Salient Features and chemical compositions of DNA and RNA

(Structure of Purines and Pyrimidines, Nucleosides, Nucleotides, Base pairing, Hoogsteen base pair & Basic structure and functions of different types of RNA)

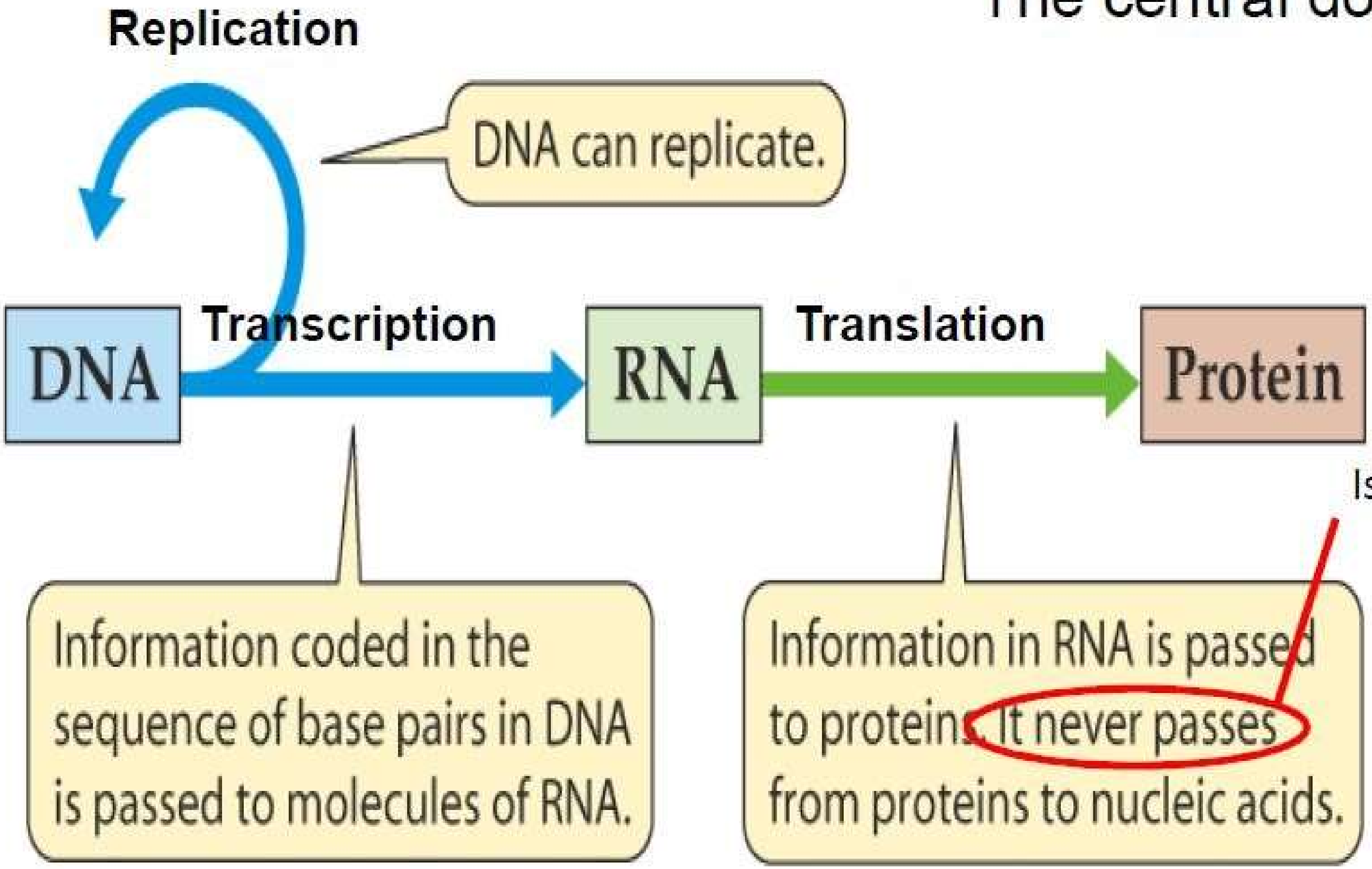
Dr. R. Prasad
Department of Zoology,
Eastern Karbi Anglong College,
Sarihajan

All Life depends on 3 critical molecules

- DNAs (Deoxyribonucleic acid)
  - Hold information on how cell works
- RNAs (Ribonucleic acid)
  - Act to transfer short pieces of information to different parts of cell
  - Provide templates to synthesize into protein
- Proteins
  - Form enzymes that send signals to other cells and regulate gene activity
  - Form body's major components

DNA, RNA, and the Flow of Information

"The central dogma"



DNA can replicate.

Information coded in the sequence of base pairs in DNA is passed to molecules of RNA.

Information in RNA is passed to proteins. It never passes from proteins to nucleic acids.

Is this true?

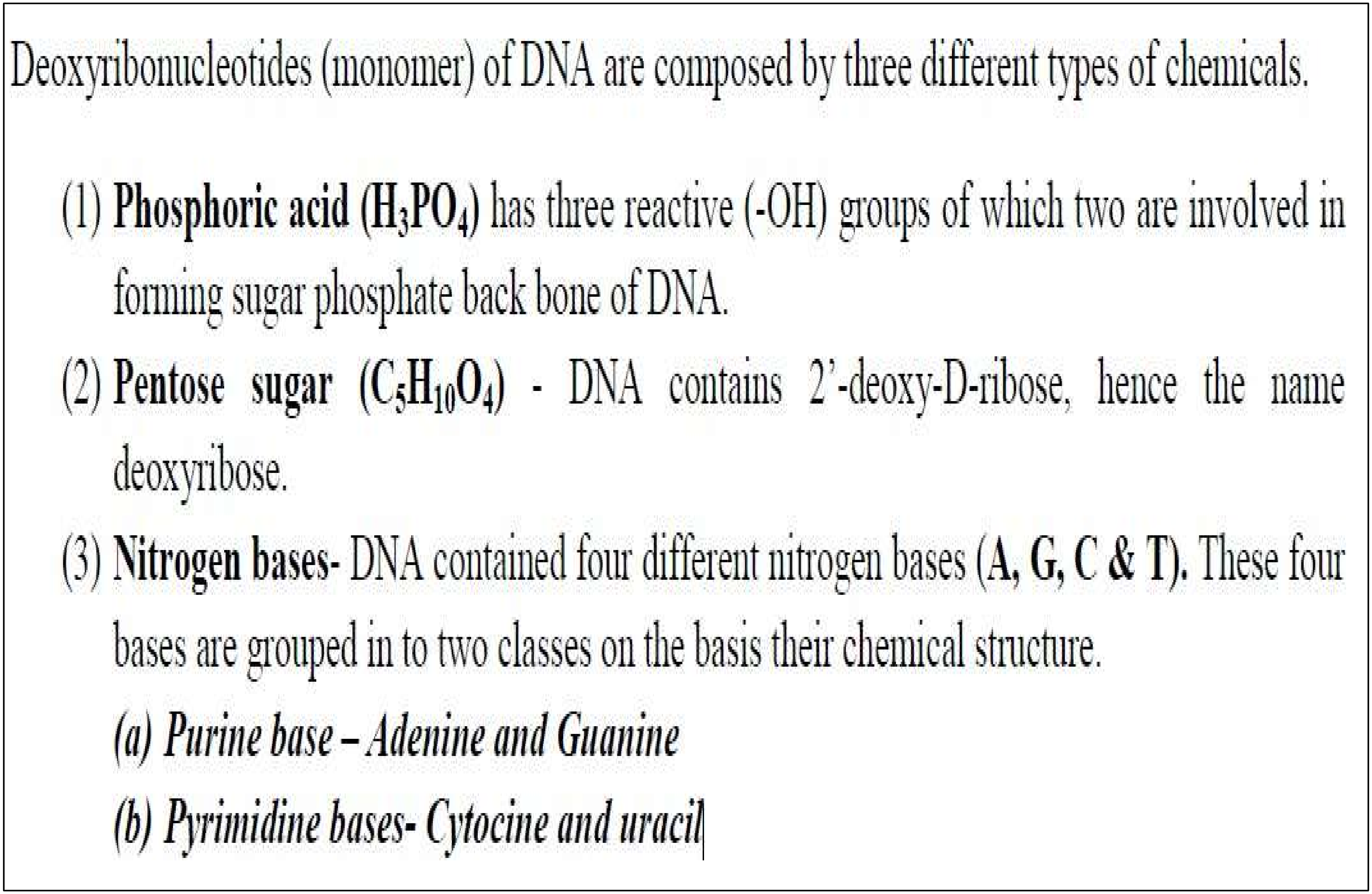
DNA

- Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) the principal **genetic materials** of living organisms are chemically called **nucleic acids**.
- Nucleic acid especially the DNA, a universal genetic material of most of the organisms, is having all the features required to be a good genetic materials.
- DNA is a macromolecule and is a helically twisted double chain of poly-deoxyribonucleotides.
- In **prokaryotes** it occurs in **nucleoid** and also as **plasmids**, both are **double stranded circular DNA**.
- In **Eukaryotes** most of the DNA is found in **chromatin of nucleus**. It is **linear**.
- Some small quantitative of DNA are found in **mitochondria and plastids** which is generally double stranded and circular.
- RNA also acts as genetic material in majority of plant viruses.

Features of DNA to act as genetic material:

- Genetic material is able to **store information** used to control both the development and metabolic activities of cell.
- It should be **chemically stable** so that it can be replicated accurately during cell division.
- It should be **transmitted for generations.**
- It should be able to undergo **mutations providing genetic variability** required for the evolution.

Chemical Composition of DNA



Deoxyribonucleotides (monomer) of DNA are composed by three different types of chemicals.

(1) **Phosphoric acid (H_3PO_4)** has three reactive (-OH) groups of which two are involved in forming sugar phosphate back bone of DNA.

(2) **Pentose sugar ($C_5H_{10}O_4$)** - DNA contains 2'-deoxy-D-ribose, hence the name deoxyribose.

(3) **Nitrogen bases-** DNA contained four different nitrogen bases (**A, G,** C **& T**)**.** These four bases are grouped in to two classes on the basis their chemical structure.

(a) Purine base – Adenine and Guanine

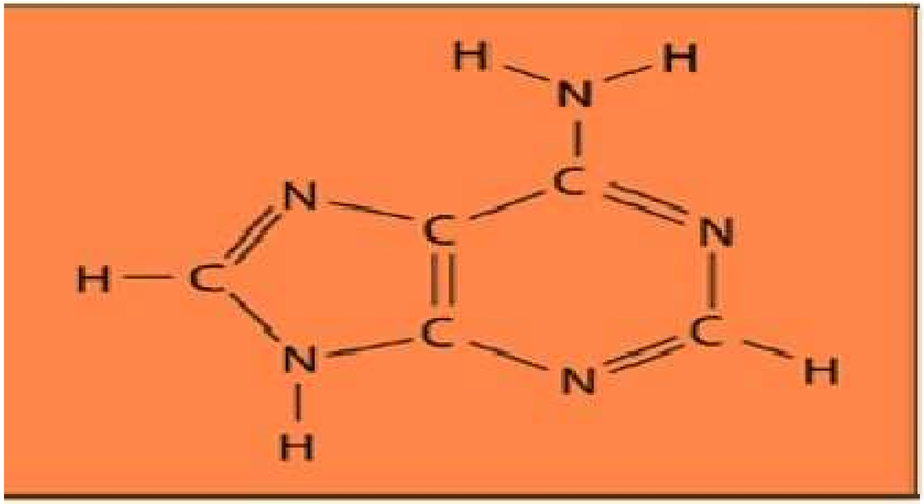
(b) Pyrimidine bases- Cytocine and uracil

(a) **Purine bases -** DNA has two types of purines (**adenine and guanine**). Each purine is a type of nitrogen base having a **double ring structure** (i.e. 9 member double rings with nitrogen at 1, 3, 7 and 9 positions).

