SysMIC
University College London

Mini-Project
The Lac Operon

—

By: Christos Efthymiou

I. ABSTRACT

Operons are a system of genetic regulation present most commonly in bacteria. They consist of a promoter, operator sequence, and a series of genes that can either all be transcribed and translated or they will all be turned off. The most famous example is the lac operon, which is an inducible operon that produces proteins needed to metabolize lactose when glucose is not present and lactose is present. The genes are normally not transcribed due to the Lac Repressor, lacI, which is constitutively expressed and binds the operator, blocking the promoter and preventing the transcription of the genes.

The purpose of this study was to utilize MatLab in order to study the nature of the interaction between the Lac Repressor (lacI), the operon (gop), and allolactose (alac). By simulating the lac operon in Matlab, the operon's behavior can be studied under different conditions to gain a deeper understanding of how it functions. This information can be used to guide experimental work and improve understanding of fundamental biological processes.

II. DESCRIPTION OF BIOLOGICAL SYSTEM

The lac operon system, discovered by Jacob and Monod, is a well-characterized protein control system in bacteria for metabolizing lactose when glucose is not present. In general, operons refer to clusters of genes that are related in their function and are under the control of a common mechanism (Ralston, 2008). The lac operon is reflective of many other bacterial operons, and therefore its structure can be used to understand the fundamentals of most operons. The lac operon consists of the catabolite activator protein (CAP) site, promoter, operator, and the genes *lac Z, Y, and A*. Beta-galactosidase is produced by the *lac z* gene, permease by the *lac y* gene, and the transacetylase enzyme by the *lac a* gene. These gene products work in concert to bring lactose into the cell and metabolise it (Ralston, 2008). There are several proteins including CAP, RNA Polymerase, and the Repressor which interact with the operon and dictate if the operon is expressed or repressed (Figure 1).

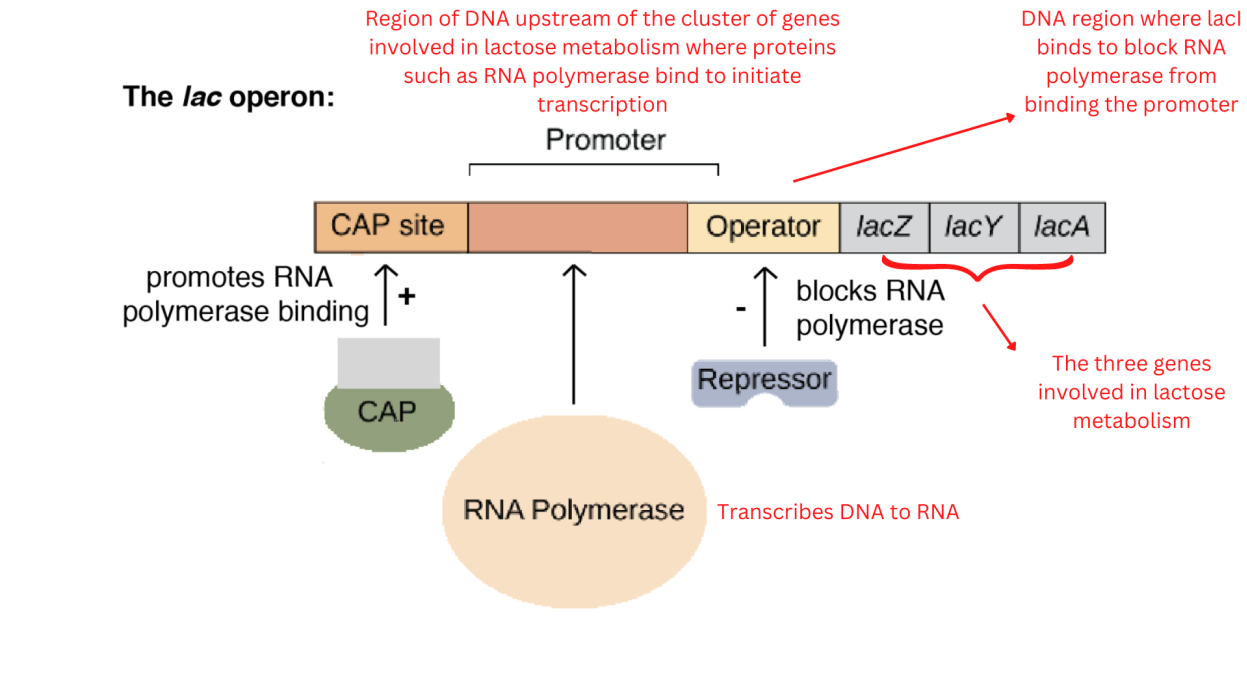

Figure 1: The structure of the lac operon. Figure adapted from Khan Academy. The LAC operon (article). Khan Academy. https://www.khanacademy.org/science/ap-biology/gene-expression-and-regulation/regulation-of-gene-expression-and-cell-specialization/a/the-lac-operon. Published 2016. Accessed January 23, 2023.

