iBiology.org Teaching Tools
Robert Tjian's Lecture Part 1:
Gene Regulation: An Introduction

Teaching Tools were prepared by Lea Witkowsky, Elisa Zhang, Nikki Kong, Chiahao Tsui with Robert Tjian.

Contents

---

1. Keywords and Terms

Transcription, gene expression, coding region, regulatory region, repressors, activators, Sp1, preinitiation complex.

2. Lecture Notes

Flow of biological information: DNA -> RNA -> protein. This is referred to as the "central dogma".

- **Transcription refers to the process of making RNA from a DNA template.**

Transcription factors: key molecular determinants that regulate the use of genetic information.

- **Transcription factors regulate gene expression by controlling transcription.**

The human genome has roughly 3 billion base pairs and roughly 22,000 genes.

- **Only ~3% of the DNA contains protein coding sequences.**
- **The majority of the DNA is used for other purposes such as regulation of the expression of protein coding regions.**
- **Transcription factors recognize these regulatory regions of DNA and dictate which genes are transcribed.**
- **Regulatory DNA sequences: small fragments of DNA that direct transcription of a gene.**

The transcription of protein coding DNA sequences to messenger RNA (mRNA) is carried out by RNA polymerase II.

- **RNA pol II needs "help"; it cannot differentiate between coding and non-coding DNA regions.**
- **Other factors help RNA pol II get to the correct place at the correct time to initiate transcription: these are the transcription factors.**

Two different types of techniques can help us identify factors that help RNA polymerase II to initiate transcription:

- **In vitro biochemistry**
- **Isolate individual parts of the cell and see how they function alone or together in a test tube.**
- **In vivo genetics (model organisms)**

- **Disrupt one component (protein, DNA sequence, etc) of the cell and see what the consequences are to the cell or the organism.**

A functional gene has various DNA regulatory sequences.

- **The upstream regulatory sequences are recognized and bound by their corresponding transcription factors.**
- **For example; a TATA box is often found upstream of a gene at its promoter and is recognized by the TATA binding protein (TBP).**
- **Transcription factors are proteins that fold into a particular shape that can often bind/recognize specific DNA regulatory sequences.**
- **Binding to specific DNA sequences upstream of a particular gene allows the transcription factor to recruit and instruct RNA polymerase II to initiate RNA synthesis at that particular gene.**

RNA polymerase II is not the only protein that is necessary for transcription.

- **Other proteins such as TFIIA, B, D, E, F, H are necessary to stabilize and interact with RNA polymerase II at most genes (in addition to gene specific transcription factors).**

Discovery of the first eukaryotic (human) transcription factor:

- **Scientists used the virus SV40 to approach this problem.**
- **SV40 was the first virus to have its genome fully sequenced.**
- **Researchers knew that the virus does not have the required transcriptional machinery to express its own genes; they knew that there are host transcription factor(s) that are needed.**
- **Despite its inability to transcribe its own genome without the host, researchers discovered Tumor Antigen, a viral protein that can regulate transcription. T-antigen is an SV40 gene that encodes for a transcription factor that, instead of activating transcription, represses transcription.**
- **To determine what sequence elements and what host factors were required to transcribe SV40 genes, researchers looked to the DNA sequence.**
- **Examination of the viral DNA showed a region containing GC-rich repeats which were found to have a regulatory function.**
- **Deletion of the GC-rich sequences ablated expression of a gene of interest.**

- **It was hypothesized that a DNA specific binding protein (transcription factor) must bind to this region to activate it.**
- **Because none of the viral proteins could fulfill this role, they knew they were looking for a host factor.**
- **Isolation of this specific protein was ultimately achieved using DNA affinity purification.**
- **DNA sequence-specific nuclear proteins can be captured with a solid surface onto which target sequences are attached. The hypothesized protein should bind to these sequences specifically, allowing its separation from the other proteins in the nuclear extract. In this case, the GC-rich sequences were attached to the solid surface of a column.**
- **Non-specific (non GC-rich) DNA were added to the solution along with the cellular extract.**
- **The addition of non-specific DNA that is NOT attached to the surface allowed the researchers to minimize binding of non GC-specific proteins to the solid surface. These non-target proteins could be washed away, leaving behind only those proteins which specifically recognize the sequence of interest.**
- **This DNA affinity purification led to the identification of the GC-rich sequence specific transcription factor Sp1.**
- **Sp1 is able to bind to this sequence regardless of its origin (virus or human).**

Most transcription factors recognize DNA by binding to the major groove of the DNA.

