1 Introduction

From how a single cell develops into a multicellular organism to how bacteria decide to go about their diet, single cells interpret the information encoded in their DNA and in their surrounding media in order to make life-changing decisions. Cellular decision making is ubiquitous in biology. As an example, most animals have the same set of genes encoded in their DNA. However, what sets them apart is when, where and how each cell decided to produce those genes.

In this Matlab tutorial we will explore simple cellular decisions in the context of the bacterium *E. coli*. We will propose a theoretical model to describe these decisions and use Matlab to generate falsifiable predictions that can be tested experimentally by quantifying the fluorescence intensity of several *E. coli* strains we will provide. We will obtain this data using fluorescence microscopy and invoke Matlab once again to analyze our microscopy images. The result will be a full cycle of the theory/experiment interplay where we go from theoretical prediction to experimental validation on our quest to test our predictive understanding of cellular decision making.

Information about these decisions flows through the so-called “central dogma of molecular biology”, shown diagrammatically in Figure 1. Here, genes are encoded on the DNA. When a gene is turned “on” it is copied by the RNA polymerase molecular machine in a process deemed transcription. Both the DNA and mRNA molecules encode information in the familiar language of ATCGs for DNA and AUCGs for mRNA. The mRNA molecule is then translated by the ribosome molecular machine into a protein made out of amino acids. Gene expression can be regulated along any of the steps of this central dogma. For the purposes of this short tutorial we will focus on regulation at the level of transcription.
Figure 1: The central dogma of molecular biology. Genes encoded in the DNA that are turned “on” are copied by the molecular machine RNA polymerase into an mRNA molecule. This mRNA molecule is then translated by the ribosome molecular machine from the ATCG language of DNA and the AUCG language of mRNA into the protein language of amino acids. The regulation of gene expression can occur at any step along the central dogma.


2 A model of transcriptional regulation by simple repression

We begin by thinking of the simple case of transcriptional regulation in bacteria. Our first task is to model the special case where a gene is not regulated and is therefore produced at a constant rate. RNA polymerase knows which genes to transcribe and make an mRNA copy of because of a DNA sequence called the “promoter” which lies upstream, in the direction of the 5’ end, from genes. This promoter sequence is basically a landing pad for RNA polymerase and gives it a signal to initiate the process of transcription.

A very basic kinetic scheme illustrating this situation is shown in Figure 2. Here, a constitutive (unregulated) promoter can be bound by RNA polymerase leading to the production of mRNA at a rate $r$. These mRNA molecules are then degraded at a rate $\gamma$. This scheme can be summarized into an equation describing the time evolution of the mRNA concentration, $m(t)$, as

$$\frac{dm(t)}{dt} = r - \gamma m(t).$$

(1)

In steady state the mRNA concentration doesn’t change such that

$$\frac{dm(t)}{dt} = 0$$

(2)

leading to the steady-state concentration of mRNA

$$m = \frac{r}{\gamma}.$$  

(3)

This expression confirms our intuition about the interplay between production and degradation in determining the mRNA concentration. Of course, this is just the simple case of a constitutive promoter where there is no regulation and, as a result, where no decisions are being made.

Let’s now introduce one of the simplest regulatory strategies, namely repression. Here, a repressor binds to a site in the vicinity of the promoter such that RNA polymerase cannot bind to it and initiate transcription. This situation is illustrated in Figure 3. When the repressor is bound to its site no transcription is present. In contrast, if the promoter is not bound by repressor, RNA polymerase can bind to the promoter and produce RNA at a rate $r$ as in the case of the constitutive promoter of Figure 2. As we see, this promoter won’t always be in the transcriptionally active state. If we
Figure 2: Transcription of a constitutive promoter. In this unregulated case RNA polymerase binds to the promoter and produces mRNA molecules at a rate $r$. These mRNA molecules are then degraded at a rate $\gamma$.

