PhySci/MiMG/CaSB M178

Homework 7

Due: 11/21/23 at 12:00PM PDT

Notes: This homework involves performing simulations of the logic gates we've been discussing in the last two class meetings. In the same folder on Bruinlearn where you obtained this document, you will also find a file called "HW7_template.ipynb" that contains a template Jupyter notebook that you can use as a starting point to complete the questions below. Please modify this notebook and use it as the starting point for answering the following problems.

To submit your homework, please answer the questions below. Note that you will have to paste in several graphs that you generate using the Jupyter notebook. After completing the questions, **save this document as a PDF and upload it to Gradescope**. You **must also upload the Jupyter notebook to Bruinlearn** to do so, navigate to the "Homework" section on the left-hand side of the course Bruinlearn website. There you will see an assignment entitled "Homework 7 Jupyter submission." You can upload your Jupyter file (which should be a .ipynb file). Make sure you upload your Jupyter notebook by the due date/time (11/21/23 at 12:00PM PDT).

Problems

In class, we talked about logic gates controlling gene expression. To begin this homework, we will start with a simple model of gene expression. In this model, a transcription factor, binds to the promoter region of a gene and converts it from an inactive to active state. In this active state, transcription can occur, and mRNA is produced. Let *TF* denote the transcription factor or input signal, whose value can change over time. The effective rate of activation of the promoter region is not only a function of the abundance of *TF*, but also a function of the binding affinity of *TF* to the promoter region, K_d. We include this information in a Hill Equation:

$$H(t) = \frac{TF(t)}{TF(t) + K_d}$$

Assuming the rates of *TF* binding to and unbinding from the promoter region are fast, *H* describes the proportion of binding sites that are occupied by the transcription factor. Hence if k_a is the maximum rate of promoter activation, the effective rate dependent on *TF* is given by the product $k_a * H(t)$.

In summary we have the following reactions:

Reactions	**Description**

$pr \xrightarrow{k\_a*H} pr\_a$	Activation of the promoter region
$pr\_a \xrightarrow{k\_d} pr$	Deactivation of the promoter region
$pr\_a \xrightarrow{k\_syn} pr\_a + tr$	Synthesis of mRNA transcript
$tr \xrightarrow{k\_deg}$	Degradation of mRNA transcript

1. (14 points) First write down the change equations for the model described above.

pr' = -k_a*H*pr +k_d*pr_a

pr_a' = +k_a*H*pr -k_d*pr_a -k_syn*pr_a +k_syn*pr_a

tr' = +k_syn*pr_a -k_deg*tr

In the section of code called "simple gene expression model" implement the change equations to simulate the model. Note how the code defines the value of the Hill Equation (H); use this value in your implementation. Once you have defined your model equations, run the cell containing the model code.

Now we will simulate the model in the section of code called "Simulation of simple model". In this section, we have defined our initial conditions and default parameter values which are also listed below for your reference.

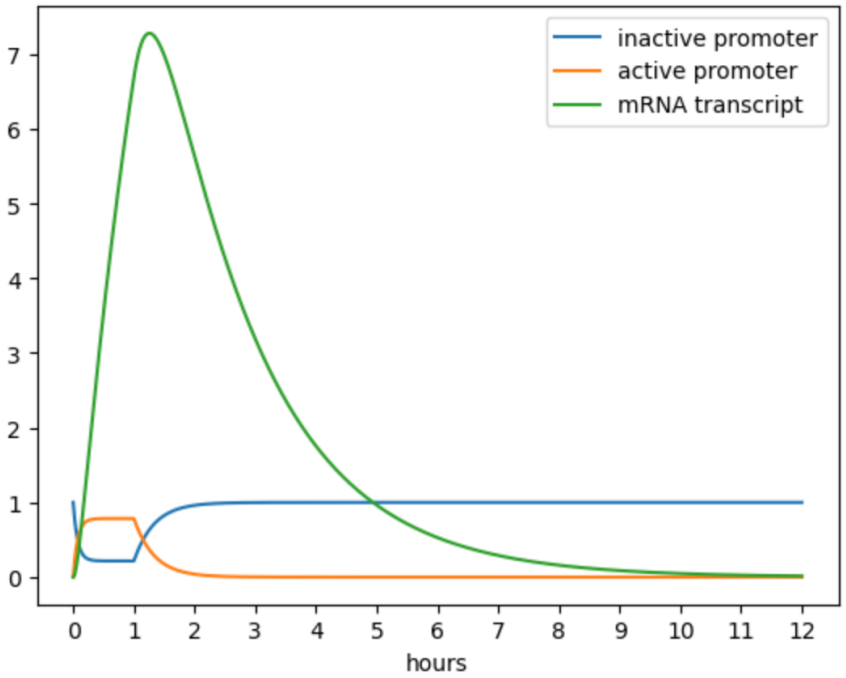
Parameter	Value
k_a	0.2 min^{-1}
k_d	0.05 min^{-1}
Kd	0.1
k_syn	0.1 min^{-1}
k_deg	0.01 min^{-1}

For the simulations in this homework, the profile of activation of *TF* will be a box function (identical to that of IKK activation from HW 6). Hence to describe the temporal dynamics of *TF*, we have to specify an on time, off time, and amplitude (TF_on, TF_off, and TF_amp respectively). By default, TF_on = 0, TF_off = 60 (minutes), and TF_amp = 1.

Run the section of code for "Simulation of simple model" and the section of code for "Checking Model Implementation". Check to see that the values on the right (from simulation of your model implementation) match the values on the

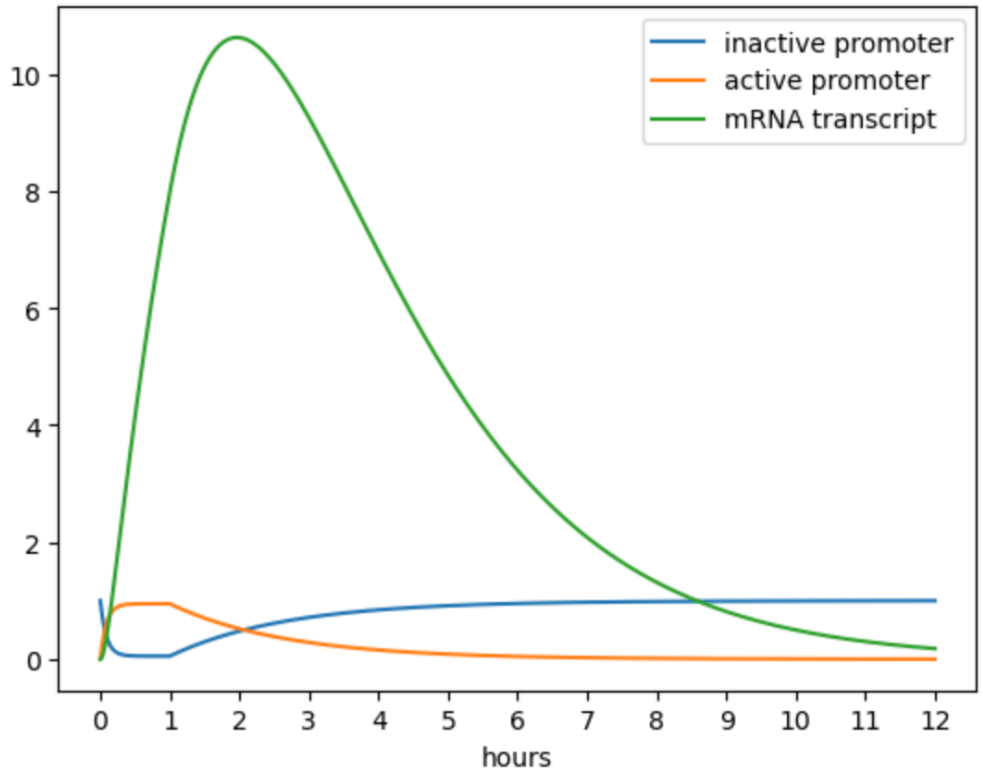
left (from simulation of the correct model implementation). If the values don't match, double check your change equations and code before proceeding.

