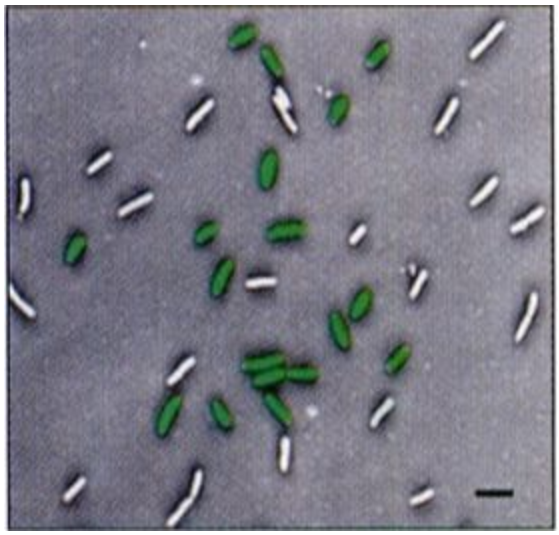
MATH 485

Multistability in Biological Systems



Group Members: Michael Volk, Rob Wallace, Katherine Smith, & Lauren Nakonechny

Mentor: J. Ruby Abrams

University of Arizona Department of Mathematics

Abstract

This report will summarize the results that Ovbudak et al. discovered in their paper entitled "Multistability in the lactose utilization network of *Escherichia coli*"[1]. These researchers varied the extracellular concentrations of glucose and TMG in *E. coli* cells and the resulting amounts of GFP and HcRed produced were measured. This gave them ample information to produce a phase diagram of the wild-type lactose utilization network of *E. coli*. After creating a phase diagram of the wild-type network, they attempted to mathematically model the lactose utilization network system. This was an attempt to show that the hysteretic nature of the wild-type network could be altered, with the result being a theoretical model that included a graded response.

Introduction - Biological Background and Multistability

The *lac* operon is a segment of DNA found in *Escherichia coli* that is responsible for regulating the metabolism and uptake of lactose. In general, glucose is *E. coli*'s preferred source of carbon; however, in the absence of glucose, *E. coli* is capable of metabolizing lactose due to the expression of the enzymes coded by the *lac* operon.

The *lac* operon is composed of several genes which code for various enzymes. The three gene segments, *LacZ, LacY, and LacA,* all code for enzymes that allow for the metabolism and uptake of lactose. *LacY* codes for lactose permease (also referred to as LacY, without italics), which allows for the uptake of thiomethylgalactoside (TMG), a molecule similar to lactose, through the cellular membrane. *LacA* and *LacZ* code for enzymes acetyltransferase and beta-galactosidase, respectively, which break down molecules into simple sugars that are easily absorbed by the body. The promoter is the site on the *lac* operon where RNA polymerase binds to begin transcription. The operator is the site where repressors, such as the LacI protein, bind to and inhibit the transcription of the *lac* operon (and therefore turn off lactose metabolism). The repressor protein, LacI, is itself inhibited by TMG and allolactose.The catabolite activator protein (CAP) binds to the CAP site to activate transcription of the *lac* operon in the presence of cyclic AMP (cAMP). (Note that the catabolite activator protein (CAP) may also be referred to as the cyclic AMP receptor protein (CRP).)

When glucose is present, cAMP levels are low. Since there is sufficient glucose in the system, the *lac* repressor (LacI) binds to the operator, preventing the transcription of the *lac* operon.

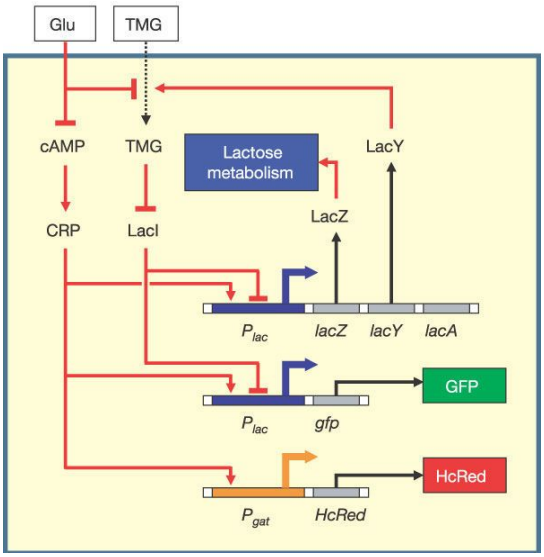


Figure 1: (Ozbudak et al.)