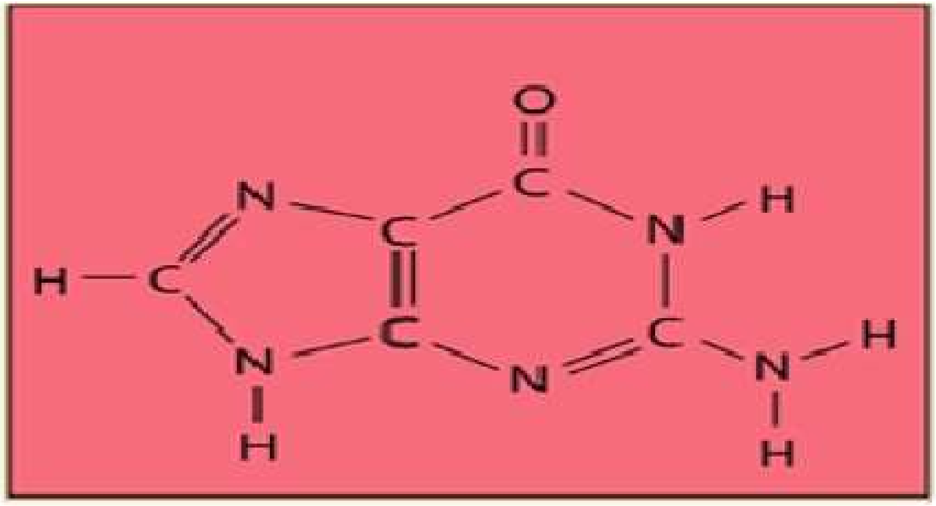
Some of the common names of these bases reflect the circumstances of their discovery. Guanine, for example, was first isolated from guano (bird manure), and thymine was first isolated from thymus tissue.

(b) **Pyrimidine bases-** DNA has two types of pyrimidine bases (**cytosine and thymine**). Each pyrimidine is a type of nitrogen containing base having a **single ring structure** (i.e. 6 member rings with nitrogen at 1 and 3 positions).

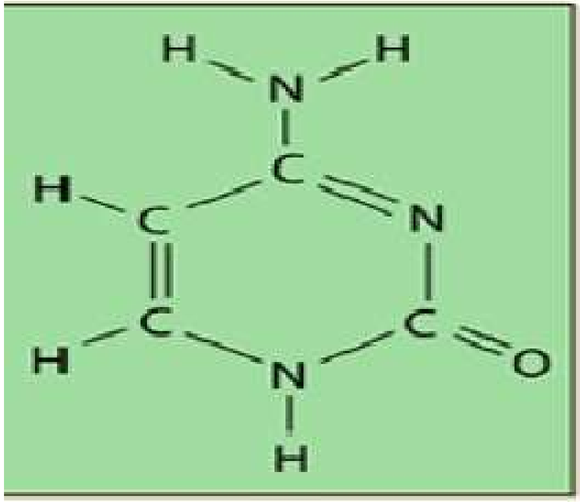
Nitrogen bases of nucleic acids



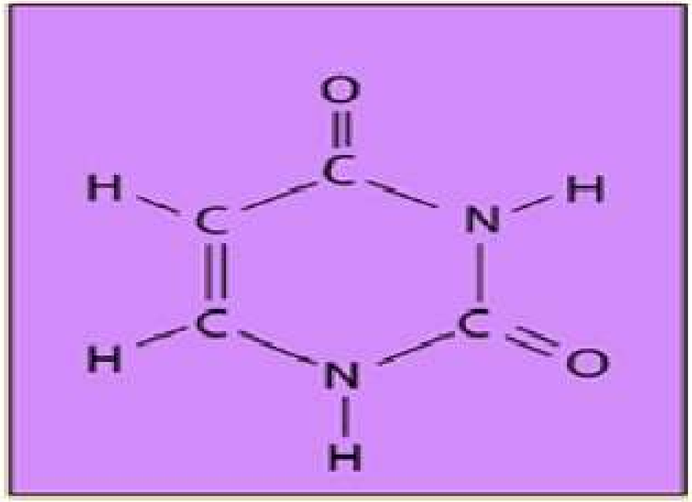
Adenine



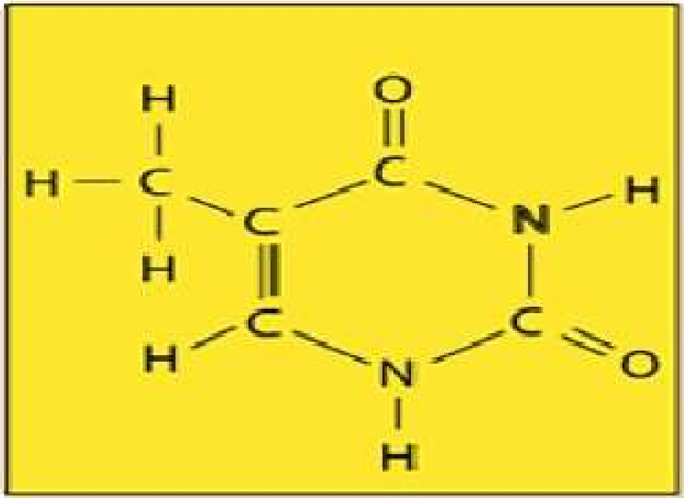
Guanine



Cytosine



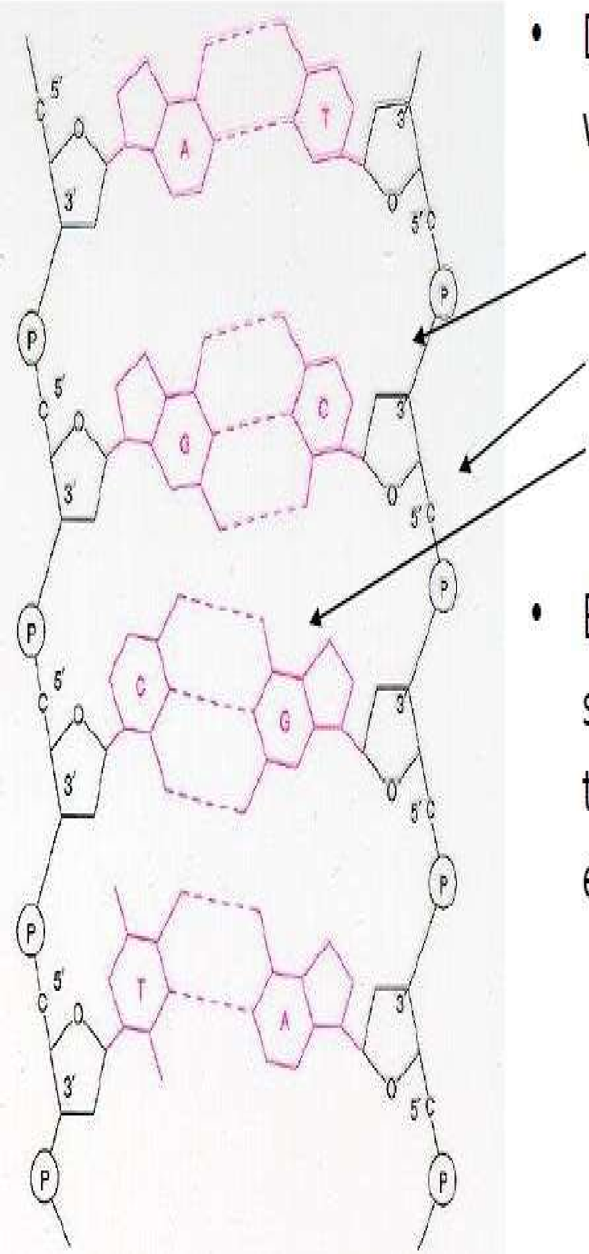
Uracil



Thymine

(**A, G and C** is common to DNA and RNA, **U** is present in RNA and **T** in DNA)

