Modeling mRNA Populations

N. Pollesch[a,b,*], M.A. Gilchrist[c,d]

[a]*Department of Mathematics, University of Tennessee, Knoxville, TN 37996-1320*
[b]*Department of ??, Environmental Protection Agency, Duluth MN XXXXX*
[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 marked for future degradation. Distribution of ribosome counts is assumed to be related to the translation initiation rate, translation completion rate, degredation marking rate, and length of the mRNA. The length of the mRNA's coding region plays the role of controlling the number of ribosome counts which, in turn, determines the number of ODEs in the system. A goal of this work is to see how the equilibrium distribution between classes as changes with coding region length. A closed form solution to the density in the i^{th} ribosomal class in a system with $i_{\max}$ states is presented for the equilibrium distribution of the marked classes in terms of the unmarked classes. The equilibrium solutions in the unmarked classes are shown to be related to the full determinant of the tri-diagonal matrix used to describe the system, as well as all the determinants of the minors associated to it. In general, there is no closed form for the determinant of a tri-diagonal matrix, only a recurrence relation that can be used to find determinants. However, in this model a closed form exists for the full determinant as it changes with changing values of $i_{\max}$ and its formula is presented . This closed form for the determinant provides a method to efficiently find equilibrium solutions for the entire system. Additionally, a continuous approximation using PDE is derived and also used to find equilibrium solutions to the system. Both of these methods for determining equilibrium solutions are utilized in an effort to find the set of parameters that maximizes the likelihood of a given data set. A process for mapping the equilibrium model results to data is also presented and used to begin preliminary estimation of model parameters and to verify model function.

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

*Corresponding author
Email addresses: pollesch@math.utk.edu (N. Pollesch), mikeg@utk.edu (M.A. Gilchrist)

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 marked and unmarked 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, marked and unmarked. - **(Nate)**
3. Model Formulation: Total model presented and then analysis of Unmarked and Marked systems - **(Nate)**
   (a) ODE/Discrete system
      i. Present system of ODEs (Total, Unmarked, and Marked)
      ii. Matrix Representation of ODE model (Total, Unmarked, and Marked)
      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 unmarked 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 marked 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.

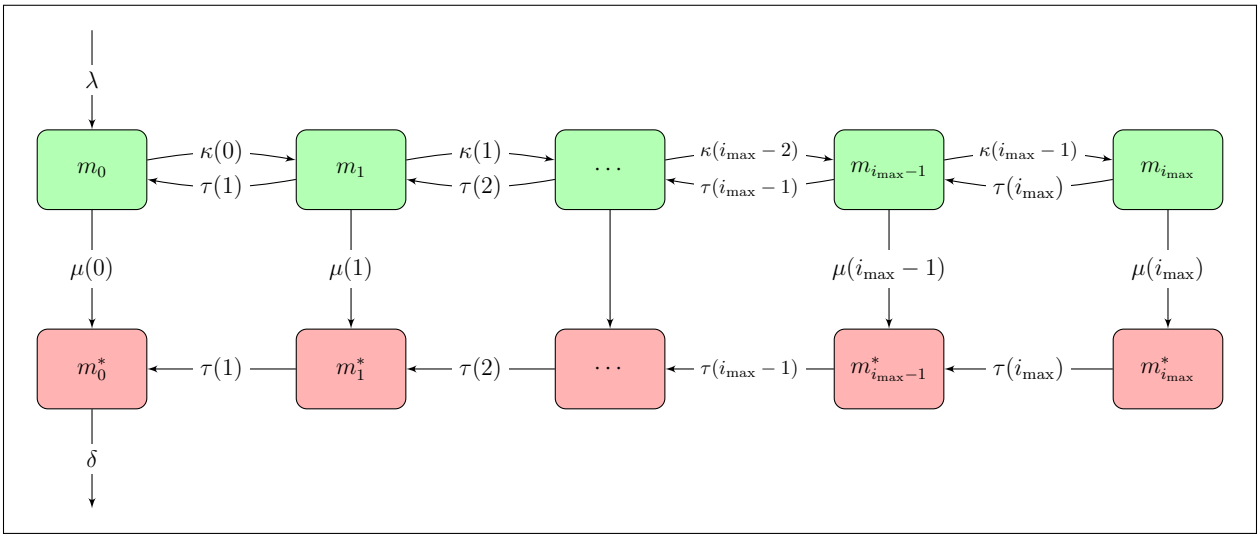
3. Current Models/Research and how our model fits in the current field

2. Model Description and Notation

The mechanistic view that is taken for building this model is as follows. Unbound, unmarked mRNA are introduced to the system at a constant rate of λ. If an unmarked mRNA has a ribosome load of 0, it is defined as being in class m_0 and contributes to that class's abundance. Ribosomal binding to mRNA in class i, m_i, is assumed to proceed at rate $\kappa(i)$. Ribosomes completing translation of the coding region are released from the mRNA at rate $\tau(i)$. mRNAs in the unmarked classes can then transition freely between the m_0 and $m_{i_{\max}}$ according to initiation and completion rates until they are marked, which ultimately leads to the degradation of the mRNA. To begin the process of mRNA degradation, marking of an mRNA in class i takes place at rate of $\mu(i)$. A marked mRNA with a ribosome load of i is denoted as being in class m_i^*. Once an mRNA has been marked, it is assumed that no more initiation can take place and only translation

completion occurs; this completion occurs until the mRNA's ribosome load drops to 0, at which point it leaves the system from state m_0^* at a constant rate of δ. Figure 1 provides a diagram of system flow to accompany the description above. Table 1 provides a summary of state variable and parameter notation.

Figure 1: Flow diagram for ODE system describing transitions within the mRNA population model. State variables and parameters are defined in Table 1. Model description and notation are described in Section 2.



3. 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.

General Model Equations

Our model consists of two sets of time dependent and coupled ODEs. Each set of ODEs describes the abundance of mRNAs that are either unmarked and marked 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 unmarked mRNAs or the degradation of marked mRNAs with a ribosome load of 0.

Specifically, new mRNA enter the 0^{th} unmarked class $m_0(t)$ at a rate.. Ribosomal bind mRNAs in the i^{th} unmarked 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. Unmarked mRNAs with ribosome load i are marked at a rate of $\mu(i)$. Accordingly, the ribosome load of marked mRNAs remains unchanged, but they are transitioned from the unmarked class $m_i(t)$ to the marked class $m_i^*(t)$.

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
State Variables	
m_i	Abundance of mRNAs with a ribosome load of i not marked for degradation.
m_i^*	Abundance of mRNAs with a ribosome load of i marked for degradation.
	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.
$\kappa(i)$	Translation initiation rate for unmarked mRNAs with a ribosome load of i.
$\tau(i)$	Translation completion rate for the marked and unmarked mRNAs with a ribosome load of i.
$\mu(i)$	Marking rate for unmarked mRNAs with a ribosome load of i.
λ	Production rate of newly produced, ribosome free, and unmarked mRNA to the m_0 class.
δ	Removal rate of marked mRNA with a ribosome load of 0 from the m_0^* class.

Ribosome movement along an mRNA is assumed to occur independent of whether or not its unmarked or marked for degradation. Thus, ribosomes complete translation of both marked and unmarked mRNAs with ribosome load i at rate $\tau(i)$, decreasing the mRNA's ribosome load to the $i-1^{th}$ class. Since mRNA's with a ribosome load of 0 have no ribosomes which can complete translation, by definition $\tau(0) = 0$. Only mRNAs marked for degradation and with a ribosome load of 0 are removed from the system. Specifically, marked mRNAs are removed from the 0^{th} class at a rate of $\delta m_0^*(t)$.

The functional form of the unmarked 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 marked subsystem is:

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

3.1. Model Equations Assuming Specific Transition Functions

As stated above, we assume mRNAs marked for degradation cannot initiate translation. Thus the initiation function for the marked set of mRNAS is $\kappa(i) = 0$. In contrast, ribosomes can initiate translation on an mRNA marked for degradation. For simplicity, we assume ribosome binding of an mRNA is independent of their ribosome load, but that successful initiation of translation is affected by ribosome load. Specifically, we assume the start codon must be unoccupied by a ribosome in order for translation initiation to be successful.

