Gene Expression

Daniel Jones

May 9, 2011

1 Introduction

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*. 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, to post-translational modification of proteins. Today, we will focus on how gene expression is regulated at the level of transcription initiation, or "transcriptional regulation," for short.

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. For instance, the *lac* repressor protein binds strongly to the sequence AATTGTGAGCGCTCACAATT. TFs can be broadly categorized as either activators or repressors. Activators are proteins which exhibit favorable energetic interactions with RNA polymerase (RNAP). When an

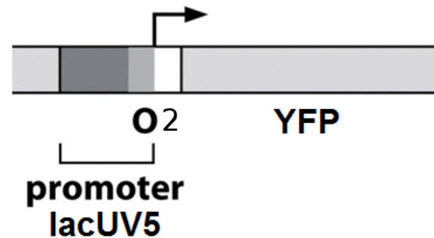
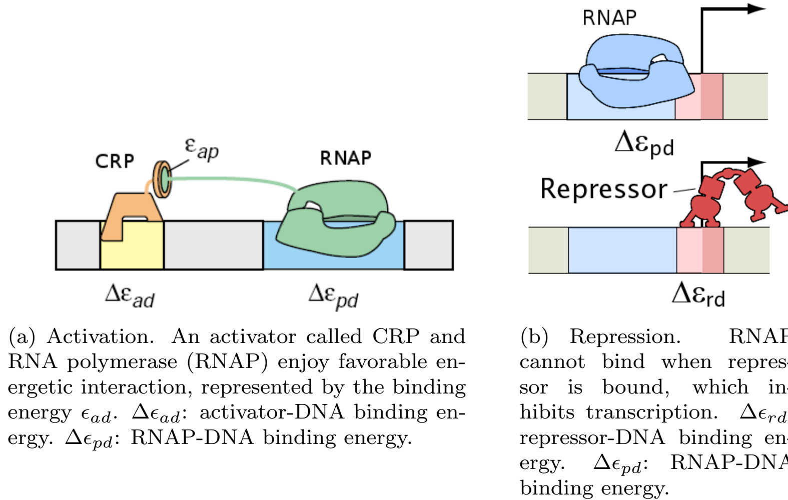


Figure 1: Schematic of the promoter architecture to be investigated in this experiment. It consists of a lacUV RNAP binding site, immediately followed by an LacI binding site (shown here as O2). The promoter is driving expression of YFP.



(a) Activation. An activator called CRP and RNA polymerase (RNAP) enjoy favorable energetic interaction, represented by the binding energy ϵ_{ad}. $\Delta\epsilon_{ad}$: activator-DNA binding energy. $\Delta\epsilon_{pd}$: RNAP-DNA binding energy.

(b) Repression. RNAP cannot bind when repressor is bound, which inhibits transcription. $\Delta\epsilon_{rd}$: repressor-DNA binding energy. $\Delta\epsilon_{pd}$: RNAP-DNA binding energy.

Figure 2: Simple transcriptional regulatory motifs.

activator is bound near an RNAP binding site, it increases the rate of transcription by recruiting RNAP to its binding site (see figure 2a). Repressors are proteins that inhibit transcription, usually by making the RNAP binding site inaccessible to RNAP and thus preventing transcription initiation (see figure 2b). The term “promoter architecture” is often used to refer collectively to the number, positions, and strengths of TF binding sites associated with a given gene.

2 Experimental System

Today, we will use fluorescence microscopy to measure expression of YFP (yellow fluorescent protein) as regulated by a very simple promoter architecture. Our promoter has an RNAP binding site, immediately followed by a lac repressor (LacI) binding site, followed by the YFP gene (see figure 1). When LacI is bound to its binding site, it blocks RNAP from transcribing YFP. Intuitively, we would expect that the higher the concentration of LacI in the cell, the lower the level of gene expression we will observe, since LacI blocks expression of YFP. We can codify this intuition mathematically by definining a quantity called the **repression**, where

$$\text{Repression} = \frac{\text{Gene expression in the absence of repressor}}{\text{Gene expression in the presence of repressor}}. \tag{1}$$

$$= \frac{\text{YFP Fluorescence in the absence of repressor}}{\text{YFP Fluorescence in the presence of repressor}} \tag{2}$$

