Gene regulation: Introduction*

Introduction to gene regulation

Regulation is all about decision making. Gene regulation is, therefore, all about understanding how cells make decisions about which genes to turn on, turn off or to tune up or tune down. In the following section we discuss some of the fundamental mechanisms and principles used by cells to regulate gene expression in response to changes in cellular or external factors. This biology is important for understanding how cells adjust changing environments, including how some cells, in multicellular organisms, decide to become specialized for certain functions (e.g. tissues).

Since the subject of regulation is both a very deep and broad topic of study in biology, in Bis2a we don’t try to cover every detail - there are simply way too many. Rather, as we have done for all other topics, we try to focus on (a) outlining some of the core logical constructs and questions that you must have when you approach ANY scenario involving regulation, (b) learning some common vocabulary and ubiquitous mechanisms and (c) examining a few concrete examples that illustrate the points made in a and b.

Gene Expression

Introduction

All cells control when and how much each one of its genes are expressed. This simple statement - one that could be derived simply from observing cellular behavior - brings up many questions that we can begin to lay out using our Design Challenge rubric.
Trying to define "gene expression"

The first thing we need to do, however, is to define what it means when we say that a gene is "expressed". If the gene encodes a protein, one might reasonably propose that "expression" of a gene means how much functional protein is made. But what if the gene does not encode a protein, but rather some functional RNA. Then, in this case, "expressed might mean how much of the functional RNA is made. Yet another person might reasonably suggest that "expression" just refers to the initial step in creating a copy of the genomic information. By that definition, one might want to count how many full-length transcripts are being made. Is it the number of end products encoded by the genomic information or is it the number of reads of the information that is important to properly describe "expression". Unfortunately, in practice we often find that the definition depends on the context of the discussion. Keep that in mind. For the sake of making sure that we are talking about the same thing, in Bis2A we'll try to use the term "expression" primarily to describe the creation of the final functional product(s). Depending on the specific case, the final product may be a protein or RNA species.

The design challenge of regulating gene expression

To drive this discussion from a design challenge perspective, we can formally stipulate that the "big problem" we are interested in understating is that of regulating protein abundance in a cell. Problem: The abundance of each functional protein must be regulated. We can then start by posing subproblems:

Let's take a moment, though, first to reload a couple of ideas. The process of gene expression requires multiple steps depending on what the fate of the final product will be. In the case of structural and regulatory RNAs (i.e. tRNA, rRNA, snRNA, etc.) the process requires that a gene be transcribed and that any needed post-transcriptional processing take place. In the case of a protein coding gene, the transcript must also be translated into protein and if required, modifications to the protein must also be made. Of course, both transcription and translation are multi-step processes and most those sub-steps are also potential sites of control.

Some of the subproblems might therefore be:

1. It is reasonable to postulate that there must be some mechanism(s) to regulate the first step of this multi-step process, the initiation of transcription (just getting things started). So, we could state, "we need a mechanism to regulate the initiation of transcription." We could also turn this into a question and ask, "how can the initiation of transcription be accomplished"?

2. We can use similar thinking to state, "we need a mechanism for regulating the end of transcription" or to ask "how is transcription terminated?"

3. Using this convention we can state, "we need to regulate the initiation of translation and the stop of translation".

4. We've talked only about synthesis of protein and RNA. It is quite reasonable to also state, "we need a mechanisms to regulate the degradation of RNA and protein."
Focusing on transcription

In this course we begin by focusing primarily on examining the first couple of problems/questions, the regulation of transcription initiation and termination - from genomic information to a functional RNA, either ready as is (e.g. in the case of a functional RNA) or ready for translation. This allows us to examine some fundamental concepts regarding the regulation of gene expression and to examine a few real examples of those concepts in action.

Suggested discussion

Why is it important to regulate gene expression why not just express all genes all of the time?

Create a list of hypotheses with your classmates of reasons why the regulation of gene expression is important for bacteria and archaea and for eukaryotes. Instead of bacteria vs. eukaryotes, you may also want to consider contrasting reasons gene regulation is important for unicellular organisms versus multi-cellular organisms or communities of unicellular organisms (like colonies of bacteria).

