CHARACTERIZATION AND OPTIMIZATION OF GENETIC TOGGLE SWITCH IN MAMMALIAN CELLS VIA POSITIVE FEEDBACK MECHANISMS

ARVIND THIAGARAJAN
DEPARTMENTS OF BIOLOGICAL ENGINEERING AND PHYSICS
MASSACHUSSETTS INSTITUTE OF TECHNOLOGY

ON BEHALF OF THE MIT 2010 IGEM TEAM

1. Introduction

Synthetic Biology is a fast growing field, and promises to accelerate our understanding of biology and biological engineering to the level at which we understand electrical engineering today. Nonetheless, despite this vast potential, synthetic biology remains primarily confined to prokaryotic and simple eukaryotic systems. While these are undoubtedly useful, both as model systems and as platforms for the development of various biotechnologies, the implementation of biological programming and synthetic circuits in medical contexts requires us to extend synthetic biology into mammalian systems. While the greater complexity of mammalian systems over simpler systems undoubtedly presents many challenges, it also provides a larger set of tools to engineer synthetic circuits.

Of course, some work has been conducted in mammalian systems already. In particular, the bistable toggle switch (1) designed by James J. Collins has been successfully implemented in embryonic stem cells of mice (2) by Sairam Subramanian. However, the operational time scale of this switch was on the order of two days, and so we sought to implement a faster switch using a different design. In particular, by incorporating both positive and negative feedback into our circuit design, we have a system that in theory should change state much more quickly. However, as might be expected, there is a price to be paid for the shorter time scale. The system we have designed is nowhere near as robust to changes in the parameters of the system as is the Collins switch. It follows, then, that in order to design a successful system, we must identify the subspace of parameter space in which our circuit will operate correctly as a toggle switch and force the system to lie within this subspace. This effort requires modeling of the system behavior with respect to the parameters of the system, measurement of the natural values of the different parameters, and biophysically modifying these natural values to bring the system into the desired subspace of parameter space. We will be detailing each of these tasks in the following sections.

2. Modeling

2.1. Circuit Designs and Fundamental Considerations.

Three different circuit designs were considered for the mammalian toggle switch. These are shown on the next two pages. For each of these circuits, models were constructed using mass action kinetics, statistical mechanics, and time scale separation arguments. The underlying reasoning will be presented in the next few pages for each component of the circuits, considering each circuit design separately. The results of trajectory simulations and sensitivity analyses using estimates for relevant parameters, and possible conclusions will be discussed.

Date: July 30, 2010.

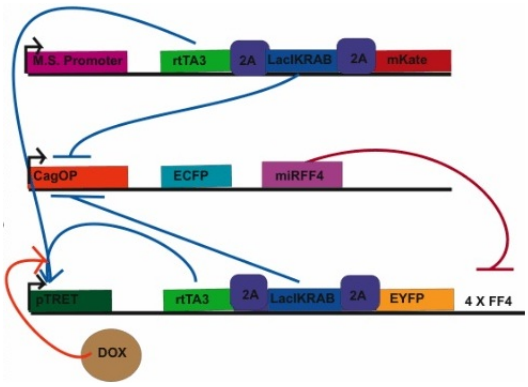

FIGURE 1. Robust Circuit A

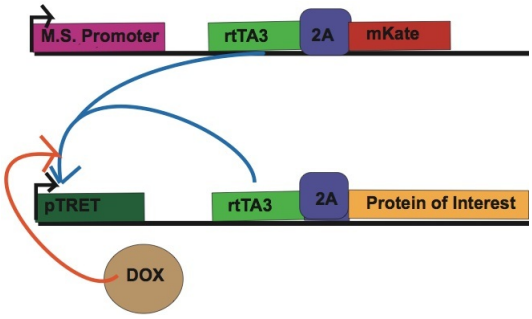

FIGURE 2. Sensitive Circuit B

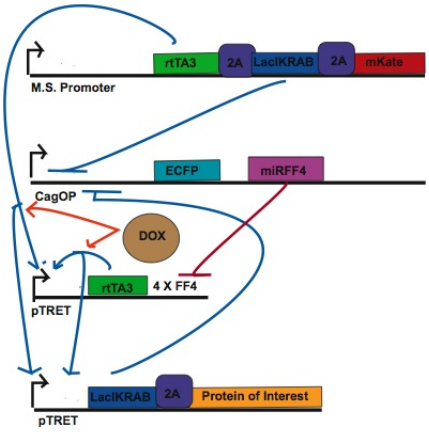

FIGURE 3. Very Robust Circuit

Some explanation as to the mechanism by which rtTA3 and Doxycycline activate the TRET promoter is in order. The TRET promoter contains 6 tetO sites, and it is on these sites that the transcription factors utilized

