Generation of a Dynamic Model for the Effect of Riboswitches in Gene Expression

BIOS570

Instructor: Dr. Aryeh Warmflash

Yifei Yang

Introduction

Riboswitches are non-translated segments of mRNA molecules commonly found in prokaryotes and some low eukaryotes. They control gene expression by forming aptamers that regulate either transcription or translation. Some of these aptamers can be stabilized upon binding to a small molecule ligand. By using these ligand-binding aptamers, either on their own or with other aptamers that do not binding to ligands, riboswitches can shift the equilibrium to favor or disfavor the expression of genes.

There exist many means through which the genetic expression is controlled by the riboswitch[1]. Some commonly found ones include inhibition of ribosome binding by protection of the ribosome binding site (RBS), termination of transcription through destabilizing the transcription complex (e.g. revealing binding site of the rho factor), and alternations of the degradation rate of the mRNA by changing the thermodynamic stability of the RNA molecule (e.g. revealing sites for an RNAse) (Figure 1. B-D)[2].

Due to the ease of manipulation of RNA and the importance of their ligands, riboswitches have been used as tools for synthetic biology and therefore modeling is important for systematically studying their behaviors. Many models construct the mass-action kinetics equations and then solve them using steady-state approximation, resulting in equations that describe only the long-term equilibrium[2,3]. However, to create more complicated genetic circuits where the purpose of a riboswitch is not merely to produce a reporter protein, it may be desirable to construct dynamic models that describe the output in time-dependent manner. While solving mass-action kinetics equation numerically appears to be straight-forward, there are details that could prevent the numerical solution to match the steady-state solution at equilibrium. I attempted at constructing a model based on numerically solving the mass-action equation and aligning its result to the solution provided by the steady-state solutions.

