Molecular Mechanisms for Human Diseases: Lab 1

Bacteria’s Brain – Part A

Peter Yibo Pan, collaborators:

(2023-02-08)

## Abstract

distillation of report (1-paragraph abstract)

Lab question 1c: Previous studies: Benzer (), Monod() supports that each individual bacterium produces at a rate representative of the entire population; only for high inducer concentrations. For low inducer concentrations, the role of the transporter and preinduced bacteria is considered to show that the concentration of transporters will also be low. Hence some bacterium will necessarily start producing enzyme before others, which conflicts with the observation at high concentration. This conclusion seems to conflict with the previous studies of Benzer and Monod, as to whether enzyme production is always uniform within the population. The two opposing hypotheses are: \* Hypothesis 1: Bacteria within a system always produce enzyme at an uniform but varying rate at the individual level. Hence as inducers are introduced, they would be uniformly induced and begin production at the same time, slowly increasing the rate of production as concentration of inducers increase. \* Hypothesis 2: Bacteria within a system, once induced, immediately start producing enzyme at a maximum level without change. Yet some bacteria within a system may be induced before others, hence the overall production of the system is dependent on the proportion of individual bacteria activated, which produces at either none or max rate. Hence, through this experiment, Novick and

Addition problems, 1: modern use

## Experiment

### Method

The experiment utilizes thiomethyl--D-galactoside (TMG), which is a “gratuitous” inducer, to trigger the production of -galactosidase. TMG has the advantage of being able to induce bacteria without being consumedas an energy source, as lactose would, which may impact the growth rate of the population. In order to differentiate between the two hypothesized mechanisms of induction and production, the inducer TMG is introduced to each condition at different concentrations.

ways of measuring, limits to the design, elegant parts (lab question 1a, b) hypotheses

Lab question 1a: Between the two trials, the concentration of inducers is altered from high to low. According to Hypotheses 2, when inducers are introduced at a high concentration, all of the cells in the population would be rapidly induced and start producing enzyme at a steady rate, meanwhile this rate of induction is much more gradual if the inducer is introduced at a low volume. However, according to Hypothesis 1, the all cells would be induced rapidly after the introduction of the inducers regardless of its concentration. However, the rate of production for each cell slowly increases with time.

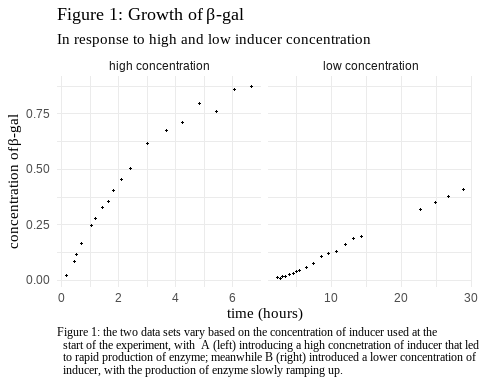
Lab question 1b: (& Lab question 6) The rate of b-gal activity is measured by the rate at which -nitrophenyl--D-galactoside (ONPG) is hydrolyzed (Novick, 555). This procedure used by Novick and Weiner is based upon the work of Joshua Lederberg in 1950. Lederberg demonstrated the merit of using spectralphotometry to obtain values for optical density for enzyme assays. ONPG is a chromogenic substrate, and can be hydrolyzed by bacteria once the bacteria starts producing -galactosidase. Once hydrolyzed, ONPG separates into free -nitrophenol (ONP) which is a weak acid and displays a yellow color in alkaline solution, with an absorption peak at 420m. The use of ONPG in tracking the activity of enzyme production can be traced back to Seidman and Link (1950) who first found its suitability, as ONP is the only compound that yield a color in the specific range, or rather in the entire visible spectrum within the experiment. Lederberg also found that ONP’s spectral frequency requires very low dilution to be accurately measured, hence enabling measurement with very low density of cells or very slow growth rate, which will be helpful in this experiment. ONP also allows the spectroscopy of the bacteria sample to directly reflect the fraction of maximum enzyme activity within the system. As for each specific condition and specific optical density of bacteria, there is a specific rate of ONP hydrolysis that corresponds to maximum enzyme production. In this experiment specifically, Novick and Weiner determined that at their condition, with a bacteria sample of optical density of 0.120, maximum bacteria activity would correspond to 225 mmoles of ONPG hydrolyzed per milliliter per minute at 28C, at pH 7 in M/10 sodium phosphate buffer.

### Setup

## Results:

The complete resulting data can be seen in tables 1 and 2 in the appendix, and as shown in table 1a and 1b:

### Visualization of Data set A and B:



## Analysis:

### Derivation A: High Concentration of Inducers

(question A.5) I will first define the different terms that will be used.These terms will be used for derivations for both data sets.