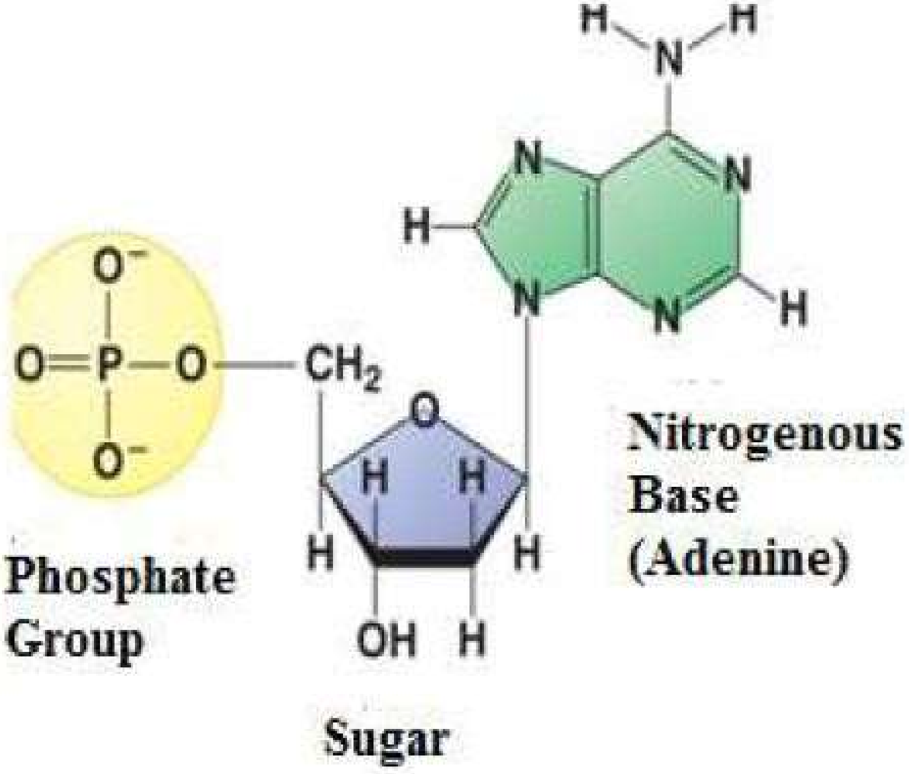
DNA structure

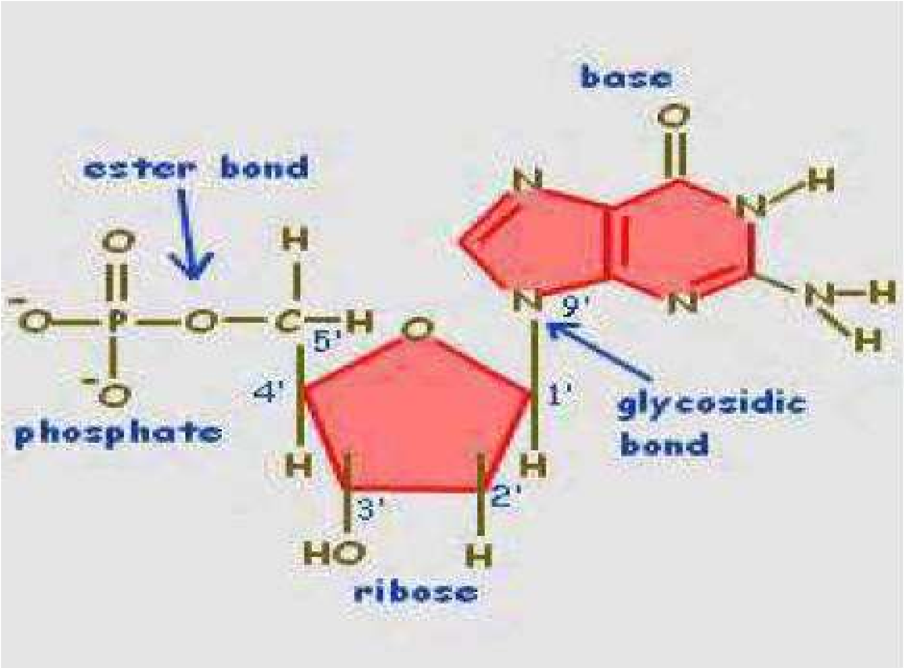


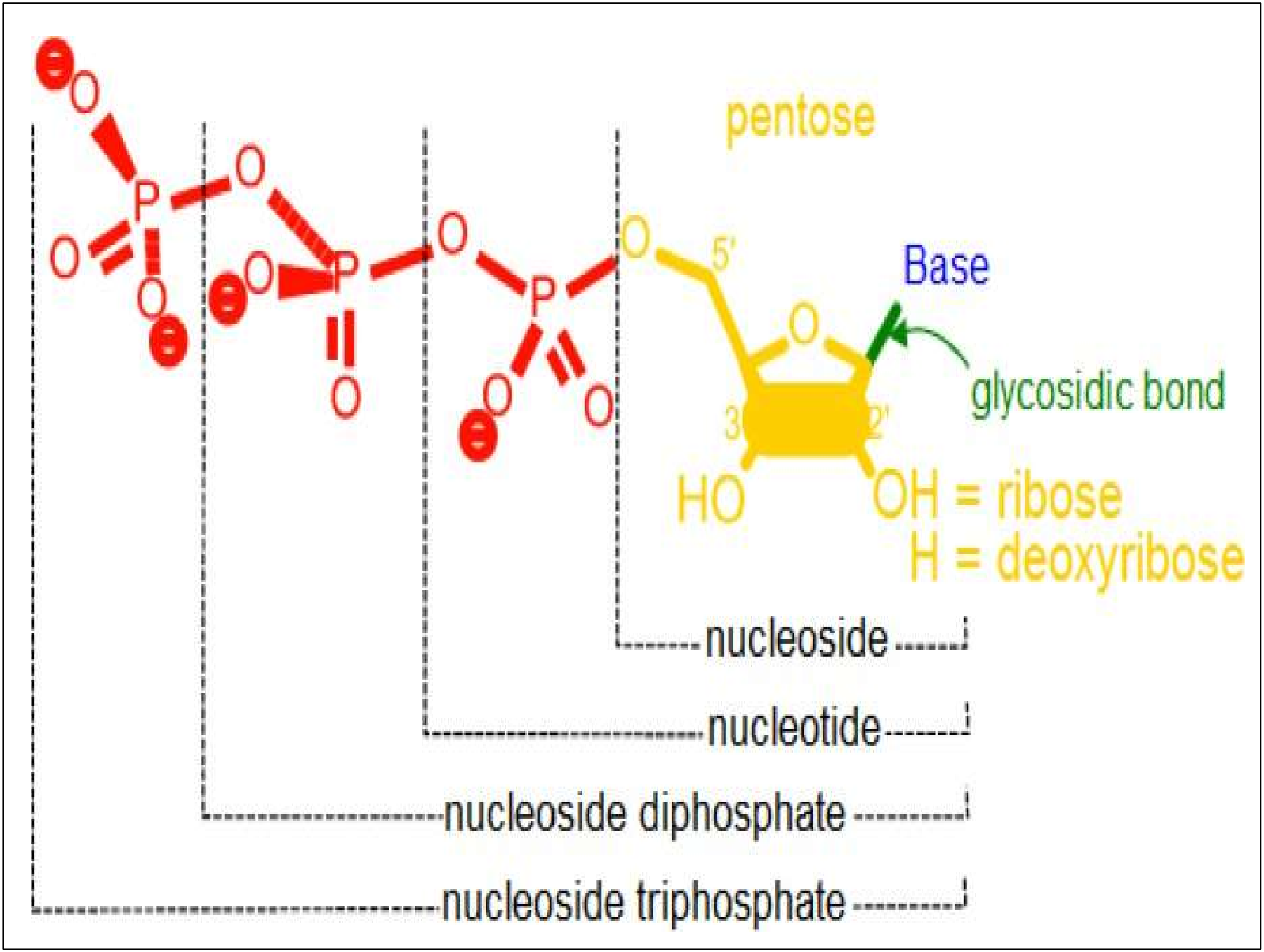
- DNA has a double helix structure which is composed of
  - sugar molecule
  - phosphate group
  - and a base (A,C,G,T)
- By convention, we read DNA strings in direction of transcription: from 5′ end to 3′ end

5′ ATTTAGGCC 3′

3′ TAAATCCGG 5′






pentose
5'
Base
glycosidic bond
3'
2'
HO
OH = ribose
H = deoxyribose
nucleoside
nucleotide
nucleoside diphosphate
nucleoside triphosphate

Nucleosides- A nitrogenous base with a molecule of deoxyribose sugar (without phosphate group) is known as nucleosides. In nucleic acids, the nitrogen bases are covalently attached to the 1’-position of a pentose sugar ring with the help of glycosidic bond.

Nitrogen base + sugar = nucleoside.

- Adenine + deoxyribose = deoxyadenosine
- Guanine + deoxyribose = deoxyguanosine
- Cytosine + deoxyribose = deoxycytidine
- Thymine + deoxyribose = deoxythymidine

Nucleotides- A nucleotide is formed of one molecule of deoxyribose sugar, one molecule of phosphoric acid and anyone of the nitrogen base. Phosphoric molecule is attached to the 5th – carbon atom of deoxyribose ring with the help of phosphoesterbond.

Nucleosides + phosphoric acid = nucleotides

Different nucleotides of DNA are as follows:

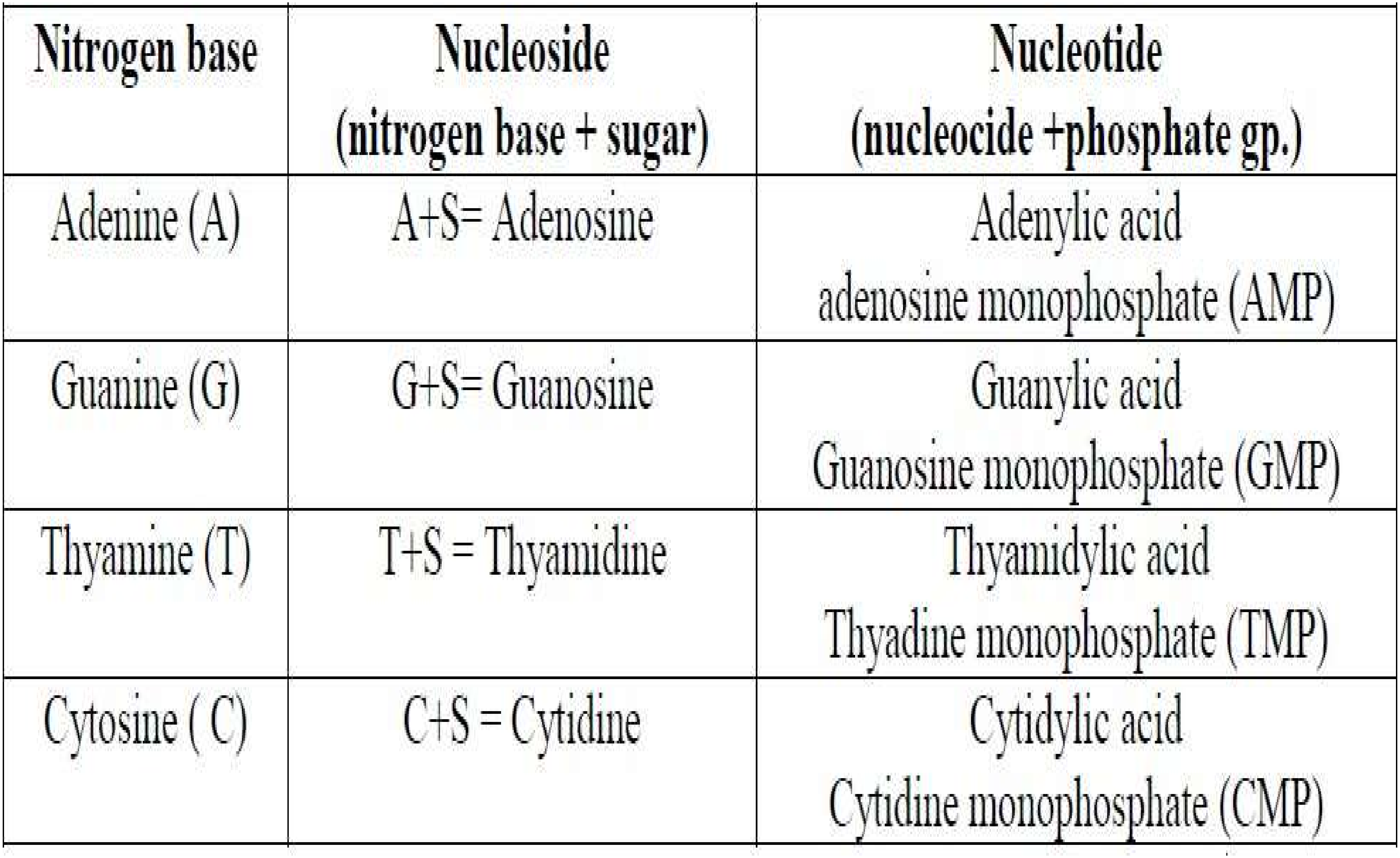
(1)Adenine + deoxyribose + phosphoric acid = deoxyadenylic acid or deoxyadenylate / dAMP

(2) Guanine + deoxyribose + phosphoric acid = deoxyguanylic acid or deoxyguanylate / dGMP

(3)Cytosine + deoxyribose + phosphoric acid = deoxycytidylic acid or deoxycytidylate / dCMP

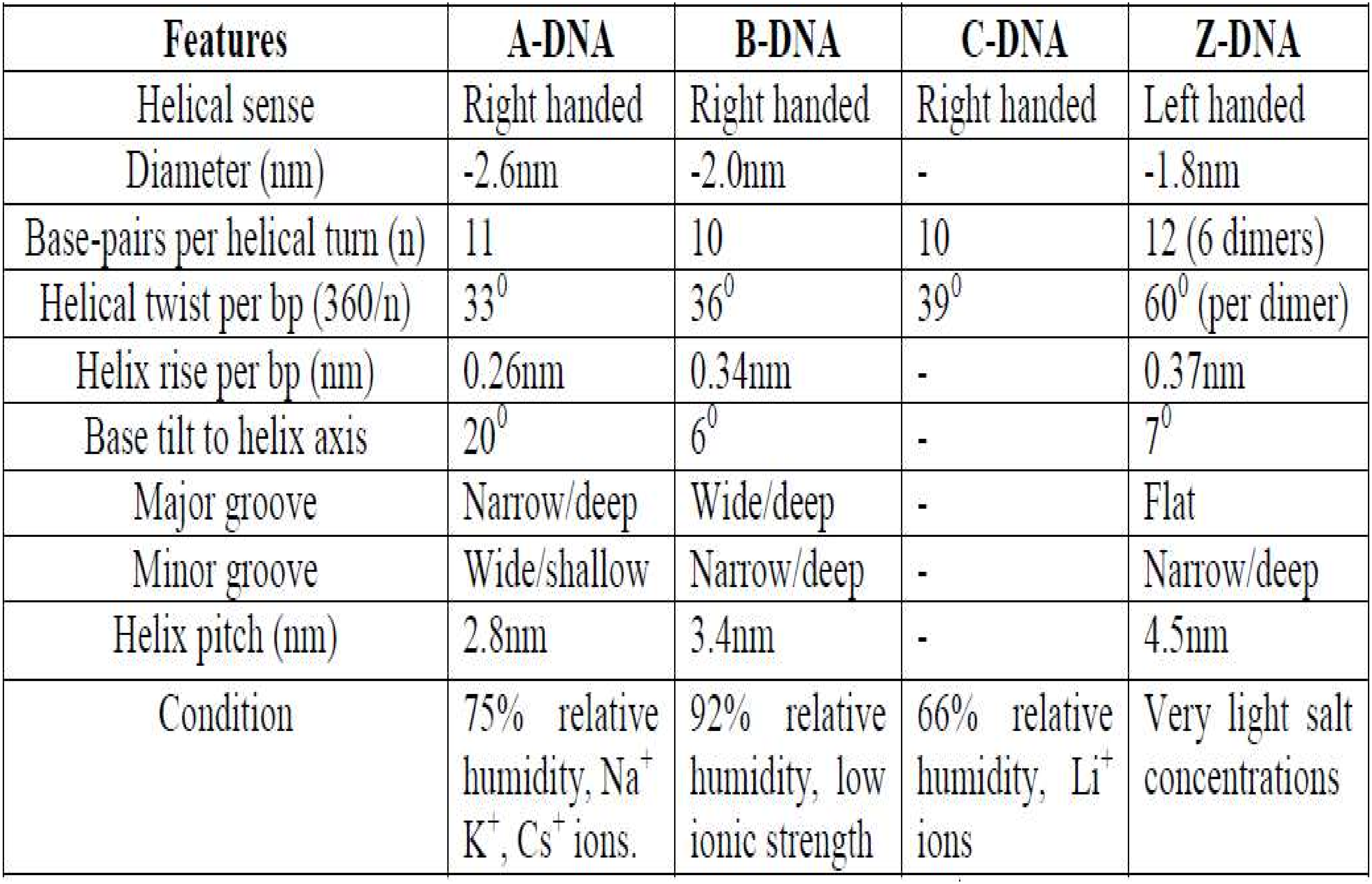
(4)Thymine+deoxyribose+phosphoric acid = deoxythymidylic acid or deoxythymidylate / dTMP