In other words, the repression tells us how many times less a gene is expressed with LacI present than it would be if LacI were not present. Using the tools

of statistical physics, we can take the next step and compute how this quantity depends on the physical parameters of our system:

$$\text{Repression} = 1 + \frac{R}{N_{NS}} e^{-\Delta\epsilon_{rd}/kT} \tag{3}$$

where R is the number of repressor proteins in the cells, N_{NS} is the number of nonspecific binding sites for LacI in the *E. coli* genome, and $-\Delta\epsilon_{rd}$ is the binding energy for LacI and its binding site.

When testing a theoretical model, it's usually best if your experiment allows you to independently tune the parameters of your model. In this experiment, you will test this model by tuning $\Delta\epsilon_{rd}$, the LacI-DNA binding energy. Each of the strains you will observe has a slightly different LacI binding *sequence*, and these differences in sequence manifest themselves as differences in binding *energy*. However, we won't tell you in advance what the binding energy is for each strain; instead, you will compute these energies as part of your homework assignment.

3 Experimental Protocol

Your TAs have prepared your samples ahead of time. Each group will be given a glass bottom dish with 6 agar pads mounted on it.

- Pad 1 is strain HG105::O2. This strain produces no LacI protein and will become the numerator of your repression measurement (see equation 1).
- Pad 2 is strain HG104::NoFluo. As its name implies, this strain does not produce YFP and will be used to correct your gene expression measurements for the presence of cellular autofluorescence.
- Pads 3-6 are the actual strains of interest. Since we're not telling you what they are, you can refer to them as Strain1, Strain2, etc. You can call them whatever you want as long as you keep track of which dish position they came from.

For each pad, you should take images at 5 different positions. Try to choose positions that have a reasonable number (~ 10) of *E. coli* in them, but without too many bacteria touching each other. At each position, you should take 1. a phase contrast image (channel "Brightfield" in MicroManager) 2. a fluorescence image (channel "FITC" or "YFP" in MicroManager). It's essential that you **use the same exposure for all fluorescence images.** Otherwise, you can't directly compare fluorescence intensities from images acquired with different exposures. To figure out what exposure to use, you'll want to start with your brightest sample, which in this case is strain HG105::O2 on pad 1 (why?). Looking for cells near the edge of the pad to avoid photobleaching on region of interest for your real data acquiring. Starting at an exposure of 200ms, adjust exposure time until your images use approximately 1/2 to 2/3 of the dynamic

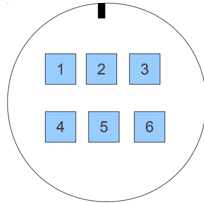


Figure 3: Schematic of pad layout on dish. This is the layout **as seen with the dish mounted on the microscope, with the black line facing up**.

range of your camera. Again, be sure to use the exposure you determine here for all fluorescence images. It will probably make your analysis easier if you use the same brightfield exposure for all your samples as well. You can use any sample to calibrate your brightfield exposures.

A tip: it will be easiest to acquire the images using the "Multi-D Acquisition" window in MicroManager. This way, you can easily set exposures for the brightfield and fluorescence images for the duration of the acquisition, and you can acquire both channels with a single mouse click. **A word of warning: be sure to set the channel back to brightfield before clicking "Live" in the main MicroManager window. Also, make sure that neither the "Time Series" nor "Use XY List" boxes are checked for the multi-D acquisition.** Finally, don't forget to change the "Directory" and "Prefix" fields when you switch to a new pad. Otherwise, you won't know which images correspond to which pad! Ask your TAs if this is unclear.

Here is a summary of the data acquisition protocol. This assumes that you have mounted the dish on your microscope with the 100x objective and the appropriate phase ring in place:

- Calibrate exposures:
  - Brightfield using any sample
  - Fluorescence using sample HG105::O2
- For each pad:
  - At 5 different positions:
    - Acquire brightfield image using exposure determined above
    - Acquire fluorescence image using exposure determined above