

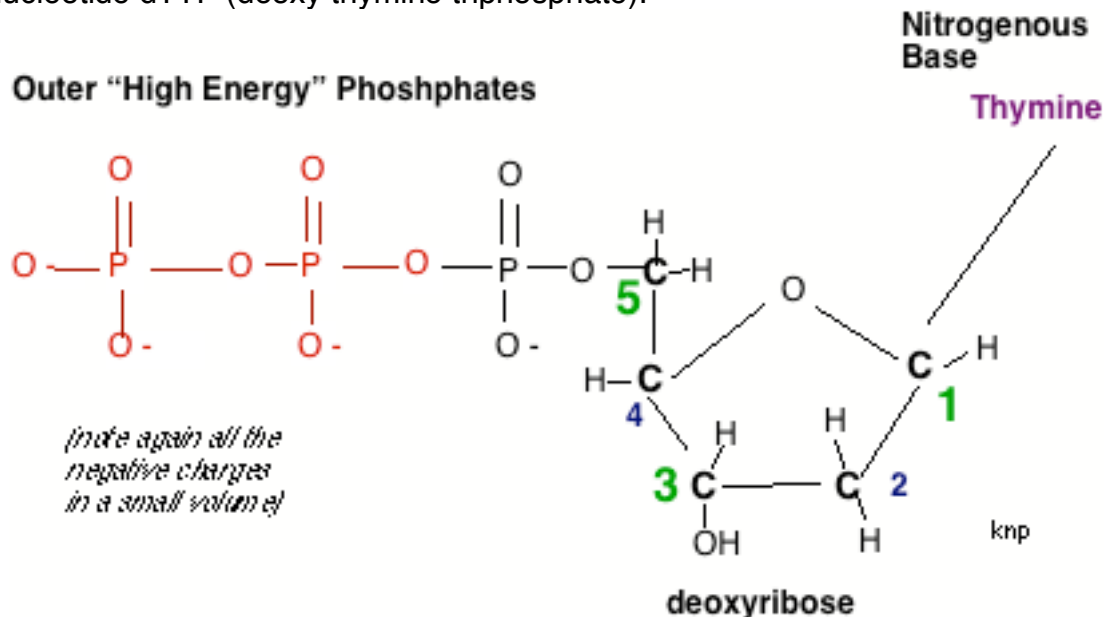
DNA and RNA SYNTHESIS¹

Introduction: For the last several weeks we have discussed the roles of proteins in cells. We have seen that their function depends entirely on their structure. Where does that structure come from? In the next few classes we will see how information relating to the structure of proteins is stored as a sequence of nucleotides in DNA. We will also see how a system decodes this information and reliably produces protein molecules with certain structures and functions. Knowledge of these processes and their meaning will form the basis for the next part of the course – inheritance and evolution.

This and the following sections assume that you are familiar with our earlier discussions of nucleotides, DNA, RNA, amino acids and protein structure. If not, go back and carefully review this material -- both in the notes and the textbook.

DNA Structure and Synthesis

With the exception of a few exotic viruses and some other strange situations, DNA always exists as a double helical structure. This means that a single DNA "macromolecule" actually consists of two strands that run parallel to each other, bind to each other via H-bonds between their nitrogenous bases, and that the entire molecule is twisted into a helix. Keep the last part in mind because I am going to depict it as a straight molecule (due to my limited artistic abilities). Let's look once again at the structure of a nucleotide -- in this case the nucleotide dTTP (deoxy thymine triphosphate):



¹ © 2006 by K.N. Prestwich and W. R. Healy, Dept. Biology, College of the Holy Cross, Worcester, MA 01610

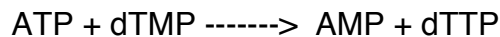
For the moment, let's focus on the deoxyribose sugar. Notice that I have numbered the carbon atoms (blue) in the sugar molecule. You need not learn this numbering, but you will need to understand its significance

In regards to the numbering, notice the following:

- The nitrogenous base is always on C1
- The phosphates are attached at C5
- There is a hydroxyl group on C3

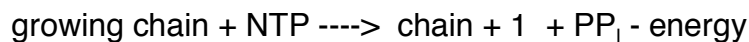
Recall from earlier that a strand of DNA (and for that matter RNA, see below) consists of a sugar-phosphate backbone. This structure is shown on the next page. Notice that the nucleotides are linked between the C3 and C5 of the sugar. This bonding pattern is the same whether the sugar is ribose or deoxyribose.