However, when glucose is low in the cell, cAMP levels are high. Allolactose, a form of lactose, and TMG bind to the repressor protein, releasing it from the operator and allowing RNA polymerase to transcribe. The cAMP molecules bind to the CAP, which attaches to the CAP binding site and activates high levels of transcription of the *lac* operon. In Figure 1, the lines with arrows denote induction pathways and the lines with blunt ends denote inhibition. cAMP binding to CAP triggers activation of the *lac* operon. HcRed and GFP are fluorescent reporter proteins that is expressed at the gat and lac promoters, respectively, are used in this experiment as direct measures of the levels of CRP and cAMP in the cell. Higher amounts of glucose inhibit cAMP-CRP, and therefore inhibit the production of GFP and HcRed (red fluorescent protein). Conversely, TMG binding inhibits LacI, the inhibitor of the *lac* operon, and therefore induces the production of GFP, a fluorescent reporter protein that is found at the *lac* promoter. TMG binding stimulates the synthesis of lactose permease (LacY), which leads to the uptake of more TMG and facilitates the uptake of more lactose. There is a positive feedback loop present, where greater concentrations of TMG produce more LacY, and subsequently imports more TMG.

The positive feedback loop described above is essential to the backbone of this experiment, because it creates the potential for multistability; more specifically, the lactose utilization network of *E. coli* expresses bistability. However, it is important to mention that the validity

of this system depends on having cells with well-defined initial states, as the bistable region has hysteretic behavior. Each cell must have been either never induced, or fully induced, as the system response is dependent upon its history.

Theoretically Modeling Positive Feedback and Bistability

In order to model the bistability of the *lac* operon, three equations are used. The first equation models the relationship between the concentration of LacI (the repressor protein) and the intracellular concentration of TMG. This equation denotes the active fraction of LacI in the system.

$$\frac{R}{R_T} = \frac{1}{1+(x/x_0)^n}$$

Equation (1): (Ozbudak et al.)

R is the concentration of active LacI, R_T is the total concentration of LacI, x is the intracellular concentration of TMG, x_0 is the half-saturation of TMG, and the exponent n is the Hill coefficient.

Equation (1) behaves as a decreasing sigmoidal function of x. This is the case because even the smallest amount of binding of TMG to LacI will interfere with its inhibitory activity, and as more TMG binds, the level of inhibition of the *lac* operon increases.

The second equation gives the rate of generation of lactose permease (LacY). Recall that as TMG binds, LacY is expressed and facilitates the uptake of more TMG. This makes up the positive feedback loop. Equation (2) shows that the generation of LacY is a decreasing hyperbolic function of LacI.

$$\tau_y \frac{dy}{dt} = \alpha \frac{1}{1+R/R_0} - y$$

Equation (2): (Ozbudak et al.)

In equation (2), *y* is the concentration of LacY, $\boldsymbol{\tau}_y$ is a time constant, α is the maximum value of growth of LacY. The minimal value achieved is α/ρ, where $\rho = 1 + R_T/R_0$, which is the repression factor. The repression factor describes how well LacI can regulate expression of the *lac* operon.

The third equation gives us the rate of change of the intracellular concentration TMG.

$$\tau_x \frac{dx}{dt} = \beta y - x$$

Equation (3): (Ozbudak et al.)

Here, β is the measure of TMG uptake per LacY molecule. TMG enters the cell at a rate proportional to the concentration of LacY in the cell, and it is diminished in a first order reaction with time constant $\boldsymbol{\tau}_x$.

