}) + (1 - p_{mtot}) \times E(\hat{\vec{m}}^*) \tag{17}$$

2.4. Numerical solution implementation in R

Code to solve the model was written and is freely available as an R package at (`https://github.com/rurquidi/Ribosome`). To solve the capped subsystem of the model, the `solve.tridiag` function from `limSolve package` (v1.5.6) (Soetaert,K 2009). The decapped solution was obtained by using the capped solutions into eq. (9). Utility functions, plots and statistics were created using `R` (v 3.6) (R core team), and `data.table` (v1.14.0) (Dowle 2021).

2.5. Data Sources

In order to biologically contextualize and illustrate our model's behavior, we use on parameter values derived from the literature. Protein lengths were extracted from the Ensembl (version 109) and Ensembl plants (version 56) respectively (Cunningham 2022, Yates 2022, Kinsella 2011). The range of $i_{\max}$ is determined from the distribution of protein lengths obtained from brewer's yeast (*saccharomyces cerevisiae*) and *Arabidopsis thaliana*. The range of $i_{\max}$ is 48 (36) (mean (SD)) for yeast and 47 (30) for Arabidopsis (Figure 2A and C). The decapping rate between the capped and uncapped system was estimated from the protein half-lives from Presnyak (2015) for yeast (Figure 2B) and Sorenson (2018)

for Arabidopsis (Figure 2D). We estimated 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 $1.3 \times 10^-3 (1.8 \times 10^-3)$ for yeast and $1.7 \times (10^-4 \pm 2 \times 10^-4)$ for Arabidopsis.

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

$$\text{initiation to elongation ratio} = \kappa' = \frac{\kappa}{\tau} \tag{18}$$

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 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. Because transcription rate λ acts as a scaling factor throughout the model and does not affect the distribution of the ribosomes, for simplicity we set $\lambda = 1$.

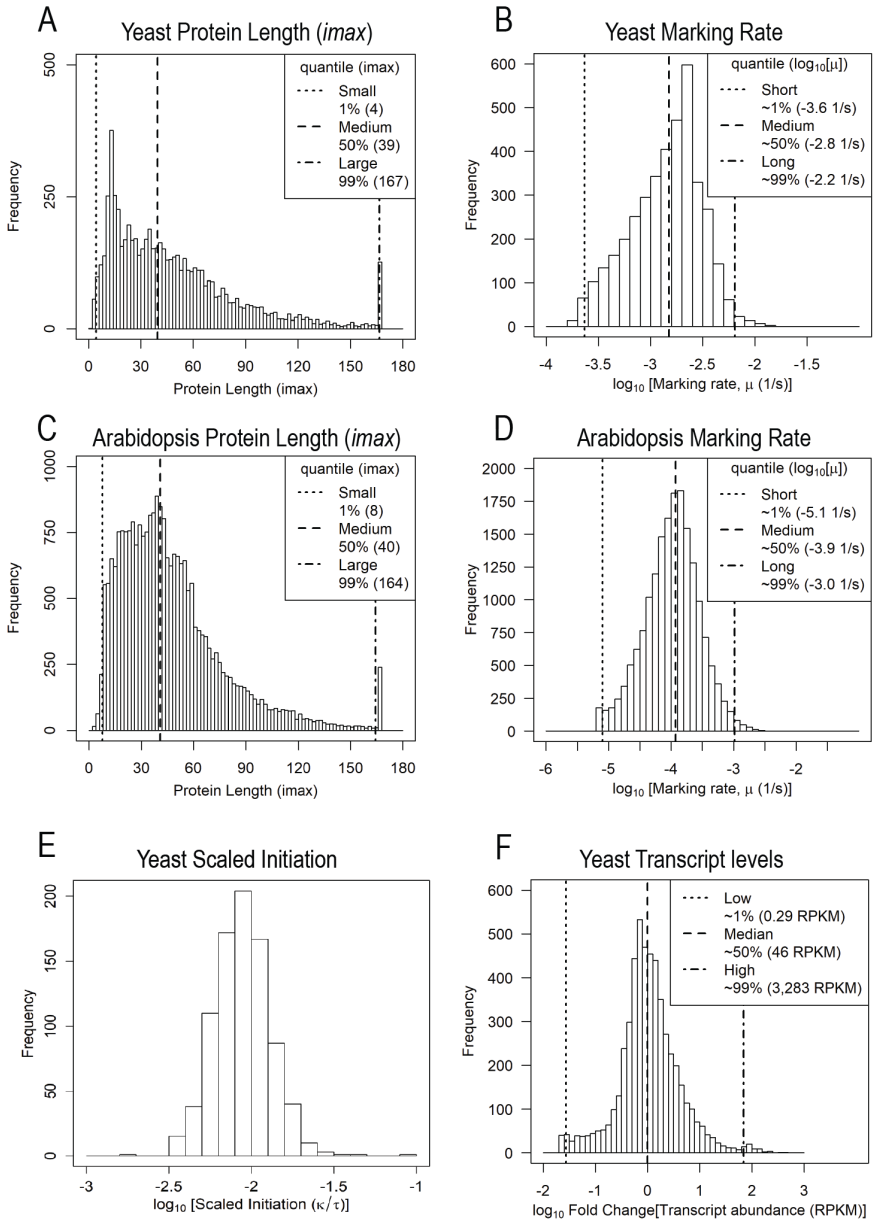
Because the mRNA clearance rate δ only determines the accumulation of transcripts in the m_0^* class, for simplicity, we set $\delta >> \tau$ and thus $m_0^* \sim 0$.

