Modeling mRNA Populations

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

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

Abstract

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

---

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

and to verify model function.

alternate abstract: Modeling Ribosomal Loading of mRNA

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

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

Paper Outline

1. Motivation - **(Mike)**

   (a) Why is this process important?

   (b) What will this model enable researchers to do?

   (c) Other modeling efforts?

2. Derivation and Assumptions

   (a) Physical processes captured (Ideally, have a quick discussion of process and inline definitions of variables used to represent process, followed by a total recap in a table) - **(Nate)**

      i. System described as population model: Dichotomy of decapped and capped mRNA. State variables based on an mRNA's ribosome load.

      ii. Process of mRNA production

      iii. Process of Marking mRNA for degradation supposed

      iv. Three processes of : Initiation, translation, and completion

   (b) Definition/Discussion of system boundaries - (**MIKE**)

      i. Physical boundaries as a cell and relation to parameters

      ii. Discussion of perceived upper and lower limits to state variables and parameters

      iii. Temporal boundaries and relation to steady state

   (c) Assumptions: Such as initial assumptions of specific functional forms, i.e. marking rate constant among classes - **(Nate)**

   (d) Justify consideration of system as two subsystems, decapped and capped. - **(Nate)**

3. Model Formulation: Total model presented and then analysis of capped and decapped systems - **(Nate)**

   (a) ODE/Discrete system

      i. Present system of ODEs (Total, capped, and decapped)

      ii. Matrix Representation of ODE model (Total, capped, and decapped)

      iii. Steady state formulations

   (b) PDE/Continuous system

 i. Explain motivation for deriving PDE

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

 iii. Explain derivation using Taylor expansion

 iv. Present PDE for capped class

 v. Present non-dimensionalized system

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

 vii. Motivate and present equation for decapped class at equilibrium

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

4. Results - **(Nate)**

 (a) Present solution strategies/methods

  i. ODE/Discrete system: Matrix inversion technique

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

  iii. Discussion of alternative solution approaches

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

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

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

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

 (b) Alternate functional forms and relaxed assumptions

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

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

1. Introduction

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

1.1. mRNA and Translation

1. Gene expression short overview

   (a) Gene expression is often stated as the central dogma in which genetic information encoded in the DNA is transcribed into mRNA which is subsequently translated into protein.

   (b) Often, a greater amount of attention is focused on explaining gene expression at the transcriptional level and prevailing changes of mRNA transcript levels.

   (c) However, multiple studies across all kingdoms of life have shown that transcript expression level is only moderately predictive of the final protein expression.

   (d) Gene expression at the post transcriptional level is controlled by mRNA transcript stability and degradation, translation and protein maturation/degradation.

   (e) The model presented in this paper encompasses gene expression regulation occuring at the translational and the mature mRNA population level.

2. Biology controlling mRNA stability and translation

   (a) Mature mRNAs in the cytosol are called the free mRNA pool, and are in one of three states.

   (b) They are actively being translated by ribosomes and will continue to initiate new rounds of translation until the transcript is degraded.

   (c) Transcripts are degraded directly from the free mRNA pool.

   (d) Transcripts are protected from degradation by RNA binding protein chaperones or are found in processesing bodies awaiting translation initiation or degradation.

   (e) Degradation of mature mRNAs is controlled by numerous processes depending on whether they are bound to ribosome, in processing bodies or in the free mRNA pool.

   (f) Free mRNAs can be decapped or deadenylated followed by exonuclease digestion.

   (g) Ribosomes can destine transcripts to degradation under multiple conditions.

   (h) The first ribosome to bind to a freshly exported transcript performs the "pioneer round of translation", which is charged with assesing the mRNA's quality.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

12.

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.
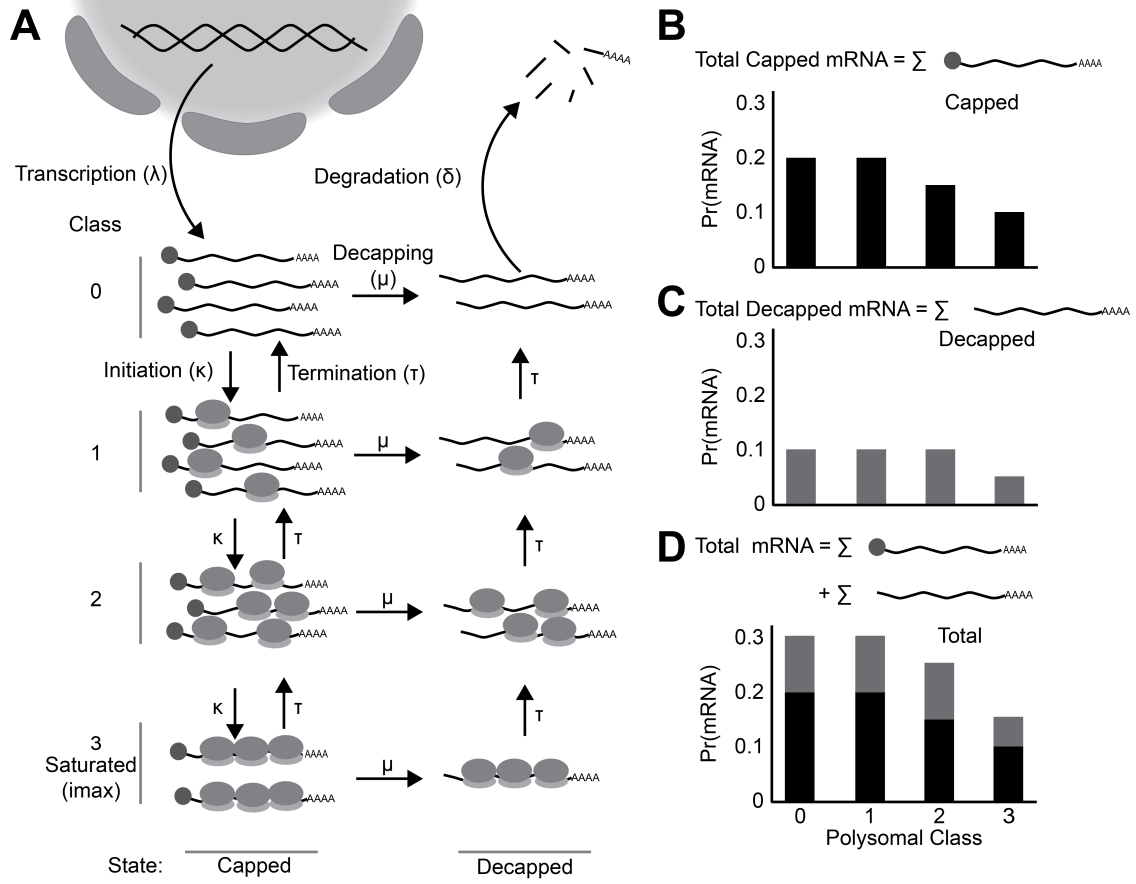
The model captures some of the basic processes governing mRNA populations: transcript production, degradation and the process of translation (Figure 1A). Transcripts can exist in one of two states: capped and decapped which captures the the role of the 5' cap in mRNA protection and translation initiation. 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 at rate $\kappa(1 - i/i_{\max})$ and be loaded onto the transcript and move it into capped class 1. Where i is the current transcript class and $i_{\max}$ is the maximal ribosomal occupancy on the transcript. The first term, κ is the average initiation rate on an empty transcript. For ribosomes initiating on transcripts already harboring ribosomes, this initiation rate is scaled to reflect the probability of a ribosome being present at or near the start codon. This attempts to account for the ribosomal density dependent effects on initiation and is called the density dependent initiation (DDI) model. As our model does not track ribosomal positions, we assume a uniform distribution of ribosomes across a transcript.

