

## 1.23: Nucleic Acids

### Functions of DNA

Last time we examined how the amino acid sequence of a peptide or protein might be discovered. We also learned how a chemical synthesis of a small peptide can be carried through.

Today we'll study the chemistry of the molecule which carries the information necessary for directing the biosynthesis of proteins and peptides. This is DNA, and we'll learn that the structure of DNA provides a very strong rationale for its function.

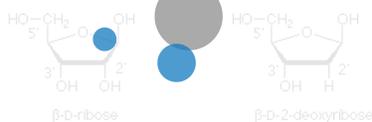
First, let's think a little about what a "molecule of heredity" needs to do. It must store an immense amount of information -- the directions for synthesizing all of the proteins necessary for the successful functioning of a living organism. It must be able to transmit that information faithfully, with an extremely low error rate, to the protein synthesis system, to both daughter cells upon cell division, and to a future generation upon reproduction of the organism.

Prior to the discovery of the structure of DNA, there was much speculation regarding possible molecules which might meet these requirements. Proteins themselves were seriously considered as candidates, since they are stable, they can be large enough to hold a large amount of information, and they are certainly intimately involved in biology. The information might be encoded in the sequence of amino acids in a protein, much like information in a word is encoded in a sequence of letters. We might think of the amino acids as a 20 character alphabet, with which we could make a very large number of words, sentences, paragraphs, books, etc.

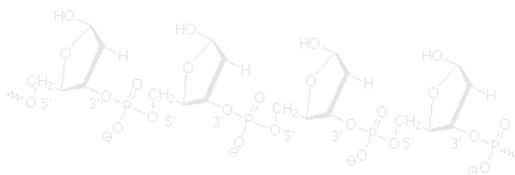
When the structure of DNA was established by James Watson and Francis Crick in 1953 it was immediately clear that it had the necessary characteristics to meet the specifications of a molecule of heredity. It is a polymer which can be extremely long. (The DNA in the single molecule which makes up a human chromosome is about 12 cm. long.) This provides the capacity to store large amounts of information. Like a protein this polymer backbone carries an alphabet. In this case, the alphabet consists of only four letters, A, C, G, and T. Let's look at the details.

### DNA Polymer Backbone

The polymer backbone is made up of two types of structure. One is a modification of the sugar ribose. The modification is that the OH group on the carbon next to the anomeric carbon has been replaced by a hydrogen. This is called 2-deoxy-D-ribose.



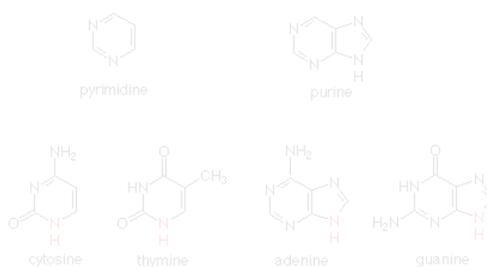
These 2-deoxyribose units are linked together by phosphate esters which link the 3' oxygen of one sugar with the 5' oxygen of the next. (The "primes" are there to differentiate atoms in the sugar ring from those in the bases, which we'll take up next.) This gives us a "backbone" for DNA which looks like this:



The "alphabet" molecules, A, C, G, and T, are attached to this backbone at the anomeric (1') carbons. Recall that these carbons are the ones where other groups can be attached by reactions such as those that convert hemiacetals to acetals. We'll next look at the structures of these molecules, which are called bases since they contain nitrogen atoms which make them mildly basic.

### The Bases A, C, G, and T

Two ring structures are found in the bases. C (cytosine) and T (thymine) have a single six membered ring, called a pyrimidine ring. A (adenine) and G (guanine) have two rings joined together. This unit is called a purine ring. C and T are called pyrimidine bases; A and G are called purine bases. Here are the structures:



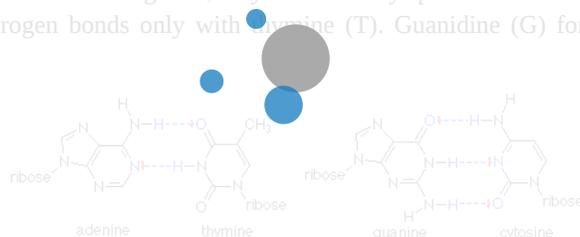
In each of these bases there is a **secondary amine** whose nitrogen forms a bond to the anomeric carbon of a deoxyribose in the DNA backbone. We can relate the chemistry of the formation of this linkage to the formation of a glycoside (acetal) from glucose (hemiacetal) and an alcohol. The difference is that in the current case the nucleophile is the secondary amine nitrogen of a base rather than the oxygen of an alcohol. An example of four bases attached in this way is:

The "word" here is CACT. Recall that the DNA backbone is very long, and it is clear that even with only a four letter alphabet, a great deal of information can be carried by DNA.