The CAP site precedes the promoter and is where the catabolite activator protein can bind when glucose levels are low. When glucose is low, the enzyme adenylyl cyclase is active and generates cyclic-AMP, or cAMP. cAMP binds to CAP, and the complex can then bind the CAP site. When it is bound to the CAP site, CAP helps recruit RNA polymerase to the promoter and therefore facilitates higher levels of gene expression. When glucose levels are high in the cell, cAMP levels remain low and CAP does not bind efficiently to the CAP site on its own, so gene expression levels are lower (Kimata et al., 1997). This design is beneficial as glucose is the preferred carbon source for *E. coli*. It takes fewer steps and less energy to metabolise glucose compared to lactose, so if both glucose and lactose are present, glucose will be preferentially metabolised as the lactose metabolism genes will not be expressed at a high level (Aidleberg et al., 2014).

The promoter refers to a region of DNA upstream of genes where proteins such as RNA polymerase can bind to initiate transcription (Ayoubi et al., 1996). However, because the bacteria do not want to waste energy producing proteins that are not needed, the lactose

metabolism genes are not normally expressed. Expression is blocked by the Lac Repressor, coded by the *lacI* gene, which binds to a sequence of DNA referred to as the operator sequence. Since the promoter sequence partially overlaps with the operator sequence, when the Lac Repressor is bound, the promoter is blocked and the RNA polymerase cannot bind and begin transcription. The Lac Repressor is constitutively expressed as it is controlled by a separate promoter. When lactose is present, it is converted to allolactose by beta-galactosidase and can then bind to the Lac Repressor. This induces a conformational change, causing the repressor to fall off the operator sequence, thereby exposing the promoter for the RNA polymerase to begin transcription of the three lactose metabolism genes so the bacteria can process the lactose (Jacob and Monod, 1961).

The lac operon is incredibly important as a model because of the widespread use of operons in bacterial and archaeal genomes. In particular, approximately half of the protein coding genes are located within operons (Price et al., 2006). Studying the mechanism of the lac operon reveals how a large portion of the genes within bacteria and archaea function. Operons are rarely found in eukaryotes; however, they do exist and they maintain some of the principles of the lac operon, though they are more complex in eukaryotes (Blumenthal, 2004). In general, the discovery and understanding of the lac operon paved the way for the field of molecular biology and introduced the concept of gene regulation for the first time (Lewis, 2011).

III. DESCRIPTION OF MODEL & CODE

Full model description

A simplified model of the lac operon consists of several key components: the operon itself (gop), the repressor protein (lacI), the inducer molecule allolactose (alac), and the bound forms of gop-lacI and lacI-alac. The operon is a series of genes that encode the enzymes needed for the metabolism of lactose, and is controlled by the lacI repressor protein. When lactose is present, it is converted to allolactose and binds to the lacI repressor, causing a conformational change that allows the operon to be transcribed and the enzymes to be produced. In the absence of lactose, the lacI repressor remains bound to the operon, preventing transcription and enzyme production. The bound forms of gop-lacI and lacI-alac represent the different states that the operon and repressor can adopt in response to the presence or absence of lactose. These

components of the model are related to one another through two reversible reactions: lacI binding gop and alac binding lacI which can be represented by the following equations:
rate1_forward = k1 * gop * lacI; % forward rate equation for lacI binding gop
rate1_reverse = k2 * gop-lacI; % reverse rate equation
rate2_forward = k3 * lacI * alac; % forward rate equation for alac binding lacI
rate2_reverse = k4 * lacI-alac; % reverse rate equation

In order to utilize these equations to model the lac operon, it is necessary to assign the 5 components of gop, lacI, gop-lacI, alac, and lacI-alac, to variables for use in the differential equations. The 5 components are represented by:
x(1) = concentration of gop
x(2) = concentration of lacI
x(3) = concentration of gop-lacI
x(4) = concentration of alac
x(5) = concentration of lacI-alac

