# Molecular Biology

## Lecture 1 & 2

Dr Jehad Al-Shuneigat

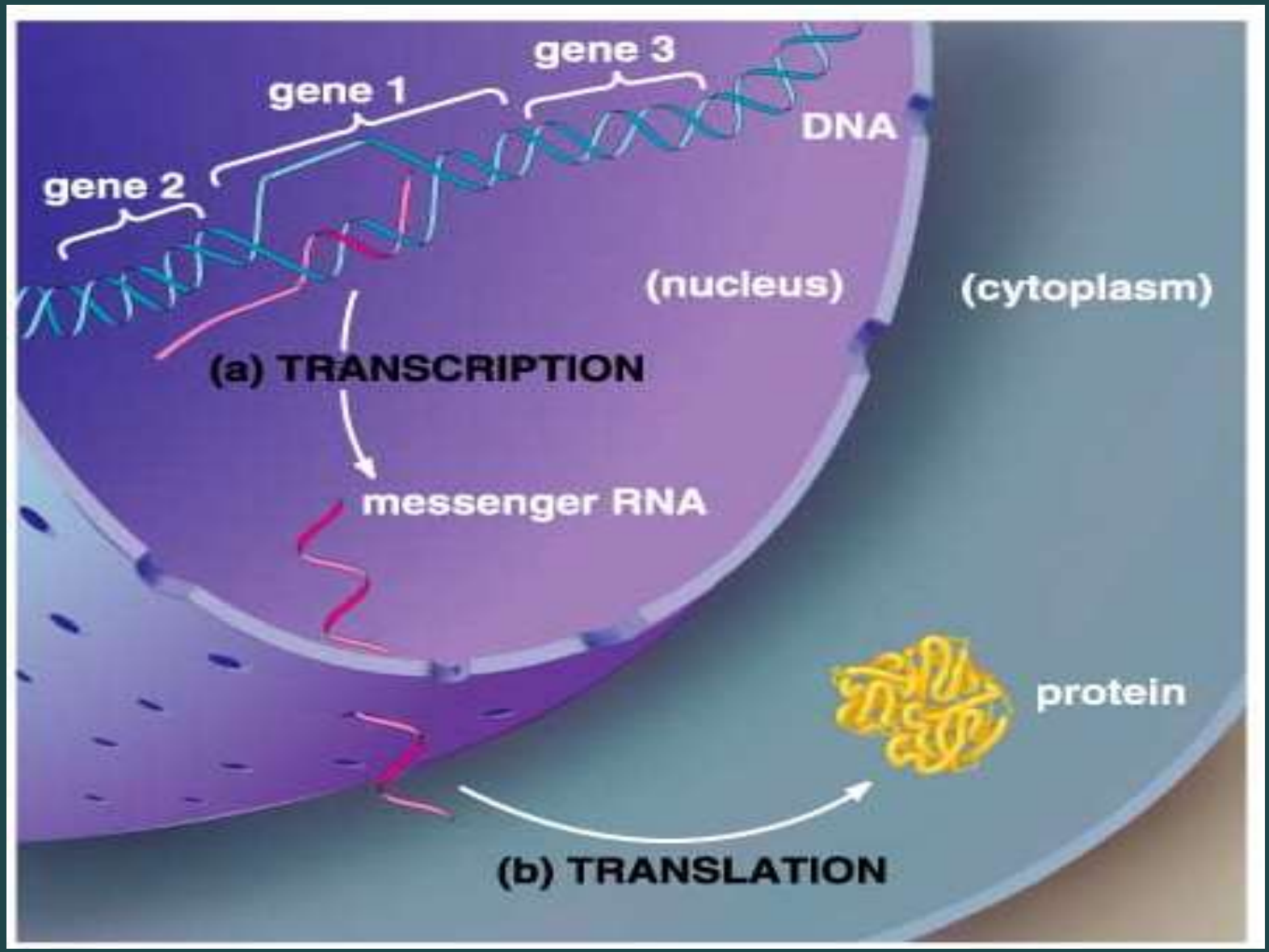
# Introduction

- Molecular biology is the study of biological molecules and the molecular basis of structure and function in living organisms.
- The field of molecular biology arose from the meeting point of work of geneticists, biochemists, physicists, and structural chemists on a common problem: the structure and function of the gene.
- The goal of molecular biology is to understand the cell growth, division, specialization, movement, and interaction in terms of various molecules that are responsible for them.

# Central Idea (Dogma) of Molecular Biology
## (DNA to RNA to Polypeptide )

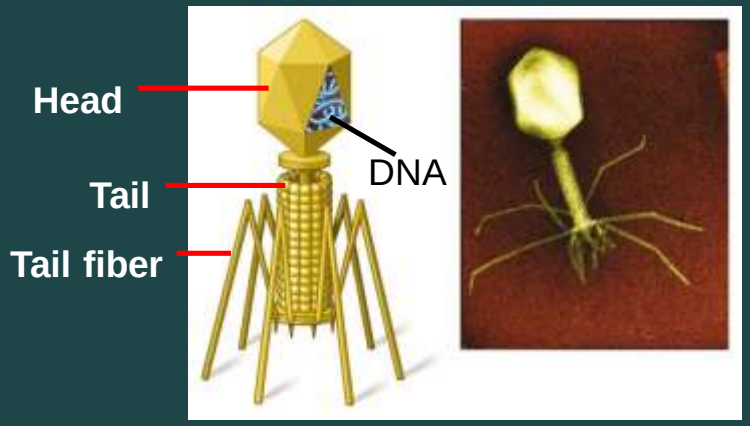
It estimated that humans has 30,000 genes
It is estimated that each gene codes for at least 4 polypeptide which can be explained by splicing.

**Genetic code** is the sequence of nucleotides in **DNA** that determines the sequence of amino acids in proteins.



# Models of biological system

- **Viruses**




- **Bacteria** prokaryote, agar (Agar is a complex polysaccharide derived from red algae), cell division, prototroph, auxotroph
- **Yeast** eukaryotes
- **Animal cells** **cell culture, cancer cell culture, stem cells**
- **Plant cells**

# Genetics and Medicine

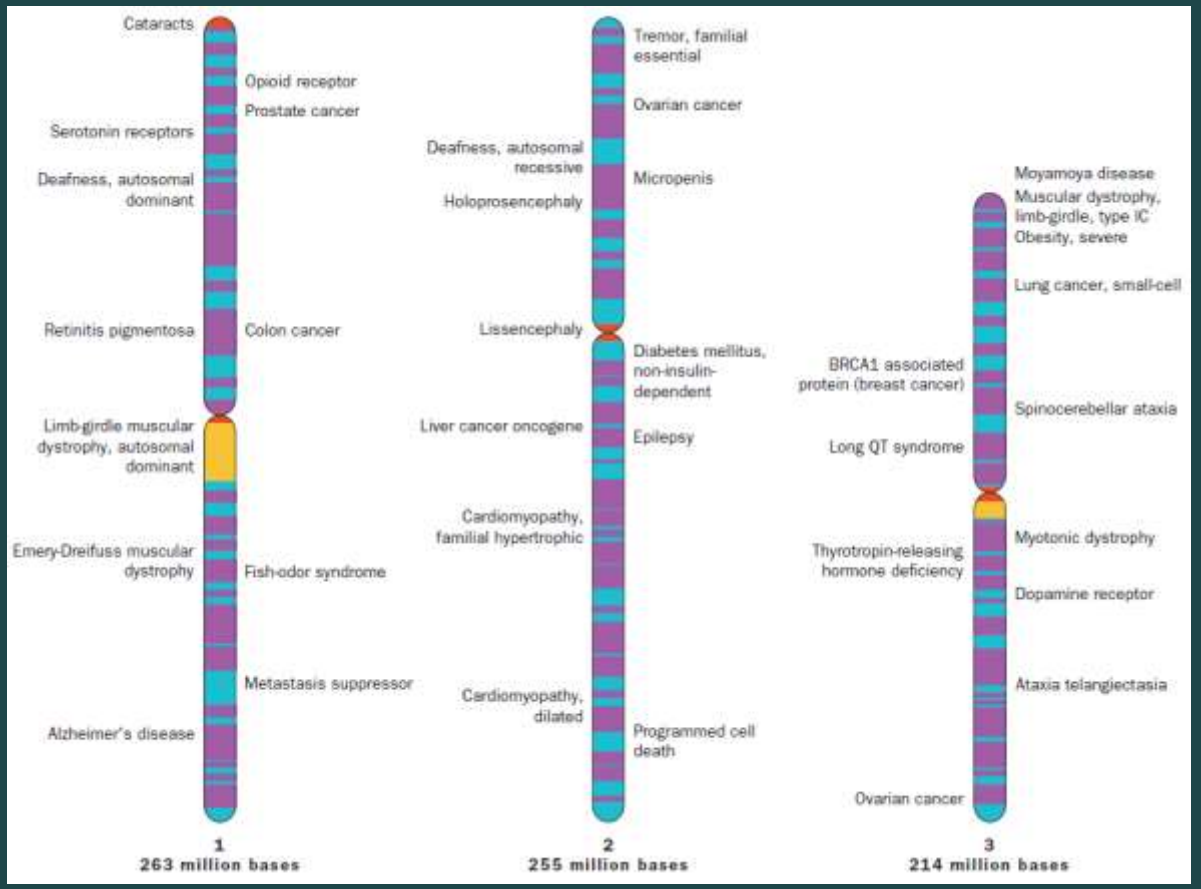
- Most disease result from environmental influences interacting with the individual genetic makeup {**genetic predisposition** (sometimes also called **genetic** susceptibility)}.
- e.g., high blood pressure, diabetes mellitus, psychiatric disorders
- More than 3000 defined human genetic diseases are known.

- Genetically determined diseases are not a marginal group.
- The total estimated frequency of genetically determined diseases of different categories in the general population is about 3.5–5.0%.

- Genetic counselling is generally offered prior to marriage or conception, in order to predict the likelihood of conceiving an affected child, during pregnancy.
- Gene therapy is used to correct defective genes that cause disease not yet an active current therapy.

