Introduction

Molecular biology is the story of the molecules of life, their relationships, and how these interactions are controlled. Its applications are wide and growing; the power of molecular biology can now be harnessed to treat diseases, solve crimes, map human history, and produce genetically modified organisms and crops. Starting with the building blocks established by Darwin, Wallace, and Mendel, and the discovery of the structure of DNA in 1953, *Molecular Biology: A Very Short Introduction* considers the wide range of applications for molecular biology today, including the development of new drugs and DNA fingerprinting, and looks forward to two key areas of evolving research: personalized medicine and synthetic biology.

Molecular biology, field of science concerned with studying the chemical structures and processes of biological phenomena that involve the basic units of life, molecules. The field of molecular biology is focused especially on nucleic acids (e.g., DNA and RNA) and proteins—macromolecules that are essential to life processes—and how these molecules interact and behave within cells. Molecular biology emerged in the 1930s, having developed out of the related fields of biochemistry, genetics, and biophysics; today it remains closely associated with those fields.

Various techniques have been developed for molecular biology, though researchers in the field may also employ methods and techniques native to genetics and other closely associated fields. In particular, molecular biology seeks to understand the three-dimensional structure of biological macromolecules through techniques such as X-ray diffraction and electron microscopy. The discipline particularly seeks to understand the molecular basis of genetic processes; molecular biologists map the location of genes on specific chromosomes, associate these genes with particular characters of an organism, and use genetic engineering (recombinant DNA technology) to isolate, sequence, and modify specific genes. These approaches can also include techniques such as polymerase chain reaction, western blotting, and microarray analysis.

Structure of nucleic acids

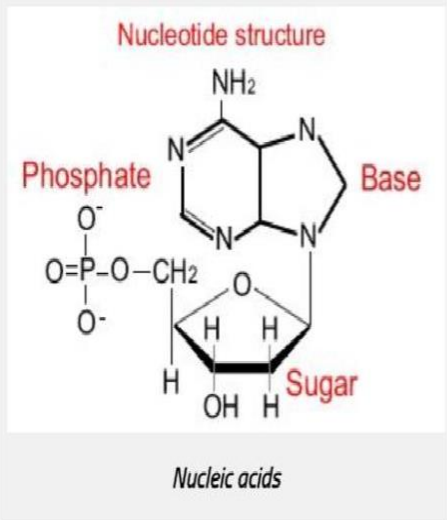
Nucleic acid, naturally occurring chemical compound that serves as the main information- carrying molecule of the cell and that directs the process of protein synthesis, thereby determining the inherited characteristics of every living thing. Nucleic acids are further defined by their ability to be broken down to yield phosphoric acid, sugars, and a mixture of organic bases (purines and pyrimidines). The two main classes of nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is the master blueprint for life and constitutes the genetic material in all free-living organisms and most viruses. RNA is the genetic material of certain viruses, but it is also found in all living cells, where it plays an important role in certain processes, such as the making of proteins.

Nucleotides: building blocks of nucleic acids

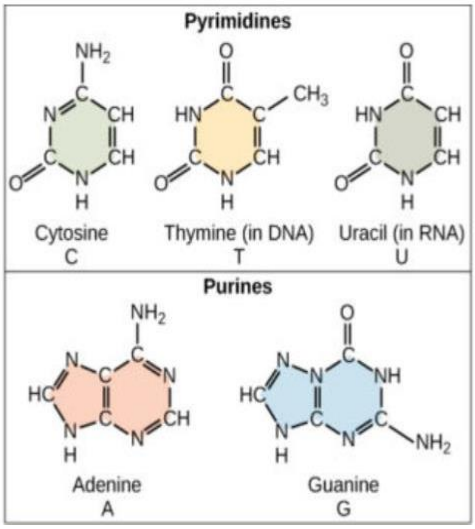
Basic structure

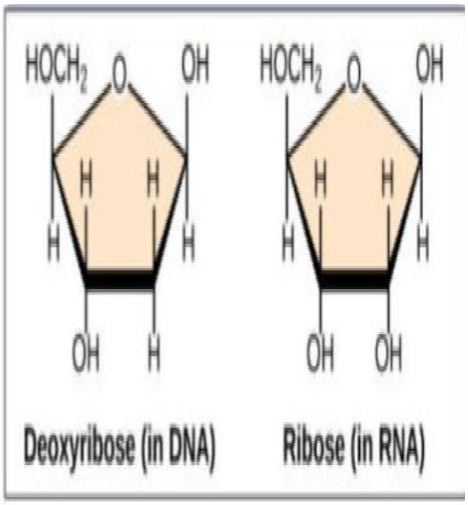
Nucleic acids are polynucleotides—that is, long chainlike molecules composed of a series of nearly identical building blocks called nucleotides. Each nucleotide consists of a nitrogen base attached to a pentose (five-carbon) sugar, which is in turn attached to a phosphate group. Each nucleic acid contains four of five possible nitrogen- bases: adenine (A), guanine (G), cytosine (C), thymine (T), and uracil (U). A and G are categorized as purines, and C, T, and U are collectively called pyrimidines. All nucleic acids contain the bases A, C, and G; T, however, is found only in DNA, while U is found in RNA.