Subproblems for transcription and the activity of RNA polymerase

Let us consider a protein coding gene and work through some logic. We start by imagining a simple case, where a protein-coding gene is encoded by a single contiguous stretch of DNA. We know that to transcribe this gene an RNA polymerase will need to be recruited to the start of the coding region. The RNA polymerase is not "smart" per se. There needs to be some mechanism, based on chemical logic, to help recruit the RNA polymerase to the start of the protein coding gene. Likewise, if this process is to be regulated, there needs to be some mechanism, or mechanisms, to dictate when an RNA polymerase should be recruited to the start of a gene, when it should not, and/or if it is recruited to the DNA whether or not it should actually begin transcription and how many times this process should happen. Note, that the previous sentence, has several distinct subproblems/questions (e.g. when is the polymerase recruited?; if recruited, should it start transcription?; if it starts transcription, how many times should this process repeat?). We can also reasonably infer, that there will need to be some mechanisms to "instruct" (more anthropomorphisms) the polymerase to stop transcription. Finally, since the role of transcription is to create RNA copies of the genome segments, we should also consider problems/questions related to other factors that influence the abundance of RNA, like mechanisms of degradation. There must be some mechanisms and these will likely be involved in the regulation of this process.
Activation and Repression of Transcription

Some basics

Let us consider a protein coding gene and work through some logic. We start by imagining a simple case, where a protein-coding gene is encoded by a single contiguous stretch of DNA. We know that to transcribe this gene an RNA polymerase will need to be recruited to the start of the coding region. The RNA polymerase is not "smart" per se. There needs to be some mechanism, based on chemical logic, to help recruit the RNA polymerase to the start of the protein coding gene. Likewise, if this process is to be regulated, there needs to be some mechanism, or mechanisms, to dictate when an RNA polymerase should be recruited to the start of a gene, when it should not, and/or if it is recruited to the DNA, whether or not it should actually begin transcription and how many times this process should happen. Note, that the previous sentence, has several distinct subproblems/questions (e.g. when is the polymerase recruited?, if recruited should it start transcription?, if it starts transcription, how many times should this process repeat?). We can also reasonably infer, that there will need to be some mechanisms to "instruct" (more anthropomorphisms) the polymerase to stop transcription.

Recruiting RNA polymerase to specific sites

To initiate transcription, the RNA polymerase must be recruited to a segment of DNA near the start of a region of DNA encoding a functional transcript. The function of the RNA polymerase as described so far, however, is not to bind specific sequences but rather to move along any segment of DNA. Finding a way to recruit the polymerase to a specific site therefore seems contradictory to its usual behavior. Explaining this contradiction requires us to invoke something new. Either transcription can start anywhere and just those events that lead to a full productive transcript do anything useful or something other than the RNA polymerase itself helps to recruit the enzyme to the beginning of a gene. The latter, we now take for granted, is indeed the case.

The recruitment of the RNA polymerase is mediated by proteins called general transcription factors. In bacteria, they have a special name: sigma factors. In archaea they are called TATA binding protein and transcription factor IIB. In eukaryotes, relatives of the archaeal proteins function in conjunction with numerous others to recruit the RNA polymerase. The general transcription factors have at least two basic functions: (1) They are able to chemically recognize a specific sequence of DNA and (2) they are able to bind the RNA polymerase. Together these two functions of general transcription factors solve the problem of recruiting an enzyme that is otherwise not capable of binding a specific DNA site. In some texts, the general transcription factors (and particularly the sigma factor varieties) are said to be part of the RNA polymerase. While they are certainly part of the complex when they help to target the RNA polymerase they do not continue with the RNA polymerase after it starts transcription.

The DNA site to which an RNA polymerase is recruited has a special name. It is called a promoter. While the DNA sequences of different promoters need not be exactly the same, different promoter sequences typically do have some
special chemical properties in common. Obviously, one property is that they are able to associate with an RNA polymerase. In addition, the promoter usually has a DNA sequence that facilitates the dissociation of the double stranded DNA such that the polymerase can begin reading and transcription the coding region. (Note: technically we could have broken down the properties of the promoter into design challenge subproblems. In this case we skipped it, but you should still be able to step backwards and create the problem statements and or relevant questions once you find out about promoters).

