Modeling mRNA Production at a Gene: The Biological Advantage of Feed Forward Loops

Olivia Fernflores

September 19, 2024

1 Introduction

Transcription regulation is of key importance biological systems because at the most basic level, it controls gene expression. By having a mechanism to control gene expression, organisms can differentially express genes under specific conditions such as receipt of an external signal, being in a particular stage of development, or being exposed to a pathogen. Because of its applications across most things that an organism would ever need to do, being able to differentially express genes is essential for survival and is a highly conserved feature. One common way of modeling transcription regulation is through networks, where each node is a gene or protein and each edge is the relationship between two nodes. Once you have a few unique nodes, there are many possibilities for how the nodes will be connected, but the possibilities are limited. Because there are a finite number of ways the nodes can be connected, we can think of each possibility as a motif. This allows us to recognize when the same pattern of node connections arises more than once despite the context being different.

In biological systems, we would expect the simplest network motif to be conserved unless there is a biological advantage to a more complicated network. Milo et al. explored this hypothesis and found that some transcription regulation networks, such as feed forward loops, are observed about 10 standard deviations more in biological systems than we would expect if this network motif occurred due to chance alone [1]. In this project, we seek to understand why these feed forward loops are biologically advantageous (as indicated by their high prevalence) by modeling two types of feed forward loops and comparing to simpler networks of direct and indirect transcription regulation.

2 Model Development

2.1 Direct Regulation of Transcription

To model transcription regulation, we started with a simple system of direct regulation. Under a model of direct regulation, a signal turns on a single transcription factor, *TF1*, and that transcription figure binds to a gene, *Gene i* with a binding constant *Kd*. When *TF1* is bound, *Gene i* is transcribed at a rate *k1*, producing *RNAi*. *RNAi* is also degraded, which happens at a rate *k2*. This is shown in Figure 1A. Together, these parameters can describe the rate at which the $[RNAi]$ is changing over time with the following equation:

$$\frac{d[RNAi]}{dt} = k_1 * f_b - k_2 * [RNAi]$$

This equation quantifies the change in $[RNAi]$ over time using the *k1* and *k2* parameters previously mentioned and includes an additional factor, *fb*, that describes what fraction of transcription factor is bound to our gene of interest. This is calculated as 1 minus the fraction unbound, which is a classic result in biochemistry called the *Hill Equation* and is written like this [2]:

$$f_b = \left(1 - \frac{1}{1 + \left(\frac{TF_1}{Kd_1}\right)}\right)$$

If we combine the two equations above, we can build our full mathematical model for direct regulation.

$$\frac{d[RNAi]}{dt} = k_1 * \left(1 - \frac{1}{1 + \left(\frac{TF_1}{Kd_1}\right)}\right) - k_2 * [RNAi]$$

According to this model, when there is no *TF1*, there will be no transcription of *RNAi*. Biologically, this makes sense. If *TF1* is the only component that facilitates transcription of *Gene i* and there is no *TF1*, we would expect to see no *Gene i* transcription. This model is biologically interesting but very simple, so we also explored three more complicated models that are also found in biological systems and build on the model of direct regulation.

2.2 Indirect Regulation of Transcription

The easiest way to add complexity to the model of direct regulation is to add an intermediate step between *TF1* and activation of *Gene i* transcription, otherwise known as indirect regulation. In this model, *TF1* is turned on by a signal and then binds to *Gene2* with affinity *Kd11* to activate transcription. *Gene2* is transcribed at a rate *k3* and produces *RNA2*, which is degraded at the rate *k4*. *RNA2* also produces *TF2* at rate *k5*, and *TF2* is degraded at the rate *k6*. *TF2* is also responsible for activating transcription of *Gene i*, which takes place when *TF2* binds to *Gene i* at an affinity of *Kd2*. *Gene i* is transcribed at a rate *k1* to produce *RNAi*, which is degraded at a rate *k2*. This model is shown in Figure 1B. Now that we have an additional RNA (*RNA2*) and an additional transcription factor (*TF2*), we can measure the concentrations of those two compounds and *RNAi* over time, which means we need three coupled differential equations to represent our system.

