

A Continuous Model of Gene Expression

California State Polytechnic University, Pomona

and

Loyola Marymount University

Department of Mathematics Technical Report

Joseph Hunt*, Lissette Laplace†, Elizabeth Miller‡, Jason Pham§

Applied Mathematical Sciences Summer Institute
Department of Mathematics & Statistics,
California State Polytechnic University, Pomona
3801 W. Temple Ave.
Pomona, CA 91768

Faculty Mentors: Erika Camacho¶, Stephen Wirkus||

August 2005

*Lamar University

†University of the Virgin Islands

‡Ferris State University

§California State Polytechnic University, Pomona

¶Loyola Marymount University

||California State Polytechnic University, Pomona

Abstract

Gene expression is the process by which a gene makes its effect on a cell or organism. Linear differential equations have been explored as a model for gene expression. We discuss the shortcomings of this model, and we propose a system of nonlinear differential equations to mathematically model gene expression in prokaryotes, specifically bacteria. We investigate this biological system using explicit functions that describe the processes of protein synthesis which includes transcription, translation, degradation, and feedback in hope of shedding light on their associated rates. We analyze the transient and steady state solutions of the model and give a biological interpretation of these results.

1 Introduction

Gene expression is the process by which a gene makes its effect on a cell, usually by directing the synthesis of protein with specific functions. This process is comprised of two basic steps: transcription and translation. Transcription is the process by which genetic information contained in deoxyribonucleic acid (DNA) is copied into a messenger ribonucleic acid (mRNA). Translation is the decoding of the genetic information in the mRNA into corresponding proteins [?]. The process of gene expression is controlled by direct or indirect interaction with other genes and their products. In both prokaryotes and eukaryotes, this operation is important in supporting the cell's survival.

Although extensive research has been conducted to find exact biological gene expression pathways and rates, most work has been primarily focused on bacterial and simple eukaryotic (yeast) transcription of mRNA. Secondly, a cell is constantly enduring intracellular changes in order to respond to its changing environment, causing the gene expression process to become extremely complicated as higher-ordered species are studied. This is the main reason why protein synthesis has not been able to be researched as a whole system. Bacterial cells have a simple but essential structure that consists of compartmentalized DNA surrounded by plasma and a protective cell wall. Due to this simple structure, our mathematical modeling efforts consider the prokaryotic bacterial cell where gene expression can take place anywhere within the cell and time delay is not a factor (as it may be in eukaryotes).

Our goals include using specific functions to formulate a mathematical model to describe gene expression of prokaryotic cells and analyze the behavior of the model by showing possible equilibrium solutions and bifurcations. We hope to gain insight into the interrelation of transcription and translation in prokaryotic cells. We also wish to provide some insight on how to possibly model protein synthesis in higher ordered eukaryotic organisms.

1.1 Biological System

Our mathematical model closely resembles prokaryotic bacterial protein synthesis that contains feedback from various protein products caused by the cell's environmental sensory. Proteins are synthesized in two main steps, transcription and translation. Transcription is the construction of mRNA from DNA genetic instructions, while translation is the process of decoding the mRNA to produce protein.

Considering the bacterial domain, transcription begins when the double stranded DNA is in contact with a protein called polymerase enzyme. The interaction between the DNA and the polymerase enzyme is initiated and terminated by proteins known as transcription factors. While in contact with the DNA, the polymerase moves along the strands of the DNA, unwinds them, and then uses one of the two as a template to make a single stranded mRNA. The mRNA is then transported to the ribosome to begin translation. The ribosome is a complex molecule that has two subunits comprised of lengthy ribosomal RNA (rRNA) and 30 to 50 proteins. Both subunits work together to decode the genetic information on mRNA into a certain sequence of amino acid. The transfer RNA (tRNA) carry the necessary amino acids to the ribosome, where these amino acids are combined into a polypeptide chain of the prescribed sequence. The ribosome releases the polypeptide chain, which then folds into itself and becomes a protein [?].

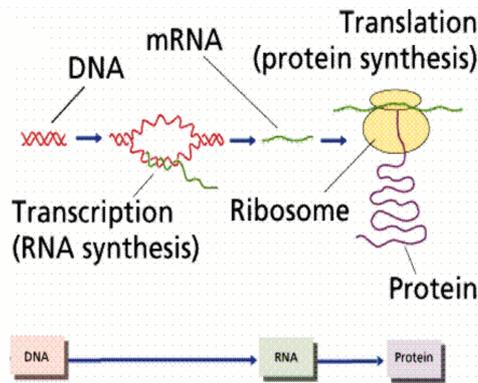


Figure 1: General Bacteria Protein Synthesis.[?]

1.2 Translating the Biological Process into Mathematical Model

In the most general case, transcription, translation, and degradation are controlled by the number of mRNA and proteins present in the cell. For instance, mRNA needs to be present in order for translation to occur, so the more mRNA there are, the more translation will occur. Likewise, the more protein there are, the more protein will decay. These relationships could be described mathematically as the rate of the change in concentrations of the respective protein or mRNA set equal to some mathematical functions that represent the cause of change for that rate.

$$\frac{d\vec{r}}{dt} = g(\vec{r}, \vec{p}, t) \quad \frac{d\vec{p}}{dt} = h(\vec{r}, \vec{p}, t)$$

This type of relationship is known as system of first order differential equations.

2 Past Gene Expression Models

2.1 Linear Transcription Model

In 1999, Chen, He, and Church proposed a simplistic linear differential equation model for gene expression that has a protein feedback loop to transcription [?]. They also regard the degradation and translation rates of the mRNA and protein as constant. This model ignores protein feedback to translation by assuming that this procedure is relatively stable. The network is shown below in Figure ??.

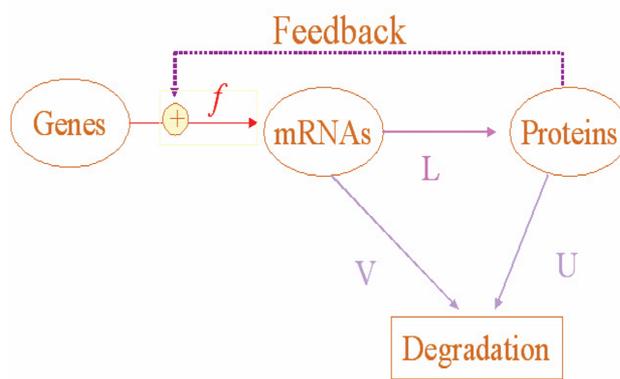


Figure 2: A Linear Transcription Model.

Based on this model, Chen et al. proposed the following system of differential equations to model the system.

$$\frac{d\vec{r}}{dt} = f(\vec{p}) - V\vec{r} \quad \frac{d\vec{p}}{dt} = L\vec{r} - U\vec{p} \quad (1)$$

Note that the variables are functions of time. Assuming the genome has n genes, we have that \vec{r} is an n -vector with r_i representing the concentration of the i^{th} mRNA, and \vec{p} is an n -vector with p_i representing the concentration of the i^{th} protein. f is a vector-valued function that represents the transcription rate of mRNA as a function of proteins (such as polymerase and transcription factors). L is a constant diagonal matrix that represents the translation rate of mRNA into proteins. V and U are constant diagonal matrices where $V_{i,i}$ and $U_{i,i}$ represent the relative degradation rates of the i^{th} mRNA and the i^{th} protein respectively.

Chen et al. argued that the function f can be represented as a constant matrix, thus making system (??) a linear one. This suggested linear differential model has been a basis for further mathematical exploration of gene expression. However, there are some discrepancies between the behavior of the biological system and the behavior that this model predicts. For example, the only equilibrium of system (??) is when $\vec{r} = \vec{0}$ and $\vec{p} = \vec{0}$, thus predicting exponential growth of the concentrations of mRNA and proteins, which does not agree with the biological process. This unrealistic prediction may be due to the fact that there are many interactions that are ignored, such as the feedback from proteins to the transcription process.

