THE UNIVERSITY *of* EDINBURGH
School of Physics
and Astronomy

SENIOR HONOURS PROJECT

UNIVERSITY OF EDINBURGH

School of Physics and Astronomy
Peter Guthrie Tait Road, Edinburgh EH9 3FD, United Kingdom
April 3, 2020

Investigating the Non-Steady State Ribosomal Dynamics in mRNA Translation

---

Author:
C. B. Abbott

Supervisor:
Dr. J. Szavits-Nossan

Declaration

I declare that this project and report is my own work.

Signature: C. Abbott Date: 03.04.2020

Abstract

The majority of current work conducted on the kinetic model of messenger RNA (mRNA) translation assumes a steady state regime. Here, we utilise the Gillespie algorithm, in tandem with the inhomogeneous totally asymmetric exclusion process (TASEP), in order to model the ribosomal dynamics on a newly produced mRNA strand in the non-steady state. Our aim is to elucidate whether the steady state assumption remains valid whilst still appreciating the finite lifetime of the mRNA strand, as it is possible that the process of degradation may already be under-way prior to the steady state being attained. Hence, due to the relative instability of E. coli mRNA, this investigation simulates an E. coli population grown in an M9 medium with a growth rate of 1.6 h^{-1}. Statistical analysis of 103 independent simulations of unique E. coli genes revealed that mRNA degradation largely operates over a larger timescale than mRNA translation, hence rendering the steady state assumption valid when modelling translation for the vast majority of genes.

Contents

1 Introduction

Perhaps the most crucial aspect in the production of biological proteins is the process of translation. In this context, translation refers to the reading of a messenger RNA (mRNA) molecule through the agency of molecular motors known as ribosomes. The mRNA strand consists of a sequence of codons[†], where each codon codes for a specific amino acid. All of which is conducted according to the universal genetic code [1]. Transfer RNA (tRNA) molecules are then responsible for the retrieval and incorporation of the amino acid into the overall polypeptide chain (protein). A holistic portrayal of mRNA translation has been presented in Figure 1, outlining the three primary phases in protein production: initiation, elongation and termination. We are motivated in simulating the dynamics of this process as identifying factors influencing the rate of protein production not only remains a fundamental open question in the field of molecular biology, but also has the potential to unlock numerous synthetic applications [3, 4].

mRNA translation is process that presents itself to be mathematically modelled as a one-dimensional (1D) driven lattice gas [5], where ribosomes translate along the mRNA strand codon by codon. This type of model is understood exceptionally well, and has even been analytically solved for the case of homogeneous elongation rates [6]. However, the rate of translation has been experimentally shown to depend on the underlying sequence of codons on the mRNA strand [3, 7, 8]. Hence, ribosome elongation rates can not be made homogeneous if we wish to investigate a system of greater biological relevance. Furthermore, it is also paramount to appreciate the finite lifetime of the mRNA strand [9, 10]. In work conducted by Esquerré *et. al*, experimental observations revealed that over a wide range of *Escherichia coli* (E. coli) growth rates, $\sim$ 90% of mRNA half-lives were $<$ 11 minutes [11, 12]. In order to appreciate this statistic, the median half-life for a human mRNA strand is widely accepted to be $\sim$ 10 hours [13]. The relative instability of E. coli mRNA therefore implies that degradation of the mRNA strand may have already commenced before the steady-state regime was attained. This notion is significant as the majority of current work concerning the kinetic model of translation often assumes a steady-state[5, 14–16]. Consequently, in this paper, dynamic Monte Carlo simulations were implemented in order to examine the stochastic movement of ribosomes along a newly produced mRNA strand; a system readily characterised to be in a non-steady state. Specifically, this investigation was conducted to elucidate the type of relationship present between the processes of translation and degradation of the mRNA strand. E. coli genes were purposefully chosen to be modelled due to the relative instability of their corresponding mRNA strands as quoted previously. This enabled effective examination of the commonly employed steady state assumption, as well as offered an environment where translation and degradation would most likely intertwine.

Our model was tested using codon specific elongation rates for an E. coli population with a growth rate of 1.6 h^{-1} [17]. Initiation rates [18], as well as the half-lives [19] for specific E. coli genes were also obtained courtesy of Gorochowski *et. al*, and Bernstein *et. al* respectively. Whilst the data utilised is explicitly for an E. coli culture grown in an M9 medium with the LacZ

[†] A codon is a triplet of nucleotides; the organic compounds constituting molecules such as DNA and RNA [2].

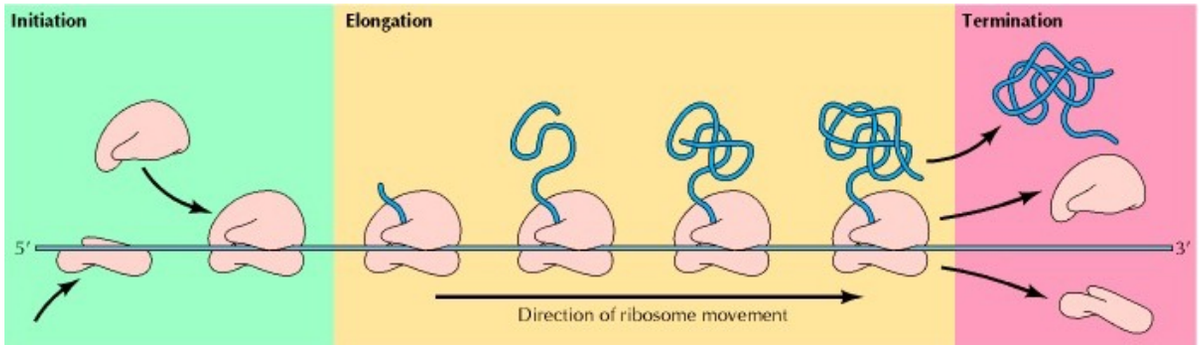


Figure 1: Initiation (left): Ribosome binds to the mRNA at the start codon. Elongation (centre): polypeptide chain (protein) begins to elongate as more amino acids are incorporated. Termination (right): Upon reaching the stop codon, the ribosome dissociates from the strand releasing the protein [1].

gene present, this has negligible consequences regarding the dynamics simulated and thus will not be discussed further.

2 Background

2.1 mRNA Translation Theory

2.1.1 Initiation

Upon the formation of a new mRNA strand, ribosomes attempt to initiate on to the molecule in order to begin the synthesis of a new protein. The position on the strand a ribosome chooses to initiate at is determined by specific start codons. In the case of E. coli, these prokaryotic[†] micro-organisms have a total of 4288 genes, 83% of which use AUG as a start codon (3542/4288), 14% (612) use GUG, whilst only 3% (103) use UUG [20]. It is paramount to establish that a single E. coli mRNA strand has the capacity to encode for multiple polypeptides. This capacity stems from the overall codon sequence of the mRNA strand possessing multiple start codons. Hence, this grants the ribosomes present in prokaryotic micro-organisms, such as E.coli, the ability to independently translate multiple polypeptides simultaneously [1].

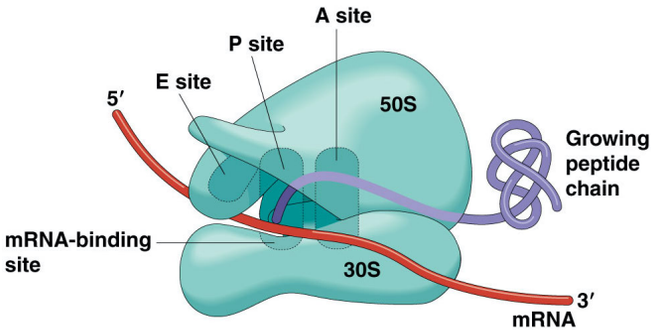


Figure 2: A schematic image presenting the internal structure of a ribosome. The 5′ and 3′ ends of the mRNA strand are also visible, indicating the effective start, and end of the mRNA strand respectively. The most vital elements depicted in the figure are the tRNA binding sites, designated the A (aminoacyl), P (peptidyl), and E (exit) sites [1, 21].

During the process of initiation, the amino-acid methionine, usually encoded by AUG, will be retrieved and incorporated by the methionyl tRNA, representing the inception of a new polypeptide chain. This remains the case even when GUG, or UUG, are the start codons [1]. The ribosome is now present and bound to the mRNA strand, ready to begin the next phase of translation.

2.1.2 Elongation and Termination

Following successful initiation of the ribosome onto the mRNA strand, translation proceeds via the process of elongation. Before tackling the mechanics of elongation however, we must first inaugurate an understanding of the internal structure of a ribosome. Figure 2 (below) presents a depiction of a ribosome bound to an mRNA strand, along with its growing polypeptide chain. This figure in particular highlights the ribosome's internal tRNA binding sites, A (aminoacyl), P (peptidyl), and E (exit), which play a vital role in the process of elongation. Once initiated, the ribosome's A site now sits over the codon proceeding the start codon, with the initiator methionyl tRNA bound at the trailing P site. Whilst abiding by the universal genetic code, the A site codon is then 'read' by the ribosome provoking an aminoacyl tRNA to become bound at the formerly free A site. A peptide bond is then formed between the occupants of the A and P sites representing the elongation of the polypeptide chain, now two amino acids in length. Following this process, the methionine associated with the methionyl tRNA, currently occupying the ribosomal P site, is then transferred over to the aminoacyl tRNA at the ribosomal A site. This results in the formation of a peptidyl tRNA at the A site, whilst leaving an uncharged initiator tRNA (formerly possessing the methionine) at the P site. Finally, the ribosome hops to the next codon transferring the peptidyl tRNA to the P site, as well as the uncharged tRNA to the E site. Subject to the 'reading' of this new codon site by the ribosome, as well as the binding of a new aminoacyl tRNA at the ribosomal A site, the uncharged tRNA is released from the E site accordingly. This allows the process of elongation to repeat thus, inducing the growth of the polypeptide chain. Ultimately, the ribosome will trans-locate, or hop, until its A site lies over one of the following stop codons UAA, UAG, or UGA [20]. This activates the termination phase of translation, provoking the release of the complete polypeptide chain from the ribosome, as well as the dissociation of the tRNA molecules and ribosome from the mRNA strand [1].