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

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

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

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

9

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

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

2.2. Formal Model Definition

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

$$\frac{dm_0}{dt} = \lambda + \frac{\tau_0}{9}\frac{1}{i_{\max}}m_1 - \left(\kappa_0\left(1 - \frac{0}{i_{\max}}\right) + \mu\right)m_0 \tag{3}$$

$$\frac{dm_1}{dt} = \kappa(0)m_0 + \frac{\tau_0}{9}\frac{2}{i_{\max}}m_2 - \left(\frac{\tau_0}{9}\frac{1}{i_{\max}} + \kappa_0\left(1 - \frac{1}{i_{\max}}\right) + \mu\right)m_1$$

$$\vdots$$

$$\frac{dm_i}{dt} = \kappa(i-1)m_{i-1} + \frac{\tau_0}{9}\frac{i+1}{i_{\max}}m_{i+1} - \left(\frac{\tau_0}{9}\frac{i}{i_{\max}} + \kappa_0\left(1 - \frac{i-1}{i_{\max}}\right) + \mu\right)m_i$$

$$\vdots$$

$$\frac{dm_{i_{\max}}}{dt} = \kappa_0\left(1 - \frac{i_{\max}-1}{i_{\max}}\right)m_{i_{\max}-1} - \left(\frac{\tau_0}{9}\frac{i_{\max}}{i_{\max}} + \mu\right)m_{i_{\max}}$$

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

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

Similarly, the functional form of the decapped mRNA sub population is:

$$\frac{dm_0^*}{dt} = \mu m_0 + \frac{\tau_0}{9}\frac{1}{i_{\max}}m_1^* - \delta m_0^* \tag{4}$$

$$\frac{dm_1^*}{dt} = \mu m_1 + \frac{\tau_0}{9}\frac{2}{i_{\max}}m_2^* - \tau(1)m_1^*$$

$$\vdots$$

$$\frac{dm_i^*}{dt} = \mu m_i + \frac{\tau_0}{9}\frac{i+1}{i_{\max}}m_{i+1}^* - \tau(i)m_i^*$$

$$\vdots$$

$$\frac{dm_{i_{\max}}^*}{dt} = \mu m_{i_{\max}}^* - \frac{\tau_0}{9}\frac{i_{\max}}{i_{\max}}m_{i_{\max}}$$

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

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:

$$m_0^* = \frac{\mu}{\delta}\sum_{j=0}^{i_{\max}} m_j$$

$$m_1^* = \frac{\mu}{\tau}\sum_{j=1}^{i_{\max}} m_j$$

$$\vdots$$

$$m_i^* = \frac{\mu}{i\,\tau}\sum_{j=i}^{i_{\max}} m_j$$

$$\vdots$$

$$m_{imax}^* = \frac{\mu}{i_{\max}\,\tau}\sum_{j=imax}^{i_{\max}} m_j$$

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 $i = i_{\max}$.

$$S_j = \sum_{i=j}^{i_{\max}} \vec{p_i} \tag{6}$$

Now the solution becomes,

$$m_0^* = \frac{\lambda}{\delta} S_0 = \frac{\lambda}{\delta} \tag{7}$$

$$m_1^* = \frac{\lambda}{\tau} S_1$$

$$\vdots$$

$$m_i^* = \frac{\lambda}{i\,\tau} S_i$$

$$\vdots$$

$$m_{imax}^* = \frac{\lambda}{i_{\max}\,\tau} S_{i_{\max}}$$

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,

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

This can be further shortened to:

$$m_{tot}^* = \lambda \left(\frac{1}{\delta} + \frac{1}{\tau} \vec{S} \cdot \vec{l} \right) \tag{8}$$

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

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,

$$p_0^* = \frac{1}{1 + \frac{\delta}{\tau} \vec{S} \cdot \vec{l}} \tag{9}$$

$$p_j^* = \frac{S_j}{j\left(\frac{\tau}{\delta} + \vec{S} \cdot \vec{l}\right)} \quad \text{for } j = 1, 2, ..., i, ..., i_{\max} \tag{10}$$

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,

$$M_{tot} = \frac{\lambda}{\mu} + \lambda \left(\frac{1}{\delta} + \frac{1}{\tau} \vec{S} \cdot \vec{l} \right) \tag{11}$$

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 you calculate the odds,

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

2.6. Calculating expected ribosomal load and protein production

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

$$E(ribosome) = \sum_{i=0}^{i_{\max}} j \times p_i \tag{13}$$

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,

$$\text{Total Ribosomal Load} = p_{mtot} \times E(ribosome)_{mtot} + (1 - p_{mtot}) \times E(ribosome)_{mtot^*} \tag{14}$$

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 7. 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 ± 36 for yeast and 47 ± 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 marking rate between the capped and uncapped system was approximated from the protein half-lives from Presnyak 2015 for yeast (Figure 2B) and Sorenson 2018 for Arabidopsis (Figure 2D). 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 This simplifies the model behavior to one generalized parameter with a unique response (Figure 2E). The scaled initiation rate ranges from $0.1s^{-1}$ to $0.001s^{-1}$.

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

Finally, the degradation 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^*.

3. Results

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

The capped solution splits the two funtions of the marking rate μ; Its effect on transcript number and its effect on transcript distribution. And allows for their separate analysis. The mRNA population is defined by the ratio of the transcription rate λ to μ.

The probability of finding a transcript in each ribosomal class for a particular parameter set is dependent on the initiation rate κ, elongation rate τ_0 and μ. In figure 3A the mRNA distribution in the capped state is presented for four different scaled initiation rates (as shown in Figure 3B) for a median length protein with a long (52 minute) half life. To summarize the model results across a range

