Imperial College
London

Planning Report for [MEng Individual] Project

Translation models that incorporate slow codons and resource limitation in E.Coli

Author name(s):

Peter Sarvari

Supervisor(s):

Dr. Guy-Bart Stan

Submitted in partial fulfilment of the requirements for the award of BEng/MEng in Biomedical Engineering from Imperial College London

January 2017 Word count: 3526

Feedback box for project markers:

WHAT I liked about the report:
WHAT could/should be improved:

1. Project Specification

The project is aimed to revise and extend the elegant whole-cell model [1] developed by Joaquín Gutiérrez as part of his Master Thesis at Imperial College London. Joaquín's model is based on the whole-cell model developed by Weiße et al. [2], but is extended to include polyribosomes and different elongation rates for different codons. The project is aimed to revise four specific points in the model, which are

- What percent of the transcripts are covered in ribosomes when the initiation rate is higher than the elongation rate and there are no slow codons (defined in Section 3)?
- What percent of the transcripts are covered in ribosomes when the initiation rate is slower than the elongation rate on the slow codon?
- How the number of the translating ribosomes changes transiently and in the steady-state?
- How do the above points modify the protein production rate and the yield?

To revise these points, we assumed that the ribosome elongation rate does not come from an exponential distribution, but is a specific number. This idea was adapted from Page 14 in the thesis, which summarizes the idea the following way: "After a ribosome has bound, it moves along the mRNA with a velocity v_{el} ("ribosomal footprints" per second) and so the distance between two ribosomes on the mRNA will be given by $\delta = v_{el} * \Delta t$, where Δt is the time elapsed between the binding of the first ribosome and the binding of the second." Using this approach, empirical equations have emerged from computational modelling that govern ribosomal queuing and dynamics on the mRNA in case of arbitrary number of slow codons at arbitrary locations. The analytical solution of these equations has been determined and computed by a developed software called DynoRibo. Note that both the model in the thesis and the DynoRibo framework assumes no unbinding of the ribosomes from the mRNA. For more information DynoRibo see section 3 and the Appendix.

The project is aimed to extend the model in the thesis the following ways:

- Incorporate more than one slow codon in the translation model
- Calculate the growth rate in the computational model based on the differential equation for the growth rate (equation (5.5)) presented in the Outlook section of the thesis.

  The differential equation takes the cell's volume dependency on growth rate into account and assumes constant density of the cell rather than constant mass of the cell in the exponential growth phase.

- Investigate if -according the modified model – the chaperone protein fraction increases when a heterologous circuit with slow codon/codons at specific locations is translated.

  This is a crucial question, since the thesis confirmed the validity of the developed model by concluding that the chaperone transcript fraction is a marker of cellular burden (burden defined in [3]), so indeed the chaperone promotors are like the burden-sensitive promoters described in Ceroni et al [4]. However, the effect of heterologous transcripts with slow codons – causing excessive burden on the cell – on chaperone transcript fraction has not yet

been studied. It needs to be confirmed the chaperone transcript fraction increases also in this case.

Additional extensions (time-permitting) include:

- Investigate which resources have greater impact on the dynamics using the graph-based framework of Qian et al [5].
- The modelling framework could be extended to account for the finite quantity of RNA polymerases (as in [5], [6]). To start, we should make Eq (1.9) in the thesis dependent on RNA Polymerase.

  Both additional extensions were suggested in the Outlook section of the thesis.

2. Ethical Analysis

The model described in the thesis already gives comparable results to that of the computationally more expensive Ribosome Flow Models (RFM, [7]), furthermore it is not just a translation, but a whole cell model considering metabolic and enzymatic processes as well as cell growth. The model permits synthetic biologists to quickly come up with a design that results in higher protein yield, which makes the field of Synthetic Biology more efficient. This can result in many environmentally beneficial consequences, such as drop in the prices of medicines, cheaper diagnostic and environmental monitoring bacterial constructs (biosensors) [8]–[10] and their more ubiquitous presence in the developing part of the world to prevent progression of many diseases.

3. Literature Review

Slow codons are codons, where the translation elongation rate is decerased locally. This is mostly because the t-RNA transporting the amino acid coded for by the slow codon has a low concentration within the cytoplasm [11]. Ribosome queues have been observed on the mRNA and they have been attributed to the effect of slow codons [12]. Ribosome queues are disadvantageous to the cell as well as for us if we would like to make the cell produce heterologous proteins (as we do in many synthetic biological constructs). Sequestered ribosomes are wasting their time on the mRNA, not contributing to translation of, for example, ribosomal proteins, housekeeping genes, transporter enzymes and (if applicable) heterologous proteins. Hence, cell growth slows down [13], which in case of bacterial cells and synthetic biological circuit means that the protein yield drops massively. In addition, in some cases because of the sequestered ribosomes, the exogenous protein can be produced in such low quantities that its effect is not visible (e.g. fluorescent protein) leading to false conclusions. Furthermore, since slow codons and sequestered ribosomes are generally not considered during the mathematical modelling of synthetic biological constructs, many times the desired effect is only achieved in theory but not in practice. Ribosome sequestration is not common in the cell's own genome, since this would put extra burden on the cell, leading to evolutional disadvantages [14]. However, when a heterologous circuit is introduced in the cell, this effect can indeed happen and hence it is important to codon optimize heterologous circuits to ensure efficient translation [15], [16]. Our method proposes fundamental equations describing this queuing process of the ribosomes on the mRNA, and we validate these

empirically derived equations via computational methods. The DynoRibo software uses these equations and their analytical solution to calculate the ribosome concentration dynamics. It yields essentially the same results (in the 45 tested cases) as those we obtain from a much slower ribosome concentration simulation method not using the analytical solution to the equations, but sequential time increment (see Appendix, Section 9.1). Note that both the DynoRibo and the sequential time increment method makes 3 significant assumptions. First, that once the ribosome is bound to the mRNA, it cannot unbind. Second, that the initiation and elongation rate is not a probability distribution, but a number (as already mentioned in Section1). This number is chosen the be the average of the exponential distribution that determines the reaction rate. The validity of this assumption will be evaluated by comparing the steady-state results of the DynoRibo with that of the mean-field approximation [17] of the Ribosome Flow Model (RFM) [7]. The third assumption is that the resources needed for translation are abundant. Previous efforts have been made to study host-circuit interactions and obtain ribosome concentration dynamics resulting from slow codons and limited resources. The most successful one was developed by the Stan lab at Imperial College using a mean-field approximation of the extended Ribosome Flow Model, in which the initiation rate is a function of the currently available ribosomes (eRFM) [17]. The eRFM models the expectation of the translation process via nonlinear differential equations. These equations were obtained by taking the average of the 1D constrained random walk of the ribosomes on the mRNA. The drawback of the eRFM is that only the steady-state values (e.g. number of ribosomes sequestered on the transcripts) can be calculated easily. Other resource limitation models include the theoretically validated model proposed by Weiße et al [2], which does not consider polyribosomes, the elegant top-down approach of Erickson et [18], which does not consider a detailed translation model, however manages to capture bacterial growth laws, and the extensive analysis from Chen et al [6] and Qian et al [5], whose complex models include transcriptional resource limitation, but does not consider slow codons. Nevertheless, it is notable that Chen et al and Erickson et al also predicts proteome composition as a function of growth rate. In addition, Chen et al considers location of heterologous genes on the plasmid.

4. Implementation Plan

The real advantage of DynoRibo is that it returns the transient as well as the steady-state number of ribosomes sequestered on the mRNA in a quick and efficient way. The fundamental equations we use in this method describe the ribosome concentration time derivatives in time intervals, where it is constant. Hence there is absolutely no need to solve nonlinear differential equations and (as long as the number of slow codons is less than the number of codons) DynoRibo is expected to execute faster than the mRFM. However, DynoRibo, in its current form, is unable to capture the resource limitation in bacterial cells (and hence quantify the burden [3]), which we already mentioned in assumption 3. Therefore, the next step is to understand how the dynamics evolves when the available ribosome concentration, amino acid concentration and energy unit (ATP) in the cytoplasm changes over time, resulting in a change in the translation initiation rate and elongation rate. Since keeping track of amino acid concentration and ATP level would require modelling of metabolism, to keep the translation model simple, we can assume abundant ATP and amino acid levels, and first extend DynoRibo to incorporate the limited number of ribosomes. The blueprint of this extension in the case of one slow codon is shown in the Appendix, Section 9.2. The results of the extension are shown in Section 6. Since multiple resource limitation requires whole-cell modelling, it seems logical to try to integrate the equations in DynoRibo

into the whole cell model developed by Joaquín to obtain a comprehensive model considering resource limitation and multiple slow codons.