[†]Prokaryotes are unicellular micro-organisms which lack a nuclear membrane [2]

2.2 mRNA Degradation Theory

2.2.1 The Importance of Degradation

In addition to possessing a firm grasp on the dynamics of mRNA translation, it is also imperative to understand the underlying mechanisms conducting mRNA degradation. This knowledge enables one to make meaningful conclusions on the relationship between variables independently associated with either phenomena, such as the translation time T[‡] (translation), and mRNA half-life $t_{1/2}$ (degradation). To further reiterate the importance of mRNA degradation, a vital factor in maintaining an organism's ability to survive is the capacity to regulate the expression of genetic information[§]. Degradation, like translation, is one process that has been shown to be highly influential regarding gene expression [23]. Since the mRNA strand acts as a template for ribosomal translation, the quantity of each protein synthesised during this process is therefore sensitive to the period of time the mRNA strand remains stable. This notion pertains its value across all domains of organism, whether uni or multi-cellular, due to the constant need for all organisms to produce the appropriate proteins necessary for an ever-changing environment. We finally note that the process of mRNA degradation varies widely between organisms of different species, hence, we will solely concern ourselves with the degradation mechanisms associated with prokaryotic bacteria such as E. coli [23].

2.2.2 mRNA Structure

We begin our analysis on the mechanisms responsible for prokaryotic mRNA degradation, by first, enhancing our knowledge on the structure of the prokaryotic mRNA strand. Preceding the start codon, the mRNA strand begins with what is known as the 5′ end. This extremity is characterised by a triplet of phosphate molecules indicated on the left of Figure 3. Following the 5′ end comes the 5′ untranslated region (UTR), as well as the Shine-Dalgarno sequence, all of which are also indicated on the left of Figure 3. An analogous 3′ end (minus the triphosphate group) and UTR are also present on the opposite end of the mRNA strand. Unlike the 5′ end however, the 3′ end possesses a *stem loop* (see Figures 4 and 5); a simple structure which turns out to have profound effects on the prokaryotic mechanism for mRNA degradation. Finally, we became familiar with the region between the 5′ and 3′ ends at the start of Section 2; this is known as the coding sequence, or CDS, due to this being where the ribosomes translate the mRNA sequence into the appropriate polypeptide chain.

2.2.3 Prokaryotic mRNA Degradation

The constituents responsible for the degradation of the mRNA strand can be generally classified into two, distinct groups: endonucleases, and exonucleases. Endonucleases are a type of enzyme which break the internal bonds of polynucleotides such as mRNA. Meanwhile exonucleases exclusively begin to break nucleotide bonds at one end of the molecule, before sequentially moving along to opposite end, breaking the other bonds in the process [2]. Historically, the mechanism proposed for prokaryotic mRNA degradation involved an endonuclease, RNase E, as well as a 3′ exonuclease[†].

[‡]The translation time, T, refers to the time elapsed for the first ribosome to complete the elongation and termination phases of translation.

[§]Expression of genetic information or *gene expression*, is simply the flow of genetic information from gene to protein [2]

[†]3′ in this context denotes which end of the mRNA strand at which degradation begins.

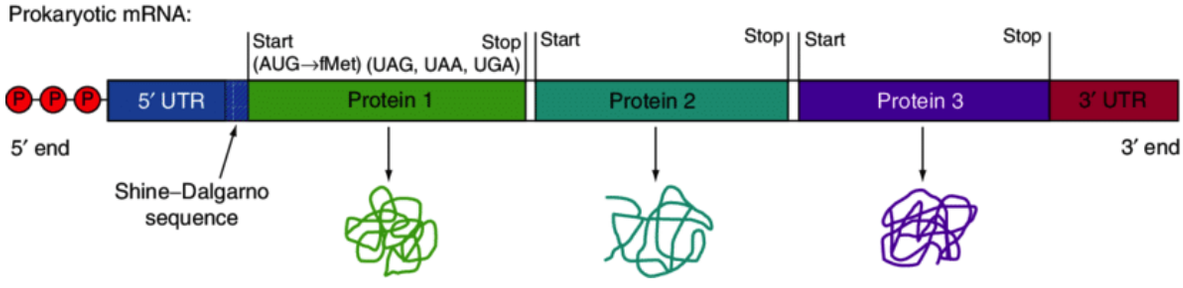


Figure 3: A simplified schematic of the general structure of prokaryotic mRNA. We observe the 5′ end, signified by the presence of a triphosphate group, along with the 5 UTR on the left of the figure. A depiction of the CDS then follows portraying the capacity for multiple proteins to be simultaneously encoded for in prokaryotes such as E. coli. We finally finish with the 3′ UTR and 3′ end depicted on the right of the figure [22].

Motivated by experimental observations stating that 3′ exonucleases are prevented from initiating degradation by the presence stem loops at the 3′ end [24]; the degradation mechanism proposed began with an RNase E internal cleavage of the mRNA strand, yielding two decay intermediates. Following this internal cut, the 5′ fragment is now free of its protective stem loop, formerly at the 3′ end, enabling exonucleases to bind and rapidly begin degrading the 5′ fragment (see Figure 4). Repetition of RNase E cleavage and 3′ exonuclease degradation subsequently conduct the decay of remaining mRNA fragment [25].

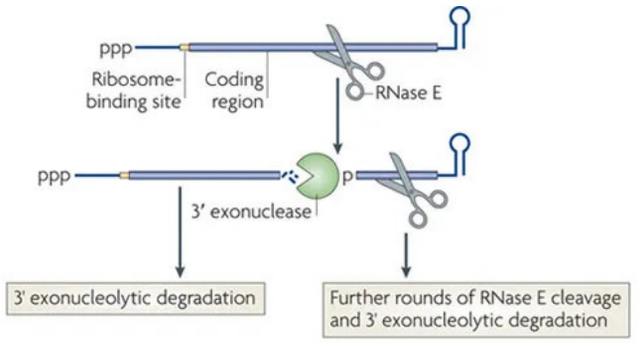


Figure 4: A schematic depiction of one of the early prokaryotic mRNA degradation pathways. In this pathway, RNase E cuts the strand generating two fragments. As a result, these degradation intermediates are susceptible to attack by the 3 exonucleases due to the absence of the stem loop [26].

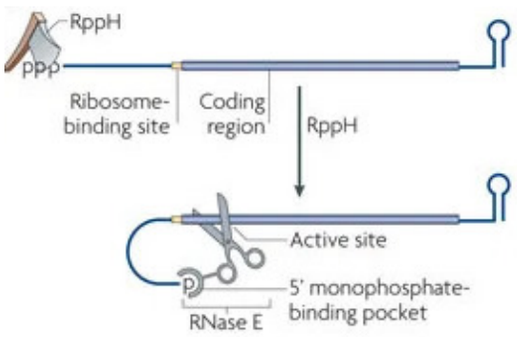


Figure 5: A schematic depiction of an alternative mRNA degradation pathway. The RppH enzyme, represented by the axe, can be seen transforming the triphosphate group into a monophosphate group at the 5′ end of the mRNA strand. RNase E, represented by the scissors, subsequently attacks the this end resulting in the degradation of the strand [26].

Proceeding the inception of this initial mechanism, whilst still viable, it was later discovered that the location of the cleavage by the RNase E enzyme was strongly influenced by the state of the 5′ end of the mRNA strand, specifically, the number of phosphate molecules present at this end. Naturally, this led to the incorporation of a new enzyme in to the mechanism, RNA pyrophosphohydrolase, or RppH, responsible for the conversion of the triphosphate group, into a monophosphate group at the 5′ end. In work conducted by Mackie [27], it was shown that this modification vastly increased the susceptibility of mRNA to degradation by RNase E. Consequently, this led to the addition of an extra initial step into the mechanism for prokaryotic mRNA degradation as depicted in Figure 5 (below). To firmly establish the influence of RppH on the process of degradation, experimental work was conducted on the degradation of E. coli mRNA strands with the RppH enzyme absent. It was observed that in this regime, the steady-state concentration of hundreds of polypeptides increased significantly, thus indicating a significant proportion of the mRNA strands are degraded via this 5′ dependent pathway [28].

2.3 Non-Equilibrium Statistical Mechanics

2.3.1 What is a Non-Equilibrium System?

We now turn our attention to the physics and mathematics which enable us to successfully encapsulate the dynamics of mRNA translation. In an effort to rigorously describe what is meant by a non-equilibrium system, we review the adverse scenario; systems which display properties of being in equilibrium. Within the field of theoretical physics, the term equilibrium is intimately linked with the idea of a system undergoing passive exchange of a relevant quantity, usually with an infinite reservoir of said quantity, until a particular condition is met. For instance, if the system can be sufficiently described by the Canonical Ensemble, the relevant quantity previously discussed would take the form of energy. Our system of interest would then exchange energy with the infinite reservoir, otherwise known as the environment, until the net current of energy from the environment into the system decays to zero. System and environment are then defined to be in thermal equilibrium sharing some common temperature, congruent with the 0^{th} law of thermodynamics. This notion of thermal equilibrium can be succinctly characterised by the Boltzmann distribution, which gives the probability of being in

micro-state C possessing an energy $E(C)$:

$$P_{eq}(C) = \frac{e^{-\beta E(C)}}{Z} \tag{1}$$

where $\beta = 1/k_b T$, and Z is known as the partition function, involving the sum over all possible micro-states C:

$$Z = \sum_{C} e^{-\beta E(C)} \tag{2}$$

This mathematical description of equilibrium can be readily extended in order to describe more complicated scenarios. For example, if the environment and system of interest now have the capacity to exchange particles, in addition to energy, we should be inclined to use the Grand Canonical Ensemble in order to characterise this regime. Total equilibrium is said to be established if the Gibbs-Boltzmann distribution is adhered to, taking the form:

$$P_{eq}(C) = \frac{e^{-\beta(E(C)-N\mu)}}{Z} \tag{3}$$

where μ represents the chemical potential. Furthermore, we must also notice that the partition function must now be extended to include a summation over all micro-states for each particle number:

$$Z = \sum_{C,N} e^{-\beta E(C)} \tag{4}$$

With this mathematical definition of what is meant by an equilibrium system, we are now equipped with the knowledge enabling us to appreciate systems said to be 'out of equilibrium'. A particular instance of this phenomenon that we must be familiar with, in the context of mRNA translation, are non-equilibrium steady states. This refers to when a system has reached a regime where its properties no longer change in time, but additionally, do not conform to either of the probability distributions mentioned previously in Equations 1 and 3.

Parallels can be drawn between the conditions prevalent in non-equilibrium steady states, and what manifests during the process of mRNA translation. We can think of the inception of a new mRNA strand as our system being prepared out of equilibrium, where each codon site on the strand starts unoccupied. As described previously, ribosomes will begin to initiate on to the mRNA strand as well as translate across it, elongating the polypeptide chain in the process. During the early stages mRNA translation, the properties of the system[†] (discussed in more detail in Section 3) are subject to change, but display behaviour of heading towards some constant value. Following the relaxation time of the system having elapsed, these constant values are attained. This does not mean our system is stationary however. The process of mRNA translation is one which is powered[‡] by its environment, rather than reaching a passive equilibrium with it. A current of ribosomes remains present as they are driven through the system, implying the presence of a net probability current; the primary characteristic of a non-equilibrium system.