# Terminology

- **Genome**
- **Chromosome**
- **DNA** (Watson-Crick model)
- **Gene**
- **RNA**
- **Proteins**



# Nucleic Acids

- **Nucleic acids** occur in two forms:

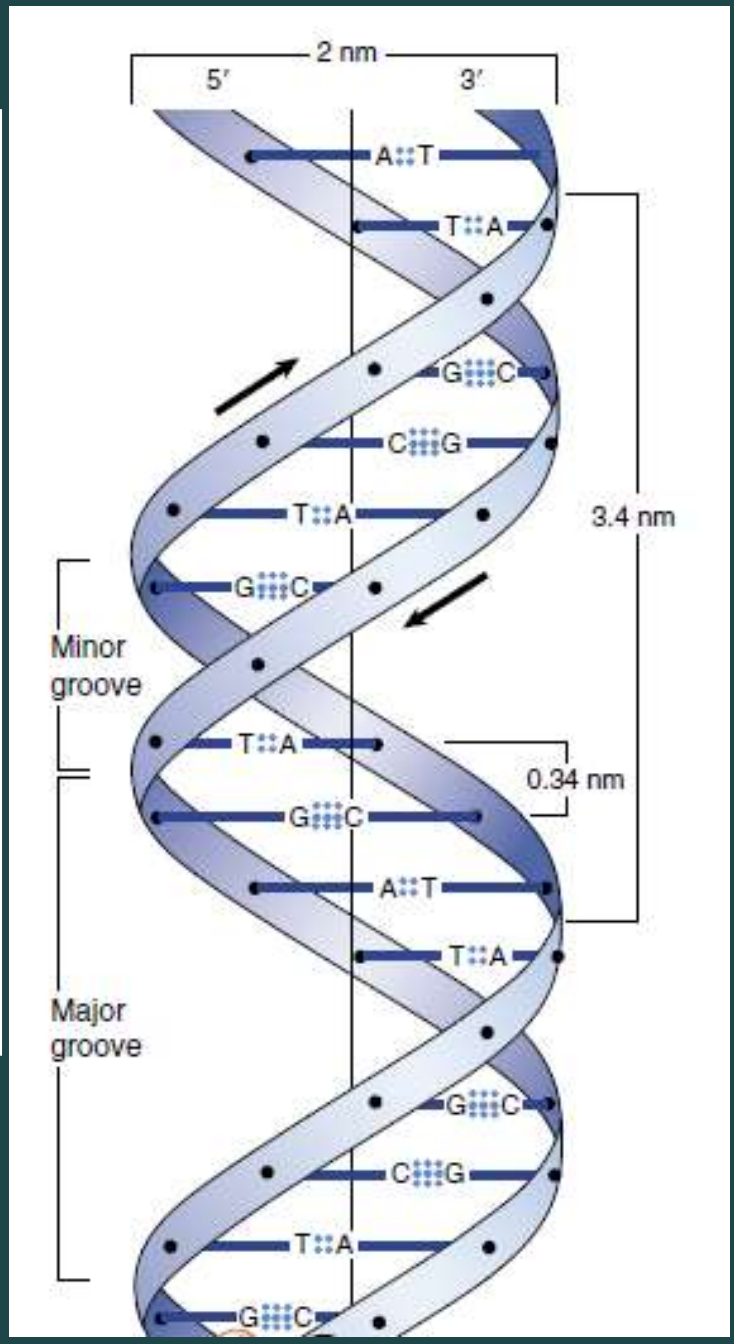
**1- DeoxyriboNucleic Acid (DNA)**

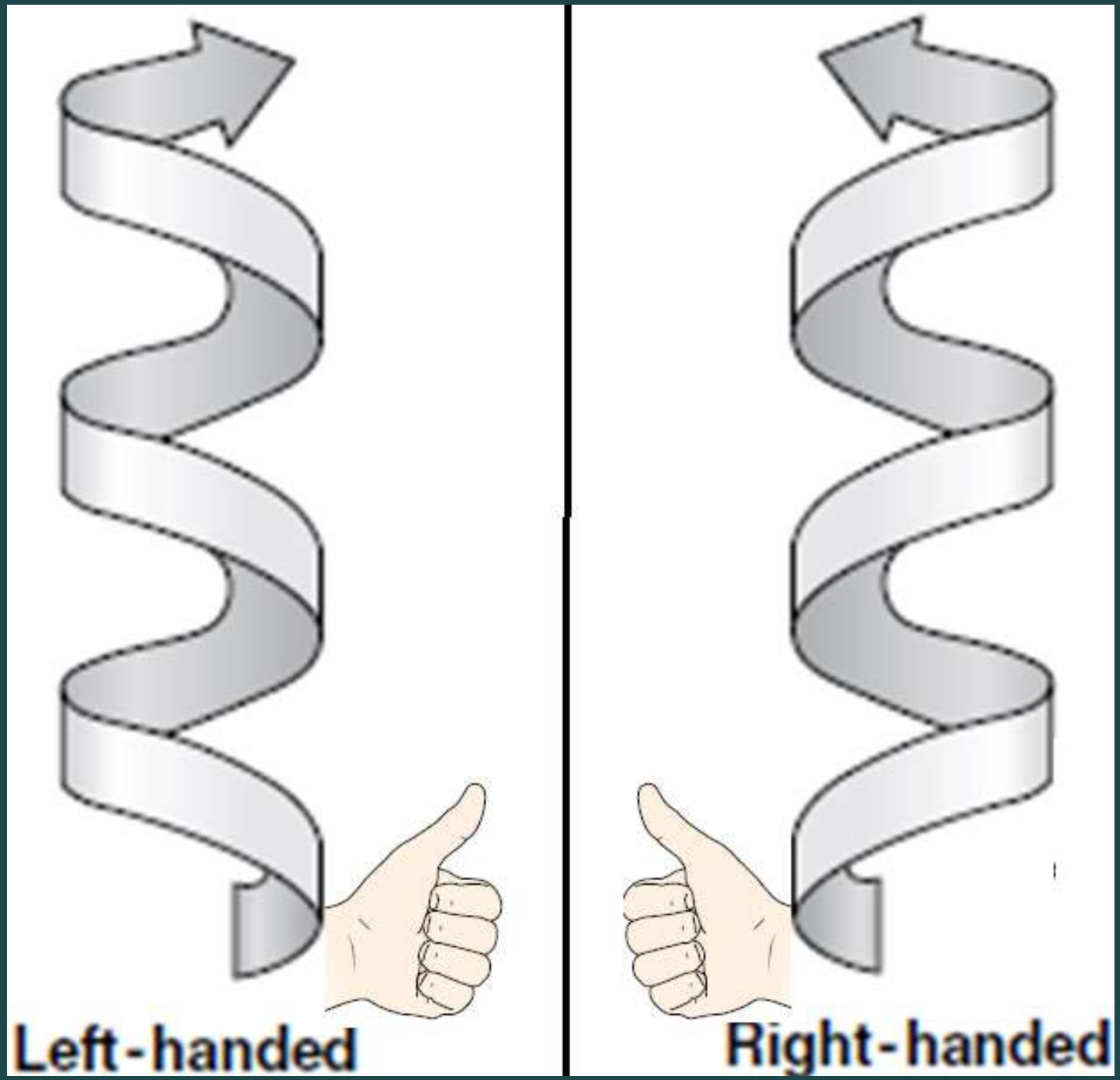
**2- RiboNucleic Acid (RNA).**

- Both DNA and RNA are polymers of **nucleotides**.
- DNA is the molecule of heredity in all cells except some viruses where RNA is the molecule of heredity.
- RNA molecules are synthesized on DNA templates and participate in protein synthesis in the cytoplasm.

# DNA

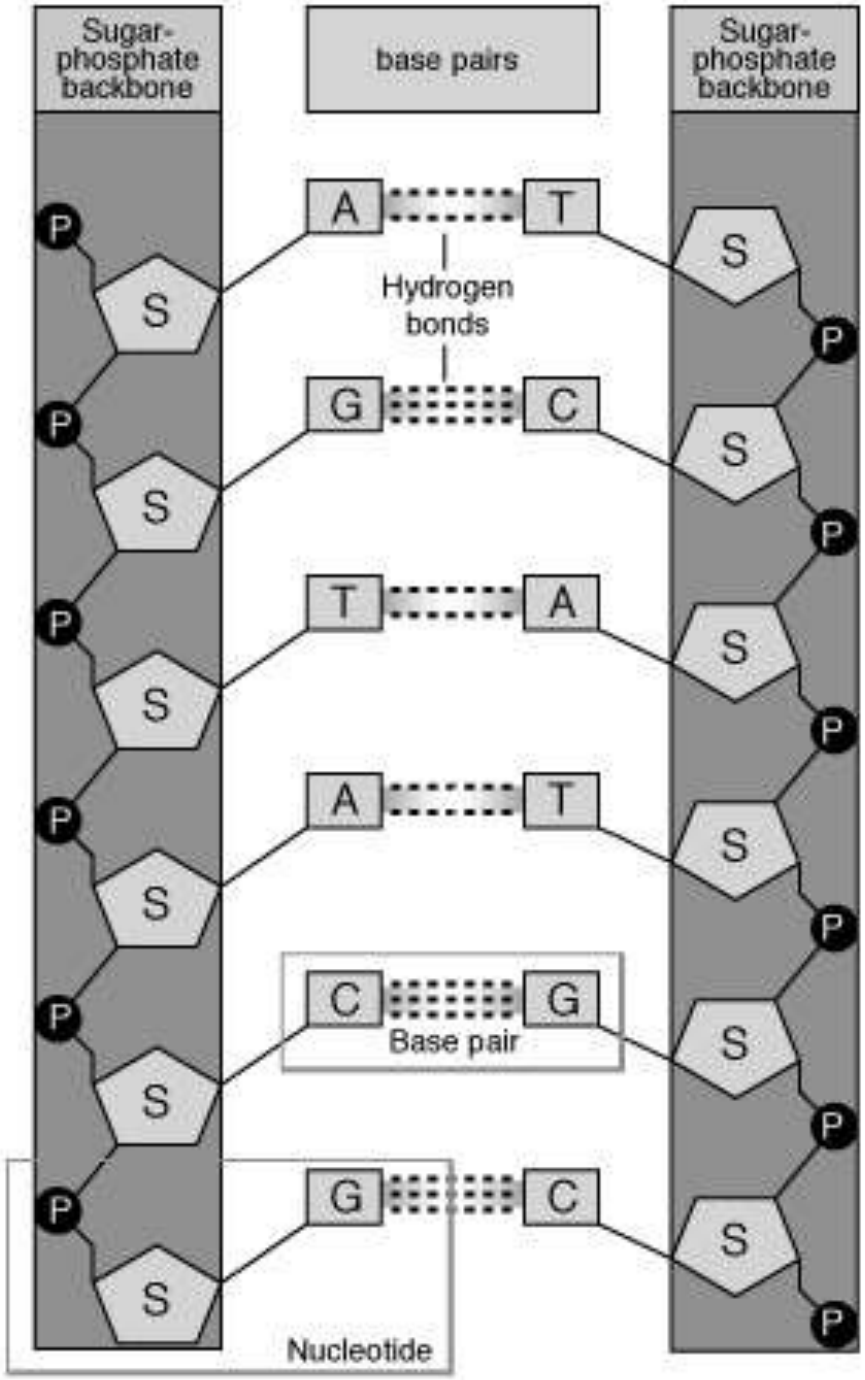
- Right-handed antiparallel double-stranded helix (spiral structure).
- Has a diameter of about 20 $A^0$ (2 nm).
- Makes complete turn every 34$A^0$ (3.4 nm)
- The distance between adjacent nucleotide is 3.4 $A^0$.
- The bases of DNA are flat and stacked above one another.