The first equation for this system is very similar to the equation used for direct regulation, but is modified slightly to show the binding of *TF2* to *Gene i* instead of the binding of *TF1* as seen with direct regulation. As with direct regulation, this equation measures the change in concentration of *RNAi* with respect to time and is written as follows:

i.

$$\frac{d[RNAi]}{dt} = k_1 * \left(1 - \frac{1}{1 + \left(\frac{TF_2}{Kd_2}\right)}\right) - k_2 * [RNAi]$$

Because we are also producing and degrading both *RNA2* and *TF2*, we also need equations to model those production and degradation rates. Similar to the production and degradation of *RNAi*, we can describe the change in concentration of *RNA2* over time as

$$\frac{d[RNA2]}{dt} = k_3 * f_b - k_4 * [RNA2]$$

where

$$f_b = \left(1 - \frac{1}{1 + \left(\frac{TF_1}{Kd_{11}}\right)}\right)$$

giving the overall equation

ii.

$$\frac{d[RNA2]}{dt} = k_3 * \left(1 - \frac{1}{1 + \left(\frac{TF_1}{Kd_{11}}\right)}\right) - k_4 * [RNA2]$$

The final equation we need for the model of indirect regulation is an equation to describe the production and degradation of *TF2*. Because there is no binding event involved with production or degradation, this is completely dependent on the rates of production and degradation, *k5* and *k6* respectively.

iii.

$$\frac{d[TF_2]}{dt} = k_5 * [RNA2] - k_6 * [TF_2]$$

Taken together, these three equations describe the model of indirect regulation detailed above and visualized in Figure 1B. Similar to the way we validated the model for direct regulation, we can check the equations by considering the production of *RNAi*, *RNA2*, and *TF2* when *TF1* is zero. Under this condition, we would expect no production of *RNAi*, *RNA2*, and *TF2*, so their derivatives should also be zero. Starting with equation ii., we see that $\frac{d[RNA2]}{dt}$ will be zero when *TF1* is zero. If this value is zero, this means that we are not producing *TF2*, so equation iii. will also be equal to zero (both terms will be zero). When there is no *TF2*, equation i. shows that there will also be no *RNAi* production, confirming that we have set up our equations correctly.

So far, our models of indirect and direct regulation assume that *Gene i* is controlled by only one transcription factor, which isn't always the case. A more complex model where *Gene i* is regulated by two transcription factors that also regulate each other is called a feed forward loop and is very commonly found in biological systems. In this paper, we explored two different methods of regulation through a feed forward loop and compared them to our models of indirect and direct regulation.

2.3 "And Gate" Model for Transcription Regulation in a Feed Forward Loop

The first model of a transcription regulation in a feed forward loop we considered is called an "and gate". The feed forward loop is basically a combination of the indirect and direct models of regulation. In the feed forward loop, *TF1* is responsible for regulating both *RNAi* and *RNA2*. Almost all of the rest of the model stays the same, with consistent parameter naming used to describe the rates of production and degradation, as well as binding affinities. In the "and gate" model, both *TF1* and *TF2* need to be bound to *Gene i* for transcription to occur. This model is shown in Figure 1C and can be described in very similar equations to the indirect regulation model.

Equations for $\frac{d[RNA2]}{dt}$ and $\frac{d[TF_2]}{dt}$ will remain exactly the same because the production and degradation of *RNA2* and *TF2* are unchanged. The only thing that does change is the equation for $\frac{d[RNAi]}{dt}$, because *RNAi* can only be produced when both transcription factors are bound. To derive this equation, we must include a factor to describe the fraction of *TF1* bound and a factor to describe the fraction of *TF2* bound, which gives us

$$\frac{d[RNAi]}{dt} = k_1 * f_{b1} * f_{b2} - k_2 * [RNAi]$$

The equations for fraction bound for each transcription factor are the same as the equations for fraction bound in the previous two models but adjusted accordingly to account for the correct transcription factor and Kd parameters.

$$f_{b1} = \left(1 - \frac{1}{1 + \left(\frac{TF_1}{Kd_1}\right)}\right)$$

,

$$f_{b2} = \left(1 - \frac{1}{1 + \left(\frac{TF_2}{Kd_2}\right)}\right)$$

If we add these substitute these factors into our equation for $\frac{d[RNAi]}{dt}$ and include our equations for $\frac{d[RNA2]}{dt}$ and $\frac{d[TF_2]}{dt}$, we can see the overall set of equations that describe the model of the "and gate".

i.