In nearly all cases, but particularly in eukaryotic systems the complex of proteins that assembles with the RNA polymerase at promoters (typically called the pre-initiation complex) can number in the tens of proteins. Each of these other proteins has specific function but this is far too much detail to dive into for Bis2A.
States of a regulated promoter

Since promoters recruit an RNA polymerase these sites and the assembly of the pre-initiation complex are obvious sites for regulating the first steps of gene expression. At the level of transcription initiation, we often classify promoter into one of three classes. The first is called constitutive. Constitutive promoters are generally not regulated very strongly. Their base state is "on". When the constitutive transcription from a promoter is very high (relative to most other promoters), we will colloquially call that promoter a "strong constitutive" promoter. By contrast, if the amount of transcription from a constitutive promoter is low (relative to most other promoters) we will call that promoter a "weak constitutive" promoter.

A second way to classify promoters by the use of the term activated or equivalently, induced. These interchangeable terms are used to describe promoters that are sensitive to some external stimulus and respond to said stimulus by increasing transcription. Activated promoters have a base state generally exhibits little to no transcription. Transcription is then "activated" in response to a stimulus - the stimulus turns the promoter "on".

Finally, the third term used to classify promoters is by the use of the term repressed. These promoters also respond to stimuli but do so by decreasing transcription. The base state for these promoters is some basal level of transcription and the stimulus acts to turn down or repress transcription. Transcription is "repressed" in response to a stimulus - the stimulus turns the promoter "off".

The examples given above assumed that a single stimulus acts to regulate promoters. While this is the simplest case,
many promoters may integrate different types of information and may be alternately activated by some stimuli and repressed by other stimuli.

**Transcription factors help to regulate the behavior of a promoter**

How are promoters sensitive to external stimuli? Mechanistically, in both activation and repression, require regulatory proteins to change the transcriptional output of the gene being observed. The proteins responsible for helping to regulate expression are generally called *transcription factors*. The specific DNA sequences bound by transcription factors are often called operators and in many cases the operators are very close to the promoter sequences.

Here's where the nomenclature gets potentially confusing - particularly when comparing across bacterial and eukaryotic research. In bacterial research, if the transcription factor acts by binding DNA and the RNA polymerase in a way that increases transcription, then it is typically called an *activator*. If, by contrast, the transcription factor acts by binding DNA to repress or decrease transcription of the gene then it is called a *repressor*.

Why are the classifications of activator and repressor potentially problematic? These terms, describe in idealized single functions. While this may be true in the case of some transcription factors, in reality other transcription factors may act to activate gene expression in some conditions while repressing in other conditions. Some transcription factors will simply act to modulate expression either up or down depending on context rather than shutting transcription "off" or turning it completely "on". To circumvent some of this possible confusion, some of your instructors prefer to avoid using the terms activator and repressor and instead prefer to simply discuss the activity of transcription various transcription factors as either a positive or a negative influence on gene expression in specific cases. If these terms are used, you might hear your instructor saying that the transcription factor in question ACTS LIKE/AS a repressor or that it ACTS LIKE/AS an activator, taking care not to call it simply an activator or repressor. It is more likely however that you will hear them say that a transcription factor is acting to positively or negatively influence transcription.

CAUTION: Depending on your instructor, you may cover a few real biological examples of positive and negative regulatory mechanisms. These specific examples will use the common names of the transcription factors - since the the examples will typically be drawn from the bacterial literature, the names of the transcription factors may include the terms "repressor" or "activator". These names are relics of when they were first discovered. Try to spend more time examining the logic of how the system works than trying to commit to memory any special properties of that specific protein to all other cases with the same label. That is, just because a protein is labeled as a repressor does not mean that it exclusively acts as a negative regulator in all cases or that it interacts with external signals in the exact same was as the example.

Suggested discussion

What types of interactions do you think happen between the amino acids of the transcription factor and the double helix of the DNA? How do transcription factors recognize their binding site on the DNA?

**Allosteric Modulators of Regulatory Proteins**

The activity of many proteins, including regulatory proteins and various transcription factors, can be allosterically modulated various factors, including by the relative abundance of small molecules in the cell. These small molecules are
often referred to as inducers or co-repressors co-activators and are often metabolites, such as lactose or tryptophan or small regulatory molecules, such as cAMP or GTP. These interactions allow the TF to be responsive to environmental conditions and to modulate its function accordingly. It is helping to make a decision about whether to transcribe a gene or not depending on the abundance of the environmental signal.