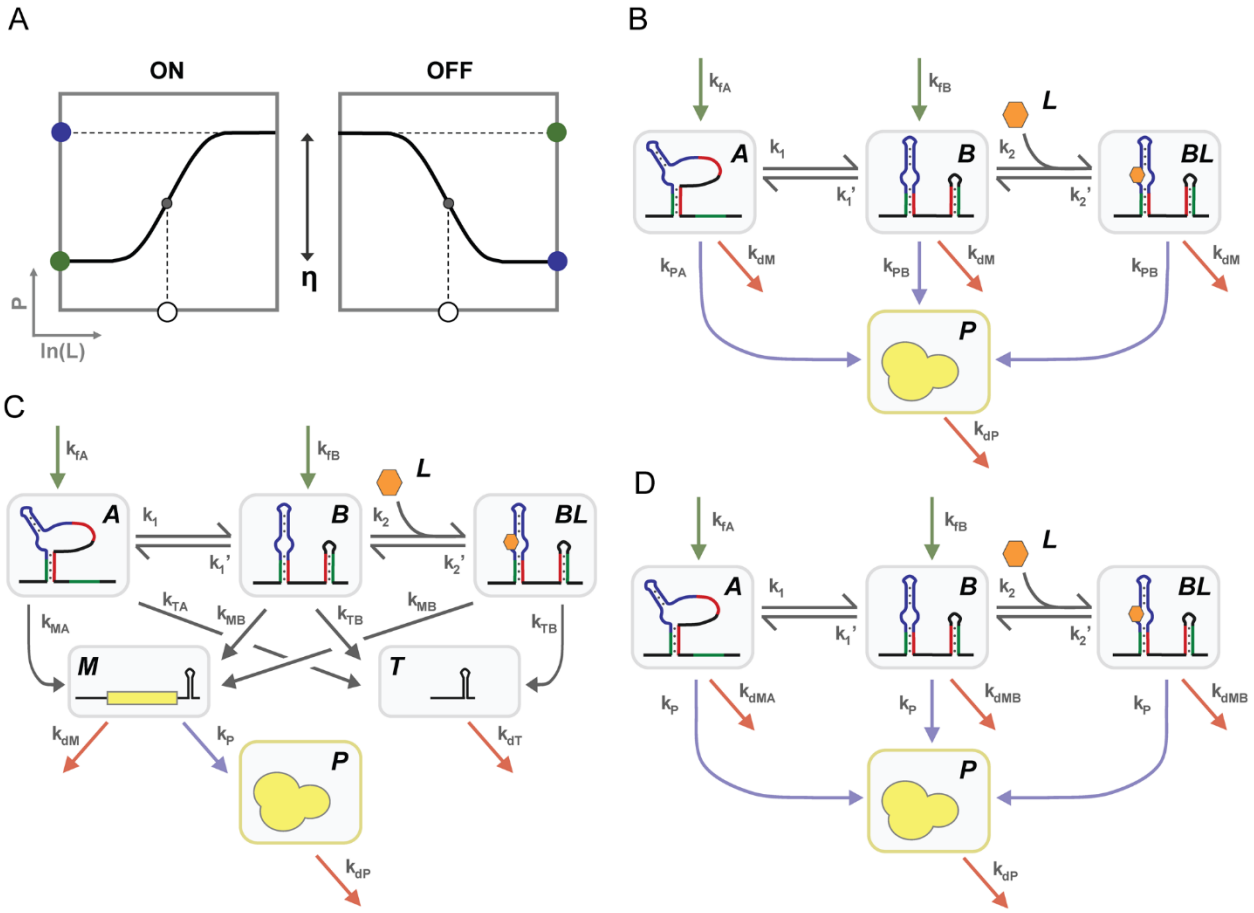

Figure 1| Some common models for riboswitches[2]. A: for a model where ligand upregulates gene expression, the relationship between the protein concentration and the natural log of ligand concentration should be sigmoidal. η is the dynamic range. B: The model where the ligand stabilizes the conformation that either hides or releases the RBS, leading to down- or up-regulation, respectively. C: The model where ligand binding leads to stabilization of a loop that promotes or inhibits transcriptional termination. D: The model where ligand binding leads to change of thermodynamic stability of the RNA molecule.

Methods

Protein production control by riboswitches were modeled with differential equations[2]. Starting from transcribed RNA, each conformation of the RNA molecule, the protein, and the ligand are each assigned a value of concentration and a differential of change.

One may assume that the transcriptional rates are constant since riboswitches should not affect the rate of RNA polymerase (Figure 1. B-D)[2]. The newly transcribed RNA molecules are assumed to

adopt different conformations with each having different rate of protein production or RNA degradation. It was assumed that there is a state that would bind to a ligand and would be stabilized as a result while the other states are grouped together as one term. The two types of states are assumed to interconvert with defined rate constants. Each of the two states would have associated rates of protein production or conversion into either mRNA or terminated transcript depending on the model. All RNA molecules that can lead to production of the whole protein would each have an associated protein production rate. All molecules, except the ligand (which is assumed to be a stable, small molecule), also have an associated degradation rate.

One relatively simple model would be that the unbound riboswitch exists commonly in a conformation of which the RBS is hidden, leading to low gene expression (Figure 1. B). Ligand binding stabilizes a conformation of which the RBS is released and upregulating the gene expression. Following the model previously described, the three states of RNA (RBS hidden, RBS released, RBS released bound to ligand) can be annotated as “A”, “B”, and “BL”, respectively. By annotating ligand as “L” and protein as “P”, a set of differential equation can be written for the system:

$$\frac{d[A]}{dt} = k_{fA} - k_1[A] + k_1'[B] - k_{dMA}[A]$$

$$\frac{d[B]}{dt} = k_{fB} + k_1[B] - k_1'[B] - k_2[L][B] + k_2'[BL] - k_{dMB}[B]$$

$$\frac{d[BL]}{dt} = k_2[L][B] - k_2'[BL] - k_{dMB}[BL]$$

$$\frac{d[P]}{dt} = k_{pA}[A] + k_{pB}([B] + [BL]) - k_{dP}[P]$$

The ligand concentration is assumed to be constant throughout the simulations.

The differential equations were written as MATLAB functions and were numerically solved using the ode23 function with a time step of 1 s. The values initially used for the rate constants are shown in table 1.

Table 1| Rate constants associated with the first model. The values are taken from Beisel and Smolke 2009 supplementary text 1[2]. Values of some constants were varied to examine the behavior of the model.

