Bi 1x: LacI Titration

1 Overview

In this experiment, you will measure the effect of various mutated LacI repressor ribosome binding sites in an *E. coli* cell by measuring the expression of a yellow fluorescent protein. Using your skills of high-magnification phase and fluorescent microscopy coupled with some image processing tools, you will be able to directly compare a theoretical description of transcription with experimental measurements!

2 Background

In this section, we will cover the basics of gene regulation as well as give a more detailed explanation of our experimental system.

2.1 Genetic Regulation

What makes a human eye cell different from a skin cell? Both have the same set of genes, but clearly, they differ substantially in both structure and function. The reason, in part, is that although both cells share the same genes, they have different profiles of *gene expression*. In our journal clubs and discussions, we talked much about how genes are inherited across the biological scales and how we can identify the *genotype* of an organism by making observations of its *phenotype*, but we did not discuss in much detail how the genes supported an observable phenotype. The question of how gene expression is regulated is a major area of inquiry in contemporary biology, both because of its importance to biological function and because of the richness and diversity of gene regulation strategies that can be found in nature. Gene regulation occurs at all steps along the central dogma, from methylation of genomic DNA (known as *epigenetics*) to the chemical modifications made to proteins after translation that modulate their function or localization. In this module, we will explore regulation of *transcription* using the LacI repressor.

Transcriptional regulation in prokaryotes is largely mediated by proteins that bind to DNA, called "transcription factors" (TFs). Most TFs bind specifically to a particular DNA sequence known as an *operator*. For instance, the *lac* repressor protein binds strongly to the sequence `AATTGTGAGCGCTCACAATT`. TFs can be broadly categorized as either *activators* or *repressors*. Activators, as you can imagine, activate gene transcription by exhibiting favorable energetic interactions with the transcription machinery, such as RNA polymerase (RNAP). When an activator is bound near an RNAP binding site (known as a "promoter"), it increases the rate of transcription by recruiting RNAP to its binding site. Repressors are proteins which inhibit gene transcription, usually by making the RNAP binding site less accessible. Many genes have multiple TF binding sites surrounding the promoter region, granting the cell exquisite control over how often and how much the gene is expressed. Often, the timing and abundance of a protein can mean life or death for the cell.

2.2 The Lac Operon

The Lac operon is the canonical example of genetic regulation at the transcriptional level. In the perfect world of an *E. coli* cell, glucose is plentiful, antibiotic is absent, and there are no pesky Bi 1x students around to squish them onto agar pads and blast them with high-intensity light. One can imagine that if glucose is present, there is no need to waste energy into building the machinery required to digest other carbon sources, such as lactose. However, if there is no glucose present but lactose is plentiful, it would be in the organism's best interest to turn on the expression of genes to help metabolize lactose. The Lac operon is sensitive to the concentrations of both glucose and lactose and initiates expression of the lactose metabolism genes (LacZ, LacY, and LacA) under the appropriate conditions. The general architecture of the lac operon is shown in Fig. 1.

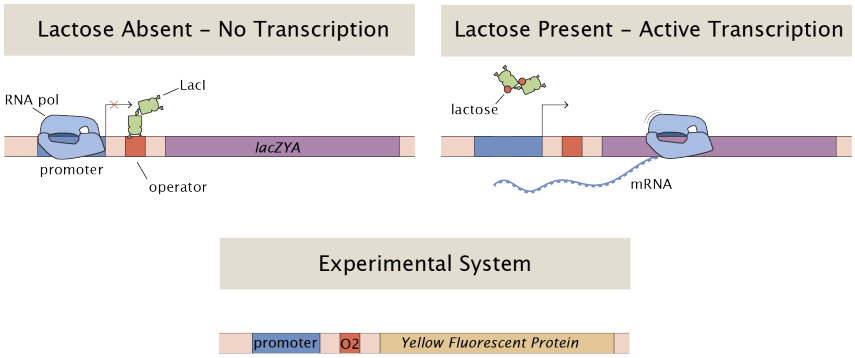

Figure 1: The lactose operator. In absence of lactose, the repressor LacI is bound to the operator, preventing transcription by RNA polymerase. In the presence of lactose (and the absence of glucose, not shown), LacI is released from the DNA and transcription can proceed. The yellow fluorescent protein (YFP) gene with the Lac repression promoter architecture has been inserted into the chromosome of *E. coli*, which will enable us to measure gene expression for different strains, each of which have a different number of LacI repressors.

The level of repression of the Lac operon is dependent on the number of LacI repressors present. If, for example, there were no LacI molecules in an *E. coli* cell, the LacZYA genes would be constitutively expressed (i.e. the gene would be read by polymerase as fast as possible). In our experiment, we will not measure the amount of LacI produced, but rather the abundance of yellow fluorescent protein (YFP). Prior to the lab, we have inserted a gene encoding YFP with the Lac operator O2 (Fig. 1).

