

Appendix A

Review of Molecular Biology

This section provides an overview and concepts with which students should be familiar in order to get the most out of this lab. Please also refer to the Glossary Section (Appendix B) for definitions of molecular biology terms.

Any living organism functions based on the complicated interactions among nucleic acids, proteins, lipids (fat), and carbohydrates. In nearly all cases, certain proteins, termed enzymes, control the almost infinite number of interactions and life processes in living creatures. Think of enzymes and proteins as all the different people on earth. Each person performs a different role, function, or job on this planet, and although people are not the actual physical make-up of buildings, documents, food, and roads, it is the people that make these buildings and roads, and write the documents, and plant and nurture the crops. In the same way, enzymes and proteins do not comprise bones, lipids, sex hormones, and sugars, but enzymes control these structures, their interactions, and processes.

Because proteins and enzymes ultimately play such a critical role in the life process, scientists have spent many lifetimes studying proteins in an attempt to understand how they work and how they can be controlled. With a complete understanding, we could cure, prevent, and overcome many diseases and physical handicaps as well as explain exactly how and why organisms exist, propagate, and die. However, the complete answers do not lie solely in the knowledge of how enzymes function; we must learn how they are made. Before we can control enzymes, we must understand where they come from and what is the basis of the molecular information that encodes proteins. That answer lies within our **genetic code**.

Each living organism has its own blueprint for life. This blueprint defines how an organism will look and function (using enzymes as a means to form the appearance and control the functions). The blueprint codes for all the different enzymes. With amazing precision, this blueprint gets passed on from generation to generation of each species.

The transfer of this blueprint from generation to generation is called **heredity**. The blueprint for any organism is called its **genome**. The hereditary code is encrypted within the sequence of the DNA molecules that make up the genome. The molecule that constitutes the genome and thus the hereditary code is DNA (**deoxyribonucleic acid**).

The genome consists of very long DNA/protein complexes called **chromosomes**. Prokaryotes, organisms lacking a true nucleus, have only one chromosome. All other species, eukaryotes, have a defined cell nucleus that contains multiple chromosomes. The nucleus is a defined, membrane-enclosed region of the cell that contains the chromosomes. The number of chromosomes varies with the organism — from 2 or 3 in some yeasts to up to 100 or so in some fish. Humans have 46.

In most cases, chromosomes come in nearly identical pairs (one member of the chromosome pair from each parent). In general, the members of a pair differ in small details from each other, since they come from different parents, but are otherwise identical or **homologous**. Cells with homologous pairs of chromosomes are called diploid. Nearly all cells of an organism are diploid. Cells that have only one chromosome of each pair are called haploid. All sperm and ova are haploid.

The process of forming sperm and ova is called **meiosis**. **Meiosis** starts with a diploid cell that divides into two haploid cells. When a sperm fertilizes an ovum, the two nuclei fuse, and thus the new nucleus contains pairs of each chromosome, one partner from each parent. The result is called a diploid zygote.

All cells of diploid organisms duplicate chromosomal pairs when they divide (except when sperm and ova are formed), so that all body cells (called somatic cells) of an organism are diploid. The process of cell division in which the chromosomes are duplicated and each daughter cell gets pairs of chromosomes is called **mitosis**. It is through the processes of mitosis and meiosis that the hereditary code is passed from cell to cell and generation to generation. Now that we know where the code is and how that code is passed on, we need to know how the code produces the enzymes that control life. The actual DNA code for a protein is contained within a segment of a chromosome called a gene. In nearly all cases, diploid organisms will have the same gene on a specific chromosome pair. Each gene on a particular chromosome of a specific chromosome pair is also called an **allele**.

To clarify, a gene encodes a particular protein that performs a particular function. An allele is a specific version of a gene on a particular chromosome. Thus, there are genes for hair color and there is an allele for the hair color gene on each chromosome pair. The gene or allele's DNA code can also be called the genotype.

When the protein is made from this code and performs its function, the physical trait or result that is seen is called the **phenotype**. In many cases the two alleles on the specific chromosome pair coding for a protein differ slightly in their respective DNA code (genotype). Any slight difference in code between the two alleles can result in two different proteins, which, although intended to perform basically the same function, may carry out that function slightly differently, causing different results and thus different phenotypes.