Prior to DNA and RNA synthesis, the cell produces amounts of NTPs sufficient for the synthesis. So, in the case of DNA, there would be large amounts of dATP, dCTP, dGTP, and dTTP. These could be made by transferring one or two high energy phosphates from ATP made in the "supply" reactions we considered earlier to the appropriate nucleotide monophosphate. Thus:



This is one case where more NTPs are made "from scratch" instead of simply being recycled (as in metabolism). This synthesis is a good example of anabolic process connected with growth and reproduction.

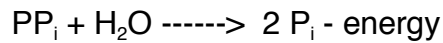
The reactions that grow DNA and RNA polymers ultimately consume two high-energy phosphate bonds per nucleotide added to the chain. Let's view the addition of a single phosphate to a pre-existing chain. The chain will grow from its 3' end (see figure on the next page). In other words, the new nucleotide is added to the chain by a reaction between the OH group on the #3 carbon of the growing chain and the phosphates on the number 5 carbon of the incoming NTP. In this process, the outer two phosphates will be split off the incoming NTP yielding what is called **pyrophosphate (PP_i -- H₂P₂O₇⁻² -- don't learn this chemical formula)**. **The two phosphates in PP_i are linked together by a high-energy bond:**



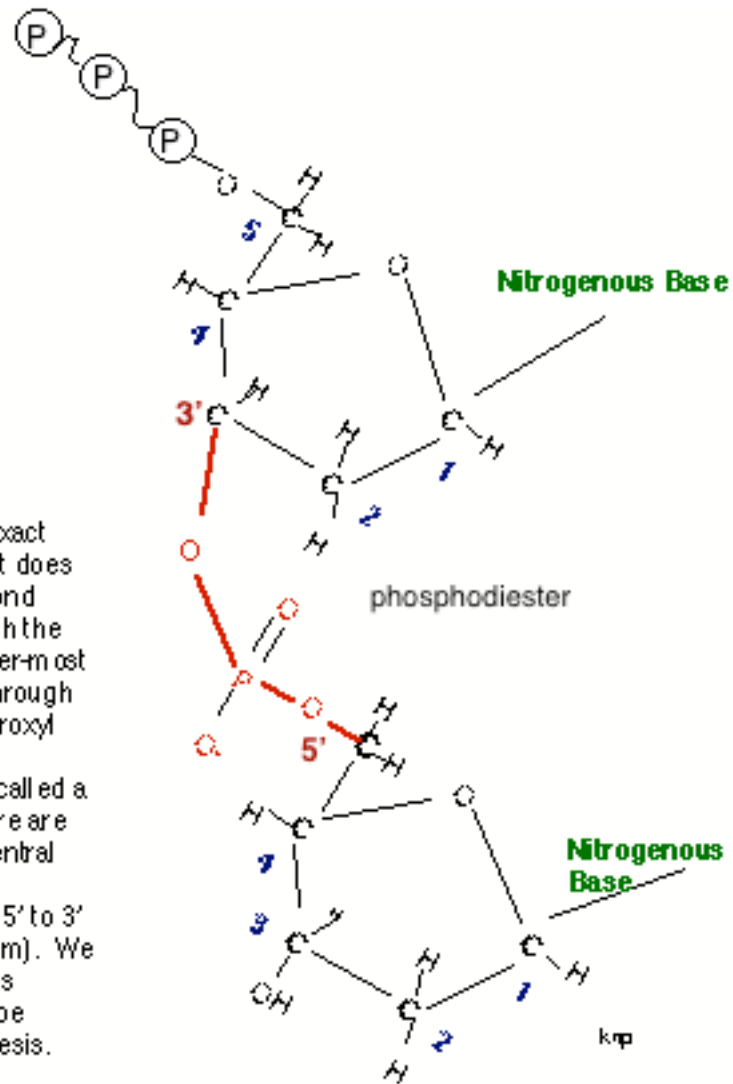
Notice that one high-energy phosphate bond was broken down in the process.