Run the section of code called "Plot Dynamics" and paste your graph here. Describe the profile of gene expression over time.



There is quick activation of the promoter leading to a spike in transcription activity and thus, gene expression activity between 0-2 hours. mRNA transcript concentration peaks at around 2 hours. However, the active promoter concentration begins declining around 1 hour, leading to an increase in inactive promoter until eventually the active promoter becomes nearly flat after 2 hours and remains that way, indicating no gene expression. There is a lingering presence of mRNA after transcription has stopped because the peak of mRNA transcript is at 2 hours, but active promoter stops at 1 hour. When the promoter is active, gene expression increases steeply, but when the promoter becomes inactive, gene expression begins to decline.

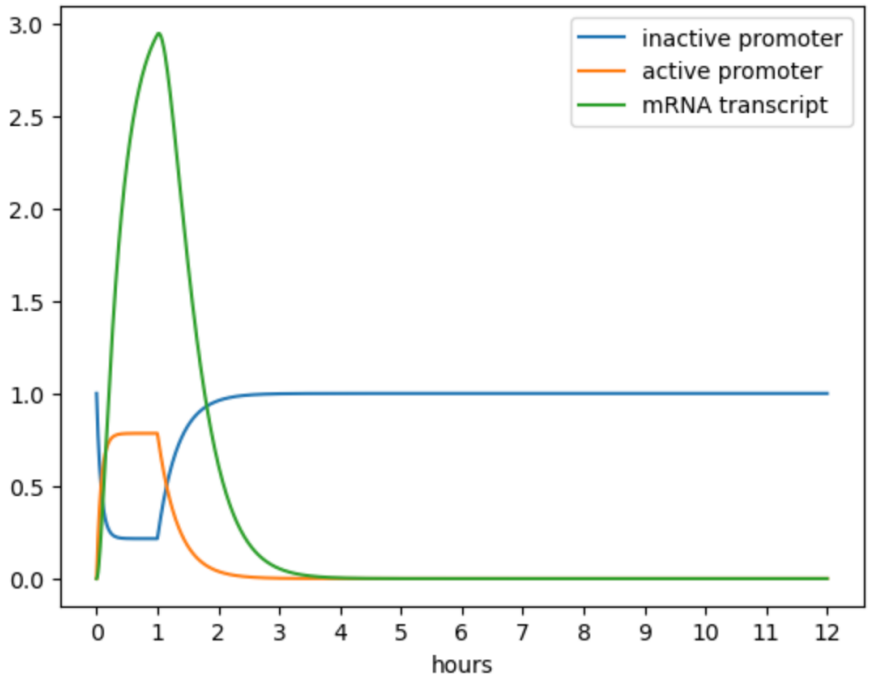
Modify the value for the parameter k_d in the section "Simulation of simple model" to 0.01 and rerun the section. Run the section of code called "Plot Dynamics" and paste your graph here. How does it affect the profile of gene expression over time?



The peak of the mRNA transcript occurs at the same time of 2 hours, but there is a higher concentration his time at its peak and it also approaches it more sharply this time. The peak mRNA transcript is now around 10.5 here, but was around 7.5 with the original parameter. mRNA transcript also doesn't entirely degrade once active promoters peak because mRNA transcript doesn't flatline at time = 12 hours with k_d set to 0.01, but it does flatline earlier with k_d set to 0.05 in the original model. There is quicker production of mRNA transcripts here, which leads to more robust gene expression as shown by a higher peak in mRNA transcript, and also a more gradual decline after the promoter becomes inactive.

Set k_d back to its default value.

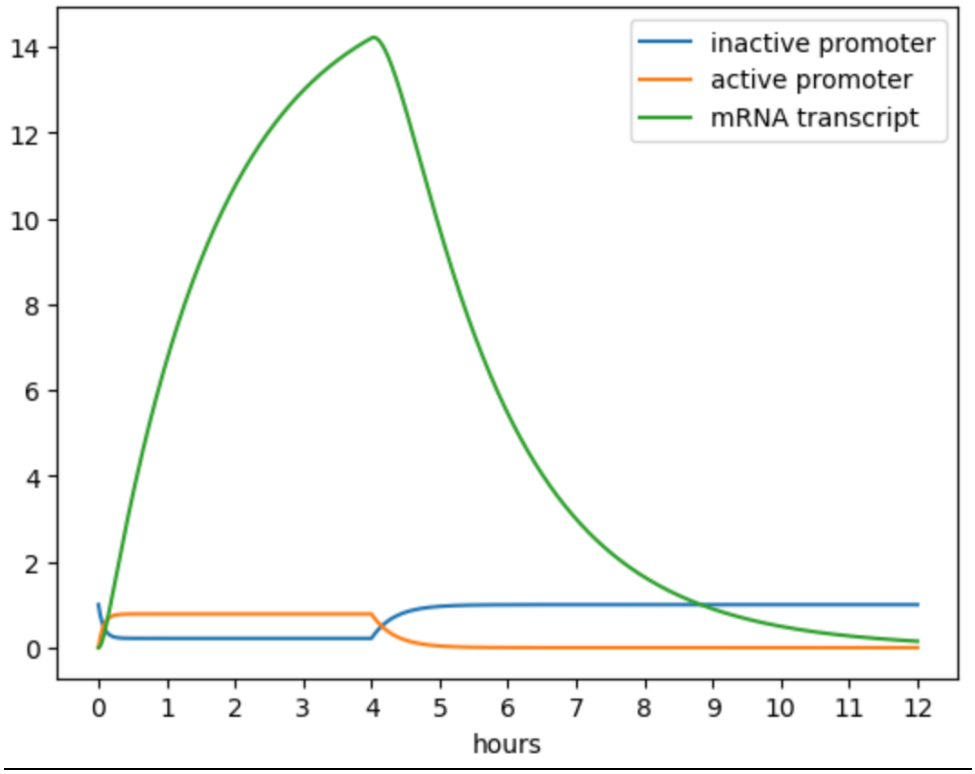
Modify the value for the parameter k_deg in the section "Simulation of simple model" to 0.05 and rerun the section. Run the section of code called "Plot Dynamics" and paste your graph here. How does it affect the profile of gene expression over time?



The overall gene expression appears to be reduced compared to the initial graph because the peak of the mRNA transcript is much lower and barely reaches 3 units whereas in the original graph, mRNA transcript peaked at around 7 units. Active and inactive promoter concentrations appear to still have the same dynamics where they flatline at 0 and 1 unit, respectively, in the long run.

Set k_deg back to its default value.

Modify the value for the parameter TF_off in the section "Simulation of simple model" to 240 and rerun the section. Run the section of code called "Plot Dynamics" and paste your graph here. How does it affect the profile of gene expression over time?



The mRNA transcript peaks at a much higher value at just over 14 units and peaks at a much later time around 4.5 hours. Active and inactive promoter also remain constant and stable for a longer time interval during transcription than before. Active promoter stays at 1 unit of concentration up to 4 hours and inactive promoter remains at 0 units of concentration up to 4 hours before they begin to decrease and increase, respectively. This would help explain the higher mRNA transcript peak since the active promoters last longer allowing gene expression to occur for a longer period of time.