2.3.2 Markov Processes and the Master Equation

Analogous to employing the Boltzmann or Gibbs-Boltzmann distribution in order to characterise a system in equilibrium, it is natural to wonder whether a similar framework can be developed for non-equilibrium systems. In order to do so, one must consider variables that are stochastic in nature, meaning they are governed by the laws of probability, rather than determinism. To aid us in gaining this ability, we begin with a review of Markov processes; a stochastic model discovered in 1906 by Russian mathematician Andrey Markov [29]. A Markov process is defined to be when a conditional probability distribution of future events depends only upon the present state of the system, hence rendering the current state independent of all events preceding it. This statement indicates that our system adheres to what is known as the Markov property [30]. To mathematically illustrate this notion, let X_n (where $n \in \mathbb{N}$) be a set of stochastic variables on a discrete space E. These variables are analogous to the events discussed previously hence, let a sequence of such events be defined as $\chi = (X_n : n \in \mathbb{N})$; this is known as a Markov chain. If P is then considered to be a probability measure of χ, the Markov property can be defined such that:

$$\begin{aligned} P(X_{n+1} = j | X_0 = i_0, ..., X_n = i_n) \\ = P(X_{n+1} = j | X_n = i_n) \end{aligned} \tag{5}$$

For all $i_0, ..., i_n, j \in E$ and $n \in \mathbb{N}$.
The mathematics of Markov chains enables us to construct an equation capable of describing the

[†]In this context, the system is comprised of the mRNA strand, as well as the environment encompassing the molecule.
[‡]The forces driving mRNA translation in E. coli originate from metabolic processes within the cell [1].

dynamics of non-equilibrium, stochastic systems. This equation takes the form of a differential equation in $P(C,t)$[†], known as the **Master Equation**[§]:

$$\frac{\partial}{\partial t}P(C,t) = \sum_{C' \neq C} W(C' \to C)P(C') - \sum_{C' \neq C} W(C \to C')P(C) \quad (6)$$

With this new mathematical framework, we are now able to formally characterise the non-equilibrium steady state present in mRNA translation. From the Master Equation, it is easy to see that steady state is defined to be where $\frac{\partial P(C,t)}{\partial t} = 0$, leaving the requirement of finding a solution to a set of linear equations:

$$\sum_{C' \neq C} W(C' \to C)P^*(C') = \sum_{C' \neq C} W(C \to C')P^*(C) \quad (7)$$

P^* now denotes the steady-state probability distribution. For a system in equilibrium, this is trivially achieved by a pair-wise cancellation of each term in the summations presented in Equation 7, yielding:

$$W(C' \to C)P_{eq}(C') = W(C \to C')P_{eq}(C) \quad (8)$$

This condition is referred to as *detailed balance*, or *reversibility*, which simultaneously emphasises an absence in the flow of probability between the two micro-states, C and C', in addition to the system exhibiting time-reversal symmetry. To further illustrate that the equilibrium condition is encapsulated by this simple equation, let us insert the Boltzmann distribution (Eq. 1) for P_{eq} into Equation 8. Combining this with some simple rearranging and we receive:

$$\frac{W(C' \to C)}{W(C \to C')} = \mathrm{e}^{-\beta[E(C')-E(C)]} \quad (9)$$

Therefore an uncompromising classification of a non-equilibrium steady state comes in the form of one whose dynamics do not adhere to the detailed balance condition expressed in Equations 8 and 9. In this regime, a general circulation of probability between all micro-states, C, will be present; a characteristic highlighted at the end of Section 2.2.1. The difficulty we are presented with for non-equilibrium systems is that although the transition rates $W(C \to C')$ may be known[‡], the same cannot be said for $P(C)$, rendering these systems arduous to consider analytically. One solution that is common throughout both computational physics and biology however, is the implementation of numerical algorithms in order to elucidate the dynamics of these complicated, non-equilibrium systems.

An example of this is presented by the Gillespie algorithm, which was the route taken here in order to study the ribosomal dynamics present in the translation of mRNA. In order to avoid confusion regarding nomenclature, we should note that work utilising this algorithm is interchangeably referred to as a dynamic Monte Carlo simulation.

3 Methods and Model Construction

3.1 TASEP Model

The mathematical model employed which successfully encapsulates the process of mRNA translation is known as the totally asymmetric simple exclusion process (TASEP). 'Totally-asymmetric' referring to the uni-directional movement of ribosomes along the mRNA strand; 'simple' denoting that the ribosomes move only one lattice site (codon) at a time. The term TASEP was coined in a more general study of Markov processes by Spitzer in 1970 [31], but was curated in 1968 by MacDonald, Gibbs and Pipkin [5] with the aim to describe the kinetics of mRNA translation. It is important to note however, that this kind model is widely applicable within the field of non-equilibrium statistical mechanics, and does not solely exist to fulfill the original purpose set out by its curators [6, 32].

3.1.1 Mathematical Employment

In the TASEP model, the mRNA strand is characterised as a one-dimensional lattice consisting of L discrete lattice sites (codons). Site 1 represents the start codon, whilst site L represents the stop codon. Ribosomes are represented by particles of

[†] $P(C,t)$ explicitly denotes the probability of being in micro-state, C, at a time, t.

[§] See Appendix A for the derivation of the Master Equation.

[‡] Generally through experiment [17].

[†] Due to the appropriate aa-tRNA being incorporated into the polypeptide chain during the initiation process, we only concern ourselves with sites $2 < i \leq L$

length l =10 [34], implying the occupation of 10 lattice sites following initiation. The position of a ribosome on the lattice, $(2 < i \leq L)^{\dagger}$, is associated with the position of its A-site, located 6 lattice sites from the trailing end of the ribosome. The population of ribosomes on the mRNA strand is monitored by assigning an occupancy variable, τ_i, to each codon site $i = 2, ..., L$ where $\tau_i \in [0, 1]$:

$$\tau_i = \begin{cases} 1, & \text{if site i is occupied} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Note that a site is defined to be occupied if, and only if, a ribosome's A site is present at said location; a lattice site can therefore remain unoccupied according to τ, even if other segments of the ribosome occupy that location. By including the occupancy variable, we can now characterise the configurational state of the system with an occupancy vector $C = \{\tau_2,...,\tau_L\}$, analogous to the micro-states defined in Section 2.

Initiation is described as a single step process by the TASEP model, where the A site of the newly recruited ribosome occupies lattice site 2, whilst the ribosomal P site occupies site 1 (start codon) following the initiation event. This encapsulation accounts for the fact that an aa-tRNA molecule is recruited and incorporated into the polypeptide chain during this pseudo-single-step process. Mathematically, this reduces initiation to a regime where ribosomes attempt to initiate translation at a rate denoted α, moving from the environment to site 2 in the process. Furthermore, since ribosomes are of finite size and thus must obey an excluded volume constraint[‡], initiation is only successful given sites $i = 2, ..., l + 1$ are unoccupied. Overall, TASEP initiation can thus be succinctly summarised as:

Initiation

$$\tau_2 = 0 \xrightarrow{\alpha} 1, \ \textit{if} \ \tau_2 = ... = \tau_{l+1} = 0 \tag{11}$$

Promptly moving onto the following phase of mRNA translation, elongation. During this phase we are concerned with sites $i = 2, ..., L - 1$, where each site is assigned a position-specific elongation rate [8], ω_i, in accordance with experimental data collected by *Lipowski* et al [17]. Therefore, translation elongation mathematically manifests according to the TASEP model as:

Elongation

$$\tau_i, \tau_{i+1} = 1, 0 \xrightarrow{\omega_i} 0, 1 \ \textit{if} \ \tau_{i+l} = 0 \tag{12}$$
$$\text{For } i = 2, ..., L - 1$$

We conclude our description of the kinetics of mRNA translation with the process of termination. The rate at which ribosomes attempt to induce this stage is denoted as β, and occurs at site

[‡]The event of one ribosome blocking another is known as *steric hinderance* [14].

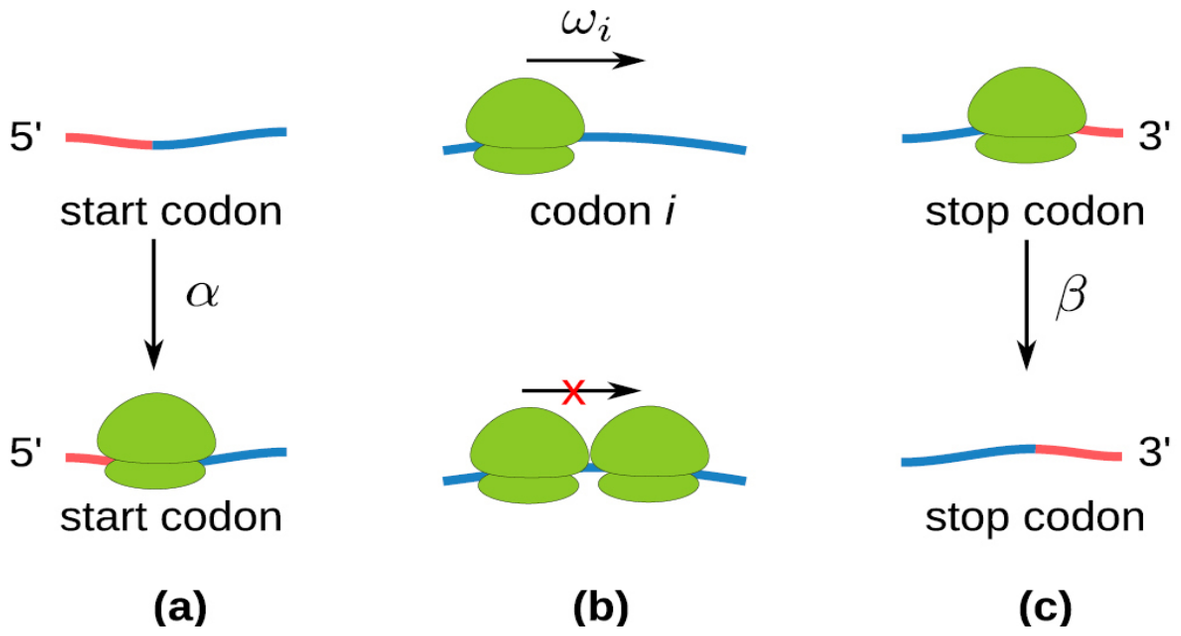


Figure 6: A schematic image of the kinetic steps involved in the TASEP model. Initiation (a): Ribosome attempts to initiate onto the lattice at rate α. Elongation (b): Ribosome translates across the mRNA strand with codon-specific rates $\{\omega_i\}$ for $i = \{2, ...L - 1\}$. Note that ribosomes are capable of preventing one another from conducting translation. Termination (C): Upon reaching the stop codon, the ribosome attempts to dissociate from the mRNA strand with termination rate, β[33].

