**Lecture 4: DNA and Protein Synthesis**

So, we have seen that all living things are complex, and to create that complexity they need energy. We have seen how life acquires energy from the environment, and convert that energy into a form that is usable by all enzymes in the cell. Now remember, enzymes are proteins that catalyze reactions in the cell. In other words, enzymes are proteins that make stuff all this complex stuff with the energy in ATP. Enzymes make phospholipids, enzymes make sugars, enzymes make DNA, and enzymes make proteins. So, a fundamental question of how life works is this: how do cells make their enzymes? Indeed, a more general question is: how do cells make all their proteins – some of which function as enzymes but others that are structural (like the muscle proteins in muscle cells that contract) or involve in transport (membrane proteins). That is what we will look at in this lecture.

Basically, DNA is a recipe for proteins. By making these proteins, a cell can make anything else it needs from what it absorbs from the environment. So, it is really that simple: DNA is a recipe for proteins. DNA is in chromosomes, and when a cell divides and passes copies of chromosomes to each daughter cell, each daughter cell receives the full recipe book for making their own proteins. But it is also very complex, as we will see. For example, although each cell in your body has the whole recipe book, those in your muscle only make muscle proteins and those cells in your eye make eye proteins. The REGULATION of reading the DNA recipe is how cells become specialized, and is one of the most exciting areas of genetic research.

So, understanding how proteins are made is fundamental to understanding biological systems. But to understand how proteins are made, one must first understand the structure of DNA.

**I. DNA, RNA, and Chromosome Structure**

DNA

**A. DNA and RNA Structure**

DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are *nucleic acids* - polymers consisting of a linear sequence of linked nucleotide monomers. We will describe the structure of the monomers first, and then describe how they are linked into linear polymers. Finally, we will describe the double-stranded structure of ds-DNA.

**1. The monomers are "nucleotides"**

three components:

**- Pentose (5 carbon) sugar:** either ribose (RNA) or deoxyribose (DNA). The carbons are numbered clockwise. The difference between the sugars is that ribose has an -OH group on the 2' carbon, whereas deoxyriboes has only 2 H groups and thus is "deoxygenated" relative to ribose. BOTH sugars have an -OH group on the 3' carbon, which will be involved in binding. The 5' carbon is a sidegroup off the ring.

**- Nitrogenous Base:** each nucleotide has a single nitrogenous base attached to the 1' carbon of the sugar. This nitrogenous base may be a double-ringed structure (purine) or a single ringed (pyrimidine) structure. The purines are adenine (A) and guanine (G). The pyrimidines are thymine (T), cytosine (C), and uracil (U). DNA nucleotides may carry A, G, C, or T. RNA nucleotides carry either A, G, C, or U.

**- The third component of a nucleotide is a phosphate group**, which is attached to the 5' carbon of the sugar. When a nucleotide is incorporated into a chain, it has a single phosphate group. However, nucleotides can occur that have two or three phosphate groups (dinucleotides and trinucleotides). ADP and ATP are important examples of these types of molecules. In fact, the precursors of incorporated nucleotides are trinucleotides. When two phosphates are cleaved, energy is released that can be used to add the remaining monophosphate nucleotide to the nucleic acid chain.

**2. Polymerization is by 'dehydration synthesis'**

As with all other classes of biologically important polymers, monomers are linked into polymers by dehydration synthesis. In nucleic acid formation, this involves binding the phosphate group of one nucleotide to the -OH group on the 3' carbon of the existing chain. For the purposes of seeing how this reaction works, we can envision an H+ on one of the negatively charged oxygens of the phosphate group. Then, a molceule of water can be removed from these two -OH groups, leaving an oxygen binding the sugar of one nucleotide to the phosphate of the next.

This creates a 'dinucleotide'. It has a polarity/directionality; it is different at its ends. At one end, the reactive group is the phosophate on the 5' carbon. This is called the 5' end of the chain. At the other end, the reactive group is the free -OH on the 3' carbon; this is the 3' end of the chain. So, a nucleic acid strand has a 5' - 3' polarity.

**3. Most DNA exists as a 'double helix' (ds-DNA) containing two linear nucleic acid chains.**

**a. the nitrogenous bases on the two strands are 'complementary' to each other**, and form weak hydrogen bonds between them. A always pairs with T, and C always pairs with G. As such, there is always a double-ringed purine pairing with a single-ringed pyrimidine, and the width of the double-helix is constant over its entire length.

