Nucleic Acids

Dr.H.B.Mahesha, Yuvaraja's College, University of Mysore, Mysuru.

Nucleic acids are biomolecules essential for life, and include DNA (deoxyribonucleic acid) and RNA (ribonucleic acid). Together with proteins, nucleic acids make up the most important macromolecules; each is found in abundance in all living things, where they function in encoding, transmitting and expressing genetic information. Nucleic acids were first discovered by Friedrich Miescher in 1871.

The term *nucleic acid* is the overall name for DNA and RNA, members of a family of biopolymers, and is synonymous with *polynucleotide*. Nucleic acids were named for their initial discovery within the cell nucleus, and for the presence of phosphate groups (related to phosphoric acid). Although first discovered within the nucleus of eukaryotic cells, nucleic acids are now known to be found in all life forms, including Bacteria, Archaea, Mitochondria, Chloroplasts, Viruses and Viroids. All living cells and organelles contain both DNA and RNA, while viruses contain either DNA or RNA, but usually not both. The basic component of biological nucleic acids is the nucleotide, each of which contains a pentose sugar (ribose or deoxyribose), a phosphate group, and a nucleobase. Nucleic acids are linear polymers (chains) of nucleotides. Each nucleotide consists of three components: a purine or pyrimidine nucleobase (sometimes termed *nitrogenous base* or simply *base*), a pentose sugar, and a phosphate group. The substructure consisting of a nucleobase plus sugar is termed a nucleoside. Nucleic acid types differ in the structure of the sugar in their nucleotides - DNA contains 2'-deoxyribose while RNA contains ribose (where the only difference is the presence of a hydroxyl group). Also, the nucleobases found in the two nucleic acid types are different: adenine, cytosine, and guanine are found in both RNA and DNA, while thymine occurs in DNA and uracil occurs in RNA.

The sugars and phosphates in nucleic acids are connected to each other in an alternating chain (sugar-phosphate backbone) through phosphodiester linkages. In conventional nomenclature, the carbons to which the phosphate groups attach are the 3'-end and the 5'-end carbons of the sugar. This gives nucleic acids directionality, and the ends of nucleic acid molecules are referred to as 5'-end and 3'-end. The nucleobases are joined to the sugars via an N-glycosidic linkage involving a nucleobase ring nitrogen (N-1 for pyrimidines and N-9 for purines) and the 1' carbon of the pentose sugar ring.

Non-standard nucleosides are also found in both RNA and DNA and usually arise from modification of the standard nucleosides within the DNA molecule or the primary (initial) RNA transcript. Transfer RNA (tRNA) molecules contain a particularly large number of modified nucleosides.

Double-stranded nucleic acids are made up of complementary sequences, in which extensive Watson-Crick base pairing results in a highly repeated and quite uniform double-helical three-

dimensional structure. In contrast, single-stranded RNA and DNA molecules are not constrained to a regular double helix, and can adopt highly complex three-dimensional structures that are based on short stretches of intramolecular base-paired sequences that include both Watson-Crick and noncanonical base pairs, as well as a wide range of complex tertiary interactions.

Nucleic acid molecules are usually unbranched, and may occur as linear and circular molecules. For example, bacterial chromosomes, plasmids, mitochondrial DNA and chloroplast DNA are usually circular double-stranded DNA molecules, while chromosomes of the eukaryotic nucleus are usually linear double-stranded DNA molecules. Most RNA molecules are linear, single-stranded molecules, but both circular and branched molecules can result from RNA splicing reactions.

Nucleic acid sequences

One DNA or RNA molecule differs from another primarily in the sequence of nucleotides. Nucleotide sequences are of great importance in biology, since they carry the ultimate instructions that encode all biological molecules, molecular assemblies, sub cellular and cellular structures, organs and organisms, and directly enable cognition, memory and behavior. Enormous efforts have gone into the development of experimental methods to determine the nucleotide sequence of biological DNA and RNA molecules, and today hundreds of millions of nucleotides are sequenced daily at genome centers and smaller laboratories worldwide.

