Prelim 1

ChemE 7770

Samavi Farnush Bint E Naser

Submitted on: March 29, 2019

Content:

PROBLEM 1

Part (a)

Expression for the kinetic limit of transcription for gene j ($r_{X,j}$) in a set of $\mathcal{N}$ genes

Table 1: Species (concentration)*

Symbol	Species
G_i	gene concentration
R_X^o	free RNAP concentration
$(G_i{:}\,R_X)_C$	closed complex concentration
$(G_i{:}\,R_X)_O$	open complex concentration
m_i	mRNA concentration
$R_{X,T}$	total abundance of RNAP

*unit of concentration

Table 2: Kinetic Parameters†

Symbol	Kinetic Parameter	Unit
k_+	on rate constant for RNAP at promoter	conc.$^{-1}$time^{-1}
k_-	off rate constant for RNAP at promoter	time^{-1}
k_E	transcription elongation rate constant	time^{-1}
k_I	transcription initiation rate constant	time^{-1}
k_A	abortive initiation rate constant	time^{-1}
$\tau_X = \dfrac{k_E}{k_A + k_I}$	time constant	dimensionless
$K_X = \dfrac{k_- + k_I}{k_+}$	saturation constant	conc.
$r_{X,j}$	kinetic limit of transcription for gene j	conc.time^{-1}

†the subscripts i,j denotes specific genes

Assumptions:

- Transcription follows the four elementary steps
- The intermediate complexes are at steady state

- Kinetic limit of transcription is directly proportional to the open complex concentration.

The four elementary steps in gene expression for any gene i can be written as:

$$G_i + R_X^o \overset{k_i}{\Leftrightarrow} (G_i{:}\,R_X)_C \tag{1}$$

$$(G_i{:}\,R_X)_C \overset{k_{I,i}}{\longrightarrow} (G_i{:}\,R_X)_O \tag{2}$$

$$(G_i{:}\,R_X)_O \overset{k_{A,i}}{\longrightarrow} G_i + R_X^o \tag{3}$$

$$(G_i{:}\,R_X)_O \overset{k_{E,i}}{\longrightarrow} m_i + G_i + R_X^o \tag{4}$$

where i =1,2,3,……j……$\mathcal{N}$

Material balances around the closed and open complexes for any gene in the system of $\mathcal{N}$ genes are given by:

$$\frac{d}{dt}(G_i{:}\,R_X)_C = k_{+,i}\,(G_i)(R_X^o) - k_{-,i}(G_i{:}\,R_X)_C - k_{I,i}(G_i{:}\,R_X)_C \tag{5}$$

$$\frac{d}{dt}(G_i{:}\,R_X)_O = k_{I,i}(G_i{:}\,R_X)_C - k_{A,i}(G_i{:}\,R_X)_O - k_{E,i}(G_i{:}\,R_X)_O \tag{6}$$

where i =1,2,3,……j……$\mathcal{N}$

At steady state, the abundance of the closed and open complexes can be estimated from the balance equations:

$$(G_i{:}\,R_X)_C \simeq \frac{k_{+,i}}{k_{-,i}+k_{I,i}}(G_i)(R_X^o) \simeq \frac{1}{K_{X,i}}(G_i)(R_X^o) \tag{7}$$

$$(G_i{:}\,R_X)_O \simeq \frac{k_{I,i}}{k_{A,i}+k_{E,i}}(G_i{:}\,R_X)_C \simeq \frac{1}{K_{X,i}}\frac{1}{\tau_{X,i}}(G_i)(R_X^o) \tag{8}$$

To estimate the free RNAP concentration, we can use the total RNAP balance:

$$R_{X,T} = R_X^o + \sum_{i=1}^{\mathrm{N}}(G_i{:}\,R_X)_C + \sum_{i=1}^{\mathrm{N}}(G_i{:}\,R_X)_O \tag{9}$$

Eqn. (9) can be re-written as:

$$R_{X,T} = R_X^o + \left(G_j{:}\,R_X\right)_C + \left(G_j{:}\,R_X\right)_O + \sum_{i=1,j}^{\mathrm{N}}(G_i{:}\,R_X)_C + \sum_{i=1,j}^{\mathrm{N}}(G_i{:}\,R_X)_O \tag{10}$$

