Molecular Mechanisms for Human Diseases: Lab 1

Bacteria’s Brain – Part A

Peter Yibo Pan, collaborators:

(2023-02-08)

## Abstract

## Experiment

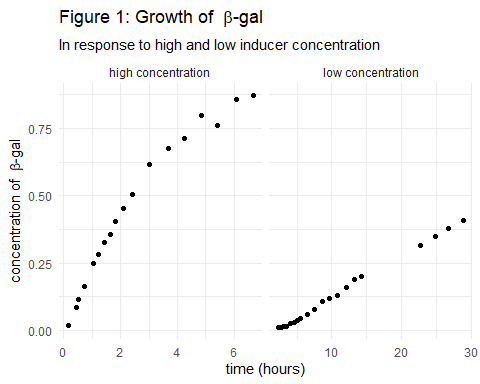
### Method

### 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:

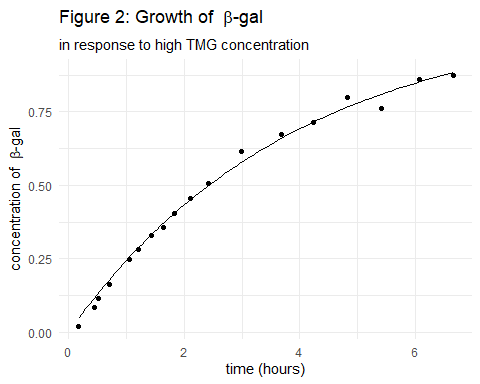
Which fit the form:

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

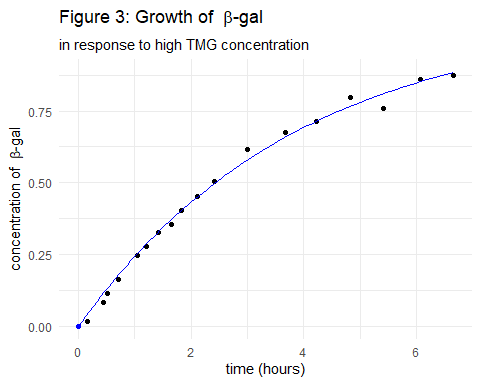
(questions A.1, A.2)  
From the derived equation A: , we set the two parameters and so the equation is of the form:



|  | 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:

We realize that this equation does not fit the initial condition of , hence we adding the term with being a free constant (not to be confused with from derivation A), which we adjust to to satisfy this initial condition, resulting in:

Now we replace and $ $ with their physical counterparts:

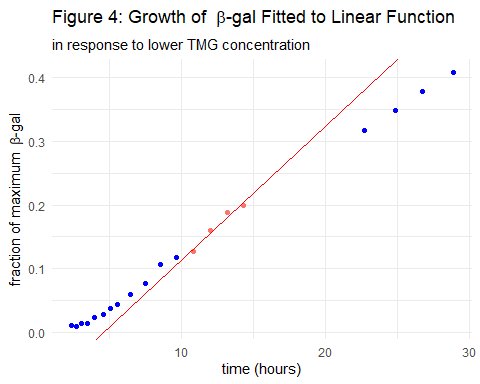
Which fit the form with and .

### 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

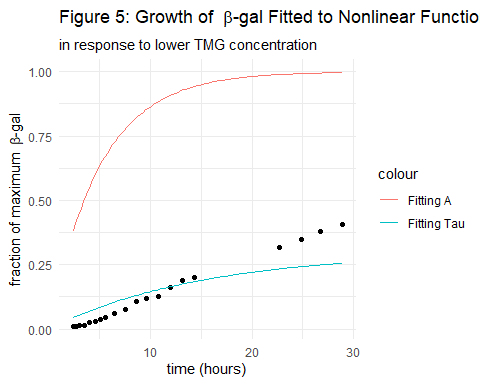


## Warning: 'newdata' had 20 rows but variables found have 4 rows

## # A tibble: 2 × 2  
## MSE data   
## <dbl> <chr>   
## 1 0.0000272 10<= t <= 15   
## 2 0.0176 entire dataset

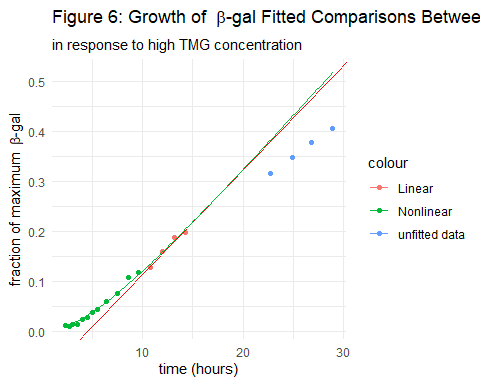
As shown in figure 4, the straight line (in red) is fitted over the data between 10 and 15 hours, over which the line fits quite well. the Mean Standard Error (MSE) for the model is much smaller within the range of $10\<= t \<= 15$ with a MSE of 2.7199507^{-5}, but not well over the entire data set with a MSE of 0.0175584. 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 (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



##   
## Formula: b\_q4\_df$concentration ~ growth\_b4(t, beta, tau)  
##   
## Parameters:  
## Estimate Std. Error t value Pr(>|t|)   
## beta 0.11755 0.02959 3.973 0.00139 \*\*   
## tau 5.34585 0.89590 5.967 3.44e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.005333 on 14 degrees of freedom  
##   
## Number of iterations to convergence: 6   
## Achieved convergence tolerance: 2.284e-06

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

## Appendix:

Data Set A:

| t | concentration |
| --- | --- |
| 0.170 | 0.0190 |
| 0.443 | 0.0855 |
| 0.511 | 0.1160 |
| 0.716 | 0.1640 |
| 1.060 | 0.2470 |
| 1.200 | 0.2800 |
| 1.430 | 0.3280 |
| 1.650 | 0.3560 |
| 1.830 | 0.4040 |
| 2.110 | 0.4540 |
| 2.420 | 0.5060 |
| 3.000 | 0.6150 |
| 3.680 | 0.6750 |
| 4.230 | 0.7130 |
| 4.830 | 0.7980 |
| 5.410 | 0.7600 |
| 6.060 | 0.8600 |
| 6.660 | 0.8740 |

Data Set B:

| t | concentration |
| --- | --- |
| 2.38 | 0.01090 |
| 2.72 | 0.00934 |
| 3.04 | 0.01460 |
| 3.52 | 0.01450 |
| 4.00 | 0.02420 |
| 4.57 | 0.02860 |
| 5.05 | 0.03760 |
| 5.54 | 0.04350 |
| 6.45 | 0.05920 |
| 7.53 | 0.07640 |
| 8.58 | 0.10700 |
| 9.63 | 0.11800 |
| 10.80 | 0.12800 |
| 12.00 | 0.16000 |
| 13.20 | 0.18800 |
| 14.30 | 0.19900 |
| 22.70 | 0.31700 |
| 24.90 | 0.34900 |
| 26.80 | 0.37800 |
| 28.90 | 0.40800 |