Prelim #1
Sam Furness
3/19/19

1 Question 1 Part A

Starting with the mental model of transcription:

$$G_j + R_X^o \rightleftharpoons (G_j : R_X)_C$$

$$(G_j : R_X)_C \rightarrow (G_j : R_X)_O$$

$$(G_j : R_X)_O \rightarrow G_j + R_X$$

$$(G_j : R_X)_O \rightarrow m_j + G_j + R_X$$

Where G_j represents some gene concentration and R_X^o represents free RNAP concentration and $(G_j : R_X)_C$,$(G_j : R_X)_O$ represent the closed and open complexes respectively. We can define the kinetic limit of transcription as being proportional to the open complex by the elongation rate rate constant k_{Ej}:

$$r_{Xj} = k_{Ej}(G_j : R_X)_O$$

Now differential equations can be written on the closed and open complexes:

$$\frac{d}{dt}(G_j : R_X)_C = k_{+j}(G_j)(R_X^o) - k_{-j}(G_j : R_X)_C - k_{Ij}(G_j : R_X)_C$$

$$\frac{d}{dt}(G_j : R_X)_O = k_{Ij}(G_j : R_X)_C - k_{Aj}(G_j : R_X)_O - k_{Ej}(G_j : R_X)_O$$

Each rate term refers to the rate of formation on the closed complex (second order k_{+j}), the dissociation of the closed complex (k_{-j}), the formation of the open complex from the closed (k_{Ij}), and the rate of abortive initiation (k_{Aj}).
Now, a mass balance on total RNAP can be made that includes every gene.

$$R_{XT} = R_X^o + (G_j : R_X)_C + (G_j : R_X)_O + \sum_{i=1,j}^{N}((G_i : R_X)_C + (G_i : R_X)_O)$$

From here, the quasi steady state assumption must be made to solve for the open and closed complexes. This means that the concentrations of the complexes after an initial change, will not change much and the rate of mRNA production will dominate. There is a separation of timescales between mRNA production and the complexes' rate of change. We can set the complexes' rate of change to 0. The solutions are thus:

$$(G_j : R_X)_C = \frac{k_{+j}}{k_{-j} + k_{Ij}}(G_j)(R_X^o)$$

$$(G_j : R_X)_O = \frac{k_{Ij}}{k_j A + k_{Ej}}(G_j : R_X)_C$$

We can define the saturation constant of the gene and RNAP binding as:

$$K_{Xj}^{-1} = \frac{k_{+j}}{k_{-j} + k_{Ij}}$$

The time constant of the gene and RNAP can be defined as:

$$\tau_{Xj}^{-1} = \frac{k_{Ij}}{k_{Aj} + k_{Ej}}$$

Taking these new parameters, the open complex can be related to the free gene and RNAP by:

$$(G_j : R_X)_O = K_{Xj}^{-1}\tau_{Xj}^{-1}G_j R_X^o$$

Substituting into the mass balance and using analogous terms for the i genes:

$$R_{XT} = R_X^o + K_{Xj}^{-1}G_j R_X^o + K_{Xj}^{-1}\tau_{Xj}^{-1}G_j R_X^o + \sum_{i=1,j}^{N}(K_{Xi}^{-1}G_i R_X^o + K_{Xi}^{-1}\tau_{Xi}^{-1}G_i R_X^o)$$

Free RNAP can be isolated to:

$$R_X^o = \frac{R_{XT}}{1 + K_{Xj}^{-1}G_j + K_{Xj}^{-1}\tau_{Xj}^{-1}G_j + \sum_{i=1,j}^{N}(K_{Xi}^{-1}G_i + K_{Xi}^{-1}\tau_{Xi}^{-1}G_i)}$$

Finally, the open conformation can be rewritten and plugged into our equation for rate:

$$(G_j : R_X)_O = \frac{R_{XT}G_j K_{Xj}^{-1}\tau_{Xj}^{-1}}{1 + K_{Xj}^{-1}G_j + K_{Xj}^{-1}\tau_{Xj}^{-1}G_j + \sum_{i=1,j}^{N}(K_{Xi}^{-1}G_i + K_{Xi}^{-1}\tau_{Xi}^{-1}G_i)}$$

$$(G_j : R_X)_O = \frac{R_{XT}G_j}{K_{Xj}\tau_{Xj} + \tau_{Xj}G_j + G_j + \sum_{i=1,j}^{N}(K_{Xj}\tau_{Xj}K_{Xi}^{-1}G_i + K_{Xj}\tau_{Xj}K_{Xi}^{-1}\tau_{Xi}^{-1}G_i)}$$

$$(G_j : R_X)_O = \frac{R_{XT}G_j}{K_{Xj}\tau_{Xj} + \tau_{Xj}G_j + G_j + \sum_{i=1,j}^{N}(\frac{K_{Xj}\tau_{Xj}}{K_{Xi}\tau_{Xi}}G_i(\tau_{Xi}+1))}$$