Substituting the expressions for open and closed complexes from Eqn. (7) & (8) in Eqn. (10):

$$R_{X,T} = R_X^o + [\frac{1}{K_{X,j}}(G_j) + \frac{1}{K_{X,j}}\frac{1}{\tau_{X,j}}(G_j) + \sum_{i=1,j}^{N}\frac{1}{K_{X,i}}(G_i) + \sum_{i=1,j}^{N}\frac{1}{K_{X,i}}\frac{1}{\tau_{X,i}}(G_i)\,] * R_X^o \quad (11)$$

Solving for free RNAP concentration R_X^o:

$$R_X^o = \frac{R_{X,T}(\tau_{X,j}K_{X,j})}{\tau_{X,j}K_{X,j} + (\tau_{X,j}+1)\,G_j + \sum_{i=1,j}^{N}\frac{\tau_{X,j}K_{X,j}}{\tau_{X,i}K_{X,i}}(\tau_{X,i}+1)(G_i)} \quad (12)$$

Eqn. (8) can be written explicitly for gene *j*, where we have substituted expression for the free RNAP concentration:

$$(G_j{:}R_X)_O \simeq \frac{R_{X,T}\,G_j}{\tau_{X,j}K_{X,j} + (\tau_{X,j}+1)\,G_j + \sum_{i=1,j}^{N}\frac{\tau_{X,j}K_{X,j}}{\tau_{X,i}K_{X,i}}(\tau_{X,i}+1)(G_i)} \quad (13)$$

Let the kinetic limit of transcription for gene *j* be directly proportional to the cconcentration of the open complex:

$$\mathrm{r_{X,j}} = k_{E,j}\,(G_j{:}R_X)_O = k_{E,j}\,R_{X,T}\,\frac{G_j}{\tau_{X,j}K_{X,j} + (\tau_{X,j}+1)\,G_j + \varepsilon_j} \quad (14)$$

where:

$$\varepsilon_j = \sum_{i=1,j}^{N}\frac{\tau_{X,j}K_{X,j}}{\tau_{X,i}K_{X,i}}(\tau_{X,i}+1)(G_i)$$

Part (b)

An $\mathcal{N}$ gene system ($\mathcal{N}$>>1) will be approximately equivalent to a 1-gene system if $\varepsilon_j \approx 0$.

- If transcription of gene *j* is elongation limited, then $\tau_{X,j} \sim \frac{k_{E,j}}{k_{I,j}} << 1$ and $\varepsilon_j \approx 0$.
- If transcription of all genes except gene j is initiation limited, then ${\tau_{X,i}}^{-1} \sim \frac{k_{I,i}}{k_{E,j}} << 1$ and $\varepsilon_j \approx 0$.
- If the saturation constant for gene *j* is small compared to the saturation constant for all other genes *i*, then $\frac{K_{X,j}}{K_{X,i}} << 1$ and $\varepsilon_j \approx 0$.
- If $\mathcal{G}_j << 1$, then also we can have $\varepsilon_j \approx 0$.

PROBLEM 2

The code for this part is in Q2\Q2.m

Instructions for running the code is given in the README file.

Part (a)

- For running the model to steady state, visual confirmation was used by generating a plot of P_1, P_2 and P_3 vs. time in the window without inducer (Q2\ss.png)
- Plot of P_1, P_2 and P_3 vs. time in Phase 1 and Phase 2 is attached (Q2\protein evolution.png)

Part (b)

- Q2.m code will generate variables sens_array1, sens_array2 and sens_array3 which are the scaled sensitivity coefficients for 20 min windows in Phase 1, Phase 2 and Phase 3 respectively. The structure of the matrices is described below:

$$s_{ij} = \begin{matrix} \frac{\partial x_1}{\partial p_1}_{t=0} & \frac{\partial x_1}{\partial p_2}_{t=0} & \cdots\cdots & \frac{\partial x_1}{\partial p_j}_{t=0} \\ \frac{\partial x_2}{\partial p_1}_{t=0} & \frac{\partial x_2}{\partial p_2}_{t=0} & \cdots\cdots & \frac{\partial x_2}{\partial p_j}_{t=0} \\ \cdots\cdots & & & \\ \frac{\partial x_1}{\partial p_1}_{t=1} & \frac{\partial x_1}{\partial p_2}_{t=1} & \cdots\cdots & \frac{\partial x_1}{\partial p_j}_{t=1} \\ \frac{\partial x_2}{\partial p_2}_{t=1} & \frac{\partial x_2}{\partial p_2}_{t=1} & \cdots\cdots & \frac{\partial x_2}{\partial p_2}_{j=1} \\ \cdots\cdots & & & \\ \cdots\cdots & & & \end{matrix}$$