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.
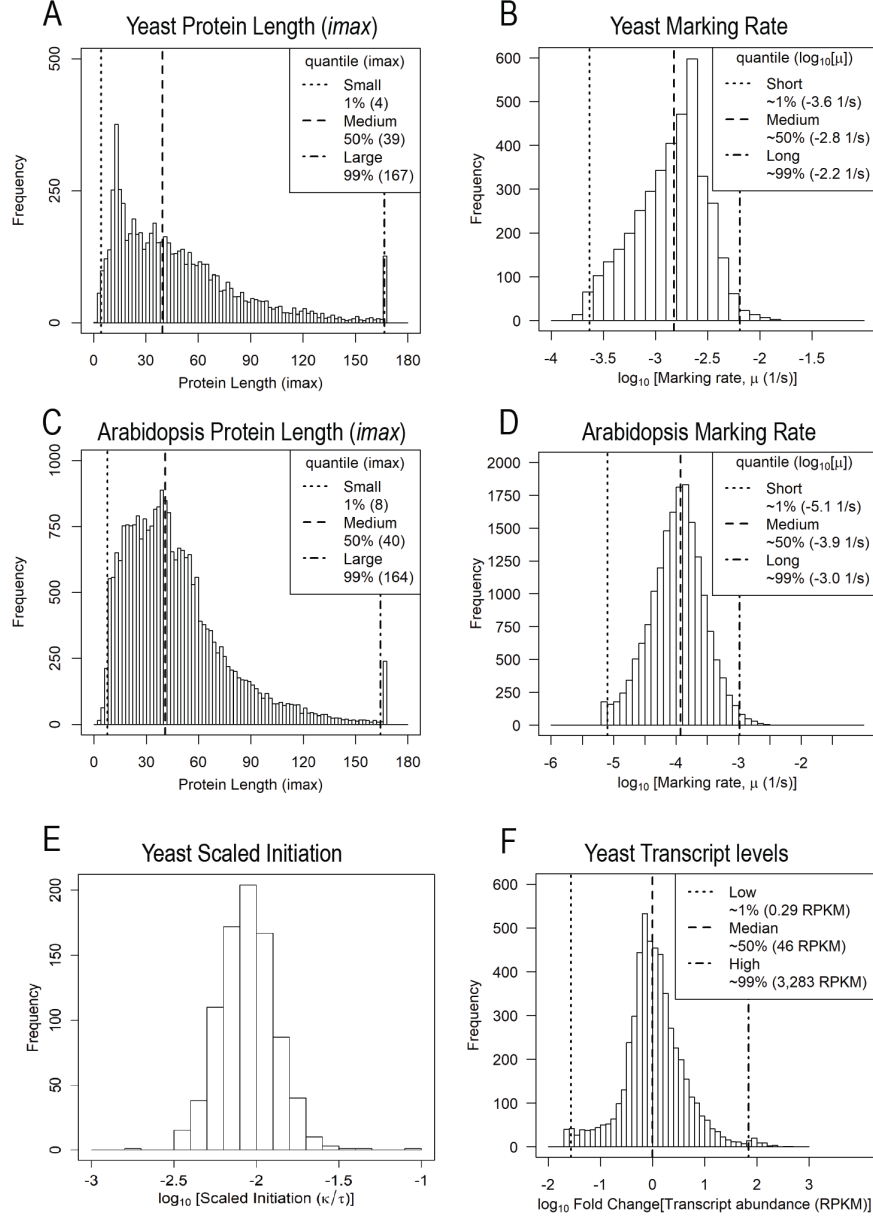
of parameters a heatmap where each row is the distribution of mRNA at a particular scaled initiation rate is shown. As the scaled initiation rate increases the density moves to the right and spreads out in the capped system. The distribution in the capped system is bounded at class 0 and class $i_{\max}$ and roughly symmetrical away from the boundaries (Figure 3B).

The decapped system is centered around the lowest classes (Figure 4). This is due to the distribution in 9 having the following arrangement $S_0 = 1$ and $S_0 \geq S_1 \geq ... \geq S_i \geq ... \geq S_{imax}$ dependant on the distribution of $\vec{m}$ of the capped state. Exploring the result we find a few properties of our system. Transcription rate (λ) again serves only to scale the entire system. The first decapped class's population m_0^* is only dependent on the degradation rate (δ). The whole system incorporates the transcript distributions from both capped and decapped systems. This is apparent in the bimodal peaks at higher scaled initiation values, with a peak at lower ribosomal load representing the decapped system and a high ribosomal load from the capped system (Figure 5).

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

On a physical level, ribosomes initiating on a transcript depend on having sufficient space around the start codon. Transcripts at with higher ribosomal loads therefore have a higher probability of having an obstructed start codon. In order to explore the effect of ribosomal load on initiation we created two versions of the model. The first is dependent on the ribosomal load of a transcript, the density dependent initiation (DDI) model. The DII model is is independent on the ribosomal load of a transcript, the density independent model (DII) presented in figure 5 A-B, with individual probability profiles presented in 5A and the summary heatmap in 5B. The DDI model is presented in Figure 5C-D. Individual probability profiles are generated at low, mid and high scaled initiation rates as determined from Duc and Song 2018. Note that in the DII model the system saturates just above the high scaled initiation rate, while the DDI model doesn't. Additionally, model profiles are very similar at low to mid scaled initiation values for both models. Figures 5 B and D suggest that the density dependent effects start appearing between mid and high scaled initiation rates. Polysome profiling experiments usually ony resolve 8-10 ribosomal peaks, with the majority of the signal arising from polysomes 2-5. This generally agrees with single molecule imaging of nascent peptides where ribosomal densities are on the range of 0.5% - 30% (Morisaki 2016, Wang 2016, Wu 2016, Yan 2016). While both the DII and DDI models show similar behavior and low density at low to mid scaled initiation values, empirical evidence and physical reality would indicate that a transcript is unlikely to ever reach saturation. This supports the DDI model. From here on out all results will be solely based on the DDI model.

### 3.3. Higher marking rates reduce capped state ribosomal loads

The interplay between the translational machinery, mRNA degradation machinery and mRNA properties such as codon optimality, secondary structure or modifications all have been reported to

Figure 3: mRNA distribution in Capped state. A) Individual distribution profiles for four scaled initiation values B) Heatmap of model output across a range of scaled initiation values. Lines represent slice represented in A). Results produced with i_{max} of 39 and a long half life of 52 minutes (99th percentile). Color bar shows probability of finding mRNA in particular ribosomal class.
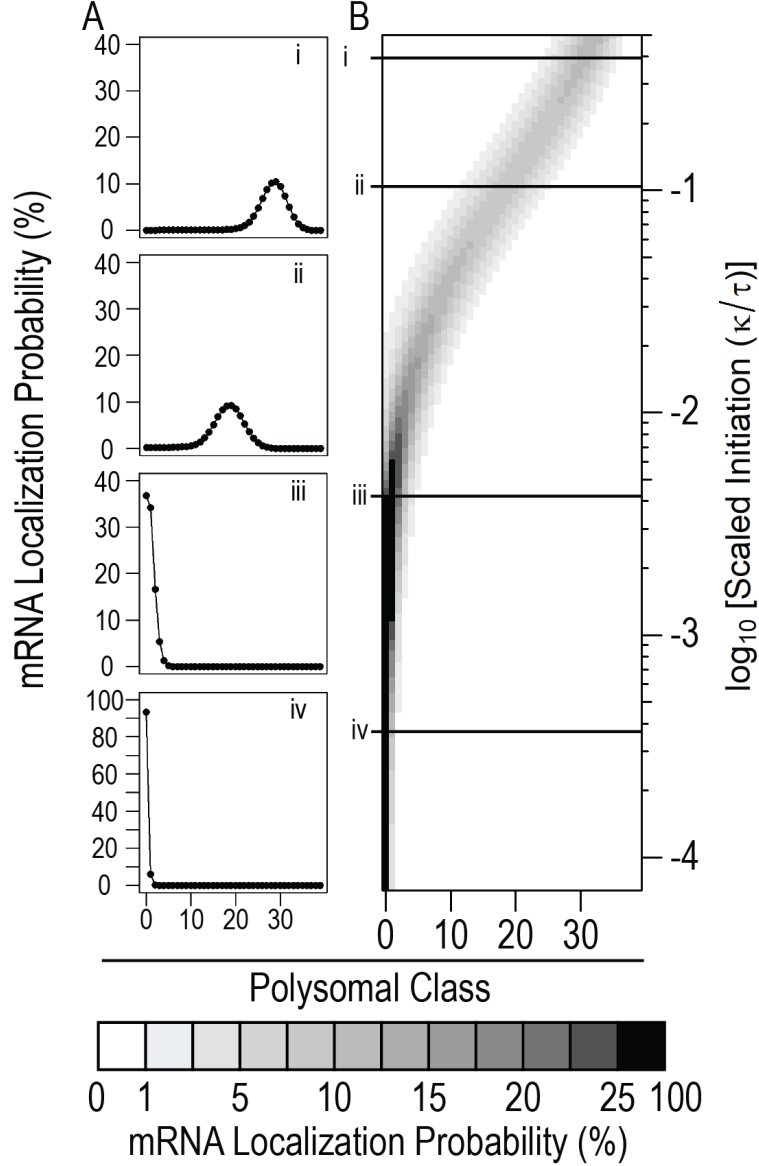
play a role in mRNA stability (Wu 2019, Medina-Munoz 2021, Bae and Collier 2022). To explore the role of mRNA stability on mRNA populations we varied the marking rate from the 1st percentile, median and 99th percentile values as determined from the half-life values in Presnyak 2015. We find that as half-life decreases the distribution of mRNAs change in two ways. First there is shift to lower ribosomal classes in the capped state (Figure 6). This is likely due to the mRNAs leaving the capped state at a higher rate and driving the equilibrium towards lower ribosomal loads. Secondly, as half-life

