A Rapid Introduction to Molecular Biology

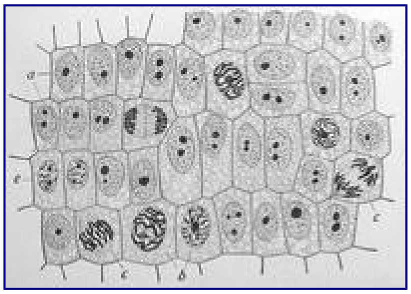

Figure 1. A 1900 drawing by Edmund Wilson of onion cells at different stages of mitosis. The sample has been dyed, causing chromatin in the cells (which soaks up the dye) to appear in greater contrast to the rest of the cell.

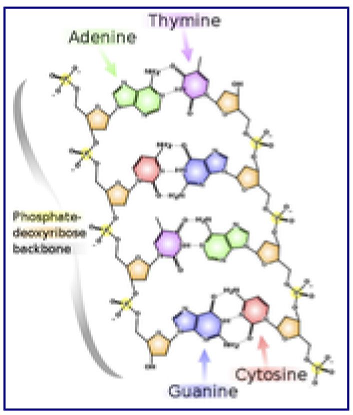

Figure 2. A sketch of DNA's primary structure.

Making up all living material, the cell is considered to be the building block of life. The nucleus, a component of most eukaryotic cells, was identified as the hub of cellular activity 150 years ago. Viewed under a light microscope, the nucleus appears only as a darker region of the cell, but as we increase magnification, we find that the nucleus is densely filled with a stew of macromolecules called chromatin. During mitosis (eukaryotic cell division), most of the chromatin condenses into long, thin strings called chromosomes. See Figure 1 for a figure of cells in different stages of mitosis.

One class of the macromolecules contained in chromatin are called nucleic acids. Early 20th century research into the chemical identity of nucleic acids culminated with the conclusion that nucleic acids are polymers, or repeating chains of smaller, similarly structured molecules known as monomers. Because of their tendency to be long and thin, nucleic acid polymers are commonly called strands.

The nucleic acid monomer is called a nucleotide and is used as a unit of strand length (abbreviated to nt). Each nucleotide is formed of three parts: a sugar molecule, a negatively charged ion called a phosphate, and a compound called a nucleobase ("base" for short). Polymerization is achieved as the sugar of one nucleotide bonds to the phosphate of the next nucleotide in the chain, which forms a sugar-phosphate backbone for the nucleic acid strand. A key point is that the nucleotides of a specific type of nucleic acid always contain the same sugar and phosphate molecules, and they differ only in their choice of base. Thus, one strand of a nucleic acid can be differentiated from another based solely on the *order* of its bases; this ordering of bases defines a nucleic acid's primary structure.

For example, Figure 2 shows a strand of deoxyribose nucleic acid (DNA), in which the sugar is called deoxyribose, and the only four choices for

> nucleobases are molecules called adenine (A), cytosine (C), guanine (G), and thymine (T).
>
> For reasons we will soon see, DNA is found in all living organisms on Earth, including bacteria; it is even found in many viruses (which are often considered to be nonliving). Because of its importance, we reserve the term genome to refer to the sum total of the DNA contained in an organism's chromosomes.

Example

A string is simply an ordered collection of symbols selected from some alphabet and formed into a word; the length of a string is the number of symbols that it contains.

An example of a length 21 DNA string (whose alphabet contains the symbols 'A', 'C', 'G', and 'T') is "ATGCTTCAGAAAGGTCTTACG."

Given: A DNA string *s* of length at most 1000 nt.

Return: Four integers (separated by spaces) counting the respective number of times that the symbols 'A', 'C', 'G', and 'T' occur in *s*

.

Sample Dataset

AGCTTTTCATTCTGACTGCAACGGGCAATATGTCTCTGTGTGGATTAAAAAAAGAGTGT
CTGATAGCAGC

Sample Output

20 12 17 21

Transcribing DNA into RNA

The Second Nucleic Acidclick to collapse

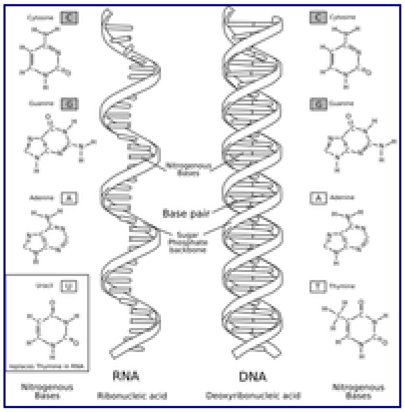

Figure 1. Structural differences between RNA and DNA

In “Counting DNA Nucleotides”, we described the primary structure of a nucleic acid as a polymer of nucleotide units, and we mentioned that the omnipresent nucleic acid DNA is composed of a varied sequence of four bases.