$$\frac{d[RNAi]}{dt} = k_1 * \left(\left(1 - \frac{1}{1 + \left(\frac{TF_1}{Kd_1}\right)}\right)\right) * \left(\left(1 - \frac{1}{1 + \left(\frac{TF_2}{Kd_2}\right)}\right)\right) - k_2 * [RNAi]$$

ii.

$$\frac{d[RNA2]}{dt} = k_3 * \left(1 - \frac{1}{1 + \left(\frac{TF_1}{Kd_{11}}\right)}\right) - k_4 * [RNA2]$$

iii.

$$\frac{d[TF_2]}{dt} = k_5 * [RNA2] - k_6 * [TF_2]$$

For validation, we can again test the system when *TF1* is equal to zero. Again, we will see that all three equations become zero under this condition, which matches the behavior we would expect biologically.

2.4 "Or Gate" Model for Transcription Regulation in a Feed Forward Loop

Our final model of transcription regulation is called an "or gate". This model has exactly the same components and relationships as the "and gate" model except that as long as either *TF1* or *TF2* is bound, *Gene i* will be transcribed. Just like the "and gate" model, the production of *RNA2* and *TF2* are unaffected by the change in how transcription of *Gene i* is activated, so the equations for $\frac{d[RNA2]}{dt}$ and $\frac{d[TF_2]}{dt}$ will remain the same. The equation for $\frac{d[RNAi]}{dt}$ will change because it needs to incorporate the possibility of just *TF1* being bound, just *TF2* being bound, or *TF1* and *TF2* being bound since all of these conditions will lead to transcription of *Gene i* in this loop. To do this, we consider the probability of both transcription factors being unbound, which is given by $f_{ub1} * f_{ub2}$. This means the probability of at least one transcription factor being bound is given by $1 - (f_{ub1} * f_{ub2})$. If we use this in the equation for $\frac{d[RNAi]}{dt}$, we get the following:

$$\frac{d[RNAi]}{dt} = k_1 * (1 - (f_{ub1} * f_{ub2}) - k_2 * [RNAi]$$

If we write out the equations for both unbound factors, we get the following:

$$f_{ub1} = \frac{1}{1 + \left(\frac{TF_1}{Kd_1}\right)}$$

,

$$f_{ub2} = \frac{1}{1 + \left(\frac{TF_2}{Kd_2}\right)}$$

Substituting these values into the first equation given in this section and adding the unmodified second and third equations from the indirect and "and gate" models gives us the following set of equations for the "or gate" feed forward loop.

i.

$$\frac{d[RNAi]}{dt} = k_1 * (1 - (\frac{1}{1 + \left(\frac{TF_1}{Kd_1}\right)}) * (\frac{1}{1 + \left(\frac{TF_2}{Kd_2}\right)})) - k_2 * [RNAi]$$

ii.

$$\frac{d[RNA2]}{dt} = k_3 * \left(1 - \frac{1}{1 + \left(\frac{TF_1}{Kd_{11}}\right)}\right) - k_4 * [RNA2]$$

iii.

