Table of contents:

By using mathematical formulations, real-world problems can be modelled and studied. Scientific investigation into the potential role of disruptions in delicate regulatory processes in the pathogenesis of disease benefits greatly from the availability of mathematical models that facilitate this inquiry. [1,2]

The standard procedure for strong inference in mathematical modelling entails two primary phases:

(1) creating a variety of models for the phenomenon in issue,

(2) comparing these models to any available experimental data and eliminating the ones that don't fit,

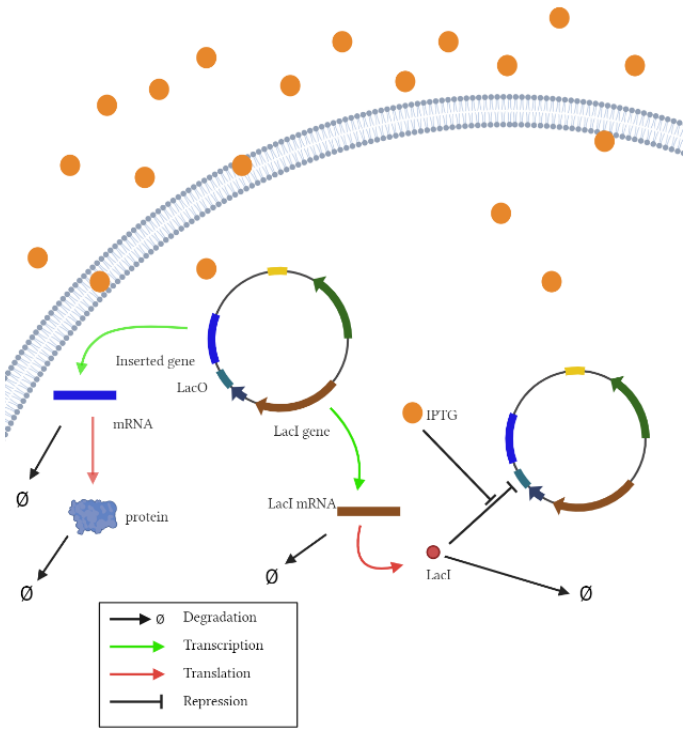
(3) to learn why unsuitable models couldn't account for the facts,

(4) to propose experiments that could distinguish amongst the remaining hypotheses. [3]

Our mathematical modeling was mainly based on 3 main function which are the assumption of Degradation, Transcription and translation steps of our main proteins, Tau and Amyloid beta. Through these 3 functions we can easily consider us to study the change and stability of the proteins expression rate.

Aim and Outline

In our project, we aim to detect the transcription, translation and degradation rates of our proteins tau and Amyloid-beta as the prediction of each rate of them will help us to predict the possible amounts of both proteins beside with helping in determination of the stability of proteins and its efficiency after expression. The parameters which will be present in the equations and codes helped us to go through this aim, in the next parts we will present each rate of them individually with its aim, usage, parameters, the code used.