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.

$$MRL_i = \frac{RPF\ RPKM_i}{mRNA\ RPKM_i \times \frac{200}{lengthmRNA_i}} \tag{19}$$

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}).

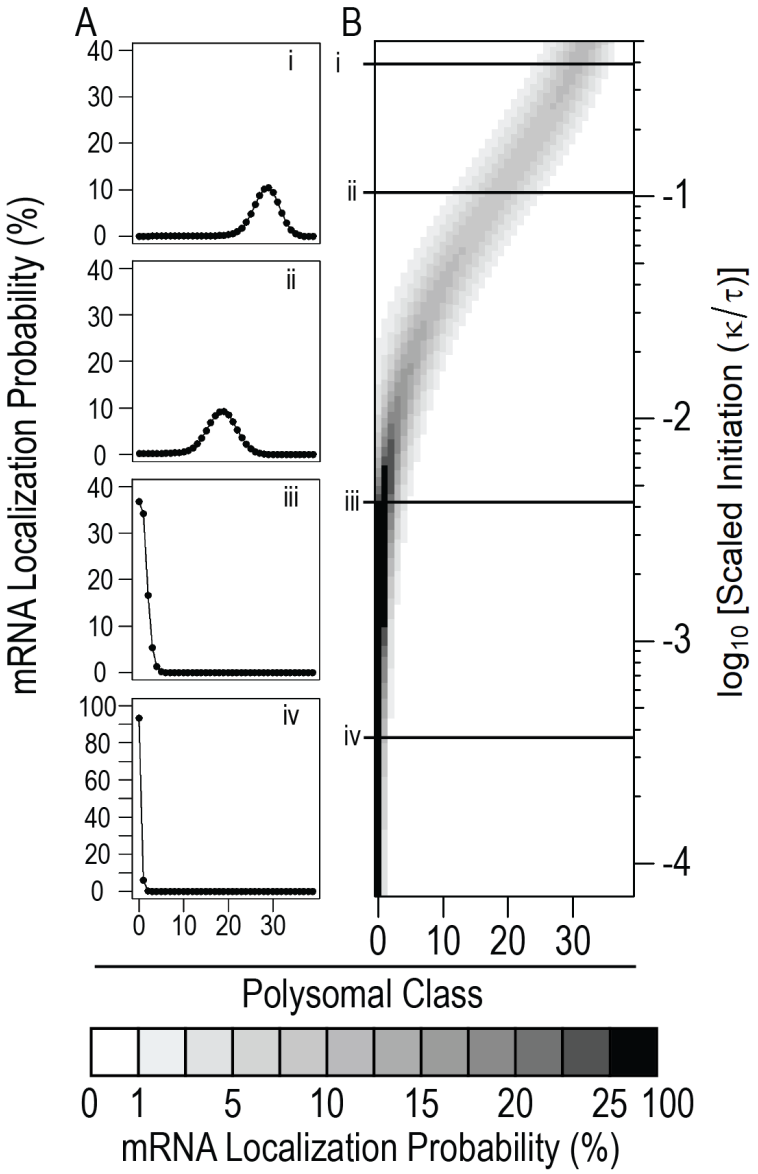
      iv. To summarize the model results across a range of parameters a heatmap where each row is the steady state distribution of mRNA at a particular κ' is shown (Figure 3B).

      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^{th} percentile). Color bar shows probability of finding mRNA in particular polysome 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. 9 can be understood in two parts.