in this circuit bind. First, rtTA3 dimerizes in almost irreversible fashion, following which each monomer binds to a Dox molecule. This complex is then able to bind a pair of tetO sites on pTRET. Now, if this description were complete, the exact transfer function for promoter activity given Dox and rtTA3 concentration could be calculated exactly up to one undetermined constant. However, this process exhibits cooperativity in an interesting manner. In particular, pTRET is initially coiled around histones, with one 2 tetO sites easily accessible, and the binding of the first rtTA3-Dox complex causes a conformational change that allows for the next rtTA3-Dox complex to bind at a greater rate to the next pair of tetO sites. The consequences of this cooperativity will be discussed shortly, but for the moment we will simply acknowledge the existence of some transfer function $f([\text{rtTA3}],[\text{Dox}])$ that encodes the aforementioned cooperativity. This transfer function gives the rate of production of the relevant mRNA transcript. Similarly, a transfer function $g(P)$ exists for the mechanical stimulus promoter that will be finally chosen, and this encodes the rate of production, as a function of pressure applied, of its own mRNA transcript. Each mRNA transcript in this design actually contains two transcript joined by a 2A site. This site allows each mRNA to be transcribed as a single transcript; however, when the mRNA is being processed by a ribosome, the 2A sequence serves to separate the portion of amino acid sequence already translated from the ribosome, and in this way ensures the separate translation of each polypeptide involved. Thus, each of the four proteins has a different translation rate, while each of the two mRNA transcripts has a different transcription rate given by the relevant transfer function. Finally, all molecules in this system gradually degrade, at a rate linearly proportional to their own concentrations.

2.2. **Catalog of Reactions for Circuit B.**

$$2\ \text{rtTA3} \xrightarrow{k_1} \text{rtTA3}_2$$

$$\text{rtTA3}_2 + \text{Dox} \underset{k_{22}}{\overset{k_{21}}{\rightleftharpoons}} \text{rtTA3}_2 \bullet \text{Dox}$$

$$\text{rtTA3}_2 \bullet \text{Dox} + \text{Dox} \underset{k_{22}}{\overset{k_{21}}{\rightleftharpoons}} \text{rtTA3}_2 \bullet \text{Dox}_2$$

$$\text{rtTA3}_2 \bullet \text{Dox}_2 + \text{pTRET}_{\text{inact}} \rightleftharpoons \text{pTRET}_{\text{act}}$$

$$\text{M.S. Promoter} \xrightarrow{g(P)} \text{M.S. Promoter} + \text{mRNA}_1$$

$$\text{pTRET}_{\text{inact}} \xrightarrow{lr_2} \text{pTRET}_{\text{inact}} + \text{mRNA}_2$$

$$\text{pTRET}_{\text{act}} \xrightarrow{tr_2+lr_2} \text{pTRET}_{\text{act}} + \text{mRNA}_2$$

$$\text{mRNA}_1 \xrightarrow{trl_1} \text{mRNA}_1 + \text{rtTA3}$$

$$\text{mRNA}_1 \xrightarrow{trl_2} \text{mRNA}_1 + \text{mKate}$$

$$\text{mRNA}_2 \xrightarrow{trl_3} \text{mRNA}_2 + \text{rtTA3}$$

$$\text{mRNA}_2 \xrightarrow{trl_4} \text{mRNA}_2 + \text{Protein of Interest}$$

$$\text{mRNA}_1 \xrightarrow{\gamma_1} \emptyset$$

$$\text{mRNA}_2 \xrightarrow{\gamma_2} \emptyset$$

$$\text{rtTA3} \xrightarrow{\gamma_3} \emptyset$$

$$\text{mKate} \xrightarrow{\gamma_4} \emptyset$$

$$\text{Protein of Interest} \xrightarrow{\gamma_5} \emptyset$$

$$\text{rtTA3}_2 \xrightarrow{\gamma_6} \emptyset$$

$$\text{rtTA3}_2 \bullet \text{Dox} \xrightarrow{\gamma_7} \emptyset$$

$$\mathrm{rtTA3_2 \bullet Dox_2} \xrightarrow{\gamma_8} \emptyset$$
$$\mathrm{Dox} \xrightarrow{\gamma_9} \emptyset$$