Types of nucleic acids

Deoxyribonucleic acid-DNA

Deoxyribonucleic acid is a nucleic acid that contains the genetic instructions used in the development and functioning of all known living organisms. The main role of DNA molecules is the long-term storage of information and DNA is often compared to a set of blueprints, since it contains the instructions needed to construct other components of cells, such as proteins and RNA molecules. The DNA segments that carry this genetic information are called genes, but other DNA sequences have structural purposes, or are involved in regulating the use of this genetic information.

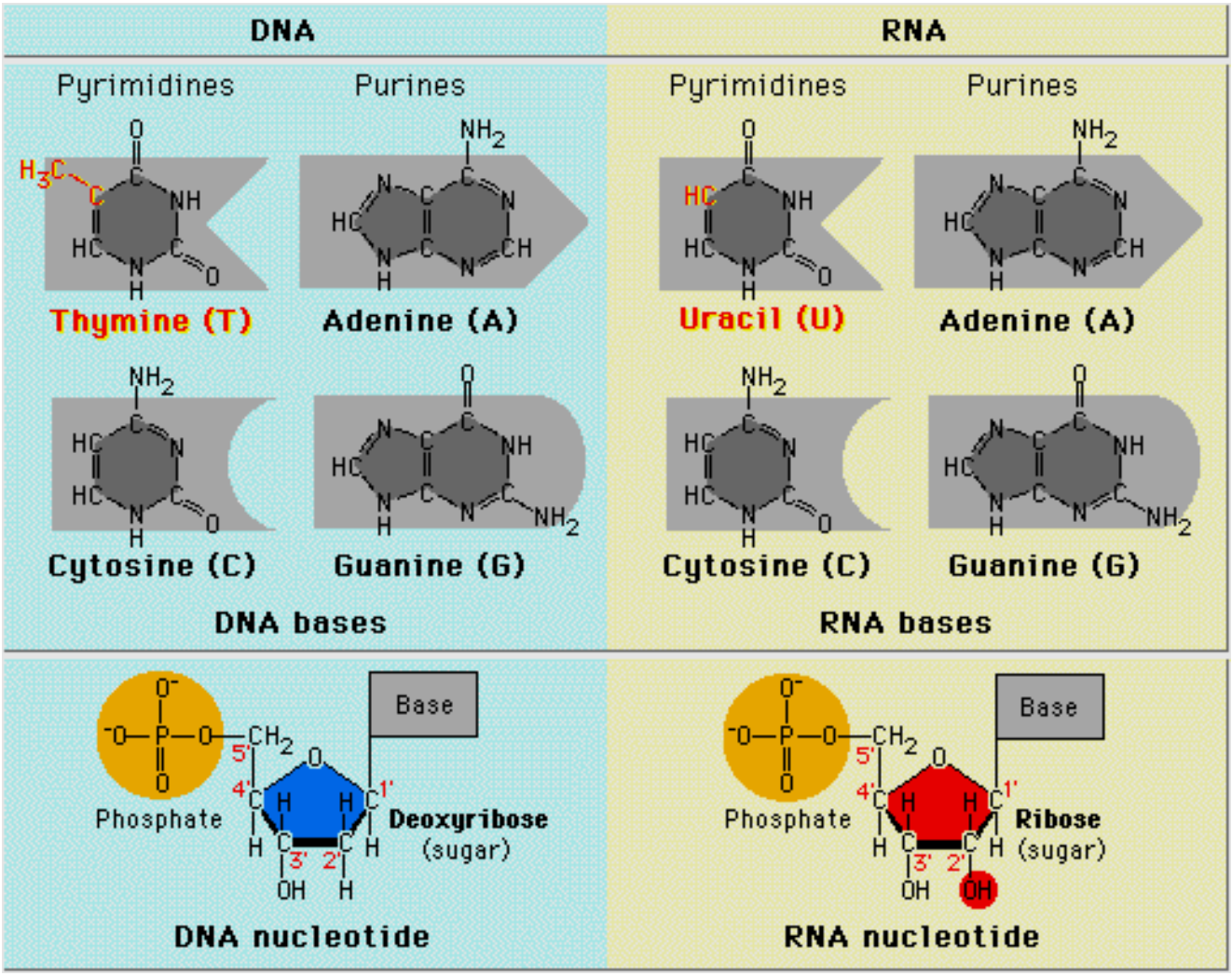
Ribonucleic acid-RNA

Ribonucleic acid (RNA) functions in converting genetic information from genes into the amino acid sequences of proteins. The three universal types of RNA include transfer RNA (tRNA), messenger RNA (mRNA), and ribosomal RNA (rRNA). Messenger RNA acts to carry genetic sequence information between DNA and ribosomes, directing protein synthesis. Ribosomal RNA is a major component of the ribosome, and catalyzes peptide bond formation. Transfer RNA serves as the carrier molecule for amino acids to be used in protein synthesis, and is responsible for decoding the mRNA. In addition, many other classes of RNA are now known.

Nucleosides are glycosylamines consisting of a nucleobase (often referred to as simply *base*) bound to a ribose or deoxyribose sugar via a beta-glycosidic linkage. Examples of nucleosides include cytidine,

uridine, adenosine, guanosine, thymidine and inosine. They are the molecular building-blocks of DNA and RNA.

The structure elements of the nucleosides and the phosphate group bearing nucleotides are as follows



Structure of DNA

DNA is a long polymer made from repeating units called nucleotides. As first described by James D. Watson and Francis Crick, the structure of DNA of all species comprises two helical chains each coiled round the same axis, and each with a pitch of 34 Ångströms (3.4 nanometres) and a radius of 10 Ångströms (1.0 nanometres). DNA polymers can be very large molecules containing millions of nucleotides. For instance, the largest human chromosome, chromosome number 1, is approximately 220 million base pairs long.