2. (12 points) Now we will simulate a logical "OR" gate. We now have two transcription factors, TF_1 and TF_2, either of which can activate the promoter region. We enforce this logic by modifying our Hill Equation:

$$H(t) = 1 - \frac{K_{d1}}{TF_1 + K_{d1}} \frac{K_{d2}}{TF_2 + K_{d2}}$$

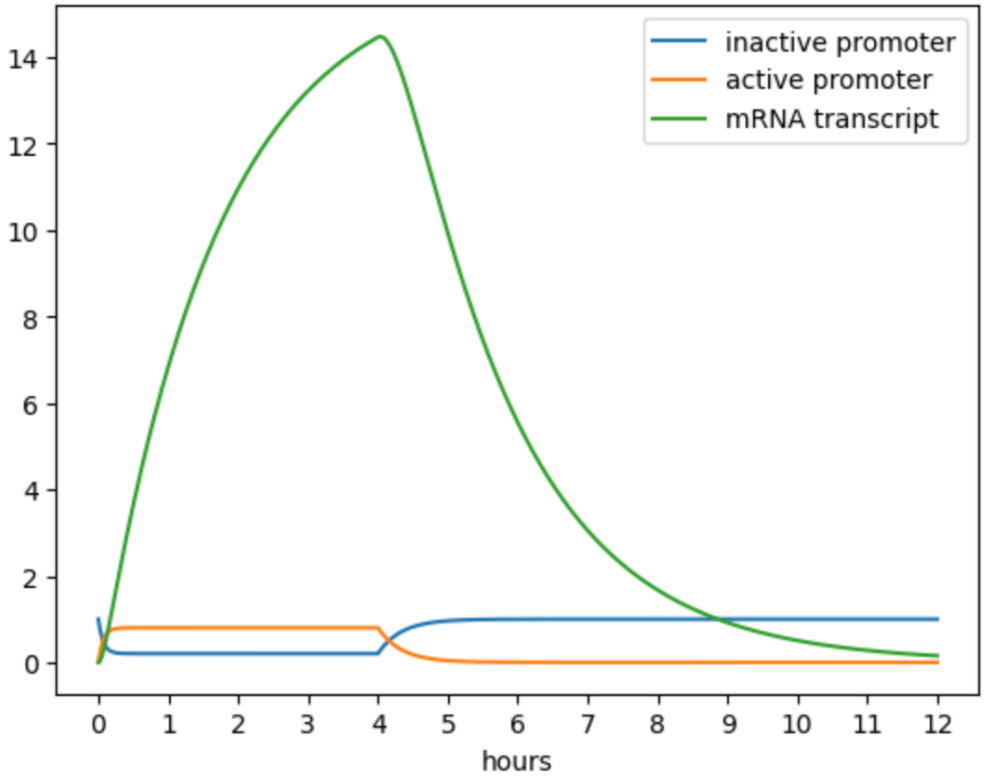
Now only if both TF_1 and TF_2 are small in value will the value of H be small. If either TF_1 or TF_2 is large in value, H will be large in value.

This model will be implemented in the section of code "OR gate gene expression model". Note that now, we again describe the dynamics of two transcription factors, TF_1 and TF_2. Additionally, we have defined our modified Hill Equation. Besides this change to the expression $H(t)$, the gene expression

model is unmodified. **Take your change equations from the "simple_model" and place them into the "OR_model". Run the cell containing the model code.**

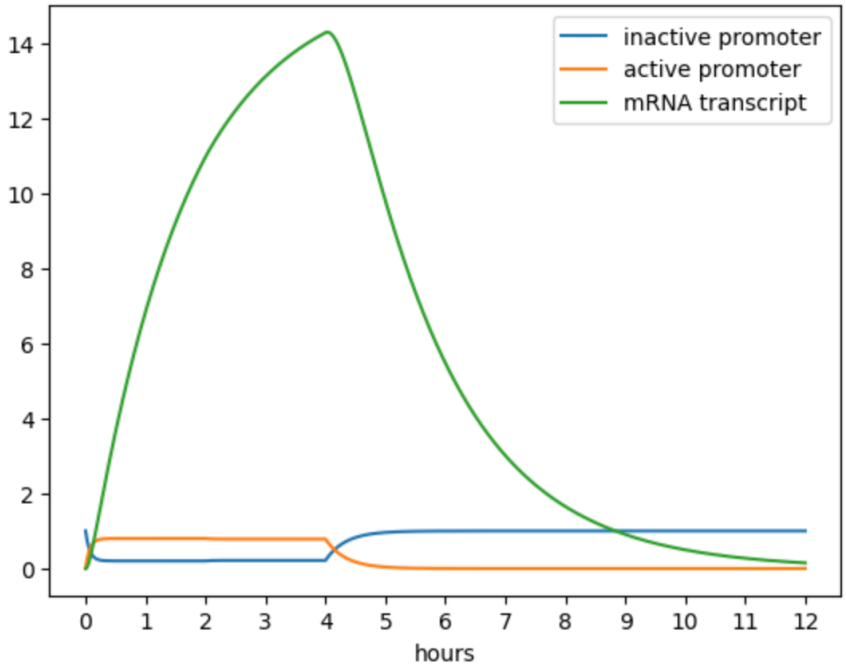
Run the section of code for "Simulation of OR model". Note that we have set the K_d of both transcription factors (Kd1 and Kd2) to the same value, 0.1. Additionally, we define the profiles of activation of TF_1 and TF_2. By default, TF_1 is active from time 0 to time 240 minutes with amplitude 1 and TF_2 is active from time 0 to time 240 minutes with amplitude 1.

Run the section of code called "Plot Dynamics – OR model" and paste your graph here. Describe the profile of gene expression over time. Compare to "simple_model" with TF_off = 240.

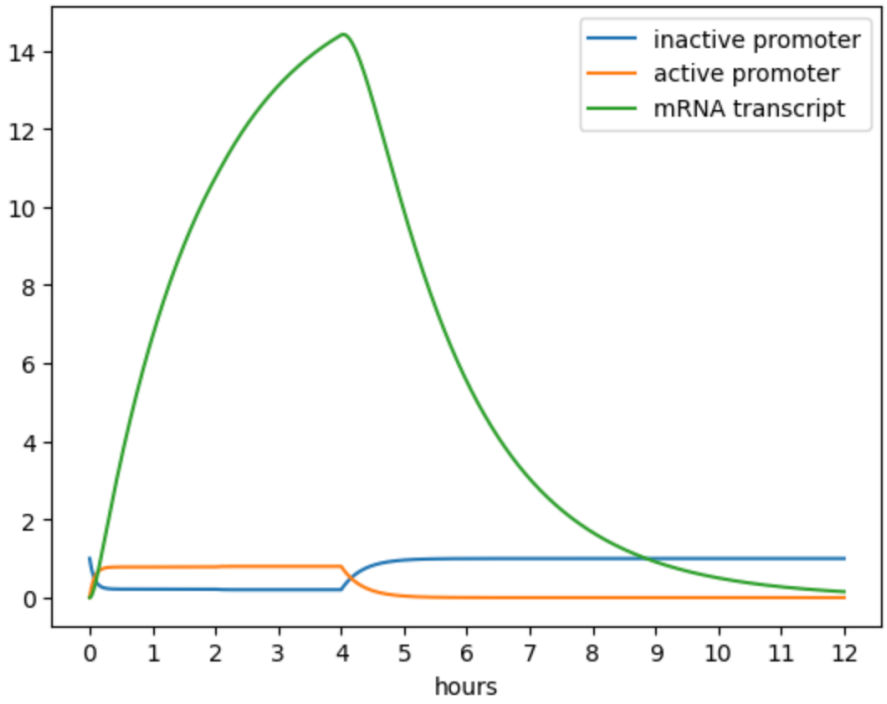


Gene expression is continually increases until 4.5 hours and then begins to decline once the promoter becomes inactive. The profile of gene expression appears to be nearly identical to the simple_model with TF_off = 240. No noticeable differences can be found since both have similar dynamics and have variable values that are indistinguishable. TF1_off and TF2_off were both set to 240 in the OR model.