- **Sp1 binds to the DNA's major groove via "zinc fingers".**
- **Zinc fingers refer to a type of 3-dimensional organization of amino acids in a protein that coordinate a zinc atom.**
- **The discovery of a zinc finger motif also led to the realization that most (if not all) transcription factors contain regions that fold into specific structural motifs to recognize and bind DNA, usually in the major groove.**
- **This type of motif is recognizable at the amino acid sequence level and allows scientists to identify genes in the genome that are likely to be transcription factors.**

Initiation of Transcription, Formation of the PIC;

- **In addition to transcription factors like Sp1, a preinitiation complex (PIC) is also needed for transcription.**
- **TATA binding protein (TBP) binds to the TATA box as part of the formation of the PIC.**
- **TBP rarely functions on its own; TBP-associated factors (TAFs) form a multi-subunit complex with TBP to form TFIID which sits on the TATA box.**
- **TFIIA, TFIIB, TFIIE, TFIIF, TFIIH, and RNA polymerase II bind with TFIID to form the transcription preinitiation complex (PIC)**
- **The complete formation of the PIC allows initiation of transcription to proceed.**

3. Recommended Reading

1. **Watson J.D., Baker T.A., Bell S.P., Gann A., Levine M., and Losick R. Molecular Biology of the Gene. Benjamin Cummings, San Francisco, CA. Chapter 12.**
2. **Lodish H., Berk A., Zipursky, S.L., Matsudaira P., Baltimore D., and Darnell J. Molecular Cell Biology. W. H. Freeman and Company, New York, NY. Chapter 10.**
3. **Nelson D.L. and Cox M.M. Lehninger Principles of Biochemistry. 2008. W. H. Freeman and Company, New York, NY. Chapter 26.2.**

4. Review Questions

1. **What is the Central Dogma?**
2. **What is a transcription factor? What is its role in the central dogma?**
3. **What is a coding region?**
4. **In addition to protein coding sequences, what is one function of the non-coding sequences in our genome?**
5. **What is the enzyme that catalyzes the transcription process? How many of them are there in humans? Which one makes mRNA?**

6. **What percent of the human genome contains protein coding information?**
7. **When scientists first began to study transcription in animal cells, why did they choose to use viruses? What advantages did this strategy give them?**
8. **What is a TATA box? What protein recognizes it, and what is its role in transcription?**
9. **Scientists have identified a new organism and have already sequenced its genome. Without having to obtain samples of this rare organism, how might you begin searching for potential sequence-specific transcription factors?**
10. **What is a zinc finger? What property does it confer to a protein?**

5. Answers to Review Questions

1. **The Central Dogma states that genetic information in the form of DNA is first transcribed into an intermediate molecule known as messenger RNA or mRNA, which is subsequently translated into the final protein product.**
2. **A transcription factor is a protein that binds DNA and either acts to activate or repress the process of making mRNA molecules from DNA.**
3. **A coding region is a portion of DNA that encodes the amino acid sequence of a protein.**
4. **One important function of our non-coding DNA is to regulate transcription of the protein coding sequences. Part of the non-coding DNA contains regulatory sequences, including promoters and enhancers, that direct the binding of the transcription apparatus. These regulatory elements are recognized by components of the preinitiation complex as well as sequence-specific activators or repressors.**
5. **RNA Polymerase is the enzyme that catalyzes transcription. There are three RNA polymerases; I and III are responsible for making rRNA and tRNA. The one that is responsible for making mRNA is RNA Pol II.**
6. **~3% of the human genome encodes proteins.**
7. **DNA sequence information was not widely available at the time, but the SV40 genome had been sequenced, and all the viral proteins had been identified. Having determined that viral proteins are not necessary for activation, these pieces of**

information allowed scientists to pin down which regions of the viral DNA and which host factors are necessary for transcription.