Time will be spent in January to carefully consider Assumption 2 and its implications. To implement the mRFM, the Python code developed by the authors of [13], [17] will be converted to Python 3 and simulated to give the number of free ribosomes in steady-state in a variety of test cases. If the results are within an acceptable range from that of DynoRibo, then the DynoRibo framework will be continued to be developed in February to include limited number of ribosomes for any number of slow codons. Simultaneously, the analytical equations used in the DynoRibo framework will be incorporated into the translation model presented in the thesis of Joaquín. Should assumption 2 turn out to be invalid, time will be spent to consider which (if any) equations should in Joaquín's model be replaced by those in the DynoRibo framework. It is important to try to combine DynoRibo with the whole cell model in order to achieve a united framework that combines the best of both individual frameworks: ability to handle multiple slow codons from DynoRibo and modelling of metabolic processes and resource limitation from the model in the thesis. The second extension to Joaquín's model outlined in Section 1 will be implemented from March on and the third from May on. These extensions are independent of the success of the merging of DynoRibo with the whole cell model, so no delay to this schedule is expected and therefore we do not provide any fallback schedule.

5. Evaluation

The project is evaluated to be successful if after the incorporation of the volume extension (second extension proposed in Section 1) into Joaquín's model, it recovers the linear relationship reported between growth rate and protein mass fractions as described by Scott et al [19]. This is the same method that was used by Weiße et al [2] to check the validity of their proposed mathematical model. If, in addition, this linear relationship is recovered with more realistic parameters values than it was before, then the project is considered to be particularly successful. It is a plus – but not essential for the success of the project – if the model is extended to incorporate more than one slow codon and if it is verified that the number of chaperone transcripts is indeed a biomarker of burden caused by heterologous transcripts with slow codons. In case the extension of Joaquín's model is not successful, the project is still believed to be successful if the assumptions made in DynoRibo are validated (as described in Section 4) and DynoRibo is extended to enable translation modelling in the limiting number of ribosomes case with multiple slow codons.

6. Preliminary Results

From the sequential time increment simulations (see Appendix, Section 9.1 for method and pseudocode) we observed that the steady-state ratio of the mRNA covered by the ribosomes (we will refer to this as ribosomal density as well as ribosome concentration) is given by the following formula in the case of one slow codon:

$$\frac{l + (L - l) * \dfrac{\dfrac{T_{el}}{T_{slow} - T_{el}}}{1 + \dfrac{T_{el}}{T_{slow} - T_{el}}}}{mRNA\ length} \qquad (1)$$

Where T_{el} is the reciprocal elongation rate, T_{slow} is the reciprocal elongation rate of the slow codon, l is the location of the slow codon and L is the length of the mRNA. If we vary T_{slow} and l, we can compute the heatmap for ribosomal density, which is shown on Figure 1. It is interesting to observe two things. First, a slow codon close to the 3' end of the mRNA significantly reduces the ribosomal density on the transcript. It has been hypothesized that this is the reason why homologous genes (optimized by evolution) tend to contain slow codons only close to the 3' end [1]. Second, given l, the slower the slow codon (the higher T_{slow} is), the less ribosome is sequestered on the mRNA. This is slightly unintuitive, but makes sense when we think about it: T_{slow} controls the flow of ribosomes into the second part of the mRNA after the slow codon. In the steady-state (if the slow codon is sufficiently slow, see below) ribosomal density before the slow codon is one. Hence the slower the ribosomes can flow into the second part of the mRNA, the fewer ribosomes are sequestered there and hence the ribosomal density on the mRNA decreases.

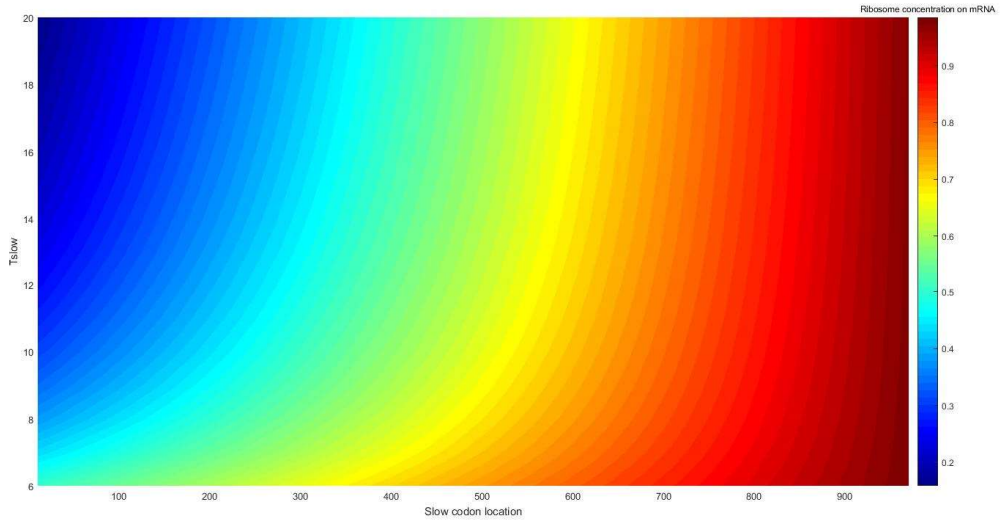


Figure 1: Heatmap for ribosome density ('concentration') on the mRNA

Equation (1) is an example of the equations that we included in DynoRibo to model the ribosome concentration dynamics on the mRNA. These equations describe both transient and steady-state concentration dynamics with arbitrary number of slow codons. A recursive mathematical method has been devised to solve this system of equations and with that uncover an underlying queuing process that affects the ribosomes on the mRNA because of the slow codons. This solution method was implemented in MATLAB and the resulting ribosomal density dynamics were compared to results of the sequential time increment method. They were the same (within a small numerical error), which confirmed the

validity of the governing equations as well as the mathematical method used to solve them. In addition, DynoRibo proved to be a more time-efficient method than the sequential time increment. Please see the Appendix for the comparison of the results (Section 8.5) from the two methods and for the pseudocodes for each of them (Section 8.4 and Section 9.1).

It was mentioned in Section 4 that an extension of DynoRibo has been implemented, where the limiting number of ribosomes are considered. Resource limitation acts as a retroactive coupling between different translation processes [1] and creates a stabilizing feedback for ribosomal density on the mRNA. The results for the parameter values in Table 1 can be seen on Figure 2. On Figure 2, the analytical result refers to the solution of the differential equation (A3) (the equation is derived in the Appendix, Section 9.2), the simulation result refers to that obtained by the sequential time increment method (see Appendix, section 9.1) and the slow codon disappearance condition is equivalent to the condition for a slow codon not to be sufficiently slow, which is:

$$T_{slow} < T_{el} + T_{ini} \qquad (2)$$

If the slow codon is not sufficiently slow, then (in the DynoRibo model) none of the ribosomes are held up by the previous ribosome on the mRNA. In other words, the slow codon is not causing queue formation, so we say that it is not sufficiently slow. In cases when because of the changing initiation rate, a slow codon stops being sufficiently slow, we say that the slow codon disappeared (does not affect the dynamics any more). It is logical that if equation (2) is fulfilled, the slow codon is not causing congestion: the following ribosome will never reach the previous one, because it has a time lag of T_{el} + T_{ini} to reach the same ribosome place: T_{el} because the RBS needs to be emptied and it takes T_{el} for the previous ribosome to move to the next place, and T_{ini}, because this is how much time the initiation process takes. In equation (A3) *num* is the number of translating ribosomes, which is the ribosomal density multiplied by the average transcript length. For the underlying assumptions and more detail, please see the Appendix, Section 9.2.

$$\dot{num} = \frac{1}{T_{el} + \frac{T_{ini}(0) * R_{total}}{R_{total} - m * num(t)}} \qquad (A3)$$

T_{el}	R_{total}	l	T_{slow}	m	$T_{ini}(0)$
2	10000	50	30	400	5

Table 1: Parameters used for Figure 2. Average transcript length is 200 ribosome places (~2000 codons), R_{total} is the number of available ribosomes for heterologous transcripts, l is slow codon location from 3' end, m is the number of heterologous transcripts and $T_{ini}(0)$ is initial reciprocal initiation rate.

