Molecular Biology Primer: All About DNA

by Rich Deem

> **How does DNA=you?**
> This page explains the "central dogma" in biology - how DNA produces all the stuff that makes up living organisms.
>
> Rich Deem

Introduction

This page is a quick introduction to molecular biology. You will find on this site numerous references to DNA, RNA and proteins. In order to understand those pages, a basic understanding of how these molecules interact is necessary. This description is a very simplified version of what actually happens within a cell. In reality, the system is quite complex, with dozens of accessory molecules and co-factors required at each step.

DNA

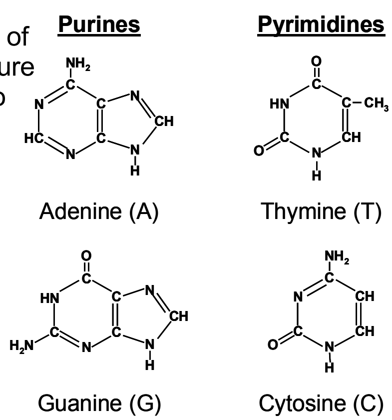
DNA (**d**eoxyribo**n**ucleic **a**cid) is the information storage molecule for the cell. Each cell contains within the DNA the entire instruction set to produce and run an entire organism (including you). In multicellular organisms, most of the DNA is stored in a central location within the cell, called the nucleus.

Bases

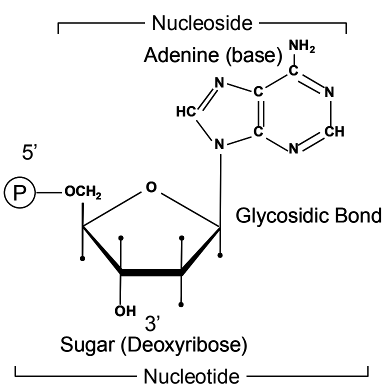
DNA is composed of only four bases, which can be thought of as letters of the DNA alphabet. The structure of these four letters is shown in the figure to the right. There are two purine bases (adenine and guanine) and two pyrimidine bases (thymine and cytosine). The bases are commonly abbreviated with the first letter of their names:

Nucleotide Bases

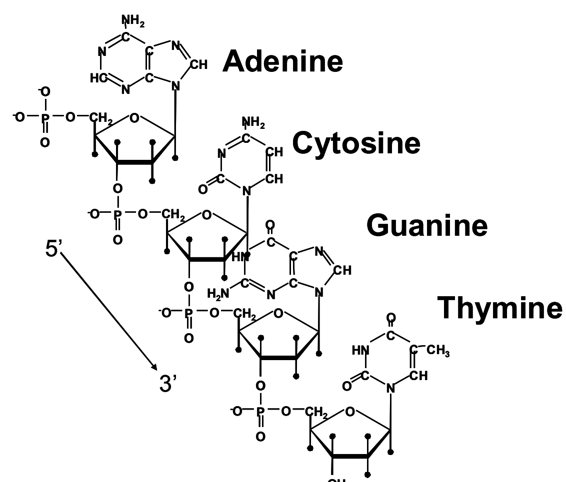
Base	Abbrev.
adenine	A
guanine	G
thymine	T
cytosine	C