* is the number of individual E.coli bacteria in the system, which is maintained at a constant value.
* is the flow rate both in- and outward from the system, which is kept at a constant value so that can be kept constant.
* is the volume of the system.
* we define as the average rate of production by any individual bacterium at a given time. Note this is not the rate of production by individual bacterium but the mean of the population. Hence this value would change based on how many bacterium are “activated” within the population.
* is defined as the concentration of -gal within the system at any moment. With these terms defined, we observe that at any moment, the growth of -gal is , which equates two terms: the rate of production by the bacteria and the rate of loss due to flow rate. Hence we get the equation:

For data set A, when the inducers are introduced at high concentration, we assume that is constant as all bacteria rapidly “switches on” and generates enzyme at its maximal rate. We can then express as a dimensionless variable as , hence:

We can prepare the function for integration by substituting :

We then integrate both sides:

Considering the fact that , we conclude that Hence:

Or:

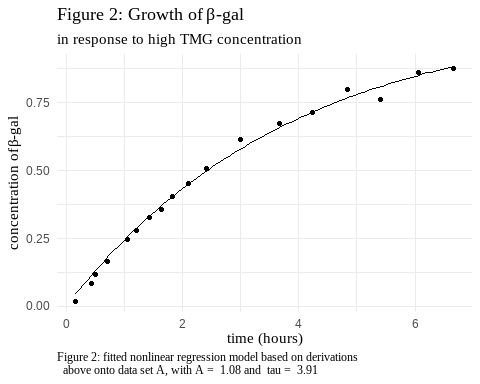
Which fit the form:

with and . The equation can be interpreted with being the theoretical maximum level of -gal that can possibly exist in the system defined by its Volume, the Number of bacteria, the individual rate of production, and the flow rate, since

In addition, can be understood as being the speed at which the system asymptotically approaches . If we return to the original equation , we see that the equation can be seen as meaning the rate of accumulation of the enzyme is negatively and linearly proportional to , which is the probability (or proportion) of existing -gal that will be lose due to system’s flow rate. This can be seen in the final equation in the term of which describes an exponential decay. In this case, the decaying factor is the growth of the rate of accumulation of -gal. As more enzyme accumulates in the system, it becomes more and more likely to be washed away by constant flow out of the system, hence would accumulate slower and slower, eventually reaching equilibrium.

### Fitting function on Data set A

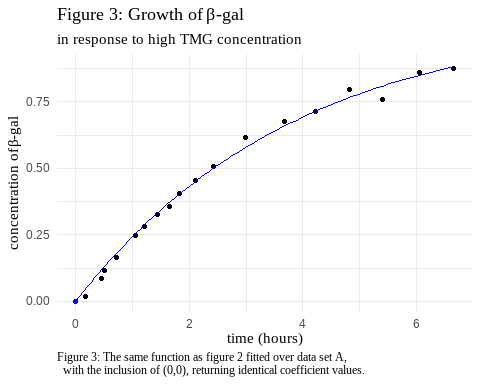
(questions A.1, A.2)  
From the derived equation A: , we set the two parameters and so the equation is of the form:



|  | data | values |
| --- | --- | --- |
| ampl | dataset A | 1.080 |
| tau | dataset A | 3.908 |

With the coefficients calculated, the nonlinear fitting of A results in the function:

With A = 1.08 and = 3.91 We observe that at , , which is valid in its physical meaning, as there would be no -gal before any inducers are introduced. Hence, the data point is included into the data and the same model is fitted again. (questions A.3, A.4)



|  | data | values |
| --- | --- | --- |
| ampl | dataset A | 1.080 |
| tau | dataset A | 3.908 |
| ampl1 | with (0,0) | 1.080 |
| tau1 | with (0,0) | 3.908 |

We see that here the two fitting parameters are identical before and after is included. This is reasonable according to the theory since there would be no -gal production before any inducer is introduced. Mathematically, . The fitting returns a residual standard error of RSE = 0.0214; the mean standard error returns as MSE = 4.11^{-4}. The two measures of error are both quite small, hence we conclude that the model is a good fit for the data. (lab question 3)

### Derivation B: Low concentration of Inducers

(questions B.6, B.7)

For data set B, when the inducers are introduced at low concentration, only a fraction of bacteria is activated, with this fraction slowly growing. Hence we assume that is a linear function of with and being the rate at which activation “spreads” through the population. Hence we again start with the equation:

However, we substitute as a function of t

Again using a dimensionless substitution, this time for as We arrive at

Which we can rearrange in the form of

Which we can use an integration factor of which we can multiply both sides of the equation and integrate:

In order to fit the initial condition of , we adjust the free constant to satisfy this initial condition, resulting in:

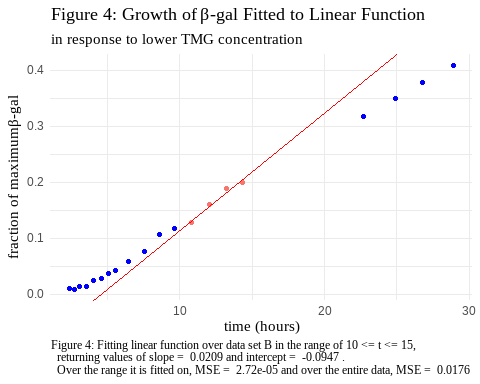
Now we replace and with their physical counterparts:

Which fit the form with and .

### Fitting function on Data set B

#### 1. Linear Function

First we start by fitting a straight line over the range

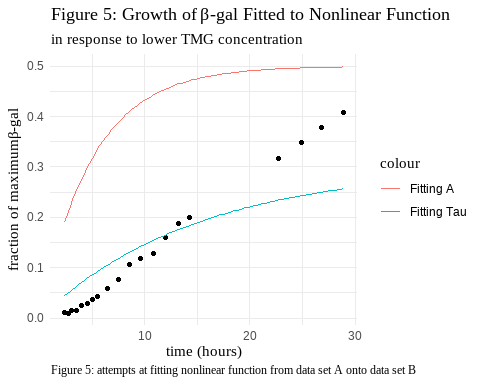


|  | coefficient | value |
| --- | --- | --- |
| (Intercept) | Intercept | -0.0947 |
| b\_q2$t | Slope | 0.0209 |

| MSE | data |
| --- | --- |
| 2.72e-05 | 10<= t <= 15 |
| 1.76e-02 | entire dataset |

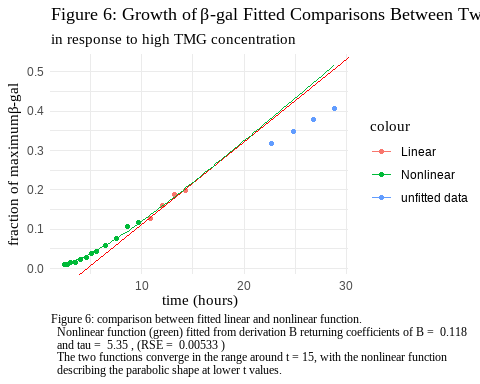
As shown in figure 4, the straight line (in red) is fitted over the data between 10 and 15 hours, over which the line fits quite well. the Mean Standard Error (MSE) for the model is much smaller within the range of $ 10t $ with a MSE of 2.72^{-5}, but not well over the entire data set with a MSE of 0.0176. Near the low end of t,

#### 2. Nonlinear Function from Data Set A



While attempting to fit data set A’s function over data set B, we realize that it is impossible to fit. If we were to fit the function around the range , the data point would inevitably exceed the value for , which is impossible since defines the theoretical maximum amount of -gal that can exist in the system as . Hence the amount of -gal should only asymptotically approach, but never reach the value for . If we try to find a value for , the value would be unreasonably high to the point where the function is irrelevant to the data at lower (See Figure 5). Hence we need to use the different function derived earlier to fit data set B.

#### 3. Nonlinear Function Derived for Data Set B



As shown in figure 6, the two fitted functions both perform well with respect to the data they are fitted to (i.e.  for nonlinear function in green, and for linear function in red).For the nonlinear function, the fitting returns a residual standard error of RSE = 0.00533, which is quite small, hence we conclude that the model is a good fit for the data. (lab question 3)

The two functions also converges in the range around . This behavior is accounted for mathematically, as we consider the nonlinear function $ z(t) = B(-1 + + e^{-})$. If we take the limit as t approaches infinity:

The function becomes linear at large , with the slope defined by (question B.5). The mathematical explanation can be complemented by an interpretation of the physical meaning of each term of the fitted function, which can explain the shape of the curve which is parabolic at small and linear at large . This interpretation can be done with a Taylor Expansion of the function: (question B.7)

From which we can get

Hence

As we can see, at small t, the growth of the enzyme is mathematically similar to .