These equations may be combined to retrieve the steady state result:

$$y = \alpha \frac{1 + (\beta y)^2}{\rho + (\beta y)^2}.$$

Equation (4): (Ozbudak et al.)

ρ, α, and β are arbitrary functions of inputs of glucose (G) and TMG (T). As these three arbitrary parameters are varied, the system responds with either one or two stable fixed points. These fixed points are separated by saddle node bifurcations.

Equation (4) can be rewritten as a cubic equation, as follows:

$$y^3 - \alpha y^2 + (\rho / \beta^2) y - (\alpha / \beta^2) = 0.$$

Equation (5): (Ozbudak et al.)

To attack this, it should be recalled how to deal with general cubic equations with two identical roots. The general cubic can be written in the following form:

$$\begin{aligned}(y-a)(y-a)(y-\theta a) \quad &= \\ y^3 - (2+\theta)ay^2 + (1+2\theta)a^2 y - \theta a^3 &\end{aligned}$$

In the above equation, θ is the dimensionless ratio of roots. By rewriting Equation (5) in this form and comparing coefficients, we can define our arbitrary functions.

$$\rho = (1+2\theta)(1+2/\theta),$$
$$\alpha\beta = (2+\theta)^{3/2}/\theta^{1/2}.$$

Equations (6) and (7): (Ozbudak et al.)

Interestingly enough, Equations (6) and (7) denote the boundary of the bistable region.

Measuring Network Parameters

As mentioned earlier, ρ, α, and β are arbitrary functions of inputs of glucose (G) and TMG (T). These parameters have physical meanings that are crucial to the interpretation of this mathematical model. Their exact derivations are steeped in experimental data, and are thus outside the scope of this report. However, some information on *how* they were derived by Ozbudak et al. will be provided.

The physical meanings of the network parameters are fairly simple:

- α refers to the level of *lac* operon expression that would be seen if every repressor molecule were inactivated. α is the maximum induction level.
- ρ refers to the ratio of the maximal induction level to the basal induction level. The basal induction level is the level of *lac* operon expression that would be see if every repressor molecule were activated. ρ is the repression factor.

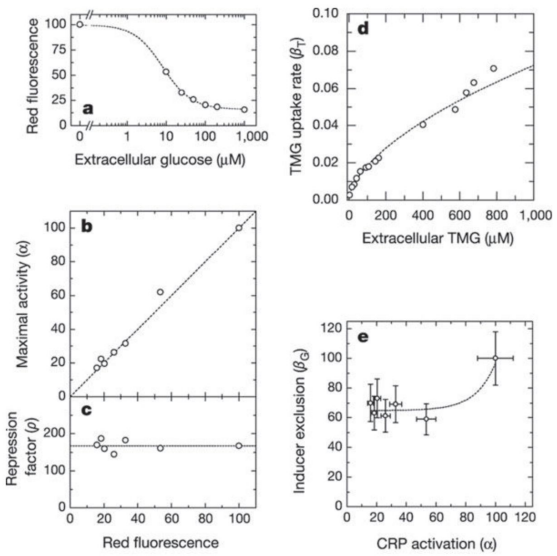


Figure 2: (Ozbudak et al.)

- β refers to the rate of TMG uptake per molecule of LacY.

Calculating the functional equations of these three parameters required numerous single-cell experiments, where their dependence on glucose and TMG levels were measured. See Figure 2 for the results of these *in vivo* experiments.

These calculations also required the application of some caveats. First, the saddle node condition was applied at each boundary of the bistable region; this was done separately at what was referred to as ON (induced) and OFF (uninduced) regions. Second, it was taken into account that α was approximately 15% higher at the

OFF threshold than the ON threshold. Third, due to low fluorescence values in the OFF region, there was a large error in the calculation of ρ. Therefore, the authors decided to estimate both α and ρ at the ON threshold only. Lastly, the net TMG uptake rate was decomposed as:

$$\beta(T,G) = \beta_T(T)\beta_G(G)$$

Recall that *T* refers to TMG concentrations and *G* refers to glucose concentrations. The power law was assumed for the first half of the equation, and the least-square fitting technique was used to extract the necessary functions.