Now subbing into the equation for kinetic limit:

$$r_{Xj} = k_{Ej}(G_j : R_X)_O$$

$$r_{Xj} = k_{Ej}\frac{R_{XT}G_j}{K_{Xj}\tau_{Xj} + \tau_{Xj}G_j + G_j + E_j}$$

Where

$$E_j = \sum_{i=1,j}^{N}(\frac{K_{Xj}\tau_{Xj}}{K_{Xi}\tau_{Xi}}G_i(\tau_{Xi}+1))$$

2 Question 1 Part B

Given that a 1 gene system has the following equation for kinetic limit of transcription (derived in class):

$$r_{Xj} = k_{Ej}R_{XT}\frac{G_j}{K_{Xj}\tau_{Xj} + (\tau_{Xj}+1)G_j}$$

The only difference between the equations is the E_j term. This term approaches zero when the gene in question G_j has a much smaller saturation binding constant K_{Xj} and time constant τ_{Xj} of the gene and RNAP than all the other genes in the system:

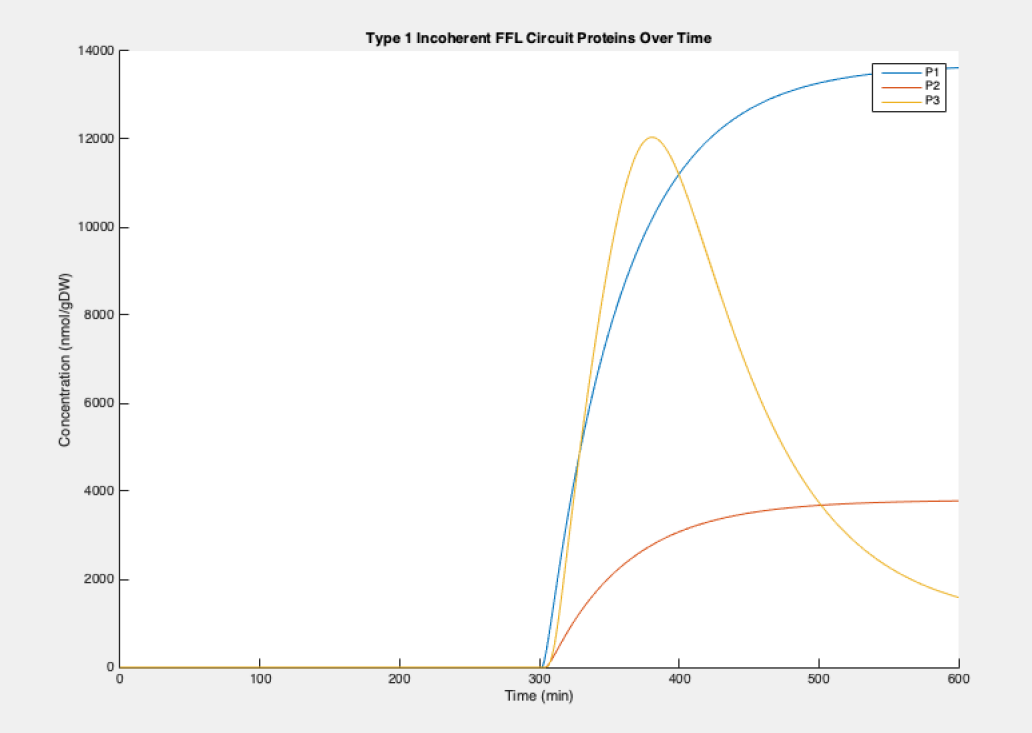
$$K_{Xj} << K_{Xi}$$

$$\tau_{Xj} << \tau_{Xi}$$

What this means is that the closed complex of this gene G_j and RNAP forms much faster than other genes (larger k_{+j} than k_{-j} and k_{Ij}). Also, the open conformation from the closed conformation is formed much faster than the other genes (larger k_{Ij} than k_{Aj} and k_{Ej}). This makes sense because it just means that the gene we are interested in forms successive complexes much faster than the rest of the genes so the other genes do not affect it's rate of transcription by taking up free RNAP. G_j transcription saturates sooner and takes less time to initiate. The other trivial situation where the two systems are approximately equivalent is when $G_i << G_j$ because the other i genes are rarely transcribed to take up resources.

