# DNA, RNA, Replication and Transcription

The metabolic processes described earlier (glycolysis, respiration, photophosphorylation, etc.), are dependent upon the enzymes present within cells. Most enzymes are proteins, (a few are RNA), and their presence within a cell is determined by the genetic information or hereditary material present. This material, contained primarily within the nucleus (eukaryotic cells) or nucleoid (prokaryotic cells), is **deoxyribonucleic acid**, commonly referred to as **DNA**.

## Background Information:

According to a National Geographic article (Vol. 150 #3, 1976), the human body contains trillions of cells and each cell contains around 100,000 genes (segments of DNA). This amount of information, if written out, would fill around 600, 1000-page books (give or take a few as influenced by font size, paper weight, etc.). Within cells, the genetic information is tightly coiled, but if the DNA from all the cells within the human body were stretched out and laid end-to-end, it would extend to the sun and back over 400 times. This same amount of DNA would fit in a box about the size of an ice cube. DNA is amazing material with respect to its information storage potential.

## Composition of DNA:

**Deoxyribonucleic acid** (DNA) is a polymer, i.e., a long, slender molecule composed of many, small, repeating units called **nucleotides**. Each cellular DNA molecule forms a **double helix** or **duplex**, i.e., includes two chains of nucleotides connected to one another by **hydrogen bonds**, and twisted into a helical configuration (like a twisted ladder). Each nucleotide (monomer) contains **deoxyribose** (a pentose monosaccharide or 5-carbon sugar), a **phosphate group** ($PO4^-$) and one nitrogenous base. The bases commonly found in DNA include **adenine**, **guanine**, **cytosine**, and **thymine**, frequently represented by the letters (A, G, C and T). Of these, adenine and guanine are **purine** bases (**purines**) and have two rings in their structure, while cytosine and thymine are **pyrimidine** bases (**pyrimidines**) and have only one ring (note that "y" words go together). The nucleotides within each strand of a DNA molecule are connected together by covalent bonds called **phosphodiester bonds**, formed between the sugar of one nucleotide and the phosphate group of the next. (Since a water molecule is removed each time one of these bonds is formed, DNA synthesis is another example of dehydration synthesis or condensation.) Each nucleotide chain has a specific orientation, determined by the positions of the phosphate ($PO4^-$) groups and hydroxyl ($OH^-$) groups associated with deoxyribose. The phosphate is connected to the number-5 carbon of the sugar and forms the 5' end of each chain. The hydroxyl group is found on the number-3 carbon of the sugar, and forms the 3' end of each chain.

The two, nucleotide chains within each DNA molecule are **antiparallel** and **complimentary** to one another. They are antiparallel because their orientation is opposite, i.e., they are up-side-down relative to one another from 5' to 3'. They are complimentary, because the nitrogenous bases forming the "rungs" of the DNA "ladder" always pair up in a specific manner. The purine base adenine is always connected to the pyrimidine base thymine by two hydrogen bonds (A=T), and the purine base guanine is always connected to the pyrimidine base cytosine by three hydrogen bonds (G≡C). Note that each **base pair** contains three rings, and that straight-sided letters pair together and curvy-sided letters pair together. Although individually, hydrogen bonds are weak, the two, nucleotide side-chains in a typical DNA molecule are held together fairly securely because there are so many hydrogen bonds present.

Most DNA molecules are twisted in a right-handed helix, i.e., when viewed from one end, wind away from the viewer in a clock-wise direction. Each turn contains about 10 base pairs, roughly perpendicular to the side chains, but each with a slight propeller-like twist between the bases. Two grooves form along the surface of each DNA double-helix, a **minor groove** (located between the two nucleotide strands) and a **major groove** (formed by the turns of the helix). In Prokaryotic chromosomes, most plasmids, mitochondria and chloroplasts, DNA molecules are covalently closed circular structures with no free ends (ccc-DNA). In eukaryotic chromosomes, DNA molecules are linear.

**Genes:**

**Genes** are hereditary units associated with **chromosomes**. As was explained during an earlier lecture, chromosomes are made up of **chromatin**, and chromatin is made of DNA and protein (recall **nucleosomes** = DNA wrapped around histone octomers). Genes are actually small sections of DNA that typically have some specific function within the cell. Some genes code for m-RNA and polypeptides, others code for t-RNA, r-RNA, s-RNA, etc., and some serve as regions involved in the regulation of gene expression. DNA forms the genetic information within cells, because cellular genes are composed of DNA, but not all genes are. The genes found within some viruses are composed of RNA.

**Composition of RNA:**

**Ribonucleic acid** (RNA) is also a polymer and like DNA is composed of **nucleotides** connected together by **phosphodiester bonds**. Cellular RNA molecules are single-stranded, i.e., contain only one chain of nucleotides (some viral RNA molecules are double-stranded). Each RNA nucleotide contains the pentose sugar **ribose**, a phosphate group ($PO4^-$) and one nitrogenous base. Although three of the bases found in DNA are also found in RNA (adenine, guanine and cytosine), the forth base, thymine is not. Instead, some RNA nucleotides contain the pyrimidine base **uracil**. (Thymine is 5'-methyl uracil). Although RNA molecules are polymers and often occur as long chains (sometimes thousands of bases in length), they are **<u>much</u>** shorter than DNA molecules.

