

Modeling mRNA Populations

N. Pollesch^{a,1,*}, Ricardo Urquidí^d, M.A. Gilchrist^{b,c,1}

^a*Department of Mathematics, University of Tennessee, Knoxville, TN 37996-1320*

^b*Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1610*

^c*National Institute for Mathematical and Biological Synthesis, University of Tennessee, Knoxville, TN 37996-3410*

^d*Genome Science and Technology Program, University of Tennessee, Knoxville, TN 37996-XXX*

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, degradation marking rate, and length of the mRNA. The length of the mRNA's coding region plays the role of controlling the number of ribosome counts which, in turn, determines the number of ODEs in the system. A goal of this work is to see how the equilibrium distribution between classes as changes with coding region length. A closed form solution to the density in the i^{th} ribosomal class in a system with i_{\max} states is presented for the equilibrium distribution of the decapped classes in terms of the capped classes. The equilibrium solutions in the capped classes are shown to be related to the full determinant of the tri-diagonal matrix used to describe the system, as well as all the determinants of the minors associated to it. In general, there is no closed form for the determinant of a tri-diagonal matrix, only a recurrence relation that can be used to find determinants. However, in this model a closed form exists for the full determinant as it changes with changing values of i_{\max} and its formula is presented. This closed form for the determinant provides a method to efficiently find equilibrium solutions for the entire system. Additionally, a continuous approximation using PDE is derived and also used to find equilibrium solutions to the system. Both of these methods for determining equilibrium solutions are utilized in an effort to find the set of parameters that maximizes the likelihood of a given data set. A process for mapping the equilibrium model results to data is also presented and used to begin preliminary estimation of model parameters

*Corresponding author

Email addresses: pollesch@math.utk.edu (N. Pollesch), 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 occurring 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 processing 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 assessing 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 interference 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 restricted to the pioneering round of translation, NGD can also occur 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

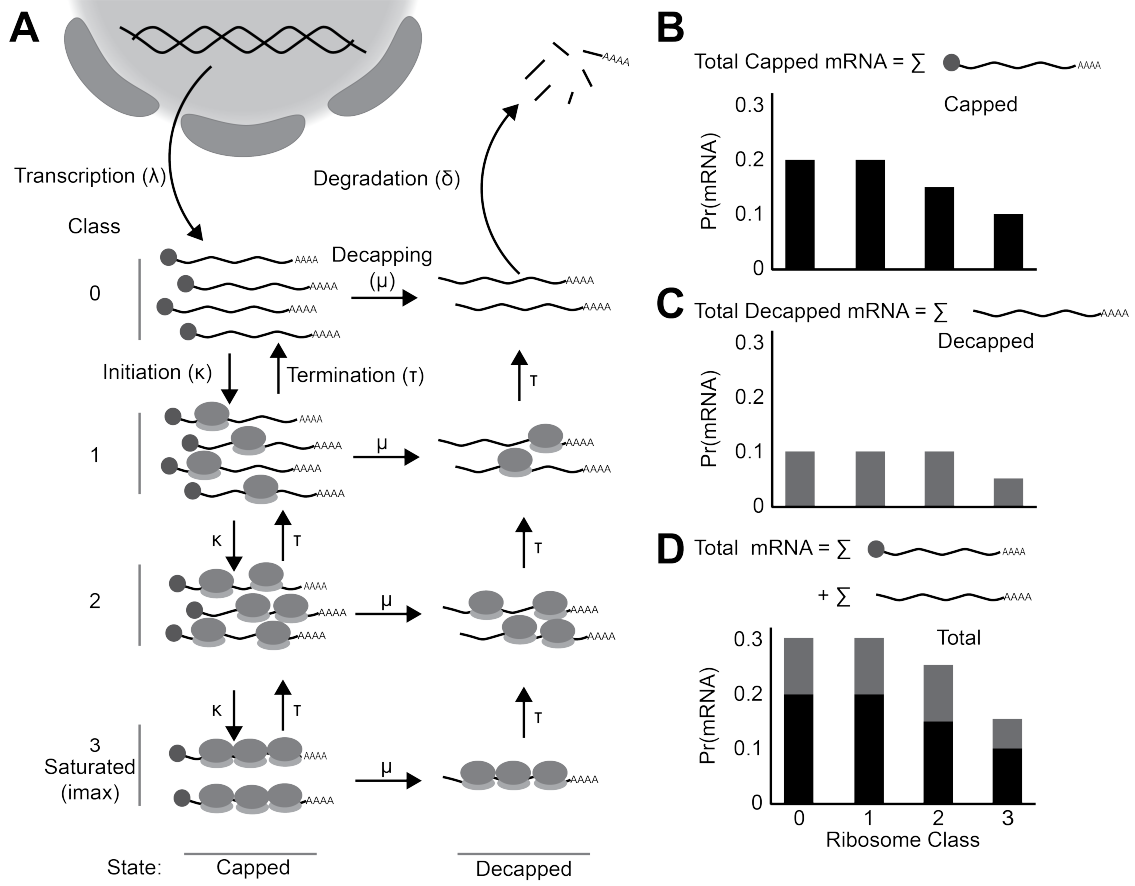
1. The basic representation of the central dogma dictates that expression of protein coding genes starts from genes encoded in DNA that are transcribed to mRNA and subsequently translated to Protein.
2. A more careful representation considers that the final protein production is dependent on both the 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 components 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 initiation 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 subsequently interacting with degradation machinery to remove the faulty transcript.
11. With some mechanisms of mRNA decay, decay can occur co-translationally. This is mainly seen in 5' decapping. When a translating transcript is decapped the 5' to 3' exonucleic degradation machinery trails the most upstream ribosome. As the ribosome translates the mRNA is degraded.
- 12.

2. Methods

2.1. Model Overview

Figure 1: Cartoon Representation of model in biological context. A) Model overview. Transcripts enter the system 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 probability of finding an mRNA in each class across each state. This reflects 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 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 ribosomal 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 (class 0) From capped class 0 a transcript can have two fates. The transcript can be decapped for degradation at rate μ and move into the decapped class 0. Alternatively, a ribosome can initiate translation at rate $\kappa * (1 - i/imax)$ and be loaded onto the transcript and move it into capped class 1. Where i is the current transcript class and $imax$ is the maximal ribosomal occupancy on the transcript A ribosome on transcript can then elongate and terminate at a rate of $\tau * i$. After a ribosome fully elongates and terminates it leaves the transcript and the transcript falls to a lower class. Capped transcripts move through rounds of translation initiation and elongation/termination and distribute along the different ribosomal 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 ribosomal 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 eukaryots (Citation) After all ribosome complete translation, the mRNA is in decapped class 0 and completely degraded at a rate δ . The saturated state of a transcript is denoted as $imax$, meaning that that transcript can no longer accept any more ribosomes. 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. Model Definition and Analysis

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.