Rate Constant	Description	unit	Initial Value
k_{fA}	Rate of production of RNA in conformation A	M/s	1*10^-11
k_{fB}	Rate of production of RNA in conformation B	M/s	0
k_1	Rate of conversion from conformation B to A	1/s	0.1
k_1'	Rate of conversion from conformation A to B	1/s	10
k_2	Rate of ligand binding to RNA in conformation B	1/M·s	1*10^6
k_2'	Rate of ligand dissociation from bound RNA	1/s	1*10^-6
k_{dMA}	Rate of RNA degradation of RNA in conformation A	1/s	1*10^-3
k_{dMB}	Rate of RNA degradation of RNA in conformation B	1/s	1*10^-3
k_{dP}	Rate of protein degradation	1/s	1*10^-2
k_{pA}	Rate of protein production from RNA in conformation A	1/s	1*10^-3
k_{pB}	Rate of protein production from RNA in conformation B	1/s	1*10^-2

For this case with $k_{fB} = 0$, the steady-state approximation has been solved to be[2]:

$$P = \frac{k_{fA}}{k_{dP}} \cdot \frac{k_{pA}K_1 + k_{pB}(1 + K_2[L])}{k_{dMA}K_1 + k_{dMB}(1 + K_2[L])}$$

Where

$$K_1 = \frac{k_1'}{k_1}$$

$$K_2 = \frac{k_2}{k_2'}$$

Results

Calibration of the Numerical Model

Initial concentrations of all molecules were set to 0 to find an equilibrium concentration. When the simulation was run, it was found that the concentration of the RNA molecules and protein varied wildly (Figure 2). Strangely, both forms of the RNA molecules and the protein have gone into negative concentrations (Figure 3). Furthermore, the mean concentration of RNA in conformation B (using the same time range as the protein) is -3.2583 nM. Nevertheless, the mean concentration of the protein at equilibrium, 1.920 nM, is roughly equal to the solution of the steady-state approximation, and 1.891 nM. When ligand was added, the protein concentration was still reasonably close to the one predicted by the steady-state solution (Figure 3, 4).