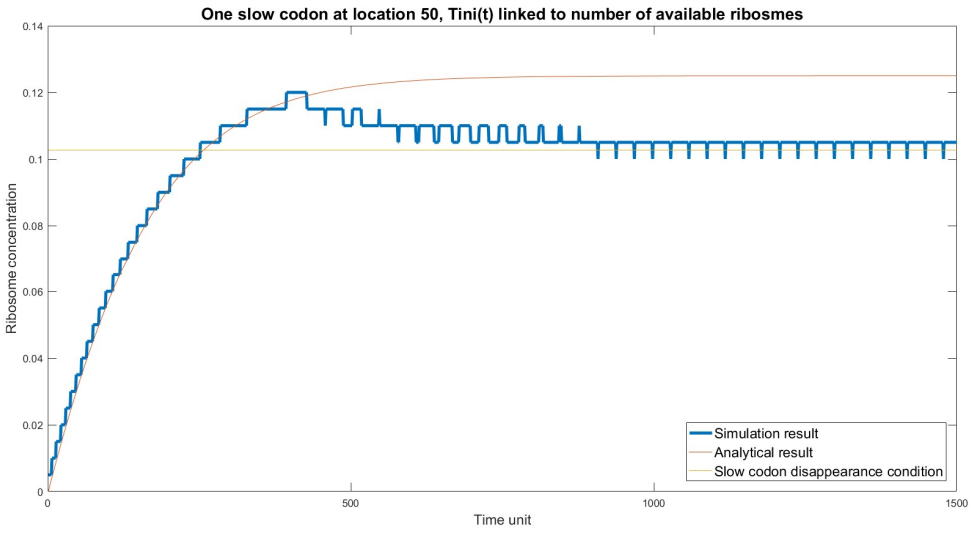


Figure 2: Demonstration of the stabilizing negative feedback that steers the ribosome concentration on the mRNA such that the resulting reciprocal initiation rate corresponds to the slow codon disappearance condition

7. References

[1] J. Gutiérrez, "Optimal synthetic biology designs for simultaneous increase in productivity and growth of engineered bacterial cells," Ludwig-Maximilians University, München, 2017.

[2] A. Y. Weiße, D. A. Oyarzún, V. Danos, and P. S. Swain, "Mechanistic links between cellular trade-offs, gene expression, and growth," *Proc. Natl. Acad. Sci.*, vol. 112, no. 9, pp. E1038–E1047, 2015.

[3] O. Borkowski, F. Ceroni, G. B. Stan, and T. Ellis, "Overloaded and stressed: whole-cell considerations for bacterial synthetic biology," *Curr. Opin. Microbiol.*, vol. 33, pp. 123–130, 2016.

[4] F. Ceroni *et al.*, "Burden-driven feedback control of gene expression," *bioRxiv*, pp. 1–17, 2017.

[5] Y. Qian, H. H. Huang, J. I. Jiménez, and D. Del Vecchio, "Resource Competition Shapes the Response of Genetic Circuits," *ACS Synth. Biol.*, vol. 6, no. 7, pp. 1263–1272, 2017.

[6] C. Liao, A. E. Blanchard, and T. Lu, "An integrative circuit–host modelling framework for predicting synthetic gene network behaviours," *Nat. Microbiol.*, vol. 2, no. December, 2017.

[7] S. Reuveni, I. Meilijson, M. Kupiec, E. Ruppin, and T. Tuller, "Genome-scale analysis of translation elongation with a ribosome flow model," *PLoS Comput. Biol.*, vol. 7, no. 9, 2011.

[8] A. J. Webb *et al.*, "A protease-based biosensor for the detection of schistosome cercariae," *Sci.*

Rep., vol. 6, no. November 2015, pp. 1–14, 2016.

[9] J. Stocker *et al.*, "Development of a set of simple bacterial biosensors for quantitative and rapid measurements of arsenite and arsenate in potable water," *Environ. Sci. Technol.*, vol. 37, no. 20, pp. 4743–4750, 2003.

[10] D. Merulla *et al.*, "Bioreporters and biosensors for arsenic detection. Biotechnological solutions for a world-wide pollution problem," *Curr. Opin. Biotechnol.*, vol. 24, no. 3, pp. 534–541, 2013.

[11] T. Ikemura, "Codon usage and tRNA content in unicellular and multicellular organisms.," *Mol. Biol. Evol.*, vol. 2, no. January, pp. 13–34, 1985.

[12] N. Mitarai and S. Pedersen, "Control of ribosome traffic by position-dependent choice of synonymous codons," *Phys. Biol.*, vol. 10, no. 5, 2013.

[13] F. Ceroni, R. Algar, G. B. Stan, and T. Ellis, "Quantifying cellular capacity identifies gene expression designs with reduced burden," *Nat. Methods*, vol. 12, no. 5, pp. 415–418, 2015.

[14] T. Tuller *et al.*, "An evolutionarily conserved mechanism for controlling the efficiency of protein translation," *Cell*, vol. 141, no. 2, pp. 344–354, 2010.

[15] E. M. Novoa and L. Ribas de Pouplana, "Speeding with control: Codon usage, tRNAs, and ribosomes," *Trends Genet.*, vol. 28, no. 11, pp. 574–581, 2012.

[16] T. E. F. Quax, N. J. Claassens, D. Söll, and J. van der Oost, "Codon Bias as a Means to Fine-Tune Gene Expression," *Mol. Cell*, vol. 59, no. 2, pp. 149–161, 2015.

[17] R. J. R. Algar, T. Ellis, and G. B. Stan, "Modelling essential interactions between synthetic genes and their chassis cell," *Proc. IEEE Conf. Decis. Control*, vol. 2015–Febru, no. February, pp. 5437–5444, 2014.

[18] D. W. Erickson, S. J. Schink, V. Patsalo, J. R. Williamson, U. Gerland, and T. Hwa, "A global resource allocation strategy governs growth transition kinetics of Escherichia coli," *Nature*, vol. 551, no. 7678, pp. 119–123, 2017.

[19] M. Scott, E. M. Mateescu, Z. Zhang, and T. Hwa, "Interdependence of Cell Growth Origins and Consequences," *Science (80-.).*, vol. 330, no. November, pp. 1099–1102, 2010.

8. Appendix

8.1. Abstract

In this report, we propose a fast method for translation modelling that given the translation initiation rate, elongation rate, slow/fast codon elongation rate and location, returns the ribosome concentration dynamics on the mRNA using analytical equations. We assume that (1) once the ribosome is bound to the mRNA, it cannot unbind, (2) the ribosome moves along the mRNA with the same speed as long as the elongation rates are the same and (3) there is no translation resource limitation. In addition, we show in the Appendix how DynoRibo can be extended to approximate the resource limitation case to get rid of assumption (3). The advantage of DynoRibo is that it uncovers fundamental queuing equations,

incorporates slow codons and provides a fast and efficient analytical translation modelling method that can be extended to model cases when translation resources (e.g. ribosomes) are not abundant. The equations discussed here can significantly help in modelling any systems - or synthetic biological phenomenon that relies on knowing the translation dynamics and wishes to incorporate the effect of slow codons. Furthermore, this method and the equations can be helpful for studying and modelling any queuing process in general.

8.2. The empirical observations

Note that we define ribosome concentration on the mRNA as the number of ribosomes on the mRNA divided by the length of the mRNA. From empirical observations of the results of the sequential time increment simulation method, we concluded the following: whenever the first ribosome on the mRNA reaches a slow codon that is sufficiently slow (see later), ribosomes on the 5'end of the slow codon start becoming congested, eventually reaching full sequestration (at steady-state). We also find that immediately after the slow codon is reached by the first ribosome, the local concentration of the ribosomes in an arbitrarily small neighbourhood starting from the 5' side of the slow codon is 1 (meaning that all ribosome places are occupied by ribosomes, full sequestration). This phenomenon starts spreading backwards eventually reaching the 5'end of the mRNA, when the concentration of the ribosomes on the 5'side of the slow codon becomes one and the steady-state is reached. We defined the speed at which this local concentration of 1 spreads backwards all the way to the 5' end as the **backwards congestion propagation** velocity. Now, if there are more than 1 slow codon, the case is more complicated. First of all, the congestion propagation doesn't necessarily reach the 5' end, but the previous slow codon. Then it increases the backward congestion propagation speed of that slow codon if its backward congestion propagation process is still ongoing. Let's have a look at two examples to better understand the terms and definitions. Figure 1 shows a three slow codon case, where the time it takes for the backward congestion process to finish is longer, the further the slow codon is from the 5'end.