One further clarification must be presented in order to completely characterize the functioning of pTRET. Since pTRET contains 6 tetO sites, it can have up to three transcription factors bound. We choose to assume that the transcriptional activity of pTRET is linearly proportional to the number of transcription factors bound. The binding of multiple transcription factors is not indicated in the catalog of reactions provided above, but it is included in the model.

Mass Action Kinetics, which can be applied to the first three reactions, indicates that the rate of the forward reaction in each case is given by the product of the forward rate constant and the concentrations of each reactant, while the rate of the backward reaction is given by the product of the backward rate constant and the concentrations of each product. For the processes of transcription and translation, the relevant rate constants signify the number of mRNA transcripts and polypeptides, respectively, produced per unit time. For the degradation processes, the rate at which a substance is depleted is given by the product of the corresponding degradation rate constant and the concentration of the substance being depleted.

2.3. **Time Scale Argument.**

Many of the rate constants involved are not known. Some of these we will be able to measure in a relatively straightforward, while for others we can at best extrapolate order of magnitude estimates based on similar systems. Thus, in order to construct a functioning model, we must use time scale arguments to separate the system into relevant modules. In particular, while transcription, translation, and degradation happen on the time scale of hours, the first four reactions listed occur on the time scale of milliseconds or smaller. Thus, since the output of our toggle switch occurs on the former time scale, the latter case can be considered to be effectively instantaneous. Thus, we can assume that the concentrations of species are such that the first four reactions are always at equilibrium. The results shown below follow from this approximation:

$$[\mathrm{rtTA3}] = 0$$
$$[\mathrm{rtTA3}]_{\mathrm{Total}} = [\mathrm{rtTA3_2 \bullet Dox_2}] + [\mathrm{rtTA3_2 \bullet Dox}] + [\mathrm{rtTA3_2}]$$
$$[\mathrm{rtTA3_2}] = \frac{[\mathrm{rtTA3}]_{\mathrm{Total}}}{1 + \frac{k_{21}}{k_{22}}[\mathrm{Dox}] + (\frac{k_{21}}{k_{22}}[\mathrm{Dox}])^2}$$
$$[\mathrm{rtTA3_2 \bullet Dox}] = \frac{k_{21}}{k_{22}}[\mathrm{rtTA3_2}][\mathrm{Dox}] = \frac{\frac{k_{21}}{k_{22}}[\mathrm{Dox}][\mathrm{rtTA3}]_{\mathrm{Total}}}{1 + \frac{k_{21}}{k_{22}}[\mathrm{Dox}] + (\frac{k_{21}}{k_{22}}[\mathrm{Dox}])^2}$$
$$[\mathrm{rtTA3_2 \bullet Dox_2}] = (\frac{k_{21}}{k_{22}})^2[\mathrm{rtTA3_2}][\mathrm{Dox}]^2 = \frac{(\frac{k_{21}}{k_{22}}[\mathrm{Dox}])^2[\mathrm{rtTA3}]_{\mathrm{Total}}}{1 + \frac{k_{21}}{k_{22}}[\mathrm{Dox}] + (\frac{k_{21}}{k_{22}}[\mathrm{Dox}])^2}$$

Since $[\mathrm{rtTA3}]_{\mathrm{Total}}$ does not give the concentration of a unique chemical species, it has no single associated degradation rate constant. However, since $[\mathrm{rtTA3}]_{\mathrm{Total}} = [\mathrm{rtTA3_2 \bullet Dox_2}] + [\mathrm{rtTA3_2 \bullet Dox}] + [\mathrm{rtTA3_2}]$, the degradation rate associated with $[\mathrm{rtTA3}]_{\mathrm{Total}}$ is simply the sum of the three relevant degradation rates, i.e.