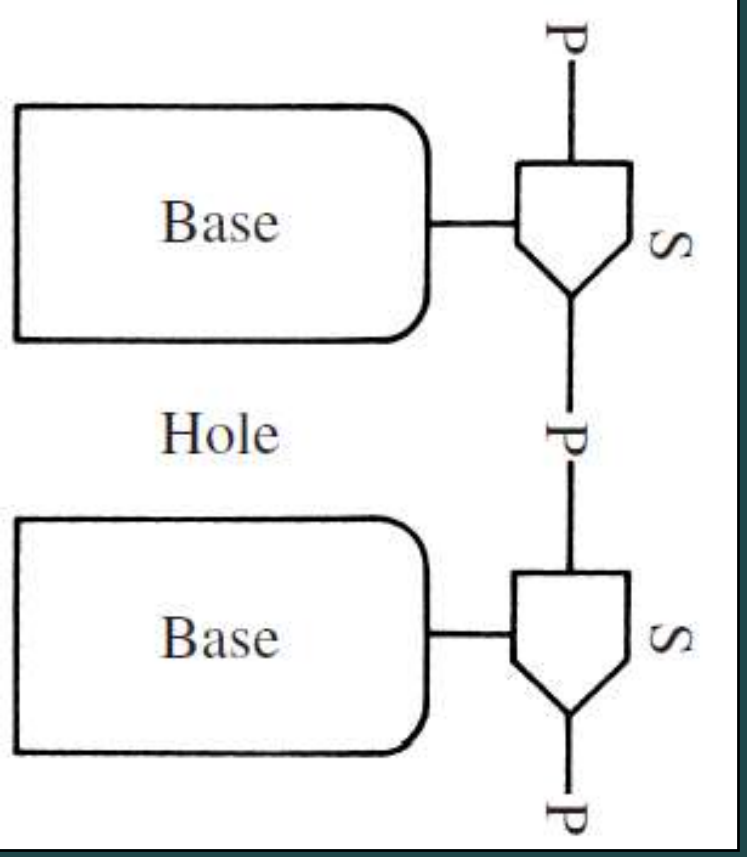


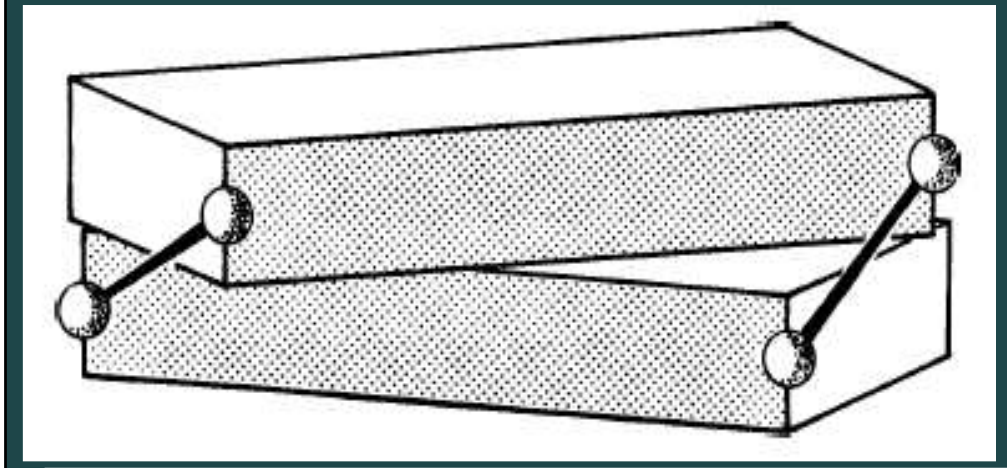

Left-handed
Right-handed

## DNA structure why it is in this shape?

The nitrogen bases are hydrophobic therefore to avoid water they are stacked in the center, while the sugar and phosphates are hydrophilic and therefore stay outside. At the end they will form a ladder shape. However the ladder shape still has holes thus to prevent water from reaching nitrogen bases through these holes the DNA twists to reduce space between the nitrogenous bases and the phosphate and sugar strands.



**Holes between DNA bases in ladder shape?**

- The distance between two sugars or phosphate is about double that of the thickness of the nitrogen bases which thus leaves a hole between the bases.

- How DNA gets rid of the ‘holes’?
- Each base pair is twisted about $36^0$ to the next base pair and this keeps the distance between the sugars and phosphate constant and at the same time get ride of the holes.
- The twisting of the two strand around one another form a double helix with a minor groove about 12 $A^0$ across and a major groove about 22 $A^0$.

# RNA

- Single strand of nucleotides
- There are three classes of RNA based on their functions:
- 1) **transfer** RNA (tRNA);
- 2) **messenger** RNA (mRNA);
- 3) **ribosomal RNA** (rRNA).

- Chemically RNA differs from DNA in two respects:
- 1- It contains ribose sugar instead of deoxyribose.
- 2- It has uracil (U) instead of thymine (T).

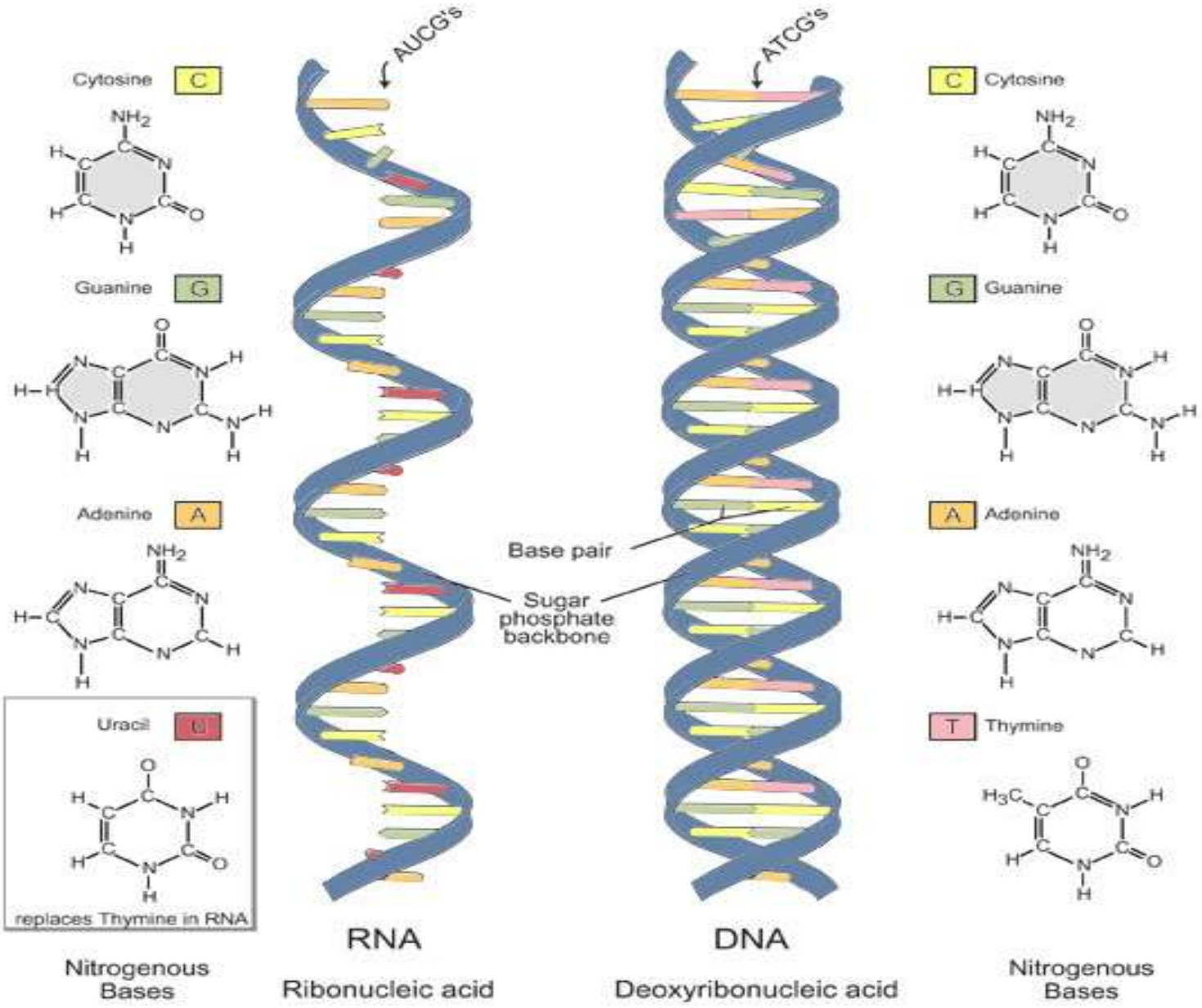
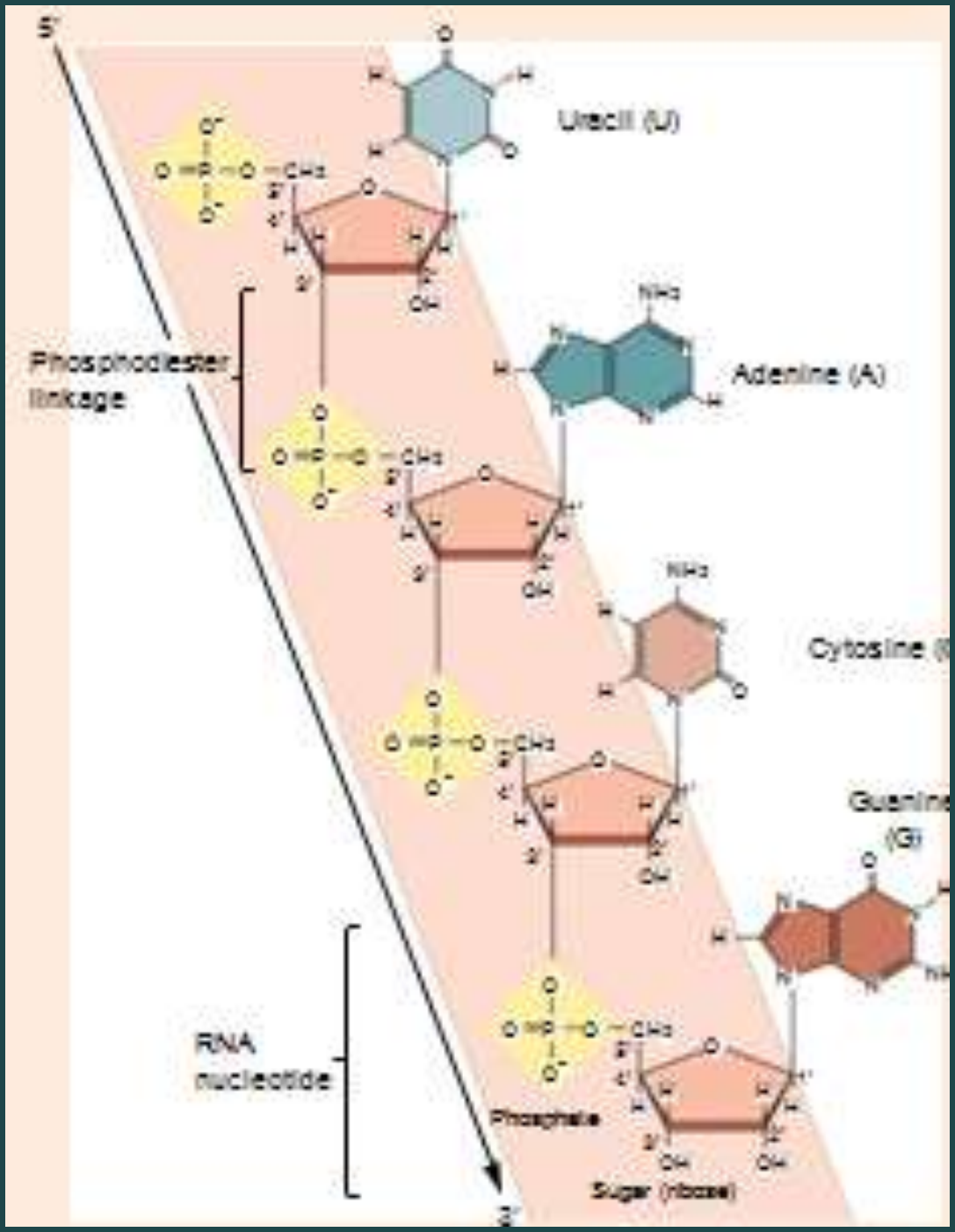



Image adapted from: National Human Genome Research Institute. Talking Glossary of Genetic Terms. Available at: www.genome.gov/Pages/Hyperion//DIR/VIP/Glossary/Illustration/rna.shtml.

