Question 1.

Part A.

Make same assumptions as in class: cells are well mixed

G_j is not degraded and produced to match growth rate → G_j is constant

we are at steady-state

$$R_{x,T} = R_x + (G_j : R_x)_c + (G_j : R_x)_o + \sum_{i=1,j}^{N} \left[(G_i : R_x)_c + (G_i : R_x)_o \right]$$

$$K_{x,j}^{-1} \equiv K_{+,j} / (K_{-,j} + K_{\Sigma,j})$$

$$\tau_{x,j}^{-1} \equiv K_{\Sigma,j} / (K_{A,j} + K_{\varepsilon,j})$$

Estimation for open or closed complexes given in class:

$$(G_j : R_x)_c \approx \left(\frac{K_{+,j}}{K_{-,j} + K_{\Sigma,j}} \right) G_j R_x = K_{x,j}^{-1} G_j R_x$$

$$(G_j : R_x)_o \approx \left(\frac{K_{\Sigma,j}}{K_{A,j} + K_{\varepsilon,j}} \right) (G_j : R_x)_c = \tau_{x,j}^{-1} K_{x,j}^{-1} G_j R_x$$

Substitute these terms into the $R_{x,T}$ RNAP balance:

$$R_{x,T} = R_x + K_{x,j}^{-1} G_j R_x + \tau_{x,j}^{-1} K_{x,j}^{-1} G_j R_x + \sum_{i=1,j}^{N} \left[K_{x,i}^{-1} G_i R_x + \tau_{x,i}^{-1} K_{x,i}^{-1} G_i R_x \right]$$

Solve for free RNAP concentration, R_x:

$$K_{x,j} \tau_{x,j} R_{x,T} = R_x K_{x,j} \tau_{x,j} \left[1 + K_{x,j}^{-1} G_j + \tau_{x,j}^{-1} K_{x,j}^{-1} G_j + \sum_{i=1,j}^{N} \left[K_{x,i}^{-1} G_i + \tau_{x,i}^{-1} K_{x,i}^{-1} G_i \right] \right]$$

$$K_{x,j} \tau_{x,j} R_{x,T} = R_x \left[K_{x,j} \tau_{x,j} + \tau_{x,j} G_j + G_j + \tau_{x,j} K_{x,j} \sum_{i=1,j}^{N} \left[K_{x,i}^{-1} G_i + \tau_{x,i}^{-1} K_{x,i}^{-1} G_i \right] \right]$$

$$K_{x,j} \tau_{x,j} R_{x,T} = R_x \left[K_{x,j} \tau_{x,j} + \tau_{x,j} G_j + G_j + \tau_{x,j} K_{x,j} \sum_{i=1,j}^{N} \left[K_{x,i}^{-1} \tau_{x,i}^{-1} (1 + \tau_{x,i}) G_i \right] \right]$$

$$R_x = \frac{K_{x,j} \tau_{x,j} R_{x,T}}{K_{x,j} \tau_{x,j} + \tau_{x,j} G_j + G_j + \tau_{x,j} K_{x,j} \sum_{i=1,j}^{N} \left[K_{x,i}^{-1} \tau_{x,i}^{-1} (1 + \tau_{x,i}) G_i \right]}$$

$$R_x = \frac{K_{x,j} \tau_{x,j} R_{x,T}}{K_{x,j} \tau_{x,j} + (1 + \tau_{x,j}) G_j + \varepsilon_j} \quad \text{where} \quad \varepsilon_j = \sum_{i=1,j}^{N} \frac{K_{x,j} \tau_{x,j}}{K_{x,i} \tau_{x,i}} (1 + \tau_{x,i}) G_i$$

Use R_x to find the open complex concentration, $(G_j : R_x)_o$:

$$(G_j : R_x)_o \approx \tau_{x,j}^{-1} K_{x,j}^{-1} G_j R_x = \tau_{x,j}^{-1} K_{x,j}^{-1} G_j \left[\frac{K_{x,j} \tau_{x,j} R_{x,T}}{K_{x,j} \tau_{x,j} + (1 + \tau_{x,j}) G_j + \varepsilon_j} \right]$$

$$(G_j : R_x)_o = \frac{G_j R_{x,T}}{K_{x,j} \tau_{x,j} + (1 + \tau_{x,j}) G_j + \varepsilon_j}$$

Using the form for the kinetic limit of transcription and $(G_j : R_x)_o$:

$$r_{x,j} = K_{\varepsilon,j} (G_j : R_x)_o$$

$$r_{x,j} = K_{\varepsilon,j} R_{x,T} \left(\frac{G_j}{\tau_{x,j} K_{x,j} + (1 + \tau_{x,j}) G_j + \varepsilon_j} \right) \quad \text{where} \quad \varepsilon_j = \sum_{i=1, i \neq j}^{N} \frac{K_{x,j} \tau_{x,j}}{K_{x,i} \tau_{x,i}} (1 + \tau_{x,i}) G_i$$

Part B.