The pentose sugar in DNA (2′-deoxyribose) differs from the sugar in RNA (ribose) by the absence of a hydroxyl group (—OH) on the 2′ carbon of the sugar ring. Without an attached phosphate group, the sugar attached to one of the bases is known as a nucleoside. The phosphate group connects successive sugar residues by bridging the 5′-hydroxyl group on one sugar to the 3′-hydroxyl group of the next sugar in the chain. These nucleoside linkages are called phosphodiester bonds and are the same in RNA and DNA.



Nucleic acids





Deoxyribonucleic acid (DNA)

DNA is a polymer of the four nucleotides A, C, G, and T, which are joined through a backbone of alternating phosphate and deoxyribose sugar residues. These nitrogen-containing bases occur in complementary pairs as determined by their ability to form hydrogen bonds between them. A always pairs with T through two hydrogen bonds, and G always pairs with C through three hydrogen bonds.

Chemical structure

In 1953 James D. Watson and Francis H.C. Crick proposed a three-dimensional structure for DNA based on low-resolution X-ray crystallographic data collected by biophysicists Rosalind Franklin and Maurice Wilkins and on Erwin Chargaff's observation that, in naturally occurring DNA, the amount of T equals the amount of A and the amount of G equals the amount of C. Watson and Crick, who shared a Nobel Prize in 1962 for their efforts, postulated that two strands of polynucleotides coil around each other, forming a double helix. The two strands, though identical, run in opposite directions as determined by the orientation of the 5′ to 3′ phosphodiester bond. The double helical structure of normal DNA takes a right-handed form called the B-helix. The helix makes one complete turn approximately every 10 base pairs. B-DNA has two principal grooves, a wide major groove and a narrow minor groove.

Biological structures

Naturally occurring DNA molecules can be circular or linear. The genomes of single- celled bacteria and archaea (collectively, the prokaryotes), as well as the genomes of mitochondria and chloroplasts (certain functional structures within the cell), are circular molecules. In addition, some bacteria and archaea have smaller circular DNA molecules called plasmids that typically contain only a few genes. Many plasmids are readily transmitted from one cell to another. For a typical bacterium, the genome that encodes all of the genes of the organism is a single contiguous circular molecule that contains a half million to five million base pairs. The genomes of most eukaryotes and some prokaryotes contain linear DNA molecules

called chromosomes. Human DNA, for example, consists of 23 pairs of linear chromosomes containing three billion base pairs.

In all cells, DNA does not exist free in solution but rather as a protein-coated complex called **chromatin**. In prokaryotes, the loose coat of proteins on the DNA helps to shield the negative charge of the phosphodiester backbone. Chromatin also contains proteins that control gene expression and determine the characteristic shapes of chromosomes. In eukaryotes, a section of DNA between 140 and 200 base pairs long winds around a discrete set of eight positively charged proteins called a histone, forming a spherical structure called the **nucleosome**

Most viruses contain linear genomes that typically are much shorter and contain only the genes necessary for viral propagation. Bacterial viruses called bacteriophages (or phages) may contain both linear and circular forms of DNA.

Ribonucleic acid (RNA)

RNA is a single-stranded nucleic acid polymer of the four nucleotides A, C, G, and U joined through a backbone of alternating phosphate and ribose sugar residues. It is the first intermediate in converting the information from DNA into proteins essential for the working of a cell. Some RNAs also serve direct roles in cellular metabolism. RNA is made by copying the base sequence of a section of double-stranded DNA, called a gene, into a piece of single-stranded nucleic acid. This process, known as transcription is catalyzed by an enzyme

called RNA polymerase.

Chemical structure

Whereas DNA provides the genetic information for the cell and is inherently quite stable, RNA has many roles and is much more reactive chemically. The 2′-hydroxyl group on the ribose ring is a major cause of instability in RNA, because the presence of alkali leads to rapid cleavage of the phosphodiester bond linking ribose and phosphate groups. In general, this instability is not a significant problem for the cell, because RNA is constantly being synthesized and degraded.