Parameter	Symbol	Value	Symbol in the code	Unit
transcription rate of parts downstream T7 promoter [4]	K_{tx}	2400 / part sequence length	Ktx	min^{-1}
IPTG diffusion rate [5]	K_t	0.92	Kt_IPTG	min^{-1}
lacI dimerization rate [5]	K_{2R}	50	k2R	$nM^{-1}.min^{-1}$
lacI dimer ($lacI_2$)dissociation [5]	k_{-2R}	10×10^{-3}	K_2R	min^{-1}
lacI transcription rate [5]	k_{sMR}	0.23	ksMR	$nM^{-1}.min^{-1}$
lacI mRNA degradation rate [5]	λ_{MR}	0.462	nMR	min^{-1}
Monomer lacI degradation rate [5]	λ_R	0.2	nR	min^{-1}
lacI translation rate [5]	k_{sR}	15	ksR	min^{-1}
lacI dimer degradation rate ($lacI_2$) [5]	λ_{R2}	0.2	nR2	min^{-1}
Association of 2nd induction pathway [5]	k_{dr2}	$3x10^{-7}$	Kdr2	$nM^{-2}.min^{-1}$
Dissociation of 2nd induction pathway [5]	k_{-dr2}	4.8×10^{3}	k_dr2	$nM^{-1}.min^{-1}$
$lacI_2$-lacO association rate [5]	k_r	960	kr	$nM^{-1}.min^{-1}$
$lacI_2$-lacO dissociation rate [5]	k_{-r}	2.4	K_r	min^{-1}
$IPTG_2$-$lacI_2$ degradation rate [5]	λ_{I2R2}	0.2	nI2R2	min^{-1}
mRNA degradation rate in *E. coli* [5]	δmRNA	0.462	delta_mRNA	min^{-1}
IPTG-lacI association rate [5]	k_{dr1}	$3x10^{-7}$	kdr1	nM-2.min-1
IPTG-lacI dissociation rate [5]	k_{-dr1}	12	K_dr1	min^{-1}
Leakage expression of lacO-lacI [6]	k_{leak}	0.001x ktx	kleak	min-1
Plasmid copy number	----	500-700	O0	----
Total number of ribosomes in a cell [7]	R_T	57000	Rt	----
Number of ribosomes per polysome[7]	S	5000	S	----
Temperature	T	310.5	temperature	K
Energy of hybridization	ΔG_R	Obtained from RBS calculator	DGh_list	$kcal.mol^{-1}$
Energy of folding	$\Delta G_{(Si)}$	Obtained from mfold	D_Gibbs_list	$Kcal.mol^{-1}$
Number of stacked nucleotides in the secondary structure	L	Obtained from mfold	L_list	-
Energy of secondary structure	$\Delta G_{(i,j)}$	Obtained from mfold	DG_SS_list	$Kcal.mol^{-1}$
Gas constant	R	1.99×10^{-3}	Gas_constant	$kcal.mol^{-1}.K^{-1}$

Process	Steps	Equations
Transcription	Passive diffusion of IPTG into and out of the cells	$I_{ext} \underset{-k_t}{\overset{k_t}{\longleftrightarrow}} I_{in}$
	Dimerization of lacI protein	$2R \underset{k_{-2R}}{\overset{k_{2R}}{\longleftrightarrow}} R_2$
	Association of lacI dimer with the lacO operator	$R_2 + O \underset{k_{-r}}{\overset{k_r}{\longleftrightarrow}} R_2\,O$
	Association of IPTG with lacI dimer	$2I + R_2 \underset{k_{-dr1}}{\overset{k_{dr1}}{\longleftrightarrow}} I_2R_2$
	IPTG replaces lacO in lacO-lacI$_2$ complex	$2I + R_2O \underset{k_{-dr2}}{\overset{k_{dr2}}{\longleftrightarrow}} I_2R_2 + O$
	mRNA degradation	$mRNA \xrightarrow{\lambda_{MR}} \phi$
	lacI degradation	$R \xrightarrow{\lambda_R} \phi$
	lacI dimer degradation	$R_2 \xrightarrow{\lambda_{R2}} \phi$
	IPTG- lacI complex degradation rate	$I_2R_2 \xrightarrow{\lambda_{I_2R2}} \phi$
	Transcription of lacI mRNA	$\phi \xrightarrow{k_{sMR}} M_R$
	Transcription o target protein mRNA under T7 promoter	$O \xrightarrow{k_{tx}} mRNA$
	Translation of lacI protein	$M_R \xrightarrow{k_{sR}} M_R + R$
	Leakage expression of target protein mRNA	$R_2O \xrightarrow{k_{leak}} mRNA$
Translation	mRNA folding probability	$p(S_i) = \frac{S_i}{T_{mRNA}} = \frac{\exp\left(-\frac{\Delta G(S_i)}{RT}\right)}{\sum_1^l exp\left(-\frac{\Delta G(S_l)}{RT}\right)}$
	RDS exposure probability	$p_{ex} = \sum_i p(S_i).p_i$
	Hybridization of anti RRS in 16s rRNA to the mRNA	$K_R = \exp\left(-\frac{\Delta G_R}{RT}\right)$