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. 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,

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

where κ_0 is a gene specific parameter that describes the rate at which unmarked 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).

Similarly, the probability a ribosome is at the penultimate codon (i.e. the codon immediately upstream from the stop codon) is $i/i_{\max}$. Assuming that the rate of translation termination is greater than the average elongation rate (i.e. translation termination is not a rate limiting step), we can define our translation completion rate function as,

$$\tau(i) = \tau_0' i/i_{\max} = \tau_0 i, \tag{2}$$

where τ_0 describes the average elongation rate of a ribosome along the protein coding sequence of that gene and, thus, could vary between genes. Finally, we assume that unmarked mRNAs are marked for degradation rate independent of their ribosome load, i.e. $\mu(i) = \mu_0$. As with τ_0, μ_0 is assumed to fixed for a given gene, but may vary between genes.

mikeg: 10/15/1
updated τ t
explicitly includ
$1/imax$. Earli
notes indicate th
term was implicitl
absorbed into τ_0
mikeg: 10/22/1
updated to introduc
τ_0' and then reabsor
$i_{\max}$ into τ_0 carryin
this term through
TODO update a
later instances
τ_0 with $\tau_0'/i_{\max}$
and then simplif
math by cancelin
out $i_{\max}$ term an
changing notation t
remove the $'$ since
becomes unnecessar
at that point.

Incorporating these explicit functional forms leads to representation of unmarked subsystem as:

$$\begin{aligned}
\frac{dm_0}{dt} &= \lambda + \tau_0 m_1 - (\kappa_0 + \mu_0)\, m_0 \\
\frac{dm_1}{dt} &= \kappa_0 m_0 + 2\tau_0 m_2 - \left(\tau_0 + \kappa_0 \left(1 - \frac{1}{i_{\max}}\right) + \mu_0\right) m_1 \\
&\vdots \\
\frac{dm_i}{dt} &= \kappa_0 \left(1 - \frac{i-1}{i_{\max}}\right) m_{i-1} + (i+1)\tau_0 m_{i+1} - \left(i\tau_0 + \kappa_0 \left(1 - \frac{i}{i_{\max}}\right) + \mu_0\right) m_i \\
&\vdots \\
\frac{dm_{i_{\max}}}{dt} &= \kappa_0 \left(1 - \frac{i_{\max} - 1}{i_{\max}}\right) m_{i_{\max}-1} - (i_{\max}\tau_0 - \mu_0)\, m_{i_{\max}}
\end{aligned}$$

and the marked subsystem as,

$$\begin{aligned}
\frac{dm_0^*}{dt} &= \mu_0 m_0 + \tau_0 m_1^* - \delta m_0^* \\
\frac{dm_1^*}{dt} &= \mu_0 m_1 + 2\tau_0 m_2^* - \tau_0 m_1^* \\
&\vdots \\
\frac{dm_i^*}{dt} &= \mu_0 m_i + (i+1)\tau_0 m_{i+1}^* - i\tau_0 m_i^* \\
&\vdots \\
\frac{dm_{i_{\max}}^*}{dt} &= \mu_0 m_{i_{\max}}^* - i_{\max}\tau_0 m_{i_{\max}}.
\end{aligned}$$

3.2. 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 marked and unmarked mRNAs can be represented as,

$$\vec{M}' = \boldsymbol{F}\vec{M} + \vec{B}, \tag{3}$$

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 4), 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 4.

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{U} & \boldsymbol{0} \\ \boldsymbol{\mu} & \boldsymbol{R} \end{pmatrix}. \tag{4}$$

The upper-left block, $\boldsymbol{U}$, corresponds to the unmarked state variables, where $\boldsymbol{U}$'s general form is provided in Equation 5. 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 marked state variables and its form is provided in Equation 6.

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,

$$\boldsymbol{U} = \begin{pmatrix} -(\kappa_0+\mu_0) & \tau_0 & & & & & & \\ \kappa_0 & \left(1-\frac{1}{i_{\max}}\kappa_0+\mu_0+\tau_0\right) & 2\tau_0 & & & & & \\ & \ddots & \ddots & \ddots & & & & \\ & & 1-\frac{(i-1)}{i_{\max}}\kappa_0 & -\left(1-\frac{i}{i_{\max}}\kappa_0+\mu_0+i\tau_0\right) & (i+1)\tau_0 & & & \\ & & & \ddots & \ddots & \ddots & & \\ & & & & \frac{1}{i_{\max}}\kappa_0 & -(\mu_0+i_{\max}\tau_0) & & \end{pmatrix}. \tag{5}$$

In the representation given in Equation 5, all blank entries are 0. The $(i_{\max}-1)^{\text{th}}$ row has been suppressed in Equation 5, but it can be generated using the formula included for the i^{th} row.

The matrix $\boldsymbol{R}$ is the lower-right block in the block lower-diagonal matrix $\boldsymbol{F}$ (Equation 4),

$$\boldsymbol{R} = \begin{pmatrix} -\delta & \tau_0 & & & & & \\ -\tau_0 & 2\tau_0 & & & & & \\ \ddots & \ddots & & & & & \\ & -(i-1)\tau_0 & (i+1)\tau_0 & & & & \\ & & \ddots & \ddots & & & \\ & & & -(i_{\max}-2)\tau_0 & i_{\max}\tau_0 & & \\ & & & & -i_{\max}\tau_0 & & \end{pmatrix}, \tag{6}$$

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

3.2.1. Unmarked Subsystem Matrix-vector Representation

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

$$\vec{m}' = \boldsymbol{U}\vec{m} + \vec{b}, \tag{7}$$

where $\vec{m} \in \mathbb{R}^{i_{\max}+1}$ is the vector of unmarked 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 unmarked subsystem (Figure 5), 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.

4. Determining steady-state solutions

A goal of this research is to confront the mechanistic model proposed with empirical gene expression data. This comparison takes place using the equilibrium distribution of polysomal classes. As such, analytical solutions to equilibrium distributions are seen as desirable for the parameter inference step. Although this model formulation has characteristics that make the derivation of analytical solutions seem attainable, a unique challenge arises from the flexibility to define the maximum number of polysomal classes through the parameter $i_{\max}$. It is likely clear to the reader at this point, but is worth explicit mention that the number of state variables in the model is controlled by parameter $i_{\max}$. Therefore whatever scheme is developed to solve for steady-state distributions among polysomal classes must be flexible enough to handle a varying number of states. Although challenging, the parameterizable number of states necessitates the use and development of a set of equilibrium solution methods that differ slightly from the standard ODE equilibrium solution fare.

For this study two different methods for finding steady-state solutions for the polysomal classes were developed. These methods include (1) an iterative scheme that leverages the tri-diagonal form of the unmarked classes in the matrix-vector representation (2) a partial differential equation approximation where the polysomal class is assumed to be continuous.

Our discussion of equilibrium solution methods begins using the original formulation of the ODEs. This scheme for the matrix-vector formulation of the original system was pursued to find analytical solutions, while the continuous approximation through the PDE was derived as an alternate approach to be solved numerically.

4.1. ODE system

Steady-state analytical solutions for equilibrium distributions of the ODE system enables efficient parameter inference. Decoupling of the unmarked subsystems allows one to focus on it and use the determined values for solving the marked subsystem.

4.1.1. Marked subsystem: $\vec{m}$

The analytical form of the equilibrium distribution in the marked subsystem is quite concise. Solved for as a function of the unmarked distribution, the marked distribution at equilibrium is as follows: Let $\hat{m}_i$ be

the steady state or equilibrium solution the i^{th} state in the unmarked subsystem. Then the solution to the i^{th} state in the marked subsystem at equilibrium, denoted $\hat{m}_i^*$, is given by:

$$\hat{m}_i^* = \frac{\mu_0}{\tau_0} \frac{\sum_{j=i}^{i_{\max}} \hat{m}_j}{\gamma(i)}, \text{ where } \quad \gamma(i) = \begin{cases} \frac{\delta}{\tau_0} & i = 0 \\ \\ i & i > 0 \end{cases} . \tag{8}$$

In contrast to the solutions for the components of the marked subsystem, the unmarked subsystem components do not have such a straight-forward analytical representation.