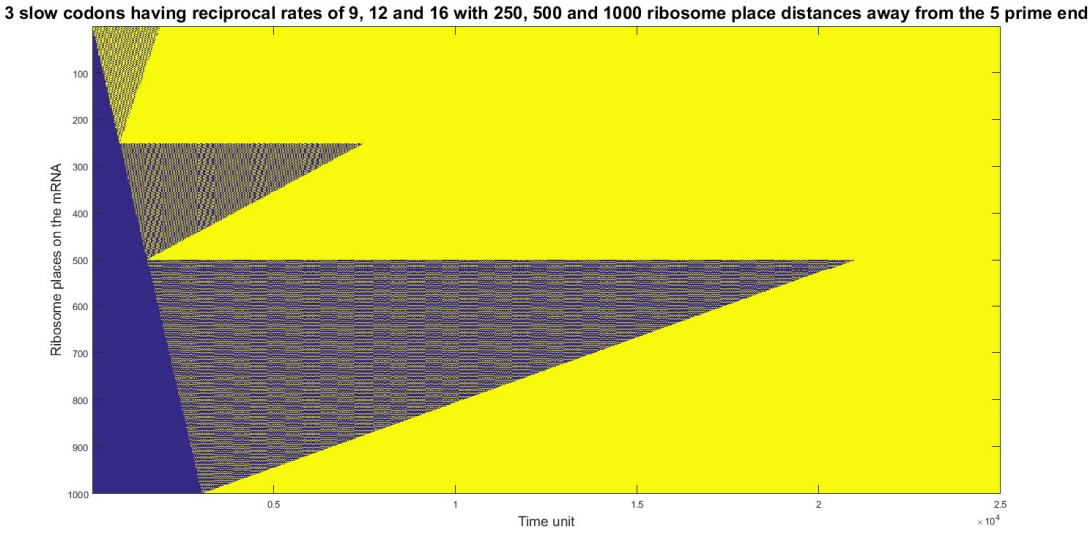


Figure 1- Blue is empty and yellow is filled

It is easy understand that the backward congestion propagation rates are the three positive slopes starting from slow codon locations. Figure 2 shows a case when the backward congestion starting from the third slow codon reaches a previous (2nd) slow codon before the congestion propagation process of that slow codon is finished. It can be seen that as soon as this happens, the backward congestion propagation rate of the 2nd slow codon is increased.

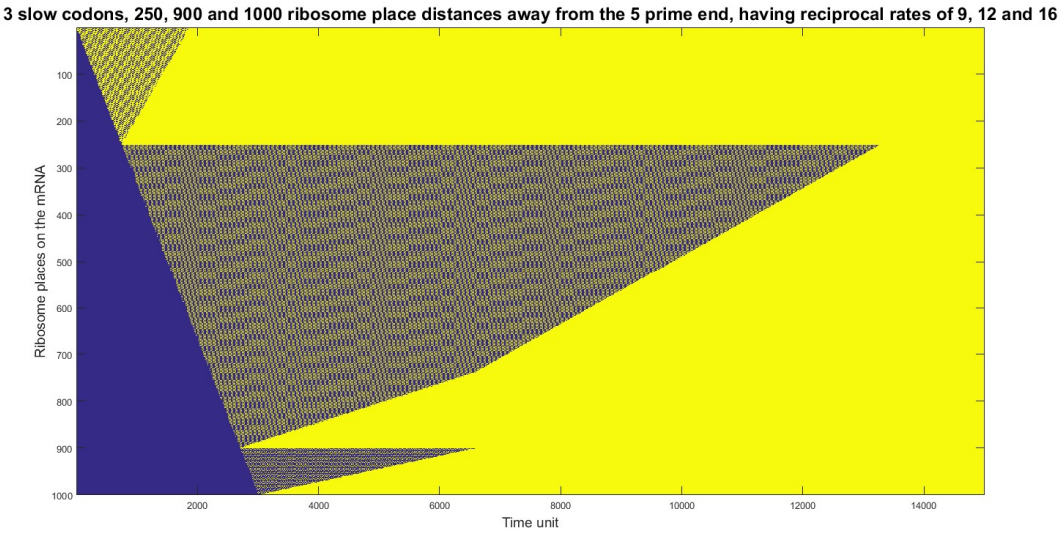


Figure 2- Blue is empty, yellow is filled

However, to determine the concentration dynamics, it is not enough to know the backward congestion propagation rates. This is because from these rates, we can work out when the slopes (for example in Figure 3) change, but we cannot work out the slopes themselves, since we do not know how much the congestion propagation changes the concentration of ribosomes on the mRNA region. We do know that the concentration changes to 1, but we do not know from what value it changes to 1.

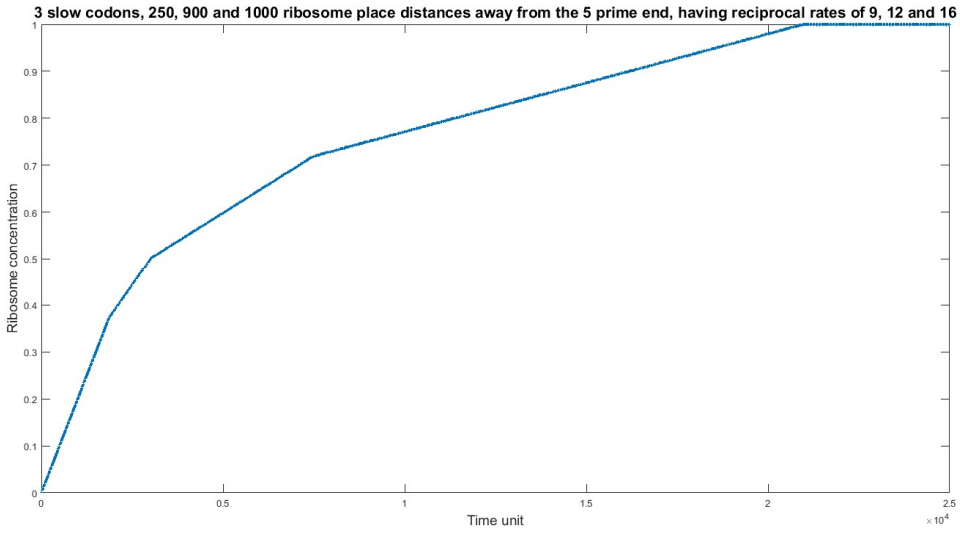


Figure 3- Ribosome concentration on the mRNA with respect to time (corresponding to the same case as Figure 1)

To understand from what value the concentration changes to 1, we need to define a term called posterior concentration. Posterior concentration is defined for each slow codon and refers to the local concentration in an arbitrarily small neighbourhood (neighbourhood expanding with time) after the first ribosome reaches the slow codon and until it reaches the next one (when this local concentration neighbourhood now reaching up to the next slow codon starts contracting because of the backward congestion propagation process of that slow codon).

8.3. The Model

Note that in the equations below, T_{ini} is the reciprocal initiation rate, T_{el} is the reciprocal normal elongation rate (neither slow, neither fast codons) and $T_{slow,n}$ is the reciprocal elongation rate of the nth slow codon. We have made many observations of the sequential time increment simulation and came up with a formula for the posterior concentration of a slow codon:

$$p_n = \frac{\frac{T_{el}}{T_{slow,n} - T_{el}}}{1 + \frac{T_{el}}{T_{slow,n} - T_{el}}} = \frac{T_{el}}{T_{slow,n}}$$

But what if there are no slow codons before? Then the posterior concentration ($posterior_0$) is essentially the slope of the concentration vs. time plot in the no slow codon case, which is:

$$p_0 = \frac{\frac{T_{el}}{T_{ini}}}{1 + \frac{T_{el}}{T_{ini}}} = \frac{T_{el}}{T_{ini} + T_{el}}$$

Hence the slope (in the plot of concentration versus time) is

$$\dot{conc} = \sum_{n=0}^{N-1} (1 - p_n) * v_{cg,n+1}$$

Where the backwards congestion propagation rate of slow codon n is

$$v_{cg,n} = \frac{T_{slow,n} - T_{slow,n-1}}{(T_{slow,n-1} - T_{el}) * T_{slow,n}}$$

And N is the number of slow codons.

The above equation for the slope is only true if

- The first ribosome reached the end of the mRNA, since this is needed to ensure that the concentration change solely comes from the backward congestion propagation processes of the slow codons. This condition can be written as

$$t > T_{total}$$

Where T_{total} is

$$\sum_{n=1}^{L-1} T_n$$

T_n is the reciprocal elongation rate on the nth codon and L is the length of the mRNA.

- None of the backward congestion propagation processes have stopped (otherwise those, which have stopped need to be taken out of the sum) or merged to another one (merger can be seen on Figure 2, around 6500 time units and further discussed in Proposition II).

- We define slow codons to be sufficiently slow if $T_{slow,1} > T_{ini} + T_{el}$ and if $T_{slow,n+1} \geq T_{slow,n}$

We observe that if the 3rd restriction is not true - some slow codons are not sufficiently slow -, they do not change the ribosome concentration dynamics (apart from increasing T_{total}), hence we can disregard them for slope calculations, so from now on we will simply not consider these as slow codons.

We have some idea already how to deal with cases when $t > T_{total}$, but we have not discussed anything regarding $t < T_{total}$, so let's do that! For this, we use two propositions: one that we came up with during the study of a single slow codon case scenario (Proposition I) and one that became obvious when we were trying to understand how the backward congestion propagation slope changes when the congestion propagation from a slow codon reaches the previous slow codon (Figure 2, around 6500 time units).

8.3.1. Proposition I

We claim that the increasing concentration due to backward congestion propagation on the 5' side of the slow codon and the increasing concentration (but this increase is smaller compared to the no slow codon case, since in the no slow codon case the posterior is $\frac{T_{el}}{T_{ini}+T_{el}}$, but after the slow codon, the posterior is $\frac{T_{el}}{T_{slow,n}}$, which is always smaller, as the slow codons considered here are sufficiently slow) on the 3' side of the slow codon summed together is exactly the increasing concentration (e.g. the slope of the curve) in the case when there is no slow codon.