	Ribosome binding probability	$P_c = \frac{m_R}{T_{mRNA}} = \frac{\alpha - \sqrt{\alpha^2 - 4\,K_R^2 \cdot p_{ex}^2 \cdot \frac{R_T}{S} \cdot T_{mRNA}}}{2 \cdot K_R \cdot P_{ex} \cdot T_{mRNA}}$
	Target protein translation	$P_c.mRNA \xrightarrow{k_{tl}} Protein$
	Target protein degradation	$Protein \xrightarrow{\delta_{prot}} \phi$

1-Transcription :

Aim and outline:

This model is built to estimate the mRNA concentration due to transcription of recombinant gene downstream T7 promoter and lac operator under the induction of IPTG at different concentration.

Assumptions and parameters:

The main points of this are:

1. IPTG enters the cell by passive diffusion only
2. lacI only exist from translation of lacI gene in our plasmid

Model explanation:

The amount of lacI protein is estimated from the amount of lacI mRNA, which is dependent on its transcription rate and mRNA degradation rate in *E. coli*.
The processes of mRNA transcription and degradation are expressed in equations 1 and 2 respectively:

$$\emptyset \xrightarrow{k_{sMR}} M_R \quad eq(1)$$

$$M_R \xrightarrow{\lambda_{MR}} \phi \qquad eq\,(2)$$

The mRNA is then translated to give lacI as in equation 3:

$$M_R \xrightarrow{k_{sR}} R \qquad eq(3)$$

Then lacI dimerizes in a reversible process:

$$2R \underset{k_{-2R}}{\overset{k_{2R}}{\longleftrightarrow}} R_2 \qquad eq(4)$$

Upon dimerization lacI can bind to lacO and stop the transcription process of our recombinant DNA. the amount of lacO is expressed as the copy number of our plasmid :

$$R_2 + O \underset{k_{-r}}{\overset{k_r}{\longleftrightarrow}} R_2\,O \quad eq(5)$$

The amount IPTG that enters the cell are diffused into and out of the cell in equal constant rates until the amount of extracellular IPTG equalize the amount of intracellular IPTG as follows:

$$I_{ext} \underset{-k_t}{\overset{k_t}{\longleftrightarrow}} I_{in} \quad eq(6)$$

The intracellular IPTG can stop transcription repression due to lacI by two pathways, in the first pathway IPTG binds to lacI dimer and prevent it from binding to lacO as follows

$$2I + R_2 \underset{k_{-dr1}}{\overset{k_{dr1}}{\longleftrightarrow}} I_2R_2 \quad eq\ (7)$$

In the second pathway IPTG substitutes lacO in lacO/I complex and free the lacO to initiate the transcription

$$2I + R_2O \underset{k_{-dr2}}{\overset{k_{dr2}}{\longleftrightarrow}} I_2R_2 + O \qquad eq(8)$$

The rates of degradation of lacI, lacI dimer and lacI dimer bonded to IPTG are expressed and integrated into the model as follows:

$$R \xrightarrow{\lambda_R} \phi \quad eq(9)$$
$$R_2 \xrightarrow{\lambda_{R2}} \phi \quad eq(10)$$
$$I_2R_2 \xrightarrow{\lambda_{I_2R2}} \phi \quad eq(11)$$

Finally, only genes that have free lacO are transcribed to give mRNA. Their transcription depends on the promoter transcription rate which can be calculated by elongation velocity divided by sequence length, copy number of genes with free lacO and the leakage expression of lacI/O complex

$$O \xrightarrow{k_{tx}} mRNA \quad eq(12)$$
$$R_2O \xrightarrow{k_{leak}} mRNA \quad eq(13)$$

Where $k_{tx} = \frac{elongation\ speed}{sequence\ length}$ and the elongation speed of parts under T7 promoter is 2400 bp/min

The model was built on the following ODEs based on the previous equations.