It was determined that:

$$\alpha = \frac{84.4}{1+(G/8.1)^{1.2}} + 16.1, \ \rho = 167.1,$$
$$\beta_T = (1.23\times10^{-3})T^{0.6}, \ \beta_G(G>10) \cong 65.$$

Equations (8) and (9): (Ozbudak et al.)

Phase Diagrams

The analysis of this system results in two important phase diagrams. The first one, included in Figure 3, was generated after the completion of a series of experiments by Ozbudak et al. This is the wild-type network phase diagram. Ozbudak et al. generated this phase diagram by measuring the behavior of cell populations for *lac* operon expression as levels of TMG were varied in the system. The cells had well-defined initial states (either uninduced or fully

induced), and were grown in media with different TMG levels. Green and red fluorescence were used to measure expression of the *lac* operon.

This phase diagram is important because it maps for which TMG and glucose levels the system is bistable. It shows that induction of the *lac* operon always takes place hysteretically (see the grey bistable region that cuts through the graph). The cells increase and decrease their expression of the *lac* operon discontinuously as they cross the boundaries of the bistable region (the so-called "switching thresholds"). In other words, initially uninduced cells are only turned on if the TMG concentration is *above* a certain level, and initially induced cells are only turned off if the TMG concentration is *below* a certain level. This tells us that the response of the system to external conditions is dependent upon the system's history.

The second phase diagram is the theoretical one derived from the mathematical model (Figure 4). This phase diagram shows us that cells may shift from being uninduced to induced, or vice versa, hysteretically or in a graded fashion. In the bistable region, the system is hysteretic and behaves in the same way as the wild-type system. However, this region of bistability decreases as the repression factor (degree of *lac* operon expression), ρ, decreases. At a critical factor of ρ=9, the bistable region hits a cusp and comes to an end, and the system response occurs in a graded fashion. That is, expression of the *lac* operon in each cell can move continuously between low and high values. This can be seen in the white regions of the graph.

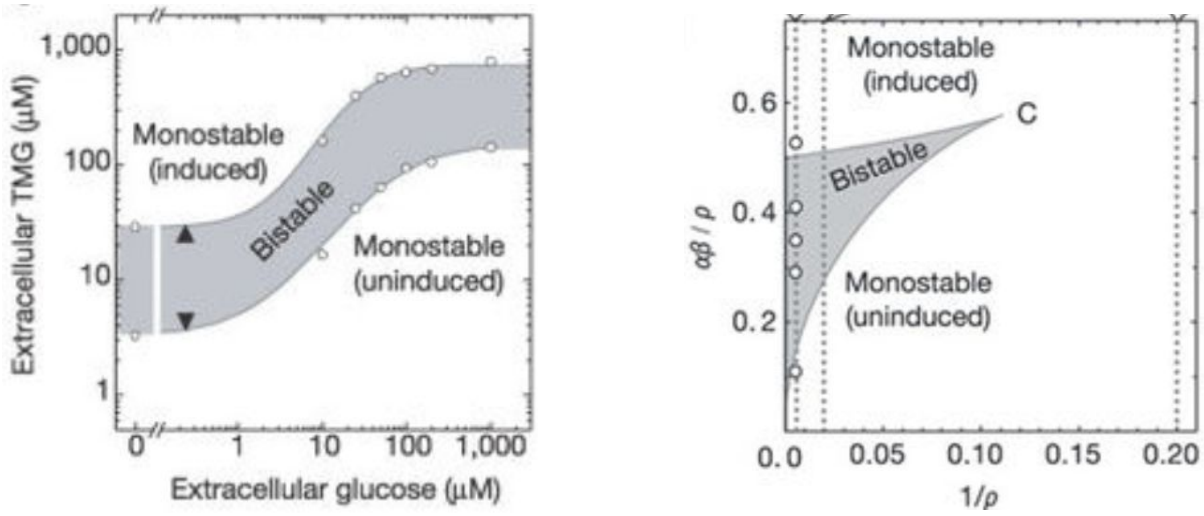


Figure 3: (Ozbudak et al.)

