Bulletin of Mathematical Biology (2001) **63**, 329–351
doi:10.1006/bulm.2000.0225
Available online at http://www.idealibrary.com on IDEAL®

A Mathematical Model for Prokaryotic Protein Synthesis

DONALD A. DREW
Department of Mathematical Sciences,
Rensselaer Polytechnic Institute,
Troy, N.Y.,
U.S.A.

A kinetic model for the synthesis of proteins in prokaryotes is presented and analysed. This model is based on a Markov model for the state of the DNA strand encoding the protein. The states that the DNA strand can occupy are: ready, repressed, or having a mRNA chain of length i in the process of being completed. The case $i = 0$ corresponds to the RNA polymerase attached, but no nucleotides attached to the chain. The Markov model consists of differential equations for the rates of change of the probabilities. The rate of production of the mRNA molecules is equal to the probability that the chain is assembled to the penultimate nucleotide, times the rate at which that nucleotide is attached. Similarly, the mRNA molecules can also be in different states, including: ready and having an amino acid chain of length j attached. The rate of protein synthesis is the rate at which the chain is completed. A Michaelis–Menten type of analysis is done, assuming that the rate of protein degradation determines the 'slow' time, and that all the other kinetic rates are 'fast'. In the self-regulated case, this results in a single ordinary differential equation for the protein concentration.

© *2001 Society for Mathematical Biology*

1. INTRODUCTION

The sequences of nucleotides for the genomes for many species are becoming available, and the push to understand and use this information has begun. A cell with a full complement of its DNA needs only favorable surroundings (including nourishment and temperature) to accomplish its biological purposes. All other processes and actions are 'programmed' in the genome. The timely deployment of the protein/enzyme machinery of the cell results from the genetic coding in the DNA.

The process by which this happens in prokaryotes is thought to be through transcription factors (TF), which are proteins or protein macromolecules that bind to transcription factor binding sites (BS) upstream of the gene, or DNA subsequence, that codes for the expression of a given protein.

Messenger RNA is synthesized from the template on DNA. This action is accomplished by the binding of a molecule of RNA polymerase to the receptor site on the DNA upstream of the code for the protein to be expressed. On or near the receptor site, there are sites where protein molecules called transcription factors

0092-8240/01/020329 + 23 $35.00/0 © 2001 Society for Mathematical Biology

can attach and either inhibit the attachment of the RNA polymerase or enhance its transcription. Thus, if the repressor molecule is attached, transcription cannot occur. If the repressor molecule is attached, there is some thermal energy of the molecule and the DNA strand, and this energy can cause the bonds between the repressor and the DNA strand to break. Similarly, for some sites, transcription can normally not occur unless a molecule called an activator attaches to an appropriate transcription factor binding site. The activator molecule enables the energetic processes necessary for the polymerase to start transcription.

The messenger RNA is the template for protein synthesis. A ribosomal body is assembled on the mRNA and is the machine that assembles the peptide chain, which subsequently folds into the useful protein. The life of a protein in prokaryotes is determined by its degradation by a protease.

Proteins act as repressors for DNA transcription. Lodish *et al*. (2000) mentions that most of the known repressors in prokaryotes are dimers. We shall assume that two protein molecules react to form a repressor molecule. Activators are also assumed to be proteins, however, in the absence of further evidence, we assume that a monomeric protein can activate transcription.

The mathematical/biological modeling of protein synthesis dates from the (Jacob and Monod, 1961) model for induction, which is cited in the literature for the lac model. One of the first works to take a general mathematical approach to protein synthesis is that of Goodwin (1963), who derives kinetic equations for protein concentrations. Brendel and Perelson (1993) derived a model for a repressor–DNA–RNA system for colE1 plasmid copy number.

The lac model is exemplary because it is quite well understood. In this model, a repressor molecule essentially stops the transcription of the lac operon by attaching to the binding site, and sterically preventing the attachment of the polymerase. The repressor is removed from the DNA by an activator molecule, allolactose, which causes a conformational change in the repressor. An enhancer, CAP + cAMP, which presumably eases the transfer of energy to drive the process, may bind near the binding site, and increase the transcription rate many-fold. The models for this process (Mahaffy, 1999) account for the rates of production of mRNA and the proteins, as well as the effects of other molecules (specifically, the activator) that affect the transcription and translation rates.

Another protein reaction system that has received much attention is the cell cycle. This system is fundamental to all cells, and is apparently quite different in eukaryotes and prokaryotes. In eukaryotes, the timing is accomplished a phosphorous–cyclin–CDK cycle (Novak and Tyson, 1997). In this system, the controls are not thought to be directly related to transcription regulation, but instead are mainly due to chemical controls in the cytoplasm.

Experimental evidence that the basic mechanisms of DNA repression and activation are correct consist of creating biological switches (Gardner *et al*., 2000) and oscillators (Elowitz and Liebler, 2000) in plasmids. Both of these works compare data with a mathematical model for protein synthesis based on systems of Hill's

equations for two (for a switch) or three (for an oscillator) proteins and their corresponding mRNAs. These examples serve to motivate a model for protein synthesis that is connected to the genes and the corresponding regulation systems.

Thieffry (1999) proposes an attack on the problem of the relation of the genome to protein expression that attempts to recognize the patterns in the proteins and relate them to the structure of the binding sites, for the purpose of identifying the repressors and activators, and the proteins that they regulate. With this information, and with equations for the rates of expression of the proteins based on the rates of transcription and translation, the temporal organization of the cell, as shown in the levels of proteins present at any time, can, in theory, be predicted.

In this paper, I shall derive a model for the expression of a protein governed by a repressor. I shall consider the probability (actually, expected value) $\mathcal{P}(\circ)$ of various states of the DNA strand that codes for the protein. The process of chain elongation is treated as if the number of bases is a continuous variable. This introduces a time delay in the equations. The mRNA chains are treated similarly, with ribosome attachment and chain growth again included in the model. A model for protein degradation is included. Assuming that the repressor is a dimer of the protein considered leads to an equation for the rate of production of a self-repressed protein. The case of co-repressors is also considered. Finally, a model for activation is derived.

2. mRNA Synthesis: Repressor Dynamics

We use a Markov model to describe the dynamics of the RNA synthesis. This model assumes that the particular gene on the DNA strand is in one of several states, and makes transitions between the states when a chemical reaction occurs. This is appropriate because the DNA site is unique in any one cell, and is either available for RNA polymerase attachment or it is not, we deal with the probabilities of events. The reactions can be written symbolically as

$$\begin{aligned} DNA + P &\to DNA_0 \\ DNA + R &\leftrightarrows DNA_- \\ DNA_0 + n_1 &\to DNA_1 \\ DNA_{i-1} + n_i &\to DNA_i \\ &\vdots \end{aligned}$$

Here DNA denotes the 'ready' DNA strand, P denotes the RNA polymerase, DNA_0 denotes the DNA–polymerase complex, DNA_- denotes the DNA with the repressor molecule R attached. DNA_i denotes the DNA–polymerase com-

plex with a nascent mRNA chain of length i attached, and n_i is the nucleotide corresponding to site i.

The mRNA is considered to be finished when the last nucleotide is attached:

$$DNA_{M-1} + n_M \rightarrow mRNA + P + DNA$$

where M is the length of the site in base pairs.

2.1. ***Rate equations.*** The probability (actually, expected value) that the DNA is available is denoted by $\mathcal{P}(DNA)$, and the rate of change of the probability can be determined by the rate at which a reaction occurs between the site and the RNA polymerase. If Δt is a small time interval, the probability that the site is available at time $t + \Delta t$ is equal to the probability that the site was available at time t, times the probability that no polymerase attaches in time Δt, plus the probability that a polymerase complex finishes, releases its mRNA, polymerase, and the DNA. Thus, assuming that the process is Markovian, we have

$$\begin{aligned}\mathcal{P}(DNA)(t+\Delta t) = {} & \mathcal{P}(DNA)(t)(1 - k_p[P](t)\Delta t)(1 - k_R[R](t)\Delta t) \\ & + \mathcal{P}(DNA_{N-1})(t)k_N[n_N](t)\Delta t + \mathcal{P}(DNA_-)(t)k_-\Delta t\end{aligned}$$

where k_p is the rate constant for the polymerase attachment reaction, $[P]$ is the concentration of the polymerase, k_R is the rate constant for the repressor attachment reaction, $[R]$ is the repressor concentration, $\mathcal{P}(DNA_i)$ denotes the probability that the polymerase has transcribed to site i on the DNA, $\mathcal{P}(DNA_-)$ denotes the probability that the binding site is repressed, k_- is the rate at which the repressor molecule is released from the binding site without the activator, $[n_i]$ is the concentration of nucleotide i, and k_i is the rate constant for attachment of nucleotide i. If we manipulate the equation, and let $\Delta t \rightarrow 0$, we have

$$\frac{d\mathcal{P}(DNA)}{dt} = -(k_p[P]+k_R[R])\mathcal{P}(DNA)+k_N[n_N]\mathcal{P}[DNA_{N-1}]+k_-\mathcal{P}(DNA_-). \tag{1}$$

A similar calculation for $\mathcal{P}(DNA_0)(t)$, the probability that the RNA polymerase is bound to the DNA strand to form the complex leads to

$$\frac{d\mathcal{P}(DNA_0)}{dt} = -k_1[n_1]\mathcal{P}(DNA_0) + k_p[P]\mathcal{P}(DNA) \tag{2}$$

and for $\mathcal{P}(DNA_-)$, the repressed strand probability

$$\frac{d\mathcal{P}(DNA_-)}{dt} = -k_-\mathcal{P}(DNA_-) + k_r[R]\mathcal{P}(DNA). \tag{3}$$

Once the DNA–polymerase starts making the mRNA chain, it proceeds according to

$$\frac{d\mathcal{P}(DNA_i)}{dt} = -k_{i+1}[n_{i+1}]\mathcal{P}(DNA_i) + k_i[n_i]\mathcal{P}(DNA_{i-1}). \tag{4}$$

Note that if we sum equations (1)–(4), we have

$$\frac{d}{dt}\left[\mathcal{P}(DNA) + \mathcal{P}(DNA^-) + \mathcal{P}(DNA_0) + \sum_{i=1}^{M-1}\mathcal{P}(DNA_i)\right] = 0$$

so that if the probabilities sum to one at the initial instant, the sum will remain one for the duration of the process.

The rate of production of mRNA is given by

$$\left.\frac{d[mRNA]}{dt}\right|_{\text{production}} = k_M[n_M]\mathcal{P}(DNA_{M-1}).$$

2.2. ***Continuous mRNA chain elongation model.*** For simplicity, suppose the abundancies and the reaction rates of the nucleotides are the same, so that $k_i[n_i] = k_j[n_j] = k[n]$. Then, we approximate

$$-k_{i+1}[n_{i+1}]\mathcal{P}(DNA_i) + k_i[n_i]\mathcal{P}(DNA_{i-1}) \cong -k[n]\frac{\partial\mathcal{P}(DNA_i)}{\partial i}$$

and the equation for elongation becomes

$$\frac{\partial\mathcal{P}(DNA_i)}{\partial t} = -k[n]\frac{\partial\mathcal{P}(DNA_i)}{\partial i}. \tag{5}$$

2.2.1. *Solution.* The equations for the characteristics for the partial differential equation (5) are given by

$$\frac{d\mathcal{P}(DNA_i)}{d\tau} = 0$$
$$\frac{dt}{d\tau} = 1$$
$$\frac{di}{d\tau} = k[n].$$

The solution is given by

$$\mathcal{P}(DNA_i)(t) = \mathcal{P}(DNA_0)(t - T_i)$$

where T_i is the time taken to make a nascent mRNA chain of length i, and is given by

$$i = \int_{t-T_i}^{t} k[n](t)dt.$$

If the concentrations of the nucleotides remains constant, we have

$$T_i = \frac{i}{k[n]}.$$

2.3. ***The reduced model.*** Substituting the elongation solution into the equations for the complex gives

$$\frac{d\mathcal{P}(DNA)(t)}{dt} = -[k_p[P] + k_R[R](t)]\mathcal{P}(DNA)(t) + k_M[n_M]\mathcal{P}[DNA_0](t - T_M) + k_-\mathcal{P}(DNA_-)(t) \tag{6}$$

$$\frac{d\mathcal{P}(DNA_0)}{dt} = -k_1[n_1]\mathcal{P}(DNA_0) + k_p[P]\mathcal{P}(DNA) \tag{7}$$

$$\frac{d\mathcal{P}(DNA_-)}{dt} = -k_-\mathcal{P}(DNA_-) + k_R[R]\mathcal{P}(DNA_0). \tag{8}$$

The equation for the sum of the probabilities becomes

$$\mathcal{P}(DNA) + \mathcal{P}(DNA_-) + \mathcal{P}(DNA_0) + \sum_{i=1}^{M-1} \mathcal{P}(DNA_0)(t - T_i) = 1$$

or, using the continuous elongation model,

$$\mathcal{P}(DNA) + \mathcal{P}(DNA_-) + k[n] \int_{T_i=0}^{T_M} \mathcal{P}(DNA_0)(t - T_i)\, dT_i = 1. \tag{9}$$

3. Protein Synthesis

The concentration of 'ready' mRNA is $[mRNA]$, which changes due to combination with the ribosome parts to form the RNA–ribosome complex. We denote this reaction symbolically by

$$mRNA + R_L + R_S \to mRNA_0$$

where R_L and R_S denote the large and small parts of the ribosome, and $mRNA_0$ denotes the messenger RNA with the ribosome attached. Next, the RNA–ribosome complex can attach amino acids to the chain. Thus,

$$mRNA_0 + a_1 \to mRNA_1$$

$$mRNA_{i-1} + a_i \to mRNA_i.$$

The reaction ends when the last amino acid is added to the chain

$$mRNA_{M-1} + a_M \rightarrow \Pi + mRNA$$

where Π denotes the polypeptide chain. If we note that the rate of production of mRNA chains in a single cell is equal to the *average* rate at which the last nucleotide is attached, that is,

$$k_M[a_M]\mathcal{P}[DNA_{M-1}](t) = k_M[a_M]\mathcal{P}[DNA_0](t - T_M)$$

then the rate equations for these reactions with a variable volume are (Brendel and Perelson, 1993)

$$\begin{aligned}\frac{dV[mRNA]}{dt} &= k_M[a_M]\mathcal{P}[DNA_0](t - T_M) + V\kappa_N[a_N][mRNA_{N-1}] \\ &\quad -V\kappa_R[mRNA][R_L][R_S] - V\kappa_-[mRNA]\end{aligned}$$

where V is the volume of the cell, κ_j is the rate at which the jth amino acid in the chain reacts with the chain, P_j is the concentration of partially polymerized chains of length j amino acids, κ_R is the rate at which the ribosome body attaches to the messenger RNA, and κ_- is the rate at which mRNA degrades. Here N is the length of the polypeptide chain, which is $N = M/3$. We shall assume that the cell volume V is constant.

Thus, the equations become

$$\begin{aligned}\frac{d[mRNA]}{dt} &= \rho k_M[n_M]\mathcal{P}[DNA_0](t - T_N) + \kappa_N[a_N][mRNA_{N-1}] \\ &\quad -\kappa_R[mRNA][R_L][R_S] - \kappa_-[mRNA] \\ \frac{d[mRNA_0]}{dt} &= \kappa_R[mRNA][R_L][R_S] - \kappa_1[a_1][mRNA_0] \\ \frac{d[mRNA_j]}{dt} &= \kappa_j[a_j][mRNA_{j-1}] - \kappa_{j+1}[a_{j+1}][mRNA_j]\end{aligned}$$

where $\rho = \frac{1}{V}$ is the concentration of one mRNA per cell volume. The rate of protein synthesis is then given by the rate at which the last amino acid is attached to the chain. Thus,

$$\left.\frac{d[\Pi]}{dt}\right|_{\text{synthesis}} = \kappa_N[a_N][mRNA_{N-1}].$$

3.1. ***Continuous peptide chain elongation model.*** Again, we assume that $\kappa_i[a_i] = \kappa_j[a_j] = \kappa[a]$ for all i, j. Further, we approximate

$$\kappa_j[a_j][mRNA_{j-1}] - \kappa_{j+1}[a_{j+1}][mRNA_j] \cong -\kappa[a]\frac{\partial[mRNA_j]}{\partial j}$$

so that the chain elongation model becomes

$$\frac{\partial[mRNA_j]}{\partial t} = -\kappa[a]\frac{\partial[mRNA_j]}{\partial j}.$$

Again, the solution is given by

$$[mRNA_j](t) = [mRNA_0](t - S_j)$$

where S_j is the time to make a chain of length j, and is given by

$$j = \int_{t-S_j}^{t} \kappa[a](t)dt.$$

If the concentrations of the amino acids remain constant, we have

$$S_j = \frac{j}{\kappa[a]}.$$

Thus, the mRNA subsystem becomes

$$\begin{aligned}
\frac{d[mRNA]}{dt} &= \rho k[n]\mathcal{P}[DNA_0](t - T_M) + \kappa[a][mRNA_0](t - S_N) \\
&\quad -\kappa_R[mRNA][R_L][R_S] - \kappa_-[mRNA] \qquad (10)\\
\frac{d[mRNA_0]}{dt} &= \kappa_R[mRNA][R_L][R_S] - \kappa[a][mRNA_0] \qquad (11)\\
\left.\frac{d[\Pi]}{dt}\right|_{\text{synthesis}} &= \kappa[a][mRNA_0](t - S_N).
\end{aligned}$$

3.2. ***Michaelis–Menten approximation.*** The reaction equations are

$$\begin{aligned}
\frac{d\mathcal{P}(DNA)}{dt}(t) &= -(k_R[R] + k_p[P])\mathcal{P}(DNA)(t) \\
&\quad +k[n]\mathcal{P}[DNA_0](t - T_M) + k_-\mathcal{P}(DNA_-)(t) \\
\frac{d\mathcal{P}(DNA_0)}{dt} &= -k[n]\mathcal{P}(DNA_0) + k_p[P]\mathcal{P}(DNA)
\end{aligned}$$

$$\frac{d\mathcal{P}(DNA^{-})}{dt} = k_r[R]\mathcal{P}(DNA) - k_{-}\mathcal{P}(DNA_{-})$$

$$\frac{d[mRNA]}{dt} = \rho k[n]\mathcal{P}[DNA_0](t - T_M) + \kappa[a][mRNA_0](t - S_N)$$
$$-(\kappa_R[R_L][R_S] + \kappa_{-})[mRNA]$$

$$\frac{d[mRNA_0]}{dt} = \kappa_R[mRNA][R_L][R_S] - \kappa[a][mRNA_0].$$

If we assume that the reaction rates for most of the reactions are fast, we can replace those differential equations by their equilibrium versions. In addition, we shall assume that the time delays for transcription and translation are small. We shall retain the derivative term in the protein dynamics equation. This procedure is justified using dimensional analysis in Appendix B. We have

$$\mathcal{P}(DNA_0) = \frac{k_p[P]}{k[n]}\mathcal{P}(DNA) \tag{12}$$

$$\mathcal{P}(DNA_{-}) = \frac{k_R[R]}{k_{-}}\mathcal{P}(DNA) \tag{13}$$

$$\mathcal{P}(DNA) = \left(1 + \frac{k_R[R]}{k_{-}} + \frac{k_p[P]}{k[n]}M\right)^{-1} \tag{14}$$

$$\rho k[n]\mathcal{P}[DNA_0] = \kappa_{-}[mRNA] \tag{15}$$

$$\kappa_R[R_L][R_S][mRNA] = \kappa[a][mRNA_0] \tag{16}$$

$$\left.\frac{d[\Pi]}{dt}\right|_{\text{synthesis}} = \kappa[a][mRNA_0]$$

$$= \frac{\rho k_p[P]\kappa_R[R_L][R_S]}{\kappa_{-}\left(1 + \frac{k_R[R]}{k_{-}} + \frac{k_p[P]}{k[n]}M\right)}.$$

4. PROTEIN DEGRADATION

In prokaryotes, protein degrades by the action of enzymes that break the oligomeric bonds, and the peptidal chains (Gottesman, 1996). These reactions are symbolized by

$$\Pi + E \longrightarrow \sum_{i=1}^{N} a_i$$

where E is the enzyme concentration. The relevant rate equation is

$$\left.\frac{d[\Pi]}{dt}\right|_{\text{degradation}} = -k_E[E][\Pi].$$

5. REPRESSOR MOLECULES

Lodish claims that repressor molecules for prokaryotes are almost always dimers. This reaction can be written as

$$[\Pi] + [\Pi] \rightleftarrows [R].$$

The repressor rate equation is

$$\frac{d[R]}{dt} = \mathcal{K}[\Pi]^2 - \mathcal{K}_-[R] - \rho k_R[R]\mathcal{P}(DNA) + \rho k_-\mathcal{P}(DNA_-)$$

and the rate of change of protein concentration due to dimerization is given by

$$\left.\frac{d[\Pi]}{dt}\right|_{\text{dimerization}} = \mathcal{K}_-[R] - \mathcal{K}[\Pi]^2.$$

5.1. ***Autorepressed model.*** Here we assume that the repressor molecule is a protein dimer, where each of the monomer units is a protein made by the above process. Using the other equations, the equation for the rate of change of the protein concentration can be written as

$$\frac{d[\Pi]}{dt} = \frac{\kappa_R \rho k_P[P][R_L][R_S]}{\kappa_-}\mathcal{P}[DNA] + \mathcal{K}_-[R] - \mathcal{K}[\Pi]^2 - k_E[E][\Pi]$$

or

$$\frac{d[\Pi]}{dt} = \frac{\kappa_R \rho k_P[P][R_L][R_S]}{\kappa_-\left(1 + \frac{k_R[R]}{k_-} + \frac{k_p[P]}{k[n]}M\right)} + \mathcal{K}_-[R] - \mathcal{K}[\Pi]^2 - k_E[E][\Pi]$$

$$\frac{d[R]}{dt} = \mathcal{K}[\Pi]^2 - \mathcal{K}_-[R] - \rho k_R[R]\mathcal{P}(DNA) + \rho k_-\mathcal{P}(DNA_-)$$

$$\frac{d[R]}{dt} = \mathcal{K}[\Pi]^2 - \mathcal{K}_-[R].$$

If we assume that the dimerization reaction is fast compared to the rate of protein degradation, we have

$$0 = \mathcal{K}[\Pi]^2 - \mathcal{K}_-[R]$$

and, consequently, the protein production equation is

$$\frac{d[\Pi]}{dt} = \frac{\kappa_R \rho k_P[P][R_L][R_S]}{\kappa_-\left(1 + \frac{k_R\mathcal{K}[\Pi]^2}{k_-\mathcal{K}_-} + \frac{k_p[P]}{k[n]}M\right)} - k_E[E][\Pi].$$

The advantage of this equation over the postulation of a Michaelis–Menten equation for the rate of change of protein concentration is that this equation relates the

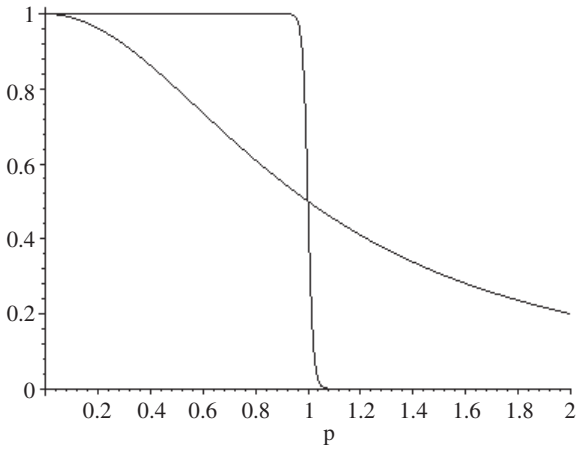

Figure 1. Rate of expression of a protein as a function of the concentration of its repressor.

parameters in the protein rate equation to more fundamental rates, such as the rates of chain elongation, protein and mRNA decay, and degradation. The equation can be written as

$$\frac{d[\Pi]}{dt} = \frac{\lambda}{(1+\nu[\Pi]^2)} - k_d[\Pi]$$

where

$$k_d = k_E[E]$$
$$\lambda = \frac{\rho k_P[P]\kappa_R[R_L][R_S]}{\kappa_-\left(1+\frac{k_p[P]}{k[n]}M\right)}$$
$$\nu = \frac{k_R\mathcal{K}}{k_-\mathcal{K}_-\left(1+\frac{k_p[P]}{k[n]}M\right)}.$$

The form of the synthesis term is shown in Fig. 1 for $\lambda = 1$, $\nu = 1$. Also shown is the step function $H(1-x)$, where H is the Heaviside function. This latter function is of interest because it represents a 'switch' in the form suggested by McAdams and Arkin (1997).

5.2. ***Coupled repressors.*** If we assume that two proteins are made, each repressing the other, we have a coupled system of differential equations. If we again make the assumption that the rates are fast, except for the protein production and degradation rates. Further, we assume that the common processes have the same rate for both reactions. Thus, the rate of attachment of the polymerase, the rate of chain elongation, the rate of attachment of the ribosome, and the rate of degradation are

common to both reactions. The rates of attachment and detachment of the repressors, the rates of degradation of the mRNA, and the rate of enzyme attachment to the proteins are assumed to be different.

$$\mathcal{P}[DNA_0^{(1)}] = \frac{k_p[P]}{k[n]}\mathcal{P}[DNA^{(1)}]$$

$$\mathcal{P}[DNA_-^{(1)}] = \frac{k_R^{(21)}[R^{(2)}]}{k_-^{(1)}}\mathcal{P}[DNA^{(1)}]$$

$$\mathcal{P}[DNA^{(1)}] = \left(1 + \frac{k_R^{(21)}[R^{(2)}]}{k_-^{(1)}} + \frac{k_p[P]}{k[n]}M^{(1)}\right)^{-1}$$

$$\rho k[n]\mathcal{P}[DNA_0^{(1)}] = \kappa_-^{(1)}[mRNA^{(1)}]$$

and

$$\mathcal{P}[DNA_0^{(2)}] = \frac{k_p[P]}{k[n]}\mathcal{P}[DNA^{(2)}]$$

$$\mathcal{P}[DNA_-^{(2)}] = \frac{k_R^{(12)}[R^{(1)}]}{k_-^{(2)}}\mathcal{P}[DNA^{(2)}]$$

$$\mathcal{P}[DNA^{(2)}] = \left(1 + \frac{k_R^{(12)}[R^{(1)}]}{k_-^{(2)}} + \frac{k_p[P]}{k[n]}M^{(2)}\right)^{-1}$$

$$0 = \rho k[n]\mathcal{P}[DNA_0^{(2)}] - \kappa_-^{(2)}[mRNA^{(2)}]$$

$$\frac{d[\Pi^{(1)}]}{dt} = \kappa_R[mRNA^{(1)}][R_L][R_S] - k_E^{(1)}[E^{(1)}][\Pi^{(1)}]$$

$$\mathcal{K}^{(1)}[\Pi^{(1)}]^2 = \mathcal{K}_-^{(1)}[R^{(1)}]$$

$$\frac{d[\Pi^{(2)}]}{dt} = \kappa_R[mRNA^{(2)}][R_L][R_S] - k_E^{(2)}[E^{(2)}][\Pi^{(2)}]$$

$$\mathcal{K}^{(2)}[\Pi^{(2)}]^2 = \mathcal{K}_-^{(2)}[R^{(2)}].$$

The two rate equations can be written as

$$\frac{d[\Pi^{(1)}]}{dt} = \frac{\lambda_1}{(1 + \nu_{12}[\Pi^{(2)}]^2)} - \kappa_d^{(1)}[\Pi^{(1)}]$$

$$\frac{d[\Pi^{(2)}]}{dt} = \frac{\lambda_2}{(1 + \nu_{21}[\Pi^{(1)}]^2)} - \kappa_d^{(2)}[\Pi^{(2)}]$$

where

$$\lambda_i = \frac{\kappa_R \rho k_p[P][R_L][R_S]}{\kappa_-^{(i)}\left(1 + \frac{k_p[P]}{k[n]}M^{(i)}\right)}$$

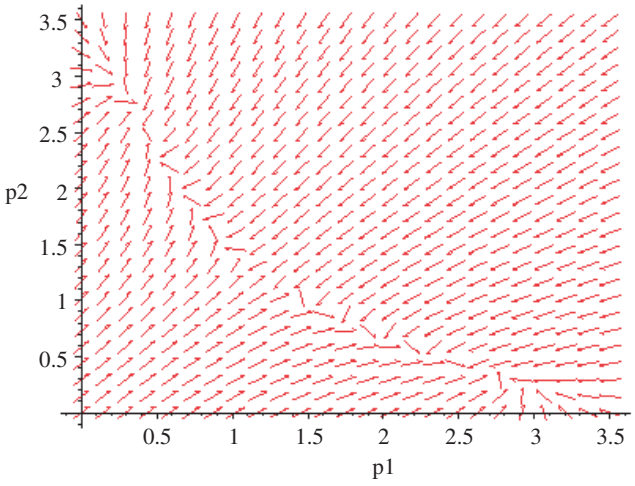

Figure 2. Field plot for the Hill's system for co-repressed proteins.

$$\nu_{ij} = \frac{k_R^{(ji)} \mathcal{K}_-^{(j)}}{k_-^{(i)} \mathcal{K}_-^{(j)} \left(1 + \frac{k_p[P]}{k[n]} M^{(i)}\right)}$$

$$\kappa_d^{(i)} = k_E^{(i)} [E^{(i)}].$$

The direction field for the coupled repressor system (Hill's system) is shown in Fig. 2. The parameter values for this figure are $\lambda_1 = \lambda_2 = 3$, $\nu_1 = \nu_2 = 1$, and $\kappa_d^{(1)} = \kappa_d^{(2)} = 1$. Note that there are three steady state solutions, two of which are stable. The separatrix is the line $y = x$.

5.3. ***An on/off approximate system.*** Although the approximation shown in Fig. 1 is qualitative at best, we examine a system corresponding to an 'on/off' model corresponding to the system represented in Fig. 2. The equations can be written as

$$\frac{d[\Pi^{(1)}]}{dt} = \lambda_1 F_1([\Pi^{(2)}]) - \kappa_d^{(1)} [\Pi^{(1)}]$$

$$\frac{d[\Pi^{(2)}]}{dt} = \lambda_2 F_2([\Pi^{(1)}]) - \kappa_d^{(2)} [\Pi^{(2)}]$$

where the translation rates F_1 and F_2 are given by

$$F_i([\Pi^{(j)}]) = H(\sqrt{\nu_{ij}} - [\Pi^{(j)}])$$

where H is the Heaviside function,

$$H(x) = \begin{cases} 1, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases}$$

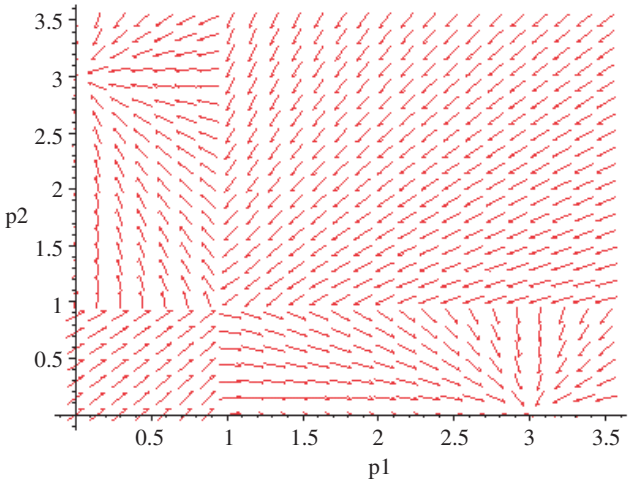

Figure 3. Field plot for the on–off model for co-repressed proteins.

The thresholds are taken to be the values where the Hill's translation functions are half their values when the repressor is not present, so that

$$1 + \nu_{ij}[\Pi^{(j)}]^2 = 1.$$

The direction field for the on–off system is shown in Fig. 3.

Note that Figs 2 and 3 are qualitatively similar. The separatrices are the same. There are three steady states, two of which are stable. Trajectories starting far from the origin have a relative minimum in the solution curve for the protein that winds up dominating. One qualitative difference is that the Hill's model shows a 'leakage', i.e., a non-zero protein level for the non-dominant protein. The McAdams–Arkin model predicts that the non-dominant protein eventually disappears. However, we note that according to the present model, there is an exponentially small amount of the non-dominant protein remaining at any finite time.

6. SIGNALERS

In some cases, a protein molecule alone cannot repress transcription from a DNA site, but after binding with a small molecule (such as iron, fe) it changes its conformation and can bind to the repressor site. We shall denote the repressor molecule without the signaler attached as R, and the signaler as S. The reactions can be written as

$$DNA + P \to DNA_0$$

$$S + R \rightleftarrows R^+$$

$$DNA + R^+ \to DNA_-$$

$$DNA_0 + n_1 \to DNA_1$$

$$DNA_{i-1} + n_i \to DNA_i$$

$$\vdots$$

Proceeding as before, we have

$$\frac{d[R^+]}{dt} = k_S[S][R] - k_D[R^+] \tag{17}$$

$$\frac{d\mathcal{P}(DNA)}{dt} = -(k_p[P] + k_R[R^+])\mathcal{P}(DNA) + k_M[n_M]\mathcal{P}[DNA_{M-1}] + k_-\mathcal{P}(DNA_-) \tag{18}$$

$$\frac{d\mathcal{P}(DNA_0)}{dt} = -k_1[n_1]\mathcal{P}(DNA_0) + k_p[P]\mathcal{P}(DNA) \tag{19}$$

$$\frac{d\mathcal{P}(DNA_-)}{dt} = -k_-\mathcal{P}(DNA_-) + k_r[R]\mathcal{P}(DNA). \tag{20}$$

We again assume that chain elongation can be approximated by

$$\mathcal{P}(DNA_i)(t) = \mathcal{P}(DNA_0)(t - T_i)$$

where T_i is given by

$$T_i = \frac{i}{k[n]}.$$

If we assume that the signal molecule reaction, as well as the Markov reactions are in equilibrium, we have

$$\frac{k_S}{k_D}[S][R] = [R^+] \tag{21}$$

$$\mathcal{P}(DNA_0) = \frac{k_p[P]}{k[n]}\mathcal{P}(DNA) \tag{22}$$

$$\mathcal{P}(DNA_-) = \frac{k_R[R^+]}{k_D}\mathcal{P}(DNA) \tag{23}$$

$$\mathcal{P}(DNA) = \left(1 + \frac{k_R k_S[S][R]}{k_- k_D} + \frac{k_p[P]}{k[n]}M\right)^{-1}. \tag{24}$$

As before, we have

$$\frac{\rho\kappa_R[R_L][R_S]k[n]\mathcal{P}[DNA_0]}{\kappa_-} = \kappa[a][mRNA_0] \tag{25}$$

so that

$$\left.\frac{d[\Pi]}{dt}\right|_{\text{synthesis}} = \kappa[a][mRNA_0]$$

$$= \frac{\rho k_p[P]\kappa_R[R_L][R_S]}{\kappa_-\left(1+\frac{k_R k_S[S][R]}{k_- k_D}+\frac{k_p[P]}{k[n]}M\right)}.$$

Note that the effect of a signaling molecule is to change the effective repressor attachment rate to $k_R \to k_R k_S[S]/k_D$.

7. ACTIVATORS

If the codon possesses an activator site, the 'clean' state of the DNA cannot bind with a polymerase molecule. Instead, an activator molecule must bind first. Then the polymerase can bind and transcription proceeds as before. In this case, the reactions can be written symbolically as

$$DNA + A \rightleftarrows DNA^+$$

$$DNA^+ + P \to DNA_0$$

$$DNA_0 + n_1 \to DNA_1$$

$$DNA_{i-1} + n_i \to DNA_i$$

$$\vdots$$

where A denotes the activator. Proceeding as before, we have

$$\frac{d\mathcal{P}(DNA)}{dt} = -k_A[A]\mathcal{P}(DNA) + k_A^-\mathcal{P}(DNA^+) \tag{26}$$

$$\frac{d\mathcal{P}(DNA^+)}{dt} = k_A[A]\mathcal{P}(DNA) + k_M[n_M]\mathcal{P}[DNA_{M-1}]$$

$$-(k_P[P] + k_A^-)\mathcal{P}(DNA^+) \tag{27}$$

$$\frac{d\mathcal{P}(DNA_0)}{dt} = -k_1[n_1]\mathcal{P}(DNA_0) + k_p[P]\mathcal{P}(DNA^+). \tag{28}$$

We again assume that chain elongation can be approximated by

$$\mathcal{P}(DNA_i)(t) = \mathcal{P}(DNA_0)(t - T_i)$$

where T_i is given by

$$T_i = \frac{i}{k[n]}.$$

Again,

$$\mathcal{P}(DNA) + \mathcal{P}(DNA^+) + k[n]\int_0^{T_M} \mathcal{P}(DNA_0)(t-t')dt' = 1.$$

With

$$\left.\frac{d[\Pi]}{dt}\right|_{\text{synthesis}} = \kappa[a][mRNA_0](t - S_N)$$

$$\mathcal{P}(DNA_0) = \frac{k_p[P]}{k[n]}\mathcal{P}(DNA^+) \tag{29}$$

$$\mathcal{P}(DNA^+) = \frac{k_A[A]}{k_A^-}\mathcal{P}(DNA) \tag{30}$$

$$\mathcal{P}(DNA_0) = \frac{k_p[P]}{k[n]}\frac{k_A[A]}{k_A^-}\mathcal{P}(DNA) \tag{31}$$

$$\mathcal{P}(DNA) = \left[1 + \frac{k_A[A]}{k_A^-}\left(1 + \frac{k_p[P]}{k[n]}M\right)\right]^{-1} \tag{32}$$

$$\left.\frac{d[\Pi]}{dt}\right|_{\text{synthesis}} = \kappa[a][mRNA_0]$$

$$= \frac{\rho k_p[P]\kappa_R[R_L][R_S]k_A[A]}{\kappa_- k_A^-\left[1 + \frac{k_A[A]}{k_A^-}\left(1 + \frac{k_p[P]}{k[n]}M\right)\right]}.$$

8. AUTOACTIVATION

The special case where Π and A are identical is the case when a protein activates itself. In this case, the differential equation for the concentration of $[\Pi]$ becomes

$$\frac{d[\Pi]}{dt} = \frac{\rho k_p[P]\kappa_R[R_L][R_S]k_A[\Pi]}{\kappa_- k_A^-\left[1 + \frac{k_A[\Pi]}{k_A^-}\left(1 + \frac{k_p[P]}{k[n]}M\right)\right]} - k_E[E][\Pi].$$

The advantage of this equation over the postulation of a Michaelis–Menten equation for the rate of change of protein concentration is that this equation relates the parameters in the protein rate equation to more fundamental rates, such as the rates of chain elongation, protein and mRNA decay, and degradation. The equation can be written as

$$\frac{d[\Pi]}{dt} = \frac{\mu[\Pi]}{(1 + \nu[\Pi])} - k_d[\Pi]$$

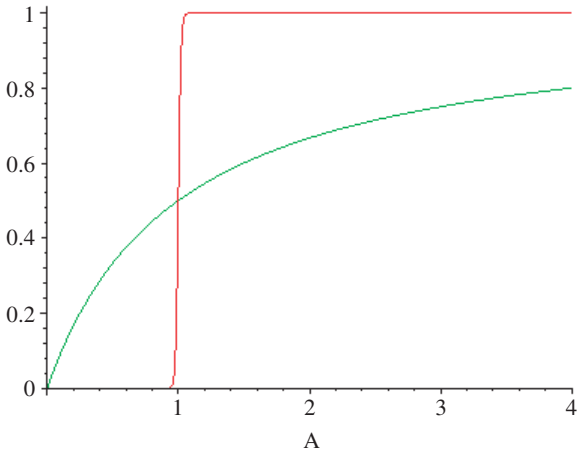

Figure 4. Rate of expression of a protein as a function of the concentration of its activator.

where

$$k_d = k_E[E]$$
$$\mu = \frac{\rho k_p[P]\kappa_R[R_L][R_S]k_A}{\kappa_- k_A^-}$$
$$\nu = \frac{k_A}{k_A^-}\left(1 + \frac{k_p[P]}{k[n]}M\right).$$

The form of the synthesis term is shown in Fig. 4 for $\mu = 1$, $\nu = 1$. Also shown is the step function $H(x - 1)$, where H is the Heaviside function. This latter function is of interest because it represents a 'switch' in the form suggested by McAdams and Arkin (1997).

It is difficult to justify the approximation of the synthesis rate as a step function in this case.

Figure 5 shows the synthesis rate for an autoactivated protein, along with the degradation rates for various values of k_d. Note that for $k_d \geq 1$, the equilibrium concentration of a self-activated protein is zero. If it degrades too fast (relative to its expression rate), the protein expression mechanism will shut down, although the degradation is an exponential decay process.

9. CONCLUSION

I have presented here a model which accounts for the kinetic modeling of the steps involved in protein synthesis in such a way that the connection to the genome is evident. The model deals with the various states of the DNA codon by deriving equations for the evolution of the probability that the codon is in various states.

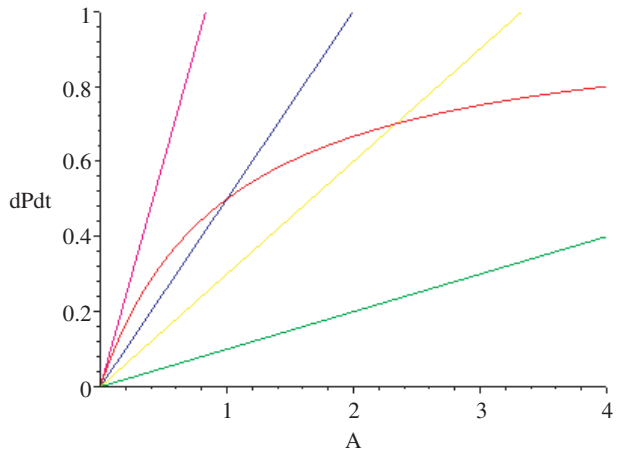

Figure 5. Equilibrium concentrations of a self-activated protein.

This assumption allows the description of the common situation where the cell has one codon for the protein considered, but it also seems appropriate for the situation where there are several codons in the genome for the protein considered. This is common in a plasmid. The steps modeled include the rates of attachment of the polymerase and an activator or repressor molecule to the DNA strand that codes for the protein, the elongation of the nascent mRNA chains, the attachment of the ribosome body to the mRNA molecules, the elongation of the nascent peptide chains, the dimerization of the protein into a repressor molecule, and the degradation of the protein. The novel feature of this model is that the single DNA strand in a cell is assumed to take on different states with probabilities that evolve according to a Markov process.

An approximate model is derived from the full system of equations assuming that the rate of protein degradation is slow compared to all the other kinetic reactions in the system. This assumption results in a Michaelis–Menten or a Hill's type equation for the evolution of the protein. This sort of result is not new (Mahaffy, 1999; Elowitz and Liebler, 2000; Gardner *et al*., 2000); what is new is that the parameters (μ_i, λ_i, ν_{ij}, and κ_d^i) are related to the rates of various basic processes in the cell (mRNA chain elongation, peptide chain elongation, polymerase attachment, ribosome attachment, mRNA degradation), along with protein-specific rates (protein degradation, activator or repressor attachment). Whether this sort of information will be of use remains to be seen.

An 'on/off' approximation to the Hill's model is presented, and its qualitative similarity to the Hill's model is discussed.

Finally a model for an activator system is presented.

9.1. ***Room for improvement.*** The model does not account for the multiple binding of ribosomes or proteases. As a practical matter, if the binding site is available, or the ribosome has moved sufficiently far that another can come and bind, multiple

chains can form. The same applies to the polymerase attached to the DNA strand.

The model does not consider the different attachment rates for different nucleotides during mRNA formation, nor for different attachment rates for different amino acids in the protein chain.

The volume of a cell changes during its growth. The dilution due to this effect is not accounted for.

APPENDIX A: STEADY STATE

If we assume that $d/dt = 0$, and neglect the time delays in equations (2), (3), (9), (10) and (11), we have

$$\overline{\mathcal{P}(DNA_0)} = \frac{k_p[P]}{k[n]}\overline{\mathcal{P}(DNA)}$$

$$\frac{k_R\overline{[R]}}{k_-}\overline{\mathcal{P}(DNA)} = \overline{\mathcal{P}(DNA_-)}$$

$$\overline{\mathcal{P}(DNA)} = \frac{1}{\left(1 + \frac{k_R\overline{[R]}}{k_-} + \frac{k_p[P]}{k[n]}M\right)}$$

$$\overline{[mRNA]} = \frac{\rho k[n]\overline{\mathcal{P}[DNA_0]}}{\kappa_-} = \frac{\frac{\rho k_p[P]}{\kappa_-}}{\left(1 + \frac{k_R\overline{[R]}}{k_-} + \frac{k_p[P]}{k[n]}M\right)}$$

$$\overline{[mRNA_0]} = \frac{\overline{[mRNA]}\kappa_R[R_L][R_S]}{\kappa[a]}$$

$$\overline{[\Pi]} = \frac{\kappa_R\overline{[mRNA]}[R_L][R_S]}{k_E[E]}$$

where the overbar denotes the steady-state value.

APPENDIX B: MICHAELIS–MENTEN DYNAMICS

We nondimensionalize the concentrations with their steady-state values

$$[\Pi] = \overline{[\Pi]}p$$

$$[R] = \overline{[R]}r.$$

Also,

$$[mRNA] = \overline{[mRNA]}m$$

$$[mRNA_0] = \overline{[mRNA_0]}m_0$$

and

$$t = T\tau.$$

We also scale the DNA state probabilities by their steady state values. Thus

$$\mathcal{P}(DNA) = \overline{\mathcal{P}(DNA)}\delta$$
$$\mathcal{P}(DNA_0) = \overline{\mathcal{P}(DNA_0)}\delta_0$$
$$\mathcal{P}(DNA_-) = \overline{\mathcal{P}(DNA_-)}\delta_-.$$

Then the scaled equations become

$$\begin{aligned}
\frac{\overline{\mathcal{P}(DNA_0)}}{T}\frac{d\delta_0}{d\tau} &= -k[n]\overline{\mathcal{P}(DNA_0)}\delta_0 + k_p[P]\overline{\mathcal{P}(DNA)}\delta \\
\frac{\mathcal{P}(DNA^-)}{T}\frac{d\delta_-}{d\tau} &= k_r\overline{R}r\,\overline{\mathcal{P}(DNA)}\delta - k_-\overline{\mathcal{P}(DNA_-)}\delta_- \\
\frac{\overline{[mRNA]}}{T}\frac{dm}{d\tau} &= \rho k[n]\overline{\mathcal{P}(DNA_0)}\delta_0(\tau-\tau_N) + \kappa[a]\overline{[mRNA_0]}m_0(\tau-\sigma_N) \\
&\quad -(\kappa_R[R_L][R_S]+\kappa_-)\overline{[mRNA]}m \\
\frac{\overline{[mRNA_0]}}{T}\frac{dm_0}{d\tau} &= \kappa_R\overline{[mRNA]}[R_L][R_S]m - \kappa[a]\overline{[mRNA_0]}m_0 \\
\frac{\overline{[R]}}{T}\frac{dr}{dt} &= \kappa[a]\overline{[mRNA_0]}m_0(t-\sigma_N) + \rho k_-\overline{\mathcal{P}[DNA_-]}\delta_- \\
&\quad -\rho k_R\overline{[R]}\,\overline{\mathcal{P}(DNA)}r\delta - k_E[E]\overline{[R]}r(t)
\end{aligned}$$

where $\tau_M = M/Tk[n]$, and $\sigma_N = N/T\kappa[a]$. We take as a time scale the time it takes for the protein to degrade. Then $T = 1/k_E[E]$, and the equations become

$$\delta + \frac{k_R\overline{[R]}}{k_-}r\delta^- + \frac{k_p[P]}{k_E[E]}M\left[\int_0^{\tau_M}\delta_0(\tau-\tau_i)d\tau_i\right] = \left(1 + \frac{k_R\overline{[R]}}{k_-} + \frac{k_p[P]}{k[n]}M\right)$$

$$\begin{aligned}
\varepsilon_1\frac{d\delta_0}{d\tau} &= -\delta_0 + \delta \\
\varepsilon_2\frac{d\delta^-}{d\tau} &= r\delta - \delta^- \\
\varepsilon_3\frac{dm}{d\tau} &= \delta_0(\tau-\tau_N) + \lambda_1 m_0 - (\lambda_1+1)m \\
\varepsilon_4\frac{dm_0}{d\tau} &= m - m_0 \\
\frac{dr}{d\tau}(\tau) &= m_0(\tau-\sigma_M) + \lambda_2(\delta_- - r\delta) - r(\tau)
\end{aligned}$$

where

$$\varepsilon_1 = \frac{k_E[E]}{k[n]}, \quad \varepsilon_2 = \frac{k_E[E]}{k_-}$$

$$\varepsilon_3 = \frac{k_E[E]}{\kappa_-}, \quad \varepsilon_4 = \frac{k_E[E]}{\kappa[a]}$$

$$\lambda_1 = \frac{\kappa_R[R_L][R_S]}{\kappa_-}, \quad \lambda_2 = \frac{\frac{k_R}{k_E[E]}}{\left(1 + \frac{k_r[\overline{R}]}{k_-} + \frac{k_p[P]}{k[n]}M\right)}$$

$$\tau_M = Mk_E[E]/k[n]$$

$$\sigma_N = Nk_E[E]/\kappa[a].$$

We shall assume that ε_1, ε_2, ε_3, and ε_4 are all small, because the protein degradation rate $k_E[E]$ is small compared to the rate of degradation of the RNA polymerase bond to the DNA, the rate of degradation of mRNA, the rate of attachment of the RNA polymerase, and the rate of elongation of the polymer chains. Indeed, let us consider τ_N and σ_M. We have

$$\tau_M = M\kappa_d/k[n] = M\varepsilon_2$$

$$\sigma_N = N\kappa_d/\kappa[a] = N\varepsilon_5.$$

If the ε_is are sufficiently small that τ_M and σ_N are also small, we can approximate the system by the system given by equations (12)–(16).

REFERENCES

Brendel, V. and A. S. Perelson (1993). A quantitative model of ColE1 plasmid copy number control. *J. Mol. Biol.* **229**, 860–872.

Elowitz, M. B. and S. Liebler (2000). A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338.

Gardner, T. S., C. R. Cantor and J. J. Collins (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342.

Goodwin, B. C. (1963). *Temporal Organization in Cells*, London and New York: Academic Press.

Gottesman, S. (1996). Proteases and their targets in *Escherichia Coli*. *Annu. Rev. Genet.* **30**, 465–506.

Jacob, F. and J. Monod (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356.

Lodish, H. B., A. Zipursky, S. L. Matsudaira, P. Baltimore and J. E. Darnell (2000). *Molecular Cell Biology*, W. H. Freeman (Ed.), New York.

Mahaffy, J. M. (1999). Stability analysis for a mathematical model of the lac operon. *Q. Appl. Math.* **LVII, 1**, 37–53.

McAdams, H. H. and A. Arkin (1997). Stochastic mechanisms in gene expression. *Proc. Nat. Acad. Sci.* **94**, 814–819.

Novak, B. and J. J. Tyson (1997). Modeling the control of DNA replication in fission yeast. *Proc. Natl. Acad. Sci. USA* **94**, 9147–9152.

Thieffry, D. (1999). From global expression data to gene networks. *BioEssays* **21**, 895–899.

Received 6 September 2000 and accepted 20 December 2000