L only. This is mathematically represented as:

Termination

$$\tau_L = 1 \xrightarrow{\beta} 0 \tag{13}$$

A representation of all steps involved in mRNA translation, as described by Equations 11, 12, and 13, is also presented schematically in Figure 6 (above).

In conclusion, the process of mRNA translation in the context of the TASEP model conforms to the non-equilibrium mathematical framework as discussed in Section 2. We can associate a certain occupancy configuration of the mRNA strand with a probability $P(C, t)$, whose dynamics are governed by the master equation (see Eq. 6). The relevant transition rates in this equation, $W(C \rightarrow C')$, are further associated to the kinetic rates most recently discussed α, β and ω_i.

3.2 Ribosomal Current & Density

Calculations of specific kinetic observables were performed in order to extract the most salient information from the model. An example is the ribosomal density denoted $\rho_i(t)$, which determines the probability of finding a ribosome at codon site i, at a given time t. Mathematically the density is defined to be:

$$\rho_i(t) = P(\tau_i(t) = 1) = \sum_C \tau_i(C) P(C, t) \tag{14}$$

Hence the total density, ρ, is determined by the mean number of ribosomes bound to the mRNA strand at a given time t, divided by the total number of codon sites, L-1:

$$\rho(t) = \frac{1}{L-1} \sum_{i=2}^{L} \rho_i(t) \tag{15}$$

An extension in order to encapsulate the steady state densities ρ_i^* and ρ^* are defined analogously, with $P(C, t)$ replaced by $P^*(C, t)$ as discussed in Section 2.2.2. Although this system is capable of reaching a steady-state, equilibrium can never be attained due to the continuous flow of ribosomes along the mRNA strand. This ribosomal current, J, can be quantified in the steady-state as:

$$\begin{aligned} J^* &= \alpha P^*(\tau_2 = \ldots = \tau_{l+1} = 0) \\ &= \omega_i P^*(\tau_i, \tau_{l+i} = 0), \ \text{i = 2,...,L - 1} \\ &= \omega_i P^*(\tau_i = 1), \quad \text{i = L - 1 + 1,...,L} \\ &= \beta P^*(\tau_L = 1) \end{aligned} \tag{16}$$

These two observables were essential in determining the validity of the stochastic TASEP model. Due to work conducted *Deridda* et al. [6], results from the simulation could be directly compared to the analytical case, reassuring the correct dynamics were being simulated.

3.3 Gillespie Algorithm Implementation

With the mathematical framework of the TASEP model in place, the Gillespie algorithm was implemented with the aim of encapsulating the time evolution of this stochastic system.

We begin the algorithm by initialising an array of values which represent all possible transitions available to the ribosomes. We label this array the *propensity*, and mathematically denote it as so: $a = \{a_2, a_3, ..., a_L\}$. Let us take the a_2 element of this array as an example to illustrate how the propensity array is constructed:

$$a_2 = \omega_2 \tau_2 (1 - \tau_{2+l}) \tag{17}$$

Since the τ variables in Equation 17 belong to the domain $\in [0, 1]$, the value of a_2 can only ever be ω_2, the transition rate at that site, or 0. A qualitative summary of the equation presented above says that a move from site 2 is only possible given that site 2 is already occupied by a ribosomal A site, in addition to the site a distance of l sites down the lattice being free. This latter statement encapsulates the steric hinderance condition obeyed by the ribosomes. The preceding notion can be readily extended across all lattice sites yielding:

$$a_i = \begin{cases} \omega_i & \text{if a move from site i is possible} \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

It is therefore readily seen that the propensity signifies an array of possible transitions available to the ribosome as mentioned previously.

The following step implemented into the algorithm was the crucial calculation of R, a quantity defined as the sum of the propensities: $R = \sum_i a_i$. R is a value that was firstly used to generate stochastic updates in time via the technique of inverse transform sampling. We begin this technique by assuming a probability distribution function (PDF) to be associated with the stochastic time variable, t. In our case, the PDF took the form of an exponential distribution[†]:

[†]See Appendix B for details on the Gillespie algorithm and the reasoning behind this choice of PDF.

$P(t) = R(a)e^{-R(a)t}$. Next, we then calculate the cumulative distribution function (CDF) associated with this variable through simple integration yielding: $F(t) = e^{-R(a)t}$. By replacing the t dependency in F with a randomly distributed variable on $\in [0, 1]$, U, the theory of inverse transform sampling states that the inverse of this new CDF, $F^{-1}(U)$, follows the same distribution as the former variable, t. Overall, applying this mathematical framework to our simulation results in the generation of successive updates in time which are sampled from the following distribution:

$$t = F^{-1}(U) = \frac{ln(U)}{R(a)} \tag{19}$$

Due to R's explicit dependence on the current propensity of the lattice, it should be noted that the increments in time presented by Equation 19 will vary between steps of the simulation, and hence will not be uniform.

An additional procedure involving the value of R, further highlighting its importance, is determining which ribosome is chosen to move. We begin this process by generating a random number between the values of 0, and the sum of the propensity array, R. Since the propensity array encapsulates all ribosome moves that are possible, it is useful to envisage the this array as a 1D length between values 0 and R as depicted in Figure 7.

Figure 7: A pictorial representation of the propensity array, a. The grey ticks indicate separations between the transition rates associated with each lattice site. We see that ω_i has a larger value than ω_j, hence rendering a transition from site $i \to i+1$ more probable than site $j \to j+1$.

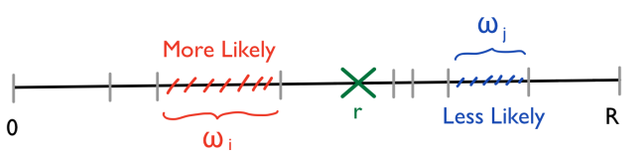


Figure 8: A continuation of the schematic presented in Figure 7. Now, a random number, r, has been generated between the values of 0 and R. The value of r coincides with a transition rate ω_k, provoking the ribosome at that site to translate to the next.

The random number generated, denoted r, will be subsequently placed at the appropriate location on this 1D length, provoking a ribosome to translate. An example of this in action is presented pictorially in Figure 8. We see that r has landed at a point corresponding to some transition rate ω_k[§], indicating the ribosome located at codon site k to move along to site $k+1$. In this pictorial representation, it clear to see that the transition rates, $\{\omega_i\}$, represent a form of probability. The larger a transition rate, the longer the probability length associated with that site is and hence, the more probable a ribosome is to translate from that site to the next. The combined process of calculating R, incrementing time, and finally determining the index of which ribosome is to translate, embodies a random sequential updating scheme. We repeat the 3 previous processes until a particular condition, set by the user, is met. This type of updating method is common throughout all types of Monte Carlo simulations.

4 Results & Discussion

4.1 Analytical Testing

Prior to applying the constructed algorithm to investigate the ribosomal dynamics on a newly produced mRNA strand, the algorithm was first tested against analytical results derived by Derrida *et al.* in 1992 [6]. In this paper, the authors present a new mathematical approach in order to describe the dynamics of "a system of particles hopping uni-directionally with hard core interactions in the case of open boundaries". This regime is formerly known as the one-dimensional fully asymmetric exclusion model, and epitomises the kind of system present in mRNA translation. However, we should note that the analytical results derived are only applicable to particles of size 1, hence, the size of the ribosomes were reduced to $l = 1$ for this testing. Furthermore, as previously mentioned, another necessary constraint that was imposed in order to obtain these results was setting all transition rates, $\{\omega_i\}$, to be equal and uniform. In the case of our simulation, all transition rates were set to unity: $\omega_2 = \omega_3 = ... = \omega_{L-1} = 1.0\ s^{-1}$. With these conditions now established, α, the initiation rate, and β, the termination rate, become the only variable parameters of our system. It is observed that if the domains of these variables are restricted between $0 < \alpha,\ \beta < 1$, the phase diagram presented in

[§]Note that this index is not $i+1$ as the propensity at this site could be 0.

Figure 10 (below) is revealed in the steady-state ($\partial P(C,t)/\partial t = 0$).

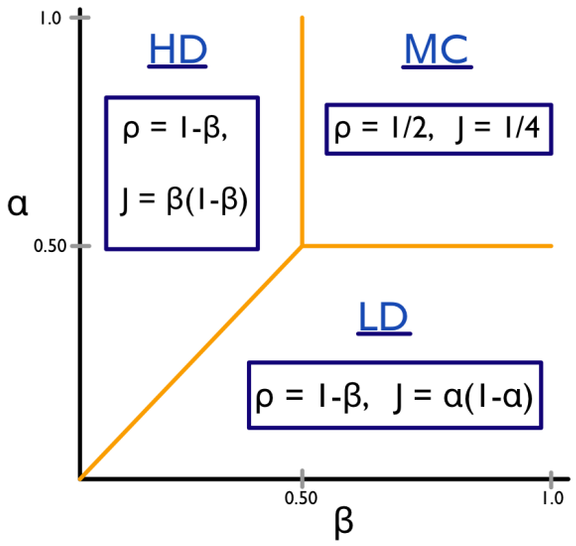


Figure 10: Phase diagram of model constructed by Derrida *et al*. The region $\alpha > 1/2$, $\beta > 1/2$ (top right) is known as the maximal current (MC) phase. In the MC phase the steady-state current and density, J and ρ, are found to be $1/4$ and $1/2$ respectively. The low density phase (LD) is obtained when $\alpha < 1/2$ and $\beta > \alpha$ (bottom right). The respective steady-state current and density values are $J = \alpha(1-\alpha)$ and $\rho = 1-\beta$. Finally, the high density (HD) phase (top left), along with its results, are achieved if one swaps $\alpha \longleftrightarrow \beta$ whilst maintaining $\rho = 1-\beta$.