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.	<i>mRNA</i>
m_i^*	Abundance of mRNAs with a ribosome load of i in decapped state.	<i>mRNA</i>
Model Parameters		
i_{\max}	Maximum number of ribosomes able to bind to mRNA; defines number of state variables and is a function of gene length.	ND
$\kappa(i)$	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.	<i>mRNA/s</i>
δ	Removal rate of marked mRNA with a ribosome load of 0 from the m_0^* class.	1/s

General model equations of the density independent initiation model

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

Specifically, new mRNA enter the 0^{th} capped class $m_0(t)$ at a rate.. Ribosomal bind mRNAs in the i^{th} capped class at a rate $\kappa(i)$, increasing the mRNA's ribsome load to the $i + 1^{th}$ class. By definition, $\kappa(i_{\max}) = 0$, i.e. mRNAs with a ribosome load of i_{\max} cannot accommodate any additional 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 ribosomal 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 mRNA'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. $\tau(1) = \text{average elongation rate} / (9 \cdot i_{\max})$. 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 $1/s$. At equilibrium every $\tau(i)$ must equal every $\kappa(i)$ by definition. Only mRNAs decapped for degradation and with a ribosome load of 0 are removed from the system. Specifically, decapped mRNAs are removed from the 0^{th} class at a rate of $\delta m_0^*(t)$. All parameters are assumed to fixed for a given gene, but may vary between genes.

The functional form of the capped subsystem is:

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

Similarly, the functional form of the decapped subsystem is:

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

2.3. The density dependent initiation model

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

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

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:

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

and the decapped subsystem is unchanged.

2.4. Matrix-vector Formulation of ODE System

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

$$\vec{M}' = \mathbf{F}\vec{M} + \vec{B}, \quad (2)$$

where $\vec{M} \in \mathbb{R}^{2(i_{\max}+1)}$ is a vector of all state variables, ordered here as $m_0, m_1, \dots, m_{i_{\max}}, m_0^*, m_1^*, \dots, m_{i_{\max}}^*$, \vec{M}' is the vector containing the first derivatives of \vec{M} with respect to time, $\mathbf{F} \in \mathbb{R}^{2(i_{\max}+1) \times 2(i_{\max}+1)}$ is the matrix representing the full system (Equation 3), 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 \mathbf{F} is block lower-diagonal and is given in Equation 3.

$$\mathbf{F} = \begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \boldsymbol{\mu} & \mathbf{R} \end{pmatrix}. \quad (3)$$

The upper-left block, \mathbf{U} , corresponds to the capped state variables, where \mathbf{U} 's general form is provided in Equation 4. The upper-right block is a matrix of all zeros, $\mathbf{0} \in \mathbb{R}^{i_{\max}+1 \times i_{\max}+1}$. Using \mathbf{I} to represent the $i_{\max} + 1 \times i_{\max} + 1$ identity matrix, the lower-left block is $\boldsymbol{\mu} = \mu_0 \mathbf{I}$, a diagonal matrix with the

constant μ_0 on the diagonal and 0s else. The lower-right block, \mathbf{R} , corresponds to the decapped state variables and its form is provided in Equation 5.

The matrix \mathbf{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,

$$\mathbf{U} = \begin{pmatrix} -(\kappa_0 + \mu_0) & \tau_0 \frac{1}{i_{\max}} & & & & & & & \\ \kappa_0 & \left(1 - \frac{1}{i_{\max}} \kappa_0 + \mu_0 + \tau_0 \frac{1}{i_{\max}}\right) & \tau_0 \frac{2}{i_{\max}} & & & & & & \\ & & \ddots & \ddots & \ddots & & & & \\ & & & 1 - \frac{(i-1)}{i_{\max}} \kappa_0 & -\left(1 - \frac{i}{i_{\max}} \kappa_0 + \mu_0 + \tau_0 \frac{i}{i_{\max}}\right) & \tau_0 \frac{i+1}{i_{\max}} & & & \\ & & & & \ddots & \ddots & \ddots & & \\ & & & & & \frac{1}{i_{\max}} \kappa_0 & -\left(\mu_0 + \tau_0 \frac{i_{\max}}{i_{\max}}\right) & & \end{pmatrix} \quad (4)$$

In the representation given in Equation 4, all blank entries are 0. The $(i_{\max} - 1)^{\text{th}}$ row has been suppressed in Equation 4, but it can be generated using the formula included for the i^{th} row.

The matrix \mathbf{R} is the lower-right block in the block lower-diagonal matrix \mathbf{F} (Equation 3),

$$\mathbf{R} = \begin{pmatrix} -\delta & \tau_0 \frac{1}{i_{\max}} & & & & & & & \\ & -\tau_0 \frac{1}{i_{\max}} & \tau_0 \frac{2}{i_{\max}} & & & & & & \\ & & \ddots & \ddots & & & & & \\ & & & -\tau_0 \frac{i-1}{i_{\max}} & \tau_0 \frac{(i+1)}{i_{\max}} & & & & \\ & & & & \ddots & \ddots & & & \\ & & & & & -\tau_0 \frac{(i_{\max}-2)}{i_{\max}} & \tau_0 \frac{i_{\max}}{i_{\max}} & & \\ & & & & & & -\tau_0 \frac{i_{\max}}{i_{\max}} & & \end{pmatrix}, \quad (5)$$

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

2.4.1. Capped Subsystem Matrix-vector Representation

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

$$\vec{m}' = \mathbf{U}\vec{m} + \vec{b}, \quad (6)$$

where $\vec{m} \in \mathbb{R}^{i_{\max}+1}$ is the vector of capped state variables ordered $m_0, \dots, m_{i_{\max}}$, \vec{m}' is the vector containing the first derivatives of \vec{m} with respect to time, $\mathbf{U} \in \mathbb{R}^{i_{\max}+1 \times i_{\max}+1}$ is the matrix representing the capped subsystem (Figure 4), and $\vec{b} \in \mathbb{R}^{i_{\max}+1}$ is the vector of λ as the first component and 0s else. With all equations defined for the full ODE system, include matrix-vector representations, the next section outlines methods for finding steady-state solutions to the system.