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

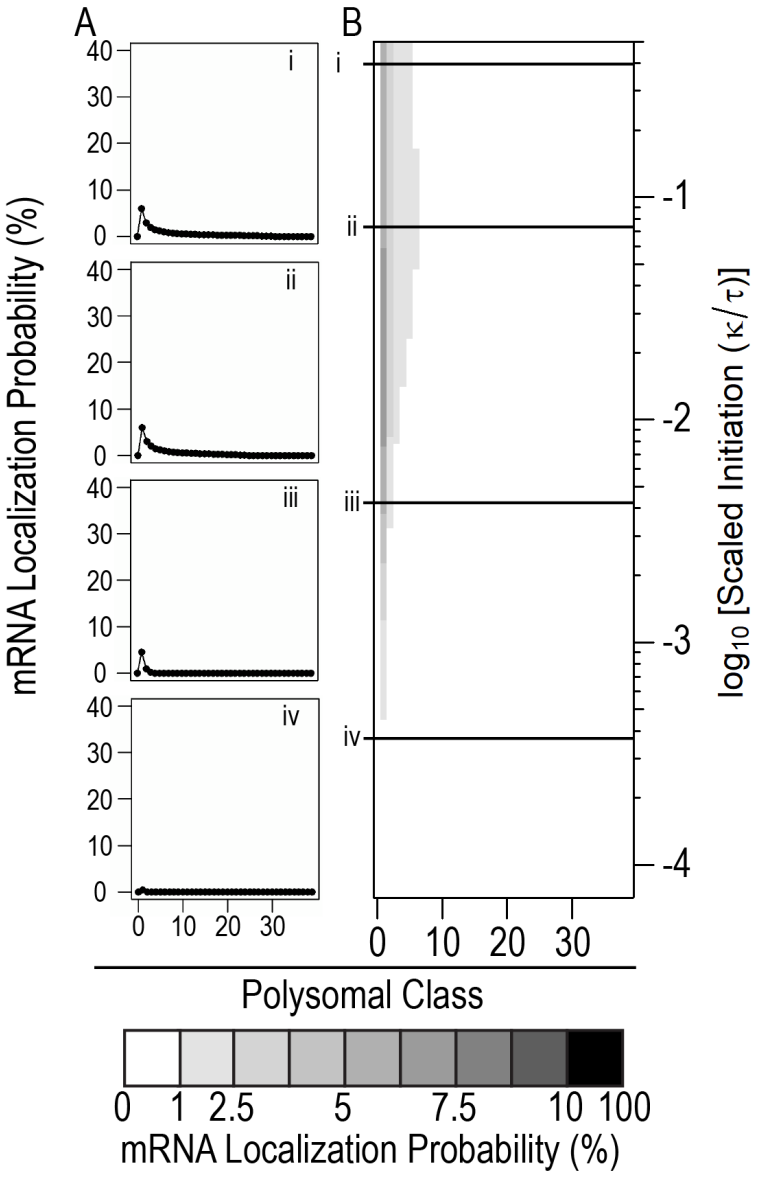
iv. m_0^* is solely determined by the ratio of the mRNA clearance rate $1/\delta$.

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^{th} percentile). Color bar shows probability of finding mRNA in particular polysome class.



(c) Steady states solution for the full model

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

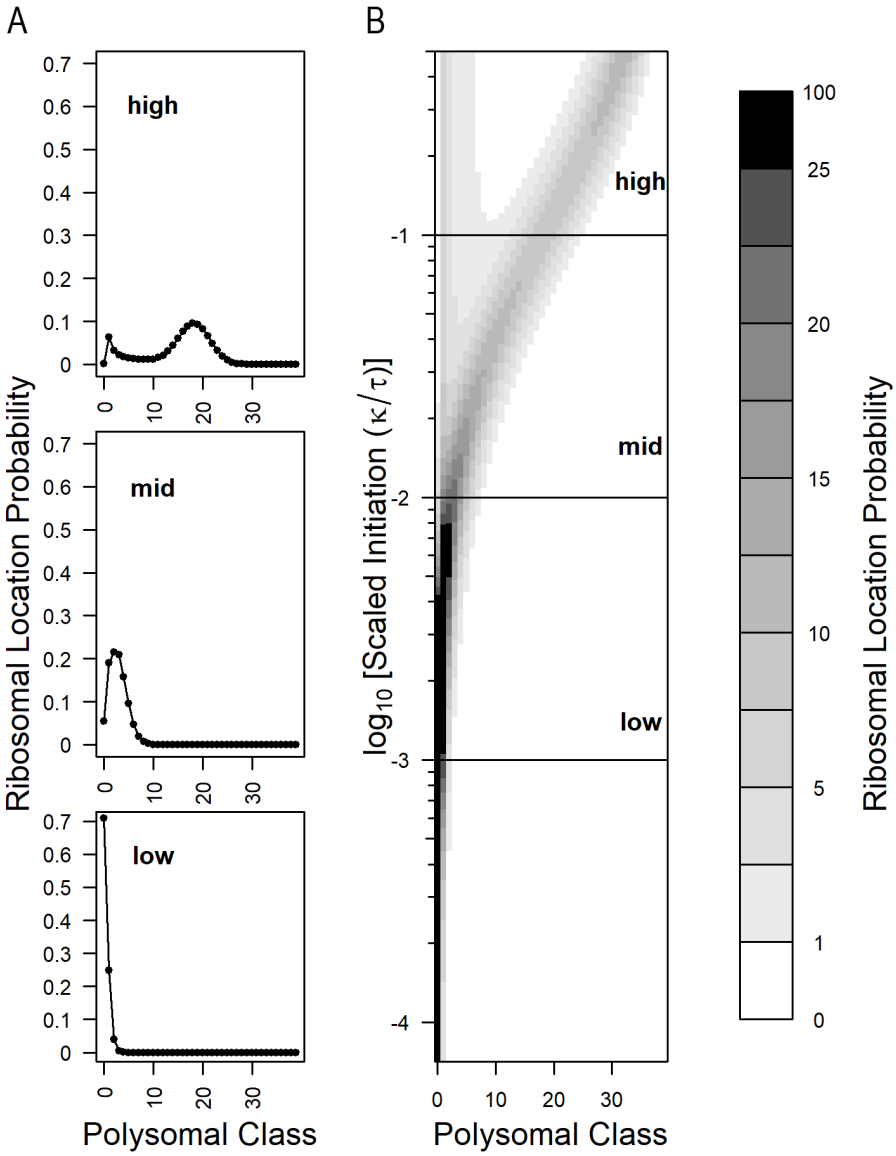
A. Figure 5A shows the mRNA distribution in the full model for three values of κ', for a median length protein with a low decapping rate (2×10^{-4}) and the full range of κ' are shown in the heatmap in Figure 5B.