Table:1 Nitrogen bases, their respective nucleosides and nucleotides of DNA



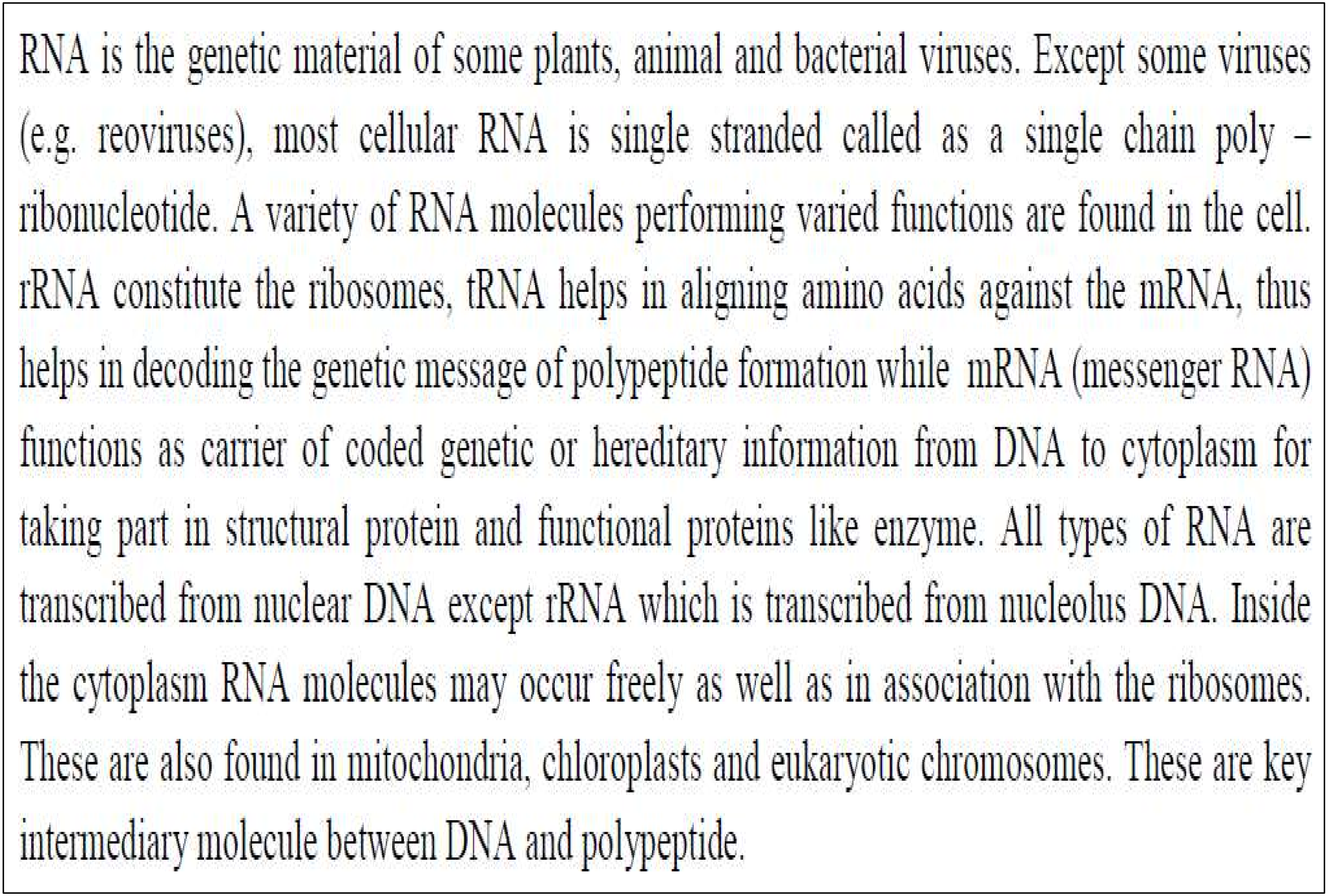
Nitrogen base	Nucleoside (nitrogen base + sugar)	Nucleotide (nucleocide +phosphate gp.)
Adenine (A)	A+S= Adenosine	Adenylic acid adenosine monophosphate (AMP)
Guanine (G)	G+S= Guanosine	Guanylic acid Guanosine monophosphate (GMP)
Thyamine (T)	T+S = Thyamidine	Thyamidylic acid Thyadine monophosphate (TMP)
Cytosine (C)	C+S = Cytidine	Cytidylic acid Cytidine monophosphate (CMP)

Comparison of different type of DNA



Features	A-DNA	B-DNA	C-DNA	Z-DNA
Helical sense	Right handed	Right handed	Right handed	Left handed
Diameter (nm)	-2.6nm	-2.0nm	-	-1.8nm
Base-pairs per helical turn (n)	11	10	10	12 (6 dimers)
Helical twist per bp (360/n)	33^0	36^0	39^0	60^0 (per dimer)
Helix rise per bp (nm)	0.26nm	0.34nm	-	0.37nm
Base tilt to helix axis	20^0	6^0	-	7^0
Major groove	Narrow/deep	Wide/deep	-	Flat
Minor groove	Wide/shallow	Narrow/deep	-	Narrow/deep
Helix pitch (nm)	2.8nm	3.4nm	-	4.5nm
Condition	75% relative humidity, Na$^+$ K$^+$, Cs$^+$ ions.	92% relative humidity, low ionic strength	66% relative humidity, Li$^+$ ions	Very light salt concentrations

RNA



RNA is the genetic material of some plants, animal and bacterial viruses. Except some viruses (e.g. reoviruses), most cellular RNA is single stranded called as a single chain poly – ribonucleotide. A variety of RNA molecules performing varied functions are found in the cell. rRNA constitute the ribosomes, tRNA helps in aligning amino acids against the mRNA, thus helps in decoding the genetic message of polypeptide formation while mRNA (messenger RNA) functions as carrier of coded genetic or hereditary information from DNA to cytoplasm for taking part in structural protein and functional proteins like enzyme. All types of RNA are transcribed from nuclear DNA except rRNA which is transcribed from nucleolus DNA. Inside the cytoplasm RNA molecules may occur freely as well as in association with the ribosomes. These are also found in mitochondria, chloroplasts and eukaryotic chromosomes. These are key intermediary molecule between DNA and polypeptide.

Chemically RNA differs from DNA in three ways-

- The sugar molecule found in RNA is ribose, rather than the deoxyribose of DNA.
- It is generally consists of only one polynucleotide strand or single stranded.
- Three nitrogen bases (A, G, C) in RNA are identical to those in DNA, the fourth base in RNA is Uracil (U), which is similar to thymine but lacks the methyl (-CH3) group.

Structure of RNA

RNA is single stranded polyribonucleotide. Each ribonucleotide is made of:

- Phosphoric acid- H_3PO_4
- Ribose sugar- $C_5H_{10}O_5$
- Nitrogen base- Adenine (A), Guanine (G), Cytocine (C) and Uracil (U)

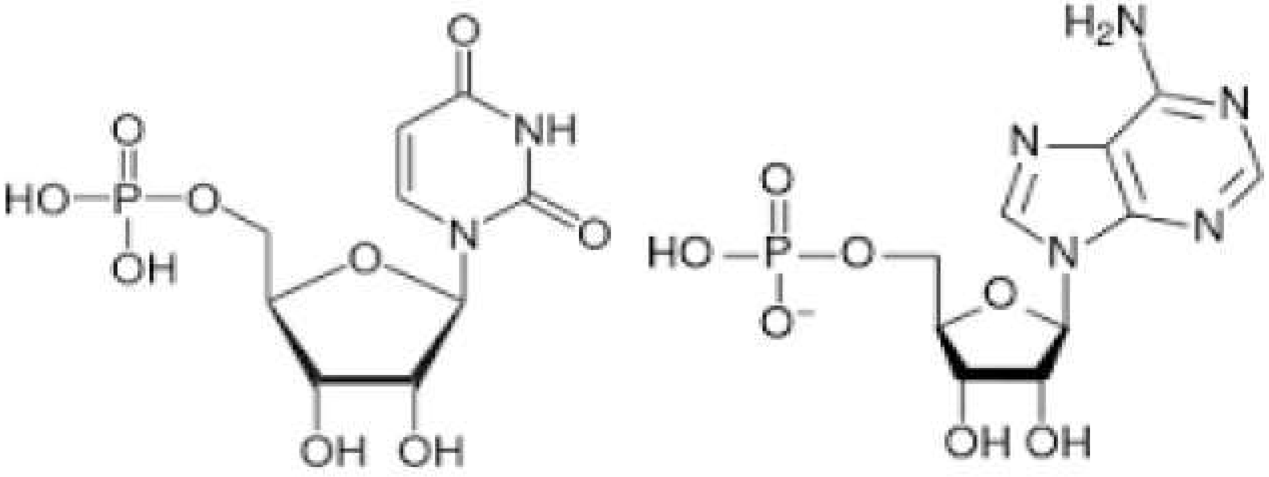


Fig. 10.2 Components of a ribonucleotide

Many ribonucleotides join with each other by phosphor-ester bonds to make a linear chain of polyribonucleotides. The chain will remain straight under all conditions in mRNA, may fold randomly in r-RNA or specifically to form t-RNA.