Prokaryotic cells typically contain three different types of RNA molecules all coded for by different genes (some occurring as multiple copies). Eukaryotic cells contain more than three types. The functions of the different types of RNA molecules will be explained later.

**DNA Replication:**

Since DNA contains the genetic information within cells, it must be reproduced before a cell undergoes fission. This insures that each new cell formed contains the information necessary to function.

**Replication**, sometimes called **semi-conservative replication**, is the process involved when DNA molecules reproduce. It is a semi-conservative process in that each new DNA molecule formed contains half of the original molecule involved in the replication process (the original or "parental" strand). This is because during replication, each strand of the DNA duplex serves as a **template** or pattern for the new strand being formed. Given this feature, one might ask, just how old is DNA?

Replication can occur by more than one mechanism, and is a complex process involving multiple factors not presented here. When considered in simplified form, replication always requires three things:
1) An existing DNA molecule to serve as a pattern or template.
2) Enzymes – The heterogeneous proteins found in chromatin.
3) Energy – Because synthesis reactions are endergonic.

Within living cells, DNA replication typically begins at a specific site called the **origin of replication**, and proceeds in both directions away from that point. Prokaryotic cells such as those of *E. coli* generally have only one origin of replication within their circular chromosome, but eukaryotic cells have many along their linear chromosomes. The origin of replication within an *E. coli* chromosome (called *ori*C), is a sequence of nucleotides 245bp in length. This region contains specific base sequences recognized by and able to interact with initiation factors and enzymes involved in the process. At the origin, the two, nucleotide strands of the DNA molecule separate (hydrogen bonds break) and individual bases are exposed between them. This separation involves enzymes e.g., helicases and gyrase (topoisomerase II).

The primary enzyme involved in DNA replication is **DNA-dependent DNA polymerase**, often referred to simply as **DNA polymerase**. Prokaryotic cells such as **E. coli** typically have three DNA polymerase enzymes designated as DNA polymerase I, II and III. Of these, DNA polymerase III is the primary builder. Polymerase enzymes catalyze chemical reactions resulting in the formation of phosphodiester bonds, i.e., they synthesize polymers; however, DNA polymerase enzymes can only add nucleotides to the free, 3' ends of existing nucleotide chains (can build from 5' to 3'). They cannot initiate the formation of nucleotide strands from individual nucleotides without the presence of **primers**.

A **primer** is a short sequence of nucleotides (often around 18-20 bases in length) and when associated with DNA replication is composed of RNA nucleotides (primers used in the PCR are often made of DNA). Enzymes called **primase enzymes** build the RNA primers associated with replication. The first of these is formed near the origin, and provides the free, 3' end DNA polymerase requires for synthesizing DNA.

Once replication has been initiated, the DNA strands involved appear to form two **replication forks**, i.e., regions where the double helix separates into two, individual strands. These will travel in opposite directions (away from one another) as the original helix unwinds and replication proceeds. There is usually only one primer synthesized at the origin of replication, and it is associated with the **leading strand**. Once this primer is in place, DNA polymerase III can add DNA-type nucleotides to it and build a new complimentary strand (the leading strand) as a continuous sequence. The opposite strand forming the replication fork is called the **lagging strand**.

Although nucleotides are also exposed along the lagging strand, replication cannot occur there in the same fashion because DNA polymerase cannot build in the 3' to 5' direction. Instead, a group of proteins including primase, form a structure called a **primosome**, and this begins to migrate along the lagging strand traveling in the same direction as DNA polymerase III (on the leading strand). Periodically, as the primosome reaches specific nucleotide sequences along the lagging strand, it synthesizes new primers (primer synthesis occurs in the 5' to 3' direction, so is opposite the direction of primosome migration). These primers (also made of RNA) serve as new start points for DNA synthesis, and initiate the formation of a series of DNA fragments called **Okazaki fragments**. Each Okazaki fragment has a short RNA sequence at its 5' end, but they are composed primarily of DNA. The Okazaki fragment formed nearest the origin serves as the beginning of the leading strand associated with the other replication fork (i.e., the one traveling in the opposite direction away from the origin).

Since DNA molecules do not contain small segments of RNA, all the RNA primers formed during replication must be removed. This is accomplished by **DNA polymerase I**. It travels along the newly formed lagging strand, degrading the RNA primers and replacing them with DNA (it also removes the primer at the beginning of the leading strand). However, although DNA polymerase can add new nucleotides to a free, 3' end of an existing nucleotide strand, it cannot form a phosphodiester bond between two existing nucleotide strands. This requires a different enzyme called **ligase**. Ligase enzymes form the phosphodiester bonds attaching the multiple Okazaki fragments together, and bind the leading strand formed with one replication fork to the lagging strand formed with the other.