Define $p_1$ as the probability of the promoter being in state 1, the production of mRNA is given by

$$\frac{d m(t)}{dt} = p_1 r - \gamma m(t)$$

and the steady-state concentration is now

$$m = p_1 \frac{r}{\gamma}.$$  

We now go ahead and calculate the probability $p_1$. A very simple way of thinking about the binding of repressor to the DNA is shown in the scheme in Figure 4. Here, repressor binds to promoter DNA in order to form a promoter-repressor complex. This reaction defines a dissociation constant $K_d$ given by

$$K_d = \frac{[P] [R]}{[P-R]}.$$  

The probability of the promoter not being bound by repressor, $p_1$, is the fraction of unbound promoters, namely

$$\text{Fraction of unbound promoters} = p_1 = \frac{[P]}{[P] + [P-R]}.$$  

If we use the definition of the dissociation constant from Equation 6 and multiply the numerator and denominator by $1/[P]$ we get

$$p_1 = \frac{[P]}{[P] + [P-R]} \frac{1/[P]}{1/[P]} = \frac{1}{1 + \frac{[P-R]}{[P]}} = \frac{1}{1 + \frac{[R]}{K_d}}.$$
Figure 3: Simple repression. A repressor can bind to the promoter excluding RNA polymerase from it. In state 1 RNA polymerase can bind and produce mRNA at a rate $r$ as in Figure 2. In state 2 the repressor is bound and no transcription is present.

From here we can also calculate the probability of the promoter being occupied by repressor since $p_2 = 1 - p_1$ leading to

$$p_2 = \frac{[R]}{K_d}. \quad (9)$$

We can now calculate the steady-state concentration of mRNA, which is given by

$$m = \frac{1}{1 + \frac{[R]}{K_d}} \cdot \frac{r}{\gamma}. \quad (10)$$

Even though this expression looks simple, it makes non-trivial predictions which we are going to explore theoretically in the next section and which we will test experimentally.

### 3 Making predictions: The fold-change

Expressions such as shown in Equation 10 make predictions about the steady-state concentration of mRNA molecules as a function of the repressor concentration $[R]$ and its binding affinity to DNA given by the dissociation constant $K_d$. Although it’s becoming more common thanks to techniques
Figure 4: Repressor binding to the promoter. The repressor present at a concentration $[R]$ binds to the available promoter which is present at a concentration $[P]$. The result is a promoter-repressor complex at concentration $[R-P]$. This simple binding reaction defines the dissociation constant $K_d$.

such as FISH, RT-PCR, and RNA-seq, the direct measurement of mRNA molecules as predicted by Equation 10 can be challenging. In order to simplify this prediction further we will define a quantity that can be measured more easily experimentally, namely the fold-change in gene expression. This fold-change is the ratio of the mRNA concentration in the presence of repressor and the mRNA concentration in the absence of repressor, namely

$$
\frac{m([R] \neq 0)}{m([R] = 0)} = \frac{1}{1 + \frac{[R]}{K_d}}.
$$

(11)

We see that the factors $r/\gamma$ cancel out resulting in

$$
\text{fold-change} = \frac{1}{1 + \frac{[R]}{K_d}}.
$$

(12)

In Figure 5 we plot this fold-change in gene expression as a function of repressor concentration, one of the experimental “knobs” we can control in order to tune regulatory response. Different $E. \ coli$ strains can be engineered to contain a varying number of repressors. This fold-change is also shown for different values of the dissociation constant of repressor and its binding site. This second “knob” can be modulated by changing the 21 base pairs that make the DNA sequence of the binding site. After generating this same plot in Matlab we will move forward to actually performing the measurements necessary to test the predictions of this simple model.
Figure 5: Fold-change and simple repression. Predictions for the fold-change in gene expression in simple repression from Equation 12. The colors correspond to different values of the dissociation constant $K_d$. The dials represent the different knobs that can be tuned in order to modulate the level of gene expression. These knobs are the intracellular concentration of repressor and the strength of the repressor binding site which can be controlled through its DNA sequence.
Figure 6: Measuring fold-change using fluorescence microscopy. The fold-change in gene expression is defined as the ratio of the levels of gene expression coming from a strain bearing the transcription factor of interest over a strain with a deletion of such transcription factor. For each one of these two strains, the average fluorescence per cell is measured at the single cell level.

4 Measuring gene expression

The theoretical version of the fold-change in gene expression from Equation 12 has an experimental counterpart shown in Figure 6. By measuring the level of gene expression in bacteria that have the repressor and normalizing it by the level of gene expression of bacteria without the repressor we can obtain this experimental fold-change magnitude. Note that instead of measuring mRNA copy number we will measure fluorescence, as the mRNA we will use codes for the fluorescent protein YFP. Hence, we will use fluorescence intensity as a proxy for the level of gene expression.