**b. the two strands (helices) are anti-parallel:** they are arranged with opposite polarity. One strands points 5' - 3', while the other points 3' - 5'. The direction of the pentose sugars and the type of reactive group at the ends of the chains show this relationship.

**4. RNA performs a wide variety of functions in living cells:**

**a. m-RNA (for "messenger")** is the copy of a gene. It is the sequence of nitrogenous bases in m-RNA that is actually read by the ribosome to determine the structure of a protein.

**b. r-RNA (for "ribosomal")** is made the same way, as a copy of DNA. However, it is not carrying the recipe for a protein; rather, it is functional as RNA. It is placed IN the Ribosome, and it helps to ‘read’ the m-RNA.

**c. t-RNA (for "transfer")** is also made as a copy of DNA, but it is also functional as an RNA molecule. Its function is to bind to a specific amino acid and incorporate it into the amino acid sequence as instructed by the m-RNA and ribosome.

**d. mi-RNA (micro-RNA) and si-RNA (small interfering RNA)** bind to m-RNA and splice it; inhibiting the synthesis of its protein. This is a regulatory function.

**e. sn-RNA (small nuclear RNA)** are short sequences that process initial m-RNA products, and also regulate the production of r-RNA, maintain telomeres, and regulate the action of transcription factors. Regulatory functions.

**B. Chromosome Structure**

**1. Prokaryotes**
- usually one circular chromosome, tethered to the membrane, with some associated, non-histone proteins.

**2. Eukaryotes**

– usually many linear chromosomes, highly condensed with histone proteins into several levels of structure.

**Level 1:** ds-DNA is wrapped around histone proteins, creating the “beads on a string’ level of organization.
**Level 2:** string is coiled, 6 nucleosomes/turn (solenoid)
**Level 3**: the coil is ‘supercoiled’
**Level 4:** the supercoil is folded into a fully condensed metaphase chromosome

To read a gene, the chromosome must be diffuse (uncondensed) in that region. Even when condensed, these ‘euchromatic’ coding regions are less condensed and more lightly staining than non-coding regions.

DNA that has few genes can remain condensed and closed (heterochromatic), and appears as dark bands on condensed chromosomes.

**II. Protein Synthesis**

As we've already mentioned, protein synthesis is fundamental to nearly everything a cell does. Protein channels are used to transport large molecules across the membrane. Almost all chemical reactions occuring in cells are catalyzed by protenaceous enzymes, including those involved in energy harvest, DNA replication, and cell division. Proteins perform important structural functions within cells and multicellular organisms, too; such as the histone proteins in chromosomes, the proteins in ribosomes, the collagen and elastin fibers that hold skin cells together, the collagen on which calcium and phosphate is deposited in bone, the protein myofibrils of actin and myosin in muscle cells, the neurotransmitters used for cell-cell communication between neurons, and the enzymes that digest food in the stomach and intestine of animals. So, proteins are fundamental to what cells and organisms ARE, structurally, and what they DO functionally. As you know, the genetic information determines the types of proteins a cell can make. The subset of proteins a cell actually DOES make, and the timing of WHEN they are made, is determined by what genes are "on" and what genes are "off" at a given time. This regulation of gene activity is ALSO co-ordinated by proteins - called transcription factors - that bind to DNA and promote or inhibit gene activity. So, proteins also regulate protein synthesis. Hopefully you see just how important proteins are to cells and organisms. So, ****the process of making these proteins is important, too.

**A. Overview**

The sequence of nitrogenous bases in a region of DNA is 'read' by a complex of enzymes that build a complementary strand of RNA. This process of reading DNA and making RNA is called 'transcription'. This is a great word for the process, as the message written in the language of nucleic acids is copied in essentially the same language - the language of nucleic acids. This RNA may be a recipe for a protein (m-RNA), or it may be an RNA that will act on its own as t-RNA, mi-RNA, si-RNA, or be complexed with proteins in the ribosome (r-RNA). Obviously, in "protein synthesis", only the m-RNA is read to make a protein. However, the other molecules all play a role. The sequence of nitrogenous bases in the m-RNA is then 'read' by a ribosome, which links a specific sequence of amino acids together into a protein based on that sequence of nitrogenous bases in the m-RNA. This process is called 'translation'. This is a great choice of a word, too. Here the sequence of information written in the language of nucleic acids is rewritten in a new language (hence, translation) - the language of amino acids.

