The Modeler's Guide to Biology

Ittai Rubinstein and Yair Maimon

October 18, 2014

Contents

Preface

In this document, we will go over all of the modeling done in our iGEM project. It is built as a guide to modeling for two reasons:

Obviously we want for this to be a potential learning tool for any future aspiring iGEM modeler.

Second, we believe that it would be easier to follow our train of thought this way.

As such, this guide is built so that it gradually builds up knowledge in molecular biology and its modeling, while at the same time slowly increasing the level of the mathematical background necessary for following and for producing the proofs provided, and their meaning.

We hope to recieve questions regarding this document at igem14il@gmail.com (is this our email?), so that we may help as many modelers as possible as they take their first step.

Part I
Introduction

1 From the Law of Mass Effect to building a model of a Promoter

1.1 Promoter Activated (or Repressed) by Several Compounds

1.1.1 Rules of the game

Let $C_1, ..., C_N$ be the concentrations of N compounds (proteins or proteins bound with various ligands). Let P be a promoter activated (or repressed) by these compounds (and cannot be connected to more than one at a time). Assume that when P is bound to C_j, then the rate of production of the mRNA regulated by P is given by v_j, and that once it has been created, the $mRNA$ molecule degrades Poissonically (see models for radioactive decay) with a rate constant γ.

Other than that, we will assume only one other law of nature - the law of mass effect: This law states that the rate at which the reaction $A, B \to C, D$, is $\alpha [A] \cdot [B]$, for some constant factor α.

Our goal is to find the differential equation characterising the concentration of R.

1.1.2 Probability of combination

Let W_j denote the concentration of P-type Promoters connected with a compound C_j (i.e. $W_j = [P - C_j]$), and W_f denote the concentration of free P Promoters.

Assume that at any given moment in time, a promoter can, either lose a compound and thus go from being counted in W_j to W_f or vice versa. Using the law of mass effect, we can obtain the rate of each reaction:

$$Rate(W_f \to W_j) \propto W_f C_j$$

$$Rate(W_j \to W_f) \propto W_j$$

Finally, we can obtain from these rates a set of diff. equations by keeping in mind the following rules:

1. The only way a promoter can join the class W_j is by being a free promoter and then meeting a C_j compound.
2. The only way a promoter can leave the class W_j is by being a W_j and losing its compound. From these two rules we obtain:
$$\frac{dW_j}{dt} = \alpha_j C_j W_f - \beta_j W_j$$
3. The total number of promoters does not change (within the scope of this model). Therefore:
$$\frac{dW_f}{dt} = -\sum_{j=1}^{N} \frac{dW_j}{dt} = \sum_{j=1}^{N} (\beta_j W_j - \alpha_j C_j W_f)$$

We will now use perhaps our only tool for the simplification of rate equations: we will choose to assume that some of them happen quickly enough, for them to be in a constant equilibrium state. The way we do this, is to assume that $\forall j \ \ \frac{dW_j}{dt} = 0$, and solve the equation that arrises from this assumption:

Therfore:

$$W_j = \frac{\alpha_j}{\beta_j} C_j W_f = k_j C_j W_f$$

Therefore:

$$P = W_f + \sum_{j=1}^{N} W_j = \left(1 + \sum_{j=1}^{N} k_j C_j\right) W_f$$

by defining

$$P_j = P(C_j \ is \ connected \ to \ P) = \frac{W_j}{P}$$

we obtain

$$P_j = \frac{k_j C_j}{1 + \sum_{j=1}^{N} k_j C_j} \tag{1}$$

1.1.3 The Equation itself

Now we are ready for the next and final step:

We will assume that the rate of production of the mRNA R is the average one.

Thus we will obtain the rate eqs:

$$Rate(\phi \to R) =< v >= \frac{\sum_{j=1}^{N} k_j C_j v_j + k_{free}}{\sum_{j=1}^{N} k_j C_j + 1} N_P$$

Where N_P is the total number of P promoters.

$$Rate(R \to \phi) = \gamma R$$

Therefore:

1.1.4 The Final Result:

$$\frac{dR}{dt} = \frac{\sum_{j=1}^{N} k_j C_j v_j + k_{free}}{\sum_{j=1}^{N} k_j C_j + 1} N_P - \gamma R \tag{2}$$

1.1.5 Paramaters

The paramaters k_j can theoretically be calculated using the Boltzman distribution, if the binding energies and entropies are known to an infinite degree of percision, but they are not, and seeing as the calculation of k_j requires an exponent of those, any small variation would result in an extremely inaccurate calculation.

