2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Modeling Gene Expression: *Lac* operon

Sarai Velazco, Delina Kambo, Kevin Yu, Anushka Saha, Emily Beckman, Nishant Mysore and Gert Cauwenberghs

***Abstract*--- Gene regulation is an essential process for cell development, having a profound effect in dictating cell functions. Bacterial genes are often regulated through inducible systems like the Lac operon which plays an important role in cell metabolism. An accurate model of its regulation can reveal the dynamics of gene expression. In this paper, a mathematical model of this system is constructed by focusing on regulation by the Lac repressor. The results show, as expected, that the concentration of lactose approaches zero while glucose concentration approaches the initial concentration of lactose by the action of β-galactosidase, expressed by the *Lac* operon. Addition of PD control improves stability of the system, with the phase margin increasing from 45° to 90°. Modeling the dynamics of gene expression in inducible operons like Lac operon can be essential for its applications in the production of recombinant proteins and its potential usage in gene therapy.**

I. Introduction

Gene expression is the fundamental process which shapes the development of organisms by controlling the spatial and temporal expression of genes into their protein products. For single-cell organisms like bacteria, regulation of gene expression plays a vital role in response to changes in their physical and nutritional environment. The mechanism of gene expression is regulated by complex pathways, an example would be that of *Lac* operon, one of the first prokaryotic gene expression systems to be well understood.

In bacteria like *Escherichia coli* the main source of energy is glucose; however, when glucose is unavailable bacteria utilize lactose as a source of energy. The *Lac* operon consists of a set of three genes encoding for enzymes that are required for lactose metabolism: β-galactosidase (*Lac* Z), which catalyzes lactose into glucose and galactose, β-galactoside permease (*Lac* Y) and galactoside-acetyltransferase (*Lac* A). The expression of these three genes is controlled by a single promoter and is regulated via two main pathways: glucose regulation of *Lac* operon by catabolite activator protein (CAP) and regulation via the *Lac1* repressor protein[9]. This paper will focus on the second regulatory pathway.

Department of Bioengineering, UC San Diego, La Jolla, CA 92093, USA.

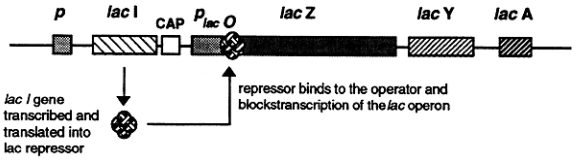

Figure 1: Overview of the Lac operon[10].

When glucose is present, there is no need to catalyze lactose to glucose. Therefore, cAMP levels are low and CAP protein is not activated so it does not bind to the regulatory site, resulting in low levels of transcription of the *Lac* operon. In addition, the Lac repressor binds to the promoter which blocks RNA polymerase from transcribing the operon as it is shown in Figure 1.

Under low levels of glucose, the CAP protein increases the transcription of the operon by binding to the CAP site. Also, lactose is converted to allolactose with the help of intramolecular levels of enzyme β-galactosidase. When allolactose binds to the repressor, it causes a conformational change to prevent the repressor from binding, allowing RNA polymerase to bind to the promoter and initiate transcription[9]. Lower levels of lactose prevents the formation of allolactose, causing the repressor to rebind to the promoter halting transcription of the *Lac* operon.

To fully model the dynamics of *Lac* operon, governing equations for the gene expression and chemical kinetics must be defined separately, before recombining the two together. A generalized model of gene expression is defined by Chen, et. al.[1]:

$$\frac{dr}{dt} = C[p] - V[r] \quad (1)$$

$$\frac{dp}{dt} = L[r] - U[p] \quad (2)$$

Where r is mRNA concentration and p is enzyme concentration. C represents the transcription rate, V represents the mRNA decay rate, L represents the translation rate, and U represents the protein decay rate. This model simplifies the rates of transcription and translation by assuming them to be constant. In addition, it is intended to be used for systems in which the protein output induces gene expression itself. The behavior of the *Lac* operon is similar in that the expressed enzyme β-galactosidase interacts with lactose to make allolactose, which induces continued transcription of the operon. Adapting the

Chen model for this type of expression, obtains the following system:

$$\frac{dr}{dt} = A[L] - V[r] \quad (3)$$

$$\frac{dZ}{dt} = B[r] - U[Z] \quad (4)$$

Here, the protein, or Z in (4) is defined to specifically refer to the enzyme β-galactosidase. The variable r refers to the mRNA encoding β-galactosidase, U is the protein decay rate, and A and B refer to transcription and translation rates respectively. Because lactose (as allolactose) plays a role in changing the conformation of the *Lac* repressor, the increase in mRNA concentration relates directly to the concentration of lactose (L).