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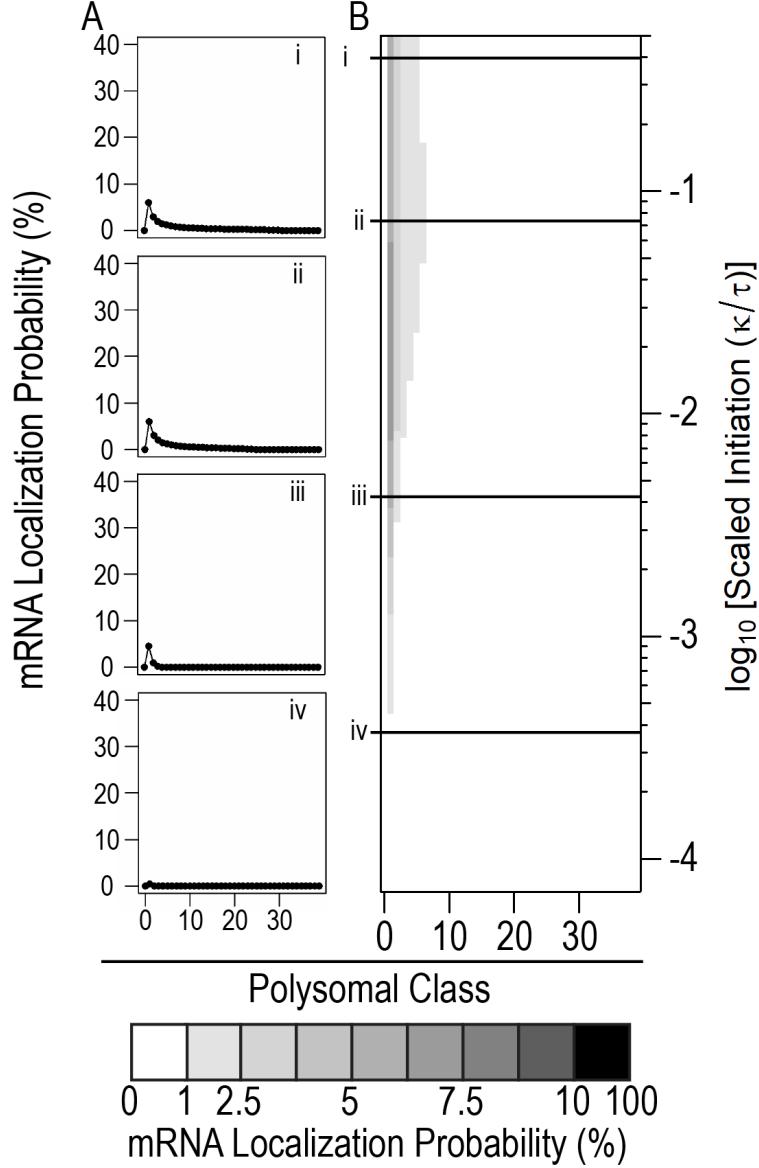
decreases, a larger proportion of the mRNA is found in the decapped state. This is further explored later.

Multicellular eukaryotes, such as plants, face a different set of environmental challenges and tend to have slower translation initiation and elongation rates as well as slower cell division when compared to single celled organisms. This is highlighted by the current gold standard study of mRNA half-lives in the model organism *Arabidopsis thaliana*, where the half-lives measured are one two two orders

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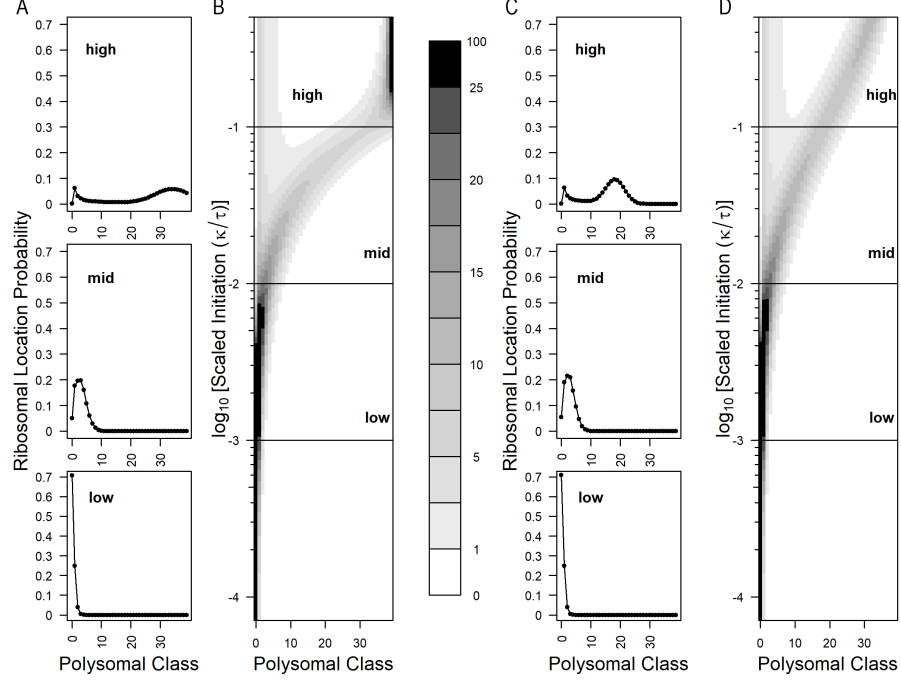

of magnitude longer than those in yeast. To explore this we ran the model using the same scaled initiation rate as in yeast, the median Arabidopsis i_{max} of 41, and Arabidopsis half-lives (long half life (1,500 minutes), median half life (115 minutes), and short half life (11.5 minutes)). As expected, the longer half-lives have higher ribosomal loads and are mostly in the capped state.