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 model. 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^{th} percentile). Color bar shows probability of finding mRNA in particular polysome 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 1^{st},

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

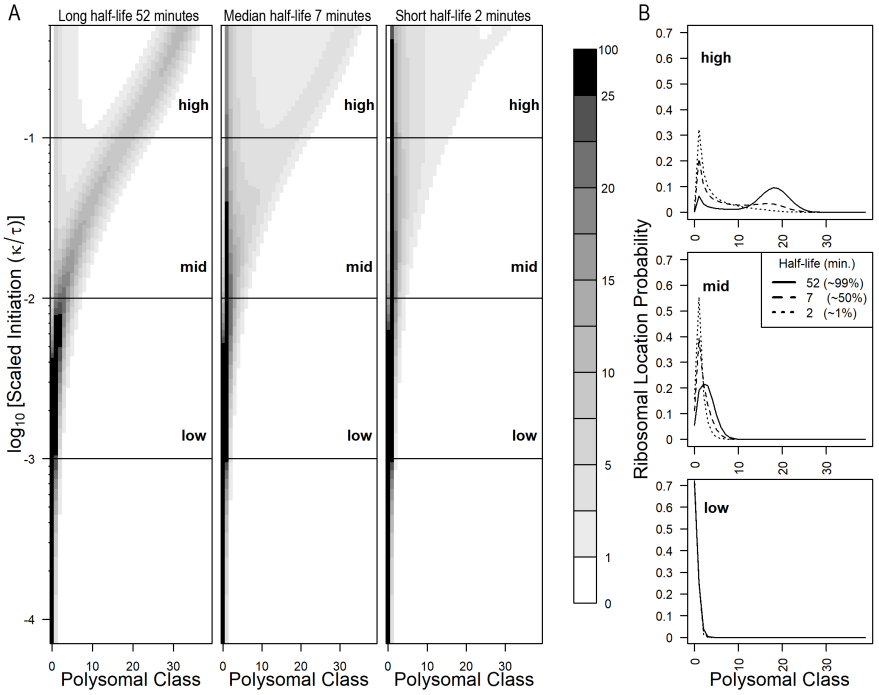
   (a) As μ increases the distribution of mRNAs changes in two ways.

      i. First there is shift to lower polysome 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 model. 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 $i_{\max} = 39$.