8. **The TATA box is a DNA sequence with a TATAAA consensus sequence. TATA box binding protein (TBP) recognizes and binds to this sequence. This is the first step in assembling the preinitiation complex.**
9. **Most, if not all, sequence-specific transcription factors recognize DNA using 3-dimensional structural motifs that are conserved among different families of transcription factors and divergent organisms. These structural motifs are encoded in the amino acid sequence and can therefore be identified at the DNA sequence level. If the genomic sequence is available, it is possible to search the sequence using computational algorithms for these known motifs. However, it is always important to still verify that the identified sequences produce a protein that has a transcriptional function.**
10. **A zinc finger is a structural motif, or domain, of a protein that is conserved. It is usually cysteine- and histidine-rich and coordinates a zinc ion. It confers DNA binding properties to a protein, usually a transcription factor.**

6. Discussion Questions

1. **You are interested in the transcriptional regulation of a particular gene and you have identified a sequence upstream of the gene that is required for its transcription. In order to identify what protein is binding to this sequence, you made a resin with DNA attached to it that contains multiple copies of the sequence that you have identified. You then added the extract you prepared from the cells' nucleus to this resin and collect the proteins that were bound. When you analyzed your results you found that you had many different proteins in your purification. Why was your purification not pure, what might you have forgotten to include in your purification?**
2. **Bacteriophage T7 uses an RNA polymerase that does not require any additional factors to transcribe DNA. In contrast, RNA polymerases from both Bacteria and Eukaryotes require additional transcription factors. Why might such a requirement exist for these organisms?**
3. **If a mutation in eukaryotic RNA polymerase were to occur which allows the polymerase to gain a sequence preference for binding, how do you think this would influence transcription? Would the cells likely be healthy or sick? Why?**

4. **How can studying transcription factors lead to improvements in human health?**

7. Answers to Discussion Questions

1. **The most likely explanation is that you forgot to include vast stoichiometric excess of competitor DNA that does not contain your regulatory sequence. Without this competitor, any DNA binding protein may stick to your column non-specifically. Since there are many DNA binding proteins other than your target protein, these will be the overwhelming majority of the proteins you obtain.**
2. **The T7 bacteriophage cannot differentially regulate its genes because it uses a single protein to recognize a single promoter sequence. However, bacteria and eukaryotes need to be able to turn certain genes on and leave others off at different times depending on their environment. Having additional factors allows bacteria and eukaryotes to use a common transcriptional machinery and confer differential regulation through these additional transcription factors.**
3. **It is likely that gaining a sequence preference would lead to severe defects in transcription. The same RNA polymerase must be able to initiate transcription from any promoter sequence in the genome. If it were to gain a sequence preference, then it may favor binding at locations that contain its recognition site. This would both disrupt its ability to slide along the DNA unencumbered during RNA polymerization and its ability to be directed by transcription factors to the correct genes at the correct time. In addition, binding to its recognition site may sequester the RNA polymerase away from the appropriate sites of transcription initiation, reducing the effective amount of RNA polymerase available. All of these effects would likely lead to very sick cells.**
4. **The expression of transcription factors is often mis-regulated in diseased states. Understanding how a particular transcription factor works can give us clues as to how to correct the effects of its misregulation. For example, if the factor is expressed more highly in a sick cell than in a normal cell, then we could find molecules that inhibit its expression or function to restore its level or activity to normalcy.**

8. Explain or Teach These Concepts to a Friend

1. **Describe the two major experimental approaches used to discover transcription factors.**
2. **Describe the pre-initiation complex. What is its function, important components, and relationships to RNA Pol II and sequence-specific activators?**
3. **Describe the technique by which Sp1 was purified.**

9. Research the Literature on Your Own

1. **How was the binding site for Sp1 identified?**
2. **How did they know that the GC-boxes were important for viral transcription?**
3. **As mentioned in the lecture, Sp1 recognizes the major groove of DNA, as do most transcription factors. Find an example of a transcription factor that recognizes the minor groove of DNA.**
4. **In addition to zinc fingers, what are two other classes of sequence-specific transcription factors based on the characteristics of their DNA-binding domains?**