Therefore, it is not only the various combinations of chromosomes a parent contributes to each offspring, but also the various combinations of alleles and how each of the enzymes coded from the alleles work together that decide how we look and allow us to function. The various combinations are nearly infinite and that is why we are all different. The study of genotypes and phenotypes is often referred to as **Mendelian genetics** (after Mendel, the individual who pioneered the study of heredity and genetics).

DNA: What Is It?

A DNA molecule is a long polymer consisting of four different components called **bases**. The four bases are also called **nucleotides**. It is the various combinations of these four bases or nucleotides that create a unique DNA code or sequence (also genotype, gene, and allele). Nucleotides are comprised of three different components:

- **Nitrogen base**
- **Deoxyribose sugar**
- **Phosphate group**

Each nucleotide contains the same ribose sugar and the phosphate group. What makes each nucleotide unique is its nitrogen base. There are four nitrogen bases:

Adenine (A)

Thymine (T)

Guanine (G)

Cytosine (C)

A DNA nucleotide chain is created by the connection of the phosphate group to the ribose sugar of the next nucleotide. This connection creates the "backbone" of the DNA molecule.

To designate the different ends of this single-stranded chain, we use some typical biochemistry terminology, in which the carbons on any sugar are numbered. The sugar of a nucleotide contains 5 carbons. The phosphate group (PO_4) of a given nucleotide is connected to the 5' carbon of the sugar. A hydroxyl group (OH) is attached to the 3' carbon of the sugar, and this 3' OH group connects to the phosphate group of the next nucleotide in the chain.

Thus, the end of a single-strand DNA molecule that has a free phosphate group (i.e., not attached to another nucleotide) is called the 5' end, and the end of the DNA molecule (with no subsequent nucleotide attached) is called the 3' end (see Figures 14 and 15).

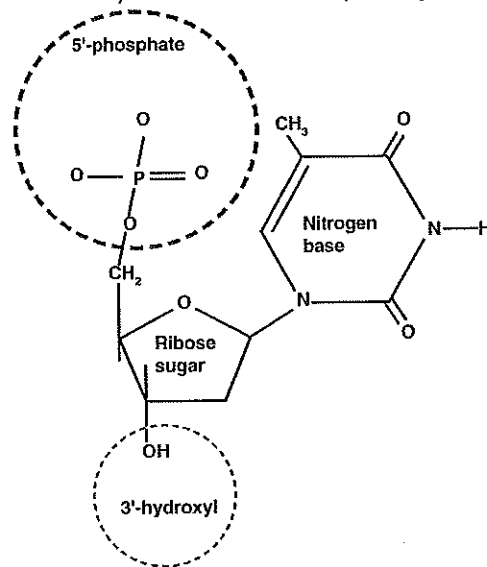


Fig. 14. Structure of one nucleotide of deoxyribonucleic acid.

It has become standard that a single-stranded DNA molecule is written with the 5' end on the left and the 3' end on the right. Therefore, a single-stranded DNA chain's sequence is represented from left to right, starting on the left with the 5' nucleotide and moving to the right until the 3' nucleotide is last. Most DNA sequences are read 5' to 3'.

However, the long DNA molecules or chains that comprise the chromosomes are not single-stranded molecules. From X-ray crystallography patterns of DNA, and some imaginative molecular model building, Watson and Crick deduced that DNA is in fact a **double-stranded** molecule with the two single strands of DNA held together by **hydrogen bonds** between the nitrogen bases (A, T, G, and C). This double-stranded molecule is often

called a duplex (Figure 15). There are several important properties of double-stranded DNA molecules.

- Chromosomal (also called genomic) DNA is double-stranded.
- The overall structure is that of a helix with two strands intertwined.
- The structure can be viewed as a twisted ladder.
- The phosphate-deoxyribose backbones are the sides of the ladder.
- The nitrogen bases (A, T, G, and C) hydrogen bonded to each other are the rungs.
- Only the nitrogen bases A and T and C and G can form hydrogen bonds to each other. When A binds to T or C binds to G this is considered **base pairing**. Neither C and T, nor A and G form hydrogen bonds.
- The two strands are antiparallel; that is, the strands are oriented in opposite directions. This means that the ladder runs 5' to 3' in one direction for one strand and 5' to 3' in the opposite direction for the other strand.