Now it is possible to write the differential equations for each of these components. As an example, the differential equation for x(1) can be determined by finding the rate equations which generate and reduce the concentration of gop. Therefore, the differential equation is given by rate1_reverse being positive and rate1_forward being negative and replacing the molecular components with the variables. A similar technique was used to determine all the differential equations:
dxdt(1) = k2*x(3) - k1*x(1)*x(2)
dxdt(2) = k2*x(3) + k4*x(5) - k1*x(1)*x(2) - k3*x(2)*x(4)
dxdt(3) = k1*x(1)*x(2) - k2*x(3)
dxdt(4) = k4*x(5) - k3*x(2)*x(4)
dxdt(5) = k3*x(2)*x(4) - k4*x(5)

Based on knowledge of the lac operon and its mechanism, it is possible to predict how the model will behave.

When the operon is inhibited:

The relative abundance of gop will be near 0 as lacI will be bound to gop to prevent transcription and translation of the genes to break down lactose. However, it will not be 0 as some expression always occurs. The relative abundance of lacI will also be near 0 as all lacI should be bound to gop rather than free in solution. Following the same logic, gop-lacI will be

abundant. Alac and lacI-alac will have relative abundance of 0 as no allolactose is present in the inhibited state.

When the operon is active:

The relative abundance of gop will be high as it is no longer inhibited by lacI. The relative abundance of free lacI will still be low as now it will be bound to alac rather than gop. The relative abundance of gop-lacI will be low for the same reason. Alac will be somewhat abundant, and lacI-alac will be highly abundant as alac will bind to lacI.

Based on biological knowledge, the values of the rate constants have been arbitrarily selected. When only lacI is present, it is known that nearly all of gop will be in its bound form (Semsey et al., 2013). Therefore, the rate constant for k1, i.e. the formation of gop-lacI, must be much greater than the rate constant for the degradation of gop-lacI. So k1 was set at 1.5 and k2 at 0.1. It is also well established that alac is capable of binding lacI and causing it to release gop (Wheatley et al., 2013). This means k3 must be set at a much greater value than k1 to model the release of lacI when alac is present. Therefore, k3 was set to be 10x faster than k1. Finally, k4 was set to a low value of 0.2 as alac should not release lacI until the lactose breakdown enzymes have been expressed and can process the alac in the cell.

Overall, this model describes the complex regulatory mechanisms that control the expression of the lac operon in response to the availability of lactose. Models similar to the one described in this study have been employed in several settings. For example, one iteration of such a model was developed to help students understand the lac operon through a free web-application known as LacOp (Charczenko et al., 2022). Users are able to vary the initial concentrations of the various components of the model including lactose. However, the web-application uses a slightly more complex model to better reflect the physiological nature of the lac operon. For example, the concentration of allolactose, glucose, and lactose inside versus outside of the cell can be changed to observe the effect. Additionally, the web-application includes the ability to study a mutant promoter, operator, and repressor to further develop a complete understanding of the lac operon's function (Charczenko et al., 2022).

Simplifying assumptions and limitations of the model

There are several simplifying assumptions and limitations of this model of the lac operon. First, the model only considers the effects of allolactose on the lac operon, ignoring other potential factors that may influence its behavior. For example, the model does not take into account the effects of other regulatory proteins or molecules that may interact with the lac operon and

affect its behavior. It is known that when glucose is also present, the lac operon is still not expressed as a result of CAP not binding the CAP site due to low cAMP levels (Santillán & Mackey, 2004). Therefore, if the model included glucose and CAP/the CAP site, at high concentrations of glucose it would be expected that the operator would be free and for the repressor to be bound to allolactose; however, the CAP site would not be bound and therefore the expression of the lactose metabolism genes would remain low. In any case, this is not considered in this model.

Additionally, the model does not consider the potential effects of mutations or other genetic variations on the lac operon, which can affect its function in different bacteria. For example, a lac repressor mutant may hinder its ability to bind allolactose, causing the lac operon to remain repressed even when allolactose is present. Other models allow for the exploration of the effect of various mutations on the behaviour of the lac operon (Charczenko et al., 2022). Furthermore, the model does not account for the potential effects of environmental factors, such as temperature or pH, on the lac operon, which can also influence its behavior. For example, changes in pH can affect the survivability of *E. coli* (Suehr et al., 2020).