Many of the initial RNA products have specific regions (introns) cut out of their sequence before they become functional. This step is known as "RNA processing" or "RNA splicing". Introns are present in nearly all eukaryotic RNA's, and are also in the DNA genes that encode them. Up until a few years ago, the only introns in prokaryotes had been found in t-RNA molecules of archaeans. More recently, however, introns have been found in m-RNA and r-RNA molecules of a few eubacteria and a few more archaeans. So, although they are rare in prokaryotes, we will describe a generic, simplified process of protein synthesis that includes introns and RNA processing.

In addition to splicing the RNA product of transcription, the initial protein product of translation may also be spliced and modified before it becomes functional. In eukaryotes, this protein processing often occurs in the Golgi apparatus.

The description presented here is a simple model of protein synthesis. You will learn more complex aspects of this process in Genetics.

**B. The Process of Protein Synthesis**

**1. Transcription**:

**a. The message is on one strand of the double helix - the sense strand:** The DNA double helix is composed of two anti-parallel complementary strands of DNA. Only one strand in a coding region ("gene") is read; this strand carries a meaningful recipe that "makes sense". This is called the "sense" strand. The other strand, limited by complementarity, is not a meaningful message - it is the "non-sense" (or "anti-sense") strand. Think about it this way. Given a meaningful message of "C-A-T" (a small furry mammal), the complementary strand is limited to the meaningless sequence of "G-T-A" (????...). As the meaningful sequence gets longer, it is even LESS likely that, just by chance, the complementary strand would be meaningful, too. Again, in all eukaryotic genes and in some rare prokaryotic ones, there will be non-functional "introns" interspersed throughout the meaningful message. The meaningful parts are called "exons". The process of transcription is continuous, so introns and exons get transcribed and these regions - if present in the DNA - will also be present in the RNA product.

**b. The cell 'reads' the correct strand based on the location of the promoter, the anti-parallel nature of the double helix, and the chemical limitations of the 'reading' enzyme, RNA Polymerase.** RNA Polymerase binds to the DNA at a specific sequence next to the gene, called the 'promoter'. It binds in a specific way, so it is pointed towards the gene. RNA polymerase can only create a strand of RNA in the 5' to 3' direction, adding a new base to the free -OH group of the preceeding nucleotide on the chain. So, from its position at the promoter, looking down the two strands in the gene, the RNA Polymerase can only 'read' one strand - the DNA strand that is 3'-5'. It must create a strand that is anti-parallel to the DNA' template', and it can only bind nucleotides in 5'-3' direction. So, only the 3'-5' DNA strand is read in this region, and only one RNA strand, 5'-3' is made. It is important to appreciate that this relationship is 'local'. In another region of the DNA, the promoter may be on the other side of the gene, and the other strand may be read.

**c. Transcription ends at a sequence called the 'terminator'.** These regions have specific sequences that destabilize the attachment of the RNA Polymerase to the DNA... it detaches and transcription stops. [VIDEO](http://www.youtube.com/watch?v=ztPkv7wc3yU)
So, the process of transcription can be summarized like this: RNA Polymerase binds at the promoter and reads the sense strand of DNA. The ploymerase links together RNA nucleotides 5--> 3, in a sequence complementary to the DNA sense strand. This process is continuous, so all DNA bases are 'read', including exon and intron sequnces. This process continues until a terminator region is reached. Reading this region destabilizes the RNA polymerase. It detaches from the DNA, and transcription stops. All types of RNA (m-RNA, r-RNA, t-RNA) are made through this process.

**2. Transcript Processing:**

At this point in the process, the cell has read the gene and synthesized a complementary copy of strand of RNA. In all eukaryote sequences and many prokaryotic ones, this RNA molecule will have non-functional introns that need to be 'cut-out'. Enzymes cut the introns out and splice the ends together. In some cases, the introns catalyze their own excision - they are RNA molecules with enzymatic activity. These are one class of "ribozymes" - a very interesting class of molecules. There are other ribozymes that cleave other RNA molecules (not themselves) and others that catalyze other chemical reactions unrelated to RNA splicing.