$$[\mathrm{rtTA3}]_{\mathrm{Total}} \frac{\gamma_6 + \gamma_7 \frac{k_{21}}{k_{22}}[\mathrm{Dox}] + \gamma_8(\frac{k_{21}}{k_{22}}[\mathrm{Dox}])^2}{1 + \frac{k_{21}}{k_{22}}[\mathrm{Dox}] + (\frac{k_{21}}{k_{22}}[\mathrm{Dox}])^2}$$

In fact, since each form of the rtTA3 dimer can be expressed in terms of the total concentration as shown here, we need only deal with this total concentration in our model. Now, it is the $\text{rtTA3}_2\text{Dox}_2$ molecule that binds to pTRET. Unfortunately, mass action kinetics cannot be applied to analyze this situation. Mass Action Kinetics assigns rates to reactions by assuming continuity of depletion and addition of molecules. However, since molecules are in fact discrete, their rate of change cannot in actuality be continuous, and for this reason mass action kinetics can only be validly applied when the concentrations of reagants involved are relatively large. Ideally, pTRET would only be present in one copy in each mammalian cell, but even supposing multiple lentiviral insertions into the mammalian genome, the concentration of pTRET would be tiny on a molar scale, and so mass action kinetics cannot be validly applied. Surprisingly, ignoring the rigor of such considerations, mass action kinetics actually produces the same equilibrium results as an analysis using statistical mechanics for the binding of pTRET.

2.4. **Derivations of Promoter Binding Transfer Function.**

2.4.1. *Statistical Mechanics.* For the purposes of this analysis, we will not specify whether we are considering only the nucleus of the cell or the whole volume. In particular, if the transcription factor is always localized to the nucleus, then we are considering the volume of the nucleus, but otherwise we are considering the volume of the whole cell. Now, let the number of transcription factor molecules be N. Let us discretize the relevant volume and consider it as a lattice with $M >>> N$ free sites. Finally, let us consider first the simpler case of a single binding site for a transcription factor molecule, and let the bound energy be $-E$, while the free energy is 0. Using the canonical ensemble, the probability that a TF is bound in equilibrium is $\frac{\binom{M}{N-1}e^{\frac{E}{kT}}}{\binom{M}{N-1}e^{\frac{E}{kT}}+\binom{M}{N}} = \frac{1}{1+\frac{M-N+1}{N}e^{-\frac{E}{kT}}} \approx \frac{\frac{N}{M}}{\frac{N}{M}+e^{-\frac{E}{kT}}}$. Since $\frac{N}{M} \propto [\text{TF}]$, we find that the proportion of time that the promoter is bound is given by the sigmoidal curve $\frac{[\text{TF}]}{[\text{TF}]+K_M}$, where K_M is a scale constant that increases as E decreases and as T increases.

Now, we must consider the given scenario, in which there are three sequential promoter binding sites. In such a case, we have three binding energies, $-E_1 > -E_2 > -E_3$, corresponding respectively to having one, two, and three sites bound. Performing the same calculation with similar approximations gives the average number of transcription factors bound to be $x\frac{a+2bx+3cx^2}{1+ax+bx^2+cx^3}$ where $x = [\text{TF}], a \propto e^{\frac{E_1}{kT}}, b \propto e^{\frac{E_2}{kT}}, c \propto e^{\frac{E_3}{kT}}$. Given the accuracy to which these parameters can usually be fit, this expression is usually approximated as $\frac{[\text{TF}]^n}{(K_M)^n+[\text{TF}]^n}$, for some fitted paramaters $n < 3, d$. We will most likely try both functional forms to interpret experimental data. Finally, our time scale argument allows us to assume that the promoter binding reaction is always in equilibrium, and thus it follows that the rate of transcription from pTRET is given by $f([\text{rtTA3}]_{\text{Total}}, [\text{Dox}]) = lr_2 + tr_2\frac{[\text{TF}]^n}{(K_M)^n+[\text{TF}]^n}$ where $[\text{TF}] = \frac{(\frac{k_{21}}{k_{22}}[\text{Dox}])^2[\text{rtTA3}]_{\text{Total}}}{1+\frac{k_{21}}{k_{22}}[\text{Dox}]+(\frac{k_{21}}{k_{22}}[\text{Dox}])^2}$.