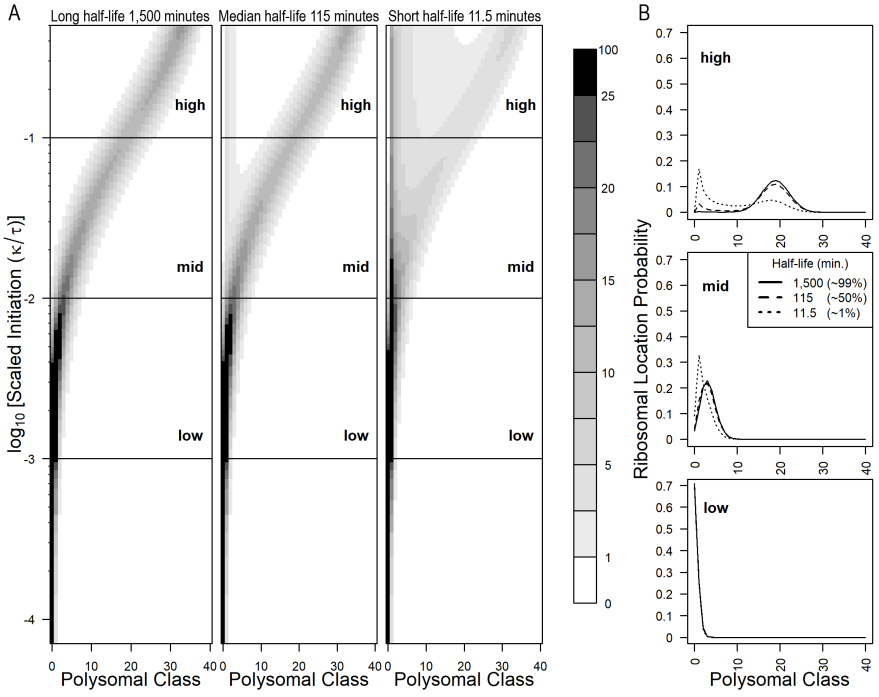
2. Plants and other multicellular eukaryotes tend to have slower translation initiation and elongation rates as well as slower cell division when compared to single celled organisms such as yeast.

   (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 model. Left) low μ (7.7×10^{-6} /s) Center) median μ ($1{\times}10^{-4}$ /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.

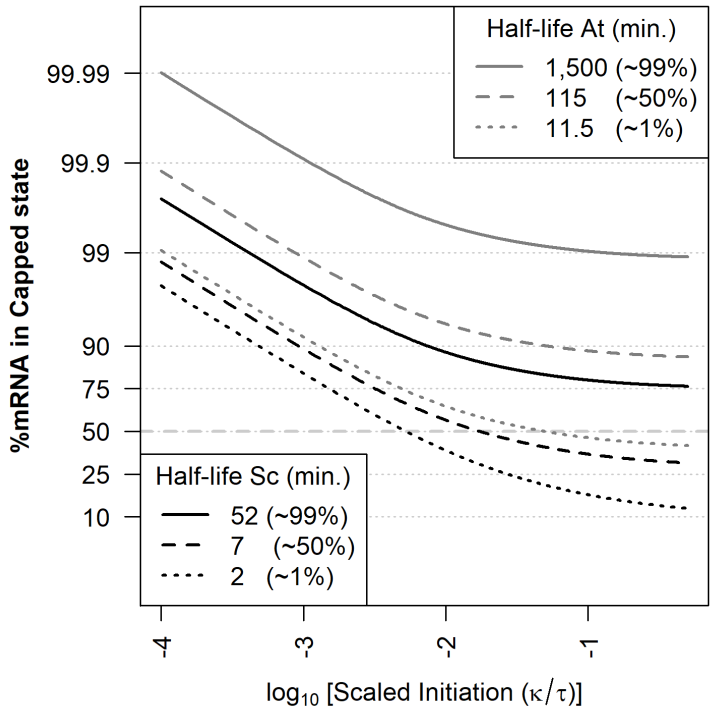
   (a) Using 14, we can determine how much of the mRNA population is in the capped state.

   (b) We produced output across all scaled initiation values κ' and under the 1^{st}, 50^{th} and 99^{th} 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 19.

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

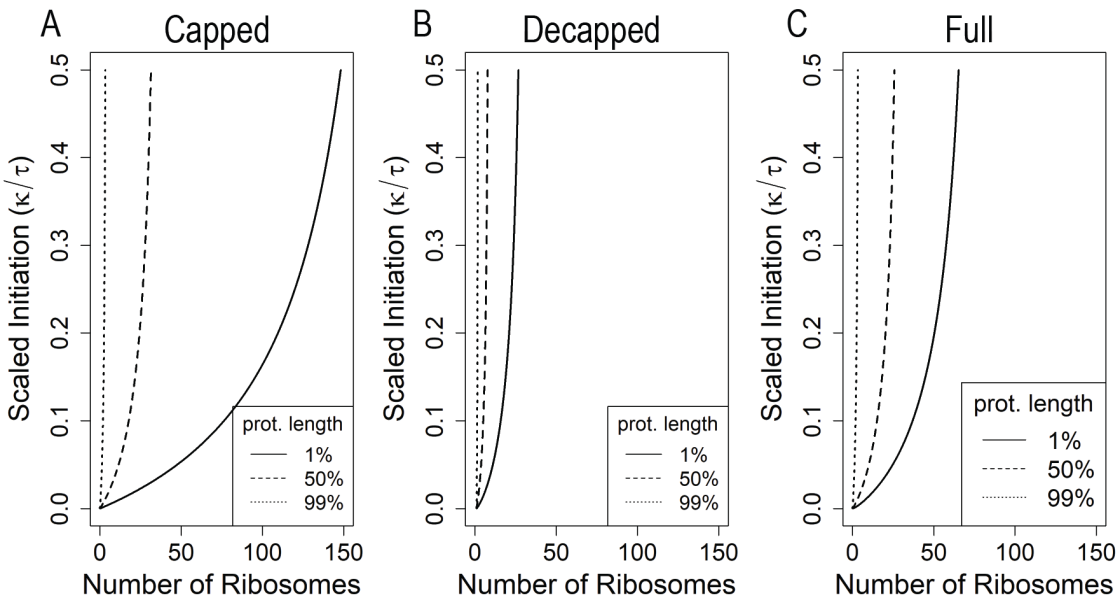
   (c) Therefore ribosomes take longer to clear the mRNA, and thus increase the contribution from the decapped state.