Let us imagine a negative transcriptional regulator. In the most simple case we've considered so far, transcription of gene with a binding site for this transcription factor would be low when the TF is present and high when the TF is absent. We can now add a small molecule to this model. In this case the small molecule is able to bind the negative transcriptional regulator through sets of complementary hydrogen and ionic bonds. In this first example we will consider the case where the binding of the small molecule to the TF induces a conformational change to the TF that severely reduces its ability to bind DNA. If this is the case, the negative regulator - once bound by its small molecule - would release from the DNA. This would thereby relieve the negative influence and lead to increased transcription. This regulatory logic might be appropriate to have evolved in the following scenario: a small molecule food-stuff is typically absent from the environment. Therefore, genes encoding enzymes that will degrade/use that food should be kept "off" most of the time to preserve the cellular energy that their synthesis would use. This could be accomplished by the action of a negative regulator. When the food-stuff appears in the environment it would be appropriate for the enzymes responsible for its processing to be expressed. The food-stuff could then act by binding to the negative regulator, changing the TF's conformation, causing its release from the DNA and turning on transcription of the processing enzymes.

![An abstract model of a generic transcriptional unit regulated by a negative regulator whose activity is modulated by a small molecule (depicted by a star). In this case, binding of the small molecule causes the TF to release from the DNA. Attribution: Marc T. Facciotti (own work)](https://bio.libretexts.org/Courses/University_of_California_Davis/BIS_2A_(2018)%3A_Introductory_Biology_(Singer)/MASTER...)
An abstract model of a generic transcriptional unit regulated by a negative regulator whose activity is modulated by a small molecule (depicted by a star). In this case, binding of the small molecule causes the TF to bind to the DNA.

Attribution: Marc T. Facciotti (own work)

Note how the activity of the TF can be modulated in distinctly different ways by a small molecule. Depending on the protein, the binding of this external signal can either cause binding of the TF-small molecule complex to DNA OR binding of the small molecule can cause the release of the TF-small molecule complex from the DNA. The same types of examples can be worked up for a positive regulator.

In both cases proposed above, the binding of a small molecule to a TF will be dependent on how strongly the TF interacts with the small molecule. This will depend on the types and spatial orientation of the protein's chemical functional groups, their protonation states (if applicable), and the complementary functional groups on the small molecule. It should not be surprising, therefore, to learn that the binding of the small molecule to the TF will be dependent on various factors, including but not limited to the relative concentrations of small-molecule and TF and pH.

Is it positive or negative regulation?

Resolving a common point of confusion

At this point, it is not uncommon for many Bis2a students to be slightly confused about how to determine if a transcription factor is acting as a positive or negative regulator. This confusion often comes after a discussion of the possible modes that stimulus (i.e. small molecule) can influence the activity of a transcription factor. This is not too surprising. In the examples above, the binding of an effector molecule to a transcription factor could have one of two different effects: (1) binding of the effector molecule could induce a DNA-bound transcription factor to release from its binding site, derepressing a promoter, and "turning on" gene expression. (2) binding of the effector molecule to the transcription factor could induce the TF to bind to its DNA binding site, repressing transcription and "turning off" gene expression. In the first case it might appear that the TF is acting to positively regulate expression, while in the second example it might appear that the TF is acting negatively.
However, in both examples above, the TF is acting as a negative regulator. To determine this we look at what happens when the TF binds DNA (whether a small molecule is bound to the TF or not). In both cases, binding of the TF to DNA represses transcription. The TF is therefore acting as a negative regulator. A similar analysis can be done with positively acting TFs.

Note that in some cases a TF may act as a positive regulator at one promoter and negative regulator at a different promoter so describing the behavior of the TF on a per case basis is often important (reading too much from the name it has been assigned can be misleading sometimes). Other TF protein can act alternately as both positive or negative regulators of the same promoter depending on conditions. Again, describing the behavior of the TF specifically for each case is advised.

A genetic test for positive or negative regulatory function of a TF

How does one determine if a regulatory protein functions in a positive or negative way? A simple genetic test is to ask "what happens to expression if the regulatory protein is absent?" This can be accomplished by removing the coding gene for the transcription factor from the genome. If a transcription factor acts positively, then its presence is required to activate transcription. In its absence, there is no regulatory protein, therefore no activation, and the outcome is lower transcription levels of a target gene. The opposite is true for a transcription factor acting negatively. In its absence expression should be increased, because the gene keeping expression low is no longer around.