$$\frac{d[TF_2]}{dt} = k_5 * [RNA2] - k_6 * [TF_2]$$

Just like with the other three models, we can test the equations at *TF1*. If *TF1* is zero, equation ii. will also be zero. This means that equation iii. and i. will be zero, which again exactly what we would expect biologically.
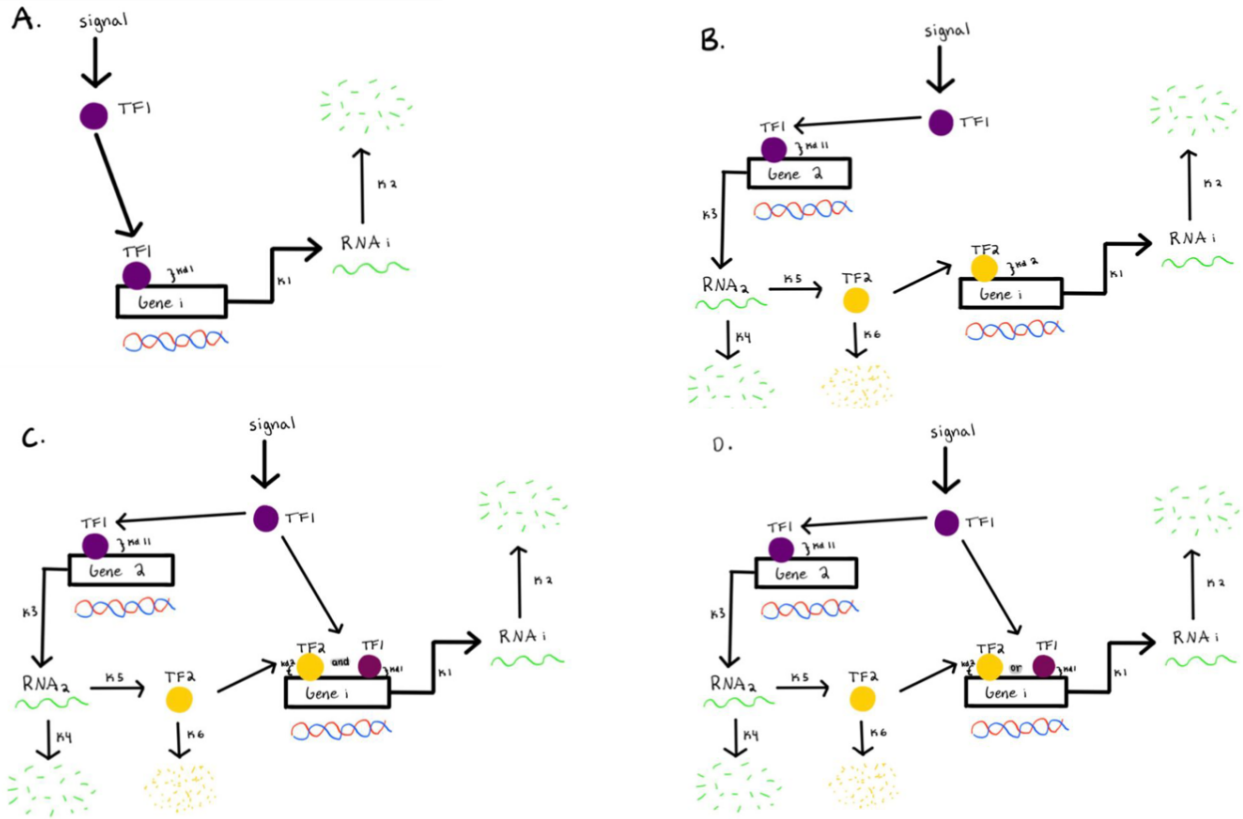
Figure 1: The four models of transcription regulation considered in this paper. **A.** Direct model of transcription regulation. **B.** Indirect model of transcription regulation. **C.** "And Gate" model of transcription regulation in a feed forward loop. **D.** "Or Gate" model of transcription regulation in a feed forward loop.

3   Results

Because of their prevalence in biological systems, we hypothesized that the "and gate" and "or gate" models for regulating transcription would have unique properties that are not possible to achieve with a simple direct or indirect regulation. To test this hypothesis, we ran continuously solved for $\frac{d[RNAi]}{dt}$ at very small time intervals under a wide range of parameter value for *k1, k2, k3, k4, k5, k6,* and *Kd1, Kd2, Kd11*. We find that the "and gate" model provides unique capabilities to have a delay when initiating production of *RNAi* and a quick stop in the production of *RNAi*. We also find that the "or gate" model behaves in the opposite manner, showing a quick initiation of production of *RNAi* and a delay in stopping production of *RNAi*. For comparison purposes, all results described here were done with all *k* and *Kd* values equal to one, although our findings appear to be robust to variation in these parameter values.

To examine the behavior when initiating *RNAi* production, we allowed all four systems to reach their steady state (achieved when the derivative is equal to zero and indicates that the concentration has reached its maximum value) and then measured the time it takes to reach half of the steady state amount and the full steady state amount (Table 1)

Table 1: Time to Half and Full Steady States

Regulation Type	Time to Half Steady State	Time to Steady State
Direct Regulation	0.9091	22.7273
Indirect Regulation	2.4242	27.8788
And Gate	2.4242	27.2727
Or Gate	1.2121	27.2727

