23.2: Gene regulation: Eukaryotic

Eukaryotic Gene Regulation

Regulation overview

As was previously noted, regulation is all about decision making. Gene regulation, as a general topic, is related to making decisions about the functional expression of genetic material. Whether the final product is an RNA species or a protein, the production of the final expressed product requires processes that take multiple steps. We have spent some time discussing some of these steps (i.e. transcription and translation) and some of the mechanisms that nature uses for sensing cellular and environmental information to regulate the initiation of transcription.

When we discussed the concept of strong and weak promoters we introduced the idea that regulating the amount (number of molecules) of transcript that was produced from a promoter in some unit of time might also be important for function. This should not be entirely surprising. For a protein coding gene, the more transcript that is produced, the greater potential there is to make more protein. This might be important in cases where making a lot of a particular enzyme is key for survival. By contrast, in other cases only a little protein is required and making too much would be a waste of cellular resources. In this case low levels of transcription might be preferred. Promoters of differing strengths can accommodate these varying needs. With regards to transcript number, we also briefly mentioned that synthesis is not the only way to regulate abundance. Degradation processes are also important to consider.

In this section, we add to these themes by focusing on eukaryotic regulatory processes. Specifically, we examine - and sometimes re-examine - some of the multiple steps that are required to express genetic material in eukaryotic organisms in the context of regulation. We want you not only to think about the processes but also to recognize that each step in
the process of expression is also an opportunity to fine tune not only the abundance of a transcript or protein but also its functional state, form (or variant), and/or stability. Each of these additional factors may also be vitally important to consider for influencing the abundance of conditionally-specific functional variants.

---

**Structural differences between bacterial and eukaryotic cells influencing gene regulation**

The defining hallmark of the eukaryotic cell is the nucleus, a double membrane that encloses the cell's hereditary material. In order to efficiently fit the organism's DNA into the confined space of the nucleus, the DNA is first packaged and organized by protein into a structure called chromatin. This packaging of the nuclear material reduces access to specific parts of the chromatin. Indeed, some elements of the DNA are so tightly packed that the transcriptional machinery cannot access regulatory sites like promoters. This means that one of the first sites of transcriptional regulation in eukaryotes must be the control access to the DNA itself. Chromatin proteins can be subject to enzymatic modification that can influence whether they bind tightly (limited transcriptional access) or more loosely (greater transcriptional access) to a segment of DNA. This process of modification - whichever direction is considered first - is reversible. Therefore DNA can be dynamically sequestered and made available when the "time is right".

The regulation of gene expression in eukaryotes also involves some of the same additional fundamental mechanisms discussed in the module on bacterial regulation (i.e. the use of strong or weak promoters, transcription factors, terminators etc.) but the actual number of proteins involved is typically much greater in eukaryotes than bacteria or archaea.

The post-transcriptional enzymatic processing of RNA that occurs in the nucleus and the export of the mature mRNA to the cytosol are two additional difference between bacterial and eukaryotic gene regulation. We will consider this level of regulation in more detail below.

---

*Depiction of some key differences between the processes of bacterial and eukaryotic gene expression. Note in this case the presence of histone and histone modifiers, the splicing of pre-mRNA, and the export of the mature RNA from the nucleus as key differentiators between the bacterial and eukaryotic systems.*

*Attribution: Marc T. Facciotti (own work)*

https://bio.libretexts.org/Courses/University_of_California_Davis/BIS_2A%3A_Introductory_Biology_(Easlon)/Readings/23.2%…

Updated: Sun, 24 May 2020 00:30:42 GMT

Powered by
DNA Packing and Epigenetic Markers

The DNA in eukaryotic cells is precisely wound, folded, and compacted into chromosomes so that it will fit into the nucleus. It is also organized so that specific segments of the chromosomes can be easily accessed as needed by the cell. Areas of the chromosomes that are more tightly compacted will be harder for proteins to bind and therefore lead to reduced gene expression of genes encoded in those areas. Regions of the genome that are loosely compacted will be easier for proteins to access, thus increasing the likelihood that the gene will be transcribed. Discussed here are the ways in which cells regulate the density of DNA compaction.

DNA packing

The first level of organization, or packing, is the winding of DNA strands around histone proteins. Histones package and order DNA into structural units called nucleosomes, which can control the access of proteins to specific DNA regions. Under the electron microscope, this winding of DNA around histone proteins to form nucleosomes looks like small beads on a string. These beads (nucleosome complexes) can move along the string (DNA) to alter which areas of the DNA are accessible to transcriptional machinery. While nucleosomes can move to open the chromosome structure to expose a segment of DNA, they do so in a very controlled manner.