where x_1, x_2, x_3 = mRNA1, mRNA2 and mRNA3 concentrations respectively

x_4, x_5, x_6 = P_1, P_2 and P_3 concentrations respectively

p_1, p_2, ……,p_{33} = model parameters

Phase 1 is as described in Q2. Phase 2/3 are early and late 20 min windows after addition of inducer respectively.

- While employing the eqn.

$$s_{ij} = \left(\frac{p_j}{x_i}\right)_* \left(\frac{\partial x_i}{\partial p_j}\right)_t$$

  For Phase 1, I used $\left(\frac{p_j}{x_i}\right)_* = 1$.

  For Phase 2 and 3, the concentration values obtained with original values of parameters along with the original parameter values were used.

- I used $\pm$ 5% deviation in the parameter values to obtain the $\left(\frac{\partial x_i}{\partial p_j}\right)_t$ terms in the sensitivity arrays using central difference scheme.

$$\left(\frac{\partial x_i}{\partial p_j}\right)_t = \frac{x_{@1.25p_j} - x_{@0.75p_j}}{p_j * (1.25 - 0.75)}$$

Part (c)

- Q2.m will generate a variable named "result" showing U_1, U_2 and U_3 values for different species. (see: U.txt)
- The time averaged sensitivity arrays were obtained using trapezoidal rule. Time averaged sensitivity arrays are generated as the variables S1, S2 and S3 for Phase 1 and early/late Phase 2.

PROBLEM 3

TABLE 1
Simplified, fundamental reaction set for protein production

Transcription initiation:	$G + \text{RNAP} \xrightarrow{v_1} G^*$
Transcription:	$G^* + n\text{NTP} \xrightarrow{v_2} \text{mRNA} + G + \text{RNAP} + 2n\text{P}_i$
mRNA decay:	$\text{mRNA} \xrightarrow{v_3} n\text{NMP}$
Translation initiation:	$\text{mRNA} + \text{rib} \xrightarrow{v_4} \text{rib}^*$
Translation:	$\text{rib}^* + a\text{AAtRNA} + 2a\text{GTP} \xrightarrow{v_5} a\text{tRNA} + 2a\text{GDP} + 2a\text{P}_i + \text{rib} + \text{mRNA} + \text{protein}$
tRNA charging:	$\text{AA} + \text{tRNA} + \text{ATP} \xrightarrow{v_6} \text{AMP} + 2\text{P}_i + \text{AAtRNA}$
Exchange fluxes:	$\text{AA}_{ext} \xrightarrow{b_1} \text{AA}$
	$\text{NTP}_{ext} \xrightarrow{b_2} \text{NTP}$
	$\text{protein} \xrightarrow{b_3} \text{protein}_{ext}$
	$\text{NMP} \xrightarrow{b_4} \text{NMP}_{ext}$
	$\text{ATP}_{ext} \xrightarrow{b_5} \text{ATP}$
	$\text{AMP} \xrightarrow{b_6} \text{AMP}_{ext}$
	$\text{GTP}_{ext} \xrightarrow{b_7} \text{GTP}$
	$\text{GDP} \xrightarrow{b_8} \text{GDP}_{ext}$
	$\text{P}_i \xrightarrow{b_9} \text{P}_{i\,ext}$

Note: The first six reactions, which are discussed in the text, occur within the virtual systemic boundary within which the machinery for protein synthesis resides. The last nine reactions correspond to the exchange of building blocks (i.e. AAs and NTPs), protein, by-products (e.g. NMPs), and energy molecules (i.e. ATP and GTP) across the systemic boundary.

The above reaction scheme has been obtained from Allen and Palsson (2003)

The code for the problem is in Q3\~. Instructions for running the code is given in the README file.