In eukaryotes, the m-RNA, t-RNA and r-RNA is shunted through the nuclear membrane to the cytoplasm. In prokaryotes, there is no nucleus so the RNA is already in the cytoplasm. In all organisms, the r-RNA is complexed with proteins to form functional ribosomes. The t-RNA's bind specific amino acids.

[VIDEO](http://www.youtube.com/watch?v=HSD1AlA1r4Y&playnext=1&list=PL909330AD83C287D4&index=30)

**3. Translation:**

In this process, amino acids are linked together into a protein. The particular sequence of amino acids that are linked together is determined by the sequence of nitrogenous bases in m-RNA. This process occurs at the ribosome.

**a. m-RNA attaches to the ribosome at the 5' end.** The ribosome has two reactive sites. The RNA moves through the ribosome until a specific sequence of nucleotides, AUG, is positioned in the first site. Three base sequence in the m-RNA are called 'codons'. This specific codon AUG, which starts the process of translation, is called the 'start codon'. All proteins made by all life forms initially begin with methionine, and use the codon AUG..

**b. a specific t-RNA molecule, with a complementary UAC anti-codon sequence, binds to the m-RNA/ribosome complex.** This t-RNA always carries the amino acid methionine. The [genetic code](http://facweb.furman.edu/~wworthen/bio111/code.htm) describes the relationship between 3-base codons in m-RNA and the amino acids they code for.

**c. Binding of the t-RNA to the first site opens a second site that reads the second 3-base codon (GCC in picture at right).** Another t-RNA binds here - one with the specific anti-codon sequence (CGG). This t-RNA, with this anti-codon, always binds with the amino acid alanine.

**d. Now a complex series of reactions occurs.** Methionine is cleaved from its t-RNA and bound to alanine (this peptide bond between amino acids forms via dehydration synthesis). The t-RNA in position 1 vacates the site, and the t-RNA in site 2 moves to site 1. This is called a 'translocation reaction'. The next 3-base codon is positioned in the second site - ready to accept the next t-RNA/ammino acid complex (for tryptophan in the picture to the right).

**e. Polymerization proceeds.** This process continues down the m-RNA strand, reading the message one codon (3-base "word") at a time. For each codon, a specific amino acid is added to the chain. Thus, the nucleotide sequence in the m-RNA - copied from the nucleotide sequence in the DNA gene - determines the sequence of amino acids in the protein.

**f. Termination.** There are some codons that have no corresponding t-RNA molecule. When these codons enter the second site, no t-RNA/amino acid is added. When the ribosome translocates, no new amino acid is added and the chain is terminated. These particular codons that stop translation are called "stop codons".

[VIDEO](http://www.youtube.com/watch?v=-zb6r1MMTkc&feature=related)

**4. Protein Processing:**

The initial protein product usually needs to be modified to become functional. These modifications are termed "post-translational modifications". First of all, the methionine is usually cut off - this relieves an important constraint on the structure of functional proteins... functional proteins DON'T all start with methionine! Then, the protein may be spliced, or it may be bound with a sugar group (glycoprotein), lipid (lipoprotein), nucleic acid (nucleoprotein), or another protein (quaternary protein). In eukaryotes, much of this processing occurs in the Golgi apparatus.

**C. Regulation of Protein Synthesis**

Aside from somatic mutations, all the cells in a multicellular organism are genetically identical. So, the cells in your retina, bone, muscle, and stomach lining all contain the same genes. These cells perform different functions because they are reading different genes and making different proteins. Your muscle has the gene for rhodopsin (a photoreceptive pigment produced in the retina), but that gene is not transcribed in muscle cells. In contrast, retinal cells have the genes for the muscle proteins actin and myosin, but these genes are not transcribed. So, cell specialization and the developmental process by which cells specialize from the fertilized egg occurs by regulating this process of protein synthesis. Regulation can occur at each of the steps described above.