2.4.2. Capped state steady state solution

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

$$\frac{dm_0}{dt} = \lambda + \mu m_0 \quad (7)$$

which, at equilibrium results in,

$$m_0 = \lambda/\mu \quad (8)$$

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 multiple classes, as the transcripts get distributed, each class contribute a weighted port of the total μ . Therefore, adding all the contributions together equals:

$$\frac{\lambda}{\mu} = \sum_{i=0}^{i_{\max}} m_i, \quad (9)$$

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:

$$labeled : eq_{general} adj_{determinant} solution \vec{m} = -\frac{1}{\det[\mathbf{U}]} Adj[\mathbf{U}] \vec{b}.$$

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

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

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.

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

Where:

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

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

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

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

$$\begin{aligned} \sum_{i=0}^{i_{\max}} m_i &= -\sum_{i=0}^{i_{\max}} \frac{a_{tot}}{\det[\mathbf{U}]} \lambda \vec{p}_m = -\frac{a_{tot}}{\det[\mathbf{U}]} \lambda = \frac{\lambda}{\mu} \\ -\frac{a_{tot}}{\det[\mathbf{U}]} &= \frac{1}{\mu} \end{aligned}$$

We finally arrive at,

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

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 obtain 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 $\mu=0$.

2.4.3. Decapped Subsystem steady state solution

Starting with the decapped subsystem of equations:

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

We get the following solutions at steady state:

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

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.

$$\frac{\lambda}{\mu} = \sum_{i=0}^{i_{\max}} m_i \quad (10)$$

Therefore,

$$1 = \frac{\mu}{\lambda} \sum_{i=0}^{i_{\max}} m_i \quad (11)$$

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

$$S_j = \frac{\mu}{\lambda} \sum_{i=j}^{i_{\max}} m_i \quad (12)$$

Now the solution becomes,

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

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

$$m_{tot}^* = \sum_{i=0}^{i_{\max}} m_i^* = \frac{\lambda}{\delta} + \frac{\lambda}{\tau} S_1 + \dots + \frac{\lambda}{i\tau} S_i + \dots + \frac{\lambda}{i_{\max}\tau} S_{i_{\max}} \quad (13)$$

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

$$m_{tot}^* = \lambda \left(\frac{1}{\delta} + \frac{1}{\tau} \vec{S} \odot \vec{l} \right) \quad (14)$$

Where \vec{S} is a vector of all the cummulative sums and \vec{l} is a vector of $1, 1/2, \dots, 1/i, \dots, 1/i_{\max}$. The S_i have the following arrangement $S_0 = 1$ and $S_0 \geq S_1 \geq \dots \geq S_i \geq \dots \geq S_{i_{\max}}$. 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 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,

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

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

2.5. Complete system mRNA population

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

$$M_{tot} = \lambda \left(\frac{1}{\mu} + \frac{1}{\delta} + \frac{1}{\tau} \vec{S} \odot \vec{l} \right) \quad (15)$$

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.

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

$$p_{mtot} = \frac{1}{\left(1 + \frac{\mu}{\tau} \vec{S} \odot \vec{l} \right)}$$

Then you calculate the odds,

$$odds_m = \frac{p_{mtot}}{1 - p_{mtot}} \quad (16)$$

$$odds_m = \frac{\frac{1}{(1 + \frac{\mu}{\tau} \vec{S} \odot \vec{l})}}{1 - \frac{1}{(1 + \frac{\mu}{\tau} \vec{S} \odot \vec{l})}} \quad (17)$$

It simplifies to,

$$odds_m = \frac{1}{\frac{\mu}{\tau} \vec{S} \odot \vec{l}} \quad (18)$$

$$\log_{10}(odds_m) = -\log_{10}\left(\frac{\mu}{\tau} \vec{S} \odot \vec{l}\right) \quad (19)$$

2.6. model implementation in R

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

2.7. Data Sources

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

$$\mu = \frac{\ln(2)}{t_{1/2}} \quad (20)$$

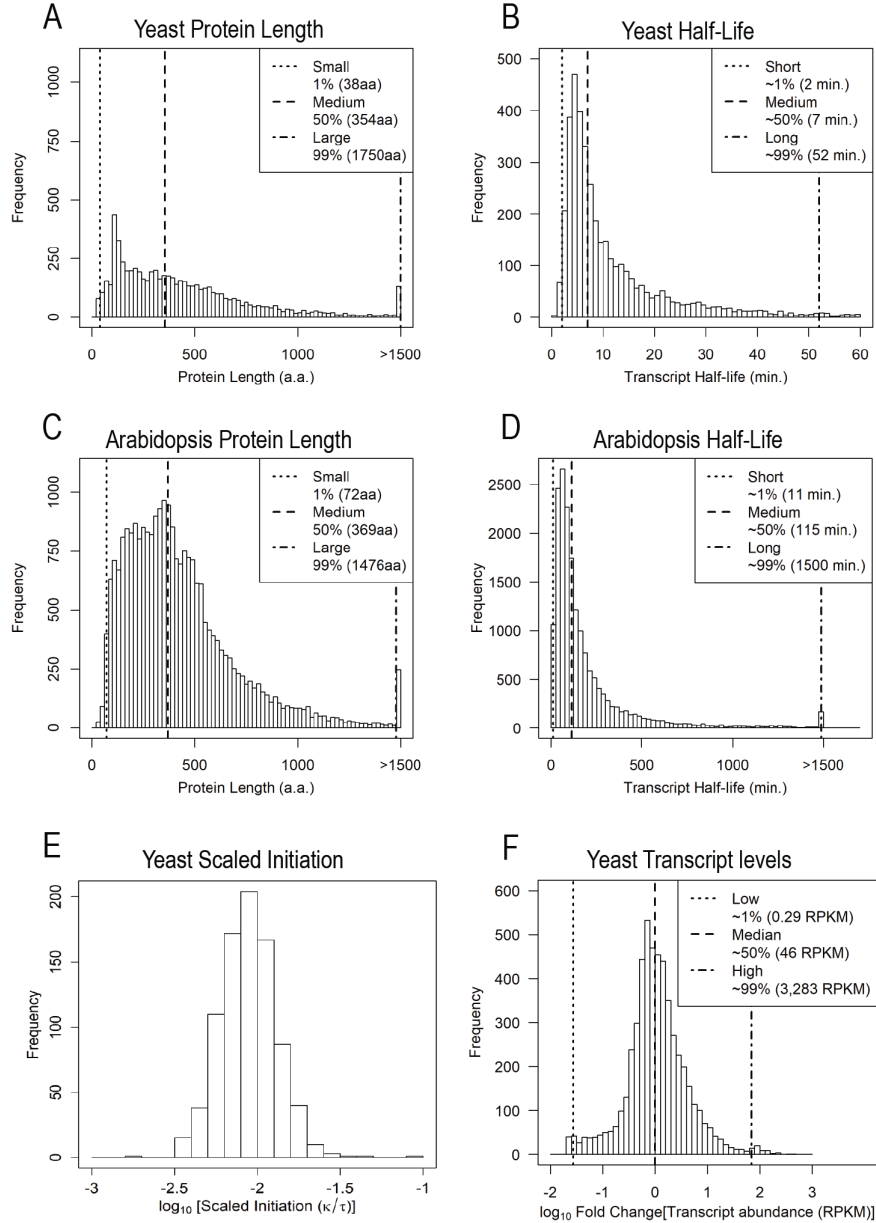
Where $t_{1/2}$ is the half-life. The resulting range of μ is from 10^{-2} to 10^{-4} for yeast and 10^{-3} to 10^{-6} for arabidopsis.