In living organisms DNA does not usually exist as a single molecule, but instead as a pair of molecules that are held tightly together. These two long strands twist together like vines, in the shape of a double helix. The nucleotide repeats contain both the segment of the backbone of the molecule, which holds the chain together, and a nucleobase, which interacts with the other DNA strand in the helix. A nucleobase linked to a sugar is called a nucleoside and a base linked to a sugar and one or

more phosphate groups is called a nucleotide. Polymers comprising multiple linked nucleotides (as in DNA) is called a polynucleotide.

The backbone of the DNA strand is made from alternating phosphate and sugar residues. The sugar in DNA is 2-deoxyribose, which is a pentose (five-carbon) sugar. The sugars are joined together by phosphate groups that form phosphodiester bonds between the third and fifth carbon atoms of adjacent sugar rings. These asymmetric bonds mean a strand of DNA has a direction. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are *antiparallel*. The asymmetric ends of DNA strands are called the 5′ (*five prime*) and 3′ (*three prime*) ends, with the 5' end having a terminal phosphate group and the 3' end a terminal hydroxyl group. One major difference between DNA and RNA is the sugar, with the 2-deoxyribose in DNA being replaced by the alternative pentose sugar ribose in RNA.

The DNA double helix is stabilized primarily by two forces that is hydrogen bonds between nucleotides and base-stacking interactions among the aromatic nucleobases. The four bases found in DNA are adenine (abbreviated A), cytosine (C), guanine (G) and thymine (T). These four bases are attached to the sugar/phosphate to form the complete nucleotide, as shown for adenosine monophosphate.

The nucleobases are classified into two types that i.e. the purines, A and G, being fused five- and six-membered heterocyclic compounds, and the pyrimidines, the six-membered rings C and T. A fifth pyrimidine nucleobase, uracil (U), usually takes the place of thymine in RNA and differs from thymine by lacking a methyl group on its ring. Uracil is not usually found in DNA, occurring only as a breakdown product of cytosine. In addition to RNA and DNA a large number of artificial nucleic acid analogues have also been created to study the proprieties of nucleic acids, or for use in biotechnology.