Proof /using mathematical induction/
For $n = 1$, $T_{slow,1}$ denoted by T_{slow}

We need to confirm that

$$\left(1 - \frac{\frac{T_{el}}{T_{ini}}}{1 + \frac{T_{el}}{T_{ini}}}\right) * \frac{\frac{T_{slow} - (T_{el} + T_{ini})}{T_{ini} * T_{slow}}}{mRNA\ length} + \frac{\frac{T_{el}}{T_{slow} - T_{el}}}{1 + \frac{T_{el}}{T_{slow} - T_{el}}} * \frac{v_{el}}{mRNA\ length}$$

$$= \frac{\frac{T_{el}}{T_{ini}}}{1 + \frac{T_{el}}{T_{ini}}} * \frac{v_{el}}{mRNA\ length}$$

We see that when we write equations where the terms are concentrations, we can easily get rid of mRNA length, so in further proofs, we will omit this term.

Using $v_{el} = \frac{1}{T_{el}}$, the second term simplifies to

$$\frac{1}{T_{slow}}$$

And the third term simplifies to

$$\frac{1}{T_{el} + T_{ini}}$$

The first term is

$$\frac{T_{ini}}{T_{ini} + T_{el}} * \frac{T_{slow} - (T_{el} + T_{ini})}{T_{ini} * T_{slow}} = \frac{T_{slow} - (T_{el} + T_{ini})}{(T_{ini} + T_{el}) * T_{slow}}$$

The third minus the second term:

$$\frac{1}{T_{el}+T_{ini}}-\frac{1}{T_{slow}}=\frac{T_{slow}-(T_{el}+T_{ini})}{(T_{ini}+T_{el})*T_{slow}}$$

Which proves our proposition.

For $n \geq 2$

Here we will prove that sum of the concentration change due to the backward congestion propagation from slow codon (n+1) and the elongation after that slow codon equals the concentration change resulting from the elongation after the previous (n^{th}) slow codon. So, in other words until the congestion propagation from slow codon (n+1) finishes or the elongation reaches the next slow codon, looking at the dynamics, it seems as if the $(n+1)^{th}$ slow codon was non-existing. This is going to be useful, since we have proven this for all n. Hence, starting from the last slow codon, we see that it does not matter in terms of the dynamics, then going backwards, one-by-one, we see that none of the slow codons matter (obviously until the first ribosome reaches the end or any[1] of the backward congestion propagation processes finish).

$$\left(1-\frac{T_{el}}{T_{slow,n}}\right)*\frac{T_{slow,n+1}-T_{slow,n}}{(T_{slow,n}-T_{el})*T_{slow,n+1}}+\frac{T_{el}}{T_{slow,n+1}}*v_{el}=\frac{T_{el}}{T_{slow,n}}*v_{el}$$

The first term simplifies to:

$$\frac{T_{slow,n+1}-T_{slow,n}}{T_{slow,n+1}*T_{slow,n}}$$

The second term simplifies to:

$$\frac{T_{slow,n}}{T_{slow,n+1}*T_{slow,n}}$$

And the third term simplifies to

$$\frac{1}{T_{slow,n}}$$

Which is indeed the sum of the first two terms, hence the claim.

Further observations regarding backward congestion propagation velocities

We observe that when the backward congestion diffusion from a later (further from the 5' end) slow codon ($n+1^{th}$) reaches the next slow codon (n^{th}), that slow codon's backward congestion diffusion dynamics changes (increases) the following way:

$$v_{cg,n+1,n}=\frac{T_{slow,n+1}-T_{slow,n-1}}{(T_{slow,n-1}-T_{el})*T_{slow,n+1}}$$

And in general

$$v_{cg,x,y}=\frac{T_{slow,x}-T_{slow,y-1}}{(T_{slow,y-1}-T_{el})*T_{slow,x}}$$

[1] we will see later (in Proposition II) that it will only change if the first ribosome reaches the 3' end or the backward congestion propagation from the first slow codon reaches the 5' end

Where x>y and x denotes the index of the slow codon where the backwards congestion propagation starts from and y denotes the index of the slow codon, which is the closest to the propagating end of the congestion.

8.3.2. Proposition II

It seems that this relation manipulates the slopes in a way that they do not change when the backward congestion diffusion from the later slow codon reaches the previous slow codon.

Proof

For a specific case when the backward congestion diffusion from a later (further from the 5' end) slow codon (n+1th) reaches the next (in backwards direction, so previous) slow codon (nth):

Slope before the merger of the queues:

$$v_{cg,n} * (1 - p_{n-1}) + v_{cg,n+1} * (1 - p_n)$$

$$\frac{T_{slow,n} - T_{slow,n-1}}{(T_{slow,n-1} - T_{el}) * T_{slow,n}} * \left(1 - \frac{\frac{T_{el}}{T_{slow,n-1} - T_{el}}}{1 + \frac{T_{el}}{T_{slow,n-1} - T_{el}}} \right) + \frac{T_{slow,n+1} - T_{slow,n}}{(T_{slow,n} - T_{el}) * T_{slow,n+1}}$$

$$* \left(1 - \frac{\frac{T_{el}}{T_{slow,n} - T_{el}}}{1 + \frac{T_{el}}{T_{slow,n} - T_{el}}} \right)$$

With the new proposal, the new slope is (which should equal to the sum above by Proposition II):

$$v_{cg,n+1,n} * (1 - p_{n-1})$$

$$\frac{T_{slow,n+1} - T_{slow,n-1}}{L * (T_{slow,n-1} - T_{el}) * T_{slow,n+1}} * \left(1 - \frac{\frac{T_{el}}{T_{slow,n-1} - T_{el}}}{1 + \frac{T_{el}}{T_{slow,n-1} - T_{el}}} \right)$$

The first term in the sum simplifies to

$$\frac{T_{slow,n} - T_{slow,n-1}}{T_{slow,n} * T_{slow,n-1}}$$

The second term simplifies to

$$\frac{T_{slow,n+1} - T_{slow,n}}{T_{slow,n+1} * T_{slow,n}}$$

The new slope according to the observation simplifies to:

$$\frac{T_{slow,n+1} - T_{slow,n-1}}{T_{slow,n-1} * T_{slow,n+1}}$$

Adding up the first and second term after making their denominator common trivially gives the fraction above.

For a general case, we need to prove:

$$v_{cg,n+1,m} * (1 - p_{m-1}) + v_{cg,m-1} * (1 - p_{m-2}) = v_{cg,n+1,m-1} * (1 - p_{m-2})$$

Using the definitions of the backward congestion propagation speed and the posterior concentration, the first term becomes:

$$\frac{T_{slow,n+1} - T_{slow,m-1}}{T_{slow,n+1} * T_{slow,m-1}}$$

The second term becomes:

$$\frac{T_{slow,m-1} - T_{slow,m-2}}{T_{slow,m-1} * T_{slow,m-2}}$$

And the right-hand side becomes:

$$\frac{T_{slow,n+1} - T_{slow,m-2}}{T_{slow,n+1} * T_{slow,m-2}}$$

Adding up the first and second term after making their denominator common trivially gives the right-hand side, hence the claim.

Before proceeding with the model, let us first define some terms: **stalled queue** is a cluster of ribosomes, such that it has local concentration 1 (occupying all ribosome places) all the way from the 5' end of the mRNA up to a slow codon, whereas **queue** is a cluster of ribosomes, such that it has local concentration 1, it starts at a slow codon and is expanding towards the 5' end of the mRNA.

8.3.3. Analysis for the case $t < T_{total}$

The slope at t=0 on the concentration vs. time plot has to be the slope equivalent to the single slope obtained in the case of no slow codons, since it is impossible to reach the first slow codon in zero time. That slope is $p_0 * \frac{v_{el}}{L}$, where L is the mRNA length. Using Proposition I, we know that when $t < T_{total}$, the slope will decrease by a term functionally equivalent to the first term in the left hand side in Proposition I. The decrease in the slope corresponds to the halting of the backwards congestion propagation process. Because of Proposition II, this process only really halts, when there is no such process closer to the 5'end of the mRNA, so in other words, when the backward congestion propagation reaches a stalled queue, or reaches the 5' end (essentially forming the first stalled queue). So, every time such a backward congestion process finishes, the slope decreases by

$$p_{y-1} * \frac{v_{cg,x,y}}{L}$$

We define this quantity of decrease as **equation (1)**, where $v_{cg,x,y}$ is the backward congestion propagation rate of the queue that just halted. X>y, and x denotes the index of the slow codon where the backwards congestion propagation starts from and y denotes the index of the slow codon, which is inside the congestion and is the closest to the propagating end of the congestion (we define such slow codon as the **queuefront**).