   (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 $1/\tau$ term in eq 14, remembering that $\tau = \tau_c/9 * i_{\max}$.

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 model. Yeast parameters were used $i_{\max}$ = 4 (1^{st} percentile), 39 (50^{th} percentile), 194 (99^{th} 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 model MRL $\times\tau$ and plots normalized to highest protein output are shown in Figure 10.

   R: may now be superfluous

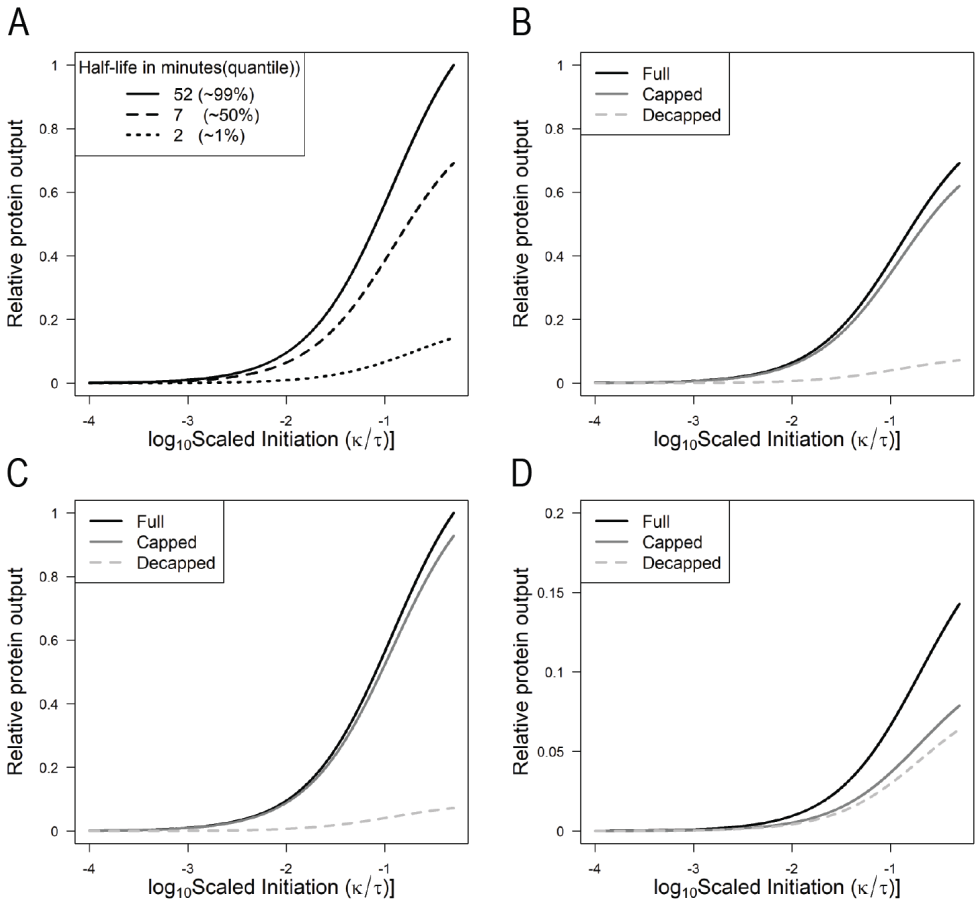
   (a) As the decapping rate μ increases it reduces the capped and uncapped MRL as well as shifting transcript abundance to the decapped state (Figure 10A).