The governing equations for chemical kinetics are easier to define. The catabolism of lactose can be summarized as:

Lactose + β-gal. -> Glucose + Galactose + β-gal.

The following differential equations model this second-order reaction as:

$$\frac{dL}{dt} = -k[Z][L] \quad (5)$$

$$\frac{dG}{dt} = k[Z][L] \quad (6)$$

Here, the conversion of lactose (L) into glucose (G) occurs at a 1:1 ratio. These conversions occur at the same rates. Thus the same rate constant k represents the consumption of lactose and the synthesis of glucose at steady state, one converted completely to the other.

II. Methods

To model the dynamics of the *Lac* operon, the following assumptions were made:

1. There is zero or minimal initial presence of glucose. The maximum expression of *Lac* operon occurs under this condition.
2. Glucose and galactose do not affect the system dynamics. In reality, the *Lac* operon has glucose-dependent regulation through CAP, but only when glucose is able to accumulate in the cell. From Assumption 1, the glucose is very quickly metabolized by other processes and thus does not accumulate.
3. All lactose is instantly converted to allolactose which binds to the *Lac* repressor, releasing it from the promoter and allowing transcription to initiate. Thus, they can be represented using the same variable.
4. The input is a small disturbance from steady-state, allowing for linearization of the biosystem.
5. Steady-state values of enzyme and mRNA are assumed to be zero. Though in reality, there is a small leaky expression, the dynamics of this model will be minimally affected; thus, it is easier to assume they equal zero.
6. Transcription and translation rates, as well as protein and mRNA decay rates, are constant with respect to time. This greatly simplifies the model, but presents significant inaccuracy compared to the real biosystem.
7. The temperature is 37°C for all reactions. A constant temperature is required for the rate constant k to remain constant.

Due to the nonlinear nature of the lactose-glucose kinetics, the function must be linearized in order to take the Laplace transform of the system. Via Taylor series approximation, (5) and (6) linearize to the forms:

$$\frac{dL}{dt} = -k\bar{Z}L(t) \quad (7)$$

$$\frac{dG}{dt} = k\bar{Z}L(t) \quad (8)$$

Applying the Laplace transform to (3) and (4), as well as (7) and (8), produces the following transfer functions:

$$H_1(s) = \frac{Z(s)}{L(s)} = \frac{AB}{s^2+(U+V)s+UV} \quad (9)$$

$$H_2(s) = \frac{G(s)}{L(s)} = -1 \quad (10)$$

Equation (9) is used for further Bode analysis (see Results). Equation (10) verifies that the conversion of lactose into glucose occurs at a 1:1 ratio..

The constants for these equations are derived from a variety of references. The rate constant k for lactose hydrolysis is determined with the Arrhenius equation:

$$k = A \cdot exp(-E/RT) \quad (11)$$

From Petzelbauer, et. al[5]., E is defined as 75000 J/mol. Selvarajan and Mohanasrinivasan[6] determine k to be 2750 L/mol*min at a temperature of 50°C. Thus, k = 853 L/mol*min at 37°C.

The transcription and translation rates were determined from results summarized by Lieve and Kollin[7], which describe a 2.5 minute average time period of transcription and a 1.75 minute average time period between transcription and the emergence of β-galactosidase. Thus, the transcription rate A = 0.4 min^{-1} and translation rate B = 0.4 min^{-1}.

The decay rate of β-galactosidase and mRNA are determined as the mean-lives of values found in literature. Miller and Zipser[8] determine the half-life of multiple variants of β-galactosidase to be between 11 to 15 min. The average of these half-lives is 12.375 min, which produces a mean-life of 17.85 min. The protein decay rate of U = 0.056 min^{-1}.

The decay rate for mRNA is determined from values described from Eriksen, et. al[9]. In their work, the mRNA stability is described in terms of half-life. This model contains the wild-type (WT) value of 116 sec. This corresponds to a mean-life of 2.78 min and an approximate mRNA decay rate of V = 0.36 min^{-1}.

