A Classical Gene Regulatory Network

Hal L. Smith*

March 9, 2005

Tryptophan is one of the 20 amino acids that link together to form proteins. For humans it is an essential amino acid, we must get it from our diet since we cannot synthesize it. E. coli bacterial cells, however, can synthesize the amino acid tryptophan when it is not provided from the environment (e.g. the gut where it lives). It would be wasteful to synthesize tryptophan when it is readily available in the environment so E. coli have evolved a genetic switch (the trp operon) that turns off synthesis in this case and turns on synthesis when tryptophan is no longer present in the environment. The following model is based loosely on the control of tryptophan production by the trp operon. Although the trp gene codes for 5 enzymes that act on five substrates, or precursors, we simplify it a bit by considering only one of each. Our treatment follows Banks and Mahaffy [2], who considered a general negative feedback genetic control model. Their model, hence ours too, does not include transcriptional attenuation of the trp operon. See Santillan and Mackey [7] for a more realistic model of the trp operon. Our references contain further related material. Check the web page http://science.nhmccd.edu/biol/operon/ton.html for an animated description of the trp operon.

1 Reactions

Name	Symbol	Description
DNA	DNA	the gene of interest
RNA Polymerase	RNAP	enzyme that reads DNA producing mRNA
mRNA	mRNA	messenger RNA-working copy of the gene
tRNA	tRNA	transfer RNA-converts mRNA code into protein E
Enzyme	E	protein product of gene
Precursor	X	substrate converted to tryptophan by catalyst E
tryptophan	T	amino acid required by the cell
Prerepressor	P	becomes a repressor when complexed with tryptophan
Repressor	R	blocks the translation of DNA when bound to DNA

Table 1: Main Players: the chemical species we will follow in the model

Its useful to begin with a verbal model of the trp operon. It begins with RNA polymerase binding to DNA, the gene, and initiating the process of copying the DNA code to messenger RNA. This is called transcription. Messenger RNA then must be translated into protein, in our case, the enzyme E. This is facilitated by various transfer RNAs which bind a particular triplet of nucleotides and its corresponding amino acid and sequentially build the protein in the molecular machine called a ribosome. This process is called translation. The enzyme E then catalyzes a reaction whereby precursor molecule X is converted to tryptophan. The cell has now synthesized the needed protein. But now things get interesting. Two molecules of tryptophan can form a complex with a prerepressor molecule forming a repressor molecule, so-called because it can bind to a site on the DNA, preventing RNA polymerase from binding there, and thus shutting down transcription of the gene and the formation of tryptophan. In this way, tryptophan controls its own synthesis. If tryptophan is readily available in the cells environment then there will be enough of it to bind with the prerepressor and shut down the cells own tryptophan synthesizing process.

*Department of Mathematics, Arizona State University, Tempe, AZ 85287, halsmith@asu.edu

However, if the external supply of tryptophan is abruptly shut off, the internal concentration will decline as the cell uses up its supply of tryptophan and eventually there will be too little bound with prerepressor forming repressor so RNA polymerase is free to bind to DNA and initiate transcription which leads to tryptophan synthesis. This is a very simple switch! Now lets say this mathematically.

We now give a set of chemical reactions describing the interaction of the above mentioned molecules. In formulating equations for the concentrations of these chemical species we must acknowledge that a single cell may not contain sufficiently many molecules of a particular chemical species to justify a continuum model. For example, a typical cell will have only one copy of the gene. Therefore, to justify our model we must focus not on a single cell but rather on a collection of synchronously dividing cells, which will contain large numbers of the relevant chemical species.

Transcription:-RNA Polymerase reads DNA copying the DNA instructions into an mRNA "transcript":

$$RNAP + DNA \rightleftarrows RNAP : DNA \rightarrow mRNA + DNA + RNAP$$

with rate constants $k1_+,\ k1_-,\ k1$. Subscript "+" indicates forward reaction and subscript "-" indicates the backward reaction in $\leftrightarrows$. RNAP:DNA denotes a complex formed when RNA Polymerase binds to DNA. It can either unbind into its constituents DNA and RNAP or proceed along the gene segment translating the DNA code to MRNA. We use this ":" notation to denote a complex hereafter.
Translation:-tRNA converts mRNA code into protein E:

$$tRNA + mRNA \rightleftarrows tRNA : mRNA \rightarrow E + mRNA + tRNA$$

with rate constants $k2_+,\ k2_-,\ k2$.
Enzyme catalyzes formation of Endproduct protein T (tryptophan) using precursor X:

$$X + E \rightleftarrows X : E \rightarrow T + E$$

with rate constants $k3_+,\ k3_-,\ k3$.
n molecules of Endproduct combine with a Prerepressor to form Repressor R:

$$P + nT \rightleftarrows R$$

with rate constants $k4_+,\ k4_-$. For tryptophan, $n = 2$.
Repressor binds to DNA making it unavailable for transcription to mRNA:

$$R + DNA \rightleftarrows R : DNA$$

with rate constants $k5_+,\ k5_-$.
Most molecules are degraded at some rate:

$$\begin{aligned} mRNA &\rightarrow \\ E &\rightarrow \\ T &\rightarrow \end{aligned}$$

with rate constants $d1,\ d2,\ d3$. Of course, they degrade to something but we will not need these. Another interpretation of these loss terms is to imagine an exponentially growing aggregate of cells in which case all chemical species suffer dilution as each cell roughly doubles in size before dividing. Thus degradation is merely dilution due to growth. Tryptophan T is also used in other cellular processes and $d3$ reflects this demand as well.
Finally, the exterior environment of the cell **may** provide tryptophan so we include this potential source

$$T_{ext} \rightarrow T$$

with rate constant $k6$. Of course, the whole point of the tryptophan gene is to be able to synthesize tryptophan when $T_{ext} = 0$.

2 Assumptions

Total DNA is constant-denoted by DNA_T.

$$DNA_T = DNA + RNAP : DNA + R : DNA \tag{2.1}$$

RNAPolymerase, tRNA, PreRepressor, Precursor concentrations are nearly constant

$$\frac{d}{dt}RNAP = \frac{d}{dt}tRNA = \frac{d}{dt}P = \frac{d}{dt}X = 0$$

Since P is constant, we have:

$$0 = \frac{d}{dt}P = k4_-(R) - k4_+(P)(T^n) \tag{2.2}$$

Complexes:

$$C1 = RNAP : DNA,\ C2 = tRNA : mRNA,\ C3 = X : E,\ C4 = R : DNA$$

are in equilibrium:

$$\frac{d}{dt}Ci = 0, \quad i = 1, 2, 3, 4$$

This means:

$$\begin{aligned} 0 &= k1_+(RNAP)(DNA) - (k1_- + k1)RNAP : DNA \\ 0 &= k2_+(tRNA)(mRNA) - (k2_- + k2)tRNA : mRNA \\ 0 &= k3_+(X)(E) - (k3_- + k3)X : E \\ 0 &= k5_+(R)(DNA) - (k5_-)R : DNA \end{aligned}$$

Using these last equations together with (2.1) and (2.2) we find:

$$DNA_T = (DNA)[1 + \frac{k1_+}{k1_- + k1}(RNAP) + \frac{k5_+}{k5_-}(R)]$$

so

$$DNA = \frac{DNA_T}{1 + K1 \cdot RNAP + K5K4 \cdot P \cdot T^n} \tag{2.3}$$

where

$$K1 = \frac{k1_+}{k1_- + k1}, \quad Kj = \frac{kj_+}{kj_-},\ j = 4, 5.$$

3 Equations

$mRNA$ is involved in both Transcription and Translation. From these two reactions we get:

$$\begin{aligned} \frac{d}{dt}mRNA &= k1(RNAP : DNA) - k2_+(tRNA)(mRNA) + (k2 + k2_-)tRNA : mRNA - d1 \cdot mRNA \\ &= k1K1(RNAP)(DNA) - d1 \cdot mRNA \\ &= k1K1(RNAP) \cdot \frac{DNA_T}{1 + K1 \cdot RNAP + K5K4 \cdot P \cdot T^n} - d1 \cdot mRNA \end{aligned}$$

The protein E is an output of translation and catalyzes formation of T. From these two reactions we have:

$$\begin{aligned} \frac{d}{dt}E &= (k3_- + k3)(X : E) - k3_+(X)(E) + k2(tRNA : mRNA) - d2 \cdot E \\ &= k2K2(tRNA)(mRNA) - d2 \cdot E \end{aligned}$$

where $K2$ is defined similarly as $K1$.

The endproduct T is used by the cell but it also combines with prerepressor to form repressor R which blocks further formation of T. T may be supplied by the exterior environment of the cell.

$$\begin{aligned}\frac{d}{dt}T &= k3(X:E) - d3(T) + nk4_-(R) - nk4_+(P)(T^n) + k6(T_{ext}) \\ &= k3K3(X)(E) - d3(T) + k6(T_{ext})\end{aligned}$$

Let's relabel $mRNA = m$ and tidy up the equations a bit. Recall that $RNAP, tRNA, X, P$ are constant so we may write:

$$\begin{aligned}\frac{d}{dt}m &= \frac{\beta}{\delta + \mu T^n} - d1 \cdot m \\ \frac{d}{dt}E &= \alpha_2 \cdot m - d2 \cdot E \\ \frac{d}{dt}T &= \alpha_3 \cdot E - d3 \cdot T + u\end{aligned}$$

Hopefully, the values of the newly introduced quantities are apparent. For example

$$\delta = 1 + K1 \cdot RNAP$$

Finally, it will be useful to scale out as many of the parameters as we can. Let

$$x_1 = m/m_0,\ x_2 = E/E_0,\ x_3 = T/T_0,\ \tau = t/t_0 \tag{3.1}$$

where m_0, E_0, T_0, τ_0 are reference values to be chosen. Check that these can be selected so as to achieve the following

$$\begin{aligned}x_1' &= \frac{1}{1 + x_3^n} - \gamma_1 x_1 \\ x_2' &= x_1 - \gamma_2 x_2 \\ x_3' &= x_2 - \gamma_3 x_3 + u\end{aligned} \tag{3.2}$$

where $' = \frac{d}{d\tau}$, $\gamma_i > 0$, $n = 1, 2, 3, \ldots$ is a positive integer and $u \geq 0$ is not the same as above. $u = 0$ corresponds to a tryptophan-free exterior environment. Equations (3.2) are sometimes referred to as the "Goodwin Oscillator" after their inventor Brian Goodwin. See [4].

4 Study of Equations (3.2)

Define $\bar{x}_3$ to be the unique positive root of

$$-\gamma_1\gamma_2 u + \gamma_1\gamma_2\gamma_3 x_3 = \frac{1}{1 + x_3^n}. \tag{4.1}$$

Figure 1 depicts that $\bar{x}_3$ is the abscissa of the unique point of intersection of the line $y = -\gamma_1\gamma_2 u + \gamma_1\gamma_2\gamma_3 x_3$ with the curve $y = \frac{1}{1+x_3^n}$. Then

$$\bar{x}_2 = \gamma_3 \bar{x}_3, \quad \bar{x}_1 = \gamma_2 \bar{x}_2$$

describes the unique steady state $\bar{x}$ of the system. Notice from Figure 1 that $\bar{x}_3 \to \infty$ as $u \to \infty$ and keep in mind that it depends on γ_i and on u.

Observe that the rate of mRNA production, a measure of the activity of the tryptophan gene, is given by $(1 + \bar{x}_3^n)^{-1}$ at steady state. It falls to zero rapidly as $\bar{x}_3$ increases and $\bar{x}_3$ increases with increasing u. Thus, an external source of tryptophan effectively shuts down gene activity.

The simulations below suggest the behavior of the system. In Figure 2 we examine how tryptophan may be produced by the cell when there is no external supply of tryptophan ($u = 0$). Note that first $mRNA$, then enzyme E, and finally tryptophan T rise and then approach the steady state $\bar{x}$ in an oscillatory fashion. In Figure 3, we start at $t = 0$ at this steady state but supply tryptophan $u = 2$ when $t > 0$ to the cell. The cell responds by immediately suppressing $mRNA$ and enzyme E production-in other words-the gene is turned off.

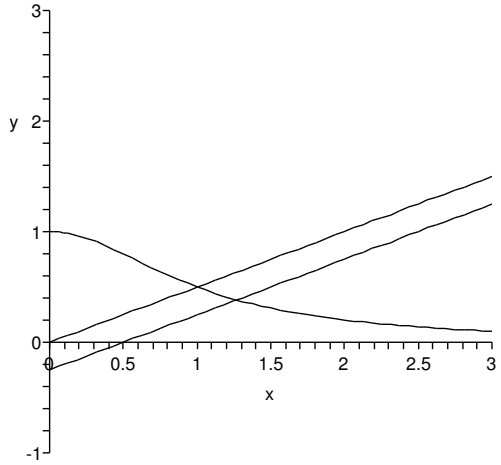

Figure 1: The x-component of the point of intersection of the two curves determines $\bar{x}_3$ with $n = 2$. The case $u = 0$ gives the top line and $u > 0$ the lower one.

The stability of $\bar{x}$ is determined by the linearized system

$$z' = \begin{bmatrix} -\gamma_1 & 0 & -q \\ 1 & -\gamma_2 & 0 \\ 0 & 1 & -\gamma_3 \end{bmatrix} z. \tag{4.2}$$

where

$$q = n\bar{x}_3^{n-1}(1 + \bar{x}_3^n)^{-2} \tag{4.3}$$

The characteristic polynomial is

$$(\lambda + \gamma_1)(\lambda + \gamma_2)(\lambda + \gamma_3) + q = 0$$

or equivalently

$$\lambda^3 + (\gamma_1 + \gamma_2 + \gamma_3)\lambda^2 + (\gamma_1\gamma_2 + \gamma_1\gamma_3 + \gamma_2\gamma_3)\lambda + \gamma_3\gamma_2\gamma_2 + q = 0$$

In order to simplify the algebra, we hereafter assume that

$$\gamma = \gamma_1 = \gamma_2 = \gamma_3.$$

Then q depends on n, γ and u.

Show that the roots are given by

$$\lambda = -\gamma - q^{1/3},\ \lambda = -\gamma + q^{1/3}[\cos(\pi/3) \pm i\sin(\pi/3)] = -\gamma + q^{1/3}[1/2 \pm i\sqrt{(}3)/2] \tag{4.4}$$

Therefore $\Re(\lambda) < 0$ for all 3 roots if and only if

$$q^{1/3}/\gamma < 2 \tag{4.5}$$

and $\bar{x}$ is asymptotically stable in this case. If

$$q^{1/3}/\gamma = 2$$

there is one negative roots and two imaginary roots $\lambda = \pm i\triangle$ and if

$$q^{1/3}/\gamma > 2$$

there is one negative root and a complex conjugate pair of roots with positive real part. In this case, $\bar{x}$ is unstable. But remember that q depends on n, γ, u.

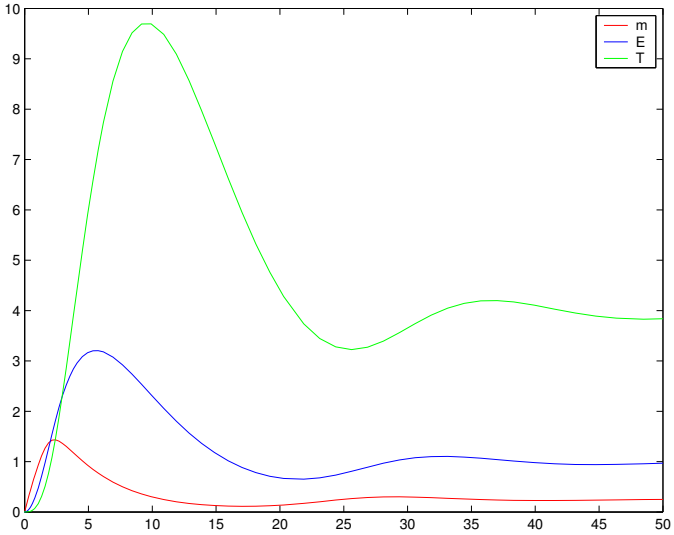

Figure 2: The time course of (3.2) with $u = 0,\ n = 2, \gamma_i = 0.25$ and $x_i = 0$ at $t = 0$.

Consider the case that $u = 0$. then use (4.1) and (4.3) to show that

$$q = n\gamma^6 x_3^{n+1} = n\gamma^3(1 - \gamma^3 x_3)$$

and $\gamma^3 x_3 < 1$ by (4.1) so

$$q^{1/3}/\gamma = n^{1/3}(1 - \gamma^3 x_3)^{1/3} < n^{1/3}$$

implying that (4.5) holds if $n < 8$.

Proposition 1. *If $u = 0$, $n < 8$ and $\gamma_i = \gamma > 0,\ i = 1, 2, 3$, then $\bar{x}$ is asymptotically stable.*

5 Discussion

There is an enormous literature on the Goodwin oscillator and related equations. The references below contain references to much of this literature.

The Goodwin model can be applied to many other gene networks with positive and negative feedback. These systems often have more than one precursor molecule and the system has higher dimension. Here is a more general form.

$$\begin{aligned} x_1' &= f(x_p) - \gamma_1 x_1 \\ x_i' &= x_{i-1} - \gamma_i x_i,\ 2 \le i \le p \end{aligned}$$

This system has positive feedback if $f' > 0$ and negative feedback when $f' < 0$.

In his famous book [5], Murray uses the Goodwin model to model testosterone production in mammals.

The secant method [10, 9] was developed to determine the stability characteristics of matrices of the form

$$\begin{bmatrix} -\alpha_1 & 0 & \cdots & 0 & -\beta_1 \\ \beta_2 & -\alpha_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \beta_p & -\alpha_p \end{bmatrix} \tag{5.1}$$

where $\alpha_i,\ \beta_i > 0$, that arise in determining the stability of steady states of the preceding equation. A sufficient condition that all eigenvalues have negative real parts is given by:

$$\frac{\beta_1\beta_2\cdots\beta_p}{\alpha_1\alpha_2\cdots\alpha_p} < [\sec(\frac{\pi}{p})]^p.$$

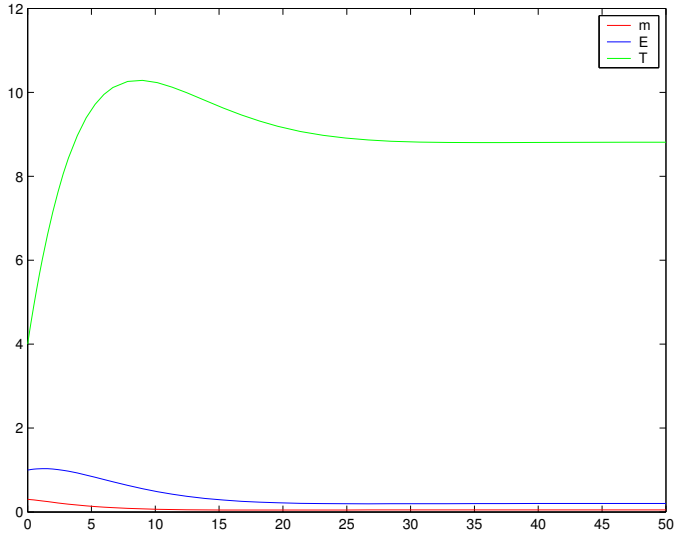

Figure 3: The time course of (3.2) with $u = 2,\ n = 2, \gamma_i = 0.25$ and x_i set to their steady state values when $u = 0$ at $t = 0$.

The article [6] shows that despite these systems having dimension higher than two, the Poincaré-Bendixson Theorem holds for them. As a consequence, it is shown that when the steady state has a pair of complex conjugate eigenvalues with positive real part, then there exists at least one stable periodic solution.

We have assumed that all the reactions described above occur instantaneously yet DNA transcription takes time, protein assembly in the ribosome also takes time. Therefore, it makes sense to include time delays in the equations leading to the system

$$\begin{aligned} x_1'(t) &= \frac{1}{1 + x_3(t - \tau_1)^n} - \gamma_1 x_1(t) \\ x_2'(t) &= x_1(t - \tau_2) - \gamma_2 x_2(t) \\ x_3'(t) &= x_2(t - \tau_3) - \gamma_3 x_3(t) + u(t) \end{aligned}$$

Time delays do not affect the steady state but do affect the stability of that steady state. See the references for more on models with time delays. In general, the external supply of tryptophan will be time dependent so we have included this by taking u time-dependent.

See Santillan and Mackey [7] for a more realistic model of the trp operon.

See the web site http://www.che.eng.ohio-state.edu/~FEINBERG/RESEARCH/ for basic lectures on the differential equations of chemical reactions.

6 Homework Problems

1. Verify that the reference values in (3.1) may be chosen to obtain (3.2).

2. Verify (4.4).

3. Verify the computations leading to Proposition 1.

4. If we also assume that $\frac{d}{dt}E = 0$, then (3.2) simplifies to two equations. Analyze the phase plane and determine the asymptotic behavior of this system.

5. Show that $\bar{x}$ is asymptotically stable for (3.2) if $n = 1$.

6. Show that there is a family of "Rectangles" of the form $R(b) := \{x : 0 \le x_i \le b_i,\ 1 \le i \le 3\}$, where $b_i > 0$, that are positively invariant for (3.2).

References

[1] Alberts et al, The Molecular Biology of the Cell, 4th ed., Taylor and Francis Group, 2002.

[2] H.T. Banks and J.M. Mahaffy, Mathematical Models for protein biosynthesis, Lefschetz Center for Dynamical Systems, Brown University, TR 79-4, 1979.

[3] H.T. Banks and J.M. Mahaffy, Global asymptotic stability of certain models for protein sythesis and repression, Quarterly of Applied Math. XXXVI (1978), 209-221.

[4] C. Fall, E. Marland, J. Wagner, J. Tyson, Computational Cell Biology, Springer 2002.

[5] J.D. Murray, Mathematical Biology, Springer 1989.

[6] J.Mallet-Paret and H.L. Smith, The Poincare-Bendixson Theorem for monotone cyclic feedback systems, J. Dynamics & Diff. Eqns.2,1990, 367-421.

[7] M. Santillan and M. Mackey, Dynamic regulation of the tryptophan operon: A modeling study and comparison with experimental data, Proc. Nat. Acad. Sciences 98, no. 4, 2001, 1364-1369.

[8] H.L. Smith, Oscillations and multiple steady states in a cyclic gene model with repression, J. Math. Biol. 25 (1987), 169-190.

[9] C.D. Thron, The secant condition for instability in biochemical feedback control, Bull. Math. Biol. 53(1991) 383-401.

[10] J. Tyson and H. Othmer, The dynamics of feedback control circuits in biochemical pathways, in Progress in Theoretical Biology (Rosen & Snell,eds.) Vol. 5, p. 1-62, Academic Press, New York 1978.

99