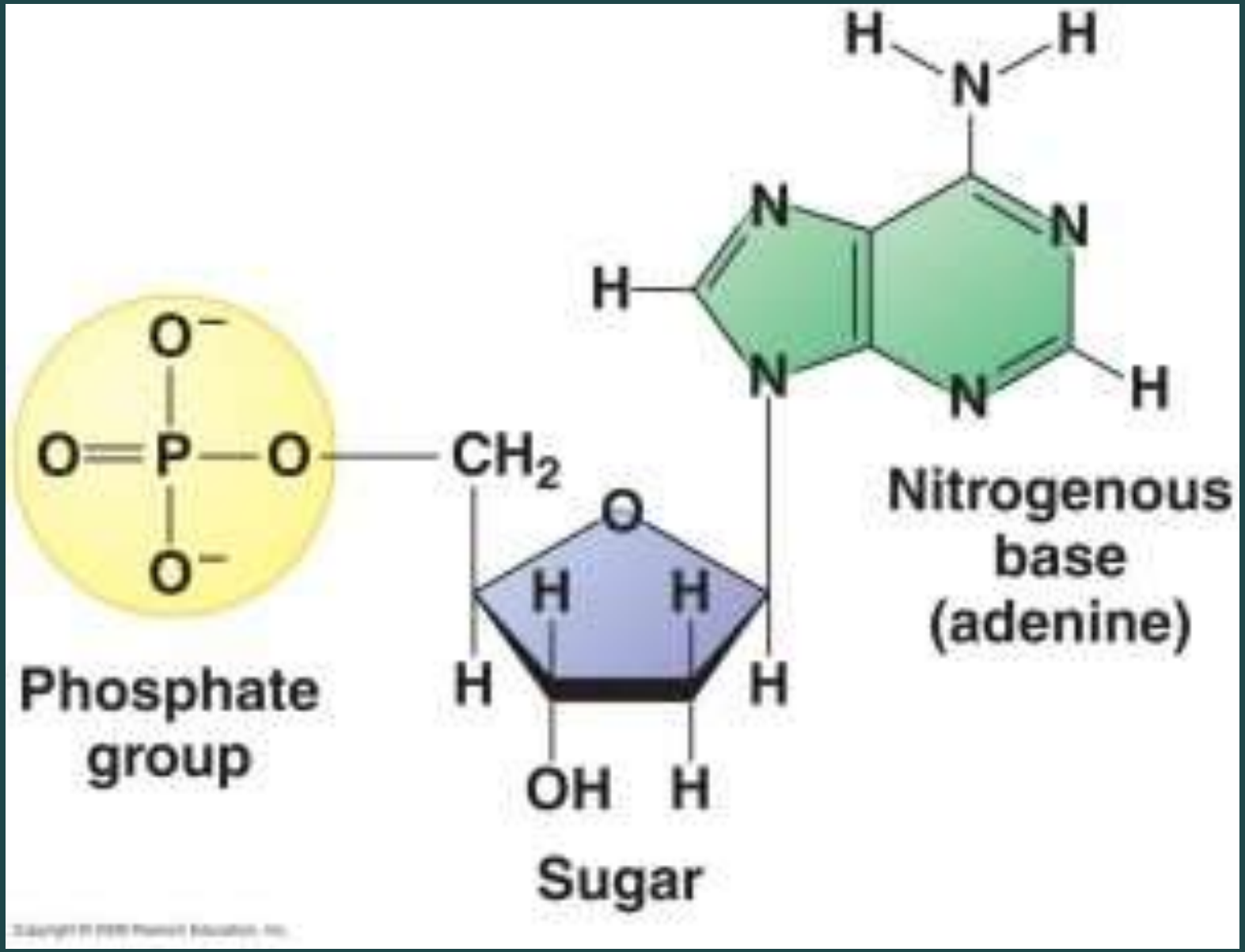

5'
Uracil (U)
Phosphodiester linkage
Adenine (A)
Cytosine (
Guanine (G)
RNA nucleotide
Phosphate
Sugar (ribose)
3'

# Nucleotides

- DNA and RNA are long, unbranched polymer of nucleotides
- **A nucleotide is made up of:**
- 1- Sugar (deoxyribose in DNA and ribose in RNA)
- 2- Phosphate
- 3- Nitrogen base.
- In DNA nitrogen bases are adenine (A), guanine (G), cytosine (C), and thymine (T), while in RNA adenine (A), guanine (G), cytosine (C), and uracil (U)).
- Nucleotides are named after the nitrogen bases present,

- The genetic information is stored in the sequence of bases
- Nucleosides are formed by joining a nitrogenous base to a sugar

# Nucleotide


H
N
H
N
N
H
N
N
H
O⁻
O
P
O
O⁻
CH2
O
H
H
H
H
OH
H
Phosphate group
Sugar
Nitrogenous base (adenine)

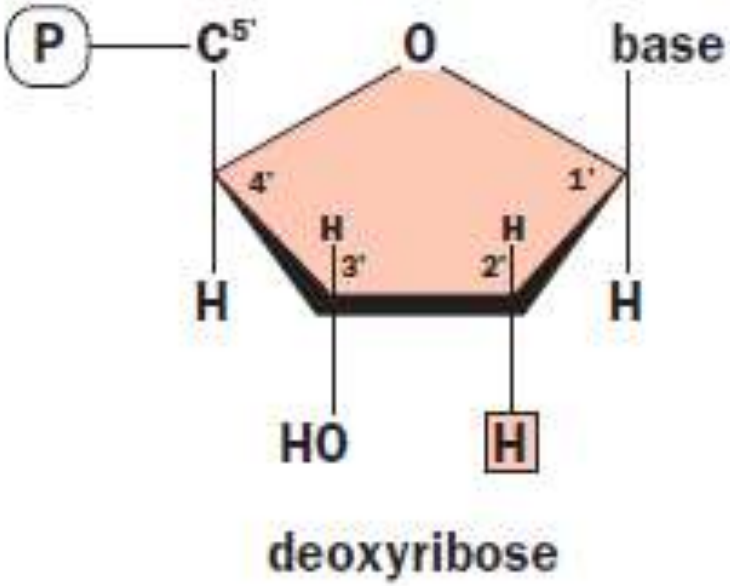
Sugar + Base ⟶ Nucleoside

Nucleoside + Phosphate ⟶ Nucleotide

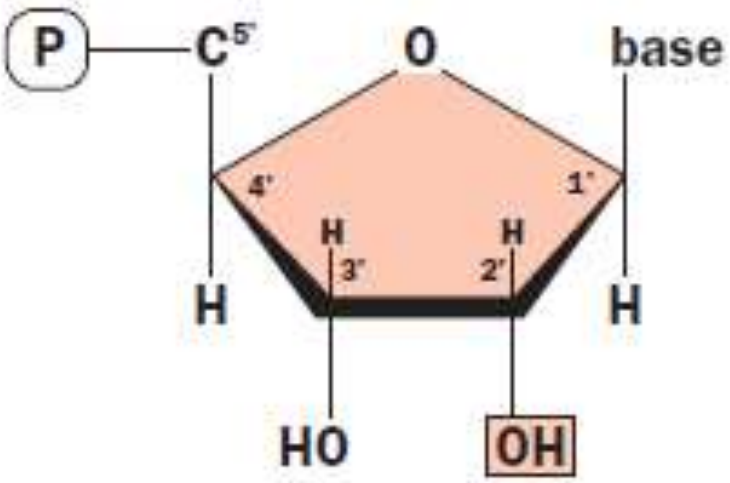
| Base | Nucleosides | Nucleotides |
|---|---|---|
| **RNA** | | |
| Adenine | Adenosine | Adenosine-5′-monophosphate (AMP) |
| Guanine | Guanosine | Guanosine-5′-monophosphate (GMP) |
| Cytosine | Cytidine | Cytidine-5′-monophosphate (CMP) |
| Uracil | Uridine | Uridine-5′-monophosphate (UMP) |
| **DNA** | | |
| Adenine | Deoxyadenosine | Deoxyadenosine-5′-monophosphate (dAMP) |
| Guanine | Deoxyguanosine | Deoxyguanosine-5′-monophosphate (dGMP) |
| Cytosine | Deoxycytidine | Deoxycytidine-5′-monophosphate (dCMP) |
| Thymine | Deoxythymidine | Deoxythymidine-5′-monophosphate (dTMP) |

# 1- Ribose ($C_5H_{10}O_5$) & Deoxyribose ($C_5H_{10}O_4$) Sugar

- The carbon atoms on the sugar ring are numbered 1’ (one prime) to 5’ to distinguish them from atoms in the bases. The ribose in DNA differs from that of RNA by the absence of oxygen at the carbon atom number 2 and is thus 2-deoxy-β-D-ribose in DNA while in RNA β-D-ribose.