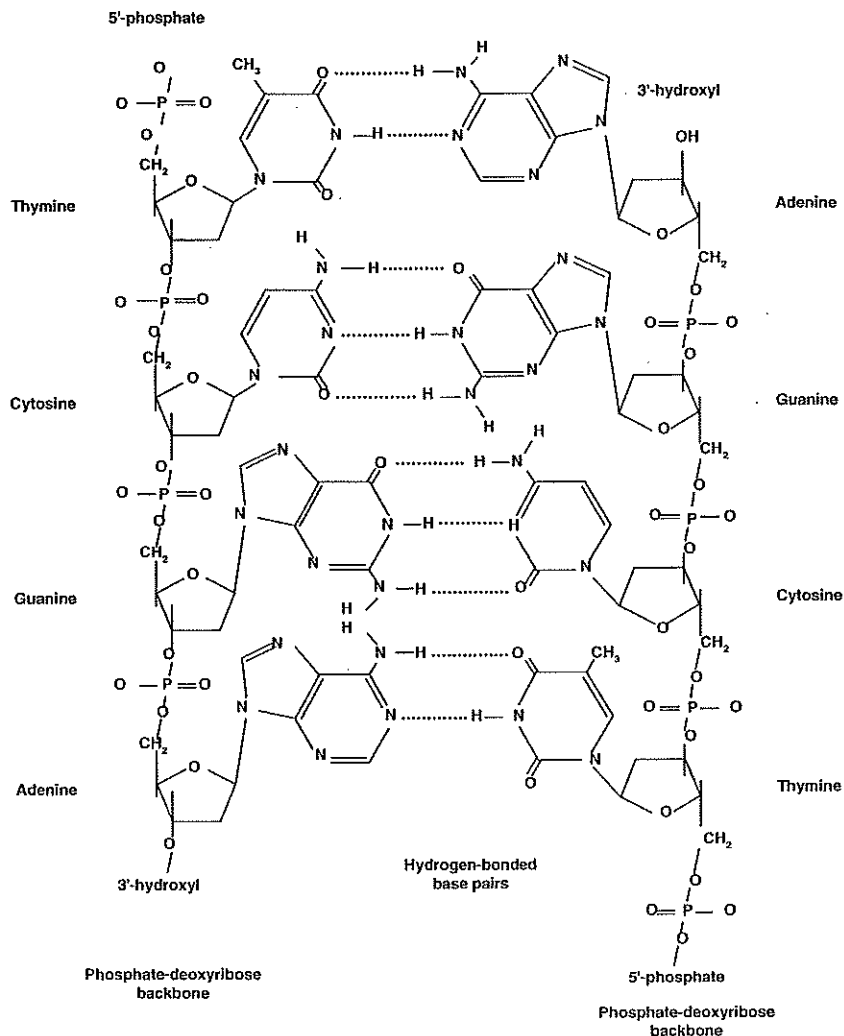


Fig. 15. Molecular structure of a portion of a double-stranded DNA molecule.

DNA Structure Conclusions

- Because A only binds to T, and G only binds to C, the two strands will have exactly the opposite, or complementary, sequence running in opposite directions (one strand 5' to 3', the other 3' to 5').
- These two complementary strands anneal or hybridize to each other through hydrogen bonds between the bases.
- A new strand of DNA can be synthesized using its complementary strand as the template for new synthesis.
- Each strand carries the potential to deliver and code for information.

The length of any double-stranded DNA molecule is given in terms of base pairs (bp). If a DNA strand contains over a thousand base pairs, the unit of measure is kilobases (1 kb = 1,000 bp). If there are over one million base pairs in a strand the unit of measure is megabases (1 Mb = 1,000 kb).