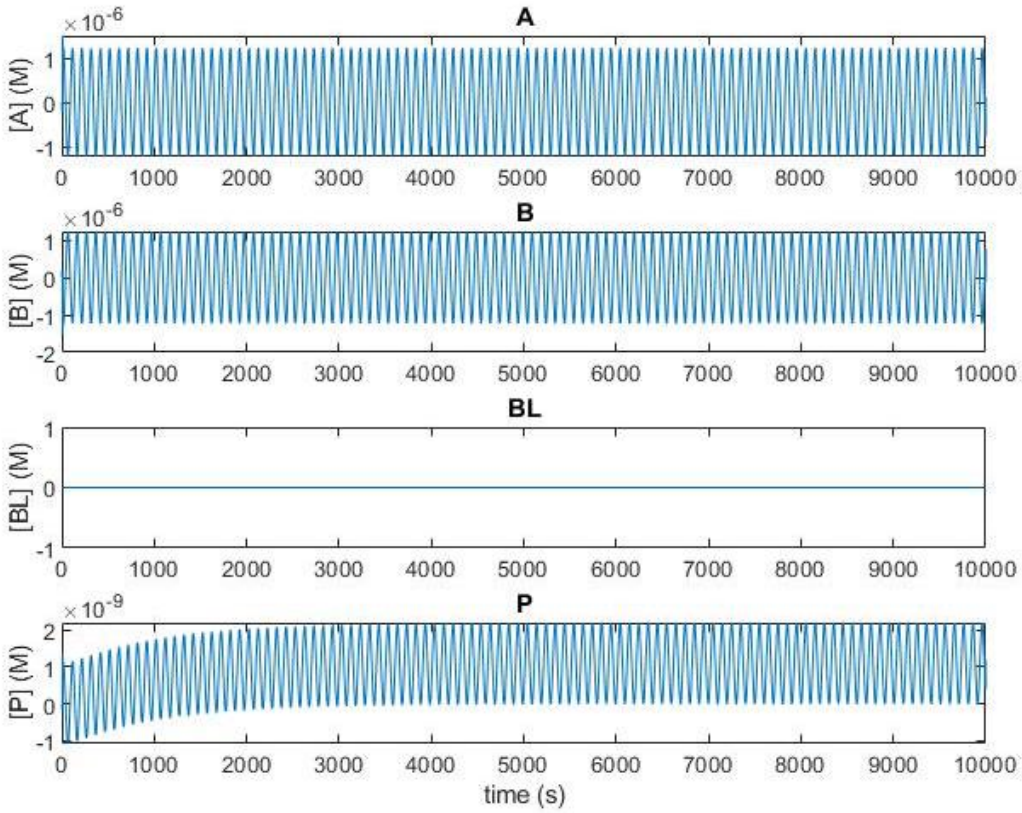

Figure 2| Initial Equilibrium Using Unaltered Differential Equations Set. A: RNA in conformation A; B: RNA in conformation B; BL: RNA in conformation B bound to ligand; P: protein.

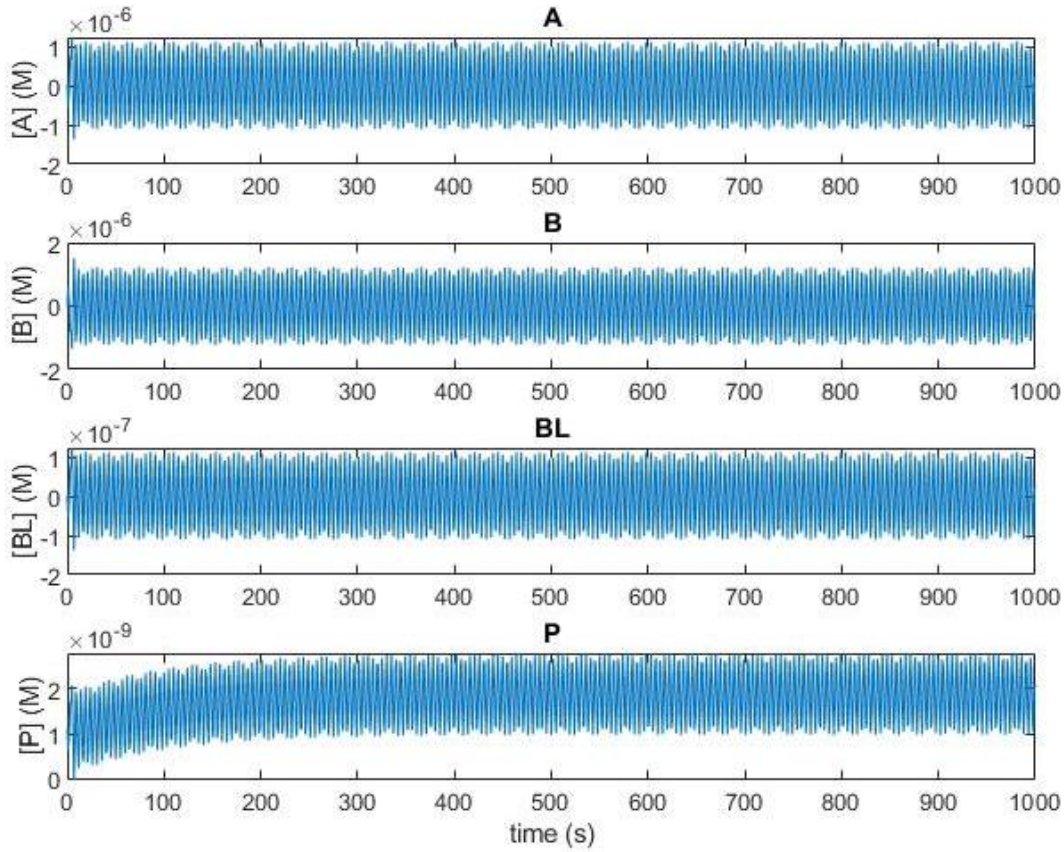

Figure 3| Simulation using [L] = 1 µM. A: RNA in conformation A; B: RNA in conformation B; BL: RNA in conformation B bound to ligand; P: protein. The initial value was the mean of the values during the equilibration between the time steps of 900 and 1000.