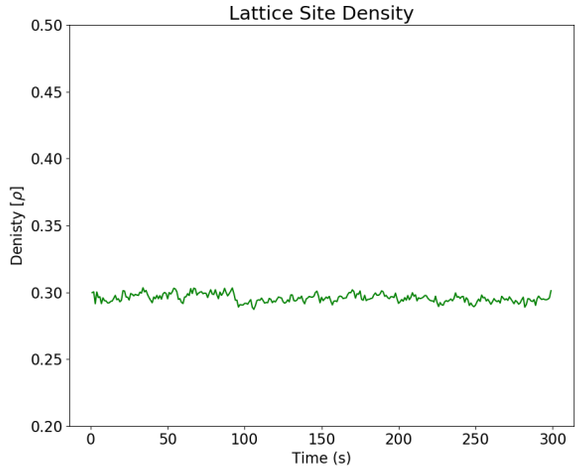
In this figure, three primary phases are presented, each of which possessing an exact, constant value across all lattice sites for the steady-state current and density:

- **Maximal Current (MC):** $\alpha > 1/2$, $\beta > 1/2$
  Steady-State Density: $\rho = 1/2$
  Steady-State Current: $J = J_{max} = 1/4$

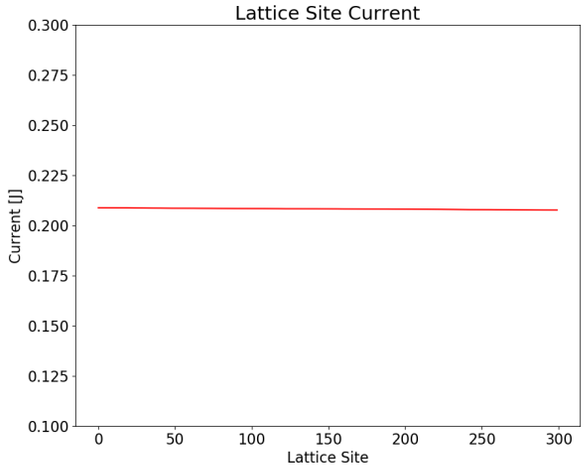
- **Low Density (LD):** $\alpha < 1/2$, $\beta > \alpha$
  Steady-State Density: $\rho = 1-\beta$
  Steady-State Current: $J = \alpha(1-\alpha)$

- **High Density (HD):** $\beta < 1/2$, $\alpha > \beta$
  Steady-State Density: $\rho = 1-\beta$
  Steady-State Current: $J = \beta(1-\beta)$

One is able to make physical sense of the nomenclature employed to characterise each of the three phases with some simple analysis. Taking the LD phase for instance, in this regime the termination rate, β, is large relative to the initiation rate, α, meaning one would expect fewer ribosomes to be occupying sites on the lattice thus yielding a lower overall density. It is also clear to see that this argument can be simply reversed ($\alpha \longleftrightarrow \beta$), whilst maintaining $\rho = 1-\beta$, in order to comprehend the naming of the HD phase. For example, in the high density phase where we are given a relatively large α and small β, one would expect to observe numerous ribosome 'traffic jams' as the ribosomes initiate on to the mRNA strand at a greater rate than they terminate from it, thus yielding a larger overall density.



(a) Overall steady-state density.



(b) Ribosomal current measured at each lattice site.

Figure 9: Plots of the density **(a)** and current **(b)** obtained in the low density phase for an mRNA and ribosome length of $L = 300$, and $l = 1$ codon site(s) respectively. Kinetic parameters were set to the following values: $\alpha = 0.30$, $\beta = 0.70$ and $\omega_2 = \omega_3 = ... = \omega_{L-1} = 1.0$, all in units of s^{-1}.

Finally, the naming of the MC phase simply originates from the system occupying a state where the ribosomal current takes its maximum value ($J = J_{max} = 1/4$).

With the aim of ensuring our algorithm adhered to these analytical results, the simulation was ran setting $\alpha = 0.30$, and $\beta = 0.70$ corresponding to the low density phase. Additional parameters such at the translation rates, $\{\omega_i\}$, and ribosome length, l, were set to unity in units of s^{-1} as quoted previously. The final parameter value worth noting is the length of the mRNA strand which was set to $L = 300$ codons, a typical value observed in E. coli genes [17]. The results of the simulation, which was ran for a total of 5,000,000 Monte Carlo steps, (equivalent to evolving time 5,000,000 times via inverse transform sampling) are presented in Figure 9 above.

In the low density phase, we expect to see a steady-state density value of $\rho = 1-\beta = 0.30$, and a steady-state current value of $J = \alpha(1-\alpha) = 0.21$, both of which are observed on average in Figures 9a and 9b respectively. These time averages were computed explicitly for both observables, with the average density of the mRNA strand in this regime determined to be $\overline{\rho} = 0.30185$ to 5 decimal places. Meanwhile, the average ribosomal current for the mRNA strand was determined to be $\overline{J} = 0.20981$, also to 5 decimal places. Due to the stochastic nature of the system at hand, the accuracy of these results begins to provide diminishing returns as the number of Monte Carlo steps is increased. This belief stems from receiving similar results for simulations which ran with fewer Monte Carlo steps, indicating the existence of a critical number of steps where both computational efficiency, and mathematical accuracy are optimised. We should add that the results for the high density and maximal current phases were also produced with similar accuracy, providing further evidence of concordance between the simulation and analytical results.

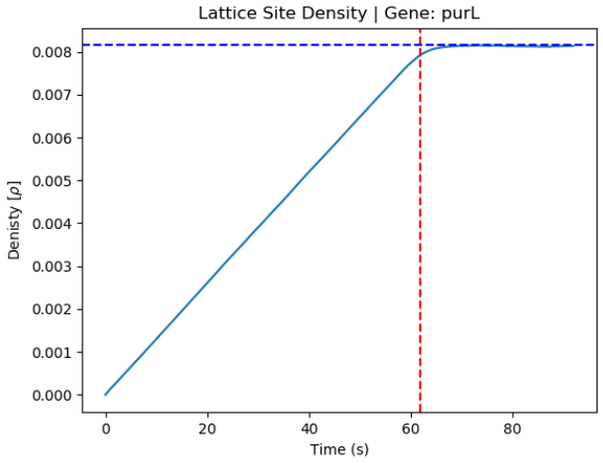
In conclusion, we believe that the results obtained were sufficient to deem the algorithm's validity and hence, press forward with our investigation.

4.2 Non-Steady State Ribosome Dynamics

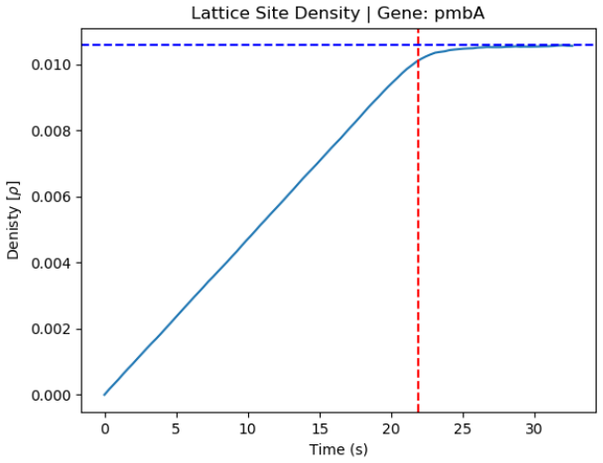
4.2.1 Temporal Density Evolution

With the computational apparatus described above now in place, we consider the ribosomal dynamics on a newly produced mRNA strand in the non-steady state. In this regime, the mRNA strand is produced with an empty lattice, $\rho(t = 0) = 0$, corresponding to no ribosomes having initiated on to the mRNA strand. The time evolution of the ribosomal density is then measured until the steady-state is attained. As discussed previously in Sections 2.3.1 and 4.1, attainment of the steady state can be characterised by the density reaching a constant value we denote ρ^*. Realization of this density value occurs following a specific relaxation time having elapsed. Interestingly, this characteristic time, t_0, seems to be intimately linked to the mean translation time (see Section 2.2.1), $\langle T \rangle$, of the first, pioneering ribosome. This link comes in the form of both the mean translation time, and characteristic time being very similar in magnitude. Across all of the 103 stochastic simulations averaged over 10^4 independent runs for each E. coli gene, this relationship was observed to be true. Three instances of which are displayed in Figure 11 below. The following genes were chosen as representatives of the whole data set due their lengths being of the largest ($L_{purL} = 1294$), intermediate ($L_{pmbA} = 449$) and smallest ($L_{rpsL} = 123$) sizes for E. coli. All 3 plots show a linear increase in the ribosomal density until $t = \langle T \rangle$, indicated by the vertical dashed red line. Now between times $\langle T \rangle < t < t_0$, linearity in the increase of $\rho(t)$ is lost until it finally reaches a constant value where the steady-state is attained $\rho(t_0) = \rho^*$. The length of time for which the increase in density remains non-linear as a fraction of the time taken to reach the steady state can be characterised as the discrepancy between times $\langle T \rangle$, and t_0 coinciding. In regards to the figures depicted below, it is possible to observe the emergence of a relationship between this fraction, and the initiation rate for particular E.coli genes. For instance, in the case of rpsL which has an initiation rate of $\alpha = 0.514495959 s^{-1}$ (**11c**), we can roughly say that $t_0 \approx 7s$ meanwhile $\delta t^{\dagger}$ is found to be around $\delta t := t_0 - \langle T \rangle \approx 1.7s$. Hence, representing δt as a percentage of t_0 we yield $\approx 25\%$. Applying this logic to the purL gene with an initiation rate of $\alpha = 0.177922176 s^{-1}$ (**11a**), we yield a much smaller percentage of $\approx$ 7%. If an E. coli gene has a relatively large initiation rate one should expect there to be an increased amount of ribosome traffic on this mRNA strand, which in turn, leads to slower relaxation dynamics. This notion is in complete congruence with the results depicted in Figure 11.

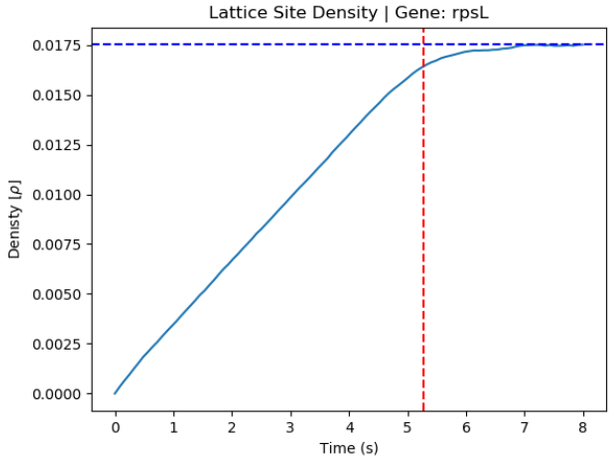
[†]We define δt as the time in which the increase in density remains non-linear.