Figure 4: (Ozbudak et al.)

Current Progress and Future Work

To begin understanding the mathematical model outlined by Ozbudak et al., the figures depicted in "Multistability in the lactose utilization network of *Escherichia coli*"[1] were recreated. The figure of greatest interest, the theoretical phase diagram, as depicted in Figure 4, was recreated in exact form. In order to recreate the phase diagram, Equation (6) giving ρ and Equation (7) giving αβ were vectorized in MatLab over a range of possible θ values. The plot in Figure 5 was created and compared to Figure 4.

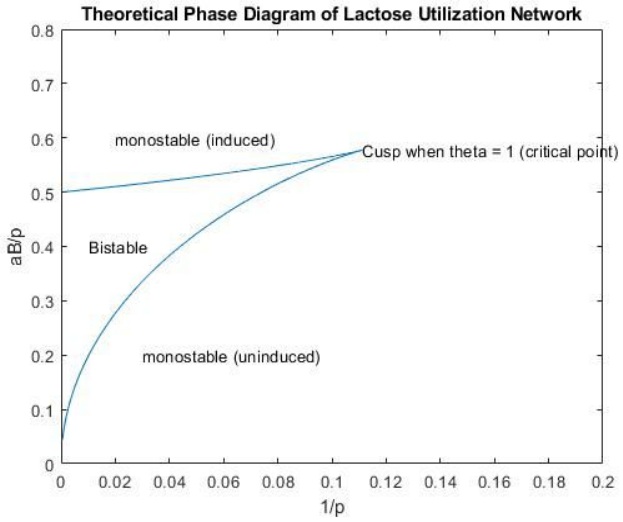


Figure 5.

In order to take the analysis a step further, the parameters from the steady state result in Equation (4) were reduced from three parameters to two parameters. The process of the reduction of the parameters follows:

$$y = \alpha \frac{1+(\beta y)^2}{\rho+(\beta y)^2}$$ Variable Change: $y = \alpha z$

$$\alpha z = \alpha \frac{1+(\alpha\beta z)^2}{\rho+(\alpha\beta z)^2}$$

$$z = \frac{1+(\alpha\beta z)^2}{\rho+(\alpha\beta z)^2}$$

Divide by: $\frac{\rho^2}{\rho^2}$

$$z = \frac{\frac{1}{\rho^2}+(\frac{\alpha\beta}{\rho}z)^2}{\frac{1}{\rho}+(\frac{\alpha\beta}{\rho}z)^2}$$

Redefine variables: $\mu = \frac{1}{\rho}$;

$$\lambda = \frac{\alpha\beta}{\rho}$$

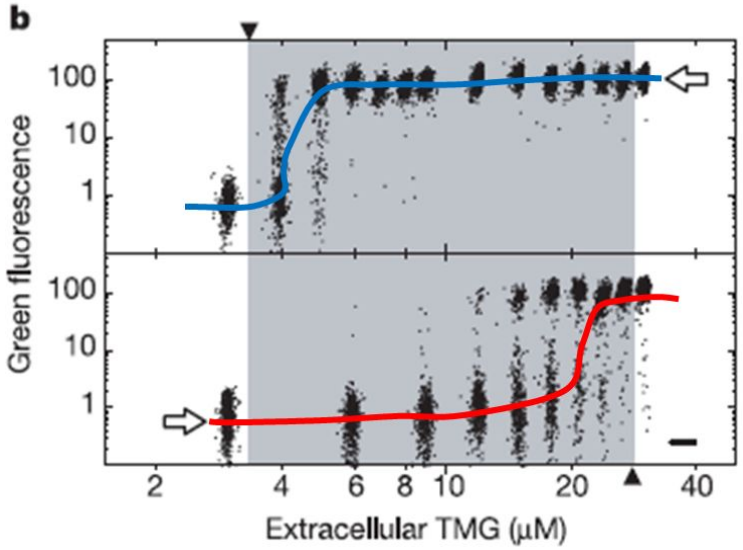


Figure 7: (Ozbudak et al.)