4.1.2. Unmarked subsystem: $\vec{m}$

Analytical steady-state solutions to the unmarked subsystem were generated for low levels of $i_{\max}$ ($i_{\max} = 1, 2, ..., 9$) using the symbolic computation capabilities of *Mathematica*. Visual inspection and hand-calculation was used to determine patterns in these solutions. Upon investigation it was found that the uppermost, $i_{\max}$, polysomal class has the closed-form equilibrium solution of:

$$\hat{m}_{i_{\max}} = \prod_{i=1}^{i_{\max}} \frac{i \frac{\kappa_0}{i_{\max}}}{\mu_0 + i\left(\tau_0 + \frac{\kappa_0}{i_{\max}}\right)}, \tag{9}$$

$$= \frac{\left(\frac{\kappa_0}{i_{\max}}\right)^i_{\max} i_{\max}!}{\prod_{i=1}^{i_{\max}} \mu_0 + i\left(\tau_0 + \frac{\kappa_0}{i_{\max}}\right)} \tag{10}$$

for any level of $i_{\max}$. As a general observation, for all $i_{\max}$ levels investigated the complexity of analytical solutions increased as one moved from higher to lower polysomal classes. In fact, beyond the uppermost polysomal class (Eq 22), no other classes were able to be determined completely in closed-form. Notably, as a side-benefit having a closed-form solution for the upper-most class is beneficial when considering other solution methods, such as PDE approximation, where such a solution can provide a boundary condition.

mikeg: 10/22/1
Note we no longe
have any GCD sinc
they cancelled ou
See text in Not
section at end
document. Nat
can you verify th
product term shoul
be from $i = 1$ an
not $i = 0$?

It follows that the formula for the determinant of $\boldsymbol{U}$ is,

$$\det[\boldsymbol{U}] = (-1)^{i_{\max}+1} \prod_{i=0}^{i_{\max}} \left(\frac{i}{i_{\max}}(\kappa_0) + (\mu_0 + i\tau_0)\right). \tag{11}$$

. It can also be shown that this formula may be rewritten in terms of the Pochhammer symbol, or rising factorial [Show in appendix or cite a source.]. Considering that in general there is no closed form for the determinant of a tri-diagonal matrix, the fact that this formula exists is a unique and desirable property of our system of ODEs.

mikeg: 10/22/1
note that by havin
product term begi
at $i = 0$ rath
than $i = 1$ as i
previous equation
we can incorporat
the orphan μ_0 int
the product term

With a closed-form expression for the determinant of the unmarked class matrix $\boldsymbol{U}$ we are now able to formulate another method to determine equilibrium solutions in the unmarked class by utilizing the matrix-vector formulation of the unmarked subsystem. At equilibrium $\vec{m}' = 0$, and so the matrix-vector equation

(Eq. 7), when solved for the vector of state variables, $\vec{m}$, becomes,

$$\vec{m} = -\boldsymbol{U}^{-1}\vec{b}. \tag{12}$$

and making determination of equilibrium solutions come down to deriving an efficient technique for inverting $\boldsymbol{U}$.

The closed-form of the determinant enables the identification of parameters in which $\boldsymbol{U}$ is invertible. Using the fact that $\boldsymbol{U}$ is invertible iff $\det[\boldsymbol{U}] \neq 0$. Equation 11 shows that the only possible scenario where $\boldsymbol{U}$ is non-invertible is when $\mu_0 = 0$. Given that μ_0 controls the rate at which unmarked mRNA become marked for deletion, the physical interpretation of $\mu_0 = 0$ is no degradation of unmarked mRNA and thus no loss in the system, leading to linear, unbounded growth in unmarked mRNA populations equal to the mRNA transcription rate λ.

With a classification of when $\boldsymbol{U}$ is invertible we now consider a particular method for matrix inversion that can make use of the closed-form of the determinant. Namely, the method considered here is the determinant-adjoint approach given by,

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

Thus inverting $\boldsymbol{U}$ now falls onto the task of simplifying the product of $Adj[\boldsymbol{U}]\vec{b}$. Owing in part to the tri-diagonal form of $\boldsymbol{U}$ and to the fact that $\vec{b}$ is only non-zero in the first component, an expression for $\vec{m}$ was able to be determined as follows,

$$\vec{m} = -\frac{1}{\det[\boldsymbol{U}]}\lambda \begin{vmatrix} \det[\boldsymbol{U}_{i_{\max}-1}] \\ -s_1 \det[\boldsymbol{U}_{i_{\max}-2}] \\ s_1 s_2 \det[\boldsymbol{U}_{i_{\max}-3}] \\ \vdots \\ (-1)^i (\prod_{j=1}^{i} s_j) \det[\boldsymbol{U}_{i_{\max}-(i+1)}] \\ \vdots \\ (-1)^{i_{\max}-2} (\prod_{j=1}^{i_{\max}-2} s_j) \det[\boldsymbol{U}_1] \\ (-1)^{i_{\max}-1} (\prod_{j=1}^{i_{\max}-1} s_j) \det[\boldsymbol{U}_0] \end{vmatrix} \tag{13}$$

Along with the convention that $\det[\boldsymbol{U}_0] = 1$, we also define $\boldsymbol{U}_i$ as $i \times i$ lower right sub-matrix, or minor, of $\boldsymbol{U}$ and s_i as the sub-diagonal entry of the i^{th} of entry of $\boldsymbol{U}$, where $s_i = \kappa(i-1) = \kappa_0 \left(1 - \frac{i-1}{i_{\max}}\right)$. Equation 13

mikeg: 10/22/1
note $s_i = \kappa(i-1)$ s
we should be able t
simplify notation b
replacing s with κ

provides a solution to the equilibrium distributions in the unmarked sub-system in terms of the determinants of $\boldsymbol{U}$ and each of its lower-right sub-matrices, $\boldsymbol{U}_i$. This result points to the determination of a closed-form for the equilibrium distributions as being linked to a determination of an efficient way to generate the sub-matrix determinants. However, at this point a closed-form for the sub-matrix determinants has yet to be found, encouraging further research into efficient methods to determine equilibrium distributions.

In order to move forward solving the system using Equation 13 above, all the sub-determinants are needed. Another property of tri-diagonal matrices is that their determinants can be calculated by solving a 2^{nd} degree partial recurrence equation. This means two things, the first is that if a closed form solution existed for that recurrence relation, it could be used to calculate all the determinants and sub-determinants of the system. Additionally, it specifies that one can work forward iteratively from $\det[\boldsymbol{U}_1]$ in order to generate all $\det[\boldsymbol{U}_i]$. Specifically, taking a tri-diagonal matrix written in the general form,

$$\boldsymbol{A} = \begin{pmatrix} d_0 & p_0 & 0 & 0 & 0 \\ s_1 & d_1 & p_1 & 0 & 0 \\ 0 & s_2 & d_2 & p_2 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & s_{n-1} & d_{n-1} & p_{n-1} \\ 0 & 0 & 0 & s_n & d_n \end{pmatrix}.$$

The determinants of the $\boldsymbol{A}_1, ..., \boldsymbol{A}_{n-1}$ sub-matrices can be solved by the following recurrence relation,

$$\det[\boldsymbol{A}_0]\, i = d_{n-i} \cdot \det[\boldsymbol{A}_0]\, i-1 - p_{n-i} \cdot s_{n-(i-1)} \cdot \det[\boldsymbol{A}_0]\, i-2, \tag{14}$$

with the initial conditions,

$$\det[\boldsymbol{A}_0]\, 1 = d_n \text{ and } \det[\boldsymbol{A}_0]\, 2 = d_{n-1} \cdot a(n,1) - p_{n-1} s_n. \tag{15}$$

In the case of the matrix $\boldsymbol{U}$, the Equations 14 along with the initial conditions in 15 can be specified with the following equations,

$$d_i = -\kappa_0 \left(1 - \tfrac{i}{i_{\max}} + \mu_0 + i\tau_0\right)$$

$$p_i = (i+1)\tau_0$$

$$s_i = \kappa_0 \left(1 - \tfrac{i-1}{i_{\max}}\right)$$

This iterative scheme provides one approach to finding the sub-matrix determinants, $\det[A_{i_{\max}-1}]$, $\det[A_{i_{\max}-2}]$, ... needed to determine the equilibrium solutions given in Equation 13.