"Broken down" means that when the bond was transferred from the NTP to the end of the pre-existing nucleic acid chain that, although covalent, is no longer easily transferred to another molecule. It is a more stable bond.

The **PP_i** that was produced in the last reaction is immediately broken down into phosphates:

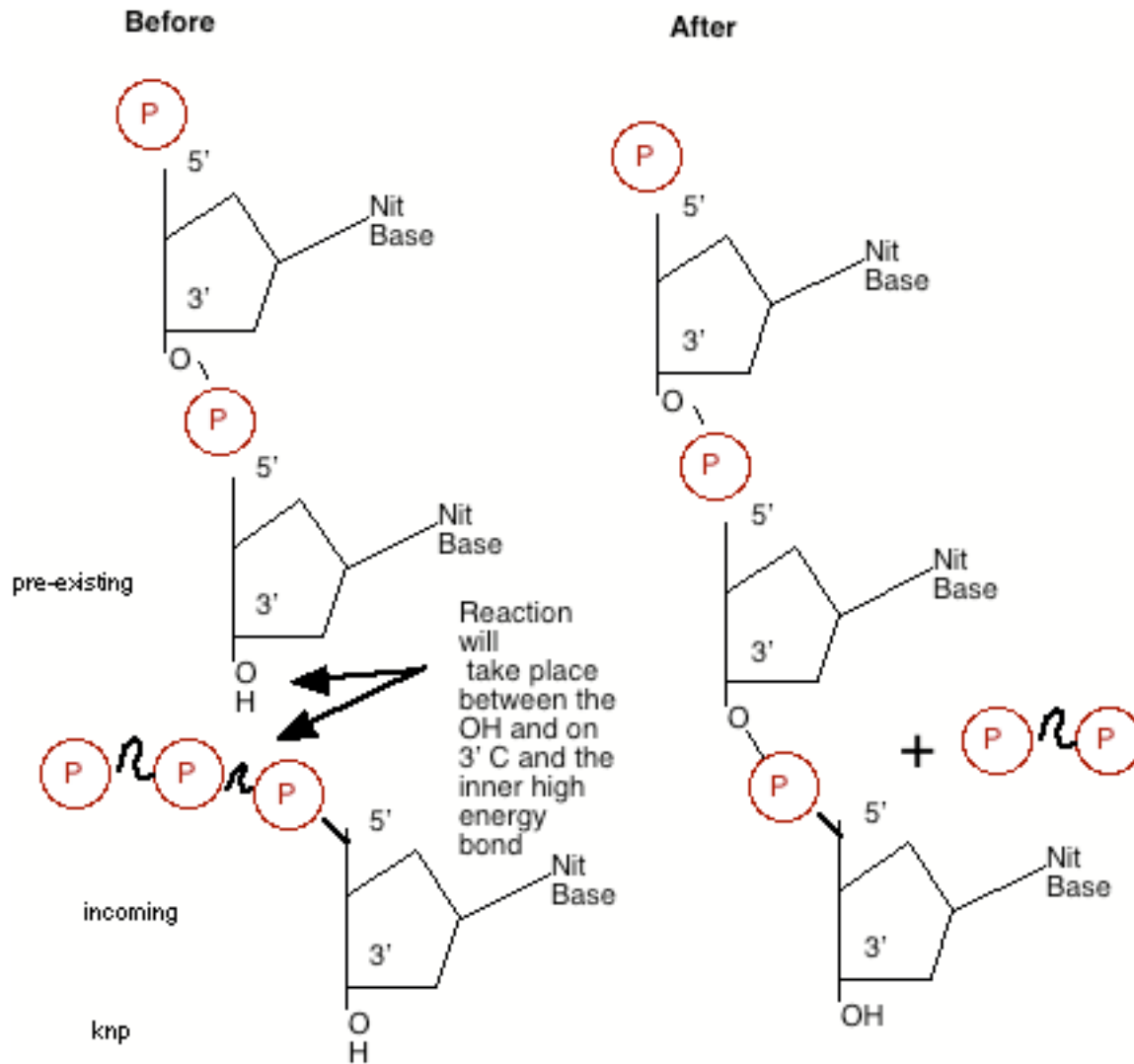


Before we go further, let's look at the linkage between two adjacent nucleotides on the backbone:



This drawing does not reflect the exact shape of the molecule. However, it does show what the sugar phosphate bond consists of. Notice that starting with the lower nucleotide, what was the inner-most phosphate on C5 is now bonded through an oxygen that was part of the hydroxyl group attached to C3 on the upper nucleotide. This particular bond is called a "phosphodiester linkage" since there are esters (-O-) on either side of the central phosphate. We can say that the deoxyribose sugars are linked in a 5' to 3' direction (downwards in this diagram). We will also see in a moment that this is direction that new nucleotides will be added during DNA and RNA synthesis.