Interactions between the nitrogen-containing bases differ in DNA and RNA. In DNA, which is usually double-stranded, the bases in one strand pair with complementary bases in a second DNA

strand. In RNA, which is usually single-stranded, the bases pair with other bases within the same molecule, leading to complex three-dimensional structures. Occasionally, intermolecular RNA/RNA duplexes do form, but they form a right-handed A-type helix rather than the B-type DNA helix.

Types of RNA

1- Messenger RNA (mRNA)

Messenger RNA (mRNA) delivers the information encoded in one or more genes from the DNA to the ribosome, a specialized structure, or organelle, where that information is decoded into a protein. In prokaryotes, mRNAs contain an exact transcribed copy of the original DNA sequence with a terminal 5′-triphosphate group and a 3′-hydroxyl residue. In eukaryotes the mRNA molecules are more elaborate. The 5′-triphosphate residue is further esterified, forming a structure called a cap. At the 3′ ends, eukaryotic mRNAs typically contain long runs of adenosine residues (polyA) that are not encoded in the DNA but are added enzymatically after transcription.

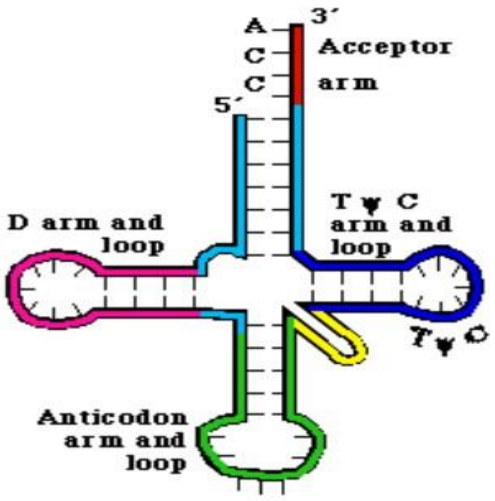
Eukaryotic mRNA molecules are usually composed of small segments of the original gene and are generated by a process of cleavage and rejoining from an original precursor RNA (pre-mRNA) molecule, which is an exact copy of the gene (as described in the section Splicing). In general, prokaryotic mRNAs are degraded very rapidly, whereas the cap structure and the polyA tail of eukaryotic mRNAs greatly enhance their stability.

Ribosomal RNA (rRNA)

Ribosomal RNA (rRNA) molecules are the structural components of the ribosome. The rRNAs form extensive secondary structures and play an active role in recognizing conserved portions of mRNAs and tRNAs. They also assist with the catalysis of protein synthesis. In the prokaryote *E. coli*, there are anywhere from one to 15 copies of the rRNA operon (an operon is a group of protein-coding structural genes in bacteria). In eukaryotes the numbers are much larger. Anywhere from 50 to 5,000 sets of rRNA genes and as many as 10 million ribosomes may be present in a single cell.

Transfer RNA (tRNA)

Transfer RNA (tRNA) carries individual amino acids into the ribosome for assembly into the growing polypeptide chain. The tRNA molecules contain 70 to 80 nucleotides and fold into a characteristic cloverleaf structure.



Secondary structure of tRNA.

Specialized tRNAs exist for each of the 20 amino acids needed for protein synthesis, and in many cases more than one tRNA for each amino acid is present. The nucleotide sequence is converted into a protein sequence by translating each three-base sequence (called a codon) with a specific protein. The 61 codons used to code amino acids can be read by many fewer than 61 distinct tRNAs.

Viral genomes

Many viruses use RNA for their genetic material. This is most prevalent among eukaryotic viruses, but a few prokaryotic RNA viruses are also known. Some common examples include polio, HIV, and influenza A H1N1,

all of which affect humans, and tobacco mosaic virus, which infects plants. In some viruses the entire genetic material is encoded in a single RNA molecule, while in the segmented RNA viruses several RNA molecules may be present. Many RNA viruses such as HIV use a specialized enzyme called reverse transcriptase that permits

replication of the virus through a DNA intermediate. In some cases this DNA intermediate becomes integrated into the host chromosome during infection; the virus then exists in a dormant state and effectively evades the host immune system.

RNA processing

Cleavage

Following synthesis by transcription, most RNA molecules are processed before reaching their final form. Many rRNA molecules are cleaved from much larger transcripts and may also be methylated or enzymatically modified. In addition, tRNAs are usually formed as longer precursor molecules that are cleaved by ribonuclease P to generate the mature 5′ end and often have extra residues added to their 3′ end to form the sequence CCA.

Splicing

The segments of DNA or RNA coding for protein are called **exons**, and the noncoding regions separating the exons are called **introns**. Following transcription, these coding sequences must be joined together before the mRNAs can function. The process of removal of the introns and subsequent rejoining of the exons is called **RNA splicing**. Each intron is removed in a separate series of reactions by a complicated piece of

enzymatic machinery called a spliceosome.