Intuitively, one would expect that the higher the concentration of LacI in the cell, the lower the level of gene expression one would observe. We can codify this intuition mathematically by defining a quantity called the **repression**, where

$$\text{Repression} = \frac{\text{Average gene expression in the absence of repressor}}{\text{Average gene expression in the presence of repressor}}$$

$$= \frac{\text{Average YFP fluorescence in the absence of repressor}}{\text{Average YFP fluorescence in the presence of repressor}}. \quad (1)$$

In other words, the repression tells us how "strong" the repressor is by comparing the amount of gene expression without any repressor to the amount of gene expression with the repressor present. Using the tools of statistical physics we will discuss in class, we can quantify repression using the physical parameters of our system:

$$\text{Repression} = 1 + \frac{R}{N_{NS}} \, \mathrm{e}^{-\Delta E_{rd}/k_B T}, \quad (2)$$

where R is the number of LacI repressor binding heads, N_{NS} is the number of non-specific binding sites for the LacI repressor, ΔE_{rd} is the energetic difference between the bound and unbound states of the LacI repressor, and $k_B T$ is the thermal energy (Boltzmann's constant multiplied by the temperature). The purpose of this lab is to test this theoretical model of repression against experimental data which we will collect.

When testing a theoretical model, it is useful to have an experiment that allows you to independently tune the parameters of your model. In this experiment, you will test this model by tuning R. You will do this by using a set of *E. coli* strains that have been genetically engineered to express varying amounts of LacI.

Below we list the strains and their corresponding repressor copy numbers. With this information, together with the values of N_{NS} and ΔE_{rd} given in Problem 4, you can theoretically predict the repression for each strain using equation (2).

3 Protocol

In the *E. coli* growth module, you imaged cells growing squished between an agarose pad and the bottom of a glass-bottom dish. We will use the same methodology to image our cells for this experiment. Your kind and skilled TAs have poured the agarose pads—made of M9 minimal media + 0.5% (w/v) glucose—onto glass slides. You will also be provided with six diluted cell mixtures to mount. These numbers correspond to the following strains:

1. *Autofluorescence* - This strain has wild type levels of LacI, but does not have the YFP reporter. This will be a control to measure autofluorescence.
2. Δ*LacI* - This strain produces no LacI and hence has 0 repressors. This measurement will become the numerator of equation (1).
3. *R22* - This strain has wild type levels of LacI and the YFP reporter. This strain has on average 22 repressors/cell.
4. *R60* - This strain has on average 60 repressors/cell.

5. *R124* - This strain has on average 124 repressors/cell.
6. *R260* - This strain has on average 260 repressors/cell.

The three mutants are versions of LacI where the ribosomal binding site sequences have been modified to alter the amount of LacI present in each cell. To prepare the samples, follow the protocol given below.

- Using a clean razor blade, cut out six different 0.5cm x 0.5cm slabs (one for each strain). **Be careful with the razor blade.**
- Organize the pads on the parafilm sheet. Add 1µL of each cell solution to an agarose pad. **Be sure to write down the order!**
- Cover with the provided empty pipette tip container and allow to dry for 5 minutes. Be sure to record which strain is on which pad.
- Using the provided mounting tool, place the pads onto the dish. **Again, write down the position of each strain!** It is incredibly important that you do not mix up the samples. To prevent any confusion, mount the pads as shown in Fig. 2 when viewed from the top. Note that you mount the pads onto the side of the microscopy dish with the glass bottom (i.e. the thicker side).
- Place the lid on the glass-bottom dish and wrap with the provided parafilm strip. This will hold the lid in place when you move the microscope stage.

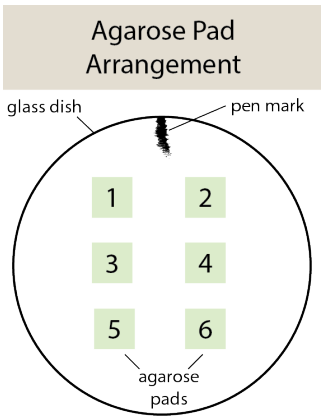

Figure 2: Agarose pad arrangement for your experiment. This should be the arrangement of agarose pads when looking from the **top** of the dish.

For each pad, your goal is to image at least 100 cells on the 100x objective (this will require between 5-10 images). Make sure to check your Köhler illuination periodically during your imaging.

To ease segmentation, it is best not to have any clumps of cells in your field of view; individual cells are best. However, having clumps of cells which you may not

be able to separate by segmentation does not invalidate the experiment. You need to think about data analysis choices if you have clumps of cells together (see Problem 3).

At each position you should take a phase contrast image (to see where the cells are with your eyes), a green FITC fluorescence image (to measure lactose gene expression), and a red fluorescence image (to allow easy segmentation).