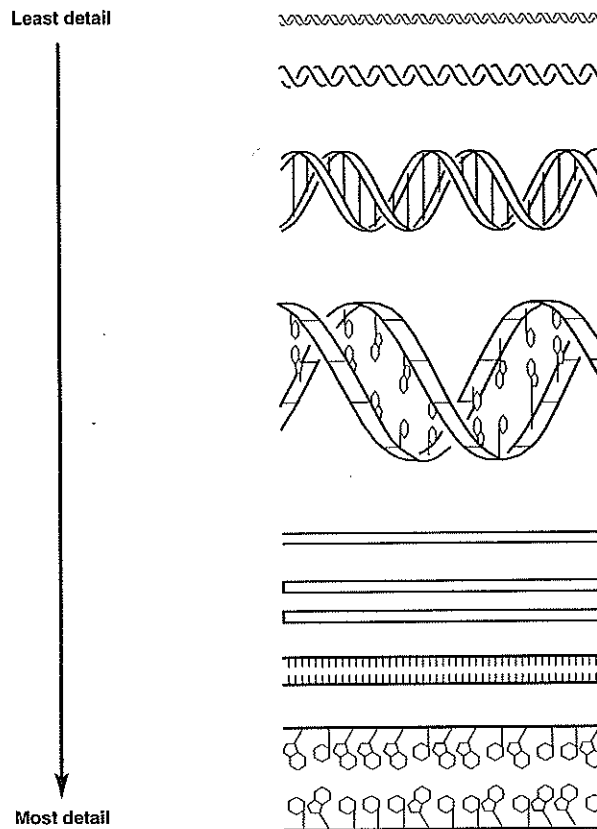


Fig. 16. DNA (deoxyribonucleic acid) — A long chainlike molecule that stores genetic information. DNA is graphically represented in a number of different ways, depending on the amount of detail desired.

DNA Replication — Strand Synthesis

New strands are synthesized by enzymes called **DNA polymerases**. New strands are always synthesized in the 5' to 3' direction. For a new single strand of DNA to be synthesized, another single strand is necessary. The single strand of DNA that will be used to synthesize its complementary strand is called the **template strand**.

However, in order for DNA polymerase to start synthesizing a new complementary strand, a short stretch of nucleotides (approximately 20 base pairs long) called an oligonucleotide primer must be present for the polymerase to start synthesis. This primer is a short stand of nucleotides complementary to the template where the researcher wants synthesis to begin. The primer must have a free 3' hydroxyl group (OH) for DNA polymerase to attach the 5' phosphate group of the next nucleotide.

The DNA polymerase grabs free (single) nucleotides from the surrounding environment and joins the 5' phosphate of the new nucleotide to the 3' hydroxyl group (OH) of the new complementary strand. This 5' to 3' joining process creates the backbone of the new DNA strand.

The newly synthesized strand maintains its complementarity with the template strand because the DNA polymerase only joins two nucleotides during new strand synthesis if the new nucleotide has its complement on the template strand. For example, the DNA polymerase will only join a G to the 3' end of the newly synthesized strand if there is the C counterpart on the template strand to form a hydrogen bond. Guanine will not be joined to the new strand if adenine, thymine, or guanine is the opposite nucleotide on the template strand.

DNA polymerase and strand synthesis allow DNA to replicate during mitosis. Both new DNA strands are synthesized simultaneously from the two original DNA template strands during mitotic DNA replication.

As you can see, DNA, RNA, and proteins are closely tied to each other. Thus, you can realize why researchers today, in an attempt to understand the mechanisms behind the various life processes, must study the nucleic acids as well as the proteins to get complete answers about the flow of information carried in the genetic code. In the last 20 years, many gains in the areas of nucleic acid techniques have finally allowed researchers the means to study the roles of nucleic acids in life processes.

Individual discoveries by many scientists have contributed the pieces that have begun to solve one of the most mysterious puzzles of life — understanding the hereditary code. In 1985, enough pieces of the puzzle were in place for a major breakthrough to occur. This understanding of how the necessary molecular components interact to faithfully replicate DNA within living cells led to the development of a technique for creating DNA in a test tube. This technique is called the **polymerase chain reaction**, or **PCR**.