Another limitation of the model is the values selected for the relative abundance of the various components and the rate constants. Since the behavior of the lac operon is already well understood, the values were arbitrarily selected to ensure that the model fits the behavior of the experimental work. The model would more accurately be developed using experimental values for the relative abundances and rate constants. For example, Zuo et al. studied the lac repressor's binding specificity in depth and these types of experiments could inform the values used for all the components in the model (Zuo et al., 2015).

The model is limited by its ability to accurately capture the complex interactions between the various components of the lac operon, and may not accurately represent its behavior in all situations. Overall, while this model provides valuable insights into the function of the lac operon, it should be used with caution and considered in the context of other available information.

IV. CODE LISTINGS

The file containing the MatLab code has been submitted along with this report on the SysMIC website.

V. EXTENDED DISCUSSION

The Matlab model of the lac operon was validated by checking that it satisfied conservation laws and that its steady states reflected biological knowledge of the system. Conservation laws are fundamental principles in biology that describe the conservation of mass, energy, and other quantities in biological systems (Podobnik et al., 2017). By ensuring that the model satisfies these laws, such as with the total amount of lacI in bound and unbound forms remaining constant, it is clear that the model is physically plausible and accurately represents the behavior of the lac operon. This can be seen in the following graph, whereby the red lacI and orange gop-lacI add up to 2, the initial concentration of lacI.

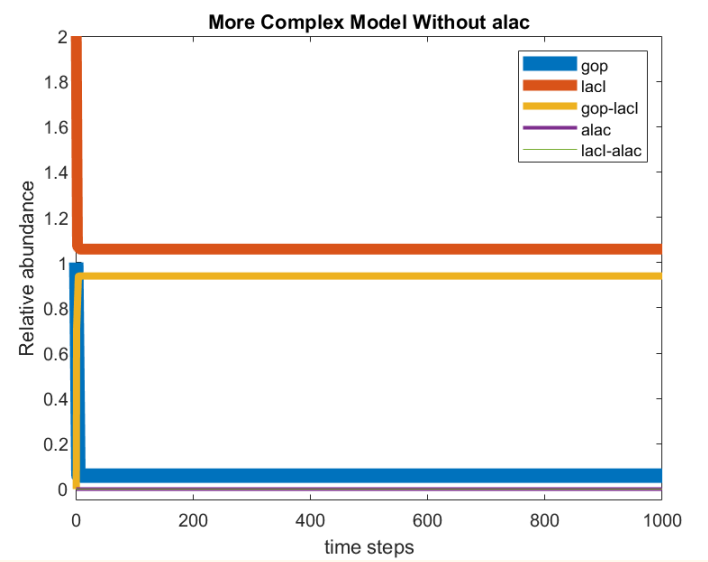

Figure 2: Validation of the model by assessing conservation laws and steady states. A check was performed to ensure that the model was behaving as expected according to physical laws and given biological knowledge. All species add up to the initial total and the majority of gop is in the bound form, thereby validating the model from both perspectives. Line widths vary solely for visualization purposes.

Additionally, the model was validated by checking that its steady states, which represent the long-term behavior of the lac operon, were consistent with biological knowledge of the system. For example, the model's steady states should reflect the known effects of allolactose on the lac operon, and should accurately capture the behavior of the lac operon in the presence and absence of lactose. In the absence of allolactose, nearly all of the gop was bound by lacI as

expected. When allolactose was introduced such as in Figure 3, a fraction of gop was free as alac bound some of the lacI, causing it to release gop.