Looking at the kinetic limit of transcription for gene j, $r_{X,j}$, the many genes ($N \gg 1$) scenario differs from the single gene system only by the inclusion of ε_j in the denominator of the many gene scenario. For the many gene scenario to be approximately equivalent to the single gene system described in class, ε_j must approach zero. Since $K_{X,i}$, $\tau_{X,i}$, and G_i can vary for each gene, there are many cases that allow for ε_j to be near zero.

For any gene i, a low concentration of gene (G_i) would override the other constants and minimize the i-ith term of the summation. Similarly, a large $K_{X,j}$ would also cause the i-ith term to be small. In the expanded form of ε_j, it is visible that there are two terms, one that depends on $\tau_{X,i}$ (in the denominator) and one term that does not depend on $\tau_{X,i}$. The $\tau_{X,i}$-dependent term can approach zero when $\tau_{X,i}$ is large enough but the second, $\tau_{X,i}$-independent term complicates things as $\frac{K_{X,j} \tau_{X,j}}{K_{X,j}}$ must also be near zero for the i-ith term to be small. As this third scenario demonstrates, there are likely many scenarios where the exact balance of $K_{X,i}$, $\tau_{X,i}$, and G_i for each gene minimizes a term in the summation and this is further complicated when $N \gg 1$. Beyond minimizing the concentration of all genes except gene j (essentially reducing this problem to the single-gene scenario), there appears to be no single scenario except setting $\varepsilon_j \approx 0$ that will ensure the many gene scenario is equivalent to the single-gene scenario we covered during class.

Question 2.

All code used for this question was prepared in MATLAB and is available in the attached GitHub repository.

Part A.

I have summarized the three-state model for PFK activity

#	Microstates	Microstate Weight	PFK Activity
0	+ PFK - 3'-5'-AMP - F6P	1	−
1	+ PFK - 3'-5'-AMP + F6P	W_1	+
2	+ PFK + 3'-5'-AMP + F6P	$W_2 f_I$	+++

Part B.

To determine W_1, I assumed that when [3'-5'-AMP] = 0 that state 2 would be completely non-active ($W_2 f_I = 0$) so $v(\ldots)_j = \frac{W_1}{1+W_1}$. This allowed for W_1 to be solved for using the initial rate provided and r_1 based on the constants provided.

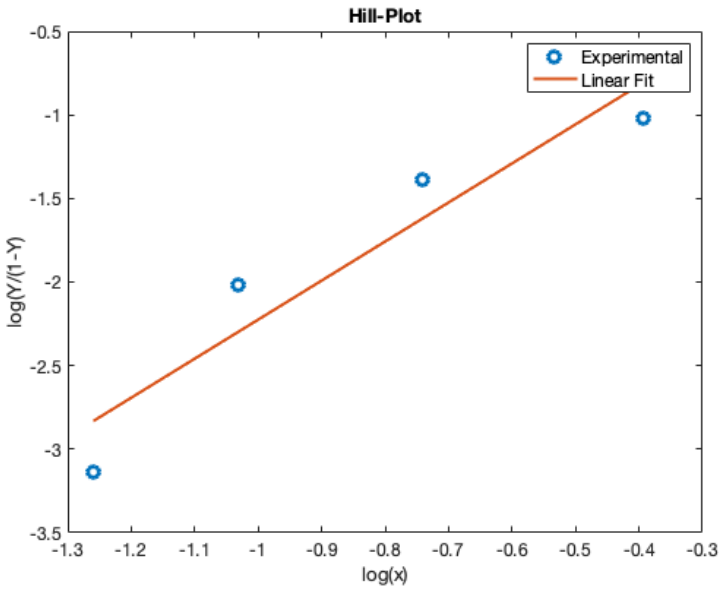
Using the r_1 and W_1 values from before, I solved for the combined $W_1 f_I$ term for each rate in the provided dataset. I assumed that f_I cannot be greater than 1 ($f_I \leq 1$) and that at [3'-5'-AMP] = .99 mM, $f_I \approx 1$ since the rate had essentially plateaued. As a result, I took the value of W_2 to be equal to the value of $W_2 f_I$ at [3'-5'-AMP] = .99 mM.