1. $\frac{dI_{in}}{dt} = k_t.I_{ext} - k_t.I_{in} - k_{dr1}.R_2.{I_{in}}^2 + k_{-dr1}.I_{in_2}R_2 - k_{dr2}.R_2O.{I_{in}}^2 + k_{-dr2}.I_{in_2}R_2.O$
2. $\frac{dM_R}{dt} = k_{sMR} - \lambda_{MR}.M_R$
3. $\frac{dR}{dt} = k_{-2R}.R_2 - k_{2R}.R^2 - \lambda_R.R + k_{sR}.M_R$
4. $\frac{dR_2}{dt} = k_{2R}.R^2 - k_{-2R}.R_2 - k_{dr1}.R_2.{I_{in}}^2 + k_{-dr1}.I_{in_2}R_2 - \lambda_{R_2}.R_2 + k_{-r}.R_2O - k_r.R_2.O$
5. $\frac{dR_2O}{dt} = k_{-dr2}.I_{in_2}R_2.O - k_{dr2}.R_2O.{I_{in}}^2 + k_r.R_2.O - k_{-r}.R_2O$
6. $\frac{dI_{in_2}R_2}{dt} = k_{dr1}.R_2.{I_{in}}^2 - k_{-dr1}.I_{in_2}R_2 + k_{dr2}.R_2O.{I_{in}}^2 - k_{-dr2}.I_{in_2}R_2.O - \lambda_{I_{in_2}R_2}.I_{in_2}R_2$
7. $\frac{dO}{dt} = k_{dr2}.R_2O.{I_{in}}^2 - k_{-dr1}.I_{in_2}R_2 - k_r.R_2.O + k_{-r}.R_2O$

8. $\frac{dmRNA}{dt} = k_{tx}.O - \delta_{mRNA}.mRNA + k_{leak}.R_2O$

- **Aim and outline**

Aim of our model to assess the protein expression

The aim of the model is to estimate the translational efficiency of a given mRNA sequence by obtaining the probability of a given mRNA being bound to a ribosome which is directly proportional to the level of protein expression

Assumptions and parameters

to calculate the TIR as mentioned by (Dokyun *et al*, 2010), the model includes three sequential events in initiation:

(1) the thermodynamic folding of all transcribed mRNAs which is predicted from UNAFold at which secondary structures and their energies are calculated.

(2) exposure of the mRNA's ribosome-docking site (RDS), a sequence of 30-nucleotide found upstream the start codon where ribosome docking occurs.

(3) the binding of a ribosome to the exposed RDS through the ribosome-recognizing sequence (RRS), which is a 10-nucleotide sequence that includes the Shine Dalgarno SD sequence and is complementary to the 3' end of the 16S rRNA, termed anti-RRS.

Model explanation:

The probability of a given secondary structure to exist is calculated using a partition function for the predicted conformations and their Gibbs free energies

$$p(S_i) = \frac{S_i}{T_{mRNA}} = \frac{exp\left(-\frac{\Delta G(S_i)}{RT}\right)}{\sum_1^l exp\left(-\frac{\Delta G(S_l)}{RT}\right)}$$

where T_{mRNA} denotes the total number of mRNAs that may be transcribed from the gene of interest, S_i denotes one of the conformations of the transcribed mRNAs, $\Delta G(S_i)$ denotes the Gibbs free energy of S_i, R denotes the gas constant, T denotes an absolute temperature, and l denotes the number of predicted conformations.

For the ribosome to bind to RDS of mRNA, folded mRNA must be unfolded. However, ribosomes cannot unwind base-paired mRNA structures during translation initiation, but the overall probability that all secondary structures in an RDS would be unfolded at any given moment (p_{ex}) can be calculated by the summation of the probability of a specific conformation to unfold (p_i)multiplying its probability to exist ($p(s_i)$).

$$p_{ex} = \sum_i p_{(Si)} . p_i$$

The probability of a specific secondary structure to unfold (p_i) is calculated from the nucleotide-unpairing probability (θ) in an RDS of conformation (S_i)

$$\theta\_(i,j) = \begin{cases} 1 & ,if\ j\ in\ loop \\ \sqrt[L]{\dfrac{1}{1+\exp\left(\dfrac{-(\Delta G_{i,j})}{RT}\right)}} & ,if\ j\ in\ stack \end{cases}$$