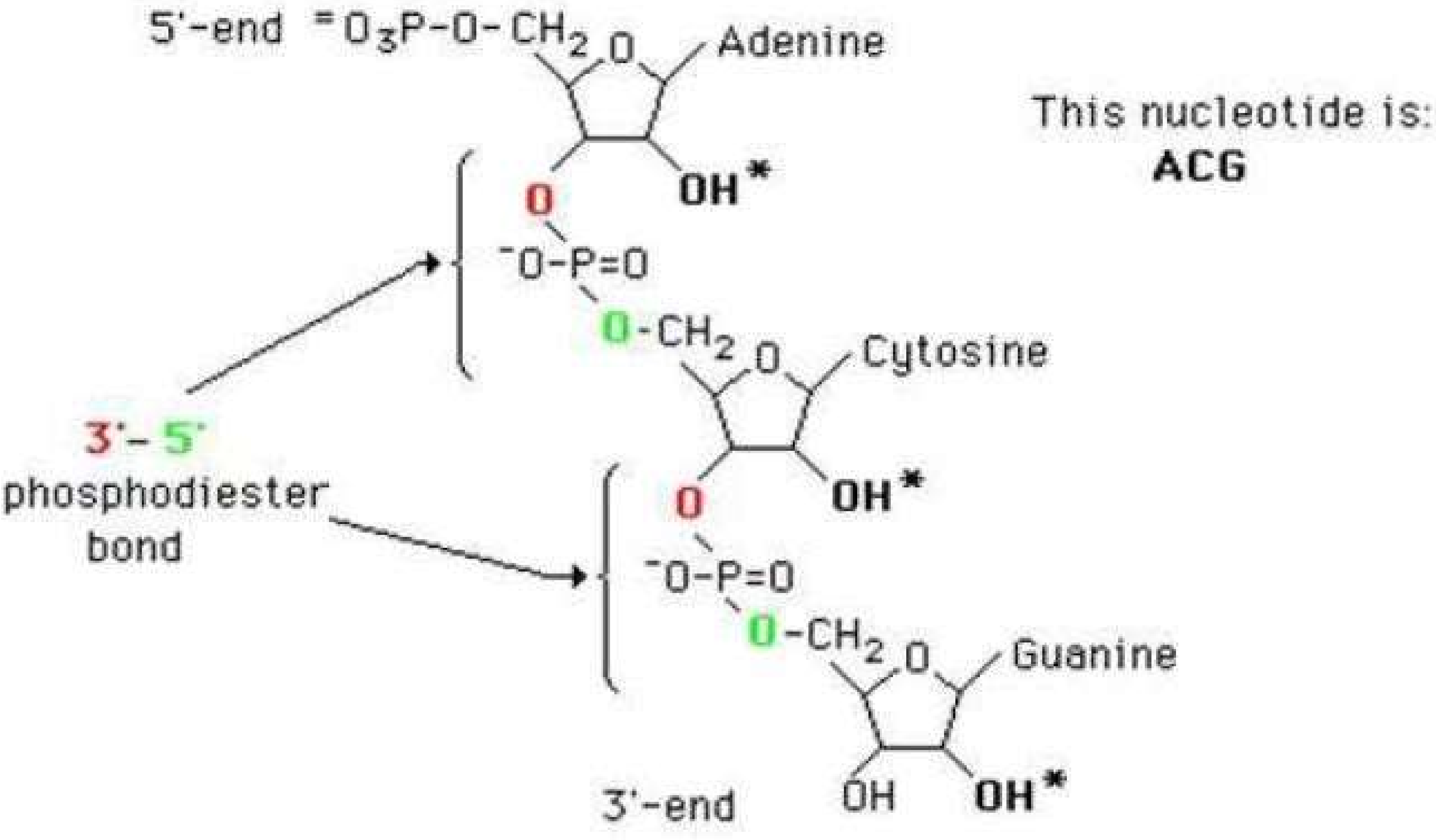
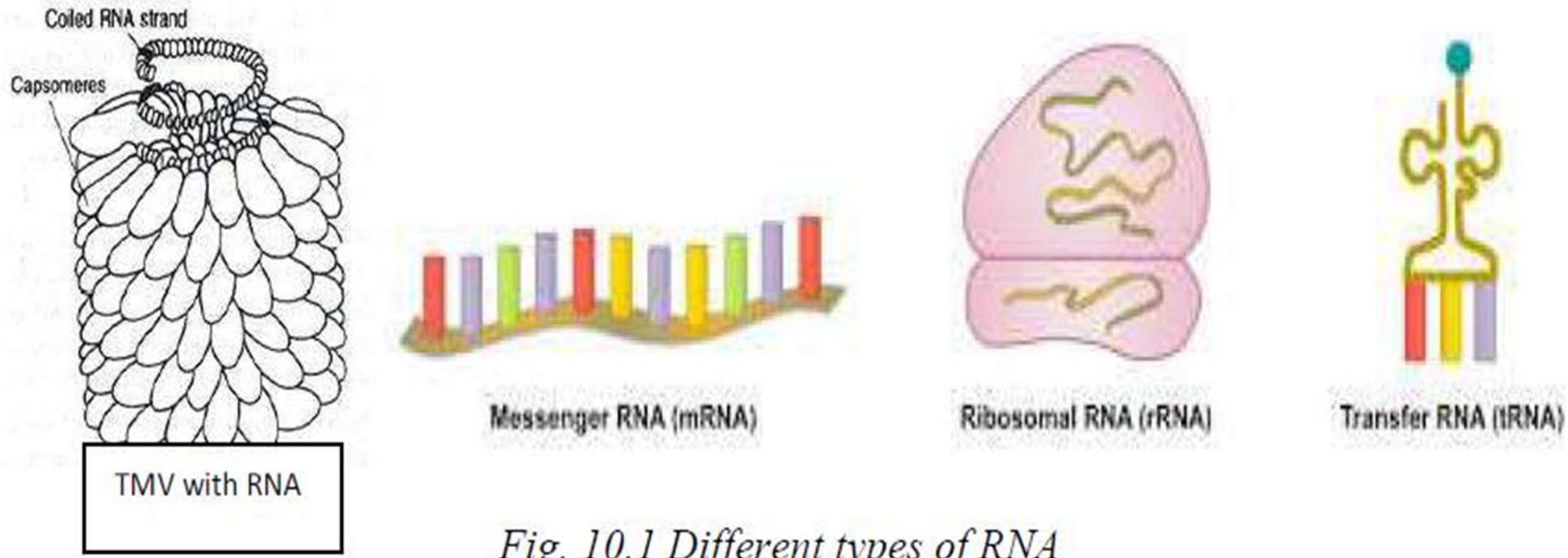


Fig. 10.3 A chain of ribonucleotides to form polyribonucleotides

RNA is generally involved in protein synthesis but in majority of plant and some animal viruses it also acts as genetic material. There are two major types of RNA:

1. **Genetic RNA**- H. Fraenkel-Conrat showed that RNA present in **Tobacco Mosaic Virus** is its genetic material and this RNA is responsible for the infection in tobacco plant.

2. Non- genetic RNA- Prokaryotes and Eukaryotes where genetic information is contained in the DNA molecule, functions of such cells are performed by a different kind of nucleic acids called non- genetic ribonucleic acid. Non-genetic RNA is synthesized on DNA template. Such non genetic RNAs can be of many types like mRNA, r RNA, & t RNA.



TMV with RNA

Messenger RNA (mRNA)

Ribosomal RNA (rRNA)

Transfer RNA (tRNA)

Fig. 10.1 Different types of RNA

t-RNA or Transfer RNA

- It is also called **soluble or s-RNA**.
- There are over 100 types of t-RNA.
- t-RNA is the smallest RNA with 70-85 nucleotides and sedimentation co-efficient of 4S.
- It is about 10-15% of the total weight of t-RNA of the cell.
- Each t-RNA has a corresponding **anticodon** that can recognize the codon on mRNA and exhibit high affinity for specific activated amino acids combine with them and carry them to the site of protein synthesis.

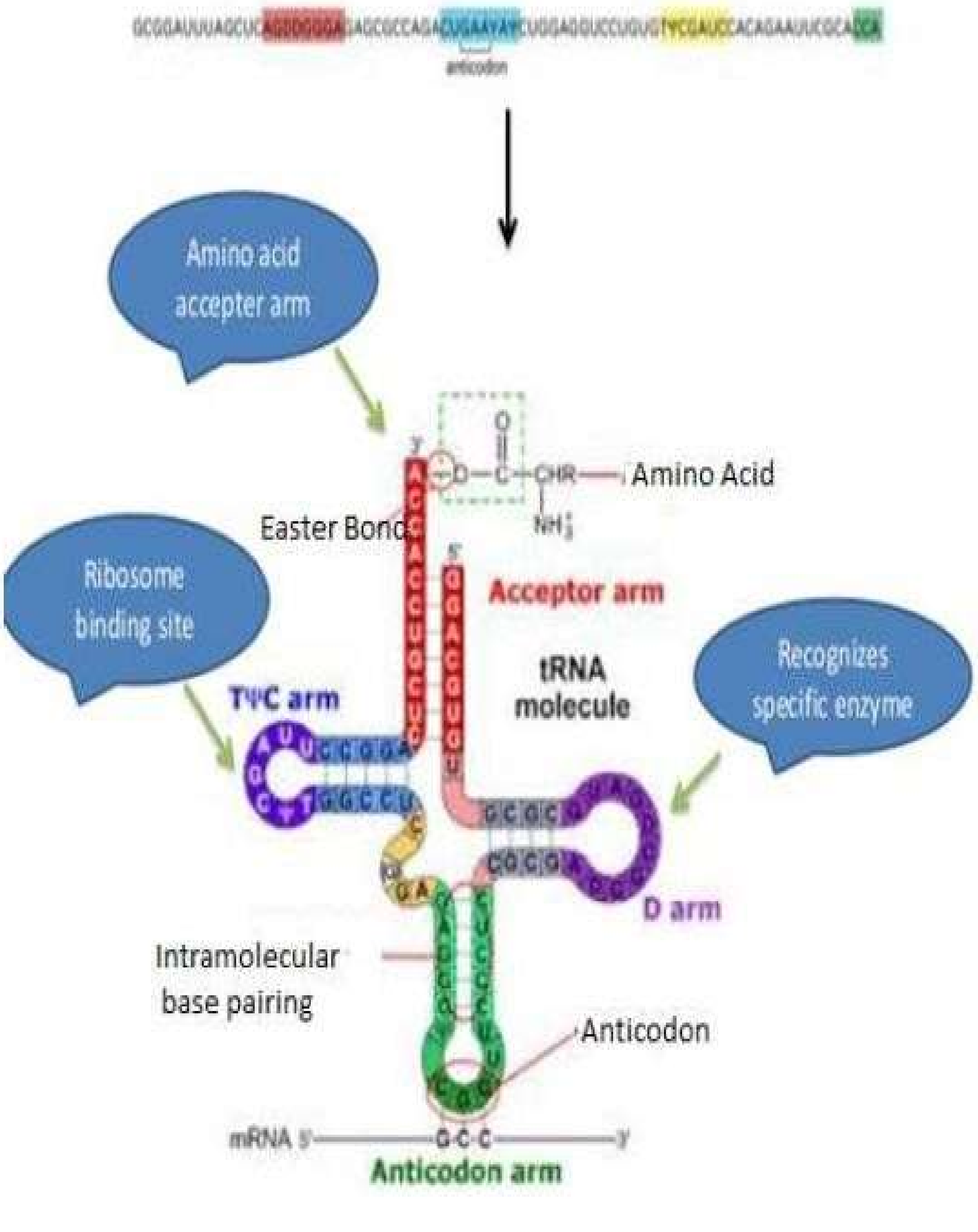
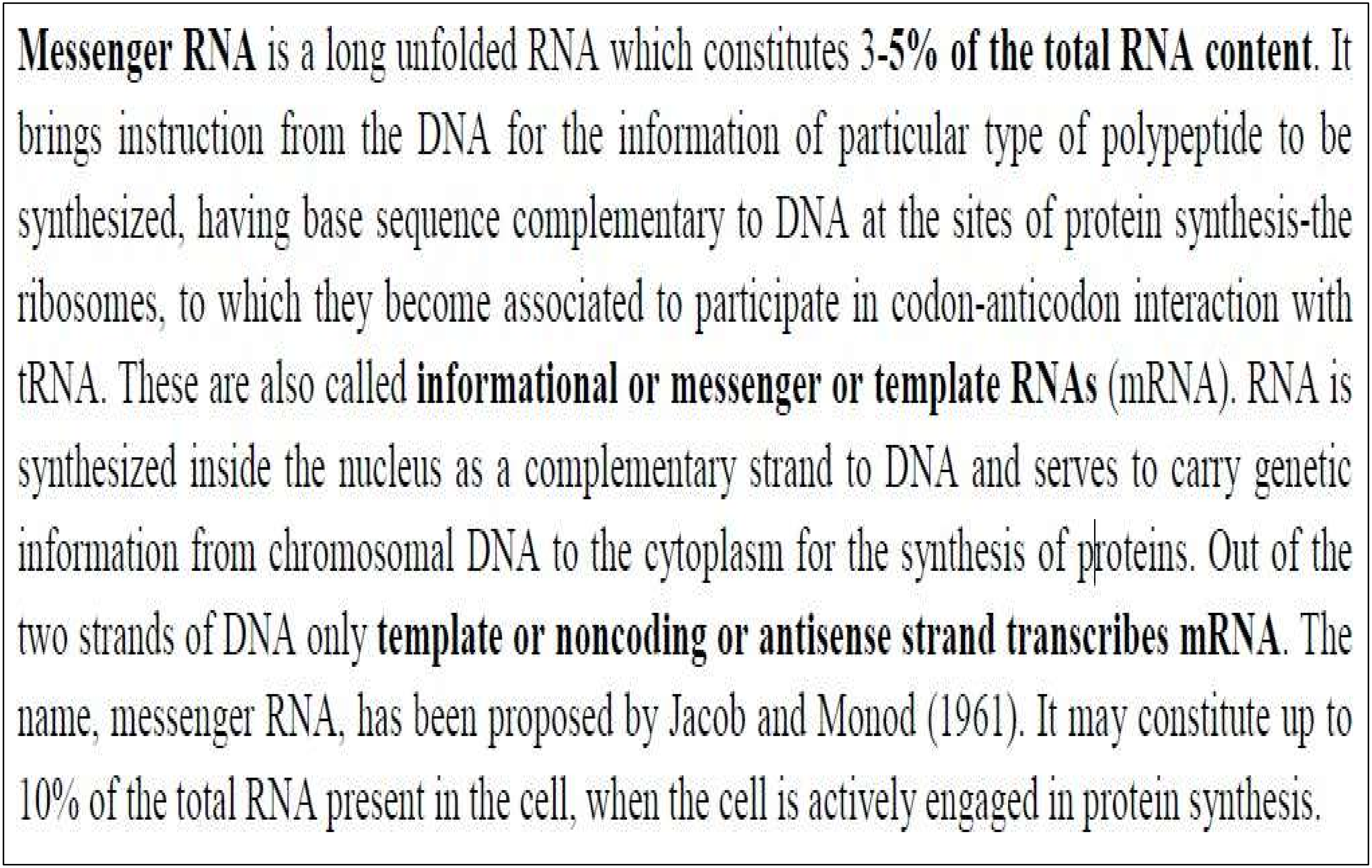