On the next page is a cartoon that shows the reaction of adding a nucleotide to a growing chain:



IN SUMMARY, in both DNA and RNA polymerization, addition of the incoming nucleotide is to the 3' end of the pre-existing nucleic acid. This means that the molecule is actually lengthening in the 3' direction and so we say that:

- **ADDITION AND GROWTH OF A NUCLEIC ACID IS ALWAYS IN THE 5' TO THE 3' DIRECTION** -- it grows towards the 3' end!
- Notice that **nothing happens on the 5' end** of the pre-existing nucleic acid molecule. This shouldn't be surprising because the 5' end has only a single phosphate and is not very reactive.

The **other things that matter here** are that:

- 2 ~P are used per addition of a nucleotide to the polymer
- The ~P (energy) source for the reaction is the incoming NTP.

Energy and Nucleic Acid Synthesis
 What does it mean to say that ~P fuels this reaction?

Notice that the reaction that adds to the polymer moved a high-energy phosphate (covalent) bond from one molecule (the incoming NTP) to another (the 3' end of the chain). This bond becomes "**low energy**" as a result. Recall that this means that it is not easily broken or transferred. We have seen this same sort of thing in other reactions involving $\sim P$ -- for example, the activation of glucose prior to glycolysis.

? Two $\sim P$ are used in the process of adding one nucleotide. What does this say about the ease of reversing the reaction? Why might that be important in the case of these molecules? Put another way, what is the significance of breaking the PP_i down – why not develop a mechanism that uses some energy to attach the PP_i to a NMP to make a NTP?

We know what happens when polymers are elongated. What remains is far more interesting.

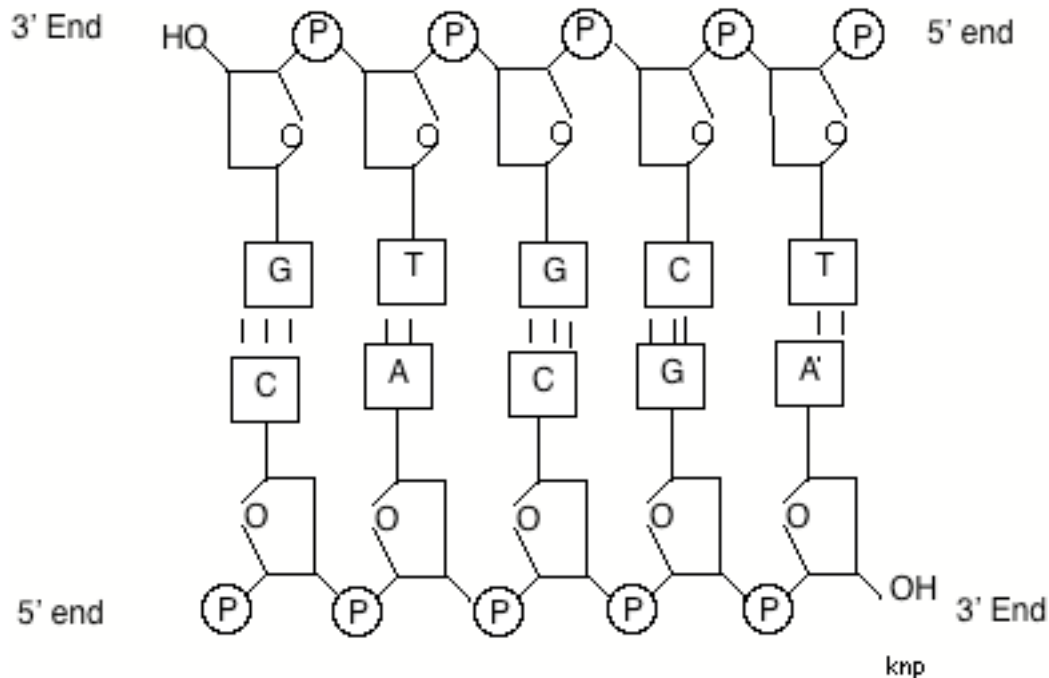
Complementary Pairing: As you know, it is common for nucleic acid polymers to interact with each other. This **interaction occurs between the nitrogenous bases**. The rules for this interaction are simple and we have learned them before.