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

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

20

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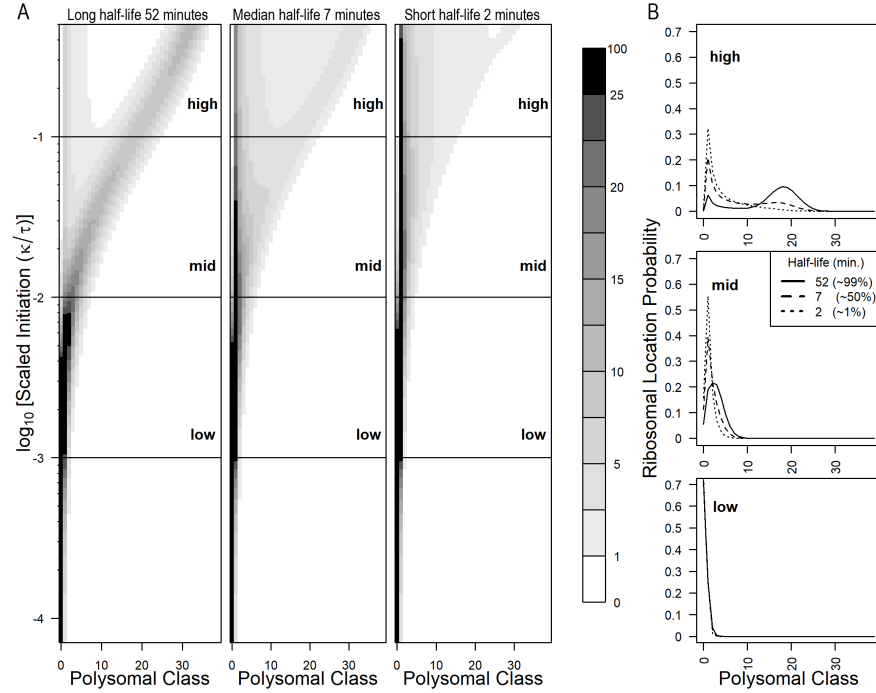

3.5. Marking rate and ribosomal load determine mRNA distribution between states

As shown in previous results, shorter half-lives (faster marking) leads to lower ribosomal load and a shift towards the decapped state. To explore this shift we can use the results in equation 5, which splits the mRNA population in the capped state and the distribution of reads within and equation 8 the total transcript population to derive the log odds in equation **??**. Using equation **??**, we can see under which parameter regimes mRNA is more abundant in the capped state. We produced output across all scaled initiation values and under the 1%, 50% and 99% percentiles for marking rates in both yeast and Arabidopsis (Figure 9). We note two patterns. First as the scaled initiation rate increases the amount of mRNA in the decapped class increases. Secondly, shorter mRNA half-life bias transcripts towards the decapped class as previously seen in Figures 6 and 7.

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

Protein production is a function of the rate at which ribosomes leave the transcript and the underlying total number of transcripts. Naively, the average protein production rate (PPR) for the capped state could be estimated by multiplying the total number of transcript by the initiation rate. However, this ignores the DDI and marking effects in the system and the actual PPR will be lower. Contributions

21

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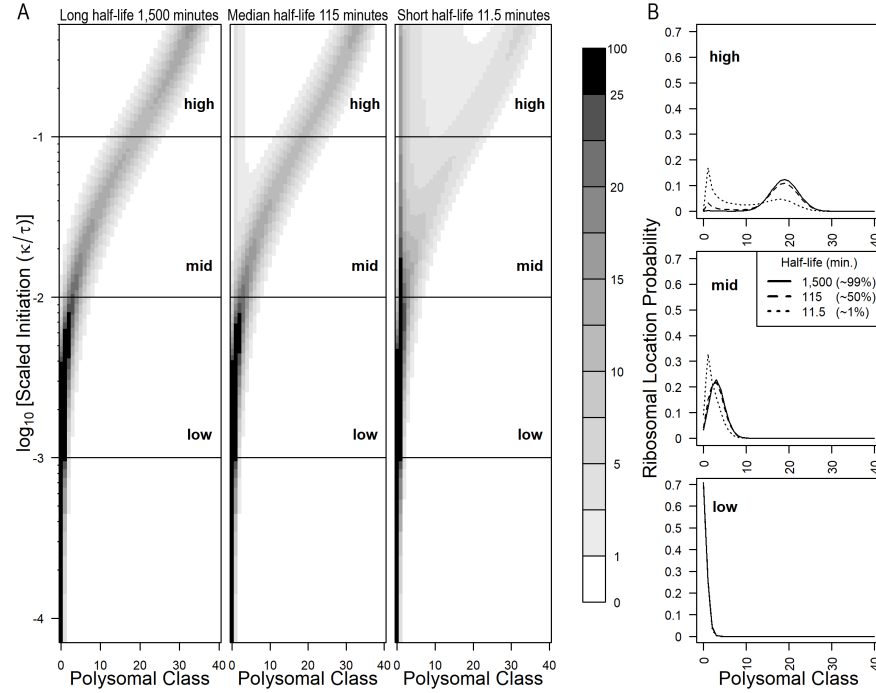

from the decapped state are harder to estimate from the parameters as the distribution is dependent on the capped state's distribution.

Another way of comparing relative protein output (RPO) is to calculate the mean ribosomal load (MRL) for each state using and multiply is by the total number of transcripts in their respective state. The full systems protein production can be found with . In our model the MRL of the capped system will always be equal to or greater than the decapped MRL (Figure 10 A). As the scaled initiation rate increases and the MRL of the capped class increases, the distributional properties of the uncapped class mean that it will always be proportionally less to the MRL of the capped class.

The RPO of a state is dependent on both the MRL of each state and the total transcript population of each state (Figure 10B). As the scaled initiation rate increases RPO will increase. The marking rate μ plays an important role in controlling both the underlying mRNA population and the distribution of transcripts in the capped state. Therefore, and increase in marking rate reduces RPO through two mechanisms. Each of the three cases in Figure 10 B, has been broken down into the RPO contributions from the capped and decapped states (Figure 10 C-E). A surprising finding from out model is that under certain biologically relavent parameter combinations, almost half of all protein production can

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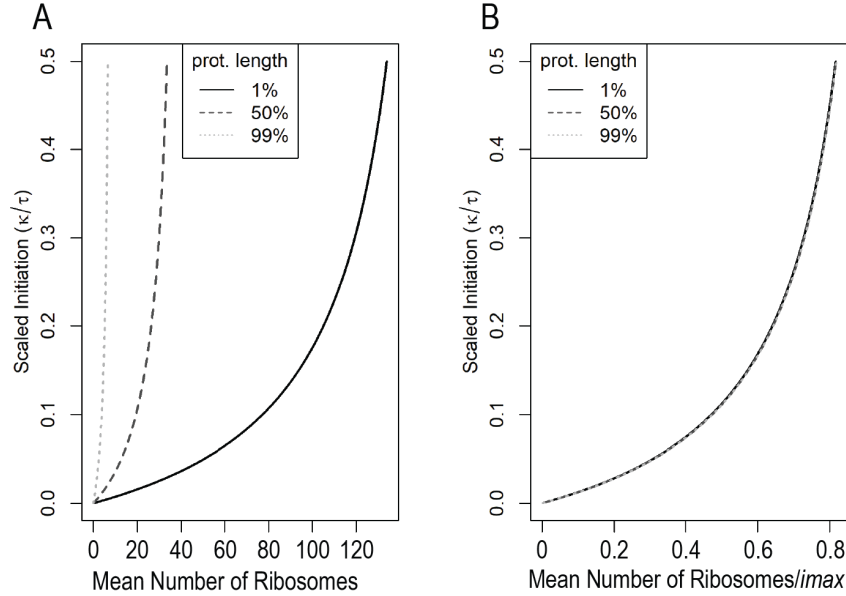

arise from the decapped state. This can be seen inf Figure 10 F, wherea substantial portion of all protein arises from the decapped state. The reason behind this is despite the the relative MRL of the decapped state being lower than the capped state as scaled initiation rises, the amount of mRNA in the decapped state rises faster.