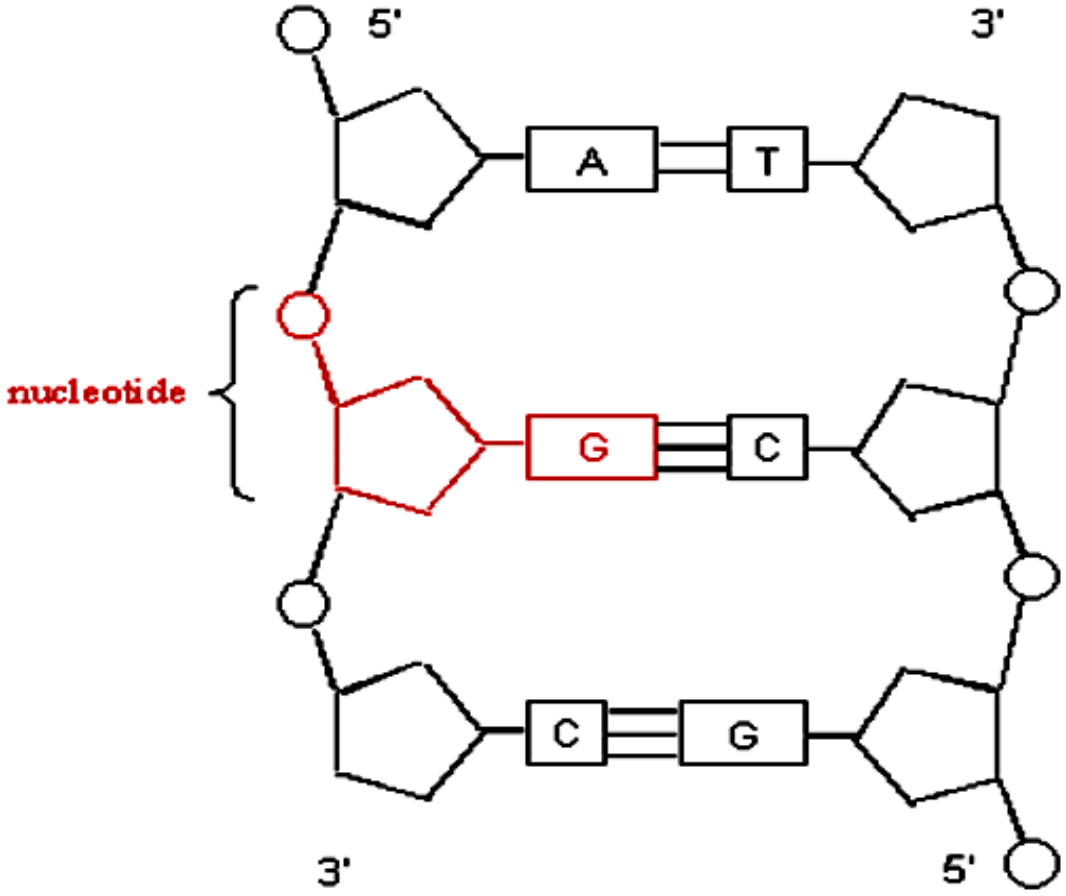
Grooves

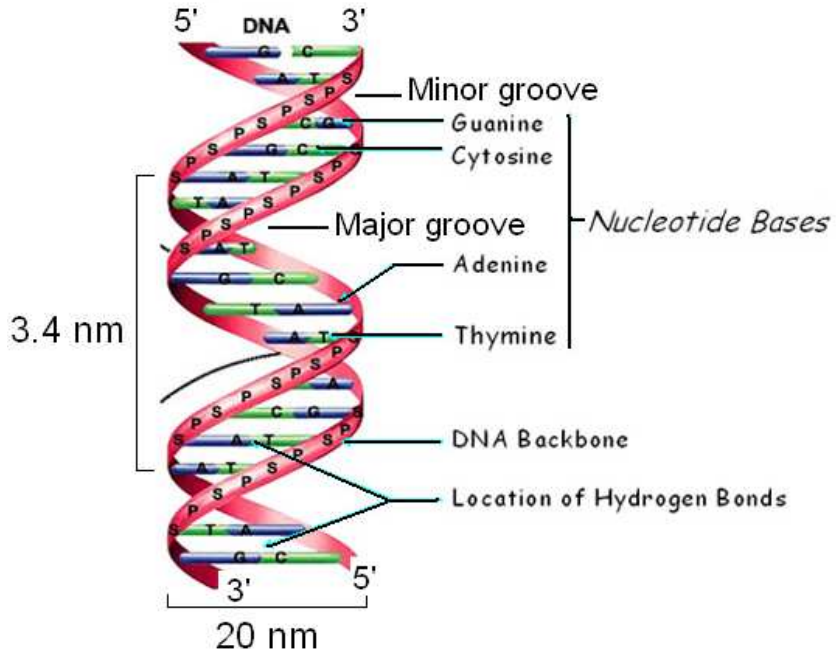
Twin helical strands form the DNA backbone. Another double helix may be found by tracing the spaces, or grooves, between the strands. These voids (empty space) are adjacent to the base pairs and may provide a binding site. As the strands are not directly opposite each other, the grooves are unequally sized. One groove, the major groove, is 22 Å wide and the other, the minor groove, is 12 Å wide. The narrowness of the minor groove means that the edges of the bases are more accessible in the major groove. As a result, proteins like transcription factors that can bind to specific sequences in double-stranded DNA usually make contacts to the sides of the bases exposed in the major groove. This situation varies in unusual conformations of DNA within the cell, but the major and minor grooves are always named to reflect the differences in size that would be seen if the DNA is twisted back into the ordinary B form.

Base pairing

In a DNA double helix, each type of nucleobase on one strand normally interacts with just one type of nucleobase on the other strand. This is called complementary base pairing. Here, purines form hydrogen bonds to pyrimidines, with A bonding only to T, and C bonding only to G. This arrangement of two nucleotides binding together across the double helix is called a base pair. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or high temperature. As a result of these complementarities, all the information in the double-stranded sequence of a DNA helix is duplicated on each strand, which is vital in DNA replication. Indeed, this reversible and specific interaction between complementary base pairs is critical for all the functions of DNA in living organisms.

The two types of base pairs form different numbers of hydrogen bonds, AT forming two hydrogen bonds, and GC forming three hydrogen bonds. DNA with high GC-content is more stable than DNA with low GC-content. Although it is often stated that this is due to the added stability of an additional hydrogen bond, this is incorrect. DNA with high GC-content is more stable due to intra-strand base stacking interactions.