Since you will be quantitatively comparing different images, **it is absolutely essential that each image is taken with the same exposure setting**. To determine the appropriate exposure time, pick a field of cells on the Δ LacI pad and test a range of exposure time ranging from 100 ms to several seconds. **Do not use this position in your data set (why?)**. An appropriate exposure time will use 1/2 to 2/3 of the dynamic range of your camera.

4 Assignment

Problem 0 (Summary)**.**
Write a brief summary of the experimental objectives, observations, and conclusions between one paragraph and one page in length.

Problem 1 (Modalities of gene expression)**.**
From your understanding of gene expression, speculate on which types of genes should be transcriptionally regulated using repressors? How about using activators? Can you think of any case in which it would be advantageous for a gene to have little to no regulation at all? There is no right or wrong answer to this question. It is meant to get you thinking about regulation of gene expression as it relates to an organism's activities and struggles.

Problem 2 (Example of your image overlay)**.**
For a phase/fluorescent image of your choice, please show the following:

1. The original phase image.
2. The mask generated by your segmentation procedure in the red channel.
3. An overlay of the segmentation mask on the original phase image.
4. An overlay of the segmentation mask and the green fluorescence image.

Be sure to include which pad your image came from as well as the code used for the processing.

Problem 3 (Integrated or mean intensity?)**.**
When analyzing the fluorescent intensity in each cell, should you use the mean or

integrated intensity in your subsequent analyses? Give arguments for your choice. In the next problems, when we refer to "fluorescence" of a cell or group of cells, we mean either the integrated or mean intensities, whichever you choose.

Problem 4 (Comparing theory with experiment)**.**

a) Using the Autofluorescence strain, compute and report the mean autofluorescence pixel value. In other words, for each of your Autofluorescence images, compute the fluorescence intensity of each cell, and then find the mean of these values.

b) For each of the other strains, calculate the autofluorescence-corrected mean fluorescence intensity of each detected cell (i.e., the mean YFP fluorescence minus the mean YFP autofluorescence).

c) For each of the four experimental strains, that is for the wild type (R22) and for the RBS mutants, compute the repression as defined by equation (1). That is, compute

$$\begin{aligned}\text{Repression} &= \frac{\langle \text{Gene expression in the absence of repressor} \rangle}{\langle \text{Gene expression in the presence of repressor} \rangle} \\ &= \frac{\langle \text{YFP fluorescence in the absence of repressor} \rangle}{\langle \text{YFP fluorescence in the presence of repressor} \rangle}. \end{aligned} \tag{3}$$

Here, the notation $\langle \cdot \rangle$ refers averaging over all of the cells in your images. Remember to do background subtraction. Be sure to compute error bars, which you can do using bootstrapping.

d) In class, we covered the basis of the statistical mechanical expression for repression defined by

$$\text{Repression} = 1 + \frac{R}{N_{NS}} \, \mathrm{e}^{-\Delta E_{rd}/k_B T}.$$

We stated that N_{NS} is the number of possible non-specific binding sites for LacI which can be approximated by the length of the *E. coli* genome, 4.6×10^6 base pairs. Why is this a good estimate? ΔE_{rd} is the energetic difference of LacI bound to the operator versus unbound and is approximately $-14.3 k_B T$ (Garcia and Phillips, *PNAS*, 2011). Plot the theoretical curve of repression as a function of repressor copy number overlaid with your experimental results. To plot the error bars, you can for a `for` loop to plot individual lines between the ends of the error bars (though there are other ways to do it). Comment on any discrepancies.

e) What are the limits of the experiment? In particular, if we gave you another RBS mutant strain with a repressor copy number of 1000, what would you expect its repression to be?

Problem 5 (Cell-to-cell variability in gene expression)**.**
As you probably noticed while taking your images, not all *E. coli* cells on the same pad exhibit the same level of fluorescence. This cell-to-cell variability in gene expression is known as the *noise* or *variability* in expression. A common measure of noise is the coefficient of variation, the ratio of the standard deviation of gene expression levels to the mean. This is generically called the *noise*, η, and is calculated by computing the variance of integrated YFP fluorescence intensity, I_{YFP}, across your cells and dividing by the mean I_{YFP}.

$$\eta^2 = \frac{\sigma^2_{\mathrm{YFP}}}{\langle I_{\mathrm{YFP}} \rangle^2} = \frac{\langle I^2_{\mathrm{YFP}} \rangle - \langle I_{\mathrm{YFP}} \rangle^2}{\langle I_{\mathrm{YFP}} \rangle^2}.$$

Compute the noise for each of your four experimental strains (WT and the RBS mutants). Make a scatter plot of the noise versus the mean for the three mutants and the wild type strain. Does the noise increase or decrease with increased gene expression?