DNA is folded around histone proteins to create (a) nucleosome complexes. These nucleosomes control the access of proteins to the underlying DNA. When viewed through an electron microscope (b), the nucleosomes look like beads on a string. (credit “micrograph”: modification of work by Chris Woodcock)

Histone Modification

How the histone proteins move is dependent on chemical signals found on both the histone proteins and on the DNA. These chemical signals are chemical tags added to histone proteins and the DNA that tell the histones if a chromosomal region should be "open" or "closed". The figure below depicts modifications to histone proteins and DNA. These tags are not permanent, but may be added or removed as needed. They are chemical modifications (phosphate, methyl, or acetyl groups) that are attached to specific amino acids in the histone proteins or to the nucleotides of the DNA. The tags do not alter the DNA base sequence, but they do alter how tightly wound the DNA is around the histone proteins. DNA is a negatively charged molecule; therefore, changes in the charge of the histone will change how tightly wound the DNA molecule will be. When unmodified, the histone proteins have a large positive charge; by adding chemical modifications like acetyl groups, the charge becomes less positive.
Nucleosomes can slide along DNA. When nucleosomes are spaced closely together (top), transcription factors cannot bind and gene expression is turned off. When the nucleosomes are spaced far apart (bottom), the DNA is exposed. Transcription factors can bind, allowing gene expression to occur. Modifications to the histones and DNA affect nucleosome spacing.

Suggested discussion

Why do histone proteins normally have a large amount of positive charges (histones contain a high number of lysine amino acids). Would removal of the positive charges cause a tightening of loosening of the histone-DNA interaction?

Suggested discussion

Predict the state of the histones in areas of the genome that are transcribed regularly. How do these differ from areas that do not experience high levels of transcription?

DNA Modification

The DNA molecule itself can also be modified. This occurs within very specific regions called CpG islands. These are stretches with a high frequency of cytosine and guanine dinucleotide DNA pairs (CG) often found in the promoter regions of genes. When this configuration exists, the cytosine member of the pair can be methylated (a methyl group is added). This modification changes how the DNA interacts with proteins, including the histone proteins that control access to the region. Highly methylated (hypermethylated) DNA regions with deacetylated histones are tightly coiled and transcriptionally inactive.
Epigenetic changes do not result in permanent changes in the DNA sequence. Epigenetic changes alter the chromatin structure (protein-DNA complex) to allow or deny access to transcribe genes. DNA modification such as methylation on cytosine nucleotides can either recruit repressor proteins that block RNA polymerase's access to transcribe a gene or they can aid in compacting the DNA to block all protein access to that area of the genome. These changes are reversible whereas mutations are not, however, epigenetic changes to the chromosome can also be inherited.

Source: modified from https://researcherblogski.wordpress...r/dudiwarsito/

Regulation of gene expression through chromatin remodeling is called epigenetic regulation. Epigenetic means “around genetics.” The changes that occur to the histone proteins and DNA do not alter the nucleotide sequence and are not permanent. Instead, these changes are temporary (although they often persist through multiple rounds of cell division and can be inherited) and alter the chromosomal structure (open or closed) as needed.

External link

View this video that describes how epigenetic regulation controls gene expression.

Eukaryotic gene structure and RNA processing

Eukaryotic gene structure

Many eukaryotic genes, particularly those encoding protein products, are encoded on the genome discontinuously. That is, the coding region is broken into pieces by intervening non-coding gene elements. The coding regions are termed exons while the intervening non-coding elements are termed introns. The figure below depicts a generic eukaryotic gene.

The parts of a typical discontinuous eukaryotic gene. Attribution: Marc T. Facciotti (own work)

Parts of a generic eukaryotic gene include familiar elements like a promoter and terminator. Between those two
elements, the region encoding all of the elements of the gene that have the potential to be translated (they have no stop
codons), like in bacterial systems, is called the open reading frame (ORF). Enhancer and/or silencer elements are
regions of the DNA that serve to recruit regulatory proteins. These can be relatively close to the promoter, like in
bacterial systems, or thousands of nucleotides away. Also present in many bacterial transcripts, 5' and 3' untranslated
regions (UTRs) also exist. These regions of the gene encode segments of the transcript, which, as their names imply,
are not translated and sit 5' and 3', respectively, to the ORF. The UTRs typically encode some regulatory elements
critical for regulating transcription or steps of gene expression that occur post-transcriptionally.