The logical progression associated with this analysis is introduced schematically in Figure 7. Note that we have images of the cells in two different channels. In particular, for each field of view, we have both a phase contrast image and a fluorescence image. Like with the example where we determined the cell cycle time of E. coli, the first step is to find the cells in an automated fashion using some segmentation scheme. Additionally, we need to choose which one of the two images we want to do the segmentation with. Detecting cells using the fluorescence image is certainly appealing due to the absence of any other fluorescent objects. However, it is clear that for dimmer cells the segmentation might not work as well. As a result we would risk biasing our segmentation based on the level of expression of the cells, which
is the quantity we are actually interested in measuring! We then choose to segment the phase contrast image which should, in principle, not be subject to bias resulting from the level of fluorescence within each cell.

Following the procedure outlined in the example on the cell division time in *E. coli*, once we have performed the thresholding, we will be left with a mask image with discrete regions that we identify as cells denoted by the different colors in Figure 7C. Once the segmentation process is complete, we can then obtain the fluorescence intensity in each of our cells. To do so, we use the segmented image from the previous step to find the individual cells and then within each such cell we ask for the fluorescence intensity of all of the pixels and sum them up. The result is a distribution of fluorescence per cell as shown in Figure 7E. However, there is an extra subtlety that has to be taken into account when obtaining such fluorescence distributions. In particular, because of the intrinsic fluorescence of the cells themselves, there is a spurious contribution to the total fluorescence we measure, namely, $F_{\text{total}}$, is given by

$$F_{\text{total}} = F_{\text{reporter}} + F_{\text{cell}},$$

where $F_{\text{reporter}}$ is the signal stemming from the fluorescent reporter, while $F_{\text{cell}}$ is the autofluorescence of the cell. As a result we need to be able to subtract the cells’ average autofluorescence if the want to report only on $F_{\text{reporter}}$. This can be easily done by following the steps outlined in Figure 7 and described above, but now for a strain of bacteria that lacks any fluorescent reporter. We will then be able to measure the mean contribution of the cell autofluorescence to the total fluorescence, $\langle F_{\text{cell}} \rangle$, which can be subtracted from the fluorescence values in the presence of the reporter.

With the fluorescence intensities in hand, we are now prepared to compute the fold-change itself so that we can examine the accord between the model of simple repression presented in Equation 12 and the data itself.

5 Experimental protocol

We will prepare samples with different strain of *E. coli*. These *E. coli* will be sandwiched between an agar pad and a coverslip, which we will show you how to prepare. Each group will be in charge of measuring the fold-change in gene expression for different values of the binding energy and the intracellular repressor concentration.

Before taking the data we need to settle on the imaging conditions, which are microscope-dependent. For example, for the fluorescence it is important to make sure that the camera is not being saturated. Pick the brightest
Figure 7: Schematic of the image segmentation algorithm to quantify levels of gene expression in bacteria. Two images of bacteria expressing a fluorescent protein are obtained, (A) one in phase contrast and (B) one in fluorescence. The phase contrast image is an imaging scheme that makes it possible to see the bacteria as dark objects. (C) These objects are automatically detected and segmented using computer software that assigns an identity to each segmented bacterium (represented by the different colors). (D) The mask generated by this procedure is applied to the fluorescence image in order to generate an overlay and integrate the fluorescence within the mask of each segmented cell. (E) By repeating this for multiple images and many cells, the distribution of fluorescence per cell can be computed.
strain, play with the exposure time and look at the pixel values in order to make sure that the images don’t have any saturated pixels.

Once you’ve converged on imaging conditions take several fields of view in both the fluorescence YFP channel and in phase contrast for each strain. We are aiming to have about 100 cells per strain. Remember that in order to make a full measurement we need to measure a strain of bacteria that has the repressor, the corresponding strain where the repressor has been deleted, and a strain with no fluorescent reporter in order to measure autofluorescence. As a result, each “measurement” will consist of three different independent measurements.

After taking the data, save it to the shared folder drive so you can retrieve it from the computers running Matlab in order to perform the image analysis on them.