$$z = \frac{\mu^2+(\lambda Z^2)}{\mu+(\lambda Z^2)}$$

z is only dependent on two parameters

With the steady state result only dependent on two parameters, a contour plot was created, with z along the vertical axis and λ and μ along the horizontal axes. The contour plot is given in Figure 6. With the plot, bistability can be viewed more easily. The folded region indicates that there are multiple z values for specific parameters

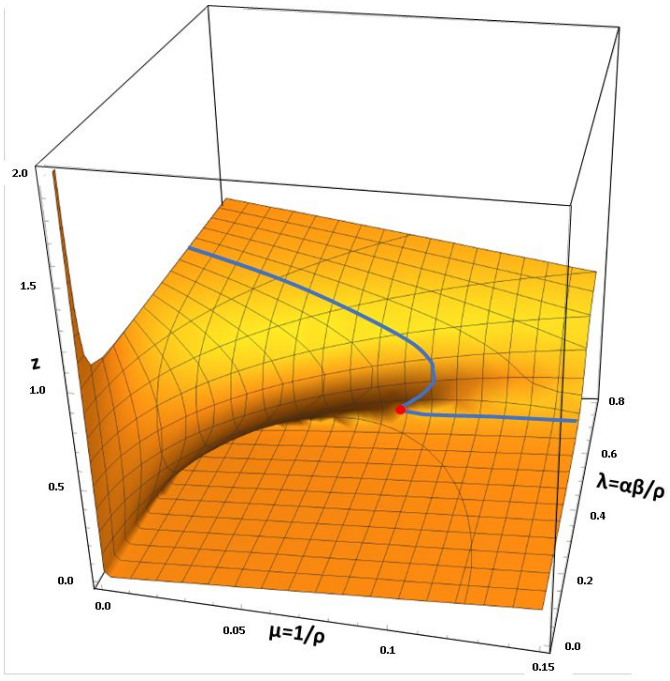


Figure 6.

of λ and μ. The blue line, depicts how a given trajectory changes with changing values of μ. The red dot depicts the inflection point in the trajectory where the blue line doubles back on itself.

In continuation of our current work trajectories from the provided differential equations will be mapped onto figure 2c from Ozbudak et al. We expect the trajectories from the theoretical model to follow paths similar to the paths roughly outlined by the blue and red lines. Trajectories will be overlaid on the figure and numerically compared with data to show consistency between the model and experimental results, thus showing that the system behaves hysteretically.

Further into the future we plan to compare the mathematical model provided by Ozbudak et al. with other first order phase transitions systems. We are specifically interested in comparing evaporation in a liquid gas system to the lac operon system, to see if mathematical trends in physical systems can be easily compared to the trends in biological systems. Ozbudak et al. made the claim that these systems behave similarly, but we would like to prove this statement. This exploration will hopefully provide some insight to future explorations in mathematical biology.

Sources

Ozbudak, Ertugrul M., Thattai, Mukund, Lim, Han N., Shraiman, Boris I., & van Oudenaarden, Alexander. Multistability in the lactose utilization network of *Escherichia coli. Nature.* **427,** 737-740 (2004).[1]

Hansen, L. H., Knudsen, S. & Sorenson, S. J. The effect of the *lacY* gene on the induction of IPTG inducible promoters, studied in *Escherichia coli* and *Pseudomonas fluorescens*. *Curr. Microbiol.* **36**, 341-347 (1998).

Yagil, G. & Yagil, E. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.* **11**, 11-27 (1971).

Khan Academy - *lac* operon
Nature Education - *lac* operon