As shown by the values in the table and visualized in Figure 2, the "and gate" model behaves most similarly to the indirect model in terms of time to half and full steady state, which both show an initial delay to reach steady state that isn't captured in the "or gate" and direct models. Contrastingly, the "or gate" model initiates production very quickly, almost mimicking the direct regulation model, which can also be seen in Figure 2.
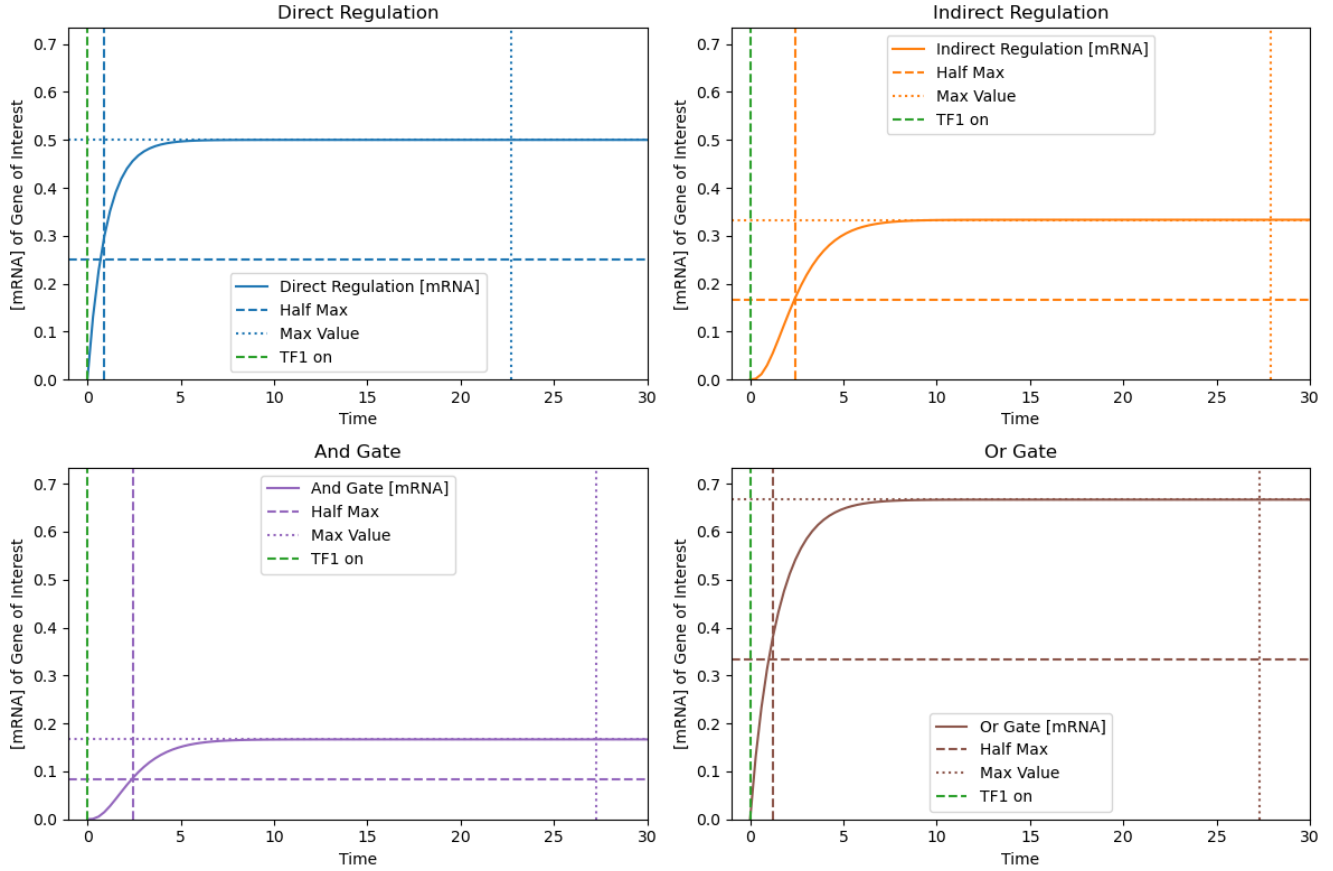

Figure 2: The four models of transcription regulation plotted under conditions of all k and Kd parameters equal to 1. The solid curves show the model over time. *TF1* equals one at time zero, which is indicated by the vertical green line on each subplot. When each model reaches it's half steady state concentration of $RNAi$ and maximum steady state concentration of $RNAi$, the time and concentration are marked with a dashed vertical line in the same color as the solid line for the corresponding model.