Nucleotides



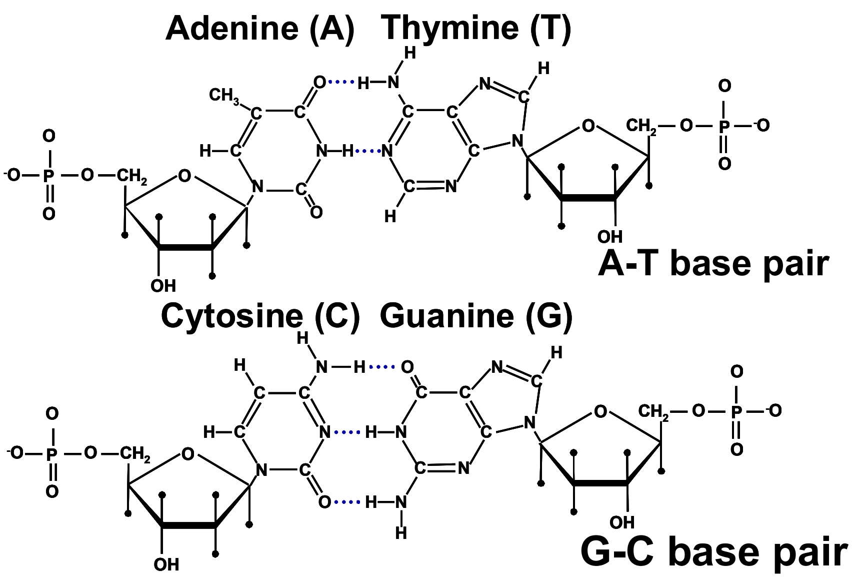
When a base, such as adenine, is linked to a deoxyribose sugar through a glycosidic bond, the structure is called a *nucleoside*. When the deoxyribose sugar is phosphorylated, on either the 3' or the 5' position (or both), the structure is a called a *nucleotide*. The precursors of DNA synthesis are deoxynucleoside-5'-triphosphates or dNTPs.

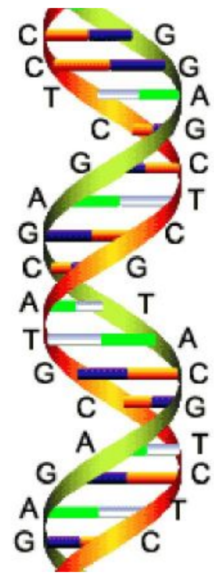


DNA structure

DNA is composed of a *polynucleotide* (multiple nucleotides) chain that is formed by linking nucleotides through 3',5'- phosphodiester bonds. In this way, DNA forms a long chain of sequential nucleotides.

DNA chains are not usually just single strands of nucleotides. Usually, nucleotides of two polynucleotide strands are base-paired to each other. The nucleotide base adenine pairs through hydrogen bonds with the nucleotide thymine forming an adenine-thymine (A-T) base pair. Guanine binds with cytosine forming a guanine-cytosine (G-C) base pair. Because of this base pairing, polynucleotide chains in double-stranded DNA are always *complementary* to each other.

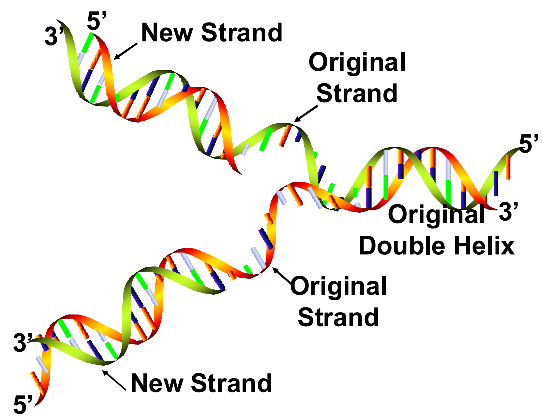




DNA Double Helix

The nature of the bonds between nucleotide bases results in the famous "double helix" DNA structure. The hydrogen bonding between strands of DNA holds the strands together until they are separated for DNA replication or RNA transcription.