8.3.4. Analysis for the case $t > T_{total}$

Now let's find an equation for the change of ribosome concentration with respect to time in the case $t > T_{total}$ without any restrictions. We define $v_{cg,queuestart:x}(t)$ to be $v_{cg,x,y(t)}$, where y is the queuefront. $v_{cg,queuestart:x}$ is constant in time, while the backward congestion propagation has not reached a **queue change point** defined as either reaching

- The 5' end of the mRNA (**type I**), or
- A stalled queue (**type II**), or
- A ribosome queue closer to the 5' end (**type III**).

In the first two cases, $v_{cg,queuestart:x}$ becomes zero and the queue starting from x stalls completely. In the third case, it merges with another queue and the queuefront of the queue starting from x changes to the queuefront of the queue it merged with.

Using the above definitions, we can define **equation (2)**:

$$\dot{conc}(t) = \sum_{for\ all\ queue\ starting\ points,x} \left(1 - p_{qf,x-1}\right) * v_{cg,queuestart:x}(t)$$

Where qf,x is the queuefront of the queue starting from x. With equation (1) and equation (2) we are now able to describe the ribosome concentration dynamics on the mRNA for any time, t.

8.4. Methods.

8.4.1. Algorithm used in queue_reduce.m

We implemented a function in Matlab called queue_reduce.m, which keeps track of the existing queues, their starting points and their queuefronts. It calculates the time needed for each queue to reach a queue change point, by keeping track of the distances the congestions have propagated from their starting slow codons. Knowing that distance, we can calculate the remaining distance to a queue change point and knowing the backward congestion propagation for the queue, we can calculate the time needed to reach a queue change point. Because of Proposition II, the function only records these times if they contribute to changing the slope (in the concentration vs. time plot), so if the queue change point is of type I or type II. Otherwise (type III queue change point), the function recalculates the queuefronts and their backward congestion propagation speeds (which is – for a queue starting at slow codon x - $v_{cg,queuestart:x}$) and the distances travelled from the starting slow codons for the two merging queues are added.

Obviously, it is not effective to calculate the time needed for each queue to reach a queue changing point, since the first queue to reach such a point may change the dynamics of another queue (if it merges with it), which in turn, may change dynamics of other queues later. However, it is very easy to calculate the time when the next queue change point occurs for any queue (since it is the next, no other queue can modify its dynamics up to that point) and we define this quantity as val. Then if that queue (e.g. starting from x) merges with another queue (starting from z, x>z),

- The distance travelled from the starting slow codon for each queue is incremented by
  - val times their respective backward congestion propagation rate if the backward congestion process started already (In other words, if the time is bigger than the time needed to reach the slow codon (let's say x), where the backward congestion propagation would start from. Mathematically: $t > \sum_{n=1}^{d_x-1} T_n$ where T_n is the reciprocal elongation rate on the n^{th} codon and d_x is the distance from the 5' end to the x^{th} slow codon)
  - (val + t - $\sum_{n=1}^{d_x-1} T_n$) times their respective backward congestion propagation rate if $t < \sum_{n=1}^{d_x-1} T_n < t + val$
  - zero if $\sum_{n=1}^{d_x-1} T_n > t + val$
- Time is incremented by val
- the new backward congestion propagation rate is calculated for the merged queue (starting from x)

- the backward congestion propagation rate for the old queue (starting from z) is discarded from the array storing the congestion propagation rate values
- the distances travelled so far for the queue starting from z and for the queue starting from x are added up and stored as the distance travelled for the new merged queue (from x)
- the distance travelled from z for the old queue is discarded from the array storing the distances travelled for queues
- the time to reach a new queue changing point for the merged queue (starting from x) is calculated using the current dynamics (this is for sure the correct time only if this queue's time to reach a queue changing point is the smallest among all queues as discussed earlier)
- the time to reach a queue changing point from z for the old queue is discarded from the array storing such times for queues
- The array keeping track of current and past queuefronts for the old queue starting from z is Minkowski added to the array keeping track of current and past queuefronts for the queue starting from x

  Technical detail: we store these arrays in a cellular array. This time, we do not delete the array corresponding to the old queue starting from z from the cellular array. We keep track which elements of the cellular array are starting points for queues via an array called auxiliary_array. This also helps us find $T_{slow,x}$ needed for the calculation of $v_{cg,queuestart:x}$. The queuefront is found for the queues by finding the minimum element in the array indexed by the auxiliary_array (vectorization) in the cellular array keeping track of current and past queuefronts. When two queues merge, the element that is not corresponding to the starting slow codon of the merged queue is deleted from the auxiliary array. So, in the example above, z would be deleted from the auxiliary_array, since there are no queues starting from z any more.

If instead of type III, the queue change point is type I or type II, no other queue's dynamics changes. The only modification we do in this case is

- The distance travelled from the starting slow codon for each queue is incremented by
  - val times their respective backward congestion propagation rate if the backward congestion process started already (time is bigger than the time needed to reach the slow codon (let's say x), where the backward congestion propagation would start from. Mathematically: $t > \sum_{n=1}^{d_x-1} T_n$ where T_n is the reciprocal elongation rate on the nth codon and dx is the distance from the 5' end to the xth slow codon
  - (val + t - $\sum_{n=1}^{d_x-1} T_n$) times their respective backward congestion propagation rate if $t < \sum_{n=1}^{d_x-1} T_n < t + val$
  - zero if $\sum_{n=1}^{d_x-1} T_n > t + val$
- Time is incremented by val
- Time is appended to an array called change_times, which keeps track of the times when the slopes change
- The first element of the array storing the congestion propagation rate values is deleted
- The first element from the array storing the distances travelled for queues is deleted

- The first element from the array storing the times to reach a queue changing point for the queues is deleted
- The first element of the auxiliary_array (see Technical detail above) is deleted
  The reason we can simply always take away the first element in the type I and type II cases is that by definition if a queue becomes a stalled queue, it was the closest queue to the 5' end and the values in all these arrays were initialized corresponding to slow codons numbered from the 5' end

8.4.2. Algorithm used in slow_main.m

In slow_main.m, we initialize the arrays used by queue_reduce.m, and introduce a new array, called slopes, which first only contains the initial slope at t = 0. Then we implement a while loop, where we call queue_reduce.m until auxiliary_array is not empty. If the current time has been appended to change_times (type I or type II queue change point occurred) then depending on whether time if smaller or bigger than Ttotal, we implement (previous_slope_value - equation (1)) or equation (2) to calculate the new slopes and add them to the array storing the slopes. Note that since Matlab indexing starts from 1, but we have p_0, all the posterior concentration indices have been incremented by 1 in equation (1) and (2). Slow_main.m returns change_times and slopes and knowing these quantities, we can easily construct a plot describing the ribosome concentration on the mRNA with respect to time.

Inputs

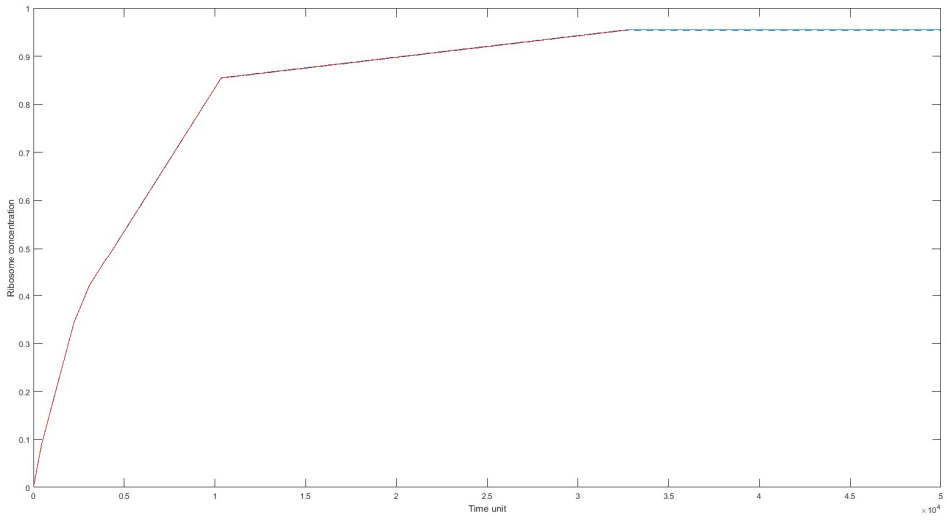
The inputs to slow_main.m are two arrays and three scalars. Tslow is the reciprocal elongation rate of slow (or fast) codons. Distance describes the distances of these codons from the 5' end of the mRNA. Tel, Tini and L store the reciprocal of the normal elongation rate, the reciprocal of the initiation rate and the length of the mRNA.

Note that since one ribosome place is approximately 10 codons long [17], the elements in Tslow are really the summed reciprocal elongation rates for 10 codons that are together defined as a slow (bunch of) codon(s). For the same reason, Tel should be set to ten times the normal reciprocal elongation rate and L should be the mRNA length divided by 10.