Fig. 10.4. t-RNA structure

m-RNA or Messenger RNA



Messenger RNA is a long unfolded RNA which constitutes **3-5% of the total RNA content**. It brings instruction from the DNA for the information of particular type of polypeptide to be synthesized, having base sequence complementary to DNA at the sites of protein synthesis-the ribosomes, to which they become associated to participate in codon-anticodon interaction with tRNA. These are also called **informational or messenger or template RNAs** (mRNA). RNA is synthesized inside the nucleus as a complementary strand to DNA and serves to carry genetic information from chromosomal DNA to the cytoplasm for the synthesis of proteins. Out of the two strands of DNA only **template or noncoding or antisense strand transcribes mRNA**. The name, messenger RNA, has been proposed by Jacob and Monod (1961). It may constitute up to 10% of the total RNA present in the cell, when the cell is actively engaged in protein synthesis.

THE STRUCTURE OF mRNA-

m-RNA is **always single stranded** having normal bases like A, G, U and C along with only a few unusual substituted bases. There is never base pairing in mRNA. It functions as a template for protein synthesis it carries genetic information from DNA to a ribosome and helps to assemble amino acids in their correct order. Each amino acid in a protein is specified by a set of three nucleotides in the mRNA called **codons.** Both prokaryotic and eukaryotic mRNA contains three primary regions:

a) 5' untranslated region (5'UTR) - the 5' untranslated region is a sequence of nucleotides at the 5' end of the mRNA that does not code for the amino acid sequence of a protein. In **prokaryotic** (bacterial cell) mRNA contains a consensus sequence called the **Shine-Dalgarno sequence** (5'AGGAGGU3'), which serves as **the ribosome binding site during translation**, it is formed of approximately 7 nucleotides upstream of the first or start codon. Eukaryotic mRNA has no such equivalent sequences in its 5' untranslated region. This is the sequence of the mRNA extending from the 5' end of the mRNA to the initiation codon. It is not translated into polypeptide sequence. It has a **function analogous to the function of a promoter on a gene**. It will direct the binding of the ribosome to the initiation codon.

b) Protein coding region- this region comprises the codon that specify the amino acid sequence of the protein. This region **begins with a start codon and ends with a stop codon.** This region has 3 regions namely initiation codon, coding region, stop codon.

- **Initiation codon**- it is always **AUG** and codes for a **methionine.** This is the triplet codon at which polypeptide synthesis begins. All polypeptides are synthesized with an amino terminal methionine.
- **Coding region**-this is the sequence of mRNA that contains **the consecutive triplet codons** that direct polypeptide synthesis. This region starts from the start codon and continue up to the stop codon. The coding region is often referred to as the open reading frame or ORF.
- **Stop codon**-this is the triplet codon that signals the **termination of translation**. There are three possible stop codon sequences **UAA, UAG, UGA**. Stop codons have no corresponding tRNA or amino acid.

c) 3' Untranslated region (3'UTR)-This region of mRNA is the 3' un-translated region, a sequence of nucleotides at the 3'end of mRNA that is not translated into protein. This is the nucleotide sequence downstream from the stop codon. It extends from the stop codon to the 3' end of the mRNA. It does not code for amino acid sequence. It may function in stabilizing the mRNA. In eukaryotes it is transcribes as hnRNA which is converted into functional mRNA in the cytoplasm by removing introns (intervening sequences) and joining together exons (expressible sequences)

For the convenience the mRNA structure can be summarized as:

1. **Cap-** at 5' end, has methylated structure, does not translate
2. **Noncoding region-1-** has 10-100 nucleotides, rich in U and A bases, does not translate
3. **The initition codon-** AUG, codes for methionine amino acid
4. **The coding region-**about 1500 nucleotides on an average, translate protiens
5. **Termination codon-** either of UAA, UAG or UGA ie present, helps in termination of translation
6. **Noncoding region-2-** made of 50-150 nucleotides, does not translate, has sequence like AAUAAA
7. **Poly(A) sequence-** 200-250 A nucleotides, does not translate, makes tail of mRNA

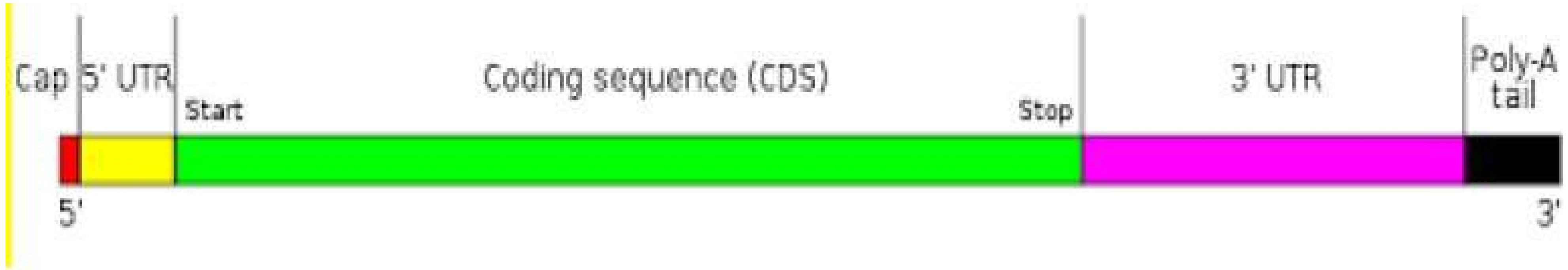


Fig. 10.6 mRNA showing different regions

r-RNA or RIBOSOMAL RNA

Ribosomal, stable or insoluble RNA constitutes the largest part (up to 80%) of the total cellular RNA. It was reported by **Kuntz.** It is found primarily in the cytoplasm as well as organelle. In prokaryotes it is transcribed from ribosomal DNA which is a part of nuclear DNA but in eukaryotes ribosome is formed on nucleolar DNA. The genetic instruction contained in mRNA is translated into the amino acid sequences of polypeptides only with the help of ribosomes. Thus ribosomes play an integral part in the transfer of genetic information from genotype to phenotype. R-RNA is most stable type of RNA.

Structure and processing of ribosome RNA- it forms about 80% of the total cellular RNA. r-RNA consists of a single stranded RNA which gets twisted over itself in certain regions due to complementary base pairing. R-RNA strand unfold on heating and refold on coiling. It is one the most stable RNA among all types of RNAs. R-RNA and ribo-proteins constitute ribosomes.

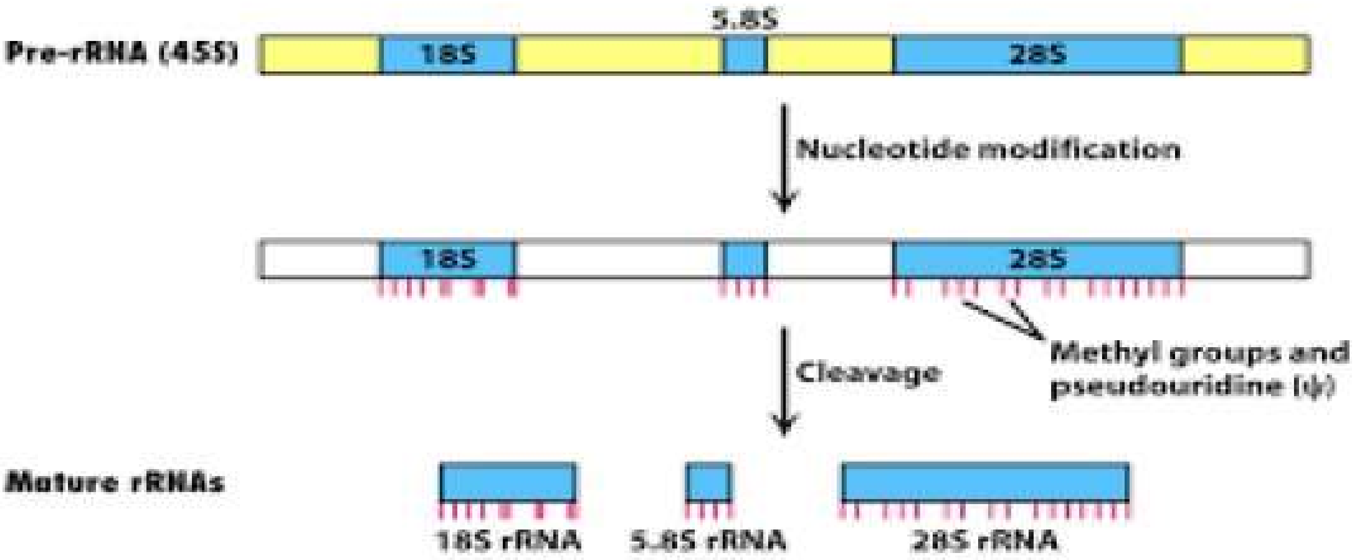
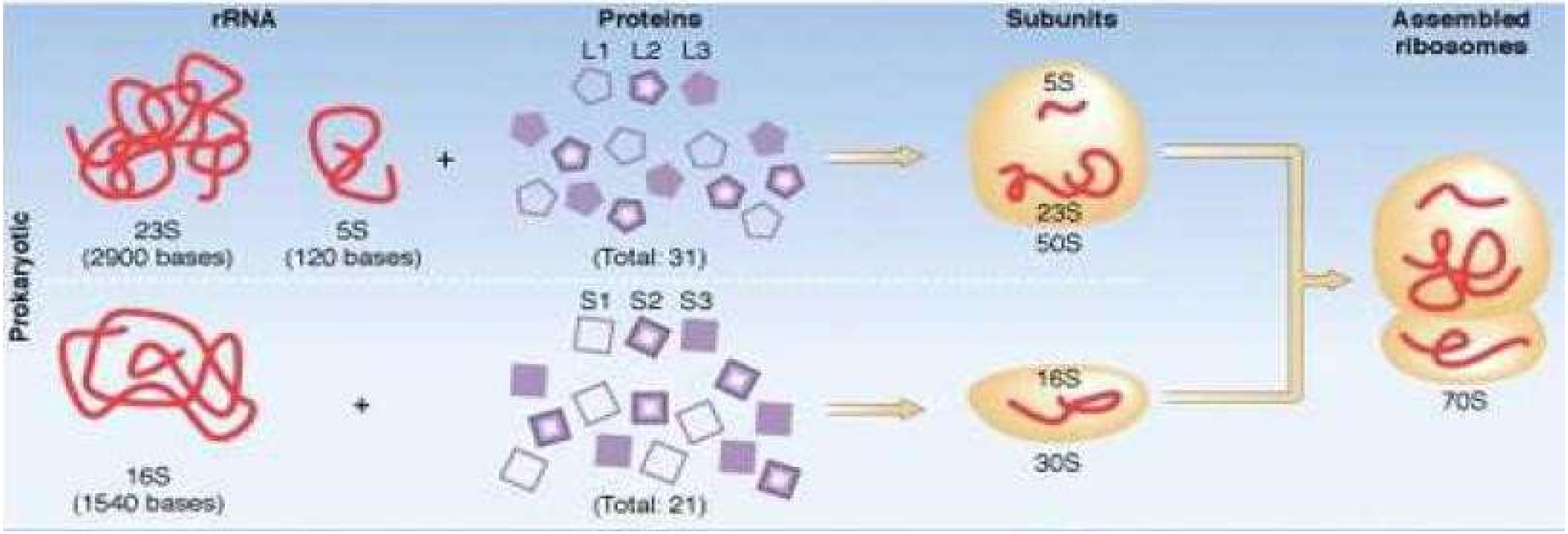
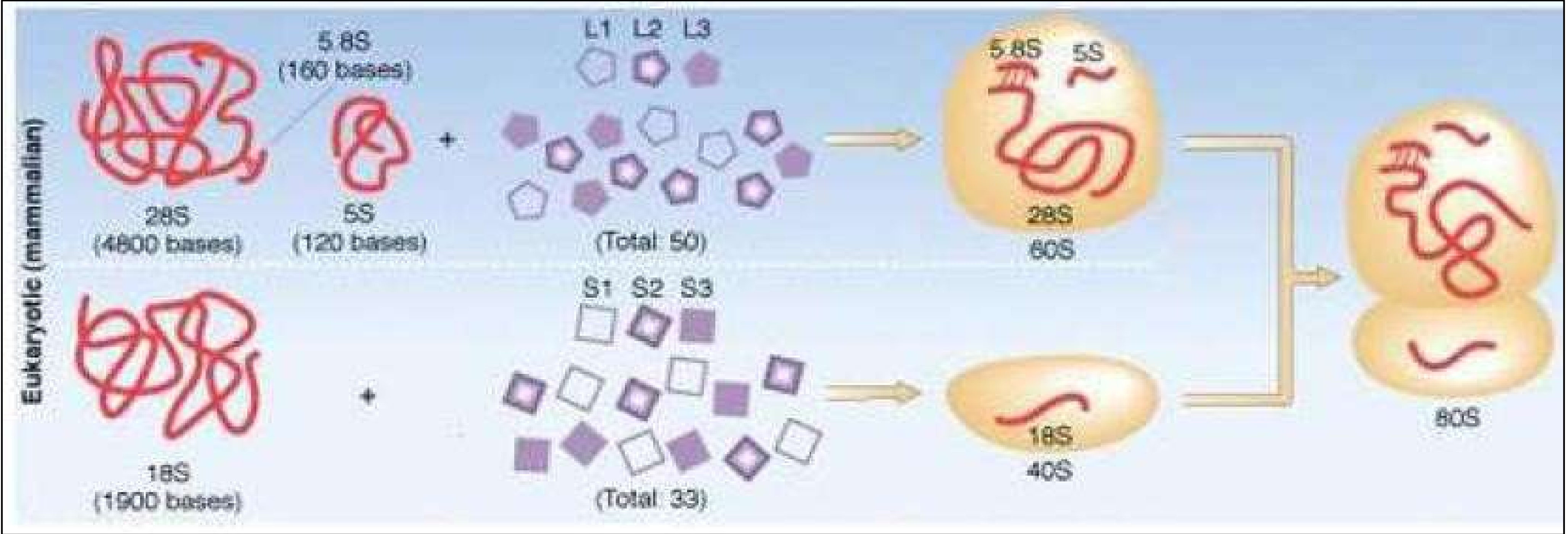