The RNA species resulting from the transcription of these genes are also discontinuous and must therefore be
processed before exiting the nucleus to be translated or used in the cytosol as mature RNAs. In eukaryotic systems this
includes RNA splicing, 5' capping, 3' end cleavage and polyadenylation. This series of steps is a complex molecular
process that must occur within the closed confines of the nucleus. Each one of these steps provides an opportunity for
regulating the abundance of exported transcripts and the functional forms that these transcripts will take. While these
would be topics for more advanced courses, think about how to frame some of the following topics as subproblems of
the Design Challenge of genetic regulation. If nothing else, begin to appreciate the highly orchestrated molecular dance
that must occur to express a gene and how this is a stunning bit of evolutionary engineering.

### 5' capping

Like in bacterial systems, eukaryotic systems must assemble a pre-initiation complex at and around the promoter
sequence to initiate transcription. The complexes that assemble in eukaryotes serve many of the same function as those
in bacterial systems but they are significantly more complex, involving many more regulatory proteins. This added
complexity allows for a greater degree of regulation and for the assembly of proteins with functions that occur
predominantly in eukaryotic systems. One of these additional functions is the "capping" of nascent transcripts.

In eukaryotic protein coding genes, the RNA that is first produced is called the pre-mRNA. The "pre" prefix signifies that
this is not the full mature mRNA that will be translated and that it first requires some processing. The modification known
as 5'-capping occurs after the pre-mRNA is about 20-30 nucleotides in length. At this point the pre-RNA typically
receives its first post-transcriptional modification, a 5'-cap. The "cap" is a chemical modification - a 7-methylguanosine -
whose addition to the 5' end of the transcript is enzymatically catalyzed by multiple enzymes called the capping enzyme
complex (CEC) a group of multiple enzymes that carry out sequential steps involved in adding the 5'-cap. The CEC
binds to the RNA polymerase very early in transcription and carries out a modification of the 5’ triphosphate, the
subsequent transfer of at GTP to this end (connecting the two nucleotides using a unique 5’-to-5’ linkage), the
methylation of the newly transferred guanine, and in some transcripts the additional modifications to the first few
nucleotides. This 5'-cap appears to function by protecting the emerging transcript from degradation and is quickly bound
by RNA binding proteins known as the cap-binding complex (CBC). There is some evidence that this modification and
the proteins bound to it play a role in targeting the transcript for export from the nucleus. Protecting the nascent RNA
from degradation is not only important for conserving the energy invested in creating the transcript but is clearly involved
in regulating the abundance of fully-functional transcript that is produced. Moreover, the role of the 5'-cap in guiding the
transcript for export will directly help to regulate not only the amount of transcript that is made but, perhaps more
importantly, the amount of transcript that is exported to the cytoplasm that has the potential to be translated.
Transcript splicing

Nascent transcripts must be processed into mature RNAs by joining exons and removing the intervening introns. This is accomplished by a multicomponent complex of RNA and proteins called the spliceosome. The spliceosome complex assembles on the nascent transcript and in many cases the decisions about which introns to combine into a mature transcript are made at this point. How these decisions are made is still not completely understood but involves the recognition of specific DNA sequences at the splice sites by RNA and protein species and several catalytic events. It is interesting to note that the catalytic portion of the spliceosome is made of RNA rather than protein. Recall that the ribosome is another example of a RNA-protein complex where the RNA serves as the primary catalytic component. The selection of which splice variant to make is a form of regulating gene expression. In this case rather than simply influencing abundance of a transcript, alternative splicing allows the cell to make decisions about which form of transcript is made.

The alternative splice forms of genes that result in protein products of related structure but of varying function are known as isoforms. The creation of isoforms is common in eukaryotic systems and is known to be important in different stages of development in multicellular organisms and in defining the functions of different cell types. By encoding multiple possible gene products from a single gene whose transcription initiation is encoded from a single transcriptional regulatory site (by making the decision of which end-product to produce post-transcriptionally) obviates the need to create and maintain independent copies of each gene in different parts of the genome and evolving independent regulatory sites. Therefore, the ability to form multiple isoforms from a single coding region is though to be evolutionarily advantageous because it enables some efficiency in DNA coding, minimizes transcriptional regulatory complexity, and may lower the energy burden of maintaining more DNA and protecting it from mutation. Some examples of possible outcomes of alternative splicing can include: the generation of enzyme variants with differential substrate affinity or catalytic rates; signal sequences that target proteins to various sub-cellular compartments can be changed; entirely new functions, via the swapping of protein domains can be created. These are just a few examples.