Modify the timing *of* TF_2 activation profile so that TF2_on = 0 and TF2_off = 120. Rerun the section of code called "Plot Dynamics – OR model" and paste your graph here.



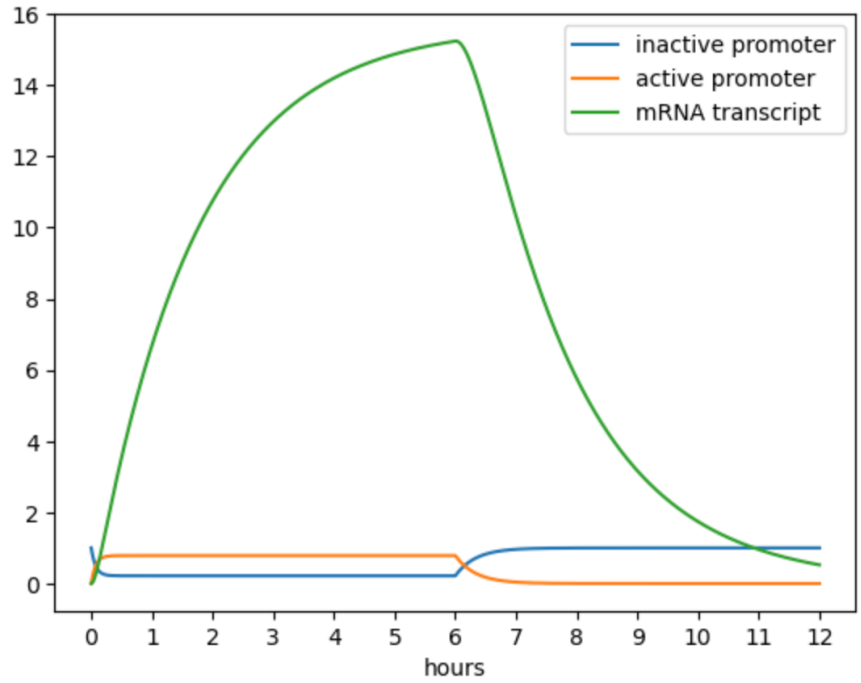
Modify the timing *of* TF_2 activation profile so that TF2_on = 120 and TF2_off = 240. Rerun the section of code called "Plot Dynamics – OR model" and paste your graph here.



Explain any differences in gene expression over time in the last two plots compared to the OR model with default parameters. If there are no differences, explain why.

There were no differences because the transcription factors possibly have a redundant function meaning the presence of either transcription factor is enough to activate the promoter and initiate transcription to its full capacity. It also implies that the parameter we adjusted is not critical to the promoter activation by these transcription factors under the OR gate conditions, which is designed to produce a consistent output as long as one transcription factor is active. The lack of difference helps reaffirm that the OR gate design helps maintain gene expression despite variations in certain parameters. Having both transcription factors present would result in the same gene expression profile.

Modify the timing *of* TF_2 activation profile so that TF2_on = 240 and TF2_off = 360. Rerun the section of code called "Plot Dynamics – OR model" and paste your graph here. Explain the difference in gene expression over time compared to the OR model with default parameters.



mRNA peaks at a higher unit concentration at just over 15 and peaks at time = 6 hours meaning there is constant and stable active promoter for a longer time interval up to 6 hours and there is constant and stable lack of inactive promoter up to 6 hours too. The mRNA transcript here also requires a longer time interval to flatline to 0 compared to the OR model with the original parameters. However, the overall dynamics appear to be similar but differ in specific values.

This result could've occurred because of the OR gate mechanism. While TF1 was on and TF2 was off, TF1 being on allowed gene expression to occur. Then as TF1 turned off and TF2 was turned on, TF2 being on allowed gene expression to continue due to only needing at least 1 TF present to keep active promoters present.

Describe in general a biological context in which a logical "OR" gate controlling gene expression would be advantageous.

An "OR" gate for gene expression would be advantageous in situations where an organism needs to respond to varying environmental conditions. For example, bacteria would benefit from an "OR" gate to active stress response genes in the presence of either high temperature OR low pH. This helps survivability by responding effectively to either stressor without the need for both signals simultaneously. Thus, it's efficient when multiple signals could cue similar adaptive responsive without needing separate signaling pathways that lead to the same result.

3. (12 points) Now we will simulate a logical "AND" gate. Again, we now have two transcription factors, *TF₁* and *TF₂*, which are both required for activation of the promoter region. We enforce this logic by this Hill Equation:

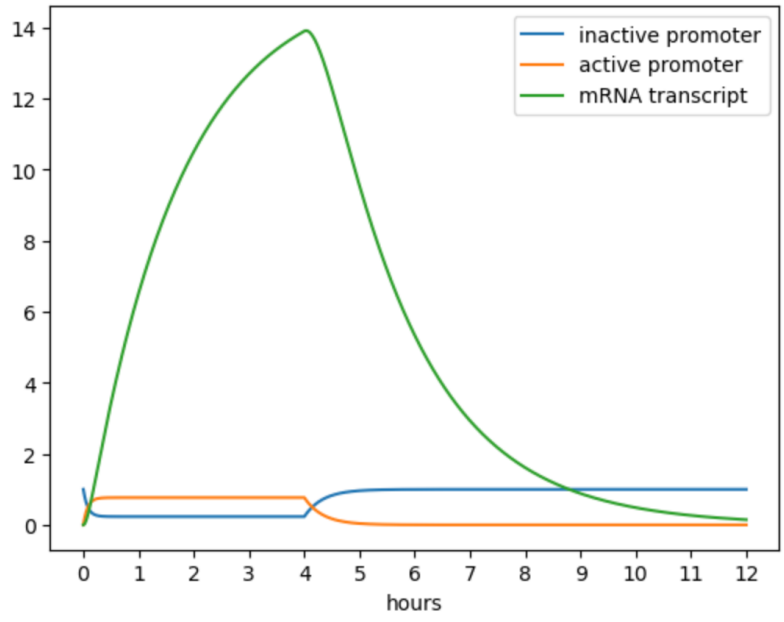
$$H(\mathrm{t}) = \frac{\mathrm{TF}_1}{\mathrm{TF}_1 + \mathrm{K}_{\mathrm{d1}}} \frac{\mathrm{TF}_2}{\mathrm{TF}_2 + \mathrm{K}_{\mathrm{d2}}}$$

Now only if both *TF₁* and *TF₂* are large in value will the value of *H* be large in value. If either *TF₁* or *TF₂* is small in value, H will be small in value.