Equations (3) through (6), and the values for k, A, B, U, and V derived here are used to construct the following block diagram for modeling:

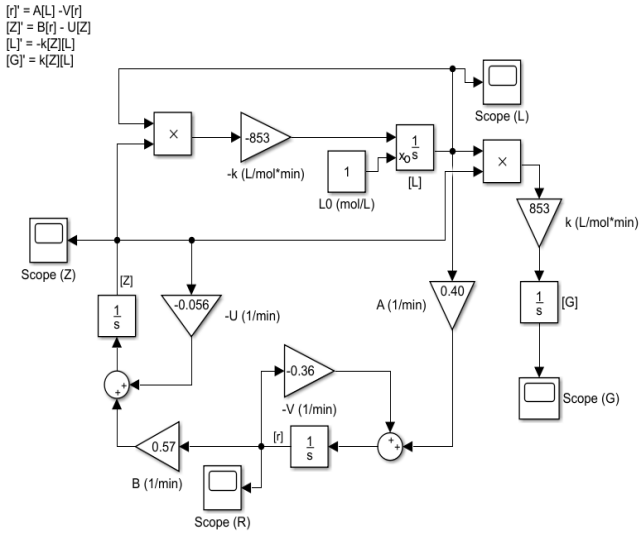

Figure 2: Block diagram of gene expression created in Simulink, illustrating the layout of the gene expression and chemical kinetic differential equations of the system.

III. Results

By simulating the *Lac* operon model in Simulink, using the block diagram in Figure 2, the output enzyme concentration was analyzed as a function of the input lactose concentration. As seen in Figure 3, the enzyme concentration does not reach a steady state value and rather continues to decrease slowly regardless of the input of lactose in the system.

When the enzyme concentration is decreasing, the cell needs to find a supply of glucose elsewhere since the enzyme induces the production of glucose. Maintaining levels of enzyme at a steady concentration of 1 mol/L ensures continued production of glucose and continuation of the system seen in Figure 8.

In Figure 4 and Figure 5, the system dynamics can be observed such that lactose gets fully consumed because the β-galactosidase cleaves the lactose into galactose and glucose. Figure 6 shows that most of the RNA concentration is transcribed within the first minute and then exponentially decreases after a minute to eventually approach 0 mol/L as expected because mRNA is degraded quickly. Furthermore, in Figure 7 the magnitude of the bode plot decreases by -40 dB/dec and the phase accordingly decreases by -180° because of the two poles in the transfer function (9).

Assuming $V > U$, the PD controller cancels out the pole at V to improve stability. Here the phase margin (PM) is 45 because at a gain of 1, the corresponding phase is at -135 so it is 45 degrees away from -180 or instability. To stabilize the system, a PD controller is needed to cancel out the pole.

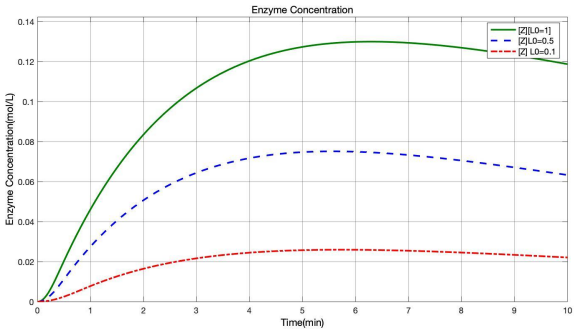

Figure 3: Enzyme concentration vs. time without PD control for different initial conditions of lactose present in the cell.

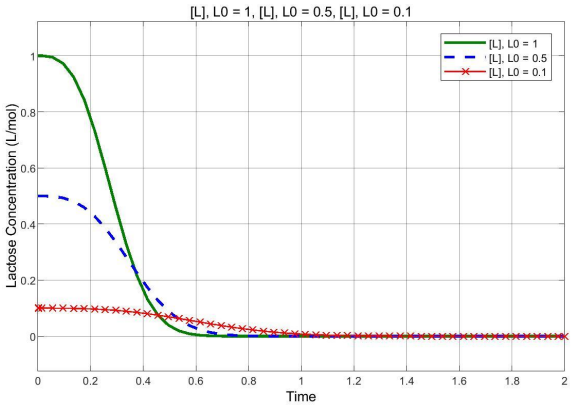

Figure 4: Lactose concentration vs time. Lactose concentrations of 1 mol/L, 0.5 mol/L, and 0.1 mol/L decrease until reaching a near-zero steady state.