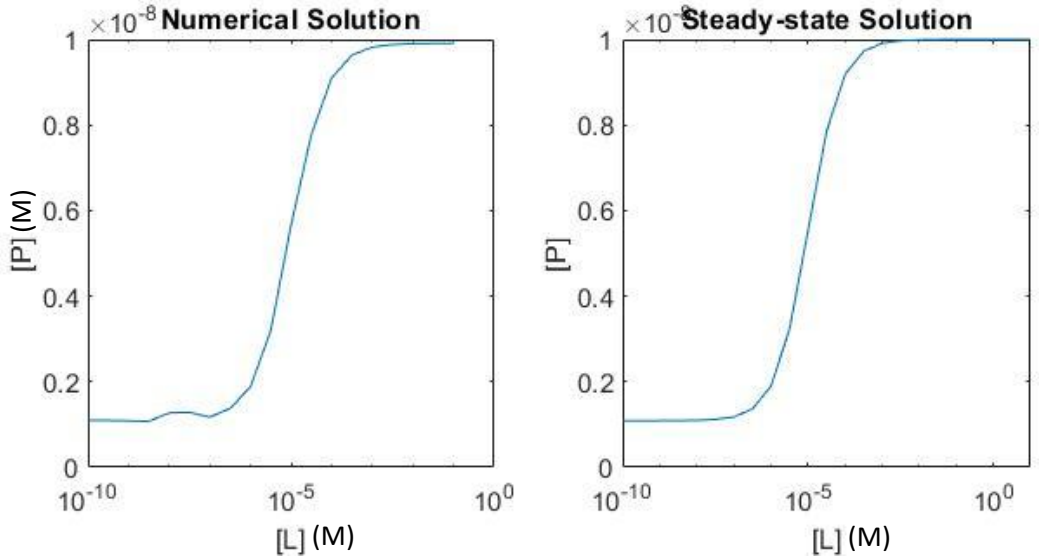

Figure 4| Comparison of the steady-state result by solving numerically and with steady-state approximation. The protein concentration for the numerical solution was a mean of the values between the time steps 900 and 1000.

Varying Ligand Concentrations

To alter the model, an obvious change is a changing ligand concentration. One could set the ligand concentration to fluctuate like a sine wave. To achieve this, one can derive the differential equation of the ligand based on the following description of a ligand concentration:

$$[L] = [L]_0 + a \cdot \sin\left(b \cdot (t + c)\right)$$

$$\frac{d[L]}{dt} = a \cdot b \cdot \cos\left(b \cdot (t + c)\right)$$

Where $[L]_o$ is the mean concentration of the ligand, a is the magnitude of the fluctuation, b is the frequency of the fluctuation, and c is the phase shift of the fluctuation.

A fluctuating ligand concentration has an obvious effect on the concentration of the protein (Figure 5).

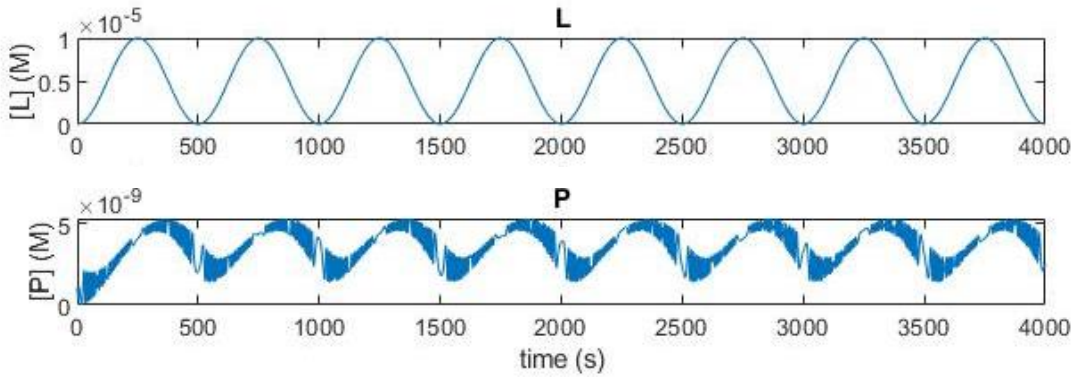

Figure 5| Variation in Protein Concentration Due to Fluctuating Ligand Concentration. L: ligand; P: protein. The ligand was set to start at 0 M, fluctuate in a concentration range between 0 and 10 µM, and with a period of 500 s.