Part (a)

- The attached file Q3\S.txt contains the stoichiometric matrix for the above reaction scheme.
- Rate of transcription,

$$\hat{r_x} = \frac{\dot{v}_x}{\mathcal{L}_X} \times \frac{\mathcal{L}_X}{gene\ length} \times R_X \times \frac{plasmid\ concentration}{(K_X \times \tau_X) + (\tau_X + 1) \times plasmid\ concentration} \times u(I)$$

where $u(I) = \frac{W_1 + W_2 f_I}{1 + W_1 + W_2 f_I}$ and $f_I = \frac{I^n}{k^n + I^n}$

- Rate of translation,

$$\hat{r}_L = \frac{\dot{v}_x}{\mathcal{L}_T} \times \frac{\mathcal{L}_T}{peptide\ length} \times R_L \times \frac{\text{mRNA concentration}}{(\text{KL} \times \tau\text{L}) + (\tau\text{L} + 1) \times \text{mRNA concentration}}$$

- In cell free steady system,

$$\frac{dmRNA}{dt} = \hat{r}_L - k_{d,X} \times mRNA\ concentration = 0$$

The above equation was used to obtain mRNA concentration for calculating $\hat{r_L}$.

Part (b)

- For solving the FBA problem, lower bounds for fluxes v_1, v_3, v_4, v_6 were set as 0 and upper bounds for these fluxes were set as 100,000. The other flux bounds were set as described the problem statement.
- The steady state protein production rate were obtained from the objective value from the solving the FBA problem.
- The code will generate a plot of protein level vs. inducer concentration on semi-log scale. Plot will be generated for a = n = 2. The values of a and n can be changed in "data.jl" file.
- The code will generate a "flux.csv" file with optimized fluxes with varying levels of inducer.

Part (c)

To evaluate which exchange flux the translation rate is most sensitive to I observed the protein production rate at two different limiting levels of the exchange fluxes, 0.1 µM/hr and 0.5 µM/hr respectively (a=n=4). Discussion is mainly based on observations made at limiting flux bound 0.5 µM/hr. Some of the observations are listed below:

- The effect of the following fluxes are equivalent:
  - b_1, b_5 and b_6: rates at which amino acids and ATP are added and AMP are taken out. While this fluxes affected the maximum production of protein attainable, translation was still on-going (Figure 1).
  - b_7 and b_8: GTP in flux and GDP out flux; these fluxes also limit the maximum production of protein attainable, translation was still on-going (Figure 2).
  - Translation of protein is the most sensitive to changes in the following exchange flux bounds: b_2, b_4, b_9: GTP, GDP and P_i fluxes respectively, where translation stops when inducer concentration goes above ~0.06 mM (Figure 3).

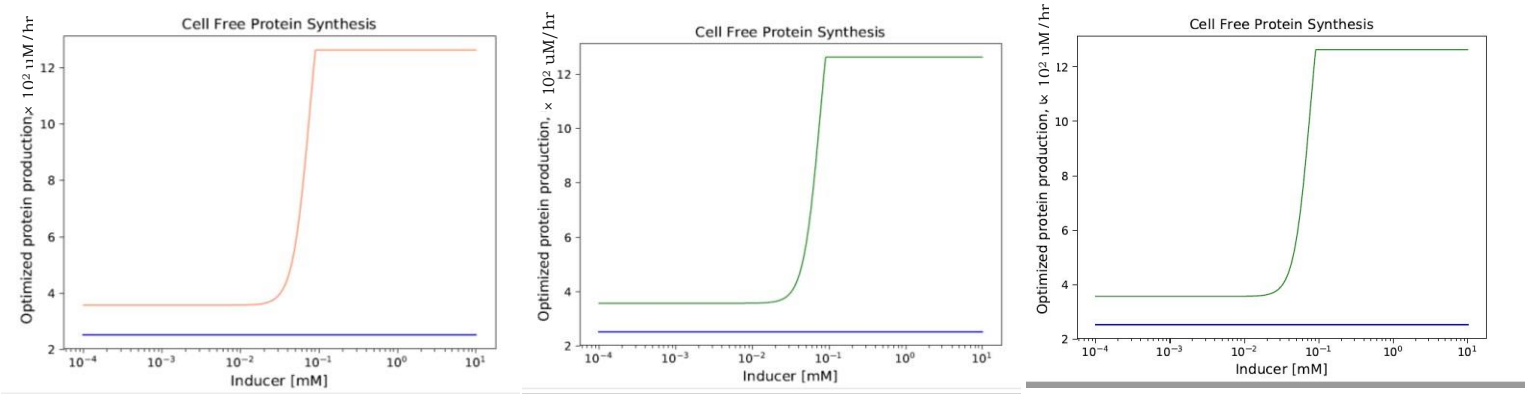

Figure 1: Effect of (left to right) b_1, b_5 and b_6 on protein level. Bottom blue line was obtained for limiting flux=0.1µM/hr.