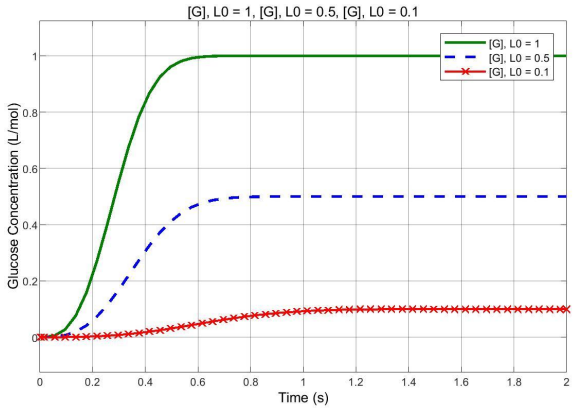

Figure 5: Glucose concentration vs time. Glucose concentrations increase over time until reaching a steady state, depending on initial conditions of lactose concentration.

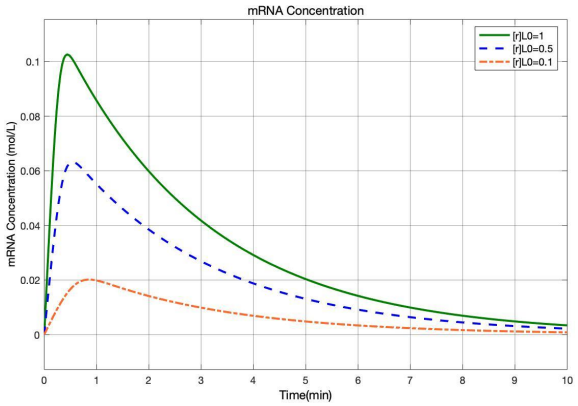

Figure 6: mRNA Concentration vs. Time. RNA concentration is in mol/L on the y-axis and time in minutes on the x-axis.

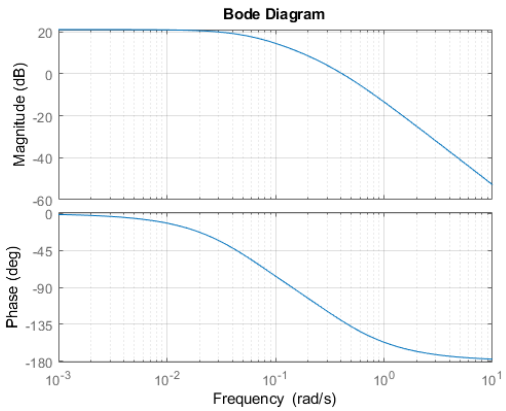

Figure 7: Bode plot of the open-loop system without PD controller. Top: magnitude (dB) vs frequency (rad/s). Bottom: phase (degrees) vs frequency (rad/s).

In order to continue transcription of the *Lac* operon and maintain a steady concentration of enzyme for any desired reporter system, a PD controller with proportional and derivative control is used. Implementing proportional control improves the settling by critically damping the response, and derivative control lowers the rise time. The block diagram is shown below in Fig.8 using the transfer function (9) and a controller with a K_p value of 1 and a K_d value of 2.7. The values of K_p and K_d were determined in a way that it would cancel out the pole at V. As a result, lactose concentration still settles to 0 because it is converted to glucose. However, the enzyme concentration remains constant instead of decreasing because the controller is using the target value of 1 mol/L to keep a steady value of the enzyme seen in Figure 9.