Continuous Dilution Due to Cell Growth and Division

Another question would be the dilution of all factors due to volume increase during cell growth. The differential equation describing dilution as a function of doubling time is:

$$\frac{d[X]}{dt} = -\frac{\mathrm{t_{double}} \cdot [\mathrm{X}]}{\ln(2)}$$

Where [X] is the concentration of the molecule of interest and t_{double} is the time for the volume to double.

The term can be simply added to the mass action equations for each molecule to represent the dilution effect. One can also solve the equations with steady-state approximation. However, if the cell doubling is not uniform across time, the numerical solution should be better able to capture the other factors that are involved (e.g. fluctuating ligand concentration). When the model that assumed constant cell division every 20 minutes was tested, however, it appeared to break down. While the portion of the model reaching equilibrium appeared like that without the dilution factor (figure not shown, see code part 2.b.), the model began behaving strangely when ligand was added (Figure 6). Regardless of the concentration of the ligand, the protein concentration would experience a sharp drop and then slowly return to fluctuate around 0. The concentration of the RNA molecules not bound to the ligand spend a significant amount of time with negative values.

Discussion

In many levels, the dynamics model of the specific case of riboswitch action was able to simulate the protein concentration in non-steady-state systems in a physically possible range. At steady state, the numerical solution closely matched the solution using steady-state approximation. It appears possible for the model to simulate realistic events such as fluctuations of ligand concentrations. However, in some situations, the model completely breaks down, returning physically impossible solution.

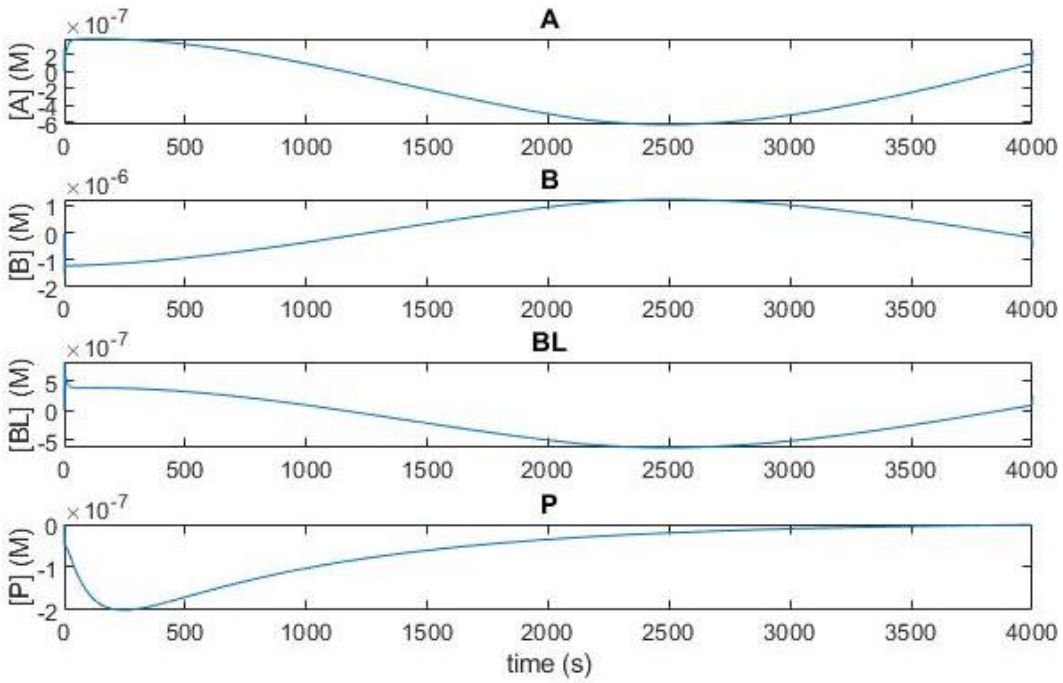

Figure 6| Numerical Solution after Ligand Addition for Constant Dilution Due to Cell Division with Constant [L] = 10 µM. A: RNA in conformation A; B: RNA in conformation B; BL: RNA in conformation B bound to ligand; P: protein.