   (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

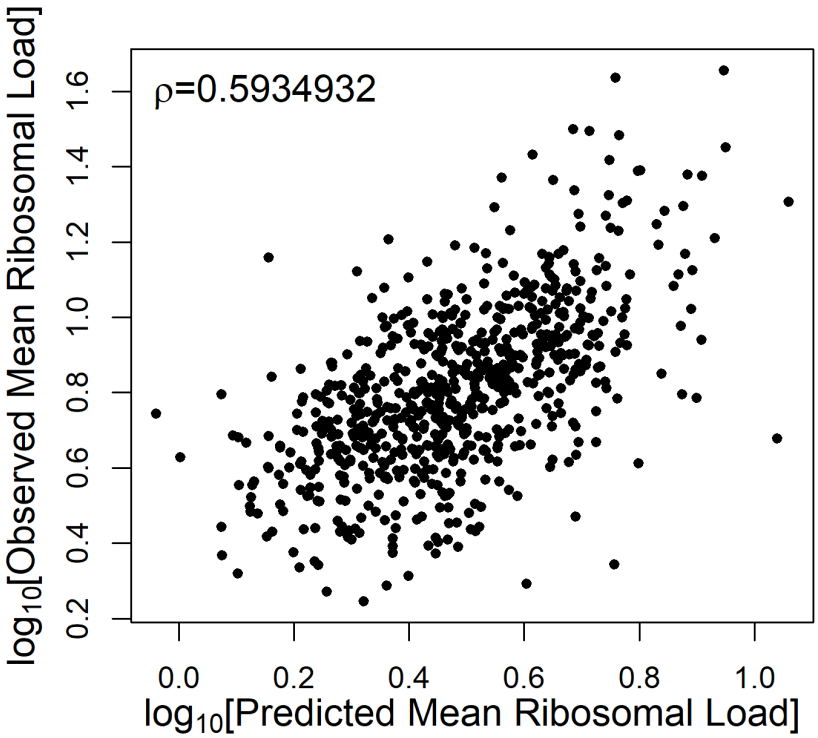
1. Gene specific MRL 19 were calculated for the genes analyzed in Duc and Song 2018 and compared to the empirical MRL calculated from raw data from Weinberg 2016 (Figure 11).

   (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 (10), 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, $\kappa' = \kappa/\tau, \mu, \text{and} \delta$ 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 9). 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 $i_{\max} >> median i_{\max}$), its contributution to the mRNA population's protein production will always be less than 50%.

The full model combines the distributions of the capped and decapped states and is equivalent to mRNA population that is often measured in translational assays. The full model distribution is strictly unimodal when the majority of trasncripts are in the capped state ($\sum \hat{\vec{m}} >> \sum \hat{\vec{m}}^*$), effectively when $\kappa'/\mu \ll 1$ (Figure 6 and 7). The distribution is also unimodal when the $\kappa' << 0.01$, meaning that the MRL is low and near the $i = 0$ bound. In all other cases the distribution is bimodal. The high MRL peak arises from the capped state, while the small MRL peak comes from the decapped states.

The model formalizes the interplay between mRNA decapping μ and initiation elongation ratio κ/τ and its effect on protein production. By multiplying the state specific MRL [19] by τ we can estimate protein production rate. Shifts in mRNA between capped and decapped states as well as changes in MRL control protein production. As expected, increasing μ raises the proportion of decapped transcripts $\hat{\vec{m}}^*$ compared to capped transcripts $\hat{\vec{m}}$. Increasing ratio of initiation to elongation rates κ/τ also results in an increase of decapped mRNAs $\hat{\vec{m}}^*$. 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^*. At high μ this shift in transcripts is enough to partly overcome the fact that the decapped MRL $\leq$ the capped MRL. One final consideration is the assumption that the mRNA clearance rate $\delta >> \lambda$, and therefore $\hat{m}_0^*$ will be negligible. If $\hat{m}_0^*$ is small our current results act as an upper bound of the protein production contribution from the decapped class.

A unique property of our model is that it can differentiate the capped and decapped states individual contributions to protein production. 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 $\sim 120\,\text{sec}$) has almost (but never more than) half of its protein production coming from the decapped mRNA polysome classes (Figure 10). The high protein production from the decapped state suggests that high decapping rate transcripts can produce more protein than would be expected from their capped MRL alone.

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 $(1-i/i_{\max})$, 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 th

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

$$\vec{M}' = \boldsymbol{F}\vec{M} + \vec{B}, \tag{20}$$

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 model (Equation **??**), 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 model 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 **??**.

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{U} & \boldsymbol{0} \\ \\ \boldsymbol{\mu} & \boldsymbol{R} \end{pmatrix}. \tag{21}$$

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, $\mathbf{0} \in \mathbb{R}^{i_{\max}+1 \times i_{\max}+1}$. 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 **??**.