(a) E. coli gene purL: $L = 1294$, $\alpha = 0.177922176\ s^{-1}$, $\langle T_{exp} \rangle = 61.8874\ s$, $\rho^* = 0.008192$



(b) E. coli gene pmbA: $L = 449$, $\alpha = 0.233617906\ s^{-1}$, $\langle T_{exp} \rangle = 21.8733\ s$, $\rho^* = 0.0105783$



(c) E. coli gene rpsL: $L = 123$, $\alpha = 0.514495959\ s^{-1}$, $\langle T_{exp} \rangle = 5.2772\ s$, $\rho^* = 0.017512$

Figure 11: Time evolution of the ribosomal density, $\rho(t)$, obtained from stochastic simulations averaged over 10^4 independent runs for three unique E. coli genes: purL (**12a**), pmbA (**12b**) and rpsL (**11c**). Red vertical dashed lines indicate the mean translation time, $\langle T \rangle$, for the first pioneering ribosome to complete the first round of translation.

The relationship between $\langle T \rangle$ and t_0 turns out to be a rather convenient result as the value of t_0 can now be quickly estimated given the transition rates of a gene. If we now consider only the translation time T for the pioneering ribosome, due to the emptiness of the mRNA strand, this time can be characterised as a sum of exponentially distributed dwell times ξ_i at each codon site $i = 2, ..., L$

$$T = \xi_2 + \xi_3 + ... \xi_L \quad For\ i = 2, ..., L \tag{20}$$

The summation of these exponentially distributed variables, ξ_i, implies that the mean translation time adheres to what is mathematically known as a hypo-exponential distribution. In work conducted by Szavits-Nossan and Evans [35], the corresponding hypo-exponential probability density function $p(T)$ (PDF), and cumulative distribution function $P(T)$ (CDF) for this system are presented as:

$$p(T) = \sum_{i=2}^{L} \omega_i e^{-\omega_i T} \left[\prod_{i=2,\ i \neq j}^{L} \frac{\omega_i}{\omega_i - \omega_j} \right],$$

$$P(T) = 1 - \sum_{i=2}^{L} e^{-\omega_i T} \left[\prod_{i=2,\ i \neq j}^{L} \frac{\omega_i}{\omega_i - \omega_j} \right] \tag{21}$$

The authors, from Eq. 22, were also able to determine an analytical value for the mean translation time along with its variance. These expressions

were shown to be of the following form:

$$\langle T_{ana} \rangle = \sum_{i=2}^{L} \frac{1}{\omega_i}, \quad \sigma^2_{ana}(T) = \sum_{i=2}^{L} \frac{1}{\omega_i^2} \tag{22}$$

It is now easily seen that the mean translation time and thus, the time for the process of mRNA translation to reach the steady-state, t_0, can be readily calculated if given the transition rates for a particular gene. The analytical values for the mean translation time for purL, pmbA and rpsL are quoted to 4 decimal places as 61.9163 s, 21.9742 s and 5.3821 s respectively. Additionally, similar accordance between values of $\langle T_{ana} \rangle$ and $\langle T_{exp} \rangle$ were observed for all 103 E. coli genes, providing further reassurance that the appropriate dynamics were being produced by the simulation.

4.3 Statistical Analysis

4.3.1 Unstable E. Coli Gene Analysis

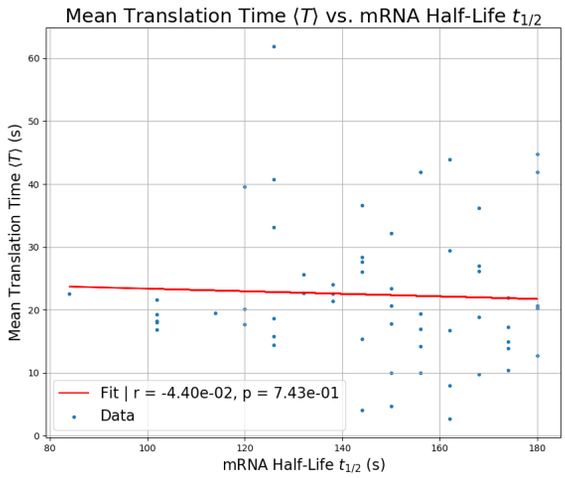
Subject to the receipt of the appropriate results for the non-steady state regime, our aim was to investigate the existence of any underlying relationships between the kinetic parameters and properties of our system. Examples of these variables include: mean translation time $\langle T \rangle$; mRNA half-life $t_{1/2}$; initiation time λ[†]; initiation rate α; and coding sequence (CDS) length L. An emphasis was placed on the relation between the mean translation time and the mRNA half-life of an E. coli gene specifically, as these variables act as a proxy for the processes of translation and degradation respectively. Our initial expectation was to receive a positive correlation between these two variables, subject to the belief that the mRNA strand must remain stable for long enough to allow a significant number of proteins to be synthesised. Whilst this belief pertains true, it was in fact incorrect to assume this would also entail a positive correlation between $\langle T \rangle$ and $t_{1/2}$.

We began our analysis by selecting 60 of the most unstable E. coli genes from the data available, each gene possessing a half life of < 3 minutes. We display the results in Figure 12a, where the red line represents a linear least squares fit through the plotted data points. Computing the Pearson correlation coefficient for this data we yield a value of $r = -0.042$, indicating a distinct lack of correlation between these two variables $\langle T \rangle$, and $t_{1/2}$. It should be noted, however, that the corresponding p-value for this calculation was 0.743, thus far too large for any conclusion drawn from this trend to be of any significance.

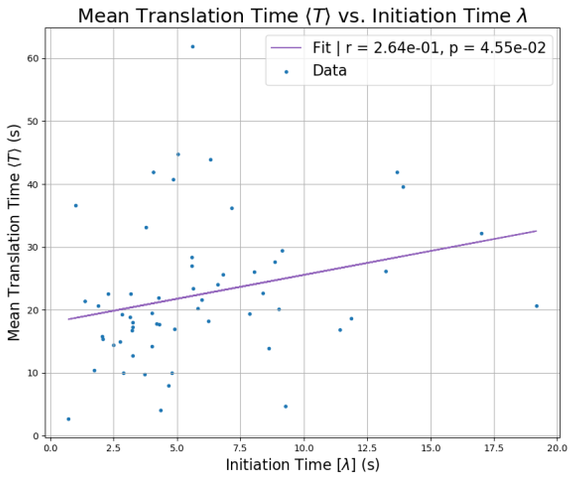
Promptly moving on to discuss the data obtained for Figures 12b & 12c, we observe a weak positive correlation between the mean translation, and initiation time, ($r = 0.264 \rightarrow \langle T \rangle, \lambda$), as well as a weak negative correlation between the mean translation time, and initiation rate, ($r = -0.257 \rightarrow \langle T \rangle, \ \alpha$). The respective p-values for these two results are 0.0455 and 0.0511, indicating these findings can be accepted with a substantial degree of confidence. On the contrary, one must be proceed with caution before bold claims are made from this data. We observed that these r and p values were very sensitive to the number of data points included in the plot, and under these circumstances, two parameters which share a correlation coefficient on the weaker side ($r \approx \pm 0.25$) could be arguably considered as uncorrelated. Therefore, in order to establish a stronger argument to support these findings, a larger number of E. coli genes ($\sim 10^3$) across all ranges of stability should be considered. Regardless of limited sample sizes as presented here, our expectation is for these weak relationships to remain across all E. coli genes. This forecast can be justified due to the translation rates for an E.coli gene usually being of an order of magnitude greater than the initiation rate in majority of cases. This illustrates that translation initiation is often the rate-limiting step in the overall process of mRNA translation, in line with other work conducted on the matter [36].

Finally diverting our attention to the data presented in Figure 12d, we observe an extremely strong positive correlation ($r = 0.996 \rightarrow \langle T \rangle, \ L$) between the mean translation time, and the length of the CDS. A corresponding p-value, further reinforcing our confidence in this intimate relationship, was also calculated to be 1.31×10^{-59}. Of course, this should come as no surprise; the longer the mRNA strand, the longer it takes for a ribosome to traverse the molecule. Although this may seem like a trivial result, there is still some value in possessing the knowledge that the length of the CDS seems to be the primary contributing factor in determining the value of $\langle T \rangle$.

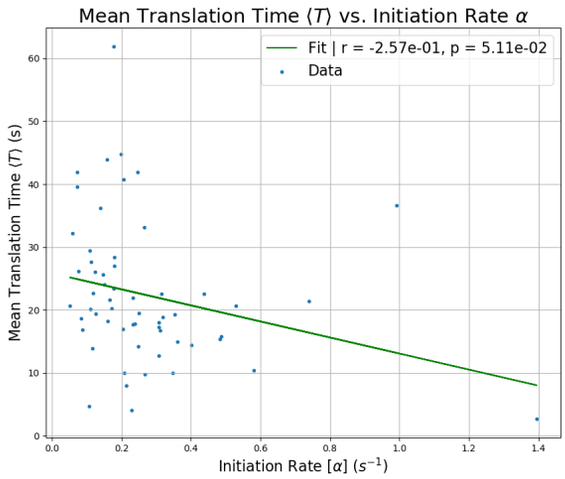
[†]We define the initiation time explicitly as the time elapsed before a ribosome initiates onto the mRNA strand.