Termination of Transcription and RNA degradation

Termination of transcription in bacteria

Once a gene is transcribed, the bacterial polymerase needs to be instructed to dissociate from the DNA template and liberate the newly made mRNA. Depending on the gene being transcribed, there are two kinds of termination signals. One is protein-based and the other is RNA-based. *Rho-dependent termination* is controlled by the rho protein, which tracks along behind the polymerase on the growing mRNA chain. Near the end of the gene, the polymerase encounters a run of G nucleotides on the DNA template and it stalls. As a result, the rho protein collides with the polymerase. The interaction with rho releases the mRNA from the transcription bubble.

*Rho-independent termination* is controlled by specific sequences in the DNA template strand. As the polymerase nears the end of the gene being transcribed, it encounters a region rich in C–G nucleotides. The mRNA folds back on itself, and the complementary C–G nucleotides bind together. The result is a stable *hairpin* that causes the polymerase to stall as soon as it begins to transcribe a region rich in A–T nucleotides. The complementary U–A region of the mRNA transcript forms only a weak interaction with the template DNA. This, coupled with the stalled polymerase, induces enough instability for the core enzyme to break away and liberate the new mRNA transcript.

Termination of transcription in eukaryotes

The termination of transcription is different for the different polymerases. Unlike in prokaryotes, elongation by RNA polymerase II in eukaryotes takes place 1,000–2,000 nucleotides beyond the end of the gene being transcribed. This pre-mRNA tail is subsequently removed by cleavage during mRNA processing. On the other hand, RNA polymerases I
and III require termination signals. Genes transcribed by RNA polymerase I contain a specific 18-nucleotide sequence that is recognized by a termination protein. The process of termination in RNA polymerase III involves an mRNA hairpin similar to rho-independent termination of transcription in prokaryotes.

**Termination of transcription in archaea**

Termination of transcription in the archaea is far less studied than in the other two domains of life and is still not well understood. While the functional details are likely to resemble mechanisms that have been seen in the other domains of life the details are beyond the scope of this course.

**Degradation of RNA**

The lifetimes of different RNA species in the cell can vary dramatically, from seconds to hours. The mean lifetime of mRNA can also vary dramatically depending on the organism. For instance, the median lifetime for mRNA in *E. coli* is ~5 minutes. The half-life of mRNA in yeast is ~20 minutes and 600 minutes for human cells. Some of the degradation is "targeted". That is, some transcripts include a short sequence that targets them for RNA degrading enzymes, speeding the degradation rate. It doesn't take too much imagination to infer that this process might also be evolutionarily tuned for different genes. Simply realizing that degradation - and the tuning of degradation - can also be a factor in controlling the expression of a gene is sufficient for Bis2a.

**Tuning termination**

Like all other biological processes, the termination of transcription is not perfect. Sometimes, the RNA polymerase is able to read through terminator sequences and transcribe adjacent genes. This is particularly true in bacterial and archaeal genomes where the density of coding genes is high. This transcriptional read-through can have various effects but the most common two outcomes are: (1) If the adjacent genes are encoded on the same DNA strand as the actively transcribed gene, the adjacent genes may also be transcribed. (2) If the adjacent genes are transcribed on the opposite strand of the actively transcribed genes, RNA polymerase read-through by interfere with polymerases actively transcribing the neighboring gene. Not surprisingly, biology has in some instances evolved mechanisms that both act to minimize the influence of read-through and to take advantage of it. Therefore, the "strength" of a terminator - and its effectiveness of terminating transcription in a particular direction - are "tuned" by Nature and "used" (note the anthropomorphism in quotes) to regulate the expression of genes.

![RNA Transcription Diagram](https://bio.libretexts.org/Courses/University_of_California_Davis/BIS_2A_(2018)%3A_Introductory_Biology_(Singer)/MASTER…)
An abstract model of a full basic transcriptional unit and the various "parts" encoded on the DNA that may influence the expression of the gene. We expect students in Bis2A to be able to recreate a similar conceptual figure from memory.

Attribution: Marc T. Facciotti (own work)

**Summary of gene regulation**

In the preceding text we have examined several ways to start solving some of the design challenges associated with regulating the amount of transcript that is created for a single coding region of the genome. We have looked in abstract terms at some of the processes responsible for controlling the initiation of transcription, how these may be made sensitive to environmental factors, and very briefly at the processes that terminate transcription and handle the active degradation of RNA. We have avoided more Each of these processes can be quantitatively tuned by nature to be "stronger" or "weaker". It is important to realize that the real values of "strength" (e.g. promoter strength, degradation rates, etc.) influence the behavior of the overall process in potentially functionally important ways.