4. Discussion

1. In this work we present a model which tracks mRNA populations and their association with ribosomes across two states, a translationally competent capped state and incompetent decapped state.

   (a) As the process of translation is entirely dependent on its mRNA substrate to produce protein, understanding the underlying fluctuation in mRNA is crucial.

   (b) We demonstrate that the process of marking an mRNA for degradation through decapping alters the total number of mRNA molecules in a system, their distribution between the two states as well as lowering the mean number of ribosomes on a transcript.

23

Figure 9: Log odds of finding mRNA in the capped state for a range of marking rates in yeast and Arabidopsis.
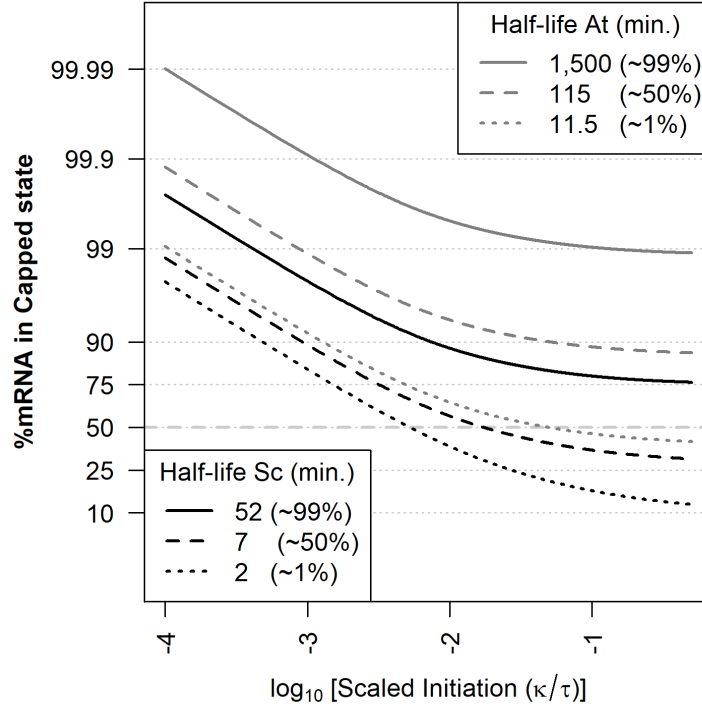
(c) A surprising outcome of this process is that, under certain biological conditions, a substantial amount of protein produced may arise from terminating ribosomes in the decapped state.

1. The model recapitulates empirical measurements of translation.

    (a) Despite the model itself not being fit to data, through use of empirically obtained parameter values it still displays ribosomal loading patterns that would be expected in polysome profiles.

    (b) In polysome profiles all transcripts in a sample are separated based on the number of ribosomes translating on them.

    (c) The majority of transcripts are found with 1-10 ribosomes and strongly bias towards lower ribosomal loads (Weinberg 2016, Lokdarshi 2020).

    (d) Throughout the scaled initiation range we observe that, for a protein of average length and transcripts with a long half-life, the transcript is below a MRL of 10 for the majority of the range.

    (e) Moreover, Duc and Song 2018, only studied the most highly translated genes (850 out of

Figure 10: Estimated average protein production in yeast. A) Ratio of the expected ribomal load of Decapped over Capped states for a protein of i_{max} 40 and low marking rate for distribution of the capped state that result in MRL of 0-40. B) Protein protein production across different marking rates. Total protein production is normalized to the maximal protein production across all parameters. C-F) Contribution of capped and decapped states to total protein production. C) Low marking D) Medium marking E) High marking F) zoom in of high marking result.
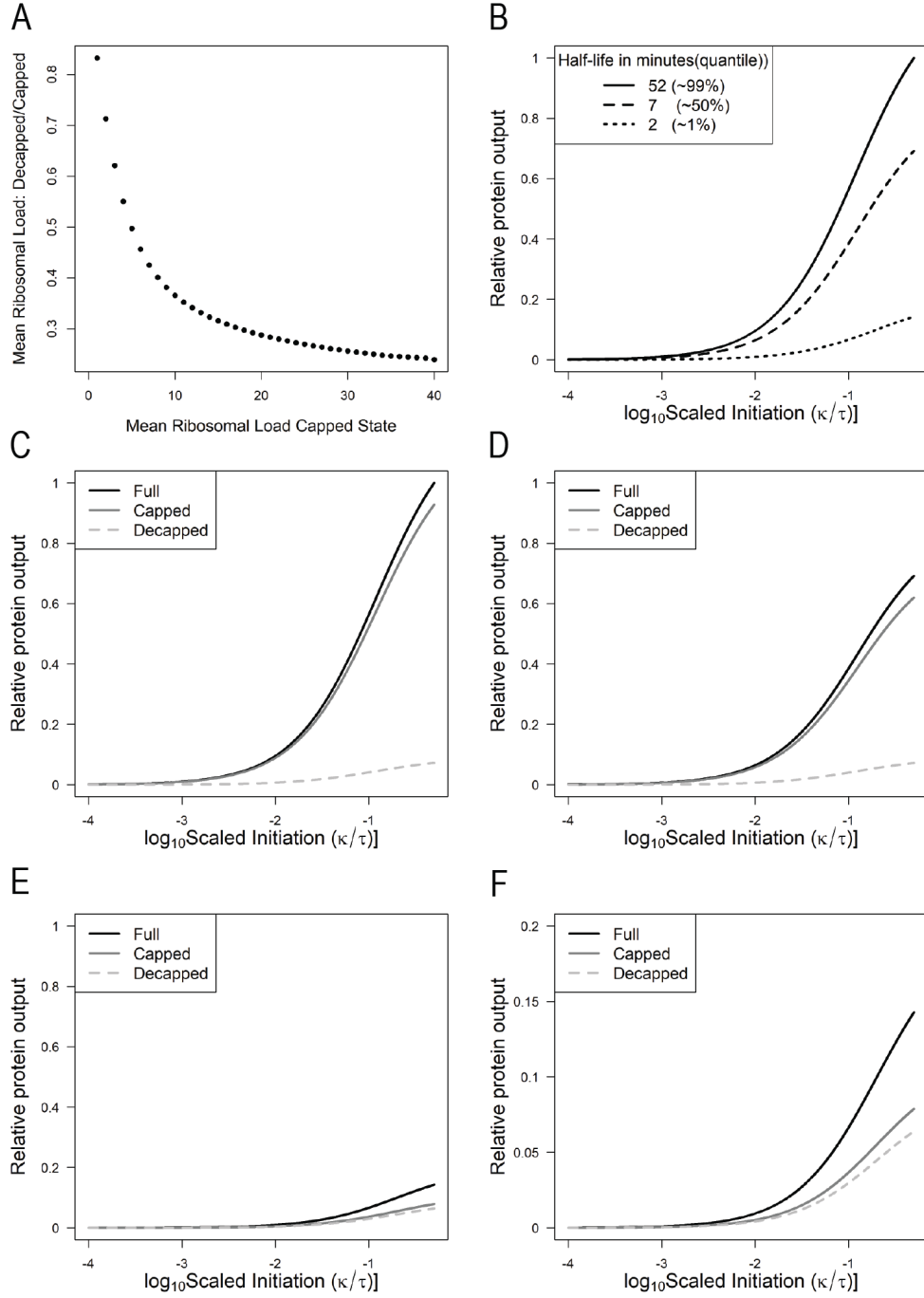
5,300 expressed genes). This means that a majority of translated genes fall below the scaled initiation range used in this study.