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,

$$\boldsymbol{U} = \begin{pmatrix} -(\kappa_0+\mu_0) & \tau_0\frac{1}{i_{\max}} & & & & & \\ \kappa_0 & \left(1-\frac{1}{i_{\max}}\kappa_0+\mu_0+\tau_0\frac{1}{i_{\max}}\right) & \tau_0\frac{2}{i_{\max}} & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & 1-\frac{(i-1)}{i_{\max}}\kappa_0 & -\left(1-\frac{i}{i_{\max}}\kappa_0+\mu_0+\tau_0\frac{i}{i_{\max}}\right) & \tau_0\frac{i+1}{i_{\max}} & & \\ & & & \ddots & \ddots & \ddots & \\ & & & & \frac{1}{i_{\max}}\kappa_0 & -\left(\mu_0+\tau_0\frac{i_{\max}}{i_{\max}}\right) & \end{pmatrix} \tag{22}$$

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

$$\boldsymbol{R} = \begin{pmatrix} -\delta & \tau_0\frac{\text{`}}{i_{\max}} & & & & & \\ & -\tau_0\frac{1}{i_{\max}} & \tau_0\frac{2}{i_{\max}} & & & & \\ & & \ddots & \ddots & & & \\ & & & -\tau_0\frac{i-1}{i_{\max}} & \tau_0\frac{(i+1)}{i_{\max}} & & \\ & & & & \ddots & \ddots & \\ & & & & & -\tau_0\frac{(i_{\max}-2)}{i_{\max}} & \tau_0\frac{i_{\max}}{i_{\max}} \\ & & & & & & -\tau_0\frac{i_{\max}}{i_{\max}} \end{pmatrix}, \tag{23}$$

$\boldsymbol{R}$ 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

$$\vec{m}' = \boldsymbol{U}\vec{m} + \vec{b}, \tag{24}$$

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 **??**), 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.

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

$$\kappa(i) = \kappa_0 \left(1 - \frac{i}{i_{\max}}\right), \tag{25}$$

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.

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 polysome 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.

$$\frac{dm_0}{dt} = \lambda + \mu m_0 \tag{26}$$

which, at equilibrium results in,

$$m_0 = \lambda/\mu \tag{27}$$

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:

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

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 polysome 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 polysome 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 polysome 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,

$$\vec{m} = \frac{\lambda}{\mu} \vec{p}_m \tag{29}$$

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.

5.2.2. Decapped Subsystem steady state solution

Starting with the decapped subsystem of equations:

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

We get the following solutions at steady state:

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

We can rearrange the solutions and simplify to find,

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

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

$$\frac{\lambda}{\mu} = \sum_{i=0}^{i_{\max}} m_i \tag{30}$$

Therefore,

$$1 = \frac{\mu}{\lambda} \sum_{i=0}^{i_{\max}} m_i \tag{31}$$

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

$$S_j = \frac{\mu}{\lambda} \sum_{i=j}^{i_{\max}} m_i \tag{32}$$

Now the solution becomes,

$$
\begin{aligned}
m_0^* &= \frac{\lambda}{\delta} S_0 = \frac{\lambda}{\delta} \\
m_1^* &= \frac{\lambda}{\tau} S_1 \\
&\vdots \\
m_i^* &= \frac{\lambda}{i\,\tau} S_i \\
&\vdots \\
m_{i_{\max}}^* &= \frac{\lambda}{i_{\max}\,\tau} S_{i_{\max}}
\end{aligned} \tag{33}
$$

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

$$m^*_{tot} = \sum_{i=0}^{i_{\max}} m^*_i = \frac{\lambda}{\delta} + \frac{\lambda}{\tau} S_1 + \ldots + \frac{\lambda}{i\tau} S_i + \ldots + \frac{\lambda}{i_{\max}\tau} S_{i_{\max}} \tag{34}$$

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

$$m^*_{tot} = \lambda(\frac{1}{\delta} + \frac{1}{\tau}\vec{S} \odot \vec{l}) \tag{35}$$

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,

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

5.3. Complete system mRNA population

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

$$M_{tot} = \lambda(\frac{1}{\mu} + \frac{1}{\delta} + \frac{1}{\tau}\vec{S} \odot \vec{l}) \tag{36}$$

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,

$$odds_m = \frac{p_{mtot}}{1 - p_{mtot}} \tag{37}$$

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

It simplifies to,

$$odds_m = \frac{1}{\frac{\mu}{\tau}\vec{S}\odot\vec{l}} \tag{39}$$

$$log_{10}(odds_m) = -log_{10}(\frac{\mu}{\tau}\vec{S}\odot\vec{l}) \tag{40}$$

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 $\tau * 9 * i_{\max} = 2$, and κ=0.04 and $\tau * 9 * i_{\max}$=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 10. 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.