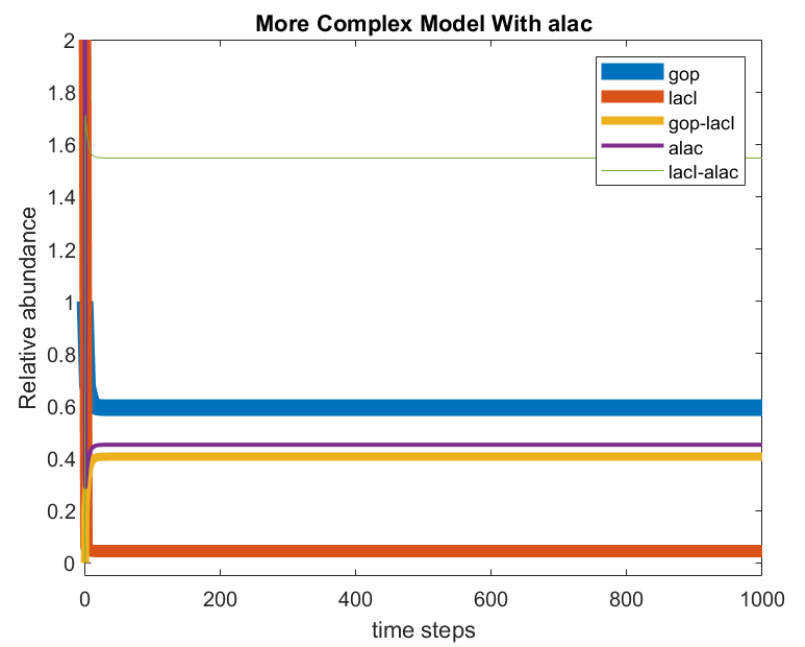

Figure 3: Graph designating the behavior of each of the species in the model when allolactose is present. A fair proportion of gop remains in its free form when allolactose is introduced as its presence causes the repressor to fall off the operator.

When a higher concentration of alac was introduced at the start of the simulation, the steady state showed even more gop in its free form (Figure 4).

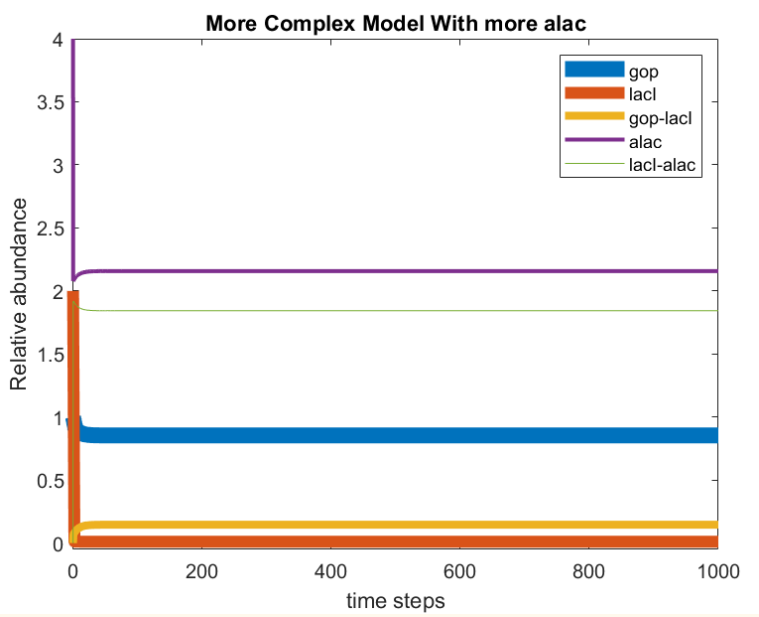

Figure 4: The behaviour of the various species in the model when a higher initial concentration of allolactose was used. The increased concentration of initial allolactose caused even more gop to be in its free form, consistent with the expectation that allolactose causes lacI to release the operator.

Therefore, the model fits the biological behavior of the lac operon. Without alac, nearly all of gop is bound by lacI as there is no need to express the lac operon genes when lactose is absent. When alac is introduced, a fraction of the lac operon becomes unbound by lacI since alac binds to lacI, inducing a conformational change. This allows transcription proteins to bind to the operon so that the lactose processing enzymes can be expressed. Furthermore, when more alac is introduced, even more gop becomes free which would allow for more lactose processing enzymes to be expressed. This is desirable as the amount of lactose processing enzymes expressed is linked to the amount of lactose present. Therefore, enzymes will only be expressed in the quantities needed to deal with the amount of lactose.

The results from the simulation could be viewed as mostly qualitative. The simulation is focused on understanding the general behavior of the lac operon, such as its response to different stimuli or the mechanisms of its regulation. It is designed to describe the different states and transitions of the lac operon, and how they are affected by different factors. If the simulation had focused on quantitatively predicting the behavior of the lac operon, such as the levels of gene expression or enzyme production under different conditions, the results would be quantitative.