We would like to make one final point concerning how to deal with multiple lentiviral insertions. Let $Z(x = \frac{N}{M}) = \binom{M}{N}\sum_{i=0}^{\infty} a_i x^i$ be the partition function for a given promoter-binding system corresponding to one insertion. Clearly, $a_i = e^{\frac{E_i}{kT}}$ where $-E_i$ is the energy associated with the first i bound transcription factors. In particular, this is true because the term $a_i x^i$ in the sum corresponds to the boltzmann factor for the first i bound transcription factors, a_i, multiplied by the number of possible arrangements which satisfy this constraint, i.e. $\binom{M}{N-i}$, all divided by $\binom{M}{N}$, thus giving $x \approx \frac{N}{M}$. Now, for a system with K insertions,

$Z_K(x) = \binom{M}{N} \sum_{i=0}^{\infty} b_i x^i$. Solving for b_i we find that $b_i = \sum_{\sum_{j=1}^{K} \sigma(j)=i} \prod_j a_{\sigma(j)}$. However, this is equivalent to stating that $Z_K(x) = (Z(x))^K$. Since the average number of transcription factors bound is given by $\frac{d}{dx} \ln Z_K(x) = K \frac{d}{dx} \ln Z(x)$, i.e. the number of transcription factors bound scales linearly with the number of lentiviral insertions, as is perhaps intuitively obvious.

2.4.2. *Mass Action Kinetics.* Suppose mass action kinetics was applicable in this scenario, and let $a, \frac{b}{a}, \frac{c}{b}$ be the equilibrium constants for the first, second and third bindings, respectively, of the transcription factor to pTRET. Then, simply by considering the concentrations of each possible state of pTRET, we find that the average number of transcription factors bound to pTRET at equilibrium is $x \frac{a + 2bx + 3cx^2}{1 + ax + bx^2 + cx^3}$, where $x = [\text{TF}]$. This is identical to the result derived above, and the remaining calculations are identical as well.

2.5. **System Time Scale and Sensitivity.**

Based on the order of magnitude estimates available to us, we have been able to estimate how long the toggle switch should take to successfully switch between steady states. This will be used to actually demonstrate the functionality of the toggle switch. However, we wish to fully characterize the system, and so in the Section 3, we will discuss experimental determination of the parameters involved here. To conclude the modeling section for Circuit B, Mathematica was used to run trajectory simulations of the system in order to determine the time scale on which the system would function. Furthermore, a sensitivity analysis was performed using the Manipulate functionality in Mathematica, and using this the effect of the parameters on the overall system topology was observed. The code for both operations is given below, as are the results of the trajectory simulations. The sensitivity analysis could unfortunately not be ported into LaTeX and so is available upon request as a Mathematica Notebook File.

3. Experimental Determination of Parameters

There are four groupings of parameters, and we will address our methods for determining each of them in succession. These four groupings are fast chemical reactions, transcriptional activity, translational activity, and degradation. The order in which these are determined is relevant, for the determination of one grouping may depends on the results for a previous grouping. Thus, we will present them in the order we intend to follow. First, we will discuss degradation.

For all degradation measurements, we will use wildtype mammalian cells that have not been transformed with lentivirus. The measurement is simplest for Dox, since Dox diffuses freely across cell membranes. We will set up 20 wells, each filled with a fixed concentration of Dox. Then we will place a fixed number of mammalian cells in each well. Every 2 hours, we will remove the cells in a given well, lyse them, measure their intracellular concentration of Dox using mass spectrometry or ELISA, both techniques to measure concentrations of proteins in cells, and record how long it takes for the Dox concentration to reach half its initial level. Thus will give us the half life τ of Dox in mammalian cells, and it follows from the mathematics of exponential decay that the relevant degradation factor is $\gamma = \frac{\ln 2}{\tau}$. We will use this same procedure repeatedly to determine all the relevant degradation constants, with just a few modifications depending on the molecule under consideration. Dox is the only diffusible molecule in our system, and for this reason we cannot simply place cells in solutions of the other substances involved. Rather, we must transfect them into the mammalian cells. For proteins, the transfection technique that best suits our needs is to mix the protein in cationic lipid mixture and allow the resultant vesicles to be taken up by the mammalian cells. By precisely determining the efficiency of