**1. Regulation of Transcription:**

The process of transcription is regulated in several ways. First, the RNA polymerase can be blocked from the promoter. This can happen because the gene is bound to histones in a nucleosome, or is in a region of condensed 'heterochromatin', or because other proteins called 'transcription factors' have bound to the DNA - either at the promoter or between it and the gene, blocking the polymerase's route. However, the binding of other transcription factors can increase the affinity of the RNA polymerase for the promoter - increasing the probability of transcription. Again, these transcription factors are proteins encoded by other genes, and affected by other cellular processes. In this way, the action of a gene can be co-ordinated with the activity of other genes in a complex and interdependent manner. In addition, environmental cues from outside the cell can, through signal transduction, affect the activity of transcription factors and turn genes on or off. So, an organism can respond genetically to environmental cues.

**2. Regulation of Transcript Processing:**

The production of a protein can be affected at the processing stage. mi-RNA's and si-RNA's are small RNA molecules encoded by their own genes. These molecules can bind to m-RNA and effectively block correct splicing. This turns off production of the correct protein. In some cases, an initial m-RNA can be spliced two ways, creating two different functional products (and eventually two different proteins) depending on the pattern of cleavage. So, one gne may code for different proteins in different cells or tissue types.

**3. Regulation of Translation:**

One way that differential splicing can affect protein production is by changing the location of stop codons. For example, suppose a stop codon occurs at the beginning of an intron. Then, suppose that the intron is spliced incorrectly, after the location of this stop codon. Now, the resulting functional m-RNA has a stop codon where it didn't before; and translation will be terminated prematurely and no functional protein will be produced.

**4. Regulation of Post-Translational Modification:**

Initial protein products can be cleaved in different ways to produce different proteins, too.

So, through all of these mechanisms, protein synthesis can be stopped or stimulated, and the product can be modified. Again, all of these regulatory pathways can be affected by environmental factors or the proteins or mi/si-RNA's produced by other genes. So, gene activity is affected by other things happening in the cell (turning other genes on and off) , in other cells of the organism (through the production of hormones that act as signal transducers), or environmental factors outside of the organism acting directly on this gene, on other genes in this cell, or on other cells..

**III.  Evolution of a Genetic System:**

**A. Problem:**

DNA requires protenaceous enzymes to be replicated and read, but proteins require a DNA recipe... neither can function without the other in living cells.  Again, we have an apparently "irreducibly complex system" - it all must be there at once to work. You need an information storage molecule (DNA), and a catalyst to access the information (proteins). Both of these functions must be present at the same time for the system to work.

**B. Possible Solution:**

But there is another component of our genetic system....RNA. And we are finding that RNA does lots of wild stuff....

a.Even in current systems, RNA is message (m-RNA), decoder (r-RNA), and transporter (t-RNA).

b. RNA's have been discovered that have catalytic ability (ribozymes), so they can act as an information storage molecule (like DNA) and as a catalyst (like proteins).

c. some RNA molecules can self-replicate.

d. some RNA molecules (mi-RNA, si-RNA) regulate the production of proteins.

So, the possibility exists that genetic systems evolved from self-replicating RNA molecules.  Coding for proteins that assist the  replication process would be adaptive.  And incorporating DNA, which is more stable than RNA, would also be adaptive, producing the system seen today. This idea has been developed into a more formal hypothesis of the evolution of genetic systems, which we will describe next.

**C. Hypothesis for the Evolution of a Genetic System:**

                       **- GENETIC SYSTEM #1:  Self-Replicating RNA Arises Spontaneously:**

                             - Self-replicating RNA forms by polymerization of spontaneously produced nucleotides.

                             - It replicates.  That's all it has to do to be a successful genetic system...replicate.  Consider this: 90% of your DNA does not code for proteins - in some cases it may be regulatory, but there is ALOT of DNA that seems to have no function at all. However, when the DNA is replicated, this non-coding DNA is passed along to the next generation, too. In that sense, it is just as successful as the genes that do code for a protein.  So, replication is REALLY all that a genetic system has to do to perpetuate itself through time.

                             - But, it could replicate more effectively if it coded for replicating proteins.  So, m-RNA and r-RNA and t-RNA interact today to make proteins.  If this step happened next, then the replicating RNA  might code for a protein (an RNA polymerase) that increased the efficiency of its replication.