(f) This is further corroborated with single molecule imaging analyses. Rescaling ribosome abundances from each study to an $i_{\max}$ of 40 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).

(g) Wu 2016, also designed constructs containing different lengths of protein sequences and found that ribosome load scales linearly with protein length. Agreeing with our results in Figure 8 and implying the measurement was performed at true steady state.

(h) Despite this general agreement, the model still represents an upper bound of translation as many other forms of translational control are not included.

1. Our model integrates the effect of mRNA decay into translation.

   (a) mRNA decay not only regulates transcript abundance, but also reduces mean ribosomal load.

   (b) This reduction can be interpreted as a inhibiting the system from reaching the unobstructed steady state. This has previously been suggested but not demonstrated by (Reuveni 2011) and explored by (Valleriani 2011).

   (c) This suggests that if left unobstructed each transcript species would reach a characteristic MRL during a set period of time.

   (d) The 5' mRNA decay pathway presented in our model most closely resembles that of co-translational decay, and in Arabidopisis 68% of mRNAs are degraded this way (Sorenson 2018).

   (e) However, other mechanisms exist with different outcomes for translation. 3' decay would result in no ribosome terminating and abruptly remove a subset of transcripts and overall protein production from the pool. Endonucleolytic decay due to NGD or NMD would potentially allow ribosomes downstream to terminate but not those upstream (Urquidi-Camacho 2020, Merchante 2017).

   (f) Extension of the model to include these mechanisms is yet to be explored. However, it is safe to speculate that they would both result in a further reduction of MRL.

   (g) Our model also used two conservative set of marking rate estimates (Presnyak 2015 and Sorenson 2018).

(h) Currently there is a debate whether mRNA stability is regulated primarily through the protective effects of ribosomal association or through the suboptimal codons causing ribosomal stalling and the subsequect ribosome associated decay pathways (Chan 2018 elife).

(i) Our model currently cannot distinguish between the protective effects of translation or

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

(k) However, biology suggests a more complex behavior.

(l) The protective effects of translation could increase per ribosome, but eventually at high loads could trigger ribosome associated decay pathways through ribosomal collisions. This leaves much to be explored further

1. In addition to the effects on the distribution of transcripts across polysomal classes, marking rate along with the scaled initiation rate determine the distribution of mRNA across the two states.

   (a) Mature mRNA populations and their degradation has been modeled before, but in absence of translation and it's quality control effects (Cao and Parker 2001, Cao and Parker 2003, Wu 2013, Wu 2016, Zupanic 2016, reviewed in Ashworth 2019).

   (b) Our model shows that even decapped populations of mRNAs can produce protein and can be the most abundant species of mRNA in the population.

   (c) Through 5'P mRNA sequencing Pelechano 2015 can track degradation intermediates. They find that degradation products in the CDS follow a 3 nucleotide periodic pattern and compose 12.4 of all reads recovered. This indicates that a significant portion of the bulk mRNA population is undergoing cotranslational decay.

   (d) Figure 9 shows that a significant amount of mRNA can be found in the decapped state. Most transcripts will have a scaled initiation rate below 0.01 and have most of the mRNA in the capped state. However, our model result is on a per gene basis. To estimate a the global proportion of decapped to capped mRNA will require a larger sample of marking, transcription and scaled initiation rates.

1. Co-translational decay may allow for decapped transcripts to provide substantial protein production.

(a) A surprising result from our model is that genes with short half-lives can have almost half of their protein produced in the decapped state.

(b) This suggests that short lived transcripts can produce more protein than expected despite having a half-life comparable to the time it takes an average length protein to get translated.

2. We also highlight that all examples in the model have been run under the same transcription rate λ. The underlying amount of transcript will scale protein output globally, which translation and mRNA degradation will subsequently tune.

(a) We have not correlated protein production to transcription as in (de Sousa Abreu 2013, Schwanhauser 2013, Edfors 2016, Brion 2020) due to not modeling protein degradation.

(b) Properly setting the transcription rate is important for determining final protein production, but not for understanding the translational dynamics.

3. In our model the remaining transcripts in class m_0^* are either only the 3' end of co-translationally degraded transcripts or full transcripts from class m_0.

(a) The role of the degradation rate is to account for the remaining transcripts.

(b) Mathematically and biologically it is essential for completing the lifecyle of an mRNA, but difficult to measure.

1. The model suggest that regulation mechanisms for translation are concerned with reducing MRL.

(a) This is first seen by comparing DII, the simplest, most unobstructed model to the DDI model.

(b) As scaled initiation rates increase, density dependence reduces MRL. We do not include bottlenecking due to translational folding, codon optimality or pausing in our model therefor density dependent effects might occur earlier in a transcript.

(c) With in the design of our model this would mean a transcript would behave as if its total length were ¡$i_{\max}$. Given a shorter effective $i_{\max}$ for translation.

(d) This particular aspect could be further modeled by splitting each transcript species into two or three regions defined biologically by pausing sites. This would resemble a nested ribosome flow model withing our model structure.

(e) As described above, mRNA degradation further reduces the MRL as can mRNA secondary structure.

(f) Other methods such as the canonical global repression of translation by eIF2α phoshorylation and upstream open reading frames are all inhibitory methods (Dever 2022, Lokdarshi 2020).

(g) Finally, modeling of the all the species of transcript (Nanikashvili 2019, Raveh 2016) and taking to account ribsome availability have been done previously (Shah 2013), but without considering mRNA degradation.

(h) All these biological phenomena provide ample room for extension of our base model.

5. Appendix

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

General model equations of the density independent initiation model

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

Specifically, new mRNA enter the 0th capped class $m_0(t)$ at a rate.. Ribosomal bind mRNAs in the ith capped class at a rate $\kappa(i)$, increasing the mRNA's ribsome load to the $i + 1^{th}$ class. By definition, $\kappa(i_{\max}) = 0$, i.e. mRNAs with a ribosome load of $i_{\max}$ cannot accommodate any addtional mRNAs. In the density independent model (DII), we assume that the current ribosomal load has no effect on the ability of another ribosome to bind to the transcript. An average ribosomal footprint covers 9 codons (27 nucleotides). Therefore for a protein of 270 amino acids in length, the maximal ribsomal load, $i_{\max} = 10$. Capped mRNAs with ribosome load i are decapped at a rate of $\mu(i)$. We assume that capped mRNAs are decapped for degradation rate independent of their ribosome load, i.e. $\mu(i) = \mu_0$. Accordingly, the ribosome load of decapped mRNAs remains unchanged, but they are transitioned from the capped class $m_i(t)$ to the decapped class $m_i^*(t)$. Ribosome movement along an

mRNA is assumed to occur independent of whether or not its capped or decapped for degradation. Thus, ribosomes complete translation of both decapped and capped mRNAs with ribosome load i at rate $\tau(i)$, decreasing the mmRNA's ribosome load to the $i - 1^{th}$ class. Where $\tau(i) = i \cdot \tau(1)$ and . This is because we not modeling the explicit movement of ribosomes along an mRNA, we assume that at steady state probability of finding a ribosome at any given codon position within the coding sequence follows a uniform distribution. Thus, the chance that a ribosome on a transcript of class i will complete translation increases as ribosome load increases. Since mRNA's with a ribosome load of 0 have no ribosomes which can complete translation, by definition $\tau(0) = 0$. It is important to note that $\tau(1)$ is not the same as the average elongation rate. $\tau(1) = average\ elongation\ rate/(9 \cdot i_{\max})$. That is, the average elongation rate in aa/s is rescaled to the average rate of total elongation and termination through a transcript in units of $1/s$.