$$H(s) = \frac{[A*B*kp]}{(s+U)} \qquad (12)$$

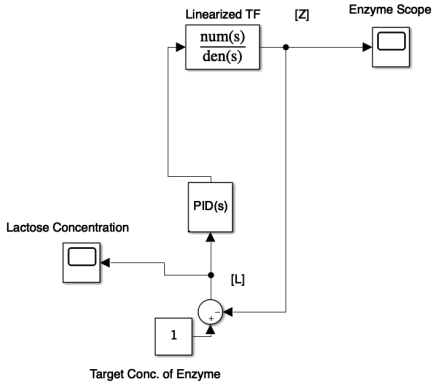

Figure 8: Linearized model of block diagram from Simulink

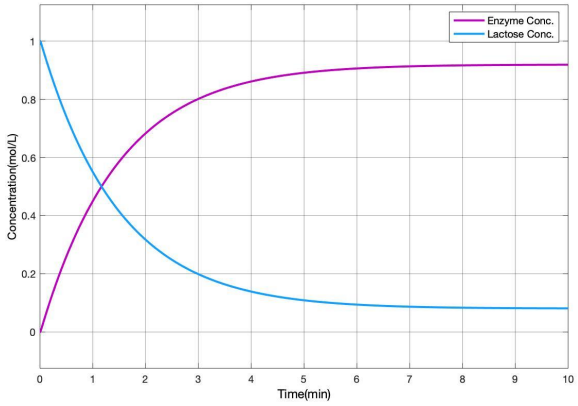

Figure 9: System Response with PD Controller. Enzyme concentration increases and settles at the target while lactose concentration decreases towards zero.

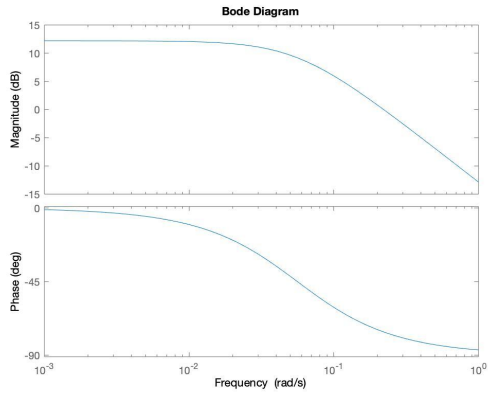

Figure 10: Bode plot of the open-loop system with PD controller. Top: magnitude (dB) vs frequency (rad/s). Bottom: phase (degrees) vs frequency (rad/s).

IV. Discussion

Based on the assumptions, the simplified model behaves as expected: while lactose and mRNA concentrations decrease, the glucose and enzyme (β-galactosidase) concentrations increase until reaching a steady state point. Lactose concentrations decrease as the operon transcribes the *Lac* genes to produce β-galactosidase. Eventually the enzyme cleaves all of the lactose to form glucose and galactose. Lower levels of lactose lead to lower levels of the inducer allolactose, allowing the repressor to bind to the *Lac* promoter halting the production of mRNA.

The PD controller utilizes proportional and derivative control to cancel out the second pole and cause first order dynamics. As seen in Fig.10, with the PD controller, the phase margin increases from 45 to 90. At a gain of 1, the phase is -90 degrees, which is 90 degrees away from a phase of -180. This occurs because the PD controller cancels the second pole so the system decays at -20 dB/dec, meaning this system is more stable with a steady enzyme concentration.

The results from the simplified model might differ from the true system output in a *Lac* operon because of the assumptions made. Assumption 2 could affect the enzyme and RNA concentrations because the simplified model does not include the effect of CAP, a protein which activates transcription of the *Lac* operon when glucose levels are low. CAP senses glucose through cAMP because cAMP is only produced when glucose is low and absent under high glucose concentrations. If low levels of glucose lead to more mRNA and enzyme production, then glucose should be considered in their Laplace equations. However, the effect of glucose is ignored which could be one source of error. Another source of error is Assumption 3 because the simplified model assumes all the lactose input is being converted into allolactose. However, allolactose is formed by the hydrolysis of lactose with initial concentrations of β-galactosidase already present in the cell, meaning sufficient enzyme concentration is needed to produce allolactose. Therefore, the simplified model assumes allolactose is only dependent on lactose concentration, but is affected by intramolecular β-galactosidase concentrations as well. In *E.Coli*, there is also the presence of Isopropyl β-D-1-thiogalactopyranoside which mimics the purpose of allolactose as an inducer. It can promote the transcription of the *Lac* operon by binding and inhibiting the repressor without being degraded by β-galactosidase like allolactose. In the simplified model, there is no effect of IPTG on any of the variables, but in the true system output, IPTG can substantially affect the mRNA and enzyme concentrations.