                       **- GENETIC SYSTEM #2: RNA systems produces useful replication enzymes.**

                             - So, now there is VARIATION IN GENETIC SYSTEMS... the old, slow, self-replicating RNA, and the new, more rapidly replicating RNA that codes for replication enzymes.  So, we have variation and DIFFERENTIAL REPRODUCTIVE SUCCESS... different RNA's that replicate (reproduce) at different rates.  What will happen????? Well, the more rapidly replicating molecules will be accumulate over time. They will increase in frequency in the "population" of chemicals...this is selection at a chemical level.

  **- GENETIC SYSTEM #3: Incorporating DNA**

                             - Now, how and why does DNA get involved?  Well, first the how.  If the RNA has a small mutation, it may make a slightly different protein... the protein might now link DNA nucleotides together, rather than RNA nucleotides (like the existing RNA replication enzymes).  This new enzyme is a 'reverse transcriptase' - it reads RNA and makes DNA. So now, the RNA is read and a DNA molecule is made. (Does this ever really happen today?  Are there enzymes that do this?  Yes, in retroviruses there are genes that code for reverse transcriptases that copy the viral RNA, make DNA, and then the DNA is inserted into the host's genome.  When this viral DNA is transcribed, the new viral genome of RNA is made.)

                             - So, there is now a DNA molecule, and a DNA polymerase that can make a copy and make it double-stranded.  There are also RNA polymerases (present already) which read a nulcei acid and make RNA.  So, here we have evolved a system where DNA codes for RNA that codes for protein, and proteins are used for DNA replication, transcription, and translation, JUST LIKE THE GENETIC SYSTEM WE HAVE TODAY.

                             - Why?  Why would DNA be advantageous?  Because its' double helix is more stable than the single helices of RNA.  Greater stability means fewer mutations, which is a good thing (adaptive). In addition, a two-step process of protein synthesis produces more proteins than a one-step process. Here's the idea in a nutshell. Suppose nucleic acids can be read at a given rate - say, 5 products per second. In a one-step process (RNA coding for proteins), 5 proteins can be made by reading the RNA template each second. So in two seconds, 10 proteins can be made. However, suppose that the first second was used to make copies of the RNA template? In a two-step process of DNA--> RNA--> protein, the first second is used to read the DNA and make 5 products... but these products are 5 RNA molecules. In the next second, each of these 5 RNA's can be used as a template for protein construction, with 5 proteins made off each of them...for a total of 25 proteins after 2 seconds. More productive.

**STUDY QUESTIONS:**

**1) Diagram the parts of an RNA nucleotide.**

**2)  Show how two nucleotides are linked together by dehydration synthesis reactions**.

**3) Why does the purine - pyrimidine structure relate to the complementary nature of double-stranded DNA?**

**4)  Draw a DNA double helix, showing three base pairs and the antiparallel nature of the helices.**

**5) Describe the higher levels of eukaryotic chromosome structure, including the terms nucleosome and solenoid.**

**6) What are two differences between euchromatin and hetochromatin?**

**7) How does the polymerase "know" to read the sense strand? (several elements to this answer....)**

**8) What 'cues' determine where transcription will start and stop?**

**9) Consider the DNA double helix, below.  Show what the RNA product will be after transcription.  Remember - the RNA Polymerase lands on the promoter, and then reads one strand..... pick the correct one and transcribe it!**

            3'\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_5'....... \_\_\_\_\_\_\_\_\_\_
                |    |    |    |    |    |    |    |    |    |    |    |    |    |            |                    |
                A  A  T  G  C  C  C  A   T  T  G  G  C   A           |                    |
                                                                                          PROMOTER
                T  T   A  C  G  G  G  T   A  A C  C  G   T           |                    |
           5'   |    |    |    |    |    |    |    |    |    |    |    |    |    | 3' ..... |                    |

**10) Describe the translocation reaction.**

**11) How is a polypeptide modified after translation to become a functional protein?**

**12) What are introns and exons, and how is protein synthesis modified to accommodate this structural change?**

**13) The process of protein synthesis and the universal genetic code provides one of the dramatic pieces of evidence of a common ancestry to all life. Explain why.**

**14) What two attributes of RNA lead biochemists to hypothesize that RNA may have been the first genetic system?
15) How and why might the synthesis of proteins by RNA have evolved?
16) How and why might have DNA been incorporated into this process?**