### Replication - Two DNA's from One

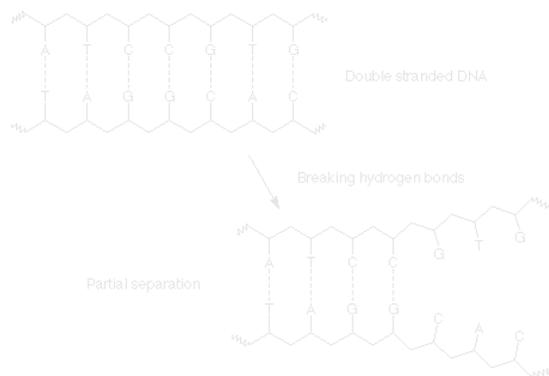
The next issue is transmission of the information to a daughter cell, to a succeeding generation, or to the protein synthesis machinery of a cell. The other key feature of the structure of DNA as discovered by Watson and Crick is that DNA molecules come in pairs, twisted together in the "double helix" (Figs. 19.8, 19.9; pp. 541,2 in Brown). Each of these molecules is a single long strand, held together by the covalent bonds along its backbone. The connections between the DNA strands are made by hydrogen bonds between the bases. Hydrogen bonds (as we learned when we studied amines) are much weaker than covalent bonds, but since there are many of them connecting the two DNA strands in the double helix, they serve very well to maintain that structure until there is a need for separation of the two chains.

Not only do the hydrogen bonds hold the chains together, they also are very specific in which bases are connected by the hydrogen bonds. Adenine (A) forms two hydrogen bonds only with thymine (T). Guanine (G) forms three hydrogen bonds only with cytosine (C).

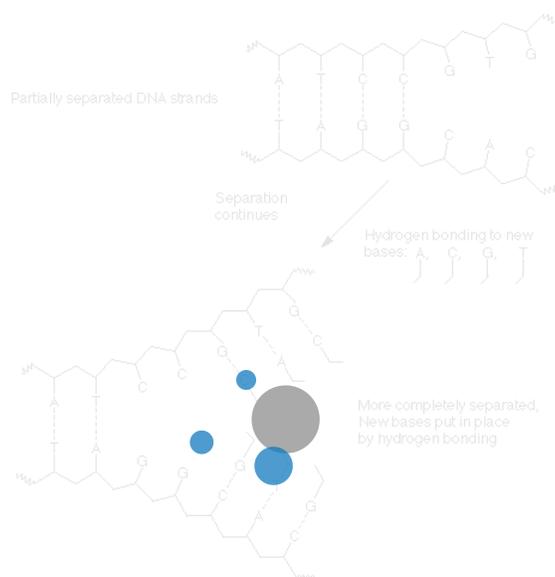


In each case, the **hydrogen bond** is formed between the positive hydrogen end of a polar N-H bond and a pair of electrons on either a nitrogen or a carbonyl oxygen. These "complementary" base pairs also have another important feature: a purine base (adenine or guanine) always bonds to a pyrimidine base (cytosine or thymine). This means that the distance between the two strands is always the same (three rings and the hydrogen bonds). Hydrogen bonding between two purine bases, for example, would put four rings into the base pair, and the fit would be poor. You can try to put together other hydrogen bonding patterns, but these two are the ones which fit best.

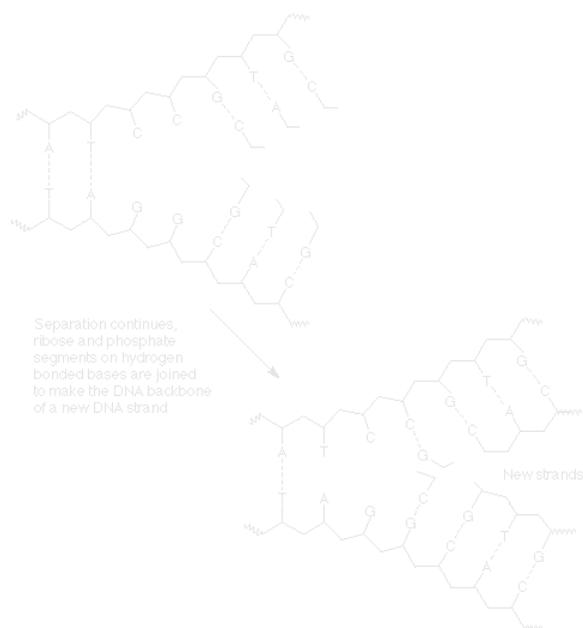
Watson and Crick realized that the specificity of this base pairing scheme was the key to replication of DNA and the transmission of information from one generation to the next. This is done in three steps. First the double helix is separated into the individual DNA strands by successively breaking hydrogen bonds between the base pairs.