4.2. Steady-state in ODE full system

The Full ODE system was described in matrix-vector form in Equation 3. Recall that $\boldsymbol{F}$ has a block matrix form given in Figure 4, where the upper-left matrix is tri-diagonal, the upper-right block is the zero matrix, the lower-left matrix is diagonal, and the lower-right block is upper-triangular. Utilizing the formula found for the determinant of $\boldsymbol{U}$, a closed form for the determinant of $\boldsymbol{F}$ was determined. This formula is given by,

$$\det[\boldsymbol{F}] = -\delta\lambda\tau_0^{i_{\max}} i_{\max}! \left(\prod_{i=0}^{i_{\max}} \mu_0 + i \left(\tau_0 + \frac{\kappa_0}{i_{\max}} \right) \right). \tag{16}$$

At equilibrium, the equations for the full system become,

$$\vec{M} = -\vec{B}\boldsymbol{F}^{-1}. \tag{17}$$

If one uses the determinant-adjoint method to invert $\boldsymbol{F}$ then the form of $\vec{B}$, being λ in the first component and 0's else, entails that only the co-factors for the first row of $\boldsymbol{F}$ need to be determined. This eases the computation of the adjoint matrix for $\boldsymbol{F}$, and once again points to the determinant-adjoint inversion as a method for determining equilibrium solutions. Unlike the previous solution methods, this method solves the for unmarked and marked classes simultaneously.

We have now presented multiple methods for solving the ODE system at equilibrium. Advantages and drawbacks for each of the methods are addressed in the discussion section later in this paper. We now turn our attention to describing a method where the discrete unmarked and marked polysomal classes are approximated as continuous and a system of PDE are used to solve for equilibrium solutions.

5. Partial differential equation approximation using continuous polysomal classes

The derivation of the system of partial differential equations utilizes the observation that the unmarked system can be viewed as a *non-linear birth and death process. Births* in this case are initiation events and *deaths* are translation completion events. The benefit of this observation about the unmarked class is that non-linear birth and death processes are a well-studied type of system and previous research can be leveraged here. In addition to simply providing another view of the system, the continuous polysomal classes assumed in the PDE derivation allow one to more easily handle high levels of $i_{\max}$.

5.1. The unmarked class PDE approximation

The rate of initiation and translation completion in our model is controlled by two functions, $\kappa(i)$ and $\tau(i)$, respectively. From a non-linear birth and death process perspective, these two functions control births

and deaths in our system, with exception of births into the m_0 unmarked non-polysomal class which is controlled by the parameter λ. Each birth represents a transition in the population of an mRNA in a given polysomal state to the next highest state and each death transitions a given mRNA from its current state to the next lower state. Following the derivation from Karev et al., (CITE) a continuous approximation for the unmarked class can be given as follows:

$$\frac{\partial m(i,t)}{\partial t} = -m(i,t)\mu(i) - \frac{\partial}{\partial i}\left(m(i,t)(\kappa(i) - \tau(i)\right) + \frac{1}{2}\frac{\partial^2}{\partial i^2}\left(m(i,t)(\kappa(i) + \tau(i))\right), \tag{18}$$

where the quantity $\kappa(i) - \tau(i)$ is considered the *drift coefficient* and $\kappa(i) + \tau(i)$ the *diffusion coefficient*. To distinguish between the discrete ODE system and the continuous PDE system, we use continue to use the subscript notation of m_i and m_i^* for the unmarked and marked ODE classes and now use $m(i,t)$ and $m^*(i,t)$ for the unmarked and marked classes in the PDE system.

Assuming the functional forms for $\kappa(i)$ and $\tau(i)$ as provided in Equations 1 and 2 and assuming equilibrium, $\frac{\partial m(i,t)}{\partial t} = 0$, Equation 18 becomes,

$$0 = \left(\frac{\kappa_0}{i_{\max}} - \mu_0 + \tau_0\right) m(i) + \left(\frac{(i-1)\kappa_0}{i_{\max}} - \kappa_0 + (i+1)\tau_0\right) m'(i) + \frac{1}{2}\left(\kappa_0 + i\tau_0 - \frac{i\kappa_0}{i_{\max}}\right) m''(i). \tag{19}$$

5.1.1. The Marked Class PDE approximation

The dynamics for the marked class can be modeled by the following nonlinear inhomogenous first-order PDE,

$$\frac{\partial m^*(i)}{\partial t} + \frac{\partial}{\partial i}\left(-\tau(i)m^*(i)\right) = \mu(i)m(i,t) \tag{20}$$

In practice, just as in the ODE system, we utilize the exact quantities in the marked classes as a function of the unmarked class abundances given by Equation 8. As such, Equation 20 is presented only to provide a complete full-system PDE approximation to the system of ODE and is not discussed further.

5.1.2. Unmarked PDE Boundary Conditions

At equilibrium, the unmarked PDE reduces to 2nd order non-linear ODE, requiring two boundary conditions for a specific solution. Two approaches were investigated for the upper boundary condition. The first is an approximation that $m(i_{\max}) = 0$ meaning that the $(i_{\max})^{th}$ unmarked class has no members at equilibrium. This boundary condition is based on the assumption that the likelihood of any mRNA being completed saturated with ribosomes is nearly zero. The second approach for the upper boundary condition utilizes a result from the ODE system. Namely, in the unmarked ODE system a closed form for $m(i_{\max})$ was determined, so Equation 22 can be used to provide an exact upper boundary for the unmarked PDE.

For the lower boundary condition, setting it exactly means knowing the value of the 0^{th} unmarked class;

unlike the $m_{i_{\max}}$ class, no simple closed form solution was able to be determined. In the absence of the closed form for m_0 an alternate approach needed to be determined. Utilizing the analytical solvers in *Mathematica*, an arbitrary initial value, $m(0) = m0$ was used and solutions were studied. From this, it was determined that $m(0)$ served only to scale the solutions in the PDE. Using this and borrowing the following observation from the discrete system,

$$\frac{\lambda}{\mu} = \sum_{i=1}^{i_{\max}} m_i, \tag{21}$$

an efficient method for supplying the appropriate scaling through the boundary condition was deduced. We originally solve the PDE for the unmarked class at equilibrium with $m(0) = 1$ and $m(i_max) = 0$ as boundary conditions, call this equilibrium solution $m_{pre}(i)$. We then let $m_{sum} = \sum_{i=0}^{n} m(i)_{pre}$, and finally, using the relationship in Equation 30, let $m0 = \left(\frac{\lambda}{\mu_0}\right) \frac{1}{m_{sum}}$, which is now the correct, unscaled lower boundary condition. At this point one can either use this value and resolve the unmarked PDE at equilibrium with the proper boundary conditions $m(0) = m0$ and $m(i_{\max}) = 0$ or, given that $m0$ is just a scaling factor, one can simply multiply the solution $m_{pre}(i)$ by $m0$ to obtain the solution the unscaled equilibrium values for the unmarked classes.

Another option to write the boundary condition for the lower boundary that is slightly more straightforward is to write $m(0)$ explicitly in terms of the sum, $m(0) = \left(\frac{\lambda}{\mu}\right) \frac{1}{\sum_{i=0}^{i_{\max}} m(i)}$. Use of this form for the boundary condition was not pursued since the results of the previously outlined method were found to be sufficient.

5.2. Nondimensionalization

In order to aid in parameter estimation and computation, a nondimensionalization was conducted for the PDE system. Investigation of a variety of parameters and parameter combinations to nondimensionalize the system with was undertaken. Although multiple options were present, the parameter τ_0 was chosen as the value to scale time, owing in part to its physical identifiability. The resulting parameters and notation for the nondimensionalized system are provided in Table 2.

Nondimensionalized Unmarked PDE: Similar to the original PDE, at equilibrium we solve the following nondimensionalized, 2nd order, non-linear ODE to describe the distribution within the Unmarked classes:

$$0 = \frac{1}{2n}(2(\bar{\kappa}_0 + n - \bar{\mu}_0 n) \cdot \bar{m}(i) + 2(\bar{\kappa}_0(-1 + i - n) + (1 + i)n)\bar{m}'(i) + (in + \bar{\kappa}_0(-i + n))\bar{m}''(i))$$

Nondimensionalized Marked PDE: Once again, the marked PDE was not used in application, however it is once again presented for completeness. At equilibrium we have the following nondimensionalized, 1st

Table 2: Nondimensionalized state variable and parameter definitions using the translation completion parameter, τ_0, for scaling the system time-scale.