- It is based on H-bonding.
- A **purine (two rings) always H-binds to a pyrimidine (one ring)**
- **Adenine (A, a purine) can form two hydrogen bonds as can the pyrimidines thymine (T) and uracil (U)**. Thus, adenine will bind with either thymine (in the case of DNA) or uracil (in the case of RNA or DNA to RNA binding)
- Likewise, **guanine (G, a purine) can form 3 H bonds as can cytosine (C, a pyrimidine)**; these two bind together also.

Pairing generally occurs:

- In DNA between the two opposite strands.
- In RNA, it either occurs with a piece of DNA when it is being synthesized or with other regions of the same or different RNA polymers. For instance, in transfer RNA (see next class) the molecule folds on itself in several places. In those regions complementary bases H-bond to each other and stabilize the resulting 3 dimensional structure. The same is true of rRNA and ribozymes.

Let's focus on DNA. We have seen that the two sides of the molecule each consist of polymers; the sides are held together by interactions between A to T and G to C nitrogenous bases. The **two strands run in opposite directions from each other**. This is because the 5C end of one side of the molecule is across from the 3C end of the other:

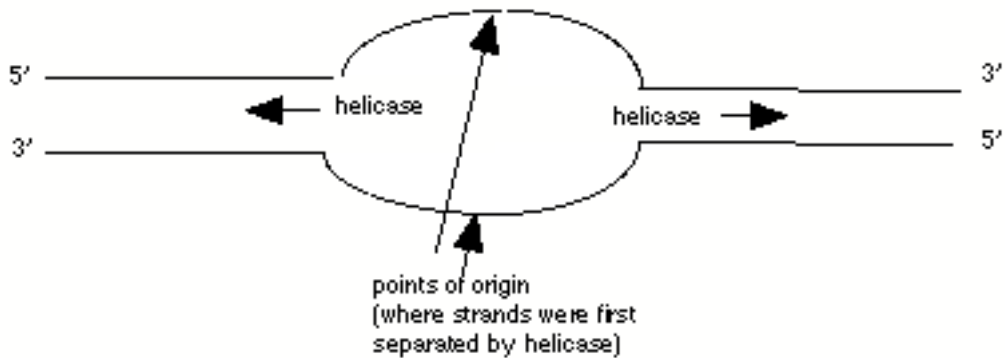


The term for this arrangement is **antiparallel**. One of the most important consequences of anti-parallel strands is that various enzymes involved in transcribing information in DNA to RNA or in synthesizing DNA will move in opposite directions when they work on the opposite strands.

DNA Replication: DNA replication is a rather complicated process. The purpose of this section is to move you beyond the simple understanding that all that is involved is that a double helix must "unzip" and then NTPs with bases that complement those on the strands being copied must line up and then be put together. That, of course, is a reasonable summary but there is more. So let's move up at least one level of sophistication.

First, recall that DNA is found in structures called chromosomes. Each chromosome consists of a single piece of double stranded DNA with lots of proteins attached to it. Normally, the chromosome is "relaxed" or "extended". It is within the nucleus and not visible.

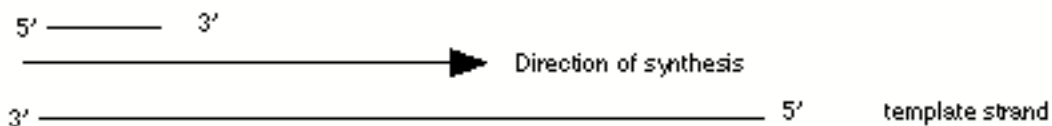
In eukaryotes, the DNA molecule in a given chromosome is quite long. It would take a long time to replicate it by going from one end to the other. So, **replication typically starts at hundreds or thousands of places along a chromosome at the same time.** Specific nucleotide sequences called **replication origin points** are recognized by enzymes called **helicases**. These can be thought of as attaching in pairs. After attaching, they begin to un-wrap the DNA spiral and separate the two strands from each other by breaking the H-bonds that pair the nucleotides. The result of the unwinding is the production of a small "**bubble**" where the strands are separated:



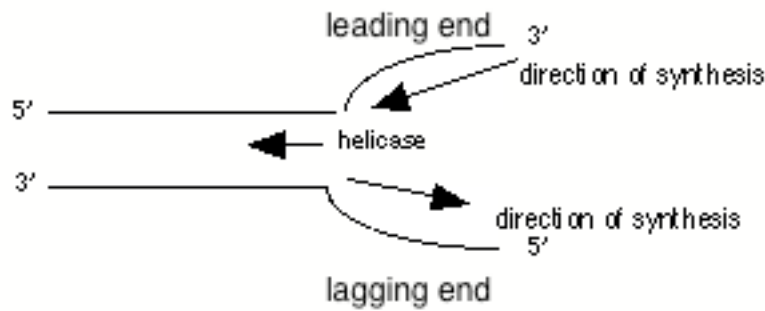
Notice that the helicase enzymes now move in opposite directions and will continue to do so until they meet up with helicases from another bubble. The areas where the helicases are working are called the **replication forks** for reasons that will be obvious in just a moment.

Now, we need to review what we learned about DNA synthesis at the start of these notes. Recall first that the pre-existing strands of DNA (the unzipped molecules shown above) will serve as **templates** to organize and direct the synthesis of new strands that bear **complementary** information.

- **Thus, the newly synthesized strand on one side (for example, the top above) will have the same information and same direction as the other strand that is being copied.** The eventual result will be two identical double helix molecules, each of which has one molecule that is new and one that is old. This is called **semiconservative replication** and it will be discussed in more detail in class.
- Furthermore, **recall that addition of new nucleotides to a strand always works by adding the 5' (phosphate) end to the 3' (OH end) of the preexisting strand** (see page 5). This means that **the template is "read" (must direct synthesis) in a 3' to 5' direction:**



Now, if you look back at the picture on the last page of the replication bubble, you'll see that in the case of one strand moving in one direction, there is no problem with synthesis. **Replication along this end of the fork is simply a matter of following the helicase.** However, in the other strand, the *movement is away from the movement of the helicase:*



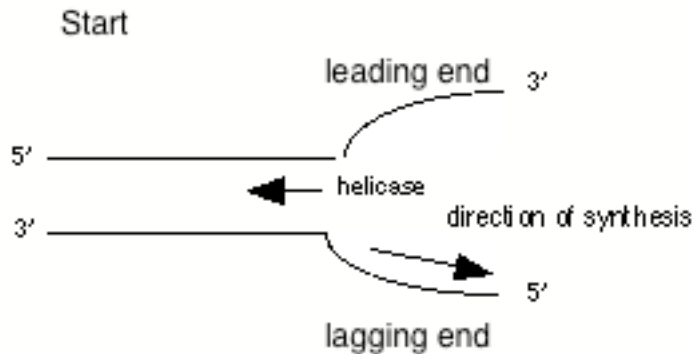
We call the part of the molecule where synthesis can follow the helicase at the **replication fork** the **leading end**. The side where it must work backwards is called the **lagging end**.

Let's consider the leading end first. Here, everything is easy. **Synthesis follows these steps:**

- First, an enzyme called **primase** comes in to the origin of the bubble and attaches to the DNA. It lines up and **polymerizes a small number of ribonucleotide triphosphates to make a very short strand of RNA** that is complementary to the DNA template. In the process, it slides along the DNA template in the 3' to 5' direction. This short piece of RNA is necessary because DNA polymerase, for some reason buried in evolutionary history, cannot start synthesis of DNA without a pre-existing nucleotide to which to attach. This piece of RNA, called an **RNA primer** will get removed and replaced later.
- Next, **DNA polymerase** comes in. **Starting with the exposed 3' end of the RNA primer begins synthesizing a molecule of DNA that is complementary to the DNA template.** It also reads this DNA along the template's 3' to 5' direction. Reading consists of lining-up the appropriate dNTPs and then catalyzing the reaction that synthesizes the sugar phosphate backbone (see earlier in these notes). The result is a new complementary DNA strand with a small pieces of RNA attached to its 5' end; Notice that **the new strand grows as always in the 5' to 3' direction** as we learned it must earlier in these notes.
- The DNA polymerase will follow the helicase until another bubble is reached. At that point the enzyme **DNA ligase** will join the strand to the one being produced at the next bubble and thereby produce one big molecule.
- One final thing needs to happen. **A different kind of DNA polymerase eventually comes in, removes the RNA primer, and replaces it with DNA.**