This model will be implemented in the section of code "AND gate gene expression model". Note that now, we describe the dynamics of two transcription factors, *TF₁* and *TF₂*. Additionally, we have defined our modified Hill Equation. Besides this change to the expression for *H(t)*, the gene expression model is unmodified. **Take your change equations from the "simple_model" and place them into the "AND_model". Run the cell containing the model code.**

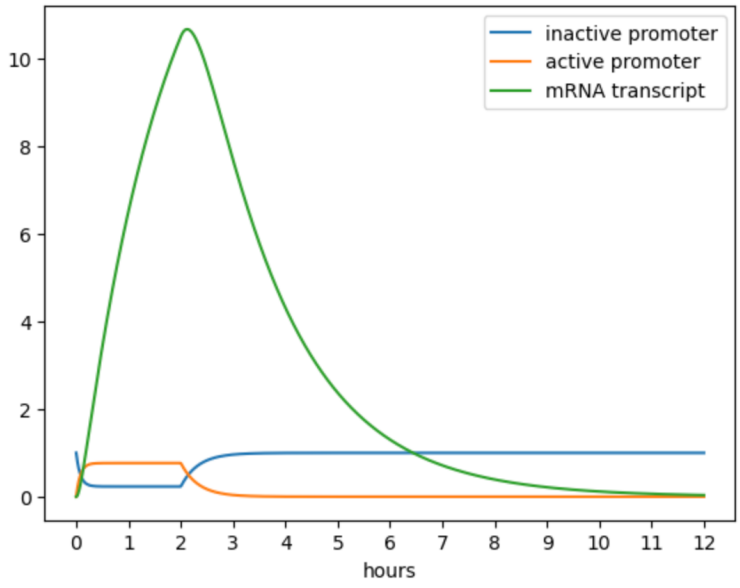
Run the section of code for "Simulation of AND model". Note that we have set the K_d of both transcription factors (Kd1 and Kd2) to the same value, 0.1. Additionally, we define the profiles of activation of *TF₁* and *TF₂*. By default, *TF₁* is active from time 0 to time 240 minutes with amplitude 1 and *TF₂* is active from time 0 to time 240 minutes with amplitude 1 (the same as the "OR" model default).

Run the section of code called "Plot Dynamics – AND model" and paste your graph here. Describe the profile of gene expression over time. Compare to "simple_model" with TF_off = 240.



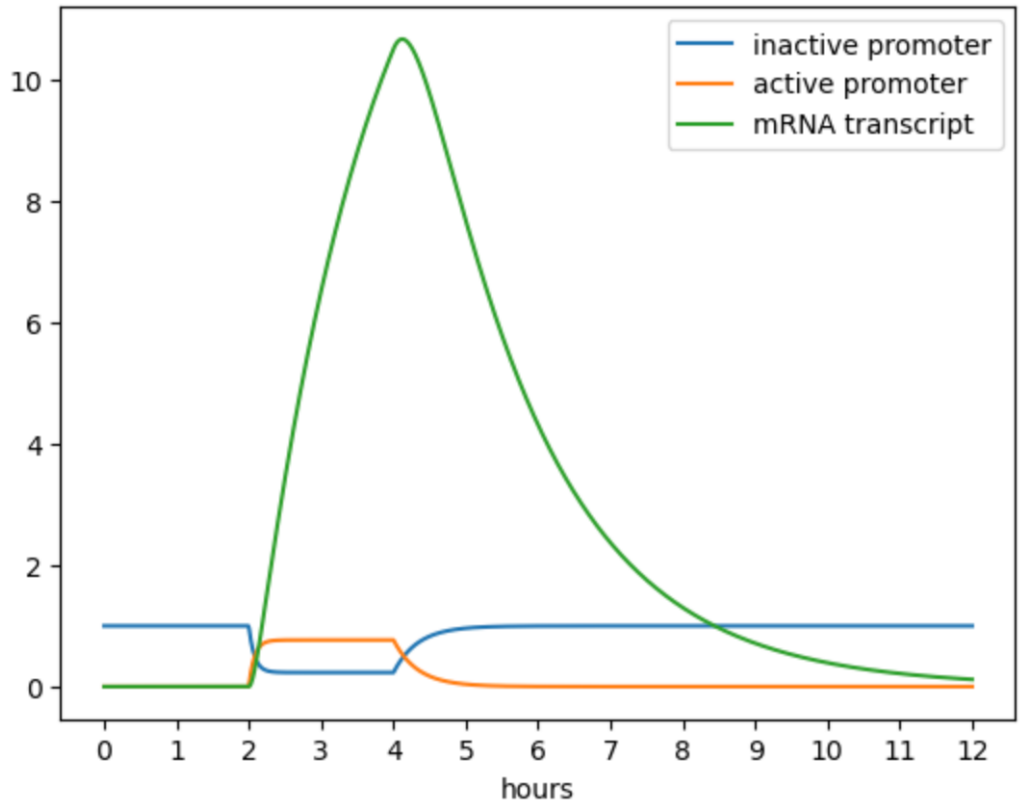
Gene expression is continually increases until 4.5 hours and then begins to decline once the promoter becomes inactive. The profile of gene expression appears to be nearly identical to the simple_model with TF_off = 240. No noticeable differences can be found since both have similar dynamics and have variable values that are indistinguishable. TF1_off and TF2_off were both set to 240 in the AND model.

Modify the timing *of* TF_2 activation profile so that TF2_on = 0 and TF2_off = 120. Rerun the section of code called "Plot Dynamics – AND model" and paste your graph here. Explain the difference in gene expression over time compared to the AND model with default parameters.



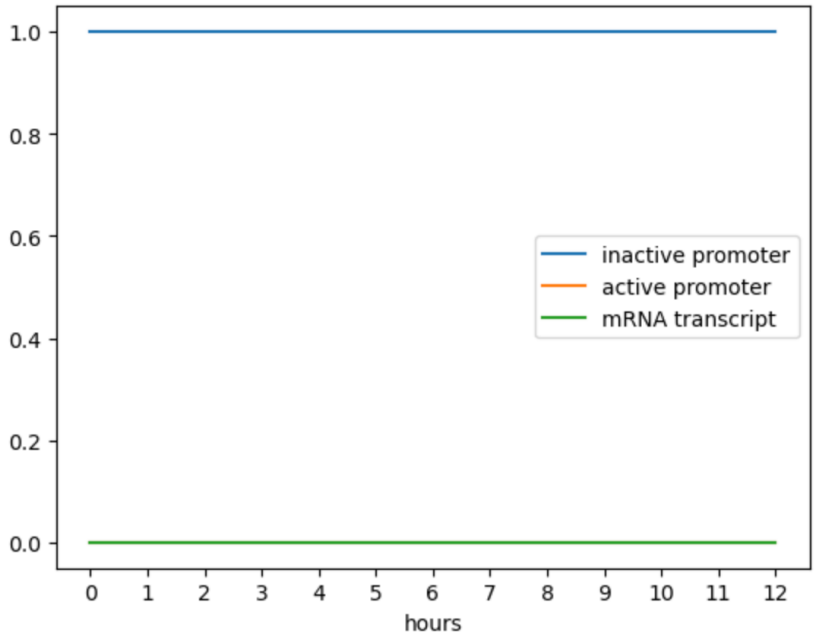
The mRNA transcript peaks earlier at 2 hours and peaks at a lower unit concentration at just above 10 compared to the original AND model, which had mRNA transcript peak at just over 14 units of concentration in 4.5 hours. The mRNA transcript level also flatlines in a smaller time interval here. The active promoter also is only present up to 2 hours here compared to it being present up to 4 hours in the original AND model.

Modify the timing *of* TF_2 activation profile so that TF2_on = 120 and TF2_off = 240. Rerun the section of code called "Plot Dynamics – AND model" and paste your graph here. Explain the difference in gene expression over time compared to the AND model with default parameters.



The mRNA transcript peaks later at around the same time at 4.5 hours but peaks at a lower unit concentration at just above 10 compared to the original AND model, which had mRNA transcript peak at just over 14 units of concentration in 4.5 hours. However, the mRNA transcript reaches its peak much faster since active promoter isn't present until 2 hours here. Transcription only occurs from 2 to 4 hours here compared to 0 to 4 hours in the original AND model.

Modify the timing *of* TF_2 activation profile so that TF2_on = 240 and TF2_off = 360. Rerun the section of code called "Plot Dynamics – AND model" and paste your graph here. Explain the difference in gene expression over time compared to the AND model with default parameters.