Through some investigation, it was suspected the primary issue of the model is the term $k_2[L][B]$ in the differential equations for [B] and [LB]. When the value of [L] is high, effectively all RNA exist in the form of BL, causing the other two forms to fluctuate very close to zero. It was hypothesized that this fluctuation close to zero also increased the number of decimal places that MATLAB needs to handle and drastically increased the computing time (Figure 7). When [L] is too high relative to [B], the positive term in the equation for [LB] becomes so large that more LB is created than amount of B available, effectively creating RNA out of thin air. It is likely that a similar mechanism takes place in the continuous dilution model where the dilution term becomes so large in magnitude that it offsets the sum of all the positive terms and drive the concentration of A and B forms of RNA, and subsequently protein, into deeply negative values.

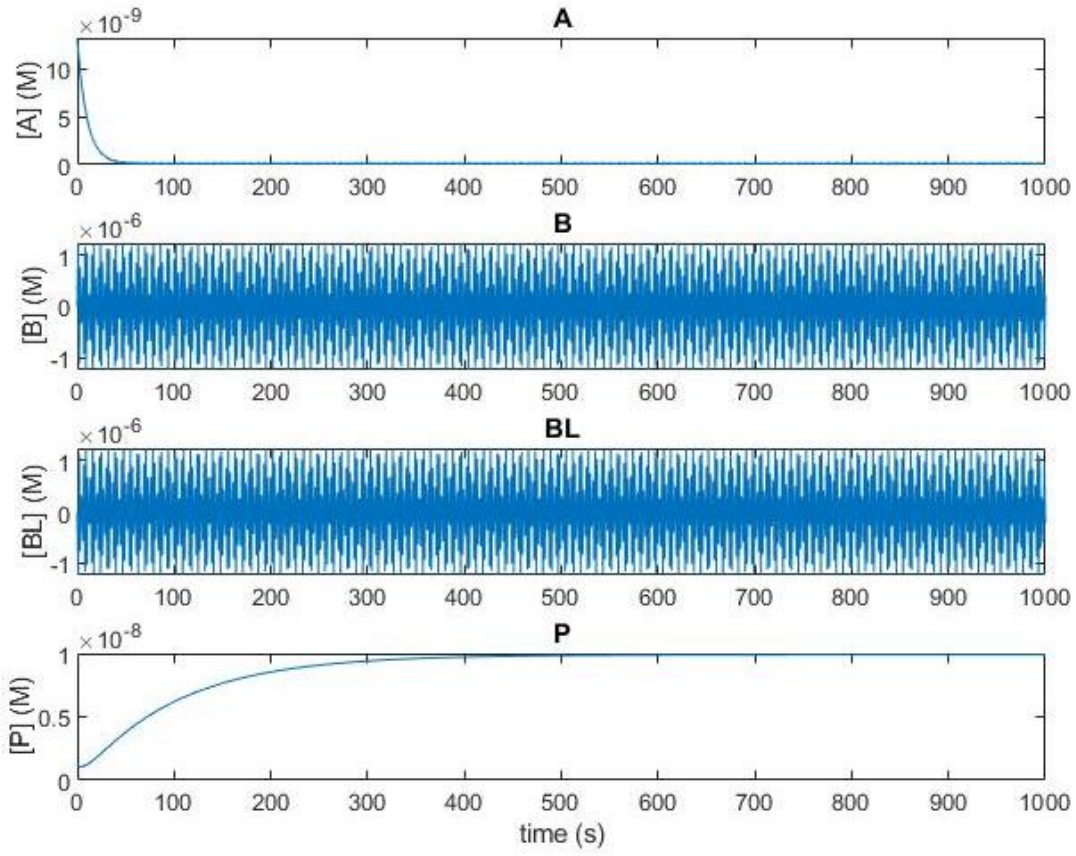

Figure 7| Simulation using [L] = 0.1 M. A: RNA in conformation A; B: RNA in conformation B; BL: RNA in conformation B bound to ligand; P: protein. The mean concentrations of the RNA molecules are very close to 0.

The attempts at correcting these issues have so far been unsuccessful. It was attempted to simply turn any negative values to zero by adjusting d[X]/dt to [X] when the magnitude of d[X]/dt is larger than [X]. However, the lack of reduction in the change of other terms (e.g. if less B is converted from A, then the decrease of A should also be decreased) caused the overall concentration of all molecules in the system to spike, resulting in ridiculous concentrations of protein larger than 100 µM in the initial equilibration (Figure 8). Other methods such as specifically limiting the $k_2[L][B]$ term or smoothing out the data have been attempted but were also ultimately unsuccessful. To model these systems correctly, better models than a simple set of mass-action kinetics equations need to be derived.