V. Conclusion

Given current understanding of the inducible gene expression control of the *Lac* operon, this model can be manipulated with today's molecular tools in numerous biomedical applications that span beyond lactose metabolism in bacteria. One of the most important usages of Lac operon is in recombinant protein production. As described by this model, its inducible nature gives scientists the flexibility to use this system for the expression of any gene that can be attached to the *Lac* operon via genetic engineering tools. An interesting aspect of such application is the fact that the amount of protein produced is tunable not only by increasing amounts of IPTG as an inducer but also by adjusting the promoter by introducing point mutations using CRISPR Cas-9.

In the human body, excessive lactose in the intestine leads to tissue dehydration and reduced calcium absorption which can cause several unwanted side effects such as diarrhea and cramps. β-galactosidase, normally found in the small intestine as a form of lactase, can help lactose-intolerant individuals. Since almost 70% of the world's population is lactose intolerant[11], many food processing companies add β-galactosidase produced as a recombinant protein to reduce lactose content and produce hydrolyzed lactose-free products. Lactose intolerant individuals are often given β-galactosidase medicine before consuming milk. Therefore, the *Lac* operon holds clinical relevance in treating lactose-heavy foods.

Finally, another promising application of the Lac operon gene expression model is in gene therapy. Potential application of the *Lac* operon would be specifically for diseases with genetic protein deficiency and growth hormone (GH) deficiency. Knowing the dynamics of gene expression control in *Lac* operon allows for the use of a Lac operon promoter to regulate a systematic expression of hGH

by attaching its gene to the *Lac* operon. This approach can be used in gene therapies for humans with inherited protein deficiencies like idiopathic hGH deficiency. However, such an application seeks a more thorough understanding of the dynamics of the gene expression mechanisms by considering multiple levels of control as well as the delays of translation and membrane transport. A successful model is a key step towards understanding the dynamics of gene expression mechanisms, its regulation and the time scale of gene expression for gene therapy applications.

REFERENCES

[1] Chen,Ting et al. "Modeling Gene Expression with Differential Equations." GC-main, 1999.

[2] Douglas F. Browning, Rita E. Godfrey, Kirsty L. Richards, Colin Robinson, Stephen J.W. Busby; Exploitation of the *Escherichia coli lac* operon promoter for controlled recombinant protein production. *Biochem Soc Trans* 30 April 2019; 47 (2): 755–763. Doi:

[3] Nickle, Todd, and Isabelle Barrette-Ng. "12.1: The Lac Operon." *Biology LibreTexts*, 13 Sept. 2020

[4] Selvarajan, E., and V. Mohanasrinivasan. "Kinetic Studies on Exploring Lactose Hydrolysis Potential of β Galactosidase Extracted from Lactobacillus Plantarum HF571129." PubMed Central (PMC), 16 Jan. 2015

[5]Petzelbauer, I et al. "Development of an ultra-high-temperature process for the enzymatic hydrolysis of lactose. I. The properties of two thermostable beta-glycosidases." Biotechnology and bioengineering vol. 64,3 (1999)

[6] Lieve, Loretta, and Virginia Kollin. "Synthesis, Utilization and Degradation of Lactose Operon MRNA in Escherichia Coli." Journal of Molecular Biology, 14 Mar. 1967

[7] Eriksen, Mette et al. "Occlusion of the Ribosome Binding Site Connects the Translational Initiation Frequency, mRNA Stability and Premature Transcription Termination." Frontiers in microbiology vol. 8 362. 14 Mar. 2017

[8] Miller, C G, and D Zipser. "Degradation of Escherichia coli beta-galactosidase fragments in protease-deficient mutants of Salmonella typhimurium." Journal of bacteriology vol. 130,1 (1977)

[9] Lodish, H. F., Molecular cell biology (4th ed.). New York: W.H. Freeman. (2000).

[10] Noel, Sturm. "Regulation of Gene Expression." Www2 Web Server, 2020, www2.csudh.edu/nsturm/CHEMXL153/RegulationofGeneExpression.htm.

[11]Saqib, Shaima et al. "Sources Of B-Galactosidase And Its Applications In Food Industry". *3 Biotech*, vol 7, no. 1, 2017. *Springer Science And Business Media LLC*, doi:10.1007/s13205-017-0645-5. Accessed 18 Dec 2020.