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,

$$\vec{M}' = \boldsymbol{F}\vec{M} + \vec{B}, \tag{15}$$

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

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

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{U} & \boldsymbol{0} \\ \boldsymbol{\mu} & \boldsymbol{R} \end{pmatrix}. \tag{16}$$

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

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

(17)

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 ith row.

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

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

(18)

$\boldsymbol{R}$ 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

$$\vec{m}' = \boldsymbol{U}\vec{m} + \vec{b}, \tag{19}$$

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,

$$\kappa(i) = \kappa_0 \left(1 - \frac{i}{i_{\max}}\right), \tag{20}$$

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:

$$\frac{dm_0}{dt} = \lambda + \tau(1)m_1 - \left(\kappa_0 \left(1 - \frac{0}{i_{\max}}\right) + \mu(0)\right)m_0$$

$$\frac{dm_1}{dt} = \kappa(0)m_0 + \tau(2)m_2 - \left(\tau(1) + \kappa_0 \left(1 - \frac{1}{i_{\max}}\right) + \mu(1)\right)m_1$$

$$\vdots$$

$$\frac{dm_i}{dt} = \kappa(i-1)m_{i-1} + \tau(i+1)m_{i+1} - \left(\tau(i) + \kappa_0 \left(1 - \frac{i-1}{i_{\max}}\right) + \mu(i)\right)m_i$$

$$\vdots$$

$$\frac{dm_{i_{\max}}}{dt} = \kappa_0 \left(1 - \frac{i_{\max}-1}{i_{\max}}\right)m_{i_{\max}-1} - \left(\tau(i_{\max}) + \mu(i_{\max})\right)m_{i_{\max}}$$

and the decapped subsystem is unchanged.

5.2.1. Capped state steady state solution

1.

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.

$$\frac{dm_0}{dt} = \lambda + \mu m_0 \tag{21}$$

which, at equilibrium results in,

$$m_0 = \lambda/\mu \tag{22}$$

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

$$\frac{\lambda}{\mu} = \sum_{i=0}^{i_{\max}} m_i, \tag{23}$$

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 $[\lambda\ 0\ 0\ ...\ 0]$. 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,

$$\vec{m} = \frac{\lambda}{\mu} \vec{p}_m \tag{24}$$

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:

$$\frac{dm_0^*}{dt} = \mu(0)m_0 + \tau(1)m_1^* - \delta m_0^*$$

$$\frac{dm_1^*}{dt} = \mu(1)m_1 + \tau(2)m_2^* - \tau(1)m_1^*$$

$$\vdots$$

$$\frac{dm_i^*}{dt} = \mu(i)m_i + \tau(i+1)m_{i+1}^* - \tau(i)m_i^*$$

$$\vdots$$

$$\frac{dm_{i_{\max}}^*}{dt} = \mu(i_{\max})m_{i_{\max}}^* - \tau(i_{\max})m_{i_{\max}}$$

We get the following solutions at steady state:

$$m_0^* = \frac{\mu\, m_0 + \tau(1)m_1^*}{\delta}$$

$$m_1^* = \frac{\mu\, m_1 + \tau(2)m_2^*}{\tau(1)}$$

$$\vdots$$

$$m_i^* = \frac{\mu\, m_i + \tau(2)m_{i+1}^*}{\tau(i)}$$

$$\vdots$$

$$m_{imax}^* = \frac{\mu\, m_{imax}}{\tau(imax)}$$

We can rearrange the solutions and simplify to find,

$$m_0^* = \frac{\mu}{\delta} \sum_{j=0}^{i_{max}} m_j$$

$$m_1^* = \frac{\mu}{\tau} \sum_{j=1}^{i_{max}} m_j$$

$$\vdots$$

$$m_i^* = \frac{\mu}{i\,\tau} \sum_{j=i}^{i_{max}} m_j$$

$$\vdots$$

$$m_{imax}^* = \frac{\mu}{i_{max}\,\tau} \sum_{j=imax}^{i_{max}} m_j$$

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

$$\frac{\lambda}{\mu} = \sum_{i=0}^{i_{max}} m_i \tag{25}$$

Therefore,

$$1 = \frac{\mu}{\lambda} \sum_{i=0}^{i_{max}} m_i \tag{26}$$

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

$$S_j = \frac{\mu}{\lambda} \sum_{i=j}^{i_{max}} m_i \tag{27}$$

Now the solution becomes,

$$m_0^* = \frac{\lambda}{\delta} S_0 = \frac{\lambda}{\delta} \tag{28}$$

$$m_1^* = \frac{\lambda}{\tau} S_1$$

$$\vdots$$

$$m_i^* = \frac{\lambda}{i\,\tau} S_i$$

$$\vdots$$

$$m_{imax}^* = \frac{\lambda}{i_{max}\,\tau} S_{i_{max}}$$

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

$$m_{tot}^* = \sum_{i=0}^{i_{\max}} m_i^* = \frac{\lambda}{\delta} + \frac{\lambda}{\tau}S_1 + \ldots + \frac{\lambda}{i\tau}S_i + \ldots + \frac{\lambda}{i_{\max}\tau}S_{i_{\max}} \tag{29}$$

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

$$m_{tot}^* = \lambda(\frac{1}{\delta} + \frac{1}{\tau}\vec{S} \odot \vec{l}) \tag{30}$$

Where $\vec{S}$ is a vector of all the cummulative sums and $\vec{l}$ is a vector of $1, 1/2, \ldots, 1/i, \ldots, 1/i_{\max}$. The S_i have the following arrangement $S_0 = 1$ and $S_0 \geq S_1 \geq \ldots \geq S_i \geq \ldots \geq S_{imax}$. This depends on the distribution of $\vec{m}$ of the capped state. Exploring the result we find a few properties of our system. Transcription rate (λ) again serves only to scale the entire system. The first decapped class's population m_0^* is only dependent on the degradation rate (δ). The total mRNA in the decapped state can wildly vary according to the value of degradation. In this work we shall set delta to be large and focus on the effects of the marking rate and elongation/termination rate. This result will be explored further in the results.

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

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

5.3. Complete system mRNA population

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

$$M_{tot} = \lambda(\frac{1}{\mu} + \frac{1}{\delta} + \frac{1}{\tau}\vec{S} \odot \vec{l}) \tag{31}$$

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,

$$odds_m = \frac{p_{mtot}}{1 - p_{mtot}} \tag{32}$$

$$odds_m = \frac{\frac{1}{(1+\frac{\mu}{\tau}\vec{S}\odot\vec{l})}}{1 - \frac{1}{(1+\frac{\mu}{\tau}\vec{S}\odot\vec{l})}} \tag{33}$$

It simplifies to,

$$odds_m = \frac{1}{\frac{\mu}{\tau}\vec{S}\odot\vec{l}} \tag{34}$$

$$log_{10}(odds_m) = -log_{10}(\frac{\mu}{\tau}\vec{S}\odot\vec{l}) \tag{35}$$

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 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 degradation rate δ would take between 1/54s = 0.019/s or 1/4.6s = 0.22/s . The total population of mRNA in m_tot^* is determined by $1/\delta$, $1/\tau$ and μ (as part of $\vec{S}$) as shown Equation 8. 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.