3 Question 2 Part A

The code CHEME5440P1_2a.m solves this question. Parameters are from the problem statement and Dr. Varner's .json file. It calls the function bind_fxn.m to calculate the binding function in the Moon/Voigt control functions. Note the change in weight values to get the correct incoherent type 1 FFL shape. The plot of protein versus time is given as follows where Phase 1 was created by just adding enough time until the curves were flat at steady state for at least 60 minutes (300 minutes was needed). Phase 2 was ran for 300 minutes.



4 Question 2 Part B

The code CHEME5440P1_2bc.m solves this question and calculates all the sensitivity coefficients. It calls the function compute.m which is basically Part A turned into a function (calculates the concentration of each species over time). Since there is an array for each parameter with dimensions #species x 600 minutes (at every time point in Phase 1-3 and each species), this data is not presented. Phase 1 is the last 20 minutes before the inducer. Phase 2 early is the first 20 minutes with inducer and Phase 2 late is the last 20 minutes of the 300 minutes containing the inducer. If you wish to look at a sensitivity array, the variable names are "s_'name of variable for that parameter'"

5 Question 2 Part C

The code CHEME5440P1_2bc.m solves this question as well and returns the first column of the U matrix for each phase (U1:Late Phase 1, U2;Early Phase 2, and U3:Late Phase 2). Each parameter's sensitivity matrix is integrated over time for a given phase to return a vector of time-averaged sensitivity for each species (dimensions #species x 1). These vectors are then concatenated into a time-averaged sensitivity array with dimensions #species x #parameters. All three of these arrays (for each phase) are put through singular value decomposition. The absolute value of the

U matrix first column vectors are:

$$U1(:,1) = \begin{pmatrix} 0.2677 \\ 0.3087 \\ 0.2612 \\ 0.4951 \\ 0.5341 \\ 0.4842 \end{pmatrix} U2(:,1) = \begin{pmatrix} 0.2171 \\ 0.2260 \\ 0.4125 \\ 0.3936 \\ 0.4171 \\ 0.6347 \end{pmatrix} U3(:,1) = \begin{pmatrix} 0.0231 \\ 0.0395 \\ 0.7758 \\ 0.0531 \\ 0.0871 \\ 0.6210 \end{pmatrix}$$

From this the rankings are thus:

$$Phase1 = \begin{pmatrix} p2 \\ p1 \\ p3 \\ m2 \\ m1 \\ m3 \end{pmatrix} EarlyPhase2 = \begin{pmatrix} p3 \\ p2 \\ m3 \\ p1 \\ m2 \\ m1 \end{pmatrix} LatePhase2 = \begin{pmatrix} m3 \\ p3 \\ p2 \\ p1 \\ m2 \\ m1 \end{pmatrix}$$

These rankings reveal that the most sensitive species moves from protein 2 and 1 to protein 3 and 2 when the inducer is introduced. The sensitivity of protein 1 drops two positions while protein 3 climbs two positions (the mRNAs follow with their protein in each case). This is probably because the inducer starts to increase the level of protein 1 rapidly making the production of protein 2 and 3 highly dependent on parameters to begin being induced. Their production is solely dependent on how much protein 1 will induce them or lead do repression with protein 3 indirectly via protein 2. Protein 1 is no longer sensitive in phase 2 because the inducer causes it's production in any case. As the system approaches steady state again in late phase 2, the most sensitive species becomes mRNA 3/protein 3 probably because it is severely repressed at this point and thus dependent on a multitude of parameters including those involved in protein 2 repression and protein 1 activation. Protein 2 is still dependent on protein 1 induction parameters so it is second most sensitive along with it's mRNA being the second most sensitive mRNA. Across the board, the proteins and mRNAs follow similar shifting patterns (p1 follows m1 etc.) and the patterns are driven by the species' sensitivity to parameters in response to induction or repression. It is worth noting that in Phase 1 with only basal transcription, protein 3 is not as sensitive because proteins 1 and 2 are not at a high enough concentration to cause significant repression or induction so it's only really depending on it's own transcription parameters.