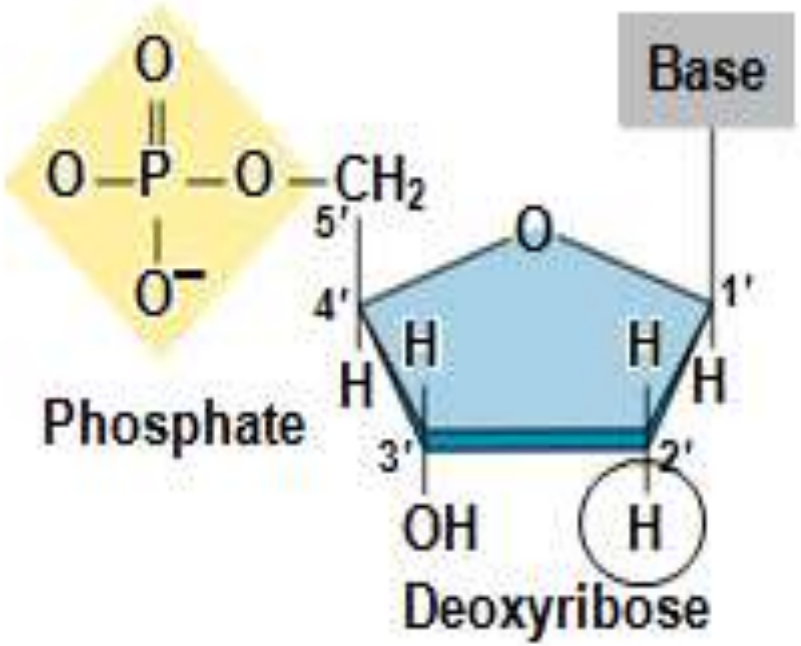



deoxyribose

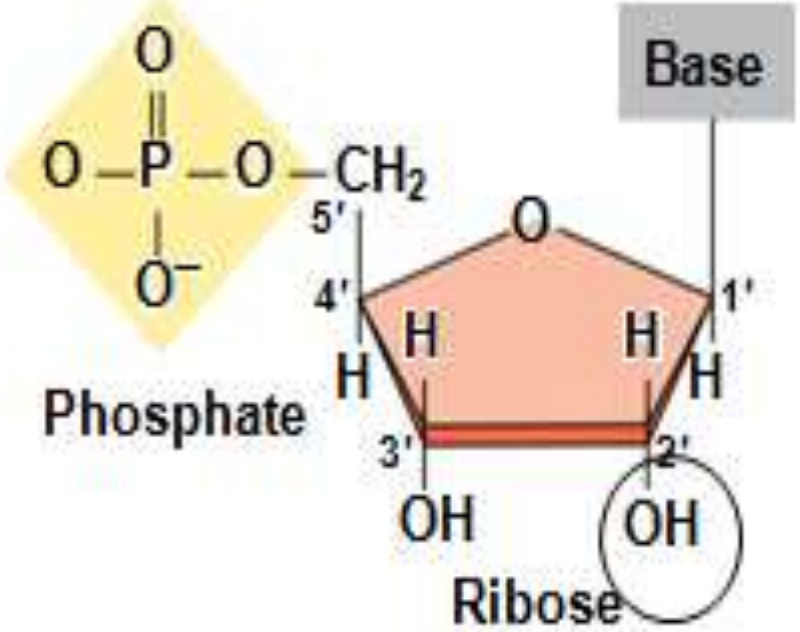



ribose

# Nucleotides of DNA and RNA




DNA nucleotide




RNA nucleotide

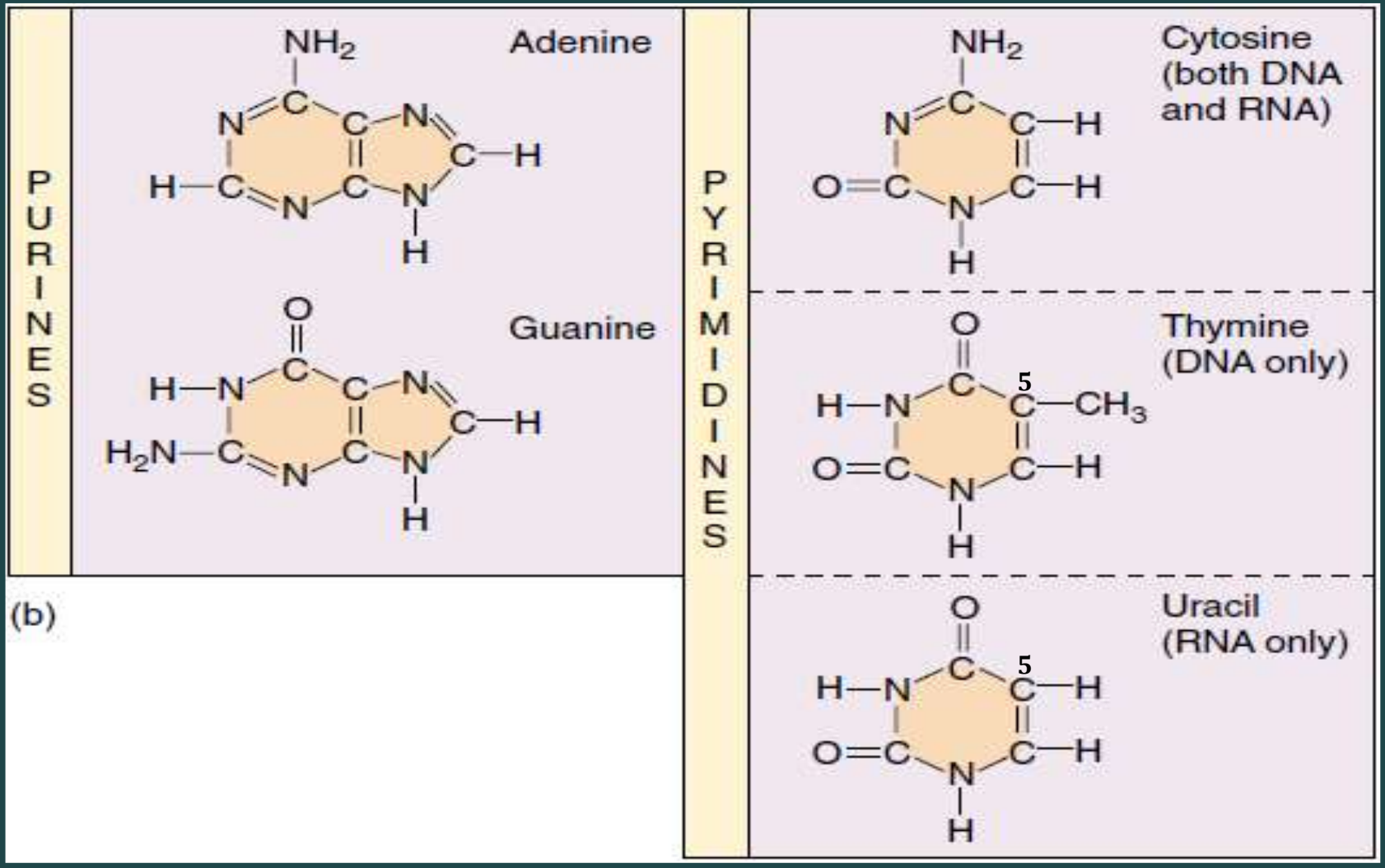
- **2- Nitrogen bases**
- Because of their nitrogen content and basic qualities (hydrophobic) they are known as **nitrogenous bases.**
- The organic bases are of two general types: **purines** and **pyrimidines**.

**A. Purines**

- The purines are **adenine** (A) and **guanine** (G).
- Each purine consists of a six-sided ring attached to a five-sided ring

**B. Pyrimidines**

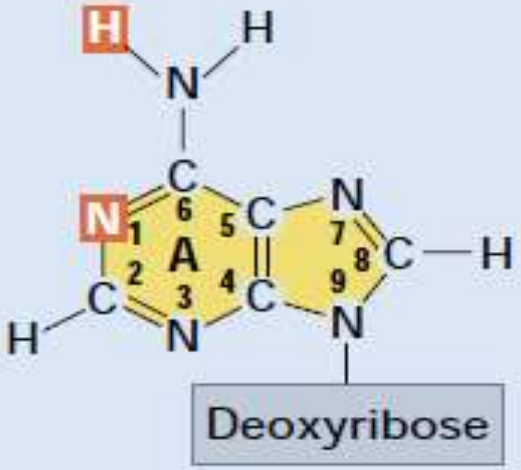
- The pyrimidines are **cytosine** (C), **thymine** (T), and **uracil** (U).
- Each pyrimidine consists of a six-sided ring only.
- Thymine is found primarily in DNA and uracil is found only in RNA.
- Uracil differs from thymine by lacking a methyl group on its C 5



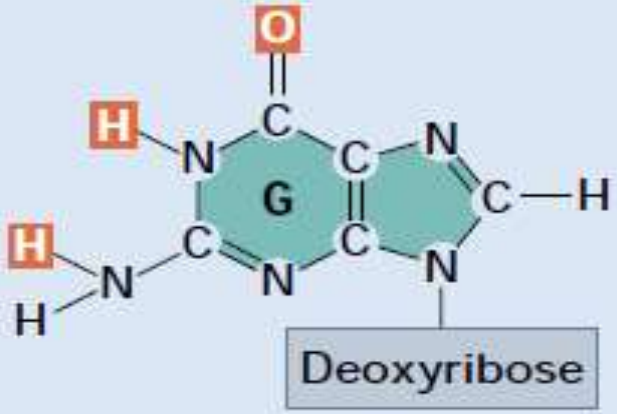

Nitrogen bases are hydrophobic with a hydrophilic edge. These hydrophilic edges are the -C=O and -N-H groups that can act as H-bond acceptors and donors. The hydrophilic edges of the bases interact in a very specific way A=T and G≡C, and thus this will leave the nitrogen bases as pure hydrophobic in the centre of the DNA.