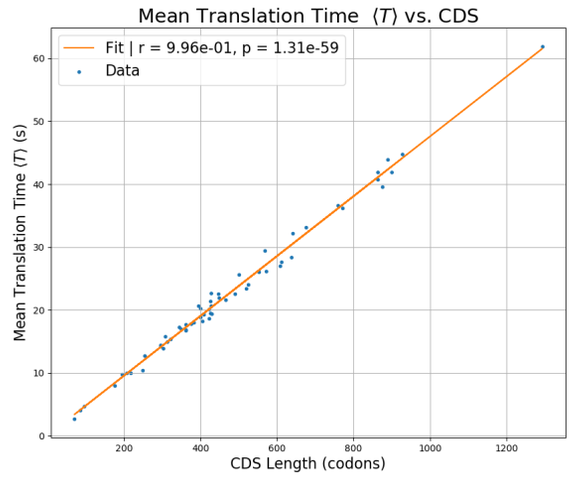
(a) Mean Translation Time $\langle T \rangle$, and mRNA Half Life $t_{1/2}$: $r = -0.0442$, $p = 0.743$



(b) Mean Translation Time $\langle T \rangle$, and Initiation Time λ: $r = 0.264$, $p = 0.0455$



(c) Mean Translation Time $\langle T \rangle$, and Initiation Rate α: $r = -0.257$, $p = 0.0511$

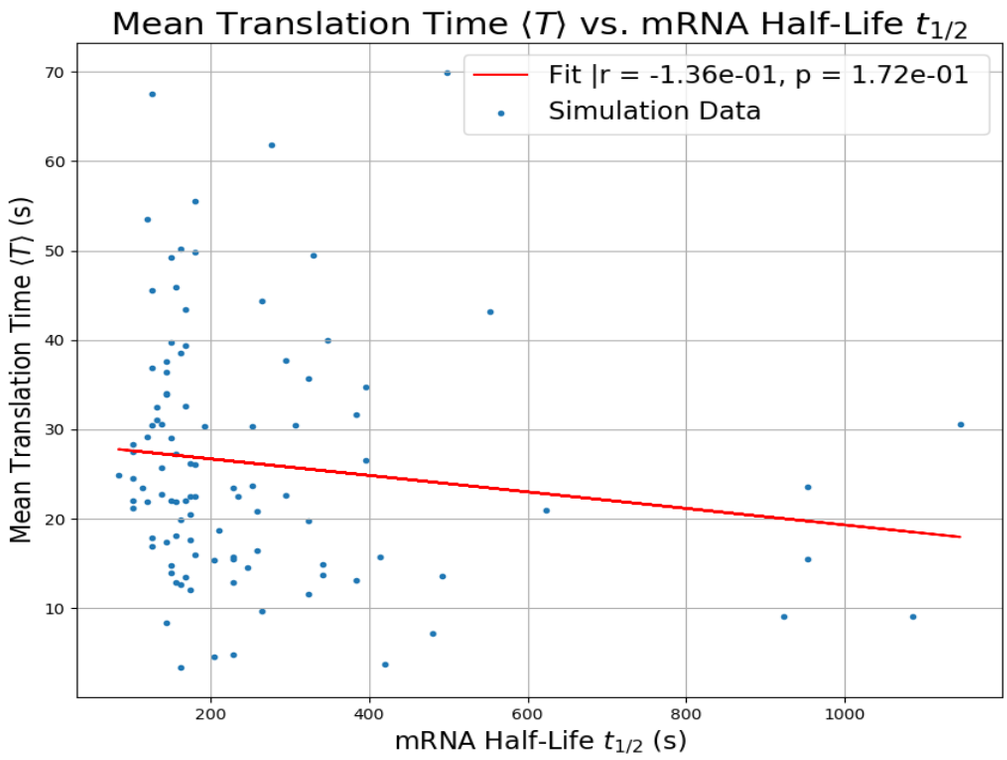


(d) Mean Translation Time $\langle T \rangle$, and CDS L: $r = 0.996$, $p = 1.31 \times 10^{-59}$

Figure 12: Plots to investigate correlations between the mean translation time, $\langle T \rangle$, and other kinetic parameters and observables such as: (a) mRNA half-life, $t_{1/2}$; (b) initiation time, λ; (c) initiation rate, α; and (d) CDS, L. Data collected for these plots was restricted to E. coli genes with half-lives < 3 minutes corresponding to the most unstable genes. The straight line drawn in each respective sub-figure represents the linear least squares fit through the plotted data points. Corresponding r and p-values for these fits can be observed either in the plot legends, or in the sub captions just below the relationships depicted.

4.3.2 The Relationship Between mRNA Translation and Degradation

Due to the extremely large p-value obtained previously for the correlation coefficient shared between $\langle T \rangle$ and $t_{1/2}$, the condition of only considering unstable E. coli genes was relaxed. This enabled a greater quantity of data to be collected in the hope of establishing a more formal conclusion on the relationship between these variables and hence, also on the relationship between the processes of translation and degradation. Consequently, an additional 43 randomly selected E. coli genes were considered resulting in the analysis of a total of 103 genes. As a collective, the genes considered possessed an average half-life of ~ 4 minutes, with the largest half-life considered being that of menA with 1146s. As presented in



(a) Mean Translation Time $\langle T \rangle$, and mRNA Half Life $t_{1/2}$:
$r = -0.136$, $p = 0.172$

Figure 13: Analogous plot to Figure 12a to further investigate the correlation between the mean translation time, $\langle T \rangle$, and half-life of an E. coli gene, $t_{1/2}$, with the condition of E. coli gene half-lives being < 3 minutes relaxed. We observe almost no correlation between these variables which were chosen specifically to elucidate whether there is any causal connection between the translation and degradation processes in E. coli. The red line depicted again represents the linear least squares fit through the plotted data with corresponding r and p-values shown in the legend and sub-caption.

Figure 13, we observe a decrease in the correlation coefficient between $\langle T \rangle$ and $t_{1/2}$ from $r = -0.042$, to $r = -0.136$, indicating a stronger, negative correlation. Furthermore, we also observe a decrease in the p-value associated with this result from $p = 0.743$, to $p = 0.172$.

Whilst this new p-value is a vast improvement on the former, it remains too large for any conclusion to be established with great confidence. This does not render these results insignificant however. Given that our initial expectation was to observe some kind of positive correlation between the mean translation time and half-life on an mRNA strand, it was surprising to observe the distinct lack of correlation between these two variables. We interpret the following results as the processes of mRNA translation and degradation being conducted as effectively independent for short lifetime, non-steady state E. coli genes. This can be explicitly portrayed as the ribosomes completing the synthesis of polypeptides on timescales significantly shorter than the enzymes RppH, RNase E, and 3'-exonucleases take to degrade the mRNA strand. Given this lack of correlation, and also considering that E. coli mRNA strands possess some of the shortest half-lives relative to any organism, this conclusion therefore strengthens the steady-state assumption employed by the majority of work concerning the kinetic model of translation.

5 Conclusion

In this paper, we have conducted stochastic simulations of mRNA translation utilising the Gillespie algorithm within the framework of the inhomogenous TASEP model. The aim of this mathematical modelling was to investigate the relationship between the mean translation time of a ribosome, and the half-life of the mRNA strand that was subject to the translation. Motivation for this work stemmed from the majority of existing work concerning the kinetic model of translation employing a steady-state assumption, originally only to ease the weight of the mathematics involved. Our results conclude that this is a very reasonable assumption to make, even for organisms with the most unstable mRNA such as E. coli.

Once accordance between the simulation and analytical results (derived by Derrida *et al*) was achieved, we then observed the time evolution of the ribosomal density on a newly produced mRNA strand. It was presented that the steady-state in this non-equilibrium regime was attained shortly following the synthesis of the first polypeptide, provided translation initiation remained the rate-limiting step. This mathematically manifests as the initiation rate, α, being $\sim$ an order of magnitude greater than the translation rates, $\{\omega_i\}$. A notion which pertained true across all 103 E. coli genes considered. Consequently, in the context of mRNA translation, one is now capable of making an estimate on the relaxation time into the steady state from the translation time of the first, pioneering ribosome.

Statistical analysis was then performed on the data obtained for 60 of the most unstable ($t_{1/2} < 3$ minutes) E.coli genes. Our aim was to test for the presence of correlations between the mean translation time, and other relevant kinetic variables. Specific emphasis was placed on the relationship between the mean translation time and the half-life of the mRNA stand. This consequently lead to the analysis of an additional 43 E. coli genes with the instability condition relaxed. A deeper investigation was required as the relationship between these two variables is paramount in justifying the steady-state assumption adopted by the majority of kinetic models aiming to describe mRNA translation. We have observed a distinct lack of correlation between these two variables, indicating the degradation time for an mRNA strand to be independent of the translation time. We infer from this conclusion that mRNA translation and degradation are processes operating over different timescales, with the latter acting more slowly. In conclusion, we note that this result provides further justification for the steady-state assumption commonly employed to study the dynamics of mRNA translation. In future work, it would be interesting to continue this kind of statistical analysis across a much larger number of genes, enabling stronger conclusions to be drawn from the data.

A Derivation of the Master Equation

Our aim is to establish a differential equation in $P(C, t)$ and hence, characterise the temporal evolution of our non-equilibrium system. Let us consider a stochastic variable, $X(t)$, which jumps between different values (or micro-states, C) at discrete, random times. The state of the system following a small time interval Δt having elapsed takes the following form [37]:

$$P(C', t + \Delta t) = \sum_{C} P(C', t + \Delta t | C, t) P(C, t) \tag{23}$$

Expanding the $P(C', t + \Delta t|C, t)$ term around $\Delta t = 0$ to first order in Δt yields:

$$P(C', t + \Delta t|C, t) = P(C', t|C, t) + \frac{\partial}{\partial t'} P(C', t'|C, t)|_{t'=t} \Delta t \tag{24}$$

Now, the first term in Equation 24 is only non zero if $C' = C$, instructing us to express this as a Dirac-Delta function $\delta_{C',C}$. In the case where $C' \neq C$, this leaves us with a solitary term on the right hand side. We can further identify this term (multiplied by Δt) as the transition rate, $W(C \to C', t)$; quantifying the rate at which our system transitions from micro-state C, to micro-state C'. If we additionally assume that our Markov chain is homogeneous in time, the time dependency of the transition rate can be dropped, thus giving:

$$W(C \to C') = \frac{\partial}{\partial t'} P(C', t'|C, t)|_{t'=t}, \quad \textit{For } C' \neq C \tag{25}$$

Moving on to the case where $C' = C$, we first must make note that $\sum_{C''} P(C'', t'|C, t) = 1$; simply expressing that at least one of the available micro-states must be occupied by the system. This therefore implies that:

$$\sum_{C''} \frac{\partial}{\partial t'} P(C'', t'|C, t)|_{t'=t} = 0 \tag{26}$$

Splitting the summation up into two terms: where $C'' = C$; and $C'' \neq C$; along with rearranging the former equation allows us to write:

$$\frac{\partial}{\partial t'} P(C, t'|C, t)|_{t'=t} = - \sum_{C'' \neq C} W(C \to C'') \tag{27}$$

We have now analysed how our system evolves after a time Δt, allowing us to conclude that:

$$P(C', t + \Delta t|C, t) = \begin{cases} W(C \to C')\Delta t, & C' \neq C \\ 1 - \sum_{C'' \neq C} W(C \to C''), & C' = C \end{cases} \tag{28}$$

Finally, by inserting our result for $P(C', t + \Delta t|C, t)$ from Equation 28 into Equation 23 and taking the limit $\Delta t \to 0$, we yield our desired differential equation in $P(C, t)$; the **Master Equation**:

$$\boxed{\frac{\partial}{\partial t} P(C, t) = \sum_{C' \neq C} W(C' \to C) P(C') - \sum_{C' \neq C} W(C \to C') P(C)} \tag{29}$$

B Gillespie Algorithm

Given a continuous-time Markov process, it is instinctive to develop a framework in order to predict which of the moves in the Markov process is next to occur, and what is the mean waiting time between these moves? This framework was presented by Daniel T. Gillespie in 1976, motivated to understand 'the time evolution of any spatially homogeneous mixture of molecular species which inter-react through a specified set of coupled chemical reaction channels' [38]. His work went on to be far more general than simulating coupled chemical reactions however, and is now common throughout computational physics and biology.