Translation initiation and elongation rates (κ and $\tau * 9 * i_{\max}$) were obtained for Yeast from Duc and Song 2018. Remember, that τ is the overall rate of translating a transcript, not the average elongation rate (see section 2.2). In Duc and 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 is due to the fact that many combinations of κ and τ can yield the same scaled initiation rate (e.g. $\kappa=0.02$ and $\tau * 9 * i_{\max} = 2$, and $\kappa=0.04$ and $\tau * 9 * i_{\max}=4$, both yield $\kappa/\tau=0.01$). This simplifies the model behavior to one generalized parameter with a unique response (Figure 2E). The scaled initiation rate ranges from $0.1s^{-1}$ to $0.001s^{-1}$.

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

Finally, δ only determines the accumulation of transcripts in the m_0^* class. In our model the remaining transcripts in class m_0^* are either only the 3' end of co-translationally degraded transcripts or full transcripts from class m_0 . Recently, the rate of degradation for the 5' - 3' exonuclease XRN1 was determined to be 26 nt/s (Atthapattu 2021). XRN1 is the primary exonuclease involved in co-translational degradation and 5' degradation pathways (Sorenson 2018, Yu 2016, Collart 2019, Pelechano 2015). For an average 3' UTR of 121 nts (Kebaara 2009) this would take 4.6s, and an average transcript of 1400nt would take 54s to degrade. This means the average degradation rate δ would take between $1/54s = 0.019/s$ or $1/4.6s = 0.22/s$. The total population of mRNA in m_{tot}^* is determined by $1/\delta$, $1/\tau$ and μ (as part of \vec{S}) as shown Equation 14. τ ranges from 0.03 to 10^{-4} . This makes $1/\delta \leq 1/\tau$. It is reasonable to explore the model with large δ since decapped transcripts are translationally incompetent.

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



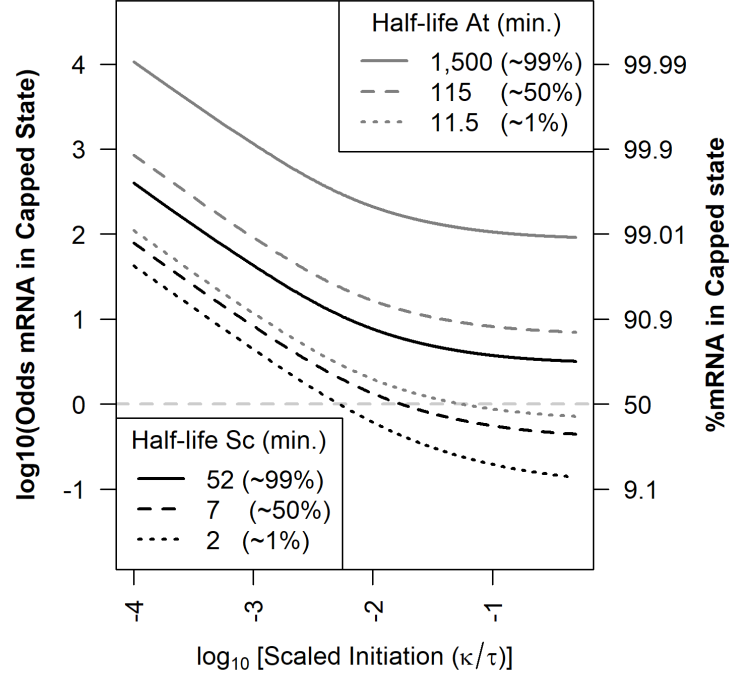
3. Results

Main results:

Supplemental Figure 1

1. The model solution at equilibrium has three outputs.

Figure 3: Comparison of density independent initiation (DII) and density dependent initiation (DDI)



2. First the total number of mRNA transcripts. Second split between the capped and decapped states. And third the distribution of the transcripts across ribosomal classes for each of the states.

3. s

4. After exploring the model solutions it was determined that the transcription rate λ acts simply as a scalar for the whole system.

Text for discussion of figure 1

1. The model presented most closely resembles the co-translational decay pathway from eukaryotes.
2. There are a few current limitations of the model. First and foremost it doesn't model endonucleolytic decay pathways such as gene silencing by siRNA and the RISC complex or the NMD, NGD and NSD pathways or degradation initiated through ribosomal collisions.
3. Implementation of endonucleolytic decay would require an ability to track ribosomal position, which is not possible in our model.

4. Alternatively a probabilistic model of ribosomal position given a particular load could allow for redistribution of capped transcripts into decapped states of lower ribosomal class.
5. Currently there is a debate whether mRNA stability is regulated primarily through the protective effects of ribosomal association or through the suboptimal codons causing ribosomal stalling and the subsequent ribosome associated decay pathways (Chan et al).
6. We explored the interaction of ribosomal loading and marking rate, noting that increased marking leads to a lower expected ribosomal load. This is particularly noticeable for genes with short half lives and is less sensitive with transcripts with halflives above 2 hours.
7. In the current implementation of the model we did not directly explore the protective effects of ribosomal loading. This could be first implemented by including a similar weighting term analogous to the weight for the initiation rate of $(1-i/i_{max})$.
8. However, biology suggests a more complex behavior.
9. The protective effects of translation could increase per ribosome, but eventually at high loads could trigger ribosome associated decay pathways through ribosomal collisions.
10. Our model does not model elongation at a single codon level like TASEP, nor at a coarse grained approximation of groups of codons like the ribo flow models. This means that it is not capable of analyzing the effect of individual non optimal codons, or stretches of mRNA which are non-optimal
11. Our model reflects the average elongation rate over a transcript. However it can be set to reflect the slowest bottle neck across a transcript.
12. Another alternative to modeling bottlenecks is increasing granularity to a very coarse grained ribo-flow model comprising of only two or three sectors, depending whether the transcript has one or two bottlenecks.
13. This splits transcripts into an initial section where the bottleneck effectively "shortens" the transcript length to the first bottleneck site. Increasing the effect of ribosome loading on initiation interference.

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

Model predicts mRNA distribution across ribosomal classes

1. The model
2. Density independent initiation and density independent initiation models
3. The model predicts the steady state total mRNA quantity in both decapped and capped states as well as the mRNA's distribution across different ribosomal classes in each state.
4. Figure X Left,

The model provides an upper bound on translational efficiency

1. The behavior of both DDI and DII models were explored at equilibrium. At equilibrium, the rate of new initiation events must be equal to the rate of ribosomes leaving the transcript.
2. The two key parameters that control this process are initiation (κ) and elongation (τ). Changes in either parameter alter the predicted distribution of transcripts across ribosomal classes.
3. To summarize their joint behavior, we chose to focus on the elongation scaled initiation rate ($\kappa \tau$) as this provides a single unique parameter for each model response.
4. Exploration of the steady state solution showed that the transcription rate only acts to scale the steady state solution. Meaning that the distribution of transcripts across the ribosomal classes is independent of the number of transcripts being produced. All model solutions presented here have the same transcription rate to allow for comparison.
5. The degradation rate (δ) only controls the transcript population in the decapped zeroth class. To reduce interference from this we set a very high degradation rate.
6. To interpret model behavior within a biologically meaningful context, empirically determined rates for translation initiation, elongation, and half-life (as a proxy for marking rate) were obtained from the literature.
7. Introduce scaled initiation rate range. Discuss upper and lower limits vis a vis the model output.

8. *** This is a very important question. How do we show in a convincing manner that the model is recapitulating actual biological results? I have previously used the argument that both polysome profiling and single molecule imaging of in-vivo ribosomes all agree with the approximate range predicted by the model, especially within the scaled initiation values. That is the model predicts that the ribosome load will be between 10-50%, with median loads of 2 - 13 ribosomes for a transcript of median length. This is reasonable and the median behavior aligns with what polysome profiles show. However, Polysome profiles contain all of the mRNAs in a cell or organism at once. This means that the fact that the model predic
- (-3 to -0.6 in log ten). The marking rate was determined from Arabidopsis mRNA half-lives with a median half-life of 115 min (short 1st percentile: 11.5 min, long 99th percentile: 1,500 min) [32] To establish the base model behavior, DII and DDI models were run with a long half-life across a broad range of scaled initiation on a transcript encoding a protein of median length (369aa) (Figure 1). In the DII model, all ribosomal classes are populated. At the high end of the biological scaled initiation transcripts reach full ribosomal occupancy (Figure 1A). The DII model ignores initiation interference caused by ribosomes already present on the transcript. The DDI model accounts for initiation interference, and as expected, a decrease in ribosomal load is observed compared to DII. Even with interference, access to most ribosomal classes is attainable in the DDI model. Ribosomal profiling experiments suggest that most transcripts have one to ten ribosomes loaded with a strong preference for smaller polysomes. Live single molecule fluorescence experiments do show higher polysome numbers rising into the 50s [33]. For the long reporters used in these experiments that equates to a 12 to 30 percent ribosomal load. Thus the DDI model probably overestimates the loading for a particular scaled initiation rate.

Marking rates have only a mild effect on ribosomal load, but a large effect on protein production.

1. From the solution of the capped system we can define two separate effects of transcript marking.
 - (a) First, marking determines how many transcripts can accumulate in the capped class, thus determining the total transcript population
 - (b) Second, marking rate shifts the distribution of transcripts in the capped class towards lower ribosomal classes.
2. Marking only has a mild effect on the distribution of transcripts on the capped class

Under equilibrium protein length affects ribosomal load but not ribosomal density

1. stuff

Figure 2 points for discussion

1.

1. As more ribosomes are loaded onto a transcript there is reduced space available for new ribosomes to occupy.

2. This in turn would reduce the ability to initiate new rounds of translation.

3. To investigate the effect of ribosomal density on initiation we created two versions of the model. A density dependent initiation (DDI) model which scales the initiation rate by the number of ribosomes on the transcript. And a density independent initiation (DII) version of the model which has a fixed initiation rate.

4. In the DII model the ribosome load increases linearly with the scaled initiation rate ($\kappa \tau$) until it reaches

5. Our model does not explicitly capture the mechanism of translation, it just keeps track of the available spaces left on a transcript. As such the model would underestimate the actual effect of ribosomal interference

Figure 4: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

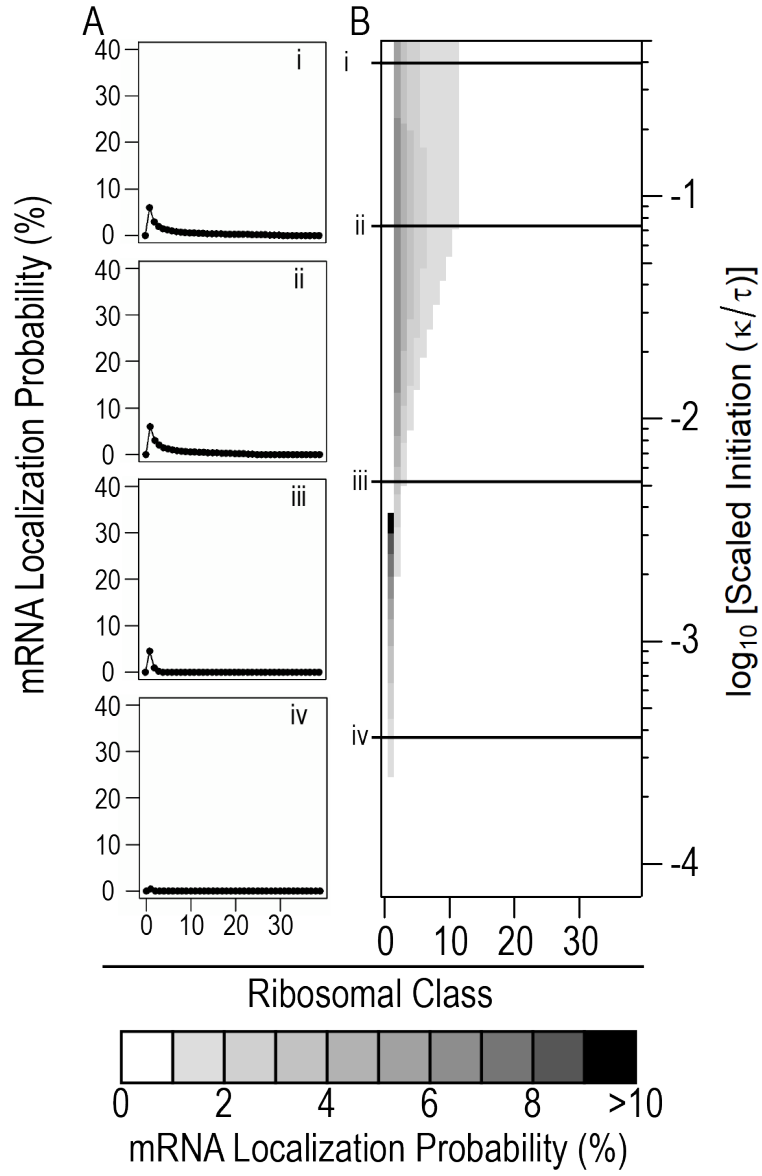


Figure 5: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

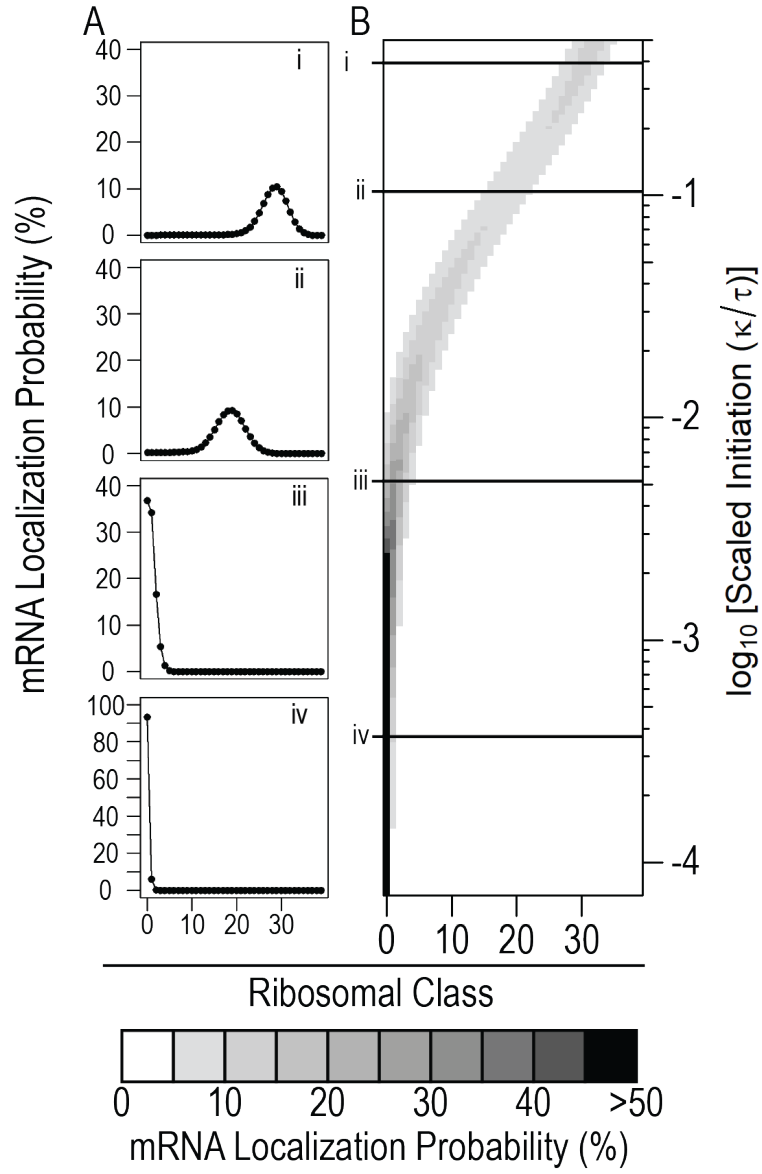


Figure 6: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

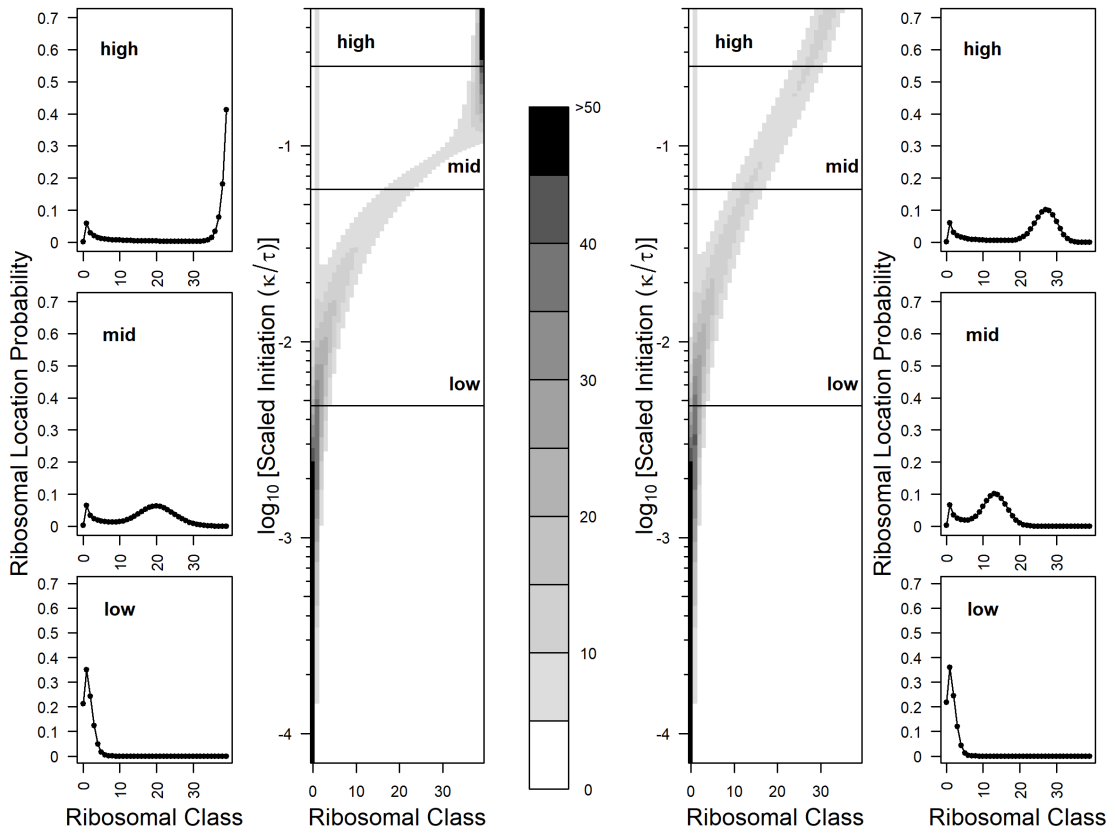


Figure 7: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

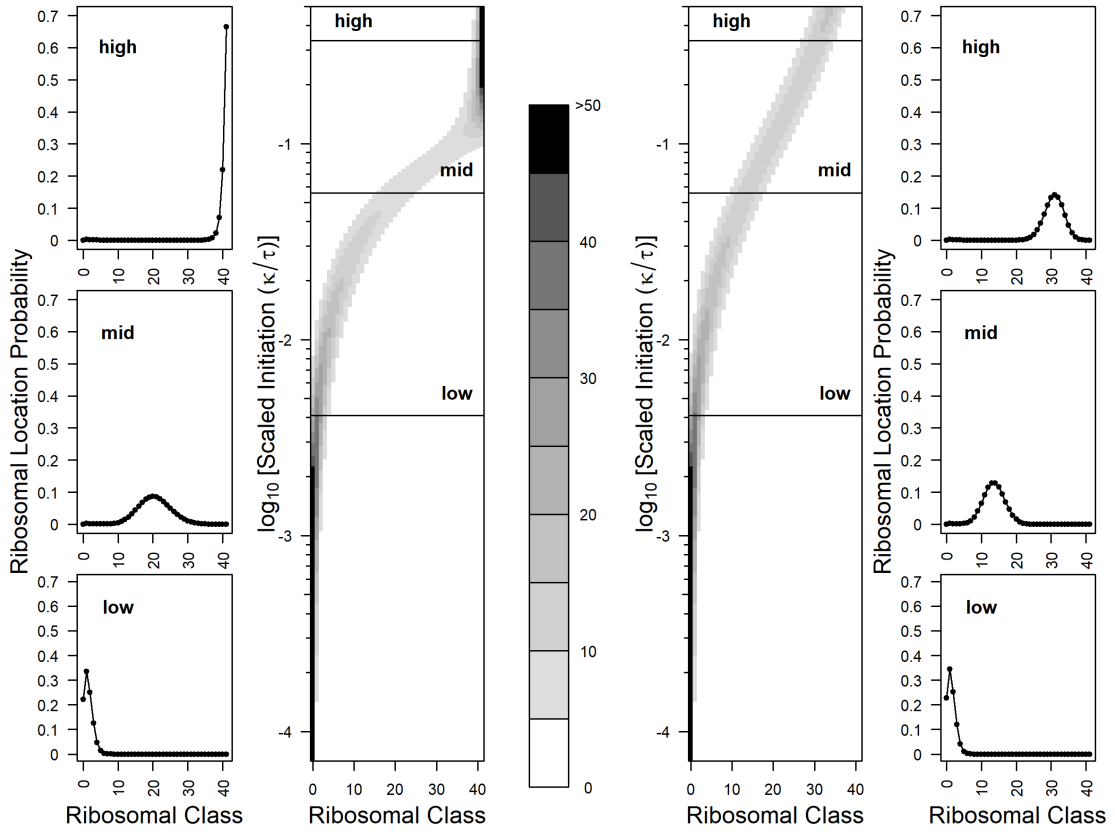


Figure 8: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

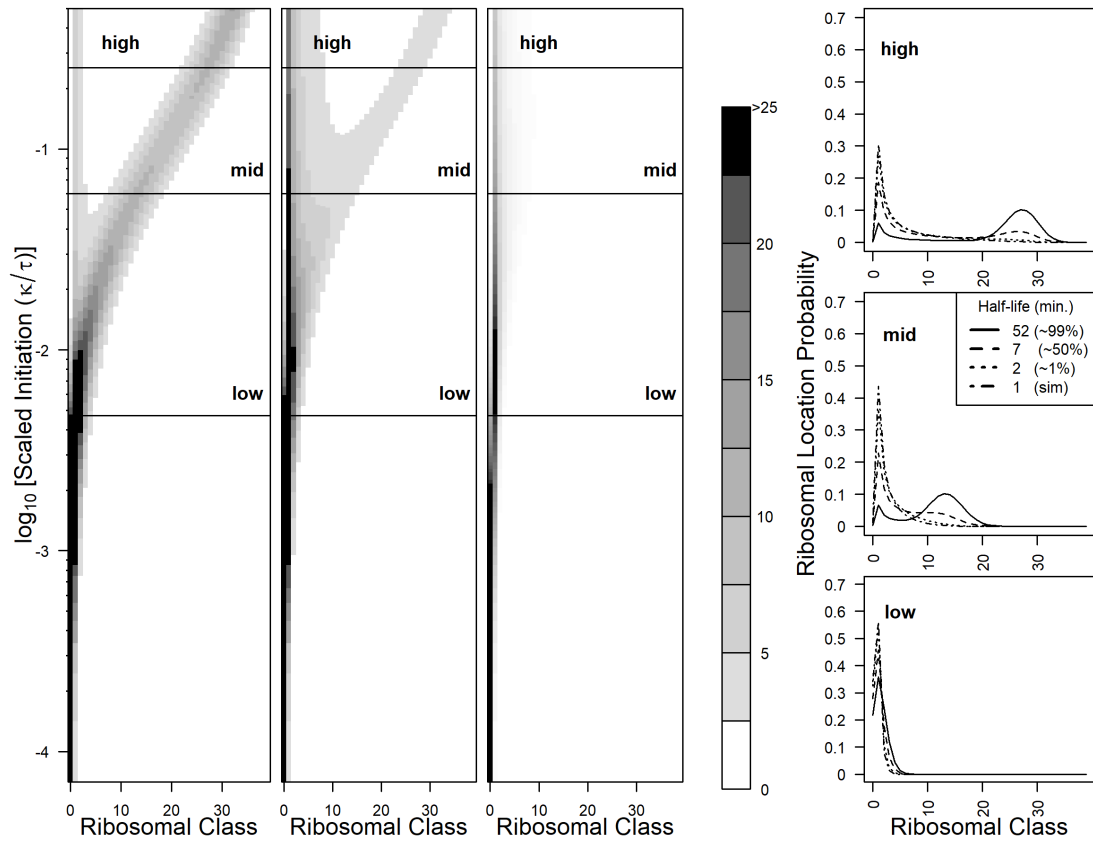


Figure 9: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

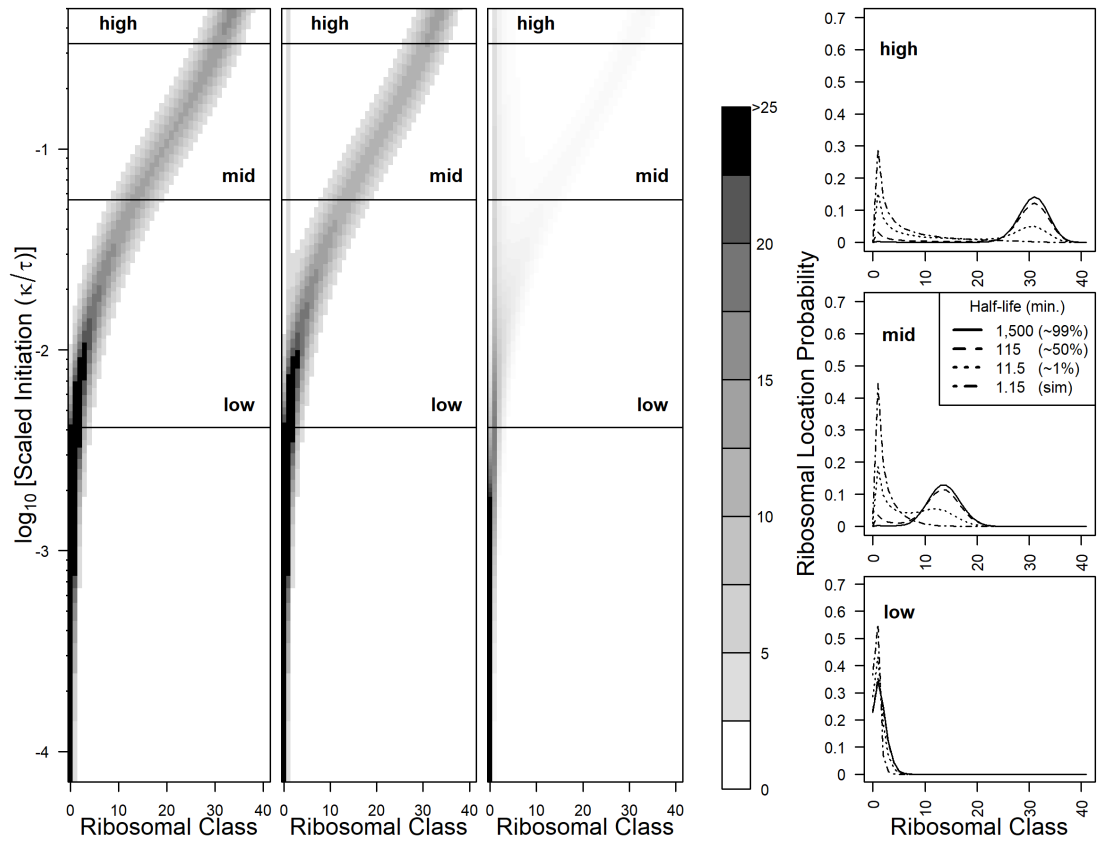


Figure 10: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

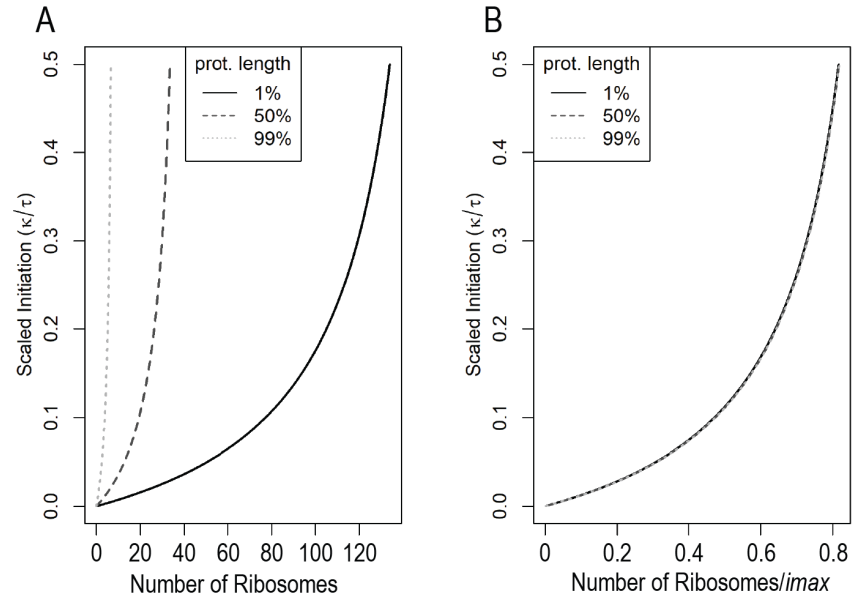


Figure 11: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