## Purines

### Adenine

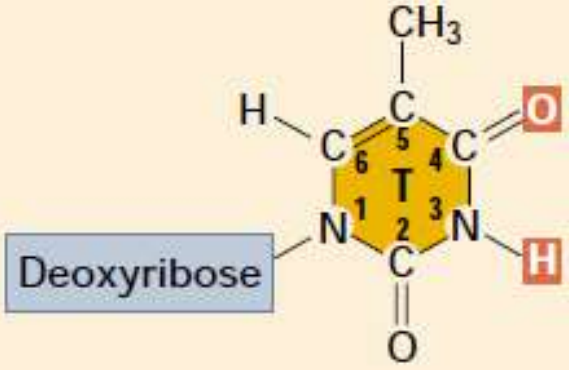



### Guanine

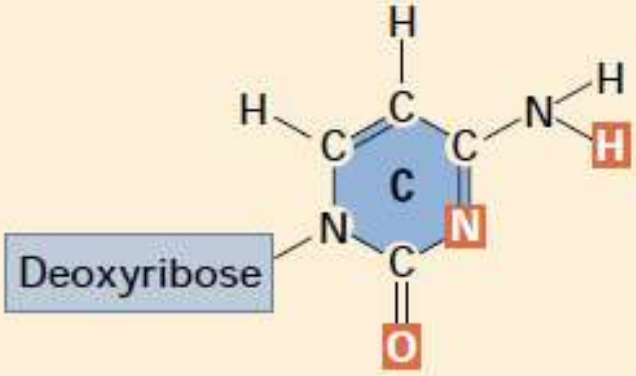



## Pyrimidines

### Thymine

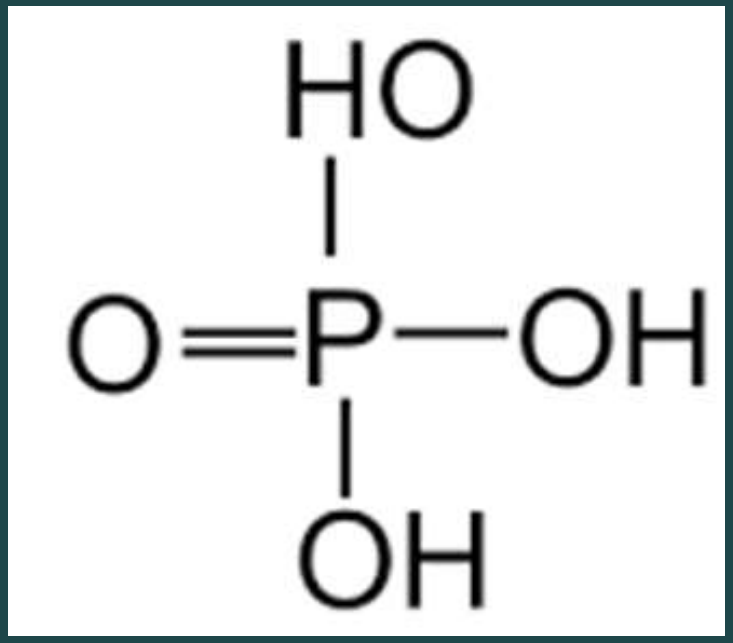



### Cytosine

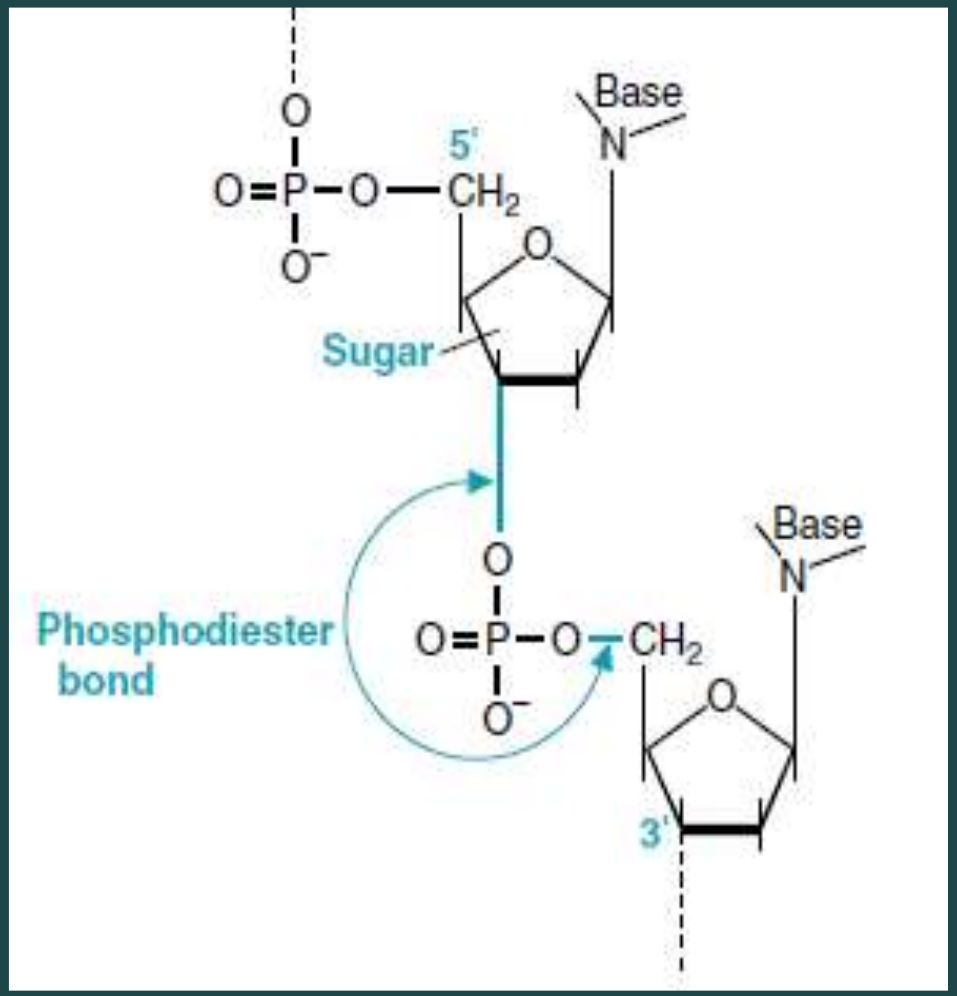


- **3- Phosphate group**
- A phosphoric acid **($H_3PO_4$)** is bounded to one oxygen by a double bond and three hydroxyl groups (–OH).
- Two of the hydroxyl groups can form covalent bonds, phosphodiester bonds, with the sugar hydroxyl groups by splitting out water.
- The third -OH group on the phosphate (its pKa about 2 is less than our pH about 7.4) dissociates a hydrogen ion (H+ ions) and leaving negatively charged oxygens at physiological pH.
- In this form, the structure is referred to as phosphate.
- Therefore, phosphate is a negatively charged which gives the polymer its acidic property and promoting their attraction to positively charged histone proteins that partially neutralized this negative charges
- because of the phosphate charges both DNA and RNA are negatively charged.