6 Question 3 Part A

The stoichiometric matrix can be found in CHEME5440P1_3.jl. The external species are ignored on Dr. Varner's instruction and there are exchange fluxes written into the reactions that use the internal species (with the given bounds). The formulation for $\hat{r}_X$ is taken from the standard equation for one gene transcription:

$$r_X = k_{EX} \frac{R_{XT} G}{K_X \tau_X + \tau_X G + G}$$

$$\hat{r}_X = r_X * u$$

Where u is the control function with basal weight $W_1 = 0.26$ and inducer weight $W_2 = 300$:

$$u = \frac{W_1 + W_2 f_I}{1 + W_1 + W_2 f_I}$$

and $f_I = \frac{I^n}{K^n + I^n}$ with order n=1.5 and binding parameter $K = 0.3mM$. Inducer I ranges from $0.0001mM$ to $10mM$. Now $\hat{r}_L$, by analogy, is the same standard translation rate. However, since mRNA is at steady state and the transcription rate is set to a constant value, we can solve for mRNA:

$$\frac{dmRNA}{dt} = 0 = k_{dX} mRNA + \hat{r}_X$$

$$mRNA = \hat{r}_X / k_{dX}$$

Therefore:

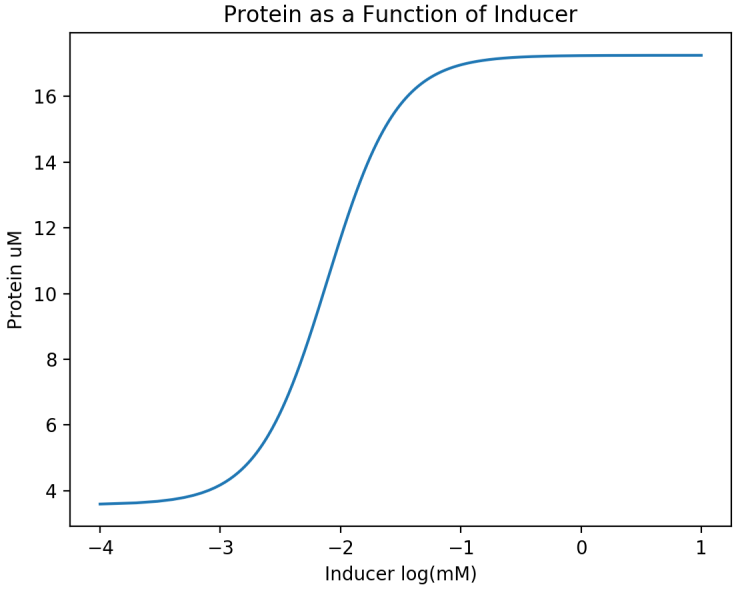
$$\hat{r}_L = k_{EL} \frac{R_{LT} mRNA}{K_L \tau_L + \tau_L mRNA + mRNA}$$

$$\hat{r}_L = k_{EL} \frac{R_{LT} \hat{r}_X / k_{dX}}{K_L \tau_L + \tau_L \hat{r}_X / k_{dX} + \hat{r}_X / k_{dX}}$$

Note that there is not control term on translation because it is assumed to be operating at it's limit. This same steady state method is used to extract protein concentration to plot. All flux bounds are given in CHEME5440P1_3.jl.

7 Question 3 Part B

The file CHEME5440P1_3a,b.jl runs FBA at each inducer concentration and plots the steady state protein concentration as a function of inducer. All bounds are in that file. An $Sv = 0$ check revealed the answer was physical.



8 Question 3 Part C

The file CHEME5440P1_3c.jl calculates shadow prices for each of the exchange fluxes and returns an array of each of them (dimension: 9). An inducer concentration of 10 mM was chosen for all shadow price calculations. A shadow price shows how much the objective (translation rate) changes due to a perturbation in a constraint. This thus reveals sensitivity. The method is as follows. The exchange flux in question was decreased to an arbitrarily small value (10/3600 $\mu M/s$) to make it the binding constraint, and then perturbed by 1 $\mu M/s$ on both ends (reversible). The objective value was then compared between the adjusted boundary (including the arbitrarily small flux bound) and the perturbed boundary. The shadow price is just the difference between these two values. The shadow price

vector given is as follows (for each $b*$):

$$ShadowPrice = \begin{pmatrix} 3.84129e-5 \\ 4.74316e-5 \\ 0.0 \\ 4.74316e-5 \\ 3.84129e-5 \\ 3.84129e-5 \\ 4.29223e-5 \\ 4.29223e-5 \\ 4.74316e-5 \end{pmatrix}$$

The highest shadow prices are for the exchange fluxes of of phosphate, NTP, and NMP. The second highest for GTP and GDP. The next highest for amino acids, ATP, and AMP. The final exchange flux that had no shadow price was protein likely because the arbitrarily low boundary condition did not even make this bound constraining. This means that the exchange fluxes that translation rate is most sensitive to are those ushering species in transcription, then those in translation energy, and then those in tRNA charging.