Fig. 10.7 processing of rRNA in a eukaryotic cell

In prokaryotes, 70S ribosome is made of two sub units- 30S and 50S. p30S subunit has 16SrRNA while 50S has 23S and 5S rRNA. An initial 30s transcript is made in ***E. coli*** by RNA polymerase. During processing p30S transcriptional unit is cleaved by RNase 111 into 25S and18S segments. Which are further reduced to p23S and p16S, further trimming results in functional 23Sa and 16S. Some modification of bases like methylation also occurs during processing of rRNA.



In **eukaryotes** 4 types of rRNAs found are **28s, 18s, 5.85s, and 5s.** In the nucleolus of eukaryotes, RNA polymerase-I transcribes the rRNA genes, which usually exit in tandem repeats to yield a long, single pre-rRNA which contains one copy each of the 18s, 5.8s and 28s sequences. Various spacer sequences are removed from the long pre-rRNA molecule by a series of specific cleavages. Many specific ribose methylations take place directed by small ribonucleoprotein particles (snRNPs) and the mature rRNA molecule fold and complex with ribosomal proteins. RNA pol. III synthesizes the 5srRNA from unlinked genes.



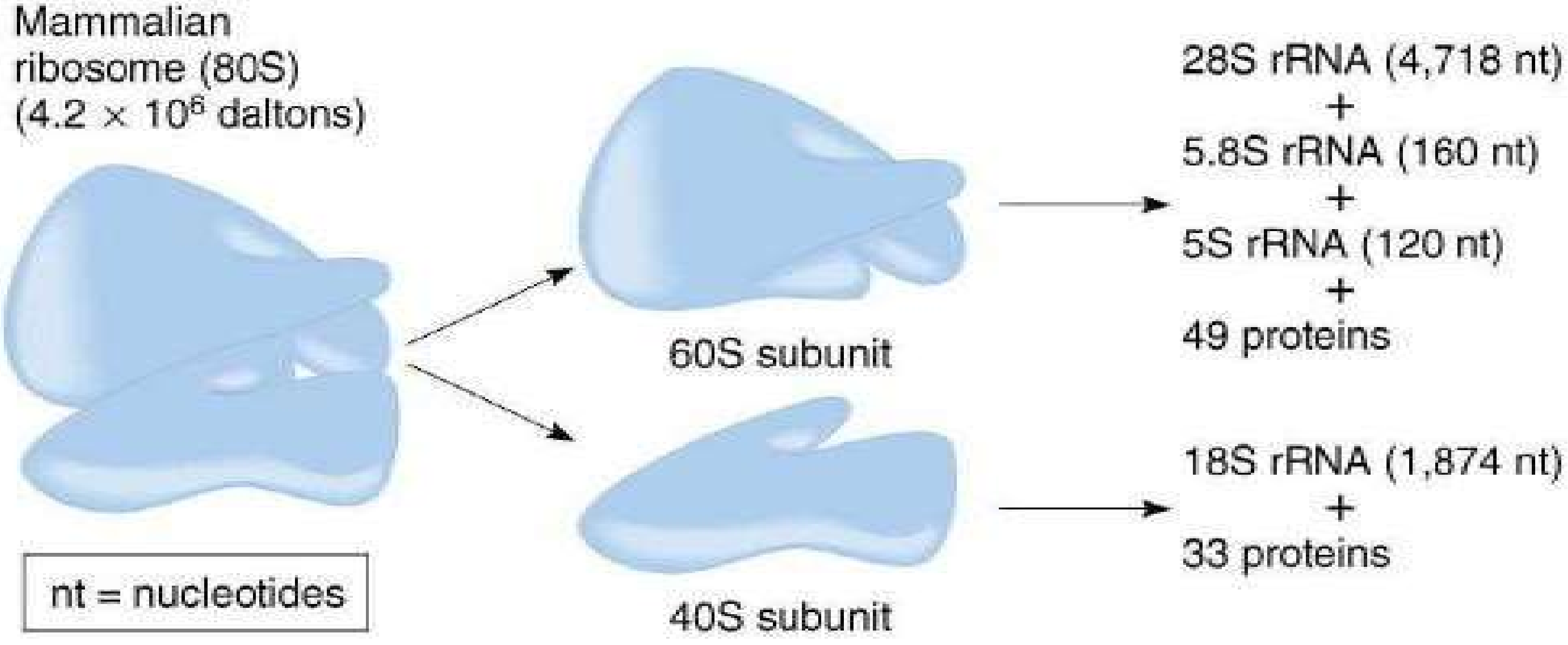


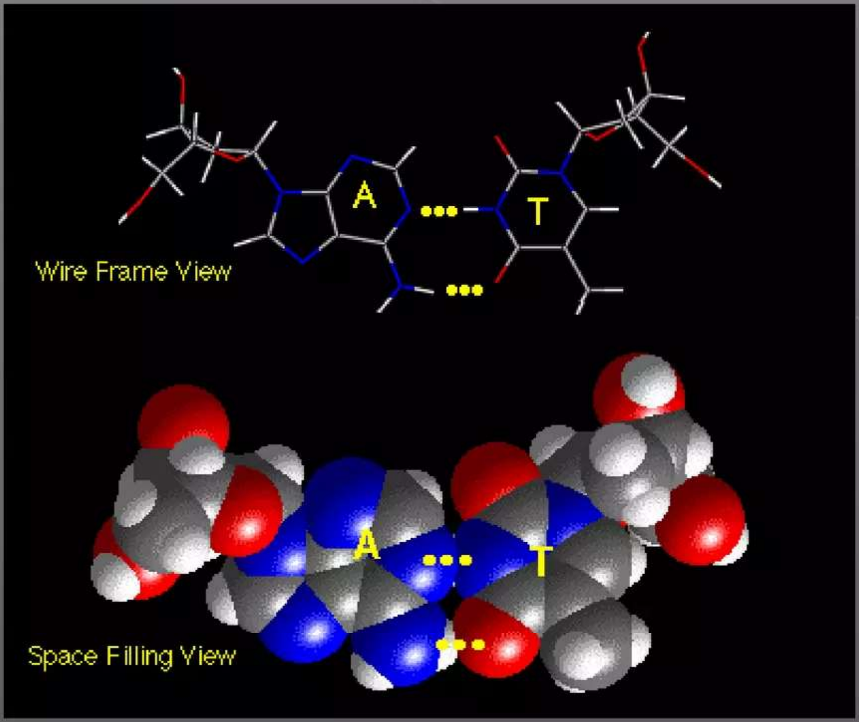
Fig. 10.9 Different types of rRNA in a eukaryotic cell

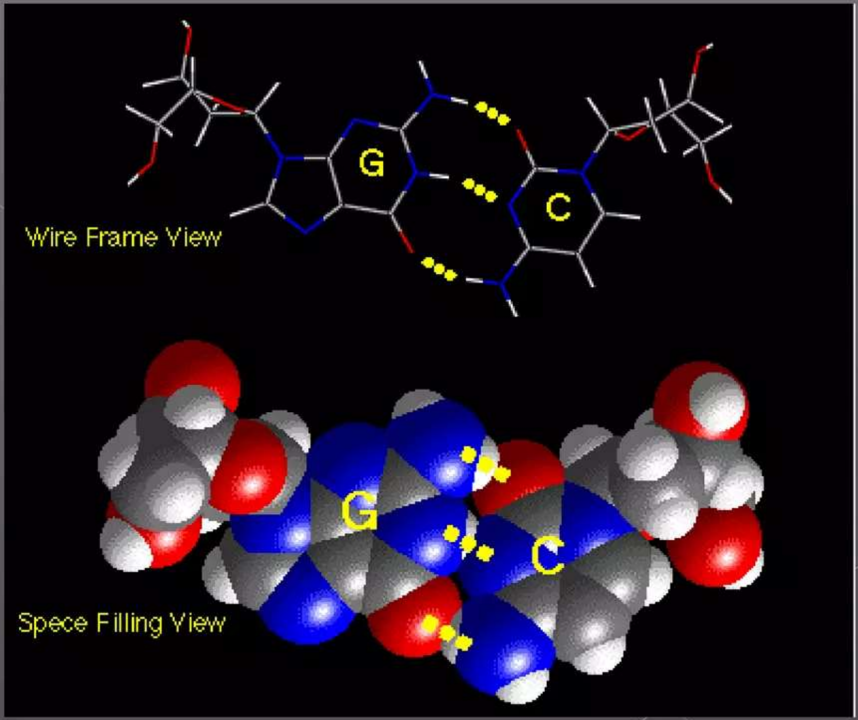
Cell type	Ribosome size	Subunit	rRNA component	Proteins
Bacterial	70S (Svedberg unit)	Large (50S)	23S (2900 nucleotides)	31
			5S (120 nucleotides)	
		Small (30S)	16S (1500 nucleotides)	21
Eukaryotic	80S	Large (60S)	28S (4700 nucleotides)	49
			5.8S (160 nucleotides)	
			5S (120 nucleotides)	
		Small (40S)	18S (1900 nucleotides)	33

RNA	DNA
RNA is single stranded except in some viruses	DNA is double stranded except in few viruses
RNA have ribose sugar	DNA have deoxyribose sugar
Bases present are adenine, guanine, cytosine and uracil.	Bases present are adenine, guanine, cytosine and thymine.
Adenine pairs with uracil	Adenine pairs with thymine
Purine is not equal to pyrimidine	Purine is equal to pyrimidine (Chargaff's rule)
Regions having complementary nucleotides, pairs, and form hair pin loop like structure and helical.	Complementary nucleotides are present throughout the length of the DNA.
RNA is genetic material in some viruses.	DNA is the genetic material in all living organisms.
Length of RNA is short consisting of only few thousands nucleotides.	Length of DNA is quite large consisting of millions of nucleotides.
Three types of RNA are present in an organism: mRNA, rRNA, tRNA.	DNA occurs only in one form in an organism.
mRNA occurs in nucleolus, rRNA and tRNA occur in cytoplasm.	DNA occurs in nucleus, nucleolus, and extrachromosomal DNA in mitochondria and chloroplast.