First consider a probability, $Q(C, t, t_0)$, the survival probability of the system remaining is micro-state C, between times t_0, and t. Let us inspect how this probability changes once a small time interval Δt has elapsed [37]:

$$Q(C, t + \Delta t, t_0) = P(C, t + \Delta t | C, t) Q(c, t, t_0) \tag{30}$$

We recognise an old friend on the right hand side of Equation 30 in $P(C, t + \Delta t|C, t)$, which was defined in Equation 24 to first order in Δt in Appendix A:

$$\begin{gathered} For\ C' = C, \\ P(C, t + \Delta t|C, t) = 1 - R(C)\Delta t, \quad R(C) = \sum_{C' \neq C} W(C \to C') \end{gathered} \tag{31}$$

Inserting this expression back into Equation 30 and then taking the limit of $\Delta t \to 0$ implies:

$$Q(C, t + \Delta t, t_0) = Q(C, t, t_0)(1 - R(C)\Delta t)$$

$$\Rightarrow \lim_{\Delta t \to 0} \frac{Q(C, t + \Delta t, t_0) - Q(C, t, t_0)}{\Delta t} = -R(C)Q(C, t, t_0)$$

$$\Rightarrow \frac{\partial}{\partial t} Q(C, t, t_0) = -R(C)Q(C, t, t_0) \tag{32}$$

Again assuming this to be a process homogeneous in time (meaning $R(C)$ has no explicit dependence on time), Equation 32 is clearly satisfied by the following exponential distribution:

$$\boxed{Q(C, t, t_0) = \mathrm{e}^{-R(C)t}} \tag{33}$$

The expectation value of this well known distribution tells us that the mean waiting time to the next event is therefore $1/R(C)$. Hence, the probability that the event happens to be the transition from micro-state, $C \to C'$, is given by $W(C \to C')/R(C)$.

In conclusion, the Gillespie algorithm allows us to assign meaning to the elements within a Markov chain. Each move, or event, in the Markov chain is chosen sequentially and at random; hence this process is generally referred to as a *random sequential updating scheme* in the context of computer algorithms. This move is then accepted with probability $W(C \to C')/R(C)$, as mentioned previously. Overall, when this process considered holistically, it is variously referred to as *dynamic Monte Carlo*, *kinetic Monte Carlo* or the *Gillespie algorithm*, and is an essential tool in the analysis non-equilibrium systems for when the solutions to the Master Equation are not analytically known.

C mrna-translation Code

All code used to complete this work is open-source and protected under the GNU General Public License v3.0. It can be found via the following URL:

`https://github.com/c-abbott/mRNA-translation`

References

[1]G. M. Cooper, *The cell : a molecular approach*, eng, Second edition.. (ASM Press ; Sinauer Associates, Washington, D.C. : Sunderland, Mass., 2000).

[2]R. Hine, *Messenger rna*, eng, 2019.

[3]H. Gingold and Y. Pilpel, "Determinants of translation efficiency and accuracy", **7** (2011).

[4]T. E. Gorochowski, I. Avcilar-Kucukgoze, R. A. L. Bovenberg, J. A. Roubos, and Z. Ignatova, "A minimal model of ribosome allocation dynamics captures trade-offs in expression between endogenous and synthetic genes", eng, **5**, 710 (2016).

[5]C. T. Macdonald, J. H. Gibbs, and A. C. Pipkin, "Kinetics of biopolymerization on nucleic acid templates", eng, Biopolymers **6**, 1–25 (1968).

[6]B Derrida, "Exact solution of a 1d asymmetric exclusion model using a matrix formulation", eng, Journal of Physics A: Mathematical and General **26**, 1493–1517 (1993).

[7]A. J. Kemp, R. Betney, L. Ciandrini, A. C. M. Schwenger, M. C. Romano, and I. Stansfield, "A yeast trna mutant that causes pseudohyphal growth exhibits reduced rates of cag codon translation", **87**, 284–300 (2013).

[8]J. Szavits-Nossan, L. Ciandrini, and M. C. Romano, "Deciphering mrna sequence determinants of protein production rate", eng, Physical review letters **120**, 128101 (2018).

[9]D. Rogers, M. Böttcher, A Traulsen, and D Greig, "Ribosome reinitiation can explain length-dependent translation of messenger rna", eng, PLoS Computational Biology (2017).

[10]T. Esquerré, A. Moisan, H. Chiapello, L. Arike, R. Vilu, C. Gaspin, M. Cocaign-Bousquet, and L. Girbal, "Genome-wide investigation of mrna lifetime determinants in escherichia coli cells cultured at different growth rates", eng, BMC genomics **16**, 275 (2015).

[11]T. Esquerré, S. Laguerre, C. Turlan, A. J. Carpousis, L. Girbal, and M. Cocaign-Bousquet, "Dual role of transcription and transcript stability in the regulation of gene expression in escherichia coli cells cultured on glucose at different growth rates", eng, Nucleic acids research **42**, 2460 (2014).

[12]D. W. Selinger, R. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow, "Global rna half-life analysis in escherichia coli reveals positional patterns of transcript degradation", English, Genome Research **13**, 216–223 (2003).

[13]E. Yang, E. Van Nimwegen, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnaso, and J. Darnell James E., "Decay rates of human mrnas: correlation with functional characteristics and sequence attributes", English, Genome Research **13**, 1863–1872 (2003).

[14]L. B. Shaw, J. P. Sethna, and K. H. Lee, "Mean-field approaches to the totally asymmetric exclusion process with quenched disorder and large particles", eng, Physical review. E, Statistical, nonlinear, and soft matter physics **70**, 021901 (2004).

[15]D Botto, A Pelizzola, M Pretti, and M Zamparo, "Dynamical transition in the tasep with langmuir kinetics: mean-field theory", eng, Journal of Physics A: Mathematical and Theoretical **52**, 045001 (2019).

[16]S. L. A. de Queiroz and R. B. Stinchcombe, "Nonequilibrium processes: driven lattice gases, interface dynamics, and quenched-disorder effects on density profiles and currents", eng, Physical review. E, Statistical, nonlinear, and soft matter physics **78**, 031106 (2008).

[17]S. Rudorf and R. Lipowsky, "Protein synthesis in e. coli: dependence of codon-specific elongation on trna concentration and codon usage", eng, PLoS ONE **10**, e0134994 (2015).

[18]T. E. Gorochowski, I. Chelysheva, M. Eriksen, P. Nair, S. Pedersen, and Z. Ignatova, "Absolute quantification of translational regulation and burden using combined sequencing approaches", Molecular Systems Biology **15**, n/a–n/a (2019).

[19]J. Bernstein, A. Khodursky, P. Lin, S Lin-Chao, and S. Cohen, "Global analysis of mrna decay and abundance in escherichia coli at single-gene resolution using two-color fluorescent dna microarrays", English, Proceedings Of The National Academy Of Sciences Of The United States Of Ame **99**, 9697–9702 (2002).

[20]F. Blattner, G Plunkett, C. Bloch, N. Perna, V Burland, M Riley, J Colladovides, J. Glasner, C. Rode, G. Mayhew, J Gregor, N. Davis, H. Kirkpatrick, M. Goeden, D. Rose, B Mau, and Y Shao, "The complete genome sequence of escherichia coli k-12", English, Science **277**, 1453–amp; (1997).

[21]*Transcription and translation*.

[22]H. Kubinyi, "Encyclopedia of molecular cell biology and molecular medicine. second edition. vols. 1–5. edited by robert a. meyers", eng, Angewandte Chemie International Edition **43**, 4689–4689 (2004).

[23]R. Rauhut and G. Klug, "mRNA degradation in bacteria", FEMS Microbiology Reviews **23**, 353–370 (1999).

[24]J. E. Mott, J. L. Galloway, and T Platt, "Maturation of escherichia coli tryptophan operon mrna: evidence for 3' exonucleolytic processing after rho-dependent termination", eng, The EMBO journal **4**, 1887 (1985).

[25]M. P. Hui, P. L. Foley, and J. G. Belasco, "Messenger rna degradation in bacterial cells", eng, Annual Review of Genetics **48**, 537–559 (2014).

[26]J. Belasco, "All things must pass: contrasts and commonalities in eukaryotic and bacterial mrna decay", English, Nature Reviews Molecular Cell Biology **11**, 467–478 (2010).

[27]G. A. Mackie, "Ribonuclease e is a 5-end-dependent endonuclease", Nature **395**, 720 (1998).

[28]"Differential control of the rate of 5-end-dependent mrna degradation in escherichia coli", eng, Journal of Bacteriology **194**, 6233 (2012).

[29]J. R. J. R. Norris, *Markov chains*, eng, Cambridge series on statistical and probabilistic mathematics ; 2 (Cambridge University Press, Cambridge, 1998).

[30]P. A. Gagniuc, *Markov chains : from theory to implementation and experimentation*, eng (John Wiley Sons, Hoboken, NJ, 2017).

[31]F. Spitzer, "Interaction of markov processes", eng, Advances in Mathematics **5**, 246–290 (1970).

[32]S. Katz, J. Lebowitz, and H. Spohn, "Nonequilibrium steady states of stochastic lattice gas models of fast ionic conductors", eng, Journal of Statistical Physics **34**, 497–537 (1984).

[33]S Scott and J Szavits-Nossan, "Power series method for solving tasep-based models of mrna translation", eng, Physical Biology **17**, 015004 (2019).

[34]N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, and J. S. Weissman, "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling", eng, Science (New York, N.Y.) **324**, 218 (2009).

[35]J. Szavits-Nossan and M. R. Evans, "Dynamics of ribosomes in mrna translation under steady and non-steady state conditions", (2020).

[36]J. Szavits-Nossan, L. Ciandrini, and M. Romano, "Predicting the steady-state rate of mrna translation in protein biosynthesis", (2017).

[37]*Non-equilibirum statistical mechanics*, University of Edinburgh (2019).

[38]D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions", eng, Journal of Computational Physics **22**, 403–434 (1976).