# Phosphoric acid

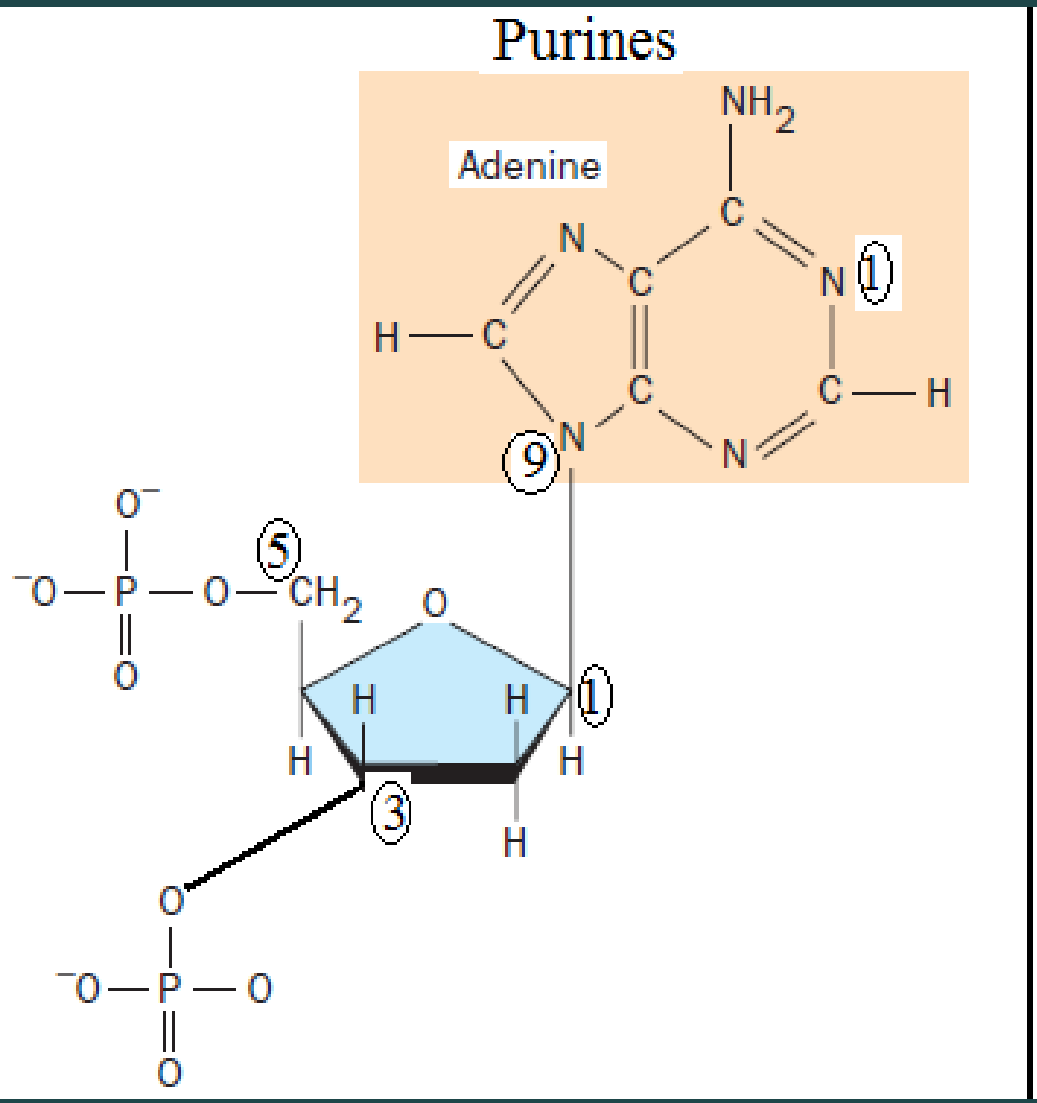


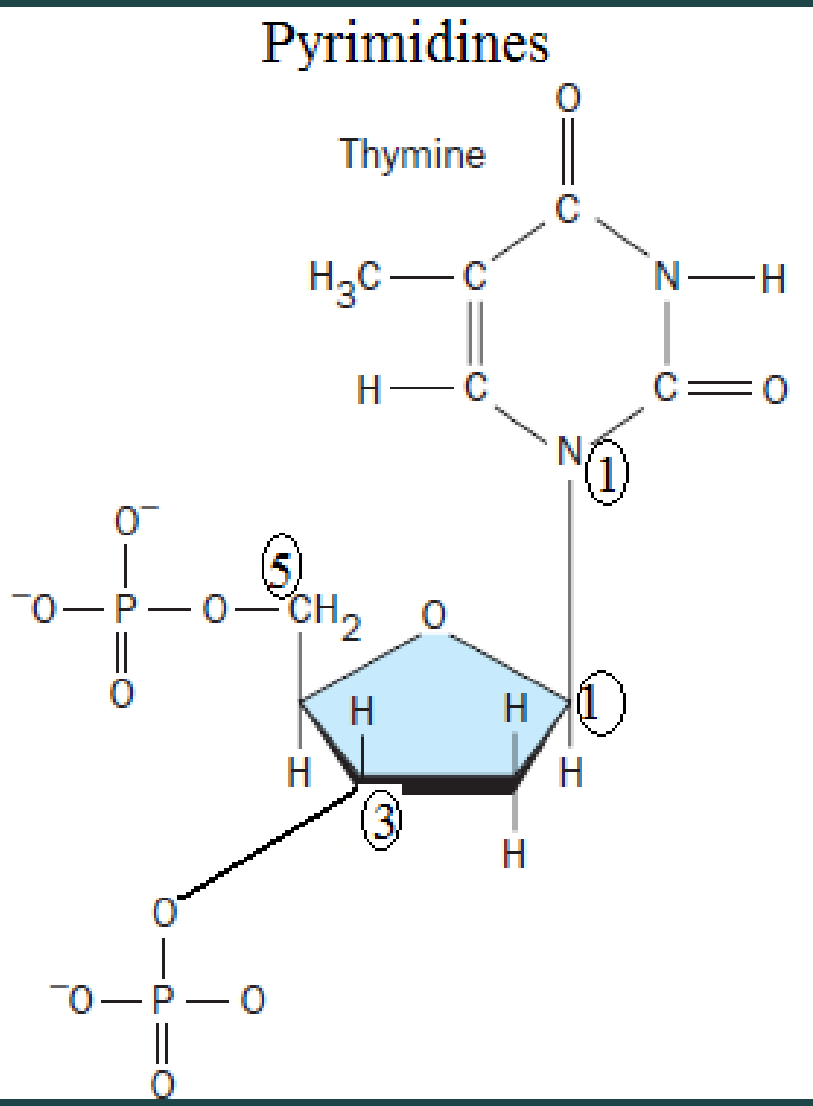

$H_3PO_4$



# • Bondings in DNA

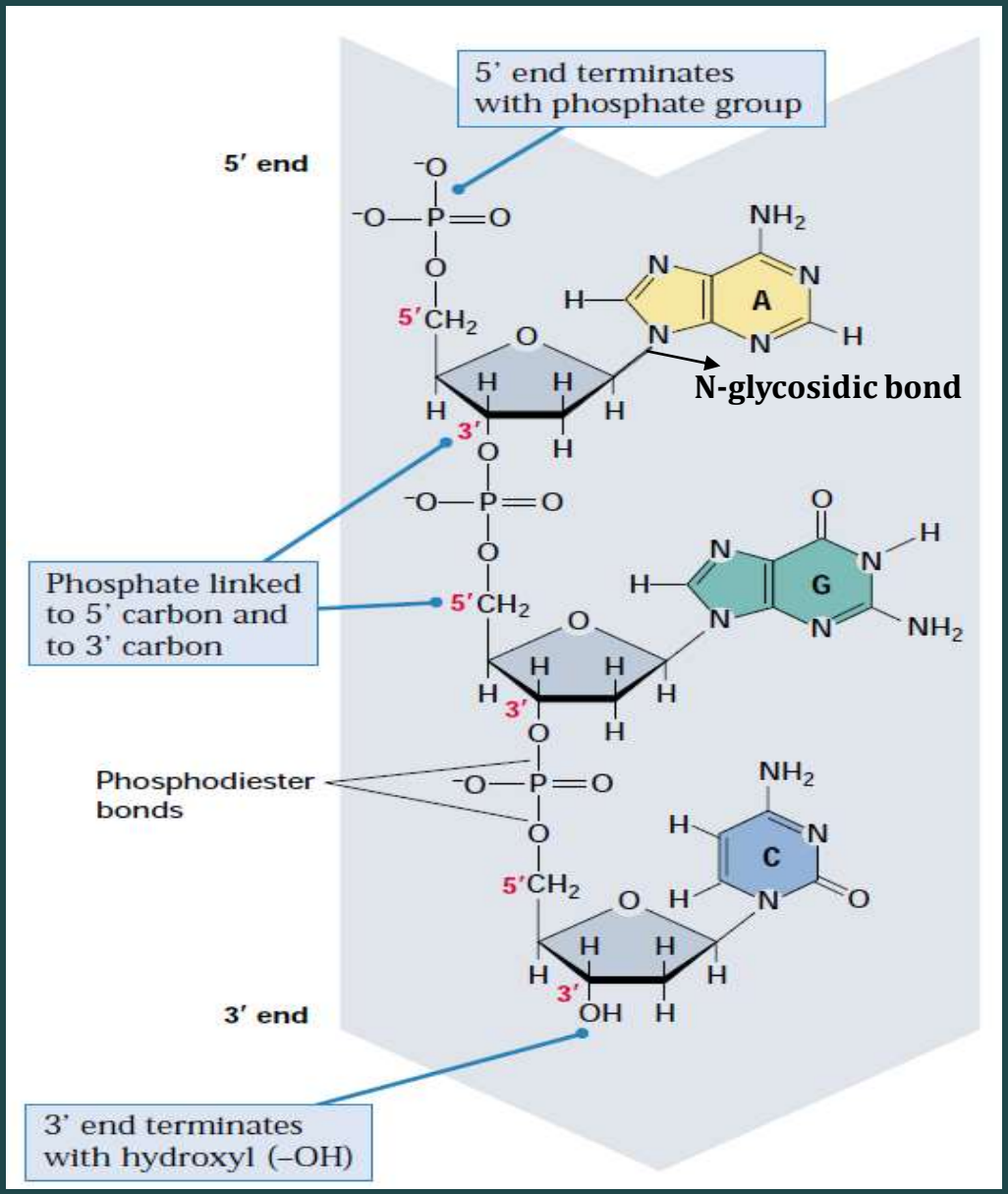
- The nitrogen atom in position 9 of a purine or in position 1 of a pyrimidine is bound by N-glycosidic bond to the carbon number 1 of the sugar.
- On the other hand, the phosphate group links the 3’ end of one nucleotide to the 5’ end of the next nucleotide through phosphodiester bonds.
- This gives the sugar-phosphate backbone directionality a 5’ end and 3’ end.
- In DNA the direction of the nucleotides in one strand is opposite to their direction in the other strand and thus called antiparallel.
- The 5' end has a terminal phosphate group and the 3' end has a terminal hydroxyl group.
- **Sequences are written and read in the 5’ to 3’ direction (from left to right); for example, the sequence AUG is assumed to be (5’)AUG(3’)**