When we just examine the behavior of initiating $RNAi$ production, it seems like the "and gate" and "or gate" behave very similarly to the indirect and direct models, respectively. This prompted us to examine the behavior of all four models when *TF1* is "turned off", which we mathematically represent from changing the binary value of *TF1* from 1 ("on state") to 0 ("off state") (Table 2).

Table 2: Time to Return to [RNAi] = 0

Regulation Type	Time to Return
Direct Regulation	23.33
Indirect Regulation	28.18
And Gate	26.36
Or Gate	28.79

When *TF1* is set to zero after all four models have had a chance to reach their steady state concentrations of $RNAi$, we see that the "and gate" turns off earlier than the indirect circuit. This suggest the unique property of the "and gate" is to initiate production with a delay and rapidly stop production. The "or gate" behaves oppositely again, showing that it can delay stopping production while rapidly initiating production. The behavior after "turning off" *TF1* is also shown in Figure 3.
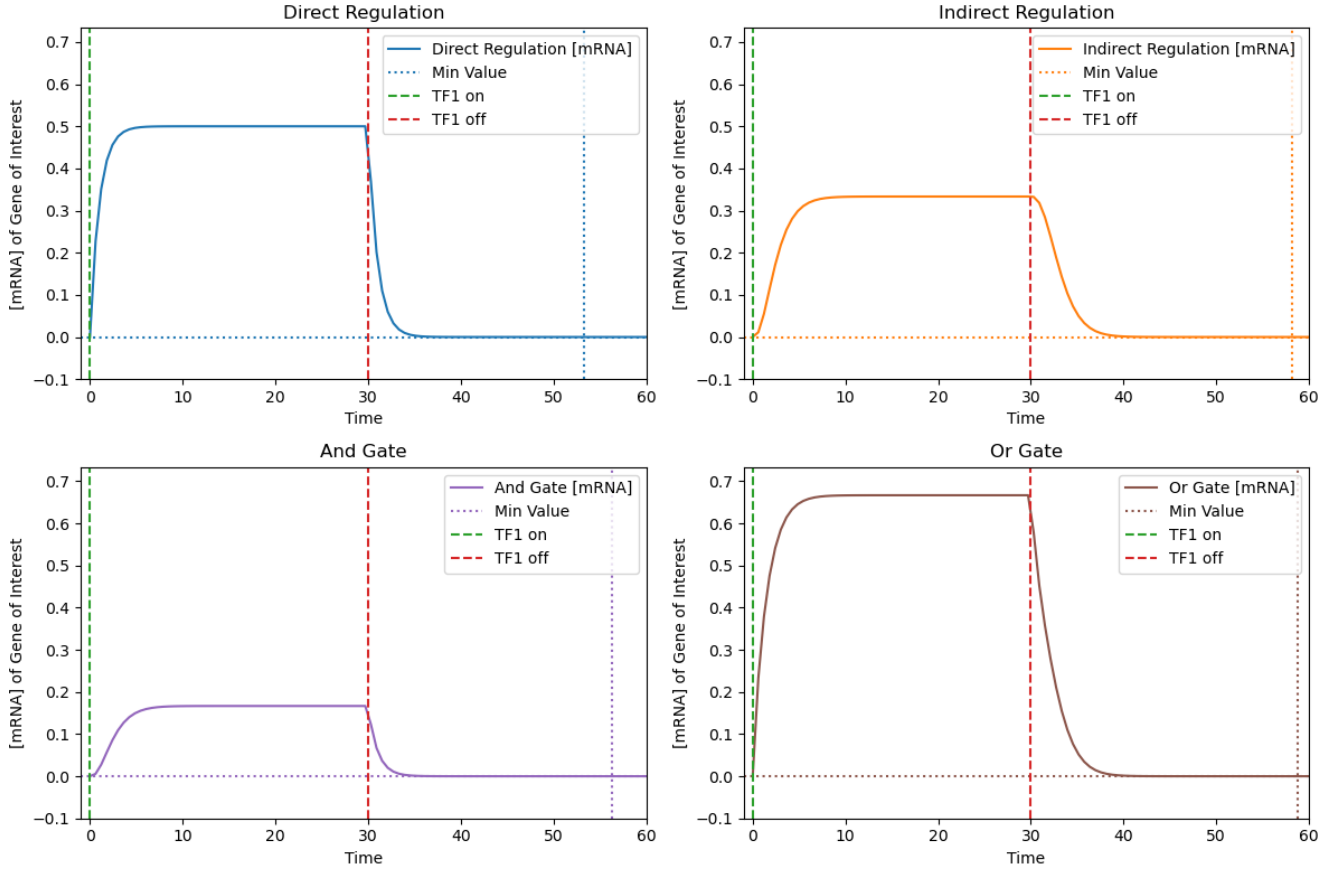
Figure 3: The four models of transcription regulation plotted under conditions of all k and Kd parameters equal to 1. The solid curves show the model over time. *TF1* equals one at time zero, which is indicated by the vertical green line on each subplot. At time 30, after all four models have reached steady state concentrations, *TF1* is equal to zero as indicated by the vertical red line on each subplot. When each model reaches it's minimum *RNAi* value of zero, the time is marked with a dashed vertical line in the same color as the solid line for the corresponding model.