2.2 Limited Nonlinear System

In 2003, Kim and Tidor described a nonlinear gene expression model that was controlled at the transcription level [?]. Moreover, they assumed a monotonic dependence on the mRNA and protein concentration levels by the transcription and translation rates. This assumption made solving for steady-states easier and they proved the existence of an asymptotically stable equilibrium solution. They further modified their model to take into account a saturation effect of protein degradation. The equations they developed to model the system are

$$\frac{dr_i}{dt} = f_i(p_i) - V_i(r_i), \quad \frac{dp_i}{dt} = L_i(r_i) - U_i(p_i), \quad (2)$$

for $i = 1, 2, \dots, n$.

The variables r_i and p_i represent the concentrations of the i^{th} mRNA and protein respectively. $f_i(p_i)$ is the transcription rate of the i^{th} mRNA, $V_i(r_i)$ is the degradation rate of i^{th} mRNA, $L_i(r_i)$ is the translation rate of the i^{th} mRNA into the i^{th} protein, and $U_i(p_i)$ is the degradation rate of the i^{th} protein. Also, $U_i(p_i)$ and $V_i(r_i)$ are assumed to be positive and strictly monotonically increasing, while $L_i(r_i)$ is strictly monotonically increasing and $f_i(p_i)$ is strictly monotonically decreasing.

These equations provide a starting point for future gene expression models, but they do not provide explicit functions to characterize their assumptions.

3 Proposed Mathematical System

To construct our mathematical gene expression model, we established our groundwork from the two previously discussed models. Furthermore, we construct a nonlinear model by incorporating multiple feedback loops throughout the protein synthesis process.

Because the bacterial cell structure consists merely of one compartment, transcription and translation simultaneously occur within the same area, which implies that there are no time delays to consider in our model. The cell is constantly enduring rapid change as it prepares for replication into a new identical cell. Since cell replication occurs on a time scale much larger than that of protein synthesis, we consider our model to represent the cell at some time interval between formation and cytokinesis (cell division). During this time, we consider the volume of the cell to be constant.

We regard tRNA and rRNA to be in excess in the cell so that the process of protein synthesis is independent of the concentrations of these types of RNA. So we consider only the effect of mRNA, as in previous models. While there are numerous types of mRNAs and proteins, the biological literature suggests that each can be grouped into three different categories according to the type of protein they produce. The categories are as follows: Type 1 mRNA produce Type 1 proteins which initiate and perform the transcription of *all* mRNA [?], Type 2 mRNA generates Type 2 proteins which help stabilize *all* mRNA against degradation [?], and Type 3 mRNA manufacture Type 3 proteins which are not directly involved in either step of protein synthesis but instead are sent out to the cell for other purposes. We assume that each type of protein regulates the production of its own protein type during the processes of both transcription and translation. This gene regulatory network is shown in Figure ??.

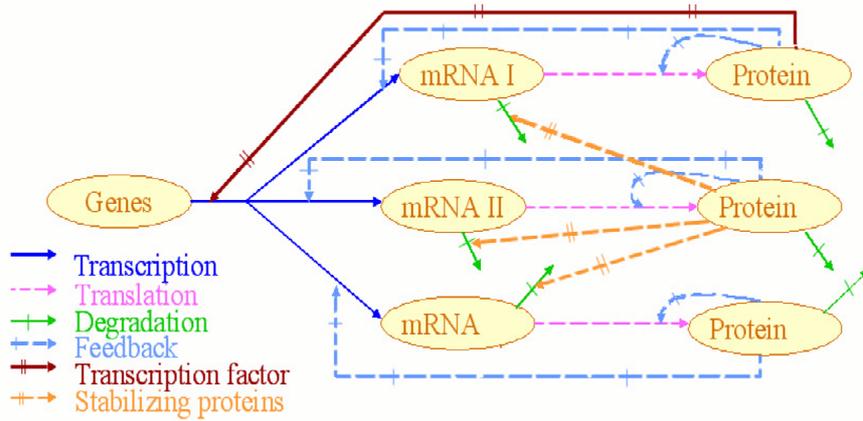


Figure 3: Proposed gene expression model where group mRNA I, mRNA II and mRNA III produce type Protein I, Protein II, and Protein III, respectively.

We now have the transcription of Type i mRNA as a function of Type 1 proteins and Type i proteins. Specifically, the transcription function is a strictly decreasing function of Type i proteins and a strictly increasing function of Type 1 proteins. Thus the presence of more Type 1 proteins will result in a further increase in this transcription while the presence of more Type i proteins will result in a decrease in this transcription function. This is simply because the transcription rate in cells is higher when more transcription factors (Type 1 proteins) are present, but if there are a lot of Type i proteins present, the cell will not need to produce many more of that type of mRNA. We assume that the dependence on Type 1 proteins is linear and somewhat logistic on Type i proteins. The general behavior of the transcription function resembles the function plotted in Figure ?? in two projections: (a) p_1 is fixed, and (b) p_i is fixed, where p_i is the concentration of Type i proteins.

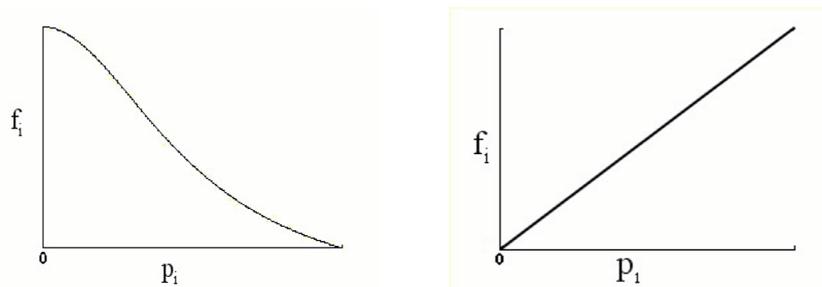


Figure 4: (a) *Left*: The transcription function m strictly monotonically decreases as p_i increases. (b) *Right*: The transcription function increases linearly with respect to p_1 .

Likewise, we want translation to be an increasing function of Type i mRNA, but a decreasing function of Type i proteins. Thus, the presence of more Type i proteins decreases the production of Type i proteins, while the presence of more Type i mRNA increases the production of Type i proteins. We again assume a linear dependence upon Type i mRNA. The translation function should look like the function plotted below in Figure ?? for (a)

fixed r_i and (b) fixed p_i , where r_i represents the concentration of Type i mRNA, and p_i is again the concentration of Type i proteins.

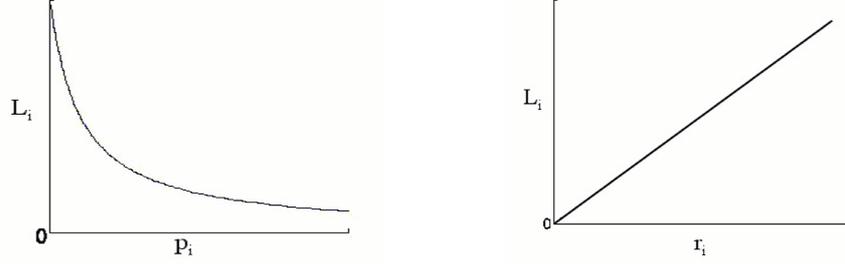


Figure 5: (a) *Left*: The translation function strictly monotonically decreases as p_i increases. (b) *Right*: The translation function of mRNA increases linearly with respect to r_i .

Similarly, we want the degradation of Type i mRNA to be an increasing function of the number of Type i mRNA, but a decreasing function of Type 2 (stabilizing) proteins. This is because the more mRNA there are, the faster they will degrade, but when more stabilizing proteins are present, the mRNA will degrade more slowly. Again assuming that the degradation rate has a linear dependence on Type i mRNA, the function should behave as described below in Figure ??, where (a) p_2 is fixed, and (b) r_i is fixed.