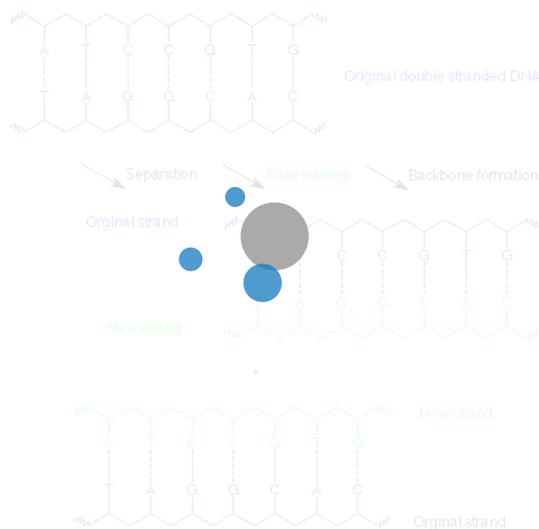
Second, as a segment of "unwound" DNA is exposed, bases from the solution encounter it, align with the complementary bases on the exposed DNA strands and form the proper base pairs, A with T and C with G. These bases are already joined to the necessary ribose and phosphate groups in molecules called nucleotides, so that as they line up in the proper arrangement, the materials for the formation of the backbone of a new polymer are in the proper locations.



Third, as separation and hydrogen bonding with new bases proceed, the individual nucleotides are joined together by the formation of new bonds between a phosphate of one nucleotide and the 3' OH group of the next nucleotide.



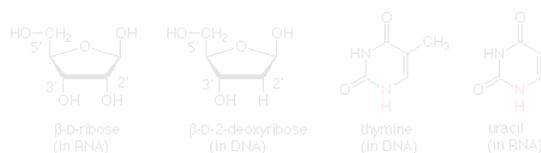
The outcome of these process is that each strand of the original DNA double helix has been used as a template upon which a copy of its former partner has been constructed. There are now two identical double helices which are the same as the original.



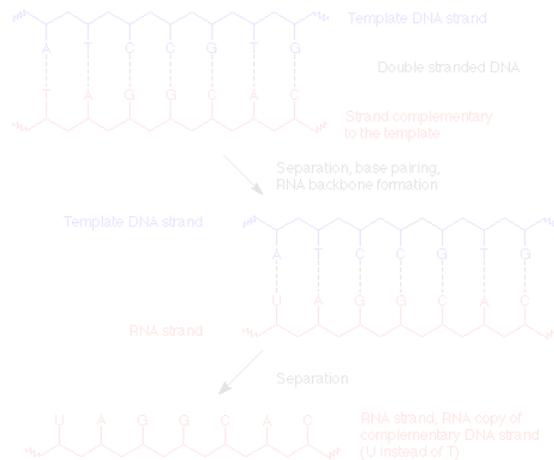
This is known as *replication*. In cell division each of these DNA copies would become part of one of the daughter cells. Each step in this process is assisted and controlled by enzymes, and there is also a "proofreading" function involved so that mismatched base pairs (such as an A-G pair) are excised and repaired.

## Transcription and Translation - DNA to mRNA to Protein

There are two successive processes by which the information contained on a DNA strand is used to determine the amino acid sequence of a protein. In the first of these, called *transcription*, a copy of the DNA strand is made, but in this case the copy is RNA. There are two structural differences between DNA and RNA. In RNA the sugar is ribose (with the 2' OH group) while in DNA it is 2-deoxyribose (without the 2' OH group). Also, where T (thymine) would occur in DNA, U (uracil) occurs in RNA.



The transcription process by which a RNA copy is made is very similar to the process by which DNA replicates. In this case, only a partial unwinding of the DNA helix occurs, and the appropriate bases hydrogen bond to the separated DNA, U with A, C with G. The RNA nucleotides which are now lined up on the DNA template are then joined together to form the RNA strand, which is a copy of the DNA strand which was **not** the template, and is complementary to the strand which was used as a template.



In this way an RNA strand which carries the genetic message (called "messenger" or mRNA) to the protein synthesis machinery (called a ribosome) is made. Its base sequence specifies the amino acid sequence in the protein to be made. The codes for each amino acid use three bases in a row and are given in table 19.3 (p 547 of Brown). Since there are 64 ways to make three letter "words" (called codons) with a four letter alphabet, many amino acids are coded for by more than one "word."

In the ribosome, the codons on mRNA are matched with anticodons (A with U, G with C) on transfer RNA (tRNA) molecules. Each tRNA molecule carries the appropriate amino acid to the enzyme which links them together to make the protein. This process is called *translation*. We won't look at the linkage process in detail, but it does include protection and activation steps much like the chemical synthesis we studied earlier.

The information flow in the overall process is: Codon sequence in DNA determines codon sequence in mRNA. Codon sequence in mRNA determines the order in which tRNA molecules line up. Order of tRNA line-up determines the sequence in which amino acids are linked to make the protein.

This page titled [1.23: Nucleic Acids](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Kirk McMichael](#).