1.2 Special Case

1.2.1 Basic Description

Let us assume that the compounds C_j are all made up of a basic protein Q which can be combined with (up to) $N-1$ Ligands of the same type L (which will henceforth denote the concentration of the ligands in the solution). At any given moment in time, a ligand can be added if there is space or removed if there is a ligand to remove.

A helpful analogy, is a coat stand (representing the protein), with many arms ($N-1$ to be precise), each of which can hold up to one item of clothing. Every now and then, someone will pass by this coat stand, and will throw a coat (representing the ligands) at some random arm (of some random coat stand).

Figure 1: A coat stand

1.2.2 Basic Operations

The combination of a ligand to a complex We will now attempt to find the rate of the combination of ligands to compounds. This will be the rate at which a C_{j-1} type object becomes a C_j object.

Let us return to the coat stand.

1. The coat stand has $N-1$ arms, $j-1$ of which are taken. Therefore, it has $N-j$ free arms. If I were to choose an available coat stand arm at random, and put my coat on it, then the odds of my placing my coat on a certain coat stand is proportional to the number of available arms it has.
2. In addition, the more coats are thrown at the stands the more likely, any given coat arm is to have a coat placed on it.
3. Of course, the whole thing is proportional to the size of the party

Using these three rules, we obtain the following equation:

$$Rate(C_{j-1} \to C_j) \propto (N-j)LC_{j-1}$$

Generally, it can be considered that, one might prefer to place their coats on an empty stand to avoid getting it wet from other coats (ligands might repel each other) or one might prefer to place their coat near other coats as a landmark (lignads might attract each other). This would cause the proportionality factors to be dependent on the index j. However, for the time being, we will not take this into account.

$$Rate(C_{j-1} \to C_j) = \alpha(N-j)LC_{j-1}$$

The breaking of a protein-Ligand bond We will now add a new player into the game:

People who placed their coats on a coat stand may one day want them back (or conversely, a ligand might fall off the protein).

This will cause our compound to go from the state C_j to the state C_{j-1}.

Let us continue with the analogy, to get a better image of the rate of this reaction:

1. If a coat stand has $(j-1)$ coats on it, and we remove a random coat from a random stand, then the odds of us removing a coat from this stand are proportional to the number of coats on it.

2. The people we are dealing with are antisocial: They don't care how many people are out there looking for a stand for their coat and won't remove their coat for another's sake. Or in Bio-terms, the rate of the reaction $C_j \to C_{j-1}$ is independent of L.

3. Of course, the whole thing is proportional to the size of the party

From these three rules alone, we obtain:

$$Rate(C_j \to C_{j-1}) \propto (j-1)C_j$$

Once again, it is possible that people will consider the number of coats on their stand when deciding whether or not to remove their coats, but we will not take this into account for now. Therefore:

$$Rate(C_j \to C_{j-1}) = \beta(j-1)C_j$$

1.2.3 Summing up the Contributions:

We consider the ways in which a coat stand can either start or stop being class C_j:

1. It goes from having $j-1$ coats to having $j-2$ coats ($C_j \to C_{j-1}$) or vice versa

2. It goes from aving $j-1$ coats to having j coats ($C_j \to C_{j+1}$) or vice versa

Summing up these contributions we get:

$$\frac{dC_j}{dt} = \alpha(N-j)C_{j-1} + \beta j C_{j+1} - \left[\alpha L(N-j-1) + \beta(j,-1)\right] C_j \tag{3}$$

1.2.4 The special special case of $N = 3$

In this case the afforementioned equations are:

$$\frac{dC_1}{dt} = \beta C_2 - 2\alpha L C_1$$

$$\frac{dC_2}{dt} = 2\beta C_3 + 2\alpha L C_1 - \alpha L C_2 - \beta C_2$$

$$\frac{dC_3}{dt} = \alpha L C_2 - 2\beta C_3$$

2 The Alpha System

Throughout the rest of this document, we will be working with several variants of the system which we have designed for practical use - The Alpha System.