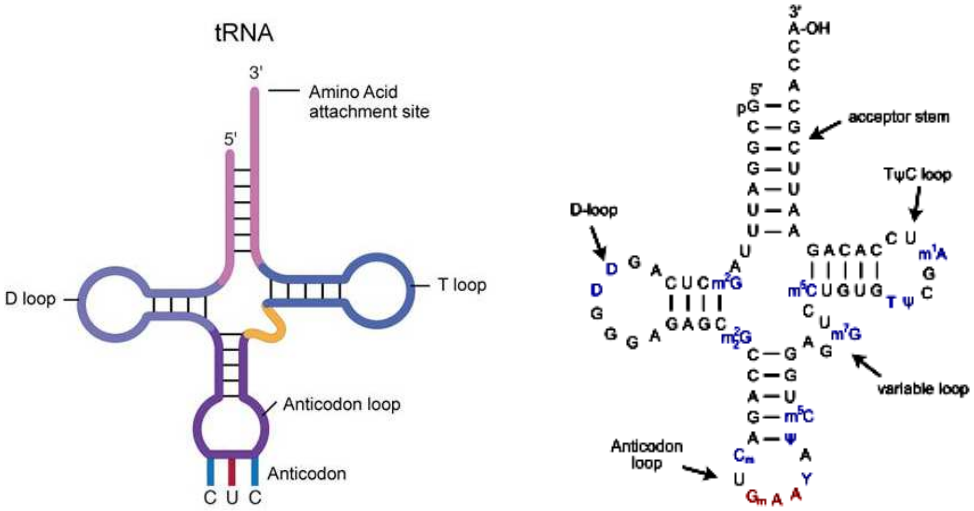
Structure of Transfer RNA

Transfer RNA (tRNA) is an adaptor molecule, typically 73 to 93 nucleotides in length, that is used in biology to bridge the three-letter genetic code in messenger RNA (mRNA) with the twenty-letter code of amino acids in proteins. The role of tRNA as an adaptor is best understood by considering its three-dimensional structure. One end of the tRNA carries the genetic code in a three-nucleotide sequence called the anticodon. The anticodon forms three base pairs with a codon in mRNA during protein biosynthesis. On the other end of its three-dimensional structure, each tRNA is covalently attached to the amino acid that corresponds to the anticodon sequence. Each type of tRNA molecule can be attached to only one type of amino acid, but because the genetic code contains multiple codons that specify the same amino acid, tRNA molecules bearing different anticodons may also carry the same amino acid.

Structure: Constitutes 10-12 % of total RNA of the cell with a molecular weight of 25- 30 kd and 0.025 % DNA codes for tRNA. The structure of tRNA can be decomposed into its primary structure, its secondary structure (usually visualized as the *cloverleaf structure*), and its tertiary structure (all tRNAs have a similar L-shaped 3D structure that allows them to fit into the P and A sites of the ribosome). The cloverleaf structure becomes the 3D L-shaped structure through coaxial stacking of the helices which is a common RNA Tertiary Structure motif.

1. The 5'-terminal phosphate group.
2. The acceptor stem is a 7-base pair (bp) stem made by the base pairing of the 5'-terminal nucleotide with the 3'-terminal nucleotide (which contains the CCA 3'-terminal group used to attach the amino acid). The acceptor stem may contain non-Watson-Crick base pairs.

3. The CCA tail is a cytosine-cytosine-adenine sequence at the 3' end of the tRNA molecule. This sequence is important for the recognition of tRNA by enzymes and critical in translation. In prokaryotes, the CCA sequence is transcribed in some tRNA sequences. In most prokaryotic tRNAs and eukaryotic tRNAs, the CCA sequence is added during processing and therefore does not appear in the tRNA gene.
4. The D arm is a 4 bp stem ending in a loop that often contains dihydrouridine.
5. The anticodon arm is a 5-bp stem whose loop contains the anticodon.
6. The T arm is a 5 bp stem containing the sequence TΨC where Ψ is a pseudouridine.
7. Bases that have been modified, especially by methylation, occur in several positions throughout the tRNA. The first anticodon base, or wobble-position, is sometimes modified to inosine (derived from adenine), pseudouridine (derived from uracil) or lysidine (derived from cytosine).



Messenger RNA (mRNA) carries the genetic information copied from DNA in the form of a series of three-base code "words" called codon, each of which specifies a particular amino acid during translation. mRNA constitutes 3-5 % of total RNA of the cell. The average molecular weight is about 500 kd. Prokaryotic and eukaryotic messenger RNAs (mRNAs) have different structures. Prokaryotic mRNAs are often polycistronic (that is, they carry the information for more than one protein) while eukaryotic mRNAs are monocistronic and almost always code for a single protein. Eukaryotic mRNAs also have structural features that prokaryotic ones do not. While prokaryotic mRNAs generally have only the common four bases A, C, G, and U.

Ribosomal RNA (rRNA) associates with a set of proteins to form ribosome. These complex structures, which physically move along an mRNA molecule, catalyze the assembly of amino acids into protein chains. They also bind tRNAs and various accessory molecules necessary for protein synthesis. Ribosomes are composed of a large and small subunit, each of which contains its own

rRNA molecule or molecules. This constitutes 80 % of total RNA of the cell 3.2 % DNA codes for rRNA.

Acknowledgements

Editor acknowledges all scientists who have pioneered and contributed information used in this article.