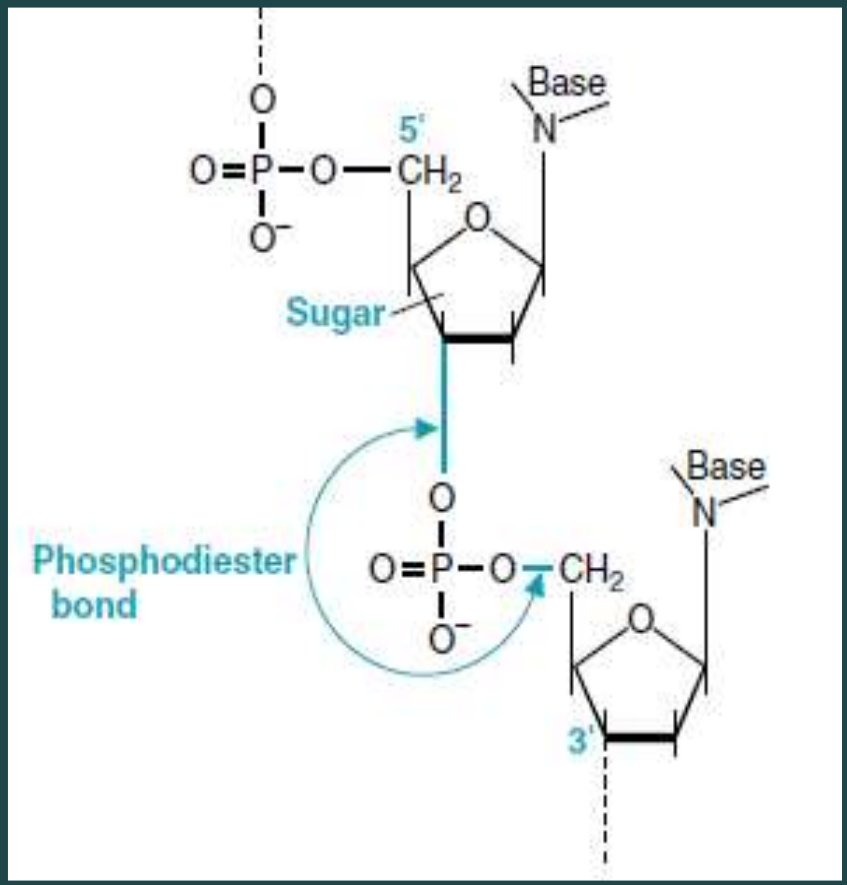


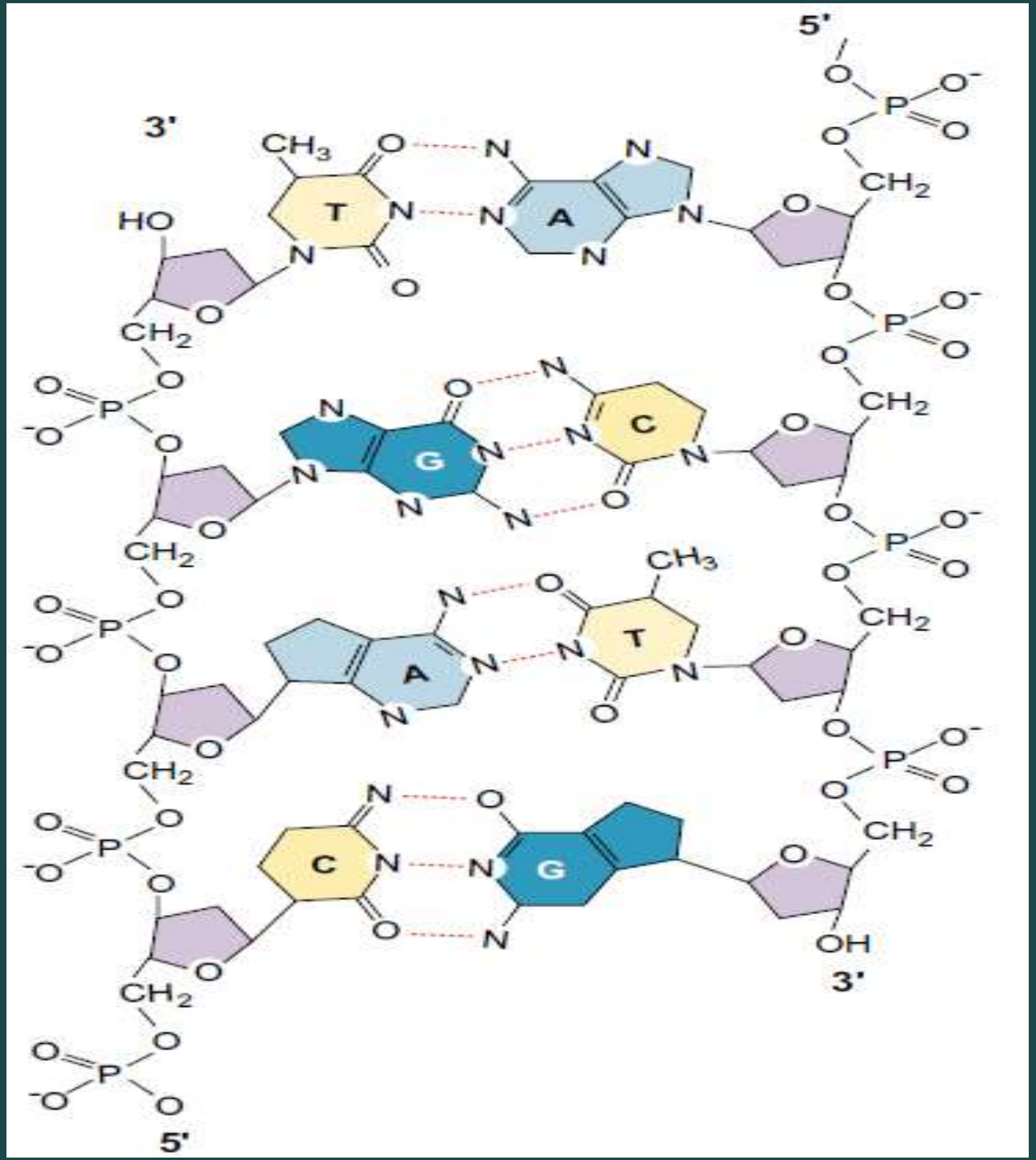



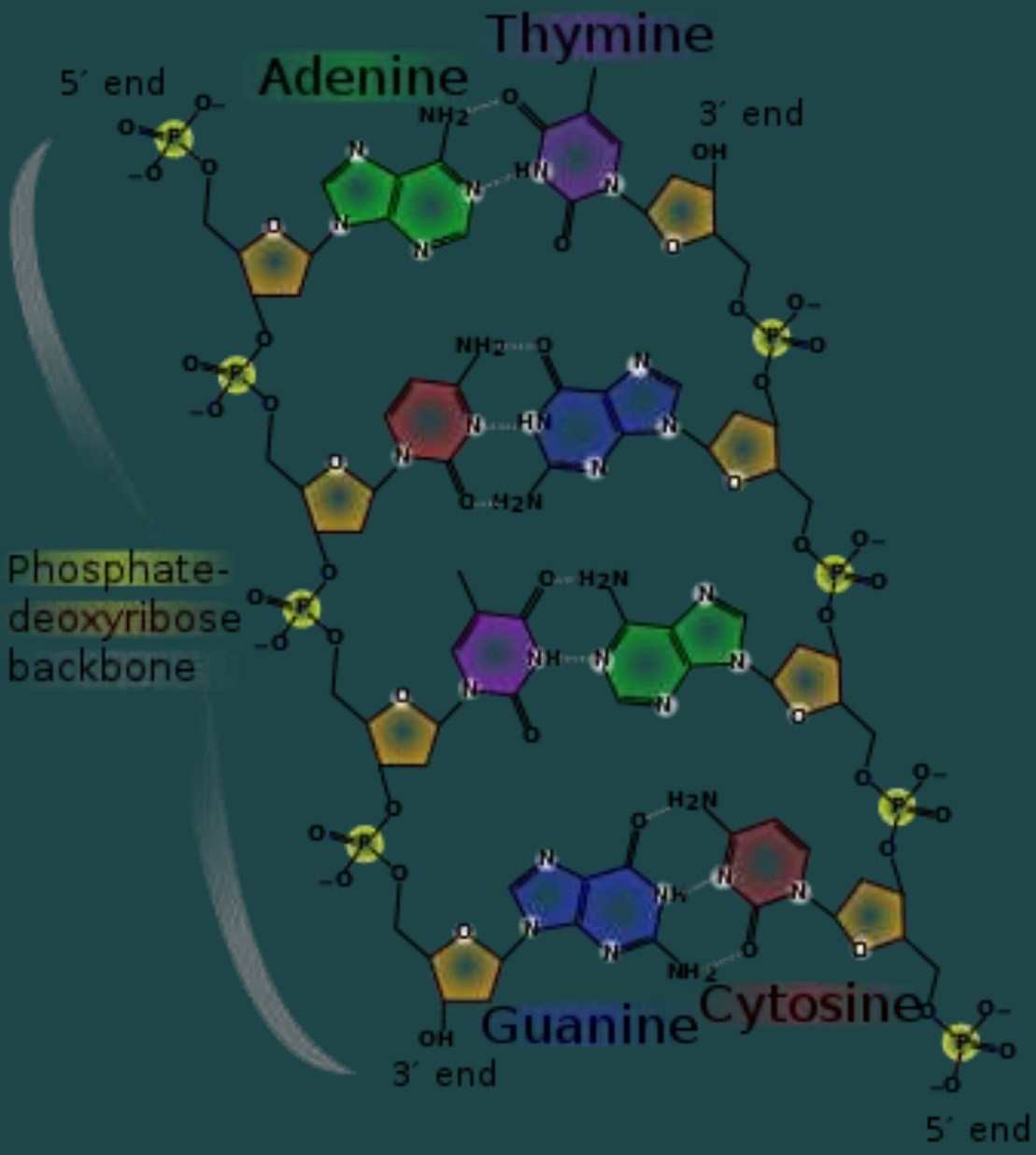

In a nucleotide, the nitrogen atom in position 9 of a purine or in position 1 of a pyrimidine is bound to the carbon number 1 of the sugar by N-glycosidic bond.

# Covalent bonds in DNA



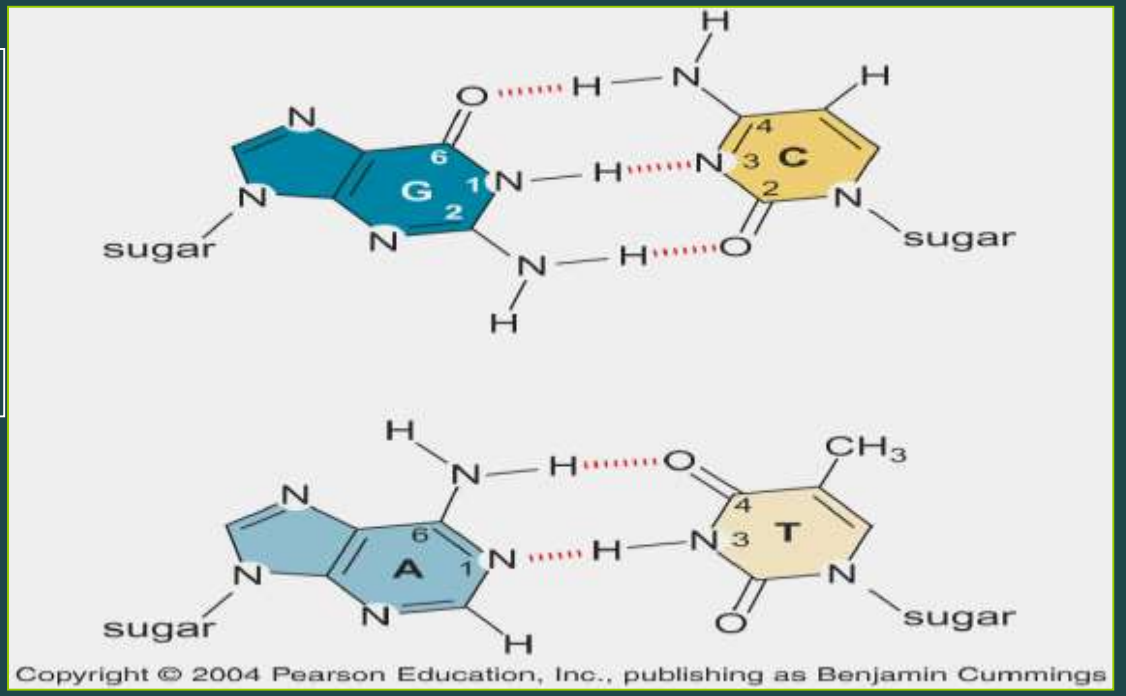

5'
3'
T
A
G
C
A
T
C
G
3'
5'

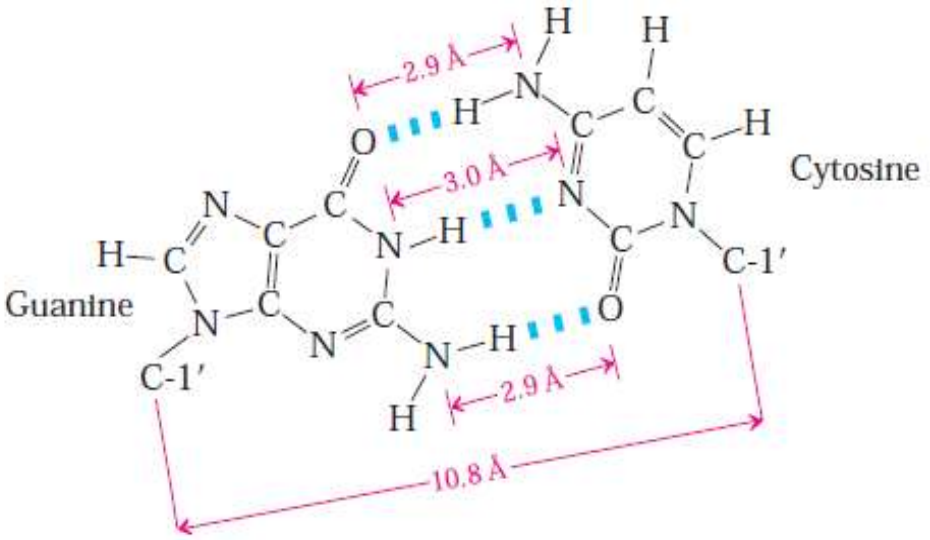

Thymine
Adenine
5′ end
3′ end
Phosphate-
deoxyribose
backbone
Guanine
Cytosine
3′ end
5′ end

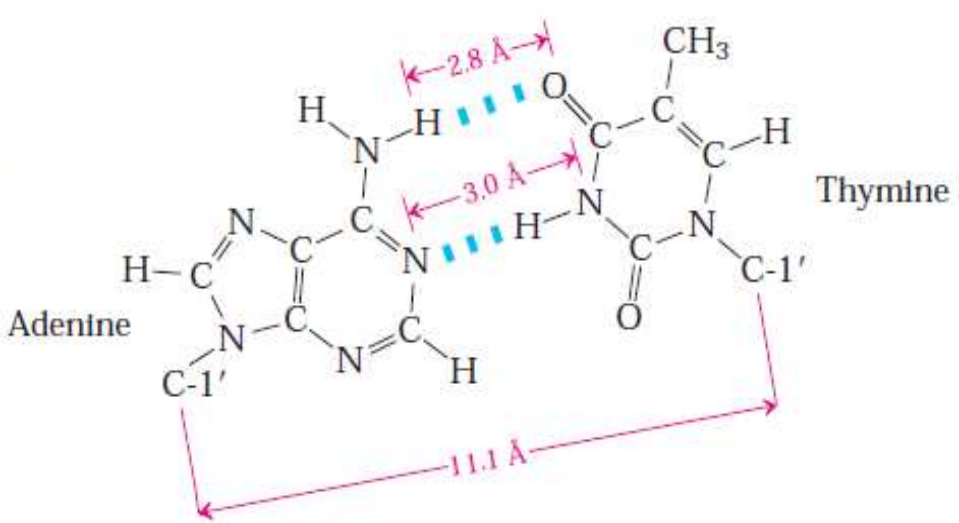
# Base Pairs and bonds in DNA

- Adenine (A) specifically binds to thymine (T) by two hydrogen bonds (A=T) and
- Cytosine (C) specifically binds to guanine (G) by three hydrogen bonds (C≡G)
- The human genome contains approximately 3 billion of these base pairs, which reside in the 23 pairs of chromosomes within the nucleus of all our cells.

**Chargaff's Rule: in any sample of double-stranded DNA, the amount of adenine equals the amount of thymine, the amount of guanine equals the amount of cytosine, and the total amount of purines equals the total amount of pyrimidines.**

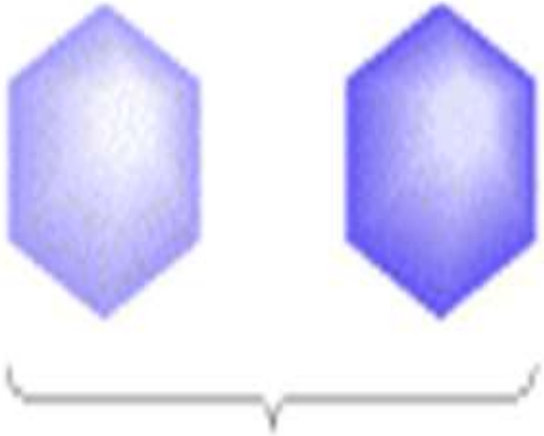