I then found the values for f_I for each data point by dividing by W_2. I adjusted the hill-type binding function provided to us to match that of other hill-type functions: $f_I = \frac{x^{n_i}}{K_i^{n_i} + x^{n_i}}$. This assumes that K_i is equivalent to K_A, the 3'-5'-AMP concentration obtaining half-maximum response, rather than the dissociation constant K_D. This is a matter of form though, where $K_D =$

$K_A^{n_i}$, but K_A ensures the hill-type binding function works. To obtain a hill-plot, I used the log-form of the hill-type function: $\log\left(\frac{f_I}{1-f_I}\right) = n_i \log(x) - n_i \log(K_i)$. I normalized the provided f_I in this way to create a linear hill-plot, from which I took the slope as the order parameter and the x-intercept as the $\log(K_i)$. In doing this, I excluded the first value where [3'-5'-AMP] = 0 and the final value where $f_I \approx 1$ as the small denominator generates a large y-value that is inconsistent with the remaining data points.

The parameter values obtained through this process are summarized below, along with the hill-plot used to obtain some of them.

Parameter	Value	Units
W_1	.0451	Dimensionless
W_2	74.028	Dimensionless
n_i	2.337	Dimensionless
K_i	.9534	mM



$$\nu(\ldots)_j = \frac{\sum_{i\in\{ij\}} W_i f_i(\ldots)}{\sum_{j\in c_j} W_j f_j(\ldots)} = \frac{W_1 + W_2 f_2}{1 + W_1 + W_2 f_2} \rightarrow \nu(\ldots)_j = \frac{W_1}{1+W_1}$$

$$\hat{r}_j([3'\text{-}5'\text{-AMP}]=0) = r_1 \nu(\ldots)_j = K_{cat} E_t \left(\frac{FGP}{K_{FGP}+FGP}\right)\left(\frac{ATP}{K_{ATP}+ATP}\right)\left(\frac{W_1}{1+W_1}\right)$$

$$3.003\ \tfrac{mM}{hr} = (69.5798\ \tfrac{mM}{hr})\left(\frac{W_1}{1+W_1}\right) \qquad a = b\frac{x}{1+x} \rightarrow a + ax = bx \rightarrow x = \frac{a}{b-a}$$

$$W_1 = \frac{3.003}{69.5798 - 3.003} = .0451$$

$$\hat{r}_j = r_1 \nu(\ldots)_j = r_1 \frac{W_1 + W_2 f_2}{1 + W_1 + W_2 f_2} \rightarrow \text{solve for } W_2 f_2,\ \frac{a}{b} = \frac{c+x}{1+c+x} \rightarrow \frac{a}{b} + \frac{ac}{b} + \frac{ax}{b} = c + x \rightarrow \frac{a}{b} + \frac{ac}{b} - c = x\left(1 - \frac{a}{b}\right) \rightarrow x = \frac{\frac{a}{b} + \frac{ac}{b} - c}{1 - \frac{a}{b}}$$

$$W_2 f_2 = \frac{\frac{\hat{r}_j}{r_1} + \frac{\hat{r}_j W_1}{r_1} - W_1}{1 - \frac{\hat{r}_j}{r_1}}, \quad \text{set } W_2 = W_2 f_2(3'\text{-}5'\text{-AMP} \times 1\ mM) \text{ so } f_2 = \frac{W_2 f_2}{W_2}$$

$$\text{Hill-type binding function: } f_i = \frac{\left(\frac{x}{K_i}\right)^{n_i}}{1 + \left(\frac{x}{K_i}\right)^{n_i}}, \text{ expanded: } \frac{\left(\frac{x}{K_i}\right)^{n_i}}{1 + \left(\frac{x}{K_i}\right)^{n_i}} = \frac{\frac{x^{n_i}}{K_i^{n_i}}}{1 + \frac{x^{n_i}}{K_i^{n_i}}} \times \frac{K_i^{n_i}}{K_i^{n_i}} = \frac{x^{n_i}}{K_i^{n_i} + x^{n_i}}$$

$$\text{log-normalized: } \log\left(\frac{f_x}{1-f_x}\right) = n_i \log(x) - n_i \log(K_i)$$

Part C.

The proposed model fits the experimental data well. The model is especially well fit at low and high concentrations and deviates when the change in PFK rate is high. The relatively low number of data points when the rate is rapidly changing likely contributes to this deviation and increasing the sampling could improve the accuracy of the model.