One of the key assumptions that make the simulation of the lac operon using Matlab qualitative rather than quantitative is the use of simplified representations of the different components of the system. For example, the model uses simplified equations to describe the binding and unbinding of allolactose to the lacI repressor, and uses simplified representations of the different genes and proteins involved in the lac operon. These simplifications make the model easier to work with, but also limit its ability to make quantitative predictions about the behavior of the lac operon.

Another assumption that makes the simulation qualitative rather than quantitative is the lack of detailed information about the values of the rate constants and other parameters of the model. In many cases, the values of the rate constants and other parameters are based on experimental data or other information, but were arbitrarily selected here to fit the general mechanism of the lac operon. This limits the model's ability to make quantitative predictions about the behavior of the lac operon, and makes the results of the simulation more qualitative in nature.

The use of simplifying assumptions and limited information about the values of the parameters of the model can make the simulation of the lac operon using Matlab qualitative rather than quantitative. While these assumptions can make the model more tractable, they also limit its ability to make precise and accurate predictions about the behavior of the lac operon.

VI. OUTLOOK

A computational model of the lac operon, such as the one created using MatLab, is useful because it allows researchers to study the behavior of this important genetic regulatory system in a flexible and powerful way. By simulating the lac operon in a computational environment, researchers can study how it responds to different stimuli, such as the presence or absence of lactose, and gain a deeper understanding of its behavior. Using a computational model allows for the rapid testing and iteration of different hypotheses and scenarios, providing a valuable tool for studying the lac operon. Compared to traditional experimental methods, a computational model allows for a more controlled and systematic approach to studying the lac operon, and can provide insights that are difficult or impossible to obtain through experimental work (Mackey et al., 2015).

However, the model has limitations and could be improved to more accurately simulate the lac operon. For example, the model could include more detailed representations of the lacI repressor protein and the allolactose inducer molecule, which would allow for more accurate

modeling of their binding and unbinding to the operon. For example, there is evidence that two lacI molecules bind a single operator, which is not considered in this model (Sadler et al., 1983). Additionally, the model could be expanded to include information about other regulatory proteins and molecules that interact with the lac operon and affect its behavior, such as glucose (Santillán & Mackey, 2004).

Another way to expand the model would be to incorporate more detailed information about the different states and transitions that the lac operon can adopt in response to different stimuli. This could involve using more complex mathematical equations to describe the dynamics of the system, and could allow for more accurate modeling of the effects of different factors, such as mutations or environmental conditions, on the lac operon. The educational web-server LacOp includes some of this functionality by allowing users to study the effect of mutations to the operator, repressor, etc. that is not feasible with the model used in this study (Charczenko et al., 2022).

Additionally, the model could be made more accurate by considering other more complex concepts such as delays caused by transcription and translation as well as the time required to import external lactose, convert lactose to allolactose, etc. A model incorporating these types of parameters was employed and compared to experimental data. The model very accurately mimicked the experimental data and demonstrates the benefit of making a more complex model (Yildirim & Mackey, 2007).

The model could be expanded to include more sophisticated optimization techniques, such as machine learning algorithms, to optimize the values of the rate constants and improve the accuracy of the model's predictions. A hybrid approach involving simulation and artificial intelligence optimization has been employed in other fields and could be applied to the lac operon (Arya et al., 2022).

Creatively, a 3D interactive model of an *E. coli* organism and the lac operon has also been created. The simulator includes cellular structures such as the plasma membrane as well as molecular components like RNA polymerase, ribosomes, etc. The various components are placed in a virtual environment and interact through simulated electrochemical and physical properties. Mathematical equations capture the bigger picture changes while localized rules reveal information about smaller scale molecular interactions (Esmaeili et al., 2015). By incorporating these and other expansions, the model could be made more accurate and comprehensive, providing a more complete representation of the lac operon and its behavior.

While using a Matlab model to study the lac operon provides valuable insights into its behavior and function, it is important to carefully consider the limitations of the model and the assumptions that it is based on. By understanding these limitations and using the model in combination with other information and experimental data, researchers can gain a more complete understanding of the lac operon and its role in bacterial metabolism.