## Dicussion

### Conclusion

(laboratory questions 4 & 5)

### Additional Questions

1. The work of Novick and Weiner contributed to the conception of multistability of gene regulation, which in modern day, is used to study how gene regulation can activate and deactivate different pathways, such as the role of repressor-operators in the transitional pathway of induction. Specifically, Bhogale et al. (2022) at the University of Cologne continues to study the transition of E. coli from uninduced to an induced state. Using more modern computational models, the team is able to simulate the effect of repressor binding and unbinding on the transition in both directions – both to induced and uninduced state. The use of lactase uptake pathway in E.coli bacteria in this experiment allows for a much more effective – yet simpler – way of building models of complex genetic regulatory circuits. These models can inform the understanding of processes such as viral reproduction. One example is a study by Santillán and Michael, who studied the stability in the virus Bacteriophage-. Phage-, which can switch between lysogenic and lysis states during the process of subverting E.coli as its host, with a circuit that is complex, but can be better understood using models tat are developed by Bhogale et al. on gene regulation.
2. Hughes et al.’s study aimed to find molecular interaction that can trigger fibril formation -amyloid peptides through their self-interaction. They manipulated strains of E.coli and yeast cells to produce either “bait” or “prey” plasmids that facilitate self-interactions of A peptides. In order to measure the formation, yeast cells were used that carry -galactosidase genes as one of the “reporters”. Neither the bait nor the prey alone have the ability to activate gene expression, but if they interact, the result can trigger conditional expression of reporter genes. Hence, rate of interaction between A is linked to the production of -gal, which in turn corresponds to the rate of hydrolysis of ONPG.
3. The main innovation of the study by Shim et al. (2009) is the use of 2D IR spectroscopy, which allows for specific and direct observation of the formation of fibrils. Shim et al. used 2D IR spectra to target specific markers at different frequencies so they can measure different rates of interaction between different points on the peptide chain, giving them much more specific insights into the mechanism of fibril formation. By contrast, Hughes et al.’s study, using ONPG, could only observe the rate of interaction between fibroid-forming peptides through multiple layers of mediums and conditional expressions. As a result, while they were able to measure the rate and genetic aspect of fibril formation, they are unable to observe the specific kinetics of fibril formation.
4. The effect of glucose on -gal production is two-fold: First, there is a default inhibitory effect of glucose, that its presence suppresses the production of -gal as well as its transporter permease. From this point, we can modify the original equation (see derivation A) by adding a multiplier that account for the rate of inhibition if glucose is present:

Where is the concentration of glucose in the medium. From this equation we can see that

Which explains the behavior of inhibition as glucose becomes increasingly present, while when it is absent, the rate of production of -gal is unaffected.

The second aspect of the relationship is the complication of “pre-induction” of bacteria that Novick and Weiner noted, where if the internal concentration of inducers is high enough, they can resist the inhibitory effect of glucose and continue to produce -gal. Hence we modify the previous equation to:

With being the maximum internal concentration of inducers in a fully induced cell. Now instead of the absolute measure of whether glucose is present, we use whether the concentration of glucose is higher than the existing concentration of inducers. As the concentration of glucose surpasses that of the inducer, -gal production halts, but the higher is, the more glucose is needed to inhibit production.

## References:

Novick paper; textbook; <https://journals.asm.org/doi/epdf/10.1128/jb.60.4.381-392.1950> Hughes et al (PNAS, 93 2065-2070, 1996) Monod experiment

## Appendix:

Data Set A:

| t | concentration |
| --- | --- |
| 0.1699 | 0.0190 |
| 0.4426 | 0.0855 |
| 0.5111 | 0.1164 |
| 0.7156 | 0.1639 |
| 1.0564 | 0.2470 |
| 1.2041 | 0.2803 |
| 1.4311 | 0.3278 |
| 1.6465 | 0.3563 |
| 1.8283 | 0.4038 |
| 2.1119 | 0.4537 |
| 2.4182 | 0.5059 |
| 2.9970 | 0.6152 |
| 3.6763 | 0.6746 |
| 4.2308 | 0.7126 |
| 4.8316 | 0.7981 |
| 5.4070 | 0.7601 |
| 6.0646 | 0.8599 |
| 6.6638 | 0.8741 |

Data Set B:

| t | concentration |
| --- | --- |
| 2.3832 | 0.0109000 |
| 2.7230 | 0.0093429 |
| 3.0358 | 0.0146000 |
| 3.5177 | 0.0145000 |
| 4.0014 | 0.0242000 |
| 4.5692 | 0.0286000 |
| 5.0528 | 0.0376000 |
| 5.5358 | 0.0435000 |
| 6.4458 | 0.0592000 |
| 7.5262 | 0.0764000 |
| 8.5809 | 0.1071000 |
| 9.6318 | 0.1182000 |
| 10.8242 | 0.1278000 |
| 12.0208 | 0.1600000 |
| 13.1884 | 0.1885000 |
| 14.2958 | 0.1989000 |
| 22.7088 | 0.3173000 |
| 24.8690 | 0.3486000 |
| 26.7739 | 0.3784000 |
| 28.9053 | 0.4075000 |