No mRNA transcript is ever made meaning there's never gene expression. There is also always inactive promoter and no active promoter. Everything remains at the same concentration in this AND model. This could be explained by the requirement of both transcription factors needing to be present in the AND model, which isn't being met here.

Describe in general a biological context in which a logical "AND" gate controlling gene expression would be advantageous.

Having a logical "AND" gate controlling gene expression could be advantageous for processes requiring precise coordination of multiple signals for proper function or differentiation. For example, during development in a multicellular organism, certain cells need to receive signals to know where they are in the tissue and which stage of development they are before specializing for specific cell functions. An "AND" gate ensures that gene expression is activated only when both conditions are met so there's no premature or inappropriate activation of genes, which could be detrimental. This could also relate to apoptosis where you want to ensure that a cell dies only when multiple conditions confirm that the cell's state is beyond repair.

4. (22 points) Now we will simulate a model in which the promoter region has to undergo two sequential transitions in order to reach its active state. The promoter region is first closed (inaccessible to RNA polymerase) and opens with the binding of a transcription factor, TF_1. The open promoter region is then converted to the active state by a second transcription factor, TF_2.

In summary we have the following reactions:

Reactions	**Description**
$pr\_c \xrightarrow{k\_a1 * H1} pr$	Opening of the promoter region
$pr \xrightarrow{k\_d1} pr\_c$	Closing of the promoter region
$pr \xrightarrow{k\_a2 * H2} pr\_a$	Activation of the promoter region
$pr\_a \xrightarrow{k\_d2} pr$	Deactivation of the promoter region
$pr\_a \xrightarrow{k\_syn} pr\_a + tr$	Synthesis of mRNA transcript
$tr \xrightarrow{k\_deg}$	Degradation of mRNA transcript

Where the hill equation terms are:

$$H_1(t) = \frac{TF_1(t)}{TF_1(t) + K_d}$$
$$H_2(t) = \frac{TF_{2(t)}}{TF_{2(t)} + K_d}$$

First write down the change equation for the new species, *pr_c*.

pr_c' = -k_a1*H1*pr_c +k_d1*pr

Do any of the other change equations from the simple model need to be modified? Write them here if so.

pr' = -k_a2*H2*pr +k_d2*pr_a +k_a1*H1*pr_c -k_d1*pr
pr_a' = +k_a2*H2*pr -k_d2*pr_a

tr' = +k_syn*pr_a -k_deg*tr (Remains the same)

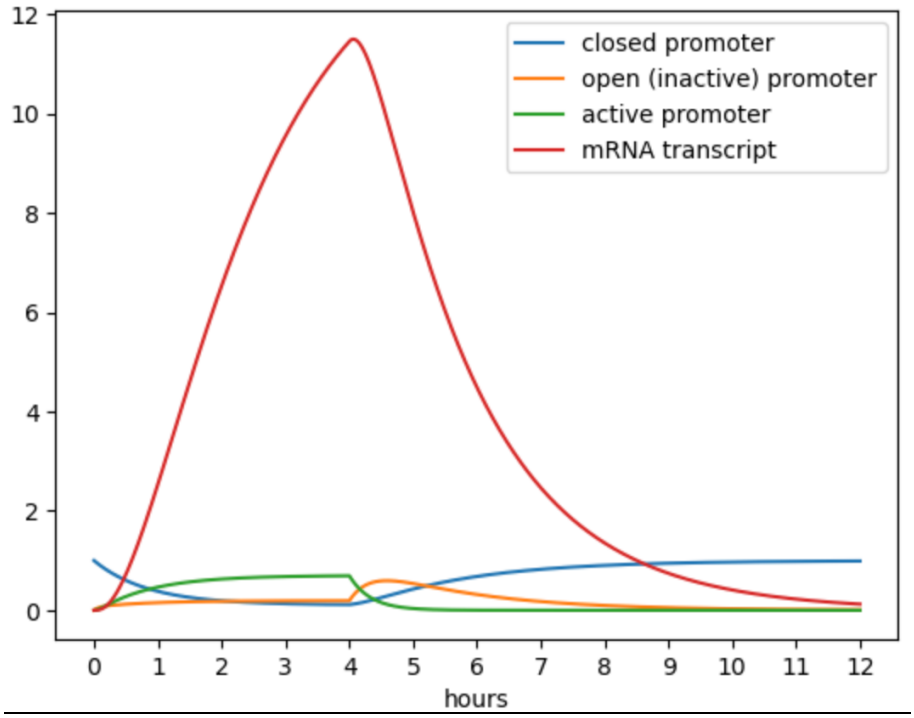
Implement the change equations for this model in the section called "Two step gene expression model". Note how the code defines the value of the Hill Equations (H1, H2); use these values in your implementation. Once you have defined your model equations, run the cell containing the model code.

Now we will simulate the model in the section of code called "Simulation of two step model". In this section, we have defined our initial conditions and default parameter values. Note that k_a1 is lower than ka_2, so the transition from

closed to open is slower than the transition from open to active promoter region. By default, both TF_1 and TF_2 are turned on at time 0 with amplitude 1 and turned off at time 240 minutes.

Run the section of code for "Simulation of two step model" and the section of code for "Checking Model Implementation" below it. Check to see that the values on the right (from simulation of your model implementation) match the values on the left (from simulation of the correct model implementation). If the values don't match, double check your change equations and code before proceeding.

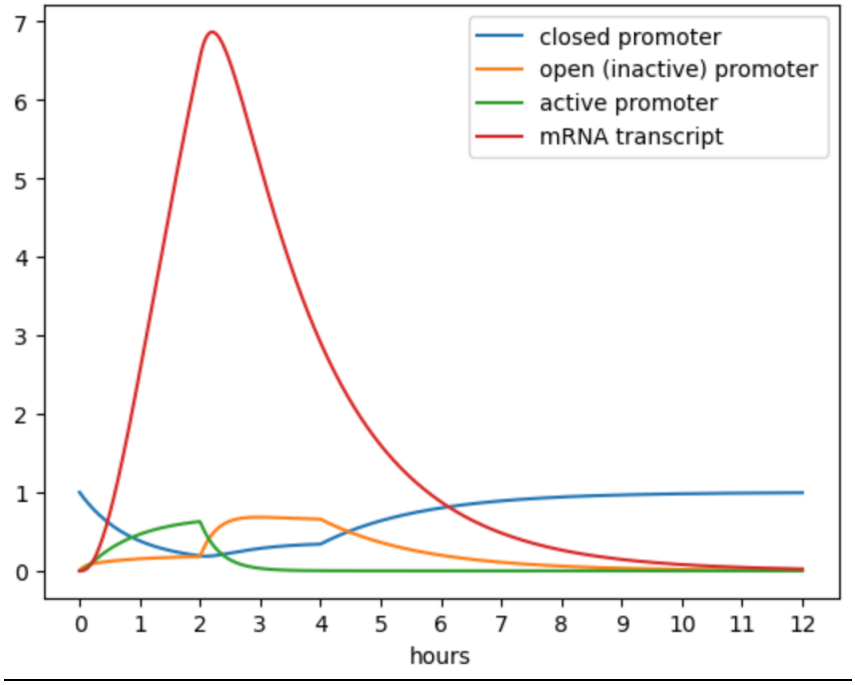
Run the section of code called "Plot Dynamics – two step model" and paste your graph here. Describe the profile of gene expression over time. How does it compare to the simple model results with TF_off = 240?