($\theta_{i,j}$) denotes the nucleotide-unpairing probability of the j-th nucleotide in an RDS of conformation (S_i), $\Delta G_{i,j}$ denotes the Gibbs free energy of a stack structure in a ribosome-docking site containing the j-th nucleotide, and L denotes the number of nucleotides in a given stack structure.

$$p_i = \prod_j \theta_{i,j}$$

The probability of binding of ribosome to the mRNA (P_c) can be calculated from the following equation

$$P_c = \frac{m_R}{T_{mRNA}} = \frac{\alpha - \sqrt{\alpha^2 - 4\,K_R^2 \cdot p_{ex}^2 \cdot \frac{R_T}{S} \cdot T_{mRNA}}}{2 \cdot K_R \cdot P_{ex} \cdot T_{mRNA}}$$

Where m_R is the ribosome bonded mRNA and T_{mRNA} is the total number of mRNA in the cell, R_T is total number of ribosomes in the cell, K_R is the hybridization constant that is calculated from the hybridization energy obtained from UNAFold as follows

$$K_R = exp\left(-\frac{\Delta G_R}{RT}\right)$$

While α is calculated from the following equation

$$\alpha = 1 + K_R \cdot P_{ex} \cdot \frac{R_T}{S} + K_R \cdot P_{ex} \cdot T_{mRNA}$$

The amount of target mRNA calculated from transcription model is multiplied by the probability of binding of ribosome to target mRNA to get the amount of active mRNA for translation and the protein then the protein is degraded due to its half-life and cell protein turn over as the following :

$$P_c.mRNA \xrightarrow{k_{tl}} Protein \xrightarrow{\delta_{prot}} \phi$$

Finally the change in the concentration of protein with respect to time is written in the following ODE:

$$\frac{dprot}{dt} = k_{tl}.P_c.mRNA - \delta_{prot}.prot$$

Code review

The code is written to import the required packages; math to get the exp() function for calculations, numpy to create the time array , SciPy to solve the ordinary differential equations, os to access files from the operating system and matplotlib to plot the results of simulation.

Then assign the constants and ask for the inputs (DNA sequence length, protein sequence length, plasmid copy number (you can input different copy numbers), IPTG concentration (you can input different IPTG concentrations), mfold thermodynamics parameters text file location, RBS calculator hybridization energy, working temperature, time of simulation and time steps) for the differential equations.

Then a for loop is designed to extract the thermodynamic parameters from the text files copied from mfold. Then six functions are defined to calculate thermodynamic energies and probabilities followed by the ODE solving function. The time array is assigned followed by another loop to assign initial concentrations and plot outputs

Code can be found in our team GitHub

Ref:

1- Edelstein-Keshet, L. (2005). *Mathematical models in biology*. Society for Industrial and Applied Mathematics.
2- Bellouquid, A., & Delitala, M. (2006). *Mathematical modeling of complex biological systems*. Birkhũser Boston.
3- Ganusov, V. V. (2016). Strong inference in mathematical modeling: a method for robust science in the twenty-first century. *Frontiers in microbiology*, *7*, 1131
4- Skinner, G. M., Baumann, C. G., Quinn, D. M., Molloy, J. E., & Hoggett, J. G. (2004). Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase: a single-molecule view of the transcription cycle. *Journal of Biological Chemistry*, *279*(5), 3239-3244.
5- Stamatakis, M., & Mantzaris, N. V. (2009). Comparison of deterministic and stochastic models of the lac operon genetic network. *Biophysical journal*, *96*(3), 887-906.
6- Kato, Y. (2020). Extremely low leakage expression systems using dual transcriptional-translational control for toxic protein production. *International Journal of Molecular Sciences*, *21*(3), 705.
7- Na, D., Lee, S., & Lee, D. (2010). Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. BMC systems biology, 4(1), 1-16.
8- iGEM Team:SZ-SHD 2020, https://2020.igem.org/Team:SZ-SHD
9- iGEM Team: Ecuador 2021, https://2021.igem.org/Team:Ecuador