Yet a second nucleic acid exists alongside DNA in the chromatin; this molecule, which possesses a different sugar called ribose, came to be known as ribose nucleic acid, or RNA. RNA differs further from DNA in that it contains a base called uracil in place of thymine; structural differences between DNA and RNA are shown in Figure 1. Biologists initially believed that RNA was only contained in plant cells, whereas DNA was restricted to animal cells. However, this hypothesis dissipated as improved chemical methods discovered both nucleic acids in the cells of all life forms on Earth.

The primary structure of DNA and RNA is so similar because the former serves as a blueprint for the creation of a special kind of RNA molecule called messenger RNA, or mRNA. mRNA is created during RNA transcription, during which a strand of DNA is used as a template for constructing a strand of RNA by copying nucleotides one at a time, where uracil is used in place of thymine.

In eukaryotes, DNA remains in the nucleus, while RNA can enter the far reaches of the cell to carry out DNA's instructions. In future problems, we will examine the process and ramifications of RNA transcription in more detail.

Example

An RNA string is a string formed from the alphabet containing 'A', 'C', 'G', and 'U'.

Given a DNA string t

corresponding to a coding strand, its transcribed RNA string u is formed by replacing all occurrences of 'T' in t with 'U' in u.

Given: A DNA string t having length at most 1000 nt.

Return: The transcribed RNA string of t.

Sample Dataset

GATGGAACTTGACTACGTAAATT

Sample Output

GAUGGAACUUGACUACGUAAAUU

The Secondary and Tertiary Structures of DNAclick to collapse

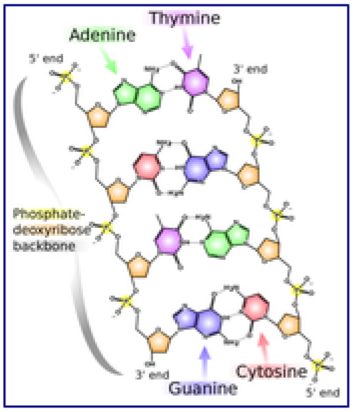

Figure 1. Base pairing across the two strands of DNA.

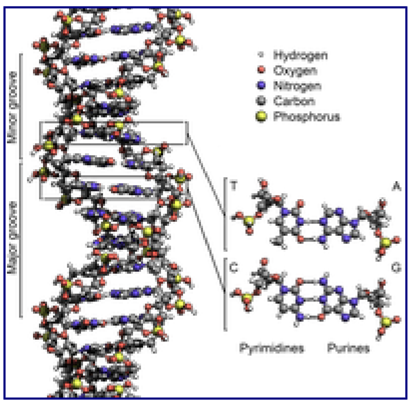

Figure 2. The double helix of DNA on the molecular scale.

In “Counting DNA Nucleotides”, we introduced nucleic acids, and we saw that the primary structure of a nucleic acid is determined by the ordering of its nucleobases along the sugar-phosphate backbone that constitutes the bonds of the nucleic acid polymer. Yet primary structure tells us nothing about the larger, 3-dimensional shape of the molecule, which is vital for a complete understanding of nucleic acids.

The search for a complete chemical structure of nucleic acids was central to molecular biology research in the mid-20th Century, culminating in 1953 with a publication in *Nature* of fewer than 800 words by James Watson and Francis Crick.

Consolidating a high resolution X-ray image created by Rosalind Franklin and Raymond Gosling with a number of established chemical results, Watson and Crick proposed the following structure for DNA:

1. The DNA molecule is made up of two strands, running in opposite directions.
2. Each base bonds to a base in the opposite strand. Adenine always bonds with thymine, and cytosine always bonds with guanine; the complement of a base is the base to which it always bonds; see Figure 1.
3. The two strands are twisted together into a long spiral staircase structure called a double helix; see Figure 2.

Because they dictate how bases from different strands interact with each other, (1) and (2) above compose the secondary structure of DNA. (3) describes the 3-dimensional shape of the DNA molecule, or its tertiary structure.

In light of Watson and Crick's model, the bonding of two complementary bases is called a base pair (bp). Therefore, the length of a DNA molecule will commonly be given in bp instead of nt. By complementarity, once we know the order of bases on one strand, we can immediately deduce the sequence of bases in the complementary strand. These bases will run in the opposite order to match the fact that the two strands of DNA run in opposite directions.

Example

In DNA strings, symbols 'A' and 'T' are complements of each other, as are 'C' and 'G'.

The reverse complement of a DNA string *s*

is the string *s*c formed by reversing the symbols of *s*

, then taking the complement of each symbol (e.g., the reverse complement of "GTCA" is "TGAC").

Given: A DNA string *s* of length at most 1000 bp.

Return: The reverse complement *s*c of *s*.

Sample Dataset

AAAACCCGGT

Sample Output

ACCGGGTTTT