What happens in the lagging end? Most steps are the same as just outlined for the leading edge. However, the process is made somewhat more complicated because on the lagging sides it is impossible for synthesis to proceed directly behind the helicase. The reason is that the helicase is going the wrong way. **On**

the lagging sides, the helicase is moving along the DNA in the 5' to 3' direction. This would mean that synthesis of a new strand would have to go 3' adding to 5' and we have already seen that nucleic acid synthesis cannot work in that direction². **To get around this problem, synthesis on the lagging side works in the opposite direction of the helicase:**



Moreover, notice that by the time synthesis has started at one place, the helicase has moved further and so new areas that need to be replicated are opened up (compare the top and bottom pictures). **The result of this is that synthesis on the lagging side is discontinuous.**

- Once synthesis starts at any given place, it works exactly like what was described for the leading end. However, synthesis moves in the opposite direction from the movement of the helicase.
- This creates the possibility for other polymerase enzymes to begin synthesis in other places, closer to the present location of the helicase.
- The result is a series of short segments of DNA called **Okazaki fragments**.

² -- if nothing else because the enzymes do not recognize the 5' end of the nucleic acid as a starting point!

- These segments, however, are eventually stitched together by the action of **DNA ligase**.

Notice that because of (i) discontinuous replication on the lagging edges and (ii) the thousands of replication forks, a lot of ligase stitching is required to complete the replication of one DNA molecule. The final result will be two molecules from one. **Each molecule will be half new and half old** and with the exception of rare errors (**point mutations**) they will be identical to each other.

RNA Synthesis: RNA synthesis is quite easy to understand. DNA serves as the template (except for a few weird cases). The RNA strand will be complementary to the half of the DNA molecule that is being used as a template with the exception that U replaces T. Thus, if the DNA strand reads ATTGC, the RNA will read UAACG.

In RNA synthesis

- **The DNA molecule is, as always, read in the 3' to 5' direction.**
- **The resulting RNA strand will be synthesized by joining the 5' (phosphate) end of an incoming ribonucleotide triphosphate to the 3' (OH) end of the pre-existing strand.**
- There are no lagging sides.
- Moreover, **there is no need for a primer** because the enzyme that catalyzes the process of RNA synthesis, **RNA Polymerase**, does not require a primer.

Here is a bit more detail:

- To start things off, a particular portion of a DNA molecule is identified by one subunit of the enzyme **RNA polymerase**. The protein begins to unwind the DNA and break up the complementary pairing.
- Another subunit begins to line up complementary NTPs across from the nitrogenous bases of the "**sense' strand of DNA**". Only one of the two separated strands contains useful information and it is the only one that is copied. Thus, it is called the **SENSE** strand of the DNA double helix. The other side, which is not copied, is called the **anti-sense strand**.
- The RNA polymerase catalyzes the same reaction we saw earlier to form the new RNA polymer. The polymerase stops when it reaches a certain sequence of DNA nucleotides. The DNA molecule closes back up behind it and the RNA polymerase releases the DNA and the new molecule of RNA.

Study Questions:

1. Explain the rules of complementary base pairing (complementarity) and their significance in both DNA replication and RNA synthesis.

2. Explain why DNA and RNA elongation work in the 5' to 3' direction while the template strand is read in the 3' to 5' direction.

3. What is the difference between leading and lagging strands? Why is DNA replication continuous on the leading end but not on the lagging end?

4. Explain where Okazaki fragments are produced and why. What happens to them?

5. Terms:

DNA and RNA polymerases

Primase, DNA ligase and helicase

Sense and anti-sense strands

Antiparallel

Pyrophosphate, PPi

Replication fork, replication bubble