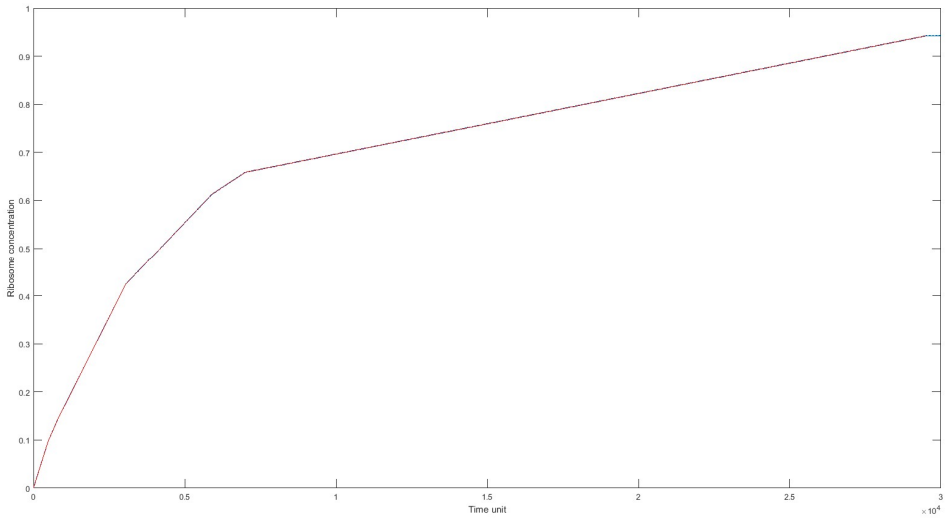
8.5. Results

We have come up with 45 rigorous testing cases in which we both plotted the result of the simulation using sequential time increment (blue line) and the result from the DynoRibo software (red line) on the same graph. We show the results for 4 out of those 45 cases here.

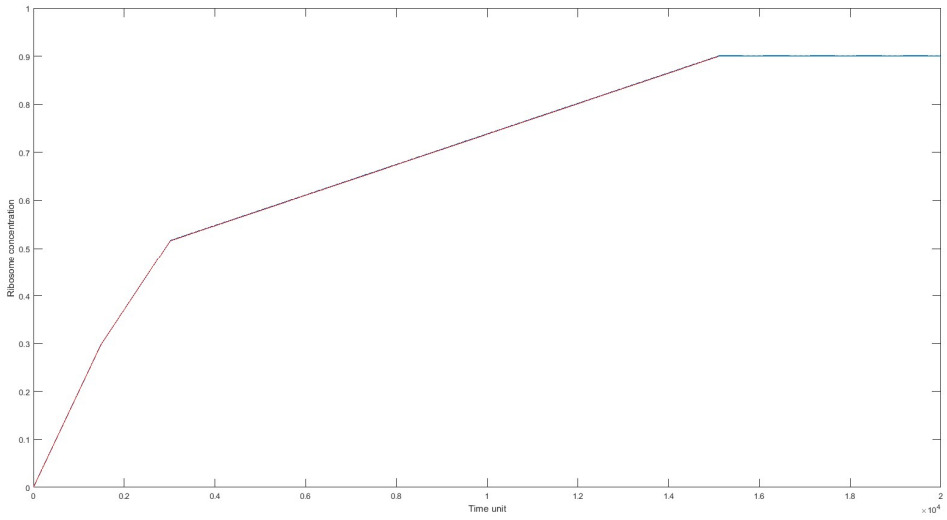
distances = [45 195 678 750 800 950];
Tslow = [7 11 20 24 28 32];
Tel = 3;
Tini = 2;



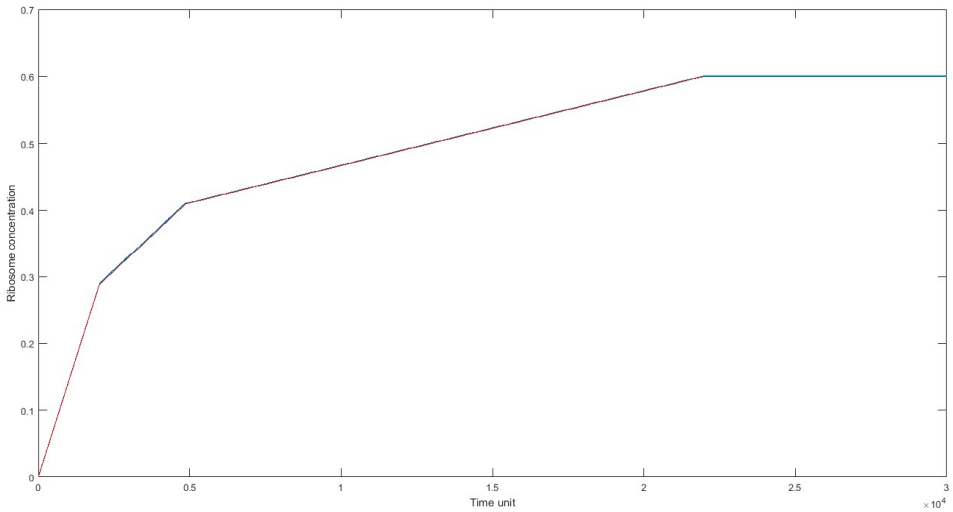
distances = [50 70 250 275 500 900 930];
Tslow = [7 8 9 10 14 16 17];
Tel = 3;
Tini = 2;



distances = [100 150 600 750 850 900];
Tslow = [5 7 7 8 9 8];
Tel = 3;
Tini = 2;



distances = [50 150 200 500 700];
Tslow = [6 7 9 10 8];
Tel = 2;
Tini = 5;



Once we know the dynamics of the ribosome concentration on the mRNA, we know how much ribosome is sequestered on the mRNA at any given time, hence we know how much less the number available ribosomes are in the cytoplasm. The rate of protein synthesis is the rate at which ribosomes are leaving the mRNA. It is hence the posterior concentration of the slow codon furthest from the 5' side of the mRNA times the normal elongation rate multiplied by the mRNA length:

$$protein = p_N * T_{el} * L = \frac{T_{el}^2}{T_{slow,N}} * L$$

Where $\dot{protein}$ is the rate of change in the number of proteins translated from the studied mRNA, N is the number of slow (sufficiently slow) codons and L is the mRNA length. Integrating the above equation using the initial condition protein = 1 @ time = Ttotal, we get

$$protein(t) = 1 + \frac{T_{el}^2}{T_{slow,N}} * L * (t - T_{total})$$

8.6. Discussion and conclusion

The results demonstrate that the fundamental equations we found by empirical methods successfully capture the dynamics of the ribosome concentration on the mRNA. These equations and the proposed theorems (Proposition I and II) can serve as very useful resources for any research group making a realistic model of translation of heterologous genetic circuits incorporating slow codons. The software shows a vectorised and very fast implementation of the method that we use to calculate the ribosome concentration dynamics from the underlying equations and propositions. The software will be extended to incorporate changing initiation and elongation rate with respect to time due to the changing concentration of available ribosomes. A blueprint for such extension can be seen in the Appendix. Once such extension is done, the software then can be used to completely model translation processes on any mRNA in any cell in a fast and realistic manner. In addition, the software then can be easily used in an agent-based system, where each agent is essentially one version of this code with unique initialized parameters to represent each available mRNA in the cell and it consumes ribosomes and energy units from the shared pool, representing the cytoplasm. Furthermore, this theoretical report provides a lot of ideas and methods that can be used to solve other problems in the field Systems Biology and Mathematical Modelling.

9. Appendix of the Appendix

9.1. Pseudocode for the sequential time increment simulation

Variables:

Array: the second dimension of the array is the number of places that ribosome can be on the mRNA (here assumed to be 1000); the first dimension is the time, so that we can see the time evolution of the ribosomes on the mRNA by going down row by row
Tini_change_array: (1/Vini) values, so the reciprocals of the initiation rate; the values are allowed to change at every time unit (since they do change indeed in the cell depending on the number of mRNAs and the number of free ribosomes, as in equation (2.3) in [1]), hence this is a one-dimensional array where the elements reflect (1/Vini) at a given time unit corresponding to the index of the array
Tel: (1/Vel) values, so the reciprocals of the elongation rate. So far, I have assumed no slow codons, so this is a constant, 1/(Vfast).
To make life and simulation easier, we assume that Tini (at all times) and Tel is a positive integer number. Since in terms of the queuing properties, only their ratio should matter, we can assume this without loss of generality.

The states:

Empty ribosome places are marked with "-1" in the array. All other numbers in the array mean that the ribosome place is occupied by the ribosome. The specific number tells us how long the ribosome has been in that position. (If zero, the ribosome has just moved there.)