DNA Replication



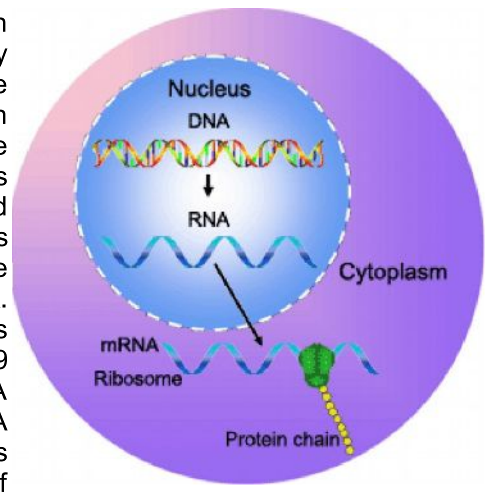
DNA is replicated (reproduced) through an extremely complicated process. This is the simplified version. The two strands of DNA are separated and a new complementary strand is synthesized for each old strand by adding complementary base pairs to the newly-synthesized strand (see figure at right). The DNA polymerase (the enzyme that makes new DNA) reads the base on the original strand and adds the complementary base to the newly synthesized strand. DNA is always synthesized in a 5' to 3' direction. Since one strand runs in the opposite direction, it must be synthesized in pieces that are later connected with DNA ligase. In reality, both strands are synthesized in pieces, since the DNA is unwound at numerous sites simultaneously on each chromosome in order to produce a complete copy in a short period of time (hours instead of weeks).

RNA

How does DNA turn in you?

Since DNA just stores information, there needs to be a way to turn the information into physical bodies. This feat is accomplished through the synthesis of an intermediate information-carrying molecule that carries the information from the nucleus to the called RNA (**ribo**nucleic **a**cid). As you can tell from the name, it differs from DNA in that part of it (the ribose sugar) has an extra oxygen compared to DNA (the "deoxy" version). In addition, RNA does not use thymine as a base, but another nucleotide known as uracil (see figure right). Uracil forms base pairs with adenine, as thymine does in DNA. RNA is transcribed from DNA through an enzyme called RNA polymerase, in a process that is known as transcription. There are several kinds of RNA, which have different functions. However, the RNA that eventually codes for proteins is called messenger RNA (*mRNA*).

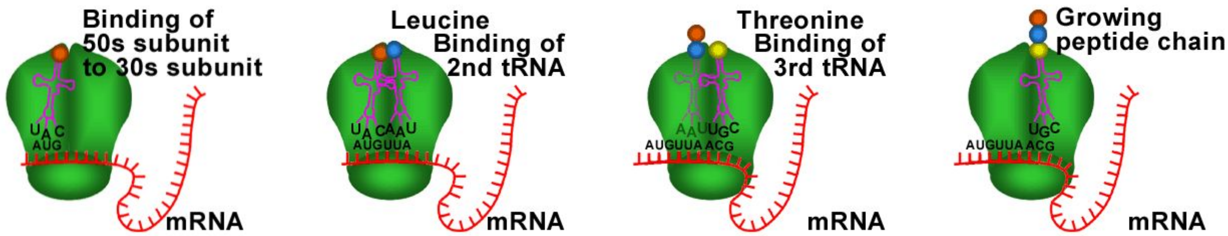
Translation



Once mRNA is synthesized, it is processed (usually through splicing) in the nucleus before moving to the cytoplasm. In many instances, alternate splicing of an RNA transcript can produce multiple proteins. It is one of the chief ways that 25,000 genes in the human genome can make ~100,000 proteins. Proteins are produced from an mRNA template using a process known as *translation*. The translation process is quite complicated and requires numerous molecular machines and co-factors. This description is a simplified version of what happens. A large molecular machine, called the *ribosome*, binds to the mRNA. Ribosomes are composed of two subunits. The large subunit is composed of three forms of ribosomal RNA (*rRNA*) and 49 different proteins. The smaller subunit is composed of one rRNA and 33 different proteins. A third kind of RNA, transfer RNA (*tRNA*), carries amino acids that are used to make proteins according to the mRNA template. One end of tRNA, consisting of three bases, called the *anticodon*, binds to a series of three complementary bases, called the *codon* on the mRNA strand. The other end of the tRNA binds to a specific amino acid. Specific sequence on the RNA are translated into proteins according to the *genetic code*. Each set of three nucleotide bases codes for one amino acid. This code is nearly universal for all organisms (with a few exceptions). The code is reproduced below:

The Genetic Code

<table>
<tr><th>Codon</th><th>AA</th><th>Codon</th><th>AA</th><th>Codon</th><th>AA</th><th>Codon</th><th>AA</th></tr>
<tr><td>UUU</td><td rowspan="2">Phe</td><td>UCU</td><td rowspan="4"></td><td>UAU</td><td rowspan="2">Tyr</td><td>UGU</td><td rowspan="2">Cys</td></tr>
<tr><td>UUC</td><td>UAC</td><td>UGC</td><td>UGC</td></tr>
<tr><td>UUA</td><td rowspan="2">Leu</td><td>UCA</td><td>UAA</td><td rowspan="2">Stop</td><td>UGA</td><td>Stop</td></tr>
<tr><td>UUG</td><td>UCG</td><td>UAG</td><td>UGG</td><td>Trp</td></tr>
<tr><td>CUU</td><td rowspan="4">Leu</td><td>CCU</td><td rowspan="4">Pro</td><td>CAU</td><td rowspan="2">His</td><td>CGU</td><td rowspan="4">Arg</td></tr>
<tr><td>CUC</td><td>CCC</td><td>CAC</td><td>CGC</td></tr>
<tr><td>CUA</td><td>CCA</td><td>CAA</td><td rowspan="2">Gln</td><td>CGA</td></tr>
<tr><td>CUG</td><td>CCG</td><td>CAG</td><td>CGG</td></tr>
<tr><td>AUU</td><td rowspan="2">Ile</td><td>ACU</td><td rowspan="4"></td><td>AAU</td><td rowspan="2">Asn</td><td>AGU</td><td rowspan="2">Ser</td></tr>
<tr><td>AUC</td><td>ACC</td><td>AAC</td><td>AGC</td></tr>
<tr><td>AUA</td><td rowspan="2">Met</td><td>ACA</td><td>AAA</td><td rowspan="2">Lys</td><td>AGA</td><td rowspan="2">Arg</td></tr>
<tr><td>AUG</td><td>ACG</td><td>AAG</td><td>AGG</td></tr>
<tr><td>GUU</td><td rowspan="4">Val</td><td>GCU</td><td rowspan="4">Ala</td><td>GAU</td><td rowspan="2">Asp</td><td>GGU</td><td rowspan="4">Gly</td></tr>
<tr><td>GUC</td><td>GCC</td><td>GAC</td><td>GGC</td></tr>
<tr><td>GUA</td><td>GCA</td><td>GAA</td><td rowspan="2">Glu</td><td>GGA</td></tr>
<tr><td>GUG</td><td>GCG</td><td>GAG</td><td>GGG</td></tr>
</table>



An model of how proteins are formed can be seen above. The 30s subunit of the ribosome binds to the mRNA. A specific tRNA (usually the first codon codes for methionine) binds to the codon of the mRNA The 50s ribosomal subunit binds to the 30s subunit. The next tRNA binds to the next codon on the mRNA and numerous initiation factors and the enzyme peptidyl transferase link the two amino acids together through a peptide bond. As the ribosomal complex moves along the mRNA, the peptide chain gets longer until the entire protein is formed. Other molecular machines are responsible for making sure that the protein folds into the correct three dimensional shape.

Conclusion

DNA is made of 4 chemical bases that are linked together sequentially. These 3 billion bases determine what you will look like and how your body will work. In order to express the information stored in the DNA, your cells have a complicated system that moves the DNA out of the cell's nucleus (through mRNA) into the cytoplasm. Once the mRNA is in the cytoplasm it is translated into proteins, which are the building blocks of the cell. Are you ready to use your new knowledge? Check out these articles:

- Is the Chemical Origin of Life (Abiogenesis) a Realistic Scenario?
- Origin of Homochirality: A Major Problem for Origin of Life Theories
- Evolution Deception in California State High School Biology Textbook Biology: Principles & Explorations
- Descent of Man Theory: Disproved by Molecular Biology
- Bad Designs in Biology? - Why the "Best" Examples Are Bad
- When Junk DNA Isn't Junk
- Pseudogenes: Argument for Evolution and Against Design?