VII. References

1. Aidelberg G, Towbin BD, Rothschild D, Dekel E, Bren A, Alon U. Hierarchy of non-glucose sugars in Escherichia coli. BMC Syst Biol. 2014;8:133. Published 2014 Dec 24. doi:10.1186/s12918-014-0133-z
2. Arya Azar N, Kayhomayoon Z, Ghordoyee Milan S, Zarif Sanayei H, Berndtsson R, Nematollahi Z. A hybrid approach based on simulation, optimization, and estimation of conjunctive use of surface water and groundwater resources. Environ Sci Pollut Res Int. 2022;29(37):56828-56844. doi:10.1007/s11356-022-19762-2
3. Ayoubi TA, Van De Ven WJ. Regulation of gene expression by alternative promoters. FASEB J. 1996;10(4):453-460.
4. Blumenthal T. Operons in eukaryotes. Brief Funct Genomic Proteomic. 2004;3(3):199-211. doi:10.1093/bfgp/3.3.199
5. Charczenko R, McMahon M, Kandl K, Rutherford R. LacOp: A free web-based lac operon simulation that enhances student learning of gene regulation concepts. Biochem Mol Biol Educ. 2022;50(4):360-368. doi:10.1002/bmb.21638
6. Esmaeili A, Davison T, Wu A, Alcantara J, Jacob C. PROKARYO: an illustrative and interactive computational model of the lactose operon in the bacterium Escherichia coli. BMC Bioinformatics. 2015;16:311. Published 2015 Sep 29. doi:10.1186/s12859-015-0720-z
7. Jacob, F., Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol. 1961;3:318-356.
8. Kimata K, Takahashi H, Inada T, Postma P, Aiba H. cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose transporter gene in Escherichia coli. Proc Natl Acad Sci U S A. 1997;94(24):12914-12919. doi:10.1073/pnas.94.24.12914
9. Lewis M. A tale of two repressors. J Mol Biol. 2011;409(1):14-27. doi:10.1016/j.jmb.2011.02.023

10. Mackey MC, Santillán M, Tyran-Kamińska M, Zeron ES. The utility of simple mathematical models in understanding gene regulatory dynamics. In Silico Biol. 2015;12(1-2):23-53. doi:10.3233/ISB-140463
11. Podobnik B, Jusup M, Tiganj Z, Wang WX, Buldú JM, Stanley HE. Biological conservation law as an emerging functionality in dynamical neuronal networks. Proc Natl Acad Sci U S A. 2017;114(45):11826-11831. doi:10.1073/pnas.1705704114
12. Price MN, Arkin AP, Alm EJ. The life-cycle of operons [published correction appears in PLoS Genet. 2006 Jul;2(7):e126]. PLoS Genet. 2006;2(6):e96. doi:10.1371/journal.pgen.0020096
13. Ralston, A. (2008) Operons and prokaryotic gene regulation. Nature Education 1(1):216
14. Sadler JR, Sasmor H, Betz JL. A perfectly symmetric lac operator binds the lac repressor very tightly. Proc Natl Acad Sci U S A. 1983;80(22):6785-6789. doi:10.1073/pnas.80.22.6785
15. Santillán M, Mackey MC. Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. Biophys J. 2004;86(3):1282-1292. doi:10.1016/S0006-3495(04)74202-2
16. Semsey S, Jauffred L, Csiszovszki Z, et al. The effect of LacI autoregulation on the performance of the lactose utilization system in Escherichia coli. Nucleic Acids Res. 2013;41(13):6381-6390. doi:10.1093/nar/gkt351
17. Suehr QJ, Chen F, Anderson NM, Keller SE. Effect of pH on Survival of Escherichia coli O157, Escherichia coli O121, and Salmonella enterica during Desiccation and Short-Term Storage [published online ahead of print, 2020 Jan 13]. J Food Prot. 2020;211-220. doi:10.4315/0362-028X.JFP-19-195
18. Wheatley RW, Lo S, Jancewicz LJ, Dugdale ML, Huber RE. Structural explanation for allolactose (lac operon inducer) synthesis by lacZ β-galactosidase and the evolutionary relationship between allolactose synthesis and the lac repressor. J Biol Chem. 2013;288(18):12993-13005. doi:10.1074/jbc.M113.455436
19. Yildirim N, Mackey MC. Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data [published correction appears in Biophys J. 2007 Jan 15;92(2):699]. Biophys J. 2003;84(5):2841-2851. doi:10.1016/S0006-3495(03)70013-7
20. Zuo Z, Chang Y, Stormo GD. A quantitative understanding of lac repressor's binding specificity and flexibility. Quant Biol. 2015;3(2):69-80. doi:10.1007/s40484-015-0044-z