Note – There is considerable "proof reading" and "repair" associated with the replication process, such that most newly formed DNA strands are identical to their "parental" compliments. Errors occur at a pace of about 1 per 100 million copies of DNA (the spontaneous mutation rate described in a later section).

In addition to requiring DNA as a template, and the enzymes described above, replication also **requires energy**. Replication of the *E. coli* chromosome (around $4.6 \times 10^6$ bp), occurs in about 60 minutes, so proceeds at a pace of about 77 thousand nucleotides per minute (over 1000 per second). The process requires considerable energy, because the chemical reaction associated with the formation of each phosphodiester bond is **endergonic**. The energy required is provided by the nucleotides used in the building process, i.e., by dNTPs and rNTPs. **Nucleoside triphosphates** (NTPs) are high-energy molecules containing **pyrophosphate bonds** (recall the structure of ATP described in an earlier section). These bonds are broken as the nucleotides are incorporated into DNA, and the energy released is used to form the phosphodiester bonds holding the nucleotides together.

**Transcription – RNA Synthesis**

The term "**transcribe"** means to write out an exact copy of something; therefore, when not applied to biological activities, **transcription** refers to the process of copying written information. When applied to biological systems, it has essentially the same meaning.

**Transcription is RNA synthesis**, or the process used to build RNA molecules within living cells. Like replication, transcription requires DNA as a template or pattern, enzymes and energy.

Transcription occurs in association with DNA, so occurs in the nucleus, in the nucleoid, in association with plasmids and inside mitochondria and chloroplasts. The process is similar to replication in that it is initiated by the breaking of hydrogen bonds, and the separation of the two strands forming the DNA double helix or duplex. This process involves enzymes, e.g., helicases and gyrases. Once the strands are separated, **RNA polymerase** enzymes can begin the building process. Since transcription does not involve the formation of primers or Okazaki fragments, primase enzymes and ligase enzymes are not required. Neither is DNA polymerase.

**DNA-dependent RNA polymerase** is the enzyme involved in transcription, though it is usually referred to simply as **RNA polymerase**. In prokaryotic cells, this is a complex composed of five subunits. Four of these, designated as $\alpha$, $\alpha$, $\beta$ and $\beta'$ bind together to form a unit called the **core enzyme**. This unit is primarily responsible for the building of new RNA molecules, i.e., comprises the "work force". The fifth unit is called **sigma factor**, and has an alternate function. When the DNA strands have been separated, the sigma factor binds to a specific region on one strand called the **promoter site**. In bacteria such as **E. coli**, there are several different sigma factors, and each recognizes a different type of promoter, but most promoters have some features in common. They typically contain highly conserved nucleotide sequences (**consensus sequences**) such as TATAAT and a specific start site for transcription (often the sequence CAT). The position and orientation of the promoter determines where transcription will begin, and in which direction it will proceed, but only with the help of sigma factor.

Once sigma factor has attached to the promoter site, the core enzyme can bind and transcription can begin. Like DNA polymerase, the core enzyme builds in the 5' to 3' direction by catalyzing the formation of **phosphodiester bonds** at free, 3' ends. Unlike DNA polymerase, it can also start the building process. The nucleotide sequence of the DNA template determines the sequence of bases incorporated into the newly forming RNA strand, just as it would during replication, except that the purine base adenine codes for **uracil instead of thymine** (because RNA molecules do not contain thymine). The DNA strand serving as the template is often referred to as the sense strand, and the opposite strand or compliment as the non-sense strand. Which strand is actually being copied?

Like replication, transcription requires energy, and this is provided by rNTPs (activated nucleotides containing the sugar ribose). In prokaryotic cells, transcription is often **polycistronic**, which means that multiple structural genes are copied as one long m-RNA molecule. This is because many promoter sites are associated with **operons** and these typically contain structural genes arranged in a sequence (as will be described in greater detail later). Polycistronic transcription does not usually occur within eukaryotic cells (except within chloroplasts).

Note – many sources divide the processes of replication and transcription into three steps identified as **Initiation**, **Elongation** and **Termination**. Initiation involves separation of the DNA strands and the binding of DNA-dependent RNA polymerase (sometimes called primase) to the DNA template. Elongation involves the binding of nucleotides to form either DNA or RNA polymers, and Termination involves the release of polymerase enzymes (and in the case of transcription, release of the newly formed RNA) from the

DNA template. Both processes are complex and involve numerous details not included here.

**Go to the following site(s) for more information and illustrations:**

http://www.pbs.org/wgbh/aso/tryit/dna/index.html
Choose "DNA Workshop Activity", then "DNA Replication" (you will need to have "Shockwave" installed.

http://www.visionlearning.com/library/science/biology-1/BIO1.1-nucleic_acids.htm

To have even more fun:
http://www.eurekascience.com/ICanDoThat/dna_intro.htm

http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookPROTSYn.html