One additional interesting possible outcome of alternative splicing is the introduction of stop codons that can, through a mechanism that seems to require translation, lead to the targeted decay of the transcript. This means that, in addition to the control of transcription initiation and 5’-capping, alternative splicing can also be considered one of the regulatory mechanisms that may influence transcript abundance. The effects of alternative splicing are therefore potentially broad -
from complete loss of function to novel and diversified function to regulatory effects.

A figure depicting some of the different modes of alternative splicing illustrating how different splice variants can lead to different protein forms.

Attribution: Marc T. Facciotti (own work)

3' end cleavage and polyadenylation

One final modification is made to nascent pre-mRNAs before they leave the nucleus - the cleavage of the 3' end and its polyadenylation. This two step process is catalyzed by two different enzymes (as depicted below) and may decorate the 3' end of transcripts with up to nearly 200 nucleotides. This modification enhances the stability of the transcript. Generally, the more As in the polyA tag the longer lifetime that transcript has. The polyA tag also seems to play a role in the export of the transcript from the nucleus. Therefore, the 3' polyA tag plays a role in gene expression by regulating functional transcript abundance and how much is exported from the nucleus for translation.

A two step process is involved in modifying the 3' ends of transcripts prior to nuclear exports. These include cutting transcripts just downstream of a conserved sequence (AAUAAA) and transferring adenylate groups. Both processes are enzymatically catalyzed.

Attribution: Marc T. Facciotti (own work)

microRNAs

RNA Stability and microRNAs

In addition to the modifications of the pre-RNA described above and the associated proteins that bind to the nascent and
transcripts, there are other factors that can influence the stability of the RNA in the cell. One example are elements called microRNAs. The microRNAs, or miRNAs, are short RNA molecules that are only 21–24 nucleotides in length. The miRNAs are transcribed in the nucleus as longer pre-miRNAs. These pre-miRNAs are subsequently chopped into mature miRNAs by a protein called dicer. These mature miRNAs recognize a specific sequence of a target RNA through complementary base pairing. miRNAs, however, also associate with a ribonucleoprotein complex called the RNA-induced silencing complex (RISC). RISC binds a target mRNA, along with the miRNA, to degrade the target mRNA. Together, miRNAs and the RISC complex rapidly destroy the RNA molecule. As one might expect, the transcription of pre-miRNAs and their subsequent processing is also tightly regulated.

**Nuclear export**

**Nuclear export**

Fully processed, mature transcripts, must be exported through the nucleus. Not surprisingly this process involves the coordination of a mature RNA species to which are bound many accessory proteins - some of which have been intimately involved in the modifications discussed above - and a protein complex called the nuclear pore complex (NPC). Transport through the NPC allows flow of proteins and RNA species to move in both directions and is mediated by a number of proteins. This process can be used to selectively regulate the transport of various transcripts depending on which proteins associate with the transcript in question. This means that not all transcripts are treated equally by the NPC - depending on modification state and the proteins that have associated with a specific species of RNA it can be moved either more or less efficiently across the nuclear membrane. Since the rate of movement across the pore will influence the abundance of mature transcript that is exported into the cytosol for translation export control is another example of a step in the process of gene regulation that can be modulated. In addition, recent research has implicated interactions between the NPC and transcription factors in the regulation of transcription initiation, likely through some mechanism whereby the transcription factors tether themselves to the nuclear pores. This last example demonstrates how interconnected the regulation of gene expression is across the multiple steps of this complex process.

Many additional details of the processes described above are known to some level of detail, but many more questions remain to be answered. For the sake of Bis2a it is sufficient to begin forming a model of the steps that occur in the production of a mature transcript in eukaryotic organisms. We have painted a picture with very broad strokes, trying to present a scene that reflect what happens generally in all eukaryotes. In addition to learning the key differentiating features of eukaryotic gene regulation, we would also like for Bis2a students to begin thinking of each of these steps as an opportunity for Nature to regulate gene expression in some way and to be able to rationalize how deficiencies or changes in these pathways - potentially introduced through mutation - might influence gene expression.