Base Pairs

Within the DNA double helix, A forms 2 hydrogen bonds with T on the opposite strand, and G forms 3 hyrdorgen bonds with C on the opposite strand.





DNA with high GC-content is more stable due to intra-strand base stacking interactions

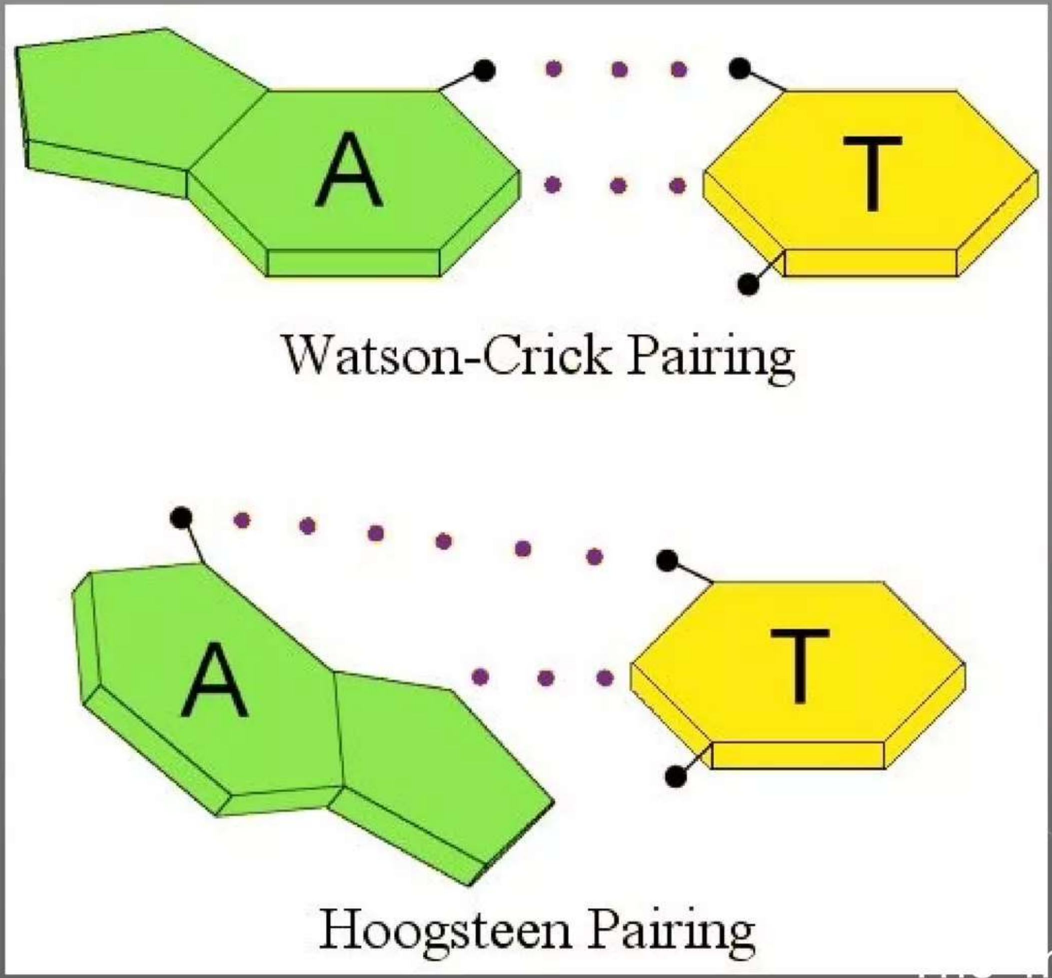
In the canonical **Watson-Crick DNA base pairing,** adenine (A) forms a base pair with thymine (T) and guanine (G) forms a base pair with cytosine (C).
In RNA, thymine is replaced by uracil (U).

Alternate hydrogen bonding patterns, such as the wobble base pair and Hoogsteen base pair, also occur—in particular, in RNA—giving rise to complex and functional tertiary structures.

A **Hoogsteen base pair** is a variation of base-pairing in nucleic acids such as the A•T pair.

In this manner, two nucleobases on each strand can be held together by hydrogen bonds in the major groove.
A Hoogsteen base pair applies the N7 position of the purine base (as a hydrogen bond acceptor) and C6 amino group (as a donor), which bind the Watson-Crick (N3–N4) face of the pyrimidine base.

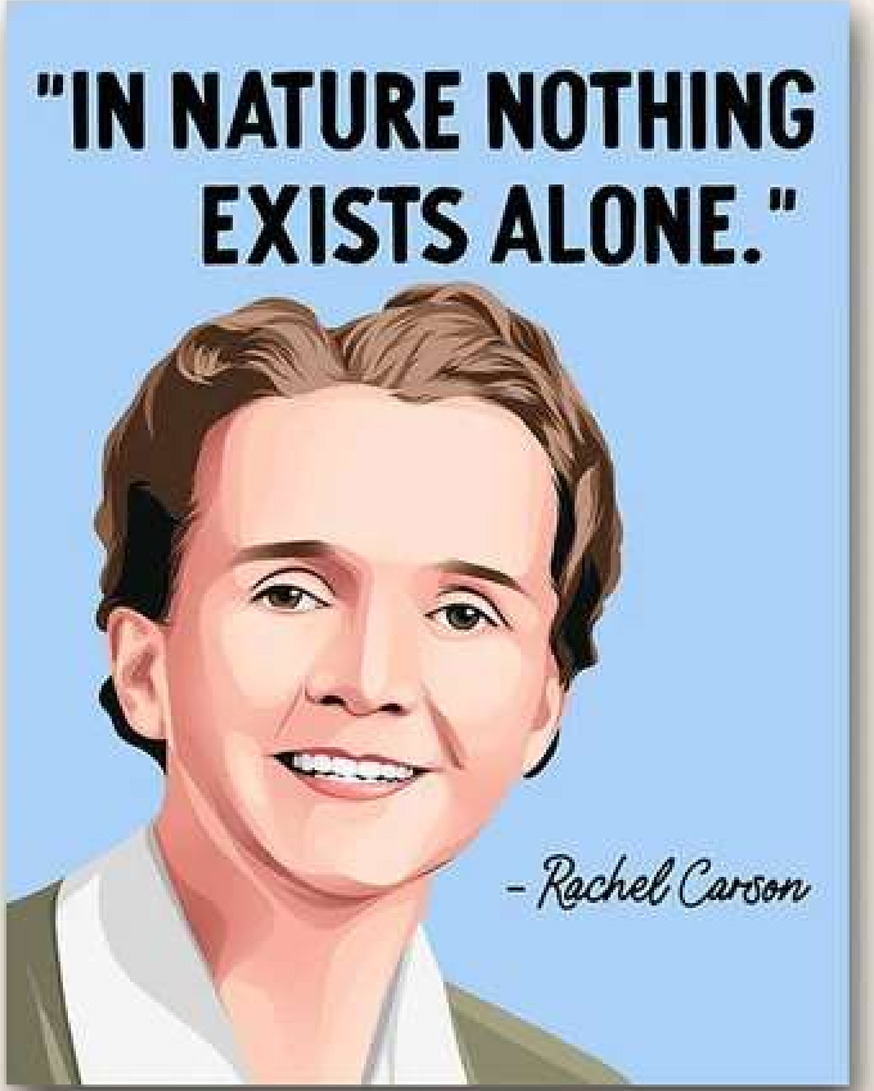
The angle between the two glycosylic bonds (ca. 80° in the A• T pair) is larger and the C1′–C1′ distance (ca. 860 pm or 8.6 Å) is smaller than in the regular geometry.
In some cases, called reversed Hoogsteen base pairs, one base is rotated 180° with respect to the other.

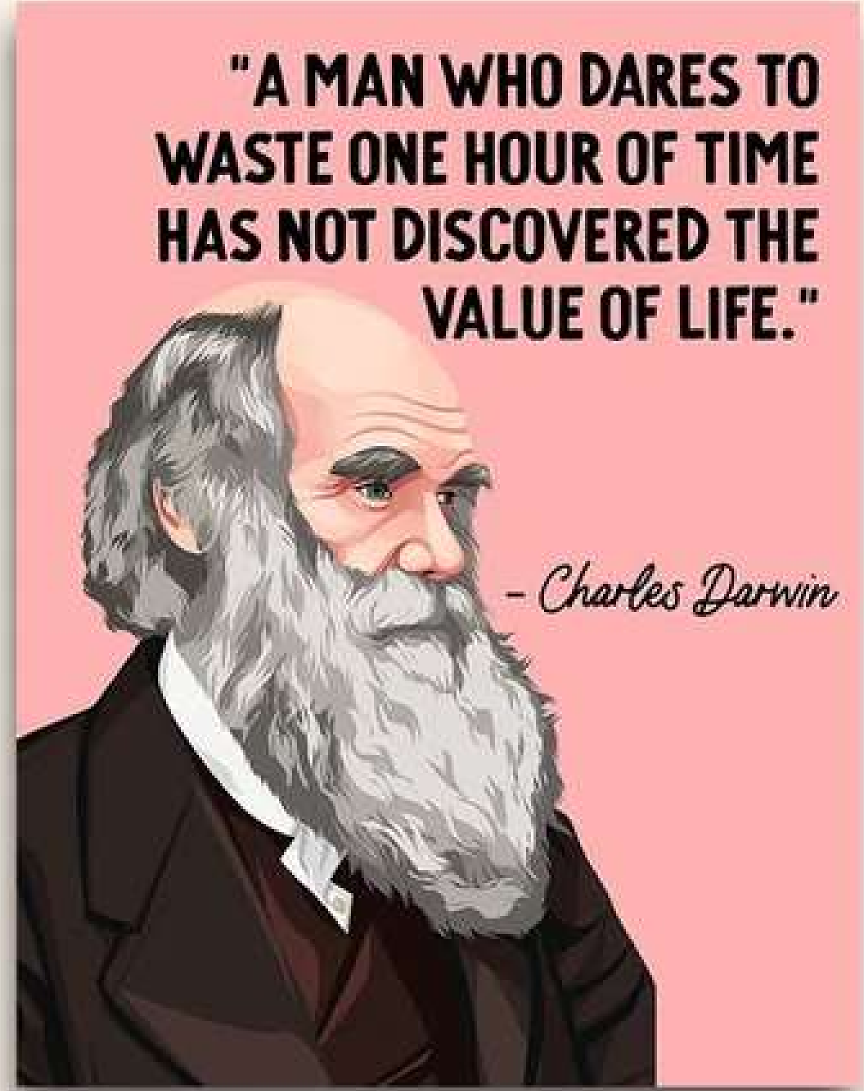


Watson-Crick Pairing

Hoogsteen Pairing

The majority of nucleotide bases in DNA link together with Watson-Crick pairing. However, there are a few that link together with Hoogsteen pairing. It has been thought for quite some time that Hoogsteen pairing only occurs in DNA when it is either damaged or bound to some other molecule (like a protein or a drug).


"IN NATURE NOTHING EXISTS ALONE."
- Rachel Carson


"A MAN WHO DARES TO WASTE ONE HOUR OF TIME HAS NOT DISCOVERED THE VALUE OF LIFE."
- Charles Darwin