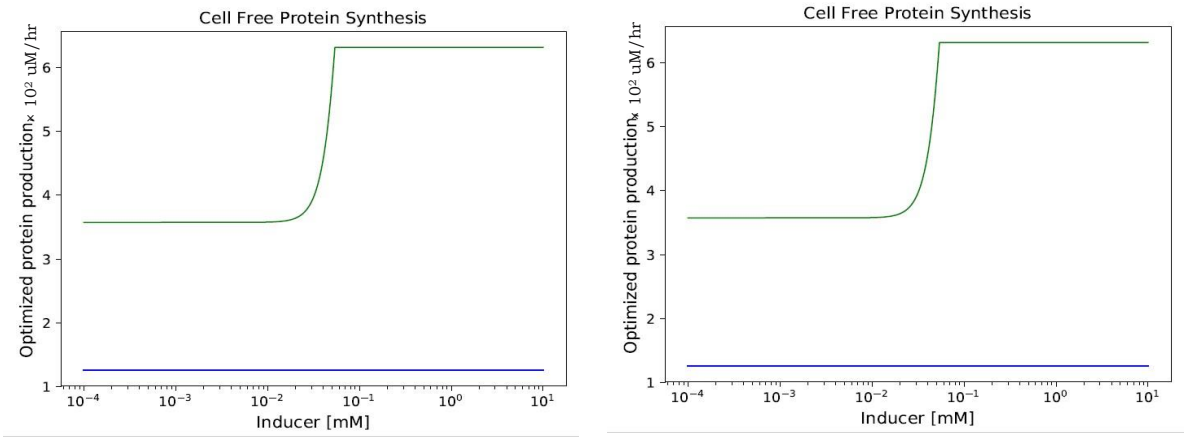

Figure 2: Effect of (left to right) b_7 and b_6 on protein level. Bottom blue line was obtained for limiting flux=0.1μM/hr.

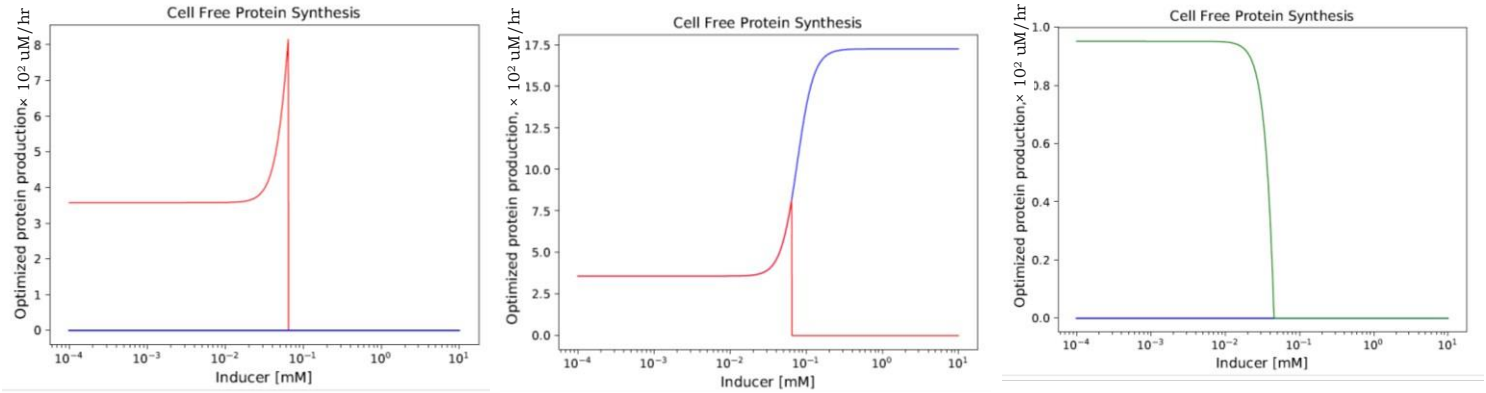

Figure 3: Effect of (left to right) b_2, b_4 and b_9 on protein level.