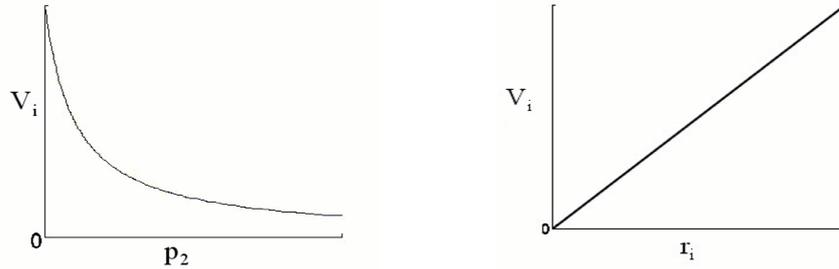


Figure 6: (a) *Left*: The degradation function of mRNA strictly monotonically decreases as p_2 increases. (b) *Right*: The degradation function of mRNA increases linearly with respect to p_i .

From these assumptions we develop the following general equations to model the rates of change of mRNA and protein concentrations:

$$\begin{aligned}
 \frac{dr_1}{dt} &= \frac{1}{1 + \frac{p_1^2}{a_1^2}} C_1 p_1 - \frac{1}{1 + \frac{p_2}{b_1}} V_1 r_1, & \frac{dp_1}{dt} &= \frac{1}{1 + \frac{p_1}{d_1}} L_1 r_1 - U_1 p_1, \\
 \frac{dr_2}{dt} &= \frac{1}{1 + \frac{p_2}{a_2}} C_2 p_1 - \frac{1}{1 + \frac{p_2}{b_2}} V_2 r_2, & \frac{dp_2}{dt} &= \frac{1}{1 + \frac{p_2}{d_2}} L_2 r_2 - U_2 p_2, \\
 \frac{dr_3}{dt} &= \frac{1}{1 + \frac{p_3}{a_3}} C_3 p_1 - \frac{1}{1 + \frac{p_2}{b_3}} V_3 r_3, & \frac{dp_3}{dt} &= \frac{1}{1 + \frac{p_3}{d_3}} L_3 r_3 - U_3 p_3
 \end{aligned} \tag{3}$$

where r_i and p_i are the concentration of the Type i mRNA and protein in unit of nMolar. Parameters C_i and L_i are the relative transcription and translation rates in the absence of feedback loops. Parameters U_i and V_i are the relative natural degradation rates of mRNA and protein. Parameters a_i , b_i , and d_i are the effectiveness factors of the respective feedback loops with the same unit as p_i . The bigger the value of parameter a_i , the smaller the effect of the feedback into the transcription term. Likewise, b_i and d_i controls the effectiveness of the feedback into degradation of mRNA and translation of protein, respectively.

Note that the p_1 and a_1 are squared in the denominator of the transcription term in the equation for $\frac{dr_1}{dt}$, thus giving the decrease suggested in Figure ??, where as p_2, p_3 are linear in the denominators of the $\frac{dr_2}{dt}$ and $\frac{dr_3}{dt}$ terms thus giving a decrease suggested in Figure ??a. This is done to ensure that for very large values of p_i , the transcription term tends towards 0 instead of towards $C_1 a_1$. We note that mathematically we will have discontinuities in our system whenever $p_2 = -a_2, p_3 = -a_3, p_2 = -b_i$ or $p_i = -d_i, i = 1, 2, 3$. Although these discontinuities, which are not biologically relevant, may appear as “false” equilibria in solving these equations.

4 Analyzing the Stability of the System

4.1 Stability of the Origin

It can easily be shown that the origin of system (??) is an equilibrium point, since $\frac{dp_i}{dt} = 0$ and $\frac{dr_i}{dt} = 0$ for $i = 1, 2, 3$ at $(r_1, p_1, r_2, p_2, r_3, p_3) = (0, 0, 0, 0, 0, 0)$. The eigenvalues of the Jacobian matrix of the system at the origin are found to be

$$\begin{aligned} \lambda_1 &= -U_2, & \lambda_2 &= -U_3, & \lambda_3 &= -\frac{1}{2} \left(U_1 + V_1 + \sqrt{(U_1 - V_1)^2 + 4C_1 L_1} \right), \\ \lambda_4 &= -V_2, & \lambda_5 &= -V_3, & \lambda_6 &= -\frac{1}{2} \left(U_1 + V_1 - \sqrt{(U_1 - V_1)^2 + 4C_1 L_1} \right). \end{aligned} \quad (4)$$

The eigenvalues $\lambda_1, \dots, \lambda_5$ are always negative since C_1, V_1, L_1 , and U_1 are positive. When $\alpha = L_1 C_1 - U_1 V_1$, the product of the production terms minus the product of the degradation terms of Type 1 mRNA and protein, is negative, we have $\lambda_6 < 0$, and thus the origin is a stable node. This implies that the concentrations of all mRNA and proteins will approach zero as time goes on when the product of the degradation rates of Type 1 products is smaller than the product of the transcription and translation rates of Type 1 mRNA and protein. The reason that the stability depends only on Type 1 products is because the transcription of all mRNA depends on Type 1 proteins, which depends on Type 1 mRNA. Thus if these products are decaying faster than they are being produced, all products will eventually die out, making the origin a stable point. However, when $\alpha > 0$, λ_6 is positive, and thus the origin becomes a saddle, meaning almost all initial concentration of mRNAs and proteins near the origin will move away from it. Only in certain conditions will the system approach the origin, such as when $r_1 = 0$ and $p_1 = 0$. In this case no mRNA will ever be produced, and the system will eventually die out.

The origin has a $\lambda = 0$ bifurcation since λ_6 is zero when $\alpha = 0$. The specific type of $\lambda = 0$ bifurcation will be discussed in a later section. On the other hand, the origin does

not undergo a Hopf bifurcation since none of the eigenvalues can ever be purely imaginary. From this, we conclude there are no limit cycles around the origin at any time [?].

4.2 Simplifying the System

Taking a closer look at the first four equations in (??), we see that they are independent of the variables r_3 and p_3 . These four equations can be solved independently of the other two. Furthermore, any equilibrium of the the six-dimensional system will also be an equilibrium of the four-dimensional one. Thus we need only to analyze the first four equations and then use the results to infer the behavior of the full system.

$$\begin{aligned} \frac{dr_1}{dt} &= \frac{1}{1 + \frac{p_1^2}{a_1^2}} C_1 p_1 - \frac{1}{1 + \frac{p_2}{b_1}} V_1 r_1, & \frac{dp_1}{dt} &= \frac{1}{1 + \frac{p_1}{d_1}} L_1 r_1 - U_1 p_1, \\ \frac{dr_2}{dt} &= \frac{1}{1 + \frac{p_2}{a_2}} C_2 p_1 - \frac{1}{1 + \frac{p_2}{b_2}} V_2 r_2, & \frac{dp_2}{dt} &= \frac{1}{1 + \frac{p_2}{d_2}} L_2 r_2 - U_2 p_2. \end{aligned} \quad (5)$$

Due to the numerous parameters in our system, it is helpful to obtain values for the constants to somehow relieve the complexity of the system. Unfortunately, it is difficult to measure the rates of production and decay of mRNA and proteins, so exact values are not available for these rates. However, we can make estimates for the relative rates of mRNA and protein production and degradation. It is also possible to reduce the number of parameters via non-dimensionalization, but we opt to leave things in dimensional form to more easily interpret rates.

To estimate the relative degradation rates of mRNA (V_i) and proteins (U_i) we consider their typical half lives. Biological literature suggests that typical mRNAs have an average half life of about 22 minutes, which we assume is the generic rate for all mRNAs in our system [?]. From this we obtain $V_i \approx .03/\text{min}$. Furthermore, protein half life is approximated in the range of a few minutes to 50 minutes. We assume that the general Type 3 protein and Type 3 has a half life of about 50 minutes [?]. Moreover, we assume that Type 1 proteins degrade quickly so that the cell can easily regulate the number of mRNA that are produced from a given number of Type 1 proteins. From this we approximate the degradation rates for proteins to be $U_1 \approx .15/\text{min}$ and $U_2 \approx .015/\text{min}$.

After gene expression begins, mRNA will appear in the cell after about 2.5 minutes with its corresponding protein appearing in an additional .5 minutes [?]. This shows that in the presence of the necessary number of Type 1 proteins, transcription will occur at a rate of 1 mRNA per 2.5 minutes. Since the typical number of proteins needed for transcription is about 15, we estimate the transcription rate as $C_i \approx .03 \text{ mRNA}/(\text{protein}\cdot\text{min})$. Also, since it takes .5 minutes for the protein to be produced once the mRNA is formed, we can estimate the translation rate as $L_i \approx 2 \text{ protein}/(\text{mRNA}\cdot\text{min})$ [?].

Here we list our approximate values for easy reference:

$$\begin{aligned} C_i &\approx .03\text{mRNA}/(\text{protein min}) & L_i &\approx 2\text{protein}/(\text{mRNA min}) \\ U_1 &\approx .15\text{min} & U_i &\approx .015\text{min, for } i = 2, 3 \\ V_i &\approx .03\text{min} & & \end{aligned} \quad (6)$$

These values are approximate, and are therefore used only to gain insight into the behavior of the system under realistic conditions. It is necessary to also point out that it may be possible to vary some of these rates through experimental methods.

4.3 Extreme Cases

The system in (??) has fourteen parameters, making it very difficult to analyze. It also has several nonlinearities making it difficult to obtain reasonable expressions for the equilibrium solutions. Therefore, we will look at several special cases of the system in which certain parameters and terms can be ignored.

Case 1. $b_i \gg p_2$ and $d_i \gg p_i$

In this case, we suppose that the stabilization of mRNA and the feedback from proteins to translation are negligible in comparison to the other interactions of the system. System (??) then simplifies to

$$\begin{aligned} \frac{dr_1}{dt} &= \frac{1}{1 + \frac{p_1^2}{a_1^2}} C_1 p_1 - V_1 r_1, & \frac{dp_1}{dt} &= L_1 r_1 - U_1 p_1, \\ \frac{dr_2}{dt} &= \frac{1}{1 + \frac{p_2^2}{a_2^2}} C_2 p_1 - V_2 r_2, & \frac{dp_2}{dt} &= L_2 r_2 - U_2 p_2. \end{aligned} \quad (7)$$

This system has an equilibrium point at the origin, which as previously stated, is a saddle (if $\alpha > 0$) or stable node (if $\alpha < 0$). To find other equilibrium points, we set the equations (??) equal to zero and solve for their respective concentration variable which yields an additional four equilibrium points. The four equilibrium points, given as (r_1, p_1, r_2, p_2) , are

$$\left(X_{-eq} \frac{V_2 U_1}{C_2 a_2 U_2 L_1} (a_2 U_2 + X_{-eq} L_2), \quad X_{-eq} \frac{V_2}{C_2 a_2 U_2} (a_2 U_2 + X_{-eq} L_2), \quad X_{-eq}, \quad X_{-eq} \frac{L_2}{U_2} \right), \quad (8)$$

$$\left(X_{+eq} \frac{V_2 U_1}{C_2 a_2 U_2 L_1} (a_2 U_2 + X_{+eq} L_2), \quad X_{+eq} \frac{V_2}{C_2 a_2 U_2} (a_2 U_2 + X_{+eq} L_2), \quad X_{+eq}, \quad X_{+eq} \frac{L_2}{U_2} \right), \quad (9)$$

$$\left(Y_{-eq} \frac{V_2 U_1}{C_2 a_2 U_2 L_1} (a_2 U_2 + Y_{-eq} L_2), \quad Y_{-eq} \frac{V_2}{C_2 a_2 U_2} (a_2 U_2 + Y_{-eq} L_2), \quad Y_{-eq}, \quad Y_{-eq} \frac{L_2}{U_2} \right), \quad (10)$$

$$\left(Y_{+eq} \frac{V_2 U_1}{C_2 a_2 U_2 L_1} (a_2 U_2 + Y_{+eq} L_2), \quad Y_{+eq} \frac{V_2}{C_2 a_2 U_2} (a_2 U_2 + Y_{+eq} L_2), \quad Y_{+eq}, \quad Y_{+eq} \frac{L_2}{U_2} \right), \quad (11)$$

where

$$X_{\pm eq} = -\frac{B_f \pm \sqrt{B_f^2 + 4A_f C_f \sqrt{\alpha}}}{2A_f}, \quad (12)$$

$$Y_{\pm eq} = -\frac{B_f \pm \sqrt{B_f^2 - 4A_f C_f \sqrt{\alpha}}}{2A_f}, \quad (13)$$

and

$$\begin{aligned} \alpha &= C_1 L_1 - V_1 U_1, & A_f &= L_2 V_2 V_1 U_1, \\ B_f &= a_2 U_2 V_2 V_1 U_1, & C_f &= \sqrt{V_1 U_1} U_2 C_2 a_2 a_1. \end{aligned} \quad (14)$$

All four equilibria will be real-valued when $\alpha > 0$ and $B_f^2 - 4A_f C_f \sqrt{\alpha} > 0$. Only the first two equilibria, (??) and (??), will be real-valued if $\alpha > 0$ but $B_f^2 - 4A_f C_f \sqrt{\alpha} < 0$. None of the equilibria will be real-valued when $\alpha < 0$. In each of these situations, the origin is still always an equilibrium solution.

In the case when $B_f^2 - 4A_f C_f \sqrt{\alpha} = 0$, we see that $Y_{+eq} = Y_{-eq}$ and thus the expressions in (??) and (??) are identical. Based on the above information, we can conclude that a saddlenode bifurcation occurs with these two equilibrium solutions for the set of parameter values satisfying the expression $\sqrt{\alpha} = \frac{B_f^2}{4A_f C_f}$.

In the case when $\alpha = 0$, we see that $X_{+eq} = Y_{+eq}$, and thus the expressions in (??) and (??) are identical. We can thus conclude based on the above information that a saddlenode bifurcations occur at this point for the set of parameter values satisfying $\alpha = 0$.

Furthermore, when $\alpha = 0$, we find that

$$X_{-eq} = -\frac{B - B}{2A_f} = 0 \quad Y_{-eq} = -\frac{B - B}{2A_f} = 0.$$

When we substitute $X_{-eq} = 0$ into (??) and $Y_{-eq} = 0$ into (??), we find that the two equilibria are both equal to the origin. We have two equilibria being born from the equilibrium at the origin, so the system undergoes a pitchfork bifurcation when $\alpha = 0$. These bifurcations are shown in Figure (??) as a bifurcation diagram for p_2 as a function α .

In order for the system to have nontrivial equilibrium points, the condition $\alpha > 0$ must be met. To analyze whether these equilibrium points are possible biologically, we see that points (??), (??), and (??) will always have a negative r_2 variable for any parameter values, and they are thus discarded. Equilibrium point (??) will always be located in the positive hyper-octant for $\alpha > 0$. Using our values in (??) and assuming $a_2 = 1$ and $a_1 = 1$, we find the eigenvalues of the Jacobian matrix evaluated at the origin are complex and the real parts are negative. This implies that a stable spiral exists at (??) when $\alpha > 0$, which is biologically relevant for the system.

To investigate the possibility of a Hopf bifurcation occurring in our system, we substitute $\lambda = i\omega$ into the characteristic polynomial of the Jacobian matrix of system (??) evaluated at the equilibria and discover the following two conditions:

$$U_2 + V_2 = 0 \quad \text{and} \quad V_1 + U_1 = 0. \quad (15)$$

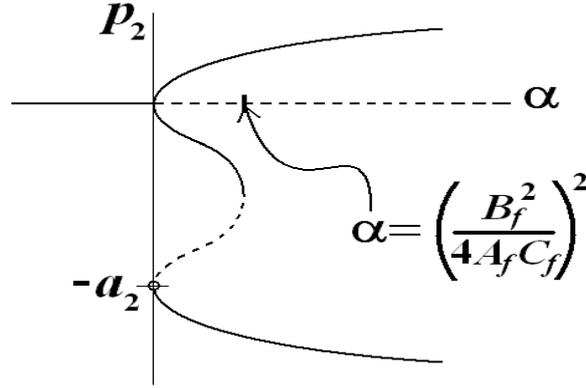


Figure 7: Equilibria and Bifurcation for Transcription.

To make sense biologically, the parameters must have positive values. Therefore, each sum is positive, and it is very unlikely that a Hopf bifurcation will occur. A very large expression of parameters gave a potential Hopf bifurcation, but we were not able to obtain useful relationships among any of the parameters.

In this case, where only feedback into transcription is considered, we obtained one biologically relevant stable point when $\alpha > 0$ because the origin is unstable. This is of interest to biologists because they can possibly study the mRNA and protein concentrations and how long it takes for them to reach saturation by adjusting the rate of transcription. This also shows that feedback into transcription is a necessary component in gene expression.

A plot of the typical trajectories for this system is found below in Figure ???. The initial conditions and parameter values for this particular graph are given by Table ??? below where all values are in appropriate units.

		I.C.		Parameters					
i	r_i	p_i	C_1	L_i	V_i	U_i	a_i	b_i	d_i
1	3	100	0.03	2	0.03	0.15	60	120	120
2	6	500	0.03	2	0.03	0.015	140	140	150
3	5	1	0.024	2	0.03	0.015	170	180	260

$\alpha = 0.055$	$\beta = -0.033$	$\delta = 2.3E6$
------------------	------------------	------------------

Table 1: Parameters and initial conditions used in numerical experiments.

Notice in the plot that the concentrations come very close to a steady state solution after about five hours.

Case 2. $a_i, d_i \gg p_i$

This is the case when the feedback from proteins to both transcription and translation is negligible. Thus, the only nonlinear effect is that of p_2 preventing the degradation of mRNA.

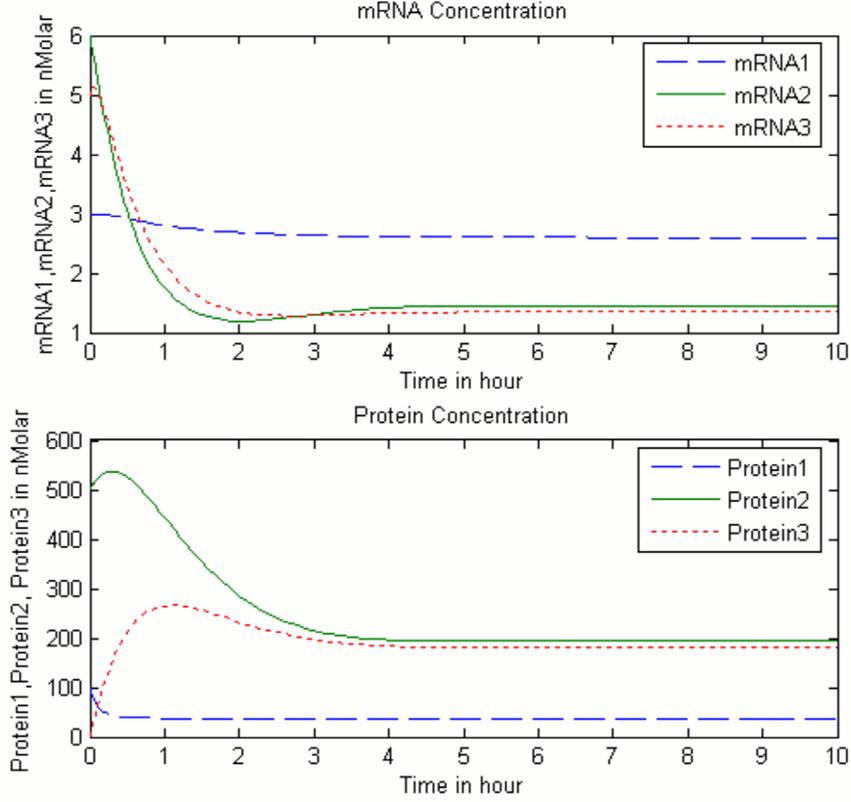


Figure 8: Numerical solutions for Case 1 with parameters and initial conditions from Table ??.

System (??) then simplifies to

$$\begin{aligned}
 \frac{dr_1}{dt} &= C_1 p_1 - \frac{1}{1 + \frac{p_2}{b_1}} V_1 r_1, & \frac{dp_1}{dt} &= L_1 r_1 - U_1 p_1, \\
 \frac{dr_2}{dt} &= C_2 p_1 - \frac{1}{1 + \frac{p_2}{b_2}} V_2 r_2, & \frac{dp_2}{dt} &= L_2 r_2 - U_2 p_2.
 \end{aligned} \tag{16}$$

For $\alpha = C_1 L_1 - U_1 V_1$, the fixed points of this system are $(r_1, p_1, r_2, p_2) = (0, 0, 0, 0)$ and

$$(r_1, p_1, r_2, p_2) = \left(\frac{U_1 V_2 b_2 U_2 b_1 \alpha}{C_2 L_2 L_1 (-b_2 C_1 L_1 + b_1 \alpha)}, \frac{V_2 b_2 U_2 b_1 \alpha}{C_2 L_2 (-b_2 C_1 L_1 + b_1 \alpha)}, -\frac{b_2 U_2 b_1 \alpha}{C_2 L_2 L_1}, -\frac{b_1 \alpha}{C_1 L_1} \right),$$

For $\alpha < 0$, the origin is stable and the nontrivial equilibrium point possibly resides in the positive hyper-octant. This can be seen in Figure ??, which is a bifurcation diagram of system (??) in one dimension, p_2 . However, this point is unstable, which is not biologically relevant. When $\alpha = 0$, the origin is the only solution, and its stability cannot be determined by linear analysis. For $\alpha > 0$, again there are two solutions, but the stability swaps between the origin and the second equilibrium point. Also, the non-trivial stable equilibrium has negative components, which is irrelevant in terms of biology. Since the origin is unstable and no stable

point exists in the positive hyper-octant, for realistic initial values, the concentrations will grow without bound. This can be seen in Figure ??.

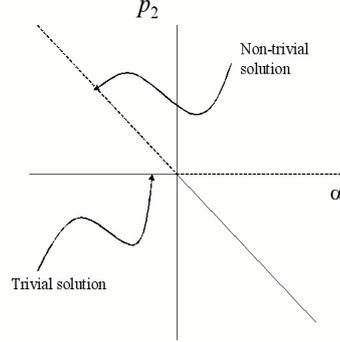


Figure 9: Behavior of the stability of equilibrium points near $\alpha = 0$ in Case 2.

$\lambda = 0$ bifurcation occurs at the origin when $\alpha = 0$. The dynamics of the equilibrium points imply that the specific type of $\lambda = 0$ bifurcation is a transcritical bifurcation. Also, the eigenvalues at each equilibrium point can never be purely imaginary; thus, there are no Hopf bifurcation and limit cycle [?].

As mentioned previously, this case which only considers feedback into mRNA degradation has no biological meaning. However, this indicates that the neglected feedbacks into transcription and translation are critical components of the system.

Case 3. $a_i \gg p_i$ and $b_i \gg p_2$

Here we consider that the feedback from proteins to transcription and the stabilization of mRNA is negligible. System (??) then simplifies to

$$\begin{aligned} \frac{dr_1}{dt} &= C_1 p_1 - V_1 r_1, & \frac{dp_1}{dt} &= \frac{1}{1 + \frac{p_1}{d_1}} L_1 r_1 - U_1 p_1, \\ \frac{dr_2}{dt} &= C_2 p_1 - V_2 r_2, & \frac{dp_2}{dt} &= \frac{1}{1 + \frac{p_2}{d_2}} L_2 r_2 - U_2 p_2. \end{aligned} \quad (17)$$

There are 3 possible equilibrium points, one of which is the origin. The other two, given as (r_1, p_1, r_2, p_2) , are

$$\left(\frac{C_1 d_1 \alpha_1}{U_1 V_1^2}, \frac{d_1 \alpha_1}{U_1 V_1}, \frac{C_2 d_1 \alpha_1}{U_1 V_1 V_2}, \frac{-B_l + \sqrt{B_l^2 + 4B_l C_2 L_2 d_1 \alpha}}{2U_1 U_2 V_1 V_2} \right), \quad (18)$$

$$\left(\frac{C_1 d_1 \alpha_1}{U_1 V_1^2}, \frac{d_1 \alpha_1}{U_1 V_1}, \frac{C_2 d_1 \alpha_1}{U_1 V_1 V_2}, \frac{-B_l - \sqrt{B_l^2 + 4B_l C_2 L_2 d_1 \alpha}}{2U_1 U_2 V_1 V_2} \right), \quad (19)$$

where

$$\begin{aligned} \alpha &= C_1 L_1 - V_1 U_1, \\ B_l &= d_2 U_2 V_2 V_1 U_1. \end{aligned}$$

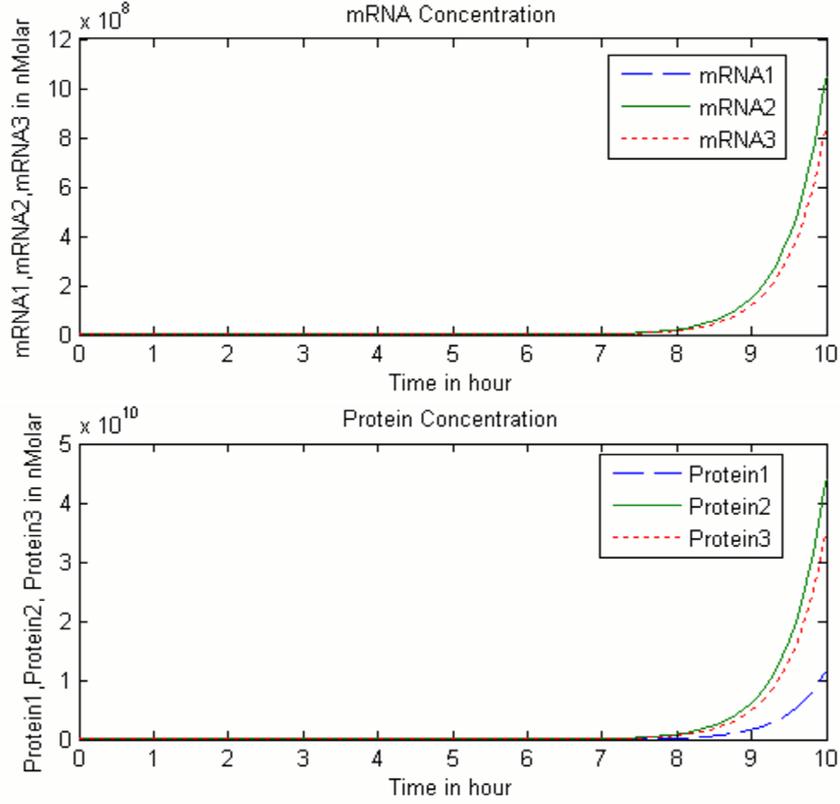


Figure 10: Numerical solutions for Case 2 with parameters and initial conditions from Table ??.

For $\alpha > 0$ there are always three equilibrium solutions, two stable non-trivial solutions and an unstable origin. For $\alpha < 0$, it seems like there are still three equilibrium solutions if $B_i > -4C_2L_2d_1\alpha$. However, a careful examination shows that under these conditions for α and β , the two non-trivial solutions are false equilibrium solutions. Thus the origin is the only equilibrium point when $\alpha < 0$, which turns out to be stable.

As seen from the bifurcation diagram, Figure ??, we see that there is a pitchfork bifurcation at the origin when $\alpha = 0$. There is no Hopf bifurcation since the eigenvalues at each equilibrium point can never be purely imaginary. Just like in Case 1, we have one relevant stable equilibrium solution when $\alpha > 0$. Thus when $\alpha > 0$, any relevant initial concentrations will eventually settle into positive concentrations. Again, using the same set of parameter values and initial condition found in Table ??, a plot of concentrations of mRNA and protein over time is shown in Figure ??.

Case 4. $d_i \gg p_i$

In this case we have that the feedback from proteins to translation is negligible. System

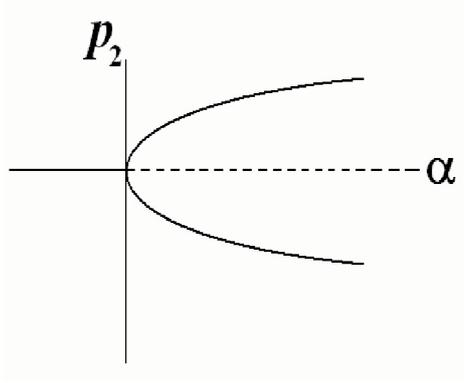


Figure 11: Behavior of the stability of equilibrium points near $\alpha = 0$ in Case 3.

(??) then simplifies to

$$\begin{aligned} \frac{dr_1}{dt} &= \frac{1}{1 + \frac{p_1^2}{a_1^2}} C_1 p_1 - \frac{1}{1 + \frac{p_2}{b_1}} V_1 r_1, & \frac{dp_1}{dt} &= L_1 r_1 - U_1 p_1, \\ \frac{dr_2}{dt} &= \frac{1}{1 + \frac{p_2}{a_2}} C_2 p_1 - \frac{1}{1 + \frac{p_2}{b_2}} V_2 r_2, & \frac{dp_2}{dt} &= L_2 r_2 - U_2 p_2, \end{aligned} \quad (20)$$

and we observe that the change in the concentration of proteins has a linear dependence on r_i and p_i .

By setting each of the equations in (??) to zero, we find that we again have an equilibrium point at the origin as well as up to four other points elsewhere. We find that the system has $\lambda = 0$ eigenvalues when $\alpha = 0$ or when $\beta = 0$, where β is an expression of parameters that will not be shown here because of its length and complexity. There is also a third condition for bifurcation: $\alpha = \frac{L_1 C_1 a_2}{b_1}$. This corresponds to a removable discontinuity in the system of equations and is not a true equilibrium of the system.

In numerical experimentation, we find four non-trivial equilibria only for very large b_1 and b_2 . In this case our bifurcation diagram is analogous to that discussed in Section 4.3.1. Otherwise, varying the parameter C_1 , which in effect varies α , and setting all other production and degradation rates to the values in (??), we find that the system has at most two equilibria other than the origin. The projections of this particular bifurcation curve as a function of C_1 are shown below in Figure ??.

When $\alpha = 0$ we have exactly one other equilibrium point, with a double root of sorts at the origin. For negative $\alpha \approx 0$, we have a stable equilibrium at the origin and an unstable equilibrium very close to the origin. For positive $\alpha \approx 0$, the origin is an unstable equilibrium and we have a stable equilibrium very close to the origin. From this we know that a transcritical bifurcation occurs when $\alpha = 0$.

When $\beta = 0$, we have exactly two equilibria with a double root at the non-zero solution. If $\beta < 0$, we have only the origin as an equilibrium point, and for very small $\beta > 0$, we have two equilibria other than the origin. We can see from this that a saddlenode bifurcation occurs when $\beta = 0$. For the values in (??), this bifurcation point is in the positive hyper-octant.

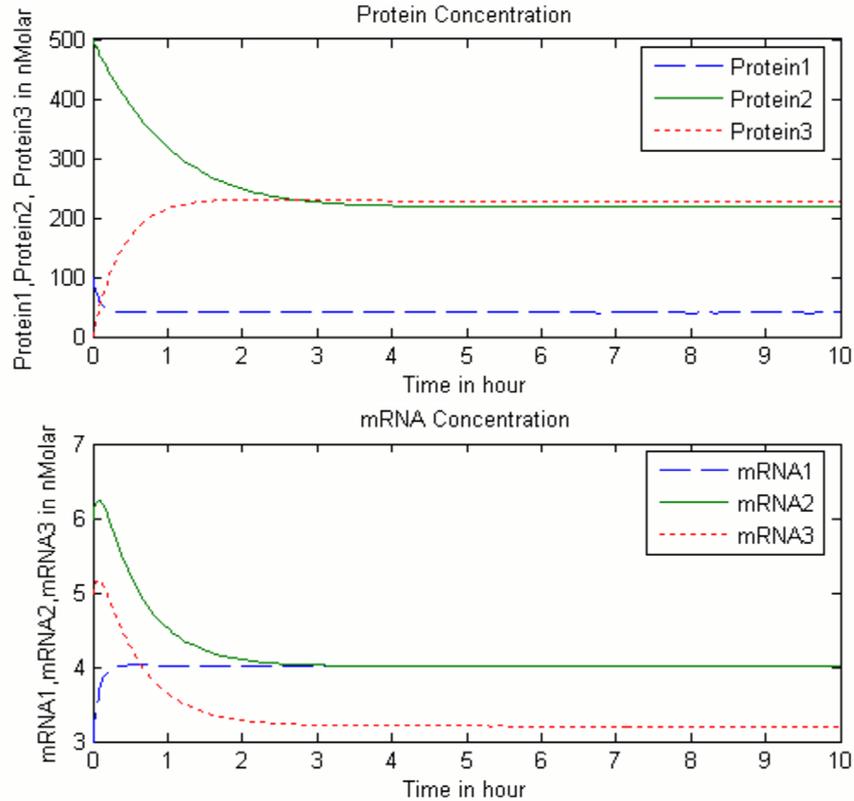


Figure 12: Numerical solutions for Case 3 with parameters and initial conditions from Table ??.

From Figure ?? we see that for parameters that satisfy the condition that $\alpha < 0 < \beta$, we have three equilibria: the origin and two in the positive hyper-octant. The origin is stable, the point closer to the origin is unstable, and the other point is stable. This is a potentially interesting case for experimentalists since, for realistic (positive) initial conditions, it is possible for the system to approach one of two stable equilibria. By fixing parameters so that $\alpha < 0 < \beta$ and by adjusting the initial conditions, this conclusion can be illustrated numerically.

A plot of the typical trajectories for this system is found below in Figure ??. Notice that the concentrations come to steady state after almost 500 hours. This is on a much larger scale than other cases. This may indicate that the stabilizing proteins are necessary in order to get realistic solutions.

4.4 The Six-Dimensional System

Biologically, the main points of interests in this system are the stable equilibrium points. If we have an equilibrium of the six-dimensional system, its projection into the four dimensional subspace described by $r_1, p_1, r_2,$ and p_2 must be an equilibrium point of the four-dimensional system in (?). Thus one can first find a four-dimensional equilibrium point, $(r_1^*, p_1^*, r_2^*, p_2^*),$

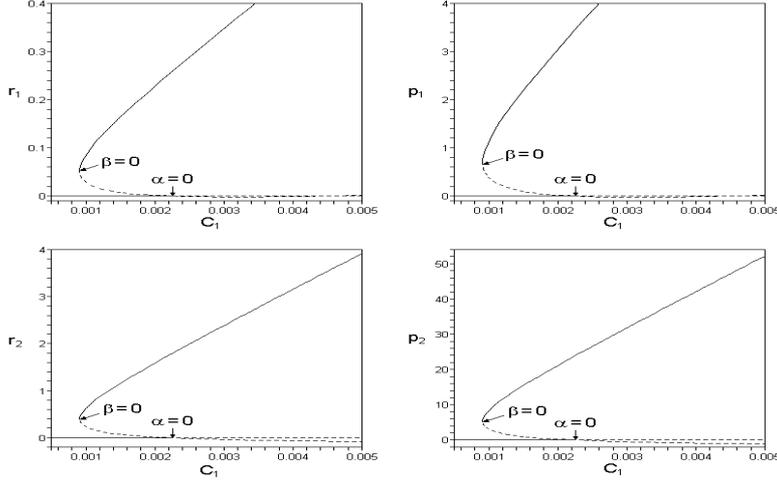


Figure 13: Bifurcation diagram for system (??) in various projections with values from (??).

then solve the following system found by setting $\frac{dr_3}{dt} = 0$ and $\frac{dp_3}{dt} = 0$:

$$0 = \frac{1}{1 + \frac{p_3}{a_3}} C_3 p_1^* - \frac{1}{1 + \frac{p_2^*}{b_3}} V_3 r_3, \quad 0 = \frac{1}{1 + \frac{p_3}{d_3}} L_3 r_3 - U_3 p_3. \quad (21)$$

There are at most three solutions to these equations. Each solution must be entered into the 6×6 Jacobian matrix and the eigenvalues must be found in order to determine the stability of the six-dimensional equilibrium point.

5 Conclusion

In Case 1, where stabilization of mRNA and protein feedback into translation is negligible, supercritical pitchfork bifurcation occurs at the origin when $\alpha = 0$. When $\alpha > 0$, which means the origin is a saddle, there exists one stable equilibrium that is biologically relevant. This can be interpreted as a saturation effect occurring for the concentration of mRNA and proteins at this point. If $\alpha < 0$ then there is no nontrivial equilibrium, but the origin is stable. These results can be tested by adjusting one or all of the production and degradation rates of Type 1 products, and thus adjusting α . The experiments could verify that if $\alpha < 0$, then the system should die out, and if $\alpha > 0$, then the system should approach a nontrivial equilibrium. These experiments may also be applied to Case 3, where feedback from proteins to transcription and the stabilization of mRNA is negligible, because it has similar supercritical pitchfork bifurcation when $\alpha = 0$ and one stable equilibrium when $\alpha > 0$.

In Case 2, where the feedback from proteins to transcription and translation is negligible and when $\alpha < 0$, the origin is stable and there is a equilibrium point in the positive hyper-octant that is unstable. If $\alpha > 0$, then the origin is a saddle with a stable point in the negative hyper-octant. Because the only stable biologically relevant equilibrium is the origin,

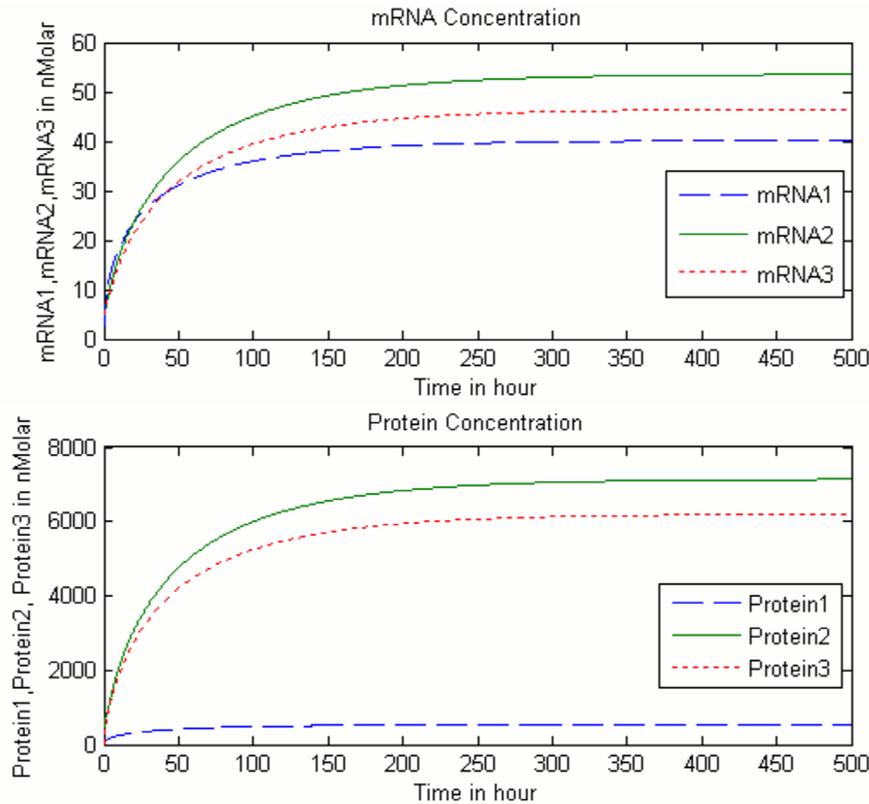


Figure 14: Numerical solutions for Case 4 with parameters and initial conditions from Table ??.

this case is not biologically relevant. This indicates that the feedback into transcription and translation are critical elements of our model and cannot be neglected.

For Case 4, where feedback from proteins to translation is negligible, there is a saddlenode bifurcation in the positive hyper-octant, and there is a transcritical bifurcation at the origin when $\alpha = 0$. This seems to be the qualitative behavior of system (??) as well. For parameter values that put the system between these two bifurcation points, the origin and a nontrivial biologically relevant point are stable equilibria. In this situation, by varying initial conditions, experimentalists can possibly verify that each stable equilibrium exists, thus giving even more credibility to our model.

The process of gene expression through the production of proteins is an important function in every organism and we have provided a rather accurate system of differential to model this procedure which includes the cell's sensory of its environment and mRNA and protein interaction. Most importantly our model uses explicit functions to describe these relationships which, at present, is hardly attempted by other researchers. Aside from bacteria, our model may be considered to represent simple eukaryotic organisms, such as yeast, and provide insight on future mathematical modelling of higher ordered eukaryotic species.

Our model does carry some fault because relationships within the cell are constantly changing from environmental stimuli and in preparation for replication. At this time, our model lacks the general consideration of overall cell health due to external factors like nu-

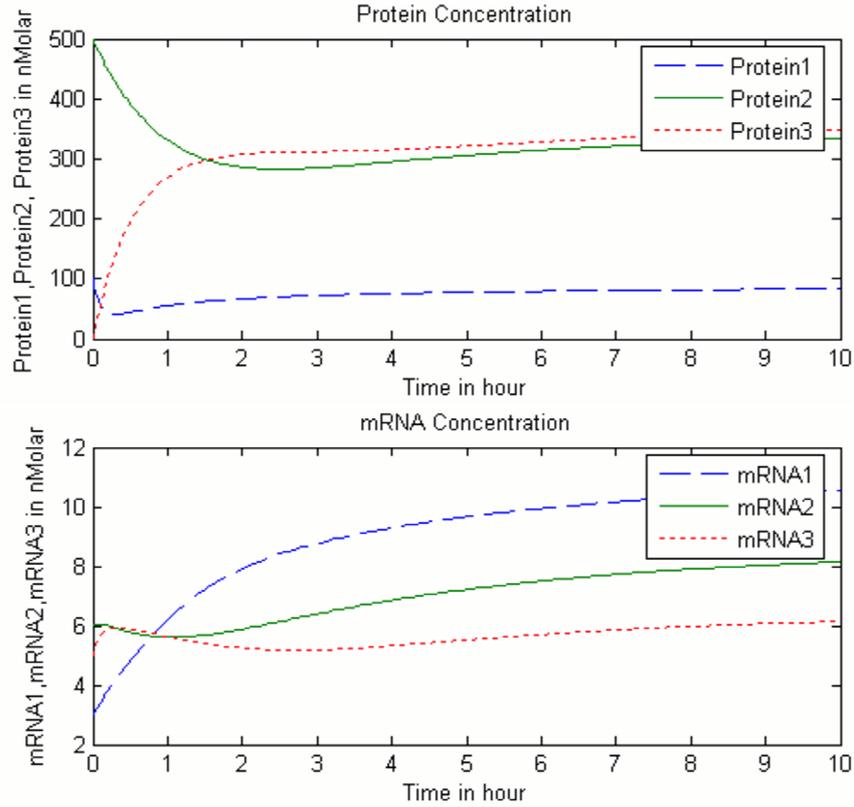


Figure 15: Numerical solution of the six dimensional system with parameters and initial conditions from Table ??.

trients and temperature. Our classification of mRNAs and proteins are slightly skewed with each other since protein Type 3 are the biggest concentration within a cell. Our estimated rates (??) seem to need some adjustment to fit a time scale relevant to the actual time of cell division. Since our model looks at gene expression in a global view, there is presently no exact data available to compare our model with in order to alter it to accommodate actual real life cellular conditions.

Overall, our individual analysis of the possible extreme cases of the four dimensional continuous differential equations gene expression provides insight, through stability and bifurcation analysis, that may be used to describe the intricate processes of gene expression.

6 Acknowledgements

This research was conducted at the Applied Mathematical Sciences Summer Institute (AMSSI) and has been partially supported by grants given by the National Science Foundation (DMS-0453602) and National Security Agency (MSPF-04IC-227). Substantial financial and moral support was also provided by Don Straney, Dean of the College of Science at California State Polytechnic University, Pomona. Additional financial and moral support was provided by the Department of Mathematics at Loyola Marymount University and the Department of

Mathematics & Statistics at California State Polytechnic University, Pomona. We would like to give a big thank you to our advisors, Erika Camacho and Stephen Wirkus for their mathematical guidance and allowing us the opportunity to experience mathematic research success. Furthermore, we are very grateful for the biological insight from Carl Urbinati and Kam Dahlquist as we prepared our gene expression model. The authors are solely responsible for the views and opinions expressed in this research; it does not necessarily reflect the ideas and/or opinions of the funding agencies and/or Loyola Marymount University or California State Polytechnic University, Pomona.

References

- [1] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter *Essential Cell Biology, Second Edition*. Garland Science, 2004.
- [2] Kristi Lynn Trogman [http : //www.saintmarys.edu/ trog7167/Proteinsynthesis.htm](http://www.saintmarys.edu/trog7167/Proteinsynthesis.htm) Aug, 2005.
- [3] Ting Chen, Hongyu He, and George M. Church *Modeling Gene Expression with Differential Equations*. 1999
- [4] Philip Kim and Bruce Tidor *Limitations of Quantitative Gene Regulation Models: A Case Study*. 2003 <http://www.genome.org/cgi/content/full/13/11/2391>
- [5] Benjamin Lewin *Genes VIII*. Pearson Education,Inc., Upper Saddle River, NJ 2004.
- [6] Jacques Lapointe and Lea Brakier-Gingras *Translation Mechanisms*. Kluwer Academic/Plenum Publishers,NY, 2003
- [7] Steven H. Strogatz *NonLinear Dynamics and Chaos*. Perseus Books Publishing, MA, 1994.
- [8] Carol J. Wilusz and Jeffrey Wilusz *Bringing the Role of mRNA Decay in the Control of Gene Expression into Focus*. TRENDS in Genetics 20(10), 2004
- [9] Nicolas E. Buchler, Ulrich Gerland, and Terence Hwa *Nonlinear Protein Degradation and the Function of Genetic Circuits*. Proceedings of the National Academy of Sciences of the United States of America, 102(27):9559-9564, 2005.