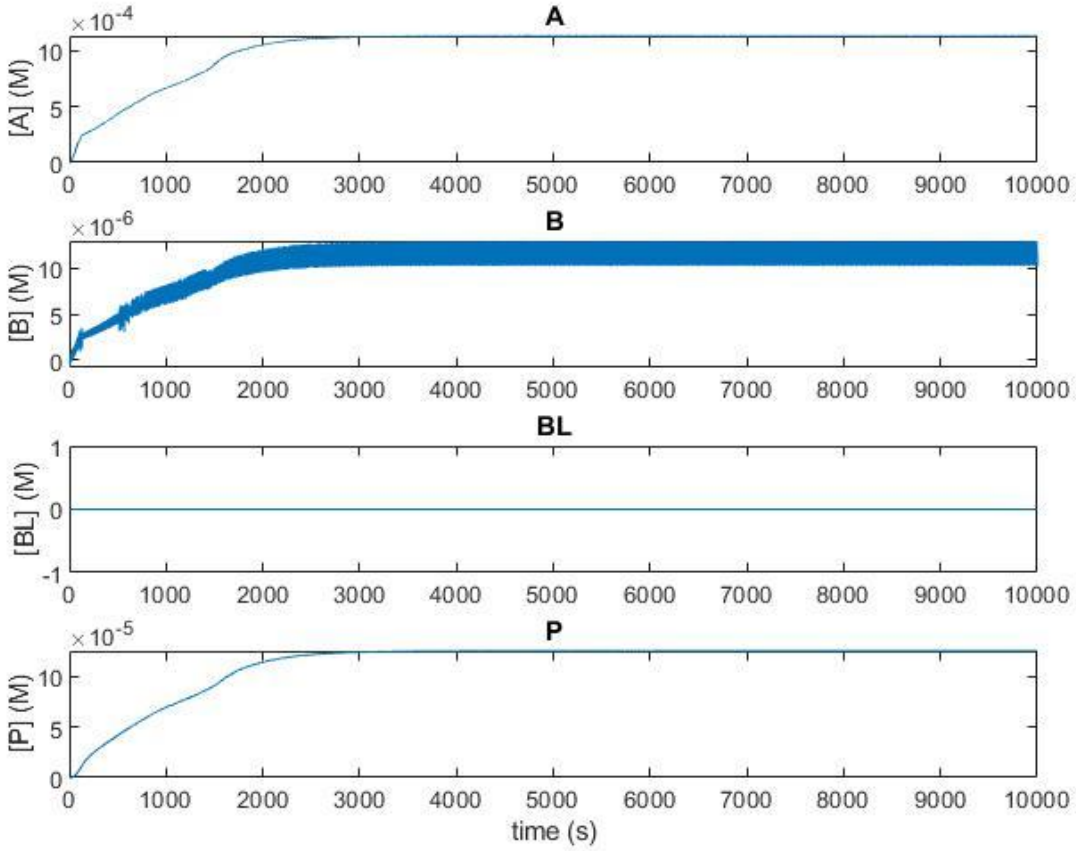

Figure 8| Initial Equilibrium by Disallowing Negative Concentrations. y-axis unit: M, x-axis unit: s. A: RNA in conformation A. B: RNA in conformation B. BL: RNA in conformation B bound to ligand. P: protein.

References

1 Bédard, A.-S. V., Hien, E. D. M. & Lafontaine, D. A. Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1863**, 194501, doi:https://doi.org/10.1016/j.bbagrm.2020.194501 (2020).

2 Beisel, C. L. & Smolke, C. D. Design Principles for Riboswitch Function. *PLOS Computational Biology* **5**, e1000363, doi:10.1371/journal.pcbi.1000363 (2009).

3 Mayo, M. L., Eberly, J. O., Crocker, F. H. & Indest, K. J. Modeling a synthetic aptamer-based riboswitch biosensor sensitive to low hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) concentrations. *PLOS ONE* **15**, e0241664, doi:10.1371/journal.pone.0241664 (2020).