such a transfection, the procedure outlined above can be followed to determine the relevant degradation rate constant. We note that it might be difficult to assay separately for the different forms of rtTA3 present, since an antibody for rtTA3_2 would likely also bind to $\text{rtTA3}_2 \bullet \text{Dox}$ and $\text{rtTA3}_2 \bullet \text{Dox}_2$. If we are fortituous, the degradation rates for these three forms will be nearly equal, and so it might suffice to simply determine a degradation rate for $\text{rtTA3}_{\text{Total}}$. For RNA transfection, more standard methods, such as electroporation or sonication, might suffice, and again we will be able to determine the relevant degradation rate constant.

Next we will want to measure translational parameters. To do this, we will use the same experimental structure as for the degradation measurements, with wells and discrete time sampling. However, when we transfect our cells with the relevant mRNA, we will measure the concentration of the protein encoded by the mRNA. Our measurements will not be entirely straightforward, unfortunately; they will be confounded by the simultaneous degradation of the transfected mRNA and the degradation of the protein product itself. Luckily, we will already know both of these degradation rates, and so we can easily correct for these and accurately calculate the translation rate. Incidentally, if we were to measure all protein degradation rates first, we could combine the degradation measurements for the mRNA and the translation measurements for the same mRNA into one experimental run.

Once we have determined the translation rates, we will proceed to measure transcriptional activity. To do this, we will transform mammalian cells with a construct pairing EGFP to the promoter in question. Then we will take the transformed cells and grow them under different concentrations of transcription factors (which for us includes pressure). We will use confocal microscopy to measure continuously the production of EGFP and use this to determine the transfer function for each promoter. We will have to account for mRNA degradation, protein degradation, and translation rates, but fortunately we will already have all these measurements.

We come now to the measurement of forward and reverse reaction rates. Initially, we had declared these to be very fast, on the order of milliseconds or smaller, and so had dismissed them as instantaneous for our purposes. Indeed, very few measurements have been made of reaction rates in the biological and chemical sciences because of their speed; at best, the equilibrium constants have been measured. The latter is a fairly trivial task, requiring the measurement only of final equilibrium concentrations of the reactants and products. We outline here a general method for measuring reaction rates, as developed by Ky-Anh Tran, another undergraduate physics student in MIT's Class of 2013, and Arvind Thiagarajan on July 30th. We will describe the theory behind the method as well as potential caveats.

The method we present was motivated by the activity of fluorescent proteins. In particular, we noted that if all reactants and products in a reaction were fluoroscent, it would be fairly easy to measure fluorescence continuously, and from these measurements determine the rate constants. However, absorption spectra are as easy to measure as emission spectra, and all substances absorb electromagnetic radiation. We posited that a homogenous solution of some chemical A would have absorbance that increased linearly with concentration, i.e. $a([\text{A}], \nu) = k_A(\nu)[\text{A}]$ for some function $k_A(\nu)$. Furthermore, we posited that a solution containing multiple chemicals A_i would have absorbance given by $a = \dfrac{\sum_i k_{A_i}(\nu)[\text{A}_i]^2}{\sum_i [\text{A}_i]}$, i.e. an average of the absorbances of the constituent chemicals weighted by the concentrations of the chemicals. Each of the $k_{A_i}(\nu)$ can be measured in advance, since they are static in time. Finally, for any given reaction involving the chemicals A_i, a general trajectory for the different concentrations can be determined theoretically in terms of the forward and reverse reaction rates, and if these concentrations are substituted into the expression for the absorbance of the mixture, a theoretical expression for absorbance as function of time can be obtained. Since this absorbance can also be measured, and since the equilibrium constant for the reaction is either known or measured easily, we need only to fit one rate to match the theoretical expression with the experimental curve. In this way we can determine the forward reaction rate, and using this and the equilibrium constant, the reverse reaction rate as well.

It is unclear how valid some of the assumptions we have made are. In particular, it is unclear if the absorbance actually scales linearly with concentration, or whether the absorbance of a mixture is in fact the

weighted average of the absorbances of the constitutent chemicals. Fortunately, both of these assumptions can be checked empirically, and corrected as necessary. If we have enough time, we will attempt to measure fast reaction rates using this method.