While we did not explicitly bring up the Design Challenge or Energy Story here these formalisms are equally adept at helping you to make some sense of what is being described. We encourage you to try making an Energy Story for various processes. We also encourage you to use the Design Challenge rubric to reexamine the stories above: identify problems that need solving; hypothesize potential solutions and criteria for success. Use there formalisms to dig deeper and ask new questions/identify new problems or things that you don't know about the processes is what experts do. Chances are that doing this suggested exercise will lead you to identify a direction of research that someone has already pursued (you'll feel pretty smart about that!). Alternatively, you may raise some brand new question that no one
Control of Protein Abundance

After an mRNA has been transported to the cytoplasm, it is translated into protein. Control of this process is largely dependent on the RNA molecule. As previously discussed, the stability of the RNA will have a large impact on its translation into a protein. As the stability changes, the amount of time that it is available for translation also changes.

The initiation complex and translation rate

Like transcription, translation is controlled by proteins complexes of proteins and nucleic acids that must associate to initiate the process. In translation, one of the first complexes that must assemble to start the process is referred to as the initiation complex. The first protein to bind to the mRNA that helps initiate translation is called eukaryotic initiation factor-2 (eIF-2). Activity of the eIF-2 protein is controlled by multiple factors. The first is whether or not it is bound to a molecule of GTP. When the eIF-2 is bound to GTP it is considered to be in an active form. The eIF-2 protein bound to GTP can bind to the small 40S ribosomal subunit. When bound, the eIF-2/40S ribosome complex, bringing with it the mRNA to be translated, also recruits the methionine initiator tRNA associates. At this point, when the initiator complex is assembled, the GTP is hydrolyzed into GDP creating an "inactive form of eIF-2 that is released, along with the inorganic phosphate, from the complex. This step, in turn, allows the large 60S ribosomal subunit to bind and to begin translating the RNA. The binding of eIF-2 to the RNA further controlled by protein phosphorylation. When eIF-2 is phosphorylated, it undergoes a conformational change and cannot bind to GTP thus inhibiting the initiation complex from forming - translation is therefore inhibited (see the figure below). In the dephosphorylated state eIF-2 can bind GTP and allow the assembly of the translation initiation complex as described above. The ability of the cell therefore to tune the assembly of the translation invitation complex via a reversible chemical modification (phosphorylation) to a regulatory protein is another example of how Nature has taken advantage of even this seemingly simple step to tuned gene expression.

An increase in phosphorylation levels of eIF-2 has been observed in patients with neurodegenerative diseases such as Alzheimer’s, Parkinson’s, and Huntington’s. What impact do you think this might have on protein synthesis?

Chemical Modifications, Protein Activity, and Longevity

Not to be outdone by nucleic acids, proteins can also be chemically modified with the addition of groups including methyl, phosphate, acetyl, and ubiquitin groups. The addition or removal of these groups from proteins can regulate their
activity or the length of time they exist in the cell. Sometimes these modifications can regulate where a protein is found in the cell—for example, in the nucleus, the cytoplasm, or attached to the plasma membrane.

Chemical modifications can occur in response to external stimuli such as stress, the lack of nutrients, heat, or ultraviolet light exposure. In addition to regulating the function of the proteins themselves, if these changes occur on specific proteins they can alter epigenetic accessibility (in the case of histone modification), transcription (transcription factors), mRNA stability (RNA binding proteins), or translation (eIF-2) thus feeding back and regulating various parts of the process of gene expression. In the case of modification to regulatory proteins, this can be an efficient way for the cell to rapidly change the levels of specific proteins in response to the environment by regulating various steps in the process.

The addition of an ubiquitin group has another function - it marks that protein for degradation. Ubiquitin is a small molecule that acts like a flag indicating that the tagged proteins should be targeted to an organelle called the proteasome. This organelle is a large multi-protein complex that functions to cleave proteins into smaller pieces that can then be recycled. Ubiquitination (the addition of a ubiquitin tag), therefore helps to control gene expression by altering the functional lifetime of the protein product.

In conclusion, we see that gene regulation is complex and that it can be modulated at each step in the process of expressing a functional gene product. Moreover, the regulatory elements that happen at each step can act to influence other regulatory steps both earlier and later in the process of gene expression (i.e. the process of chemically altering a transcription factor can influence the regulation of its own transcription many steps earlier in the process). These complex sets of interactions form what are known as gene regulatory networks. Understanding the structure and dynamics of these networks is critical for understanding how different cells function, the basis for numerous diseases, developmental processes, and how cells make decisions about how to react to the many factors that are in constant flux both inside and outside.