Nondimensionalized variables and parameter definitions	
$\bar{m}(i)$	Dimensionless unmarked state variable
$\bar{m}^*(i)$	Dimensionless marked state variable
$\bar{t} = \frac{1}{\tau_0}$	New time unit scaled by τ_0
$\bar{\kappa}_0 = \frac{\kappa_0}{\tau_0}$	Dimensionless parameter associated to translation initiation rate
$\bar{\mu}_0 = \frac{\mu_0}{\tau_0}$	Dimensionless parameter associated to marking rate
$\bar{\lambda} = \frac{\lambda}{\tau_0}$	Dimensionless parameter associated to mRNA input rate into class $m_(0)$ in the system
$\bar{\delta} = \frac{\delta}{\tau_0}$	Dimensionless parameter associated to the loss of marked mRNA from class $\bar{m}^*(0)$ out of the system

order, ODE to describe the distribution within marked classes:

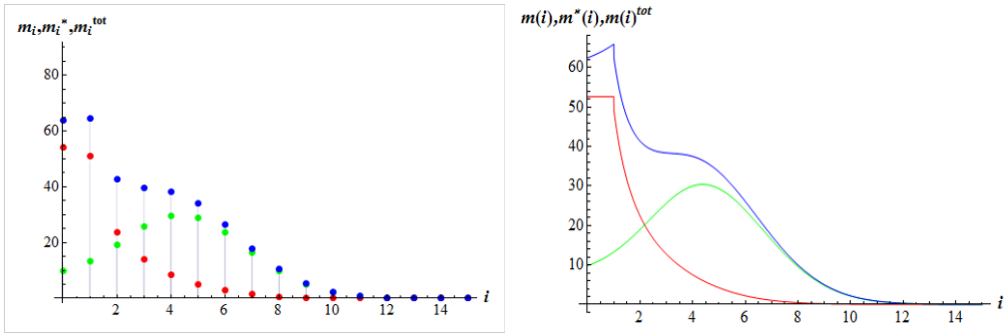
$$0 = -\bar{\mu}_0 \bar{m}(i) - \bar{m}^*(i) - i\bar{m}^{*\prime}(i)$$

With the standard and nondimensionalized PDE system defined we now present results that compare solution methods and show concordance among the varying approaches for determining equilibrium solutions.

6. Comparison of Solution Methods

This section compares the exact ODE solutions at equilibrium to the approximate PDE solutions at equilibrium. Although both are based on the same mechanistic underpinning

Figure 2: Model results for parameters: $i_{\max} = 30, \lambda = 5.4, \kappa 0 = 0.66, \tau 0 = 0.1, \mu 0 = 0.0292, \delta = 0.1$



(a) Discrete model solutions

(b) Continuous approximation model solutions

Although solutions for only one set of parameter values is shown, the same behavior has been observed for all sets of parameters simulated. Additionally, since the two solution were derived using different methods, they serve to validate each other. Possessing two different methods to generate equilibrium distributions for the polysomal classes is advantageous in that it provides flexibility to researchers for model fitting and parameter estimation when investigating empirically derived data.

Additional Notes

Getting Rid of GCD function

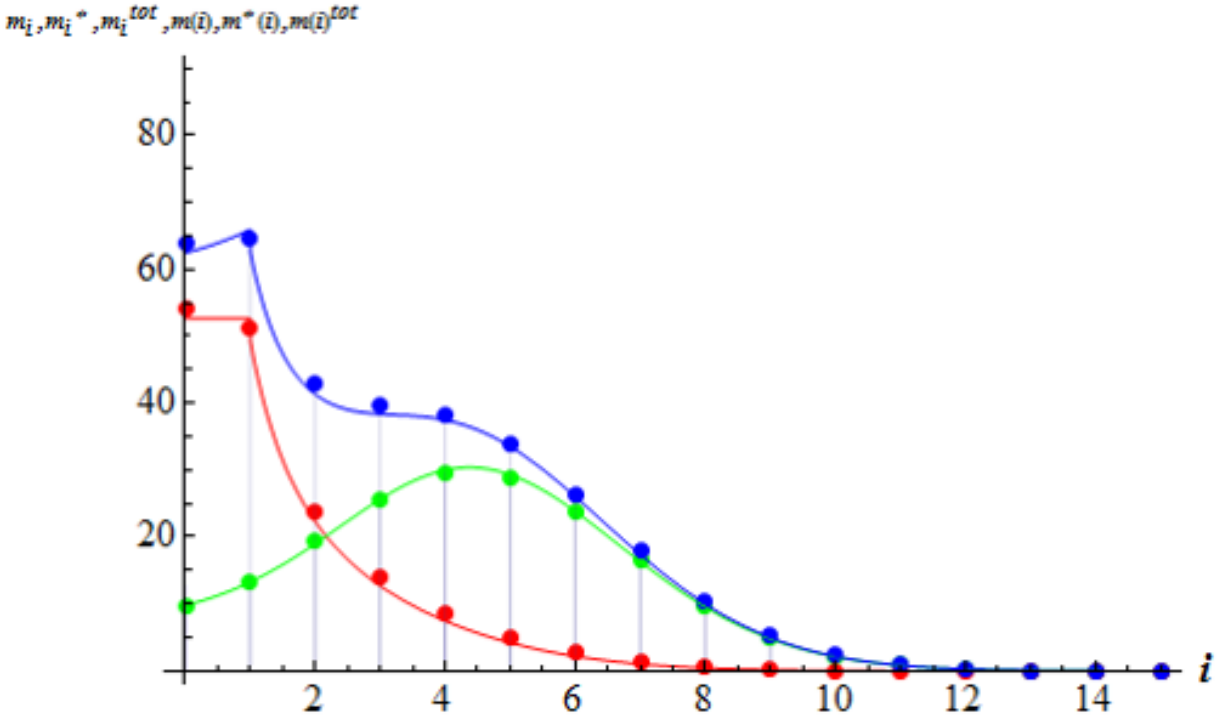
If equation

$$m_{i_{max}}^{eq} = \frac{\prod_{i=1}^{i_{max}} \left(\frac{i}{GCD[i_{max}, i]} \kappa_0 \right)}{\prod_{i=1}^{i_{max}} \left(\frac{i_{max}}{GCD[i_{max}, i]} (\mu_0 + i\tau_0) + \frac{i}{GCD[i_{max}, i]} \kappa_0 \right)}, \tag{22}$$

in previous versions of the write up is correct, then it seems we've missed a very simplification that would simplify our calculations *immensely*. Let's imagine three generic functions of i, a_i, b_i, and c_i such that

$$F = \frac{\prod_i (a_i b_i)}{\prod_i (a_i c_i + a_i b_i)} \tag{23}$$

Figure 3: Comparison of discrete and continuous approximation model Solutions for Parameters: $i_{\max} = 30, \lambda = 5.4, \kappa 0 = 0.66, \tau 0 = 0.1, \mu 0 = 0.0292, \delta = 0.1$



Based on my understanding of mathematics,

$$F = \frac{\prod_i a_i \prod_i b_i}{\prod_i (a_i (c_i + b_i))} \tag{24}$$

$$= \frac{\prod_i a_i \prod_i b_i}{\prod_i a_i \prod_i (c_i + b_i)} \tag{25}$$

$$= \frac{\prod_i b_i}{\prod_i (c_i + b_i)} \tag{26}$$

$$\tag{27}$$

If this is correct and we define $a_i = 1/\text{GCD}[i_{\max}, i]$ and $b_i \kappa_0$ and $c_i(\mu_0 + i\tau_0)$, then the equation 22 can be rewritten as,

$$\hat{m}_{i_{\max}} = \frac{prod_{i=1}^{i_{\max}} i \frac{\kappa_0}{i_{\max}}}{prod_{i=1}^{i_{\max}} \mu_0 + i \left(\tau_0 + \frac{\kappa_0}{i_{\max}}\right)} \tag{28}$$

If this is the case, then I'm both elated and embarrassed as hell we didn't see it earlier.