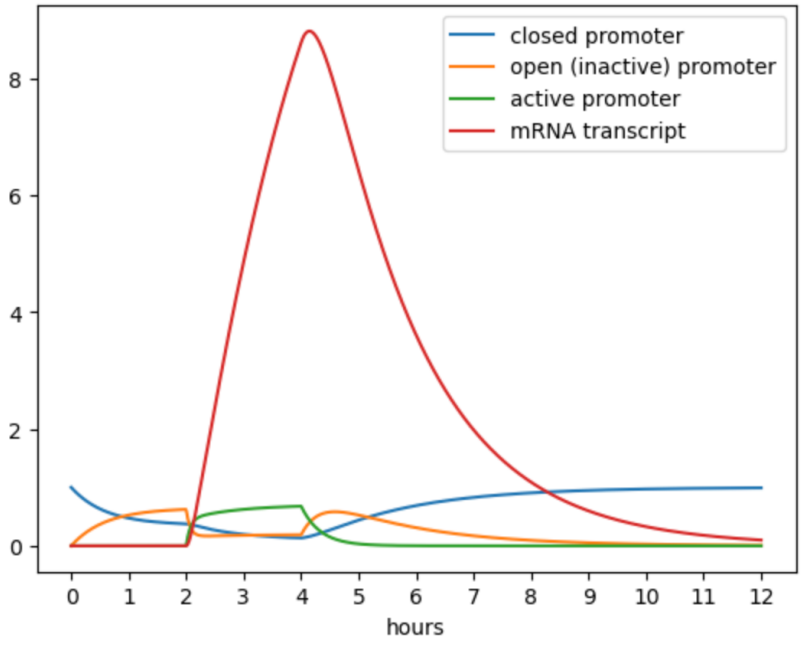
mRNA transcript peaks at a slightly smaller concentration here at just above 11 units of concentration, while mRNA transcript peaked at just over 14 units of concentration in the simple model with TF-off = 240. There are also 3 promoter states in this two step model compared to only 2 in the simple model. Initially, only closed promoter is present, but closed promoter beings declining as active promoter increases. Then, at 4 hours, both closed promoter and open (inactive) promoter increase as active promoter decreases. Open (inactive) promoter increases for a small duration of time and peaks at around 0.8 units until gradually decreasing to 0, but closed promoter concentration increases to 1 and remains stable at that concentration.

We will now run several simulations varying the on and off times for *TF_2*. For each set of values, modify and run the code in "Simulation of two step model". Then paste the resulting graph from "Plot Dynamics – two step model".

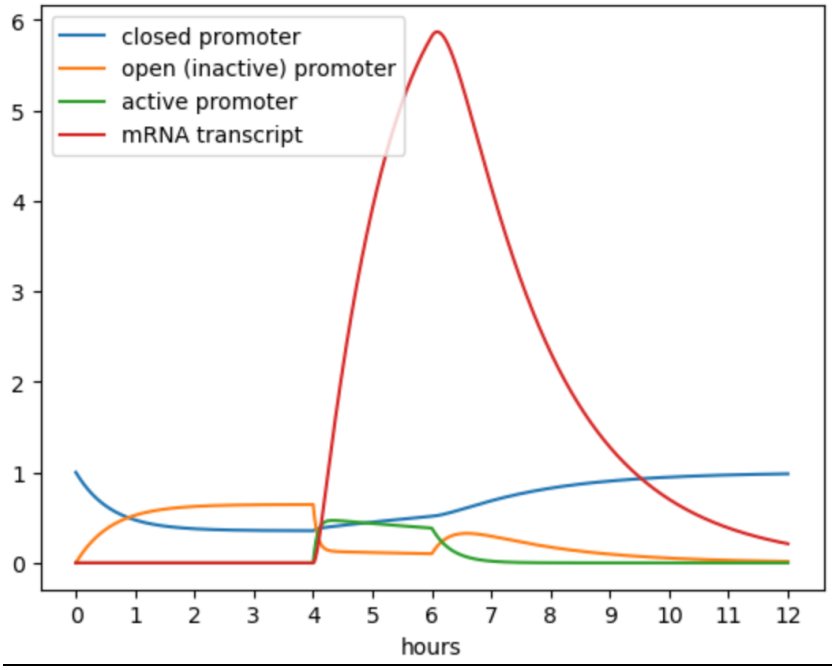
TF2_on = 0 TF2_off = 120



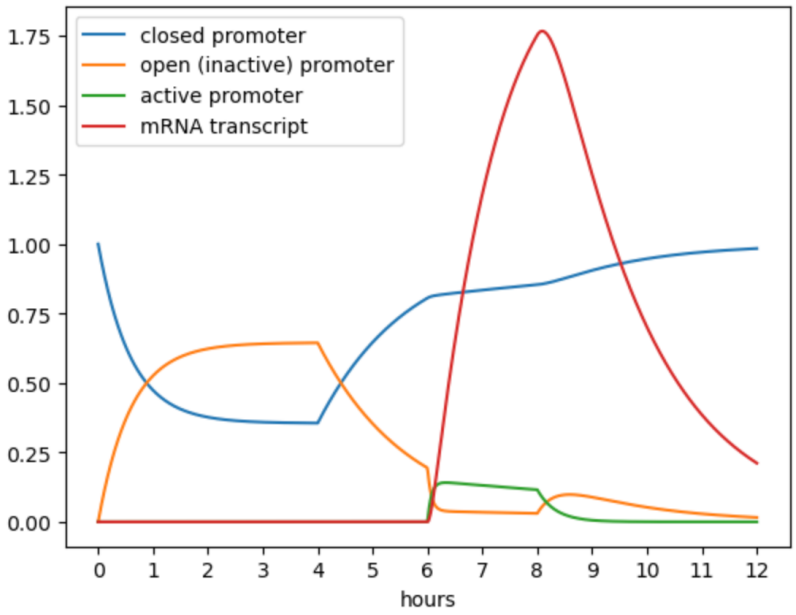
TF2_on = 120 TF2_off = 240



TF2_on = 240 TF2_off = 360



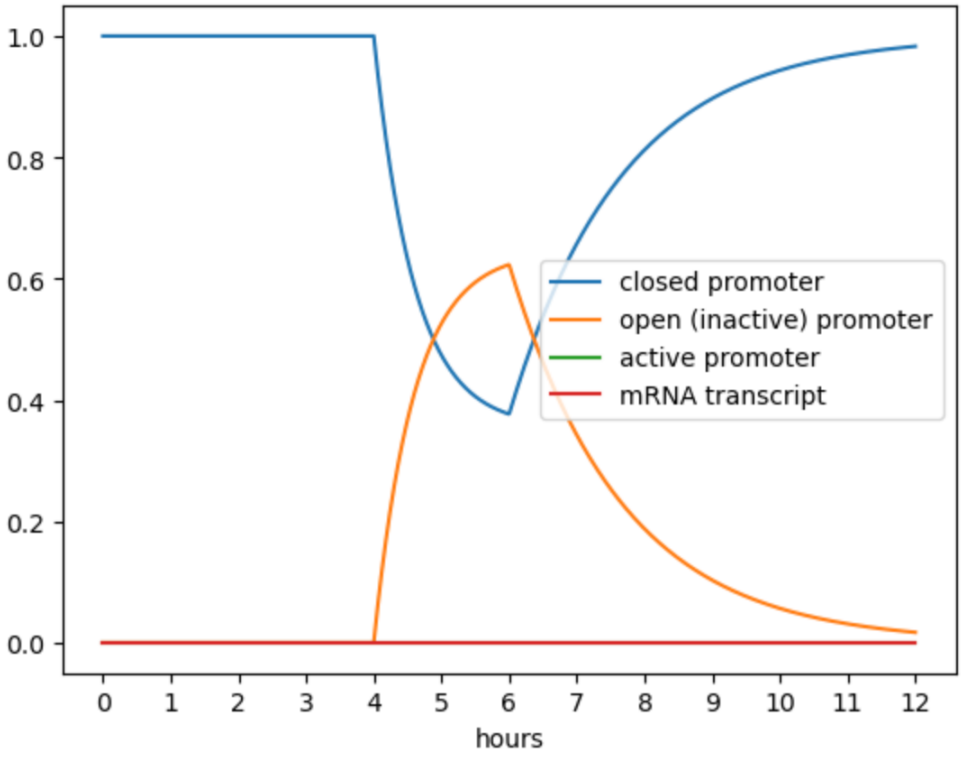
TF2_on = 360 TF2_off = 480



Explain the differences in the profile of gene expression over time for the prior four plots.