4 Discussion

In this project, we have determined that both the "and gate" and "or gate" feed forward loop models of regulating transcription have unique properties that are not captured in a simple direct or indirect regulation system. Although you can adjust parameters such as *k2*, *k4*, and *k6* to increase the rate of degradation or production in the direct or indirect models, you cannot adjust the parameters such that only the production rate increases without the degradation rate increasing or vice versa. For example, doubling *k2* will increase the rate of production and increase the rate of degradation. In a feed forward loop like the "and gate" or "or gate", you can have a quick switch to production or degradation and the and the slow response to the opposite, which will never be possible under the direct or indirect model.

The "and gate" model shows a delay in initiating RNA production and a rapid decline in RNA production when signal is lost. Biologically, this could be relevant in stress response. The delay we see with the "and gate" happens because it takes time to build up enough of both transcription factors so they can simultaneously bind to the gene and initiate transcription. Because of the time delay, any noise in signaling will not be enough to initiate full levels transcription but an immediate stop in signaling will abruptly stop transcription. In a stress response, you need a way to filter out noise. If the cell responds to every small amount of noise and immediately activates all stress response pathways, the cell will frequently engage in stress response behavior despite not being under conditions of stress. With the "and gate" feed forward loop, this dysfunction isn't possible due to the time delay. So far, this property could be achieved by the indirect model but not the direct model. Where the "and gate" model stands out is that it can also rapidly stop RNA production, which the indirect model cannot do under any parameter combinations. If the RNA this process regulates is part of a stress response pathway, we need production to stop as soon after the stress signal stops as possible so that the cell isn't behaving in the stress condition for longer than necessary. This is possible with the "and gate" and also the direct regulation model, but not the indirect model. Thus, the "and gate" model provides unique functionality that could be useful in conditions like stress response.

7

The "or gate" model showed a rapid initiation of RNA production and delay in declining RNA production. Biologically, this could be relevant in an immune response. During an immune response, you want a cell to rapidly activate genes that will help with immune function and once the signal that starts the immune cascade is turned off, a delay will help make sure the cell can still respond to any remaining pathogenic activity. The "or gate" model works perfectly for this, as it allows a very quick initiation of RNA production, which is also seen in the direct regulation model but not the indirect regulation model. What sets the "or gate" apart is that when signal is lost, the "or gate" shows a delay that's only seen at the same time point as the indirect model. This means that the "or gate" is the only model that would allow a cell to respond rapidly to a signal to activate an immune response and also delay turning off the immune pathway when signal is lost, increasing the cell's immune function at initial exposure to a pathogen and as the pathogen clears.

This experiment showed the biological usefulness of the "and gate" and "or gate" feed forward loop models of regulating transcription have properties that cannot be captured in simple direct or indirect models. The properties of the "and gate" may be useful in stress response conditions, while the properties of the "or gate" may be useful in immune response conditions. Further experimentation in biological systems is needed to confirm these hypothesis that have been developed by our models, though the known prevalence of both feed forward loops in biological systems [1] suggests that these models do serve a purpose in cell survival.

References

[1] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U. (2002). *Network motifs: Simple building blocks of complex networks.* Science, 298(5594), 824–827. https://doi.org/10.1126/science.298.5594.824.

[2] Goutelle, S., Maurin, M., Rougier, F., Barbaut, X., Bourguignon, L., Ducher, M., Maire, P. (2008). *The hill equation: A review of its capabilities in pharmacological modelling.* Fundamental Clinical Pharmacology, 22(6), 633–648. https://doi.org/10.1111/j.1472-8206.2008.00633.x.