Pseudocode:

Initialization:
array = -1*ones(length(Tini_change_array), 1000); #empty mRNA
array(1,1) = 1; #Assume at time unit 1, the first ribosome has been at the first place for 1 time unit
For time = 1:length(Tini_change_array)

1. FIND NONEMPTY RIBOSOME PLACES
2. COPY THERE VALUES TO THE NEXT (time+1) ROW OF THE ARRAY AND INCREMENT THEM BY ONE
3. FIND RIBOSOME PLACES WHERE THE SPECIFIC NUMBER IS BIGGER OR EQUAL THAN Tel
4. SORT THESE INTO DESCENDING ORDER #this is important to allow synchronized movement of ribosomes down the mRNA
5. FOR THESE PLACES, AT ROW time+1, REPLACE THE SPECIFIC NUMBER BY "-1" AND MAKE THE SPECIFIC NUMBER OF THE NEXT PLACE DOWN THE MRNA 0 BUT ONLY IF THE SPECIFIC NUMBER OF THAT PLACE WAS "-1" (or if it was the end on the mRNA) #meaning that the ribosome can only move if the place next to it is empty
6. EXAMINE THE LAST Tini (pick Tini to be the element of the Tini_change_array matrix that corresponds to the time in the for loop) ELEMENTS OF THE FIRST COLUMN OF THE ARRAY AND IF THEY WERE ALL EMPTY, THEN PLACE A 0 TO THE FIRST COLUMN IN THE NEXT (time+1) ROW #meaning that we have waited long enough, and a new ribosome should bind to the mRNA

END FOR

9.2. Blueprint for extension of DynoRibo to incorporate the limited number of ribosomes available in the cytoplasm

Slow, but not sufficiently slow codons do not affect the ribosome dynamics in this model, because the movement of a ribosome is never affected by the one in front of it, since Tslow < Tel + Tini. (Of course, in reality, where reaction rates are exponential distributions and not numbers, a not sufficiently slow codon does alter the ribosome dynamics in long enough time.) We also discussed that hence we do not consider not sufficiently slow codons to be slow codons at all. As a result, it depends on the initiation rate whether a codon is a slow codon or not. If initiation rate increases, (Tini decreases) we have more slow codons, which makes ribosome sequestration on the mRNA more likely, which in turn results in less ribosomes and a decrease in initiation rate. If the initiation rate decreases (Tini increases), we have less slow codons, making more ribosome available, resulting in an increase in initiation rate. This seems to be a stabilizing negative feedback.
We have seen that posterior concentration in case of no slow codons (p0) is Tel/(Tel+Tini). The number of ribosome on the mRNA in steady-state is hence L*Tel/(Tel+Tini). If there are no slow codons, there is only one slope we have on the number of ribosomes on mRNA versus time plot. Then the steady-state number of ribosomes on the mRNA is reached when the first ribosome reaches the end of the mRNA (at t = L*Tel). Hence the slope must be L*Tel/(Tel+Tini) divided by L* Tel giving 1/(Tel+Tini). Let's call the number of ribosomes sequestered on one heterologous mRNA **num**. The slope is the time derivative of num. Once Tini is a function of time, we have the following differential equation:

$$num = \frac{1}{T_{el} + T_{ini}(t)} \qquad (A1)$$

Let's determine how Tini changes with time! We know that Vini = k*[mRNA]*[Rfree] (equation 2.3 in [1]), where [] denotes 'the number of' and [Rfree] is number of the free ribosomes, which is the total number of ribosome minus the ribosomes sequestered on all heterologous transcripts. By rearrangement, we get that Tini*[Rfree] = constant, where the constant is 1/(k*[mRNA]). Since Tini(t)*[Rfree](t) is a constant, it has to equal Tini(0)*[Rfree](0) = Tini(0)*Rtotal. Using this, we can express Tini(t) as Tini(0)*Rtotal/[Rfree](t). Note that the number of ribosomes sequestered on all heterologous transcripts if there are m such transcripts is m*num(t). Using this, we can express Tini(t) as

$$T_{ini}(t) = \frac{R_{total}}{R_{total} - m * n(t)} * T_{ini}(0) \qquad (A2)$$

If we substitute this into the equation (A1), we yield:

$$num = \frac{1}{T_{el} + \frac{T_{ini}(0) * R_{total}}{R_{total} - m * num(t)}} \qquad (A3)$$

So far, we assumed no slow codon. We know that at some point, the slope changes when we have slow codons, but from Preposition I, we also know that the change happens when

a. The first ribosome reaches the end of the mRNA
b. Backward congestion propagation process stalls (but not in case of merger as stated by Preposition II)

Now if we assume that none of these happen and we choose the parameter values in such a way, we can safely use (A3). Note that since we know how the slopes change in case a) or b), we could even take this further. We plan to do that in the future, now we only demonstrate the result of a simple case.

One example of parameter values such that we can use (A3) is:
Tel = 2;
L = 200; %length of heterologous mRNA
distances = [50];
Tslow = [30];
no_heterologous_transcript = 400;
Tini_0 = 5; %Tini(0)
R = 10000; %total number of ribosomes, Rtotal

The solution of this differential equation (divided by L to convert the number of ribosomes on the mRNA to our definition of concentration) is the orange curve on the diagram.
The yellow line is the ribosome "concentration" that makes the slow codon just about to disappear, that is

Tslow -Tel = Tini(t)

Which gives the following condition on num(t) /using (A2)/

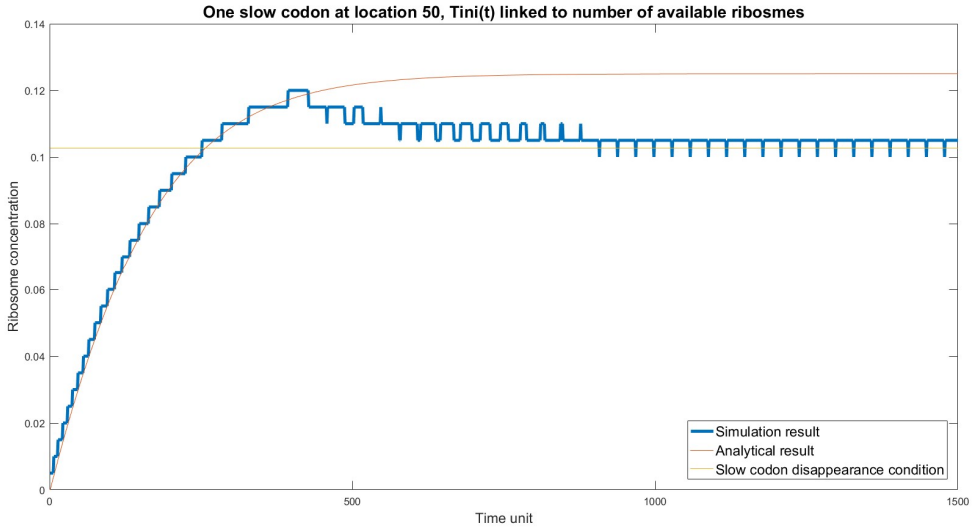
$$num_{critical} = \frac{R_{total}}{m} - \frac{R_{total} * T_{ini}(0)}{m * (T_{slow} - T_{el})} \qquad (A4)$$

Hence, the critical concentration marked by the yellow line is

$$c_{critical} = \frac{n_{critical}}{L}$$

The blue line is the simulation result using the sequential time increment method.
Note that $c_{critical}$ becomes the steady-state concentration (only when $m*L > R_{total}$, in which case the steady-state where the ribosomes are queuing in front of the slow codon is not reachable, because there are not enough ribosomes), which makes sense: when the ribosome concentration on the mRNA is higher, the internal feedback pushes the ribosome concentration on the mRNA down (due to lack of free

ribosomes in the pool reducing T_{ini}) and when it is lower, it gets higher because of the backwards congestion propagation process associated with the slow codon, which sequesters the ribosomes until the slow codons start disappearing due to the lack of free ribosomes.



Appendix 1 – Analytical and numerical result of the translation model for a one slow codon case where limiting amount of ribosomes are available

10. Glossary

For simplicity, here we provide the definition of all terms introduced in this report.

Sufficiently slow codon	$T_{slow} > T_{el} + T_{ini}$ is satisfied
Slow codon disappears	$T_{slow} > T_{el} + T_{ini}$ stops being satisfied
Queue	cluster of ribosomes, such that it has local concentration 1, it starts at a slow codon and is expanding towards the 5' end of the mRNA
Backwards congestion propagation	The process as the queue starts forming from a slow codon and spreads backwards (towards the 5' end)
Stalled queue	cluster of ribosomes, such that it has local concentration 1 (occupying all ribosome places) all the way from the 5' end of the mRNA up to a slow codon

Type I queue change point	The queue reaches the 5’ end of the mRNA
Type II queue change point	The queue reaches a stalled queue
Type III queue change point	The queue reaches an (ongoing) queue closer to the 5’ end of the mRNA