In terms of mRNA transcript peak, they peaked at 7 units in hour 2,5, 8.5 units in hour 4.5, 6 units in hour 6.5, and 1.75 units in hour 8.5 for plots 1, 2, 3, and 4, respectively. All the plots showed that as active promoter increased, open (inactive) promoter decrease and as active promoter decreased, open (inactive) promoter increased. All the plots had similar peaks of open (inactive) promoter concentration at around 0.60 units of concentration. All the plots had closed promoter concentration begin and stabilize at 1.0 units of concentration, and had open (inactive) promoter and active promoter flatline at 0 in the long run. They all differed in the time interval in which active promoter was nonzero and open (inactive) promoter was zero, but the time interval in which closed promoter was decreasing was similar from 0 to 4 hours. All the plots also look very sporadic at a glance.

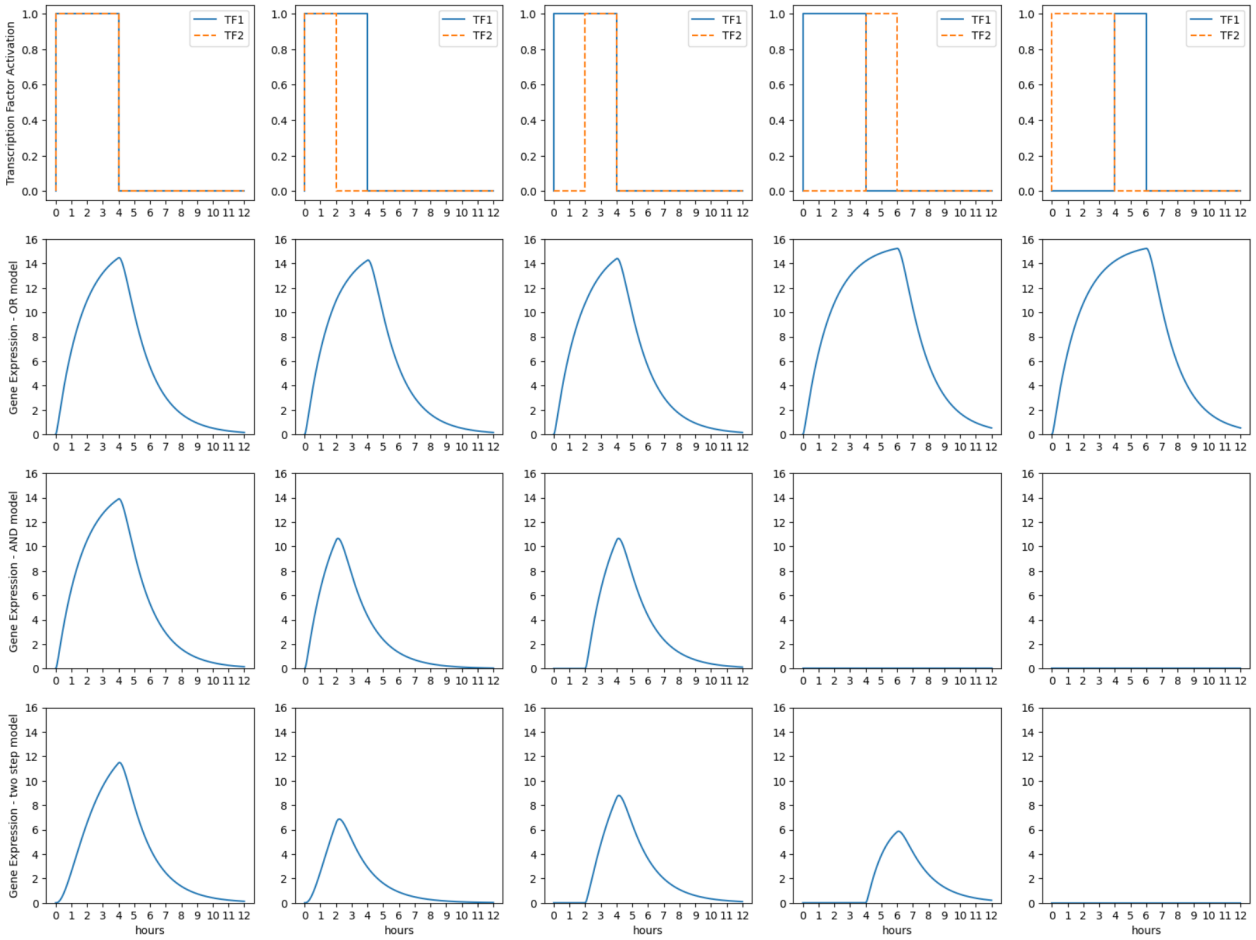
Unlike in the “OR” and “AND” models, the roles of TF_1 and TF_2 are distinct in the “two-step” model. Choosing to alter the timing of TF_1 instead of TF_2 can change the results. **In the section “Simulation of two step model” reset TF2_on = 0 and TF2_off = 240 and set TF1_on = 240 and TF1_off = 360. Run the code in “Simulation of two step model” and then paste the resulting graph from “Plot Dynamics – two step model”.**



How does the profile of gene expression over time compare to that when TF2_on = 240 and TF2_off = 360 ? What does this tell you about the order of transcription factor activation required for gene expression in the "two-step" model if their activation is sequential?

Here, mRNA transcript is never made and there is never any active promoter. This suggests that the promoter never transitions from an open (inactive) state to an active state so the while the promoter is accessible, it's never actually activated. So, sequential activation requires both transcription factors to be present and active within a certain time frame otherwise either TF1 or TF2 will fail to bind and activate the promoter even when it becomes active. If they are not active simultaneously, or at least within a similar time window where their activities overlap, the necessary sequential activation doesn't occur, and gene expression doesn't happen. This tells us that for gene expression to occur in a two-step model, the timing of TF1 and TF2 activation is important. One transcription factor might be responsible for opening the promoter, making it accessible, while the other is required for the actual initiation of transcription.

Finally run the section of code called "Plotting all results". This will return a plot summarizing all of your model results in one figure. So far, we have focused on comparing results within one model. Compare results across the models for different TF activation curves. How could you use this information in general to determine the regulatory logic controlling a promoter region?



We can see that magnitude, duration, and timing of TF activation influence the amount and duration of gene expression.
When both TF are active at the same time, there is often no difference in gene expression profiles among the OR, AND, and two-step model.
When one TF is active for a longer duration than the other TF, gene expression doesn't change in the OR model, but gene expression becomes greatly reduced in terms of magnitude and duration or entirely absent in the AND and two-step model.
When TF activity don't overlap, there is no difference in gene expression in the OR model, but in the AND, and two step model, gene expression is absent.
If gene expression occurs regardless of simultaneous activity of both TFs, it suggests "OR" logic, where the activation of either TF is sufficient for gene expression. However, if gene expression is present only when both TFs are active concurrently, it implies "AND" logic, meaning the presence

of both TFs for activation is necessary. Lastly, the absence of gene expression when TF activities don’t overlap indicates a sequential or two-step model, where the order and timing of TF activation are crucial.