7. Results

1. The unmarked system can be split into two components: Total transcript count and the distribution of transcripts.

(a) The unmarked class transcript number is determined by λ, μ

(b) In the unmarked class the total transcript abundance is determined by a similar relationship as Equation 21.

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

(c) 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.

(d) μ affects both the total transcript abundance and the distribution of ribosomal classes across a particular species of transcript.

(e) First μ controls the rate of outflow from unmarked unto marked, and second it shifts mRNAs to lower ribosomal classes.

(f) The solution to the system, as presented previously, can be expressed in the determinant-adjoint form:

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

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

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

and

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

(h) 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.

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

(i) 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.

(j) This formulation has three interesting properties

i. First it gives a determinant free solution to our system. Now, to obatin a full solution of the unmarked solution to our model we only need the first column of the Adjugate matrix.

ii. 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.

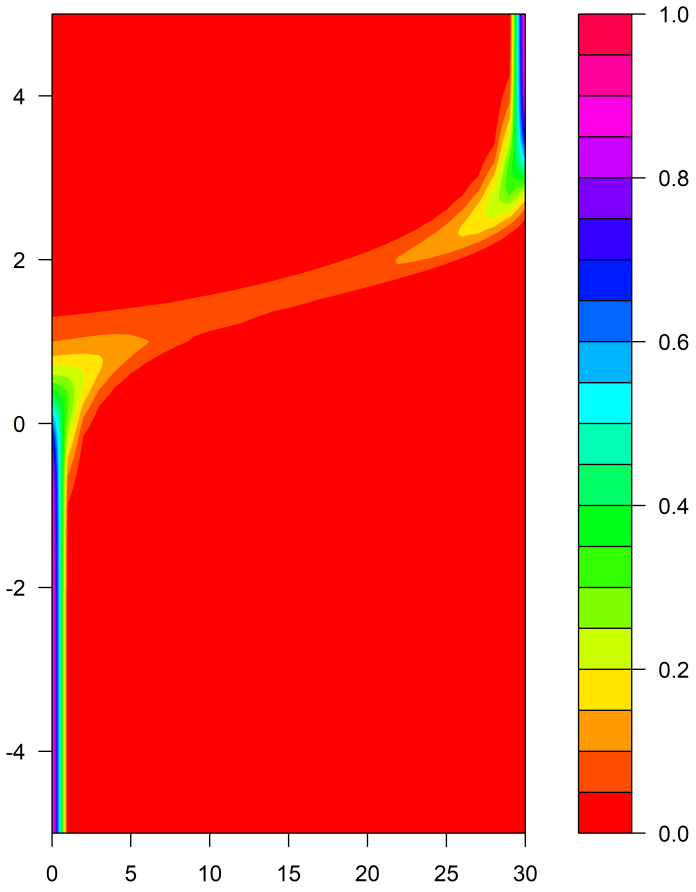
iii. Third, it permits analysis of the underlying transcript distribution even under conditions where the model has no solution. For example, when $\mu = 0$, both solutions are indeterminate. However, the determinant free solution allows for us to explore what the transcript distribution would be when μ=0.

2. Ribosomal distribution on the marked class is determined by the ratio of κ/τ and μ

   (a) Figure 4 shows a representative contour plot of model solutions for a broad range of κ values.

   (b) For $\kappa < 0.1$ there is no ribosome loading, while loading maxes out beyond 10^4, with a smooth transition between boundaries.

Figure 4: Contour plot of unmarked class over a range of kappa values: $\kappa = 10^-5 - 10^5.i_{\max} = 30, \lambda = 1, \tau 0 = 1, \mu 0 = 0.1, \delta = 100,000$. Color scale is percent ribomose load for each kappa value across all 31 ribosome classes.



   (c) In order to better interpret the model we will focus on parameter ranges dervied from literature.

i. To approximate μ transcript half-lives calculated by Sorenson (2018) in cordycepin treated *Arabidopsis thaliana* seedlings. where μ is:

$$\mu = \frac{ln(2)}{t_{1/2}} \tag{30}$$

Where $t_{1/2}$ is the half-life. Model solutions were produced for a range of half-lives: 1, 5, 25, 50, 75, 95 and 99 percentile values of the half-lives found in Sorenson (2018)

ii. The resulting range of μ is from $10^{-}3to10^{-}6$.

iii. The range for τ is based from (insert citations here) and ranges from 4-20 aa/s depending on the organism. A plausible biological range of 0.1-20 aa/s was explored.

iv. The

v. As λ is only a scaling factor and does not affect the distribution of the ribosomes it was set to 1.

3. The marked class depends on the distribution of the unmarked class, τ, μ and δ

   (a) The marked subsystem can be further divided into the m_0^* class and the m_i^* classes for $i > 0$.

   (b) the m_0^* class in the current model design represents only two forms of degradation, free mRNA degradation and the degradation of the remanents of XRN1 mediated degradation.

   (c) δ only determines the accumulation of transcripts in the m_0^* class. It is reasonable to explore the model with large δ since marked transcripts are translationally incompetent.

Data Sources

1. Protein Length data

   (a) Saccharomyces cerevisiae

   (b) Arabidopsis thaliana

   (c) Homo Sapiens

2. mRNA half life

   (a) Saccharomyces cerevisiae

   (b) Arabidopsis thaliana

   (c) Homo Sapiens

3. Defining parameter space

(a) Saccharomyces cerevisiae

(b) note on At and Hs

Main results:

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

5. When not not bound by ribosomes, mRNA degradation relies on removing protective and translation enhancing componenets of the mRNA. These include the 5' mG cap and the 3' polyadenosine tail.

6. Additionally mRNA can be started through endonucleolytic cleavage promoted by siRNAs.

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. In the plant model or

13. Model predicts mRNA distribution across ribosomal classes

    (a) Density independent initiation and density independent initiation models

    (b) The model predicts the total mRNA quantity in both marked and unmarked states as well as the mRNA's distribution across different ribosomal classes in each state.

    (c) Figure X Left,

14. The model provides an upper bound on translational efficiency

    (a) 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.

    (b) 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.

    (c) To summarize their joint behavior, we chose to focus on the elongation scaled initiation rate (κ τ) as this provides a single unique parameter for each model response.

    (d) 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.

    (e) The degradation rate (δ) only controls the transcript population in the marked zeroth class. To reduce interference from this we set a very high degradation rate.

    (f) 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.

    (g) Introduce scaled initiation rate range. Discuss upper and lower limits vis a vis the model output.

    (h) *** 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 arguement 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.

15. Marking rates have only a mild effect on ribosomal load, but a large effect on protein production.
    (a) From the solution of the unmarked system we can define two separate effects of transcript marking.
        i. First, marking determines how many transcripts can accummulate in the unmarked class, thus determining the total transcript population
        ii. Second, marking rate shifts the distribution of transcripts in the unmarked class towards lower ribosomal classes.
    (b) Marking only has a mild effect on the distribution of transcripts on the unmarked class
16. Under equilibrium protein length affects ribosomal load but not ribosomal density
    (a) stuff

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 intedpendent 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 ($\kappa \quad \tau$) 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 e ffect of ribosomal interference

Figure 5

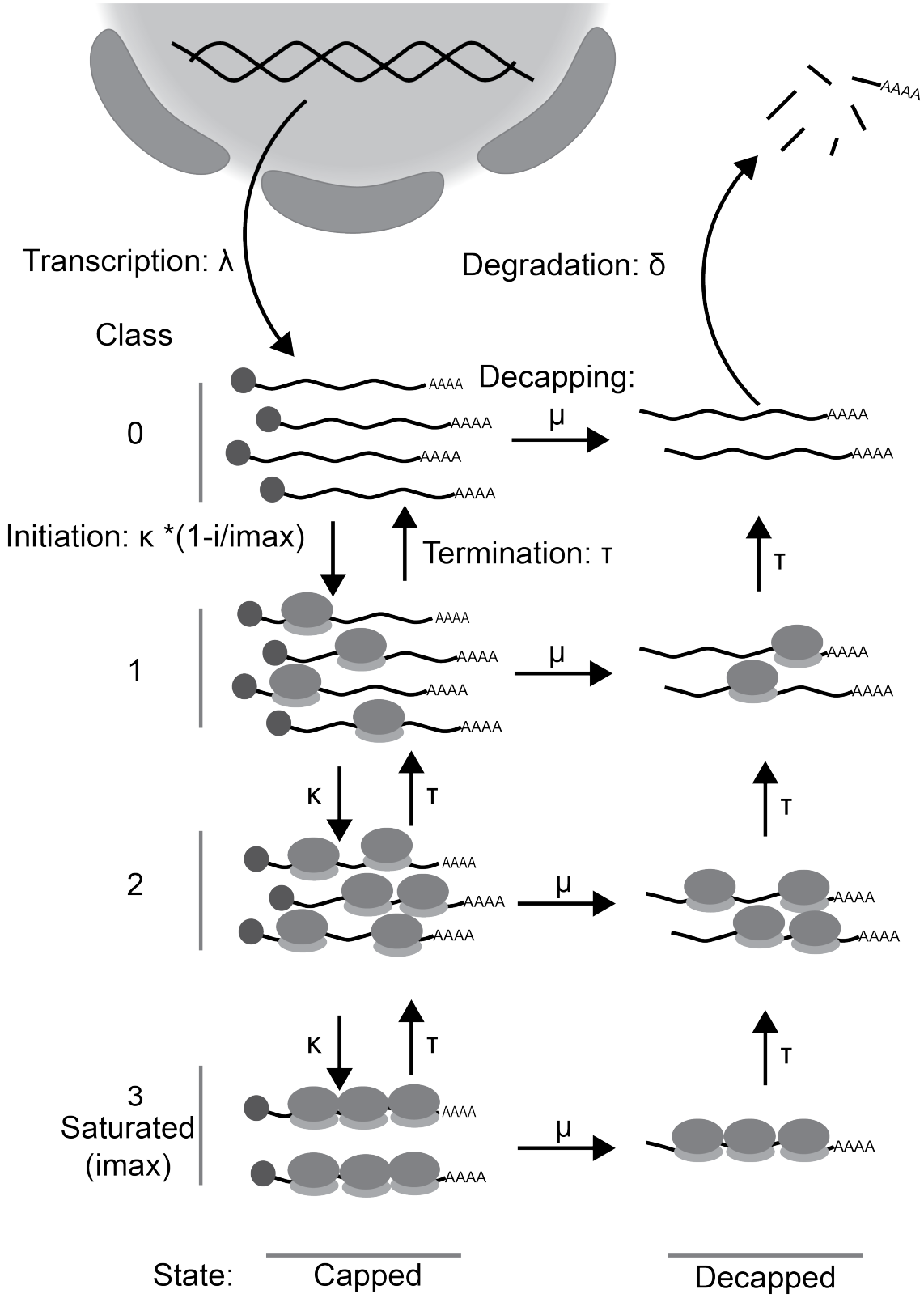

Transcription: λ
Degradation: δ
Class
Decapping:
μ
0
AAAA
Initiation: κ *(1-i/imax)
Termination: τ
τ
1
κ
2
3
Saturated
(imax)
State:
Capped
Decapped
Pr(mRNA)
0.3
0.2
0.1
Lorem ipsum

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

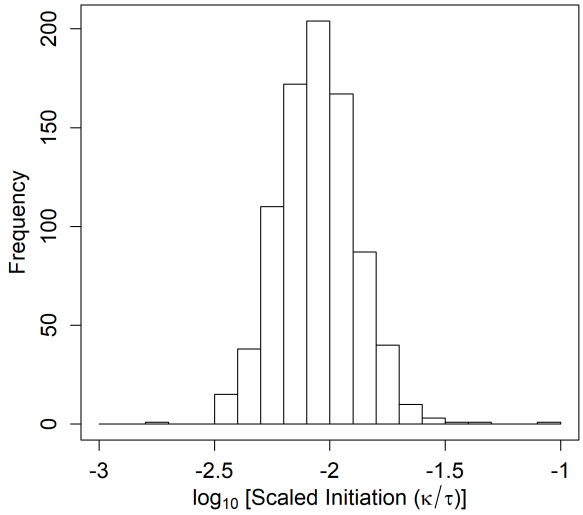


Figure 7: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

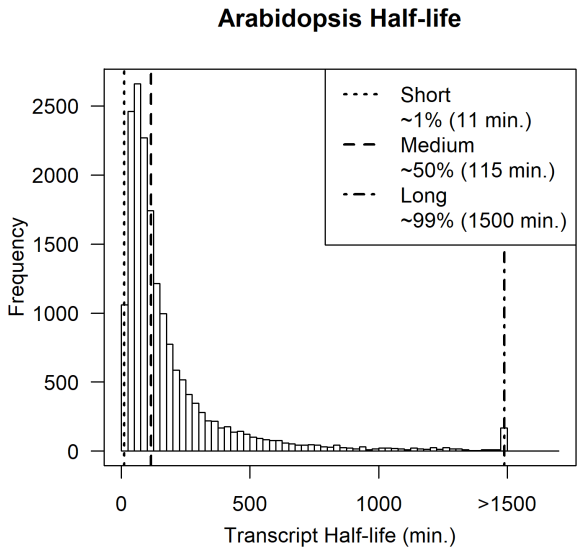


Figure 8: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

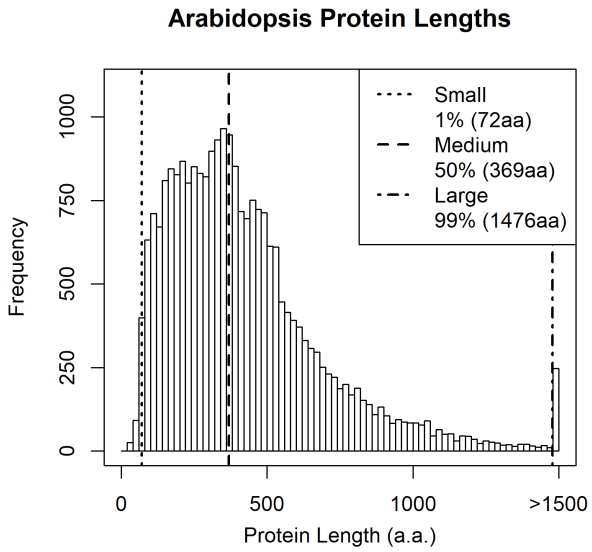


Figure 9: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

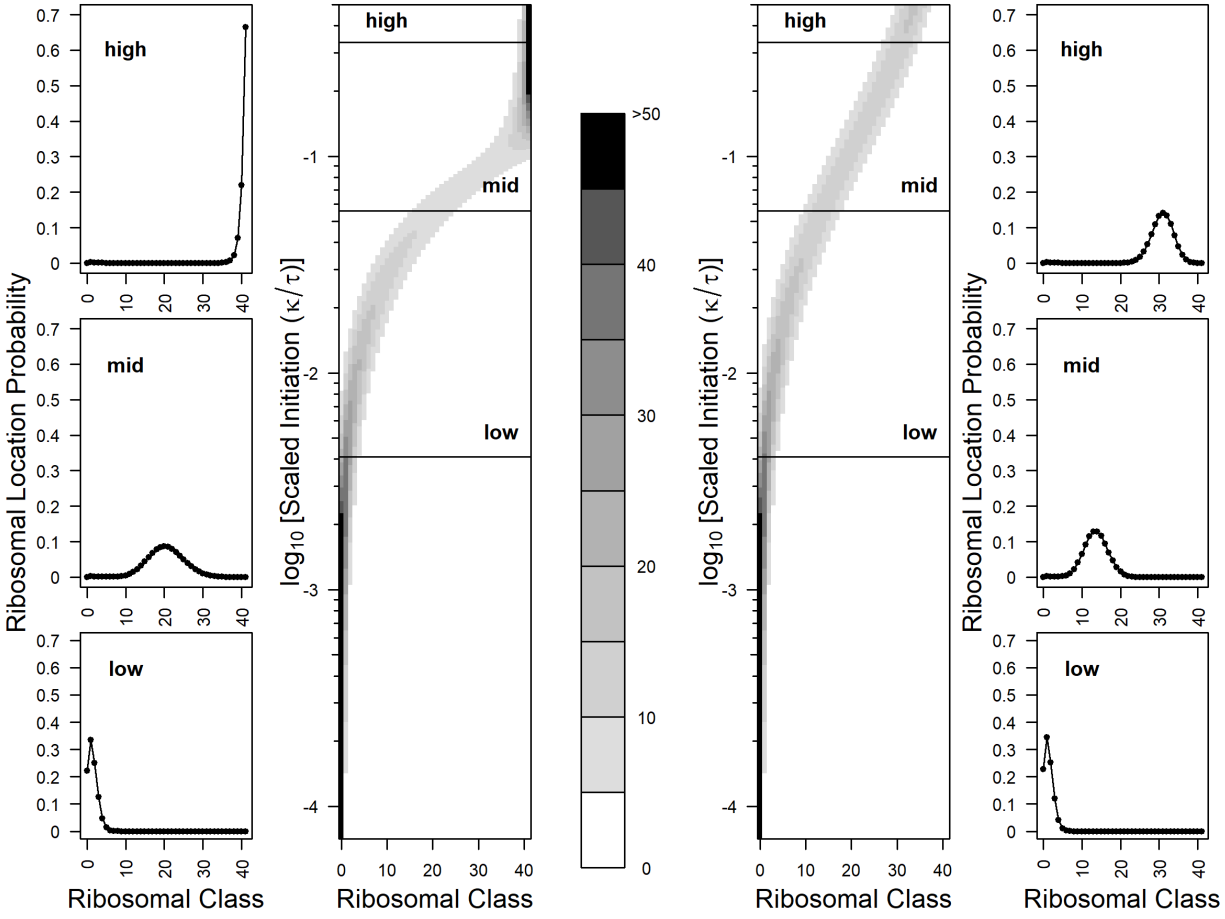


Figure 10: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

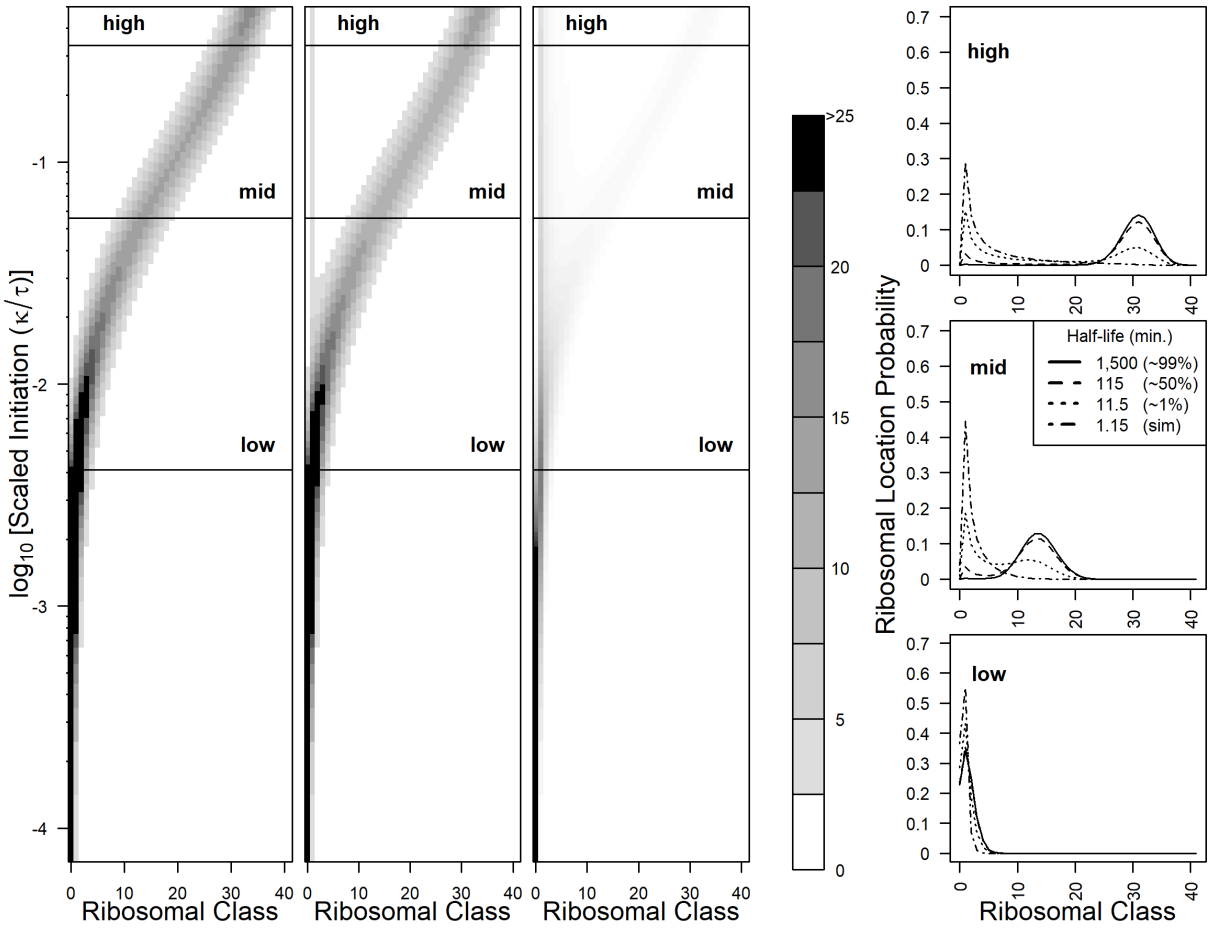


Figure 11: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

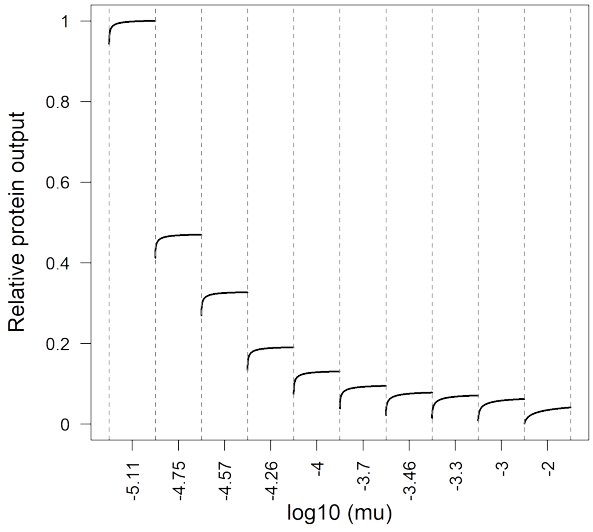


Figure 12: Comparison of density independent initiation (DII) and density dependent initiation (DDI)

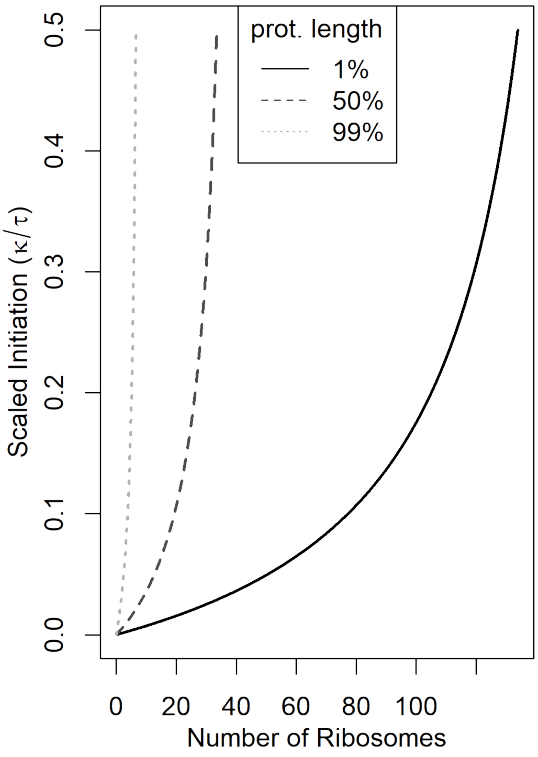


Figure 13: Comparison of density independent initiation (DII) and density dependent initiation (DDI)