2.1 What does our system contain?

The easiest way to answer this is to visit our wiki page at: http://2014.igem.org/Team:Technion-Israel/Project#alpha

For those of you who want to skip the somewhat long read, we have designed a quorum sensing based positive feedback loop which can be bypassed by a promoter activated by the presence of a harmful substance. The feedback loop centers arround Gate 3, which produces the AHL which activates it, as a dimer - i.e. it takes two copies of the ligand to activate the promoter.

2.2 What should we look at when modelling our system?

Notice, that the only part of the system, which is acts not simply as a function of its surroundings is Gate 3: It produces the AHL which activates itself. This part is the center of the positive feedback loop, and requires two ligands for the activation of the promoter.

2.3 Modeling Gate 3

Let us start out by writing down all of the equations necessary:

We know that the promoter is activated by two ligands, which means that there exist some binding constant k, and some transcription factors $v_{A,}v_B$ (for Active and Basal), for which:

$$\frac{d\left[mRNA_{LuxI}\right]}{dt} = \frac{v_B + v_A k_A \left[AHL\right]^2}{1 + k_A \left[AHL\right]^2} - \gamma_{mRNA_{LuxI}} \left[mRNA_{LuxI}\right] + GateI \tag{4}$$

This mRNA can then be tranlated into a protein:

$$\frac{d\left[LuxI\right]}{dt} = \alpha_{LuxI} \left[mRNA_{LuxI}\right] - \gamma_{LuxI} \left[LuxI\right] \tag{5}$$

which then produces AHL at a constant rate giving us:

$$\frac{d\left[AHL\right]}{dt} = \alpha_{AHL} \left[AHL\right] - \gamma_{AHL} \left[AHL\right] \tag{6}$$

Once again, in order to simplify our system of equations, we will assume a steady state where necessary, and obtain

$$\left[AHL\right] \propto \left[LuxI\right] \propto \left[mRNA_{LuxI}\right]$$

and by sweeping the factors of proportionality under the "rug" which is the ambiuity in the choice of units for k, v_A, v_B we obtain:

$$\frac{d\left[AHL\right]}{dt} = \frac{v_B + v_A k_A \left[AHL\right]^2}{1 + k_A \left[AHL\right]^2} - \gamma_{AHL} \left[AHL\right] + GateI \tag{7}$$

We can now advance to the next step with this equation, but first we will need to do two other things:

1. Figure out what we want for the system to be able to do.
2. See how a few small changes to our system can affect this equation, so that from now on, we can "carry" all of the equations along (proving each claim once for all of the models).

2.4 Variation of the Alpha System

We will now deal with a variation of the Alpha System called the RNA Splint. For precise information about the biological mechanism behind it, feel free to visit our wiki at: http://2014.igem.org/Team:Technion-Israel/Project#rna

As a general overview, we the RNA splits the task of building the mRNA of LuxI into three parts, which have to be combined in one reaction. This means that:

$$\left[AHL\right] \propto \left[LuxI\right] \propto \left(\left[mRNA\right]^3\right)$$

From here on out, there are two ways to continue:

Either we simply change the AHL to being concentration to being $\left[AHL\right]^3$ in the production term, or we can raise the construction rate to the third power, to obtain one of the following equations:

$$\frac{d\left[AHL\right]}{dt} = \frac{v_B + v_A k_A \left[AHL\right]^6}{1 + k_A \left[AHL\right]^6} - \gamma_{AHL} \left[AHL\right] + GateI \tag{8}$$

$$\frac{d\left[AHL\right]}{dt} = \left(\frac{v_B + v_A k_A \left[AHL\right]^2}{1 + k_A \left[AHL\right]^2}\right)^3 - \gamma_{AHL} \left[AHL\right] + GateI \tag{9}$$

2.5 General Alpha System, Goals, and our plans for the next part of the model

So far, we have seen, that the models for our system are all of the form

$$\frac{d\left[AHL\right]}{dt} = \left(\frac{v_B + v_A k_A \left[AHL\right]^n}{1 + k_A \left[AHL\right]^n}\right)^m - \gamma_{AHL}\left[AHL\right] + GateI$$

For positive $v_{A,B}, k, \gamma$. To simplify the future analysis, we will slightly change the form of the equation:

$$\frac{dx}{dt} = \left(\frac{v_1 + v_2 k x^n}{1 + k x^n}\right)^m - \gamma x + G$$

We will now discuss what it is our system is trying to accomplish, and how we hope to get there:

We want to be able to detect substances at low concentrations and without false-positives. This requirement of having a sharp resolution between on and off, has led us to look for bi-stability in our system. Throughout the next two parts, two methods for gouging bi-stability will be obtained, and results for our three models will be shown.

Part II
Determining the Bi-Stability of Positive Feedback Loops

1 Limitations

1.1 The Limitations of our Method for Determining the Bi-Stability of a Positive Feedback Loop

We will assume that we are dealing with a system which has only one material in it, which's concentration fits the following ODE:

$$\frac{dx}{dt} = f(x) = c(x) - d(x) \tag{10}$$

Where, c is the rate at which the material is produced. We will assume that c is a real analytic function (if you are not from a pure-mathematics background you can read this as: is a "nice" function), with one inflection point at most, which has a finite upper bound, and which fits the following conditions:

$$\forall x > 0 \quad c(x), c'(x) > 0 \quad \&\& \quad c(0) > c'(0) = 0$$

We will also assume that $\exists \gamma \forall x \geq 0 \quad d(x) = \gamma x$.

1.2 Proof that our System Fits These Limitations

The only limitation which is not obvious for our system is our c contains one inflection point at most. To show this, we will find the second derivative of c, and solve $c'' = 0$.

We know that

$$c(x) = \left(\frac{v_1 + v_2 k x^n}{1 + k x^n}\right)^m$$

So we define:

$$g(x) = \frac{v_1 + v_2 k x^n}{1 + k x^n}; \quad c = (g^m)$$

$$c'(x) = m g^{m-1} g' = mn(v_2 - v_1) g^{m-1} \frac{k x^{n-1}}{(1 + k x^n)^2}$$

$$c'' = m(m-1) g^{m-2} (g')^2 + m g^{m-1} g'' = m(m-1) g^{m-2} \left(n(v_2 - v_1) \frac{k x^{n-1}}{(1 + k x^n)^2}\right)^2 + mn(v_2 - v_1) g^{m-1} \frac{(n-1) x^{n-2} (1 + k x^n) + 2}{(1 + k x^n)^3}$$

$$c'' = m \left[(m-1) \left(nk(v_2 - v_1) \frac{x^{n-1}}{(1 + k x^n)^2}\right)^2 \left(\frac{v_1 + v_2 k x^n}{1 + k x^n}\right)^{m-2} + nk(v_2 - v_1) \left(\frac{v_1 + v_2 k x^n}{1 + k x^n}\right)^{m-1} \frac{(n-1) x^{n-2} (1 + k x^n) + 2nk}{(1 + k x^n)^3} \right.$$

$$(m-1) \left(nk(v_2 - v_1) \frac{x^{n-1}}{(1 + k x^n)^2}\right)^2 \left(\frac{v_1 + v_2 k x^n}{1 + k x^n}\right)^{m-2} = -nk(v_2 - v_1) \left(\frac{v_1 + v_2 k x^n}{1 + k x^n}\right)^{m-1} \frac{(n-1) x^{n-2} (1 + k x^n) + 2nk x^{2n-2}}{(1 + k x^n)^3}$$

Multiply by $\frac{\left(\frac{v_1 + v_2 k x^n}{1 + k x^n}\right)^{2-m} \cdot \left((1 + k x^n)^4\right)}{nk(v_2 - v_1)}$, to obtain:

$$(m-1) nk(v_2 - v_1) x^{2n-2} = -(v_1 + v_2 k x^n) \left[(n-1) x^{n-2} (1 + k x^n) + 2nk x^{2n-2}\right]$$

Divide by x^{n-2}, and substitute $q = k x^n$ to obtain:

$$(m-1) n(v_2 - v_1) q = (v_1 + v_2 q) \left[(n-1)(1 + q) + 2nq\right]$$

This is a quadratic equation in q. Therfore, it has at most two real positive solutions (it has at most two solutions, but we are not interested in the complex and imaginary ones). Since it is obvious that $c''(\epsilon) > 0$ for any sufficiently small $\epsilon > 0$, and that for sufficiently large M, $c''(M) < 0$, it must have an odd number of inflection points between ϵ, and M, which means that it cannot have 2 inflection points along the positive axis. Which goes to show that it has up to one such inflection point. QED

2 Analysis of the General Case

We will now return to the general case, and try to find a necessary and sufficient condition for bi-stability.

2.1 The Condition

We claim that the system is bi-stable, iff, there exist $0 < z_1 < z_2$, for which $c'(z_i) = \gamma_i$, $c(z_i) = \gamma_i z_i$, and $\gamma_1 < \gamma < \gamma_2$ (i.e. that the contour of c is tangent to lines $\gamma_{1,2}x$, and the line γx passes between them).

2.2 Proof

2.2.1 First Direction - from bi-stability to tangents

We will begin by using a well known theorem from the field of qualitative solutions for ODEs:

The system $\frac{dx}{dt} = f(x)$ has a stable equilibrium point at $x = x_0$, iff f changes its sign from positive to negative at x_0, for a smooth enough f (in our case, f is a real anlytic function).

This means that f changes its sign from positive to negative at at least two points x_1, x_3 along the positive real axis. For this to be possible, there has to be another equilibrium point x_2 between x_1, x_3 (as a result of the intermediate value theorem), and such that in x_2, f changes its sign from negative to positive. Moreover, from considerations of convexity of f, there can only be one point x_2 as described.

This means that x_1, x_2, x_3 are the only points where the line γx crosses the contour c. In addition, we know that $f'(x_1), f'(x_3) \leq 0 \leq f'(x_2)$. Let us proove by contradiction that $f(x_i) \neq 0$ for all i. Assume otherwise. Therefore, from considerations of convexity, this crossing must occur at the single inflection point, and $\forall x < x_i \;\; f(x) > 0$ and $\forall x > x_i \;\; f(x) < 0$, which is contradiction with the claim that there are three points for which $f(x) = 0$. Therefore:

$$f'(x_1), f'(x_3) < 0 < f'(x_2)$$

We now define:

$$q(x) = \frac{c(x)}{c'(x)\,x}$$

First of all, the points z_1, z_2, we are looking for have to maintain that $q(z_i) = 1$, and that $f(z_1) < 0 < f(z_2)$. The second condition is promised if we show that $z_i \in (x_i, x_{i+1})$.

As for the first condition notice that:

$$q(x_1) > 1 > q(x_2)$$

$$q(x_2) < 1 < q(x_3)$$

And so the existance of the points z_1, z_2 required is proven.

2.2.2 Second Direction

Now let us assume that z_1, z_2, exist.

We need to show that there exist x_1, x_3 as in the the theorem about ODEs, and we're done.

But notice that by our assumptions:

$$f(0) > 0 > f(z_1)$$

$$f(z_2) > 0 > -\infty > \lim_{x \to \infty} f(x)$$

Which proves that such x_1, x_3 exist as a result of the convexity of c, and of the intermediate value theorem.

QED

3 Analysing the Condition for Our System

We will now look to see what happens when c is the function we defined for our system:

We want to find the points z_1, z_2 if they exist, and from there on our it is a simple arithmetic calculation to find $\gamma_{1,2}$, and find out whether or not this is bi-stable.

We do this by solving the equation

$$q(x) = 1$$

$$xc' = c$$

$$mn\left(v_2 - v_1\right) g^{m-1} \frac{kx^{n-1}}{\left(1 + kx^n\right)^2} x = g^m$$

$$mn\left(v_2 - v_1\right) \frac{kx^n}{\left(1 + kx^n\right)} = \frac{v_1 + v_2 kx^n}{1 + kx^n}$$

Substitute $\zeta = kx^n$, to obtain

$$mn\left(v_2 - v_1\right)\zeta = \left(1 + k\zeta\right)\left(v_1 + v_2\zeta\right)$$

which is a quadratic equation which can easily be solved, arithmetically.

To obtain more data, we also defined a normalized bi-stability parameter:

$$N = \frac{\gamma_2 - \gamma_1}{\gamma_2}$$

and plotted both for each of the configurations of our system.

To see more of our results visit our wiki at: http://2014.igem.org/Team:Technion-Israel/Modeling

Part III
Stochastic Models

1 The General Case: From the Master Equation to the Fokker-Planck Equation

1.1 The Master Equation

For the sake of simplicity, we will model a system containing only one material M, of which there are x copies, and such that the probability that a new molecule will appear in the infinitesimal time interval $(t, t+dt)$, given that there are x such molecules at the time t, is $c(x)$, and that the probability that such a molecule will be destroyed is $d(x)\,dt$ (c for created, and d for destroyed).

Let $p_n(t)$, be the probability that $x = n$, at time t. We know that:

$$\frac{p_n(t+dt) - p_n(t)}{dt} = \frac{dt}{dt} \cdot (c(n-1)\,p_{n-1}(t) - c(n)\,p_n(t)) + (d(n+1)\,p_{n+1}(t) - d(n)\,p_n(t)) \tag{11}$$

1.2 The Fokker Planck Equation

Now expand the RHS of equation 11, and by assuming that the functions p, c, d are slow-changing enough, as functions of n, neglect the third order and so on. Also define $f(y,\tau) = \lim_{dx\to 0^+} \frac{P(x\in(y,y+dx)|t=\tau)}{dx}$, and obtain:

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial\left[(c(x) - d(x))\,f(x,t)\right]}{\partial x} + \frac{1}{2}\frac{\partial^2\left[(c(x) + d(x))\,f(x,t)\right]}{(\partial x)^2} + O\left(\frac{\partial^3\left[(c(x) - d(x))\,f(x,t)\right]}{(\partial x)^3}\right) \tag{12}$$

Notice, that unlike a rate equation, the Fokker Planck equation is:

1. A PDE instead of an ODE (making it unlikely that it will be solvable analytically).
2. Highly dependent upon the units of x (in a rate equation, if you change the units of x, then the whole equation is multiplied by the same factor, but here, this is not true). The units assumed in the development of this equation, are such that x is measured in single molecules. If these are not the units used, then an appropriate factor must be appended to the last two terms.

2 In Our System

It is known, that for all the models for our system, the following holds true:

$\exists m, n, k, v_1, v_2, \gamma$:

$\forall x \geq 0$

$$(c(x), d(x)) = \left(\left(\frac{v_1 + v_2 k x^n}{1 + k x^n} \right)^m, \gamma x \right) = ((g(x)^m), \gamma x)$$

~In a previous (deterministic) model, we showed a way of characterizing whether or not this equation describes a bi-stable system when analyzing it with the rate eqaution (i.e. when the rate equation, which is an ODE, has two stable equilibrium points, and one unstable equilibrium point between them).~

We will now substitute this into equation 12, to obtain:

$$c'(x) = m g^{m-1} g' = mnk(v_2 - v_1) \frac{x^{n-1}}{(1 + k x^n)^2} \left(\frac{v_1 + v_2 k x^n}{1 + k x^n} \right)^{m-1}$$

$$c''(x) = m(m-1) g^{m-2} (g')^2 + m g^{m-1} g''$$

$$c''(x) = m \left[(m-1) \left(nk(v_2 - v_1) \frac{x^{n-1}}{(1 + k x^n)^2} \right) \left(\frac{v_1 + v_2 k x^n}{1 + k x^n} \right)^{m-2} + nk(v_2 - v_1) \left(\frac{v_1 + v_2 k x^n}{1 + k x^n} \right)^{m-1} \frac{(n-1) x^{n-2} (1 + k x^n) + 2}{(1 + k x^n)^3} \right.$$

$$d'(x) = \gamma$$

$$d''(x) = 0$$

$$\frac{\partial f}{\partial t} = -(c - d) \frac{\partial f}{\partial x} - (c' - d') f + \frac{1}{2} \left[(c + d) \frac{\partial^2 f}{(\partial x)^2} + (c' + d') \frac{\partial f}{\partial x} + (c'' + d'') f \right]$$

It is clear from this equation, that the element which enables the appearance of a standard deviation in the solution, is proportional to a monotonically increasing function of k. Since k's effect on the bi-stability is known, we will hereby assume that $k = 1$.

We produced a numerical approximation of the solution of these equations for several different configurations of this system, and for each of the possible stable equilibrium points.

For plotted results feel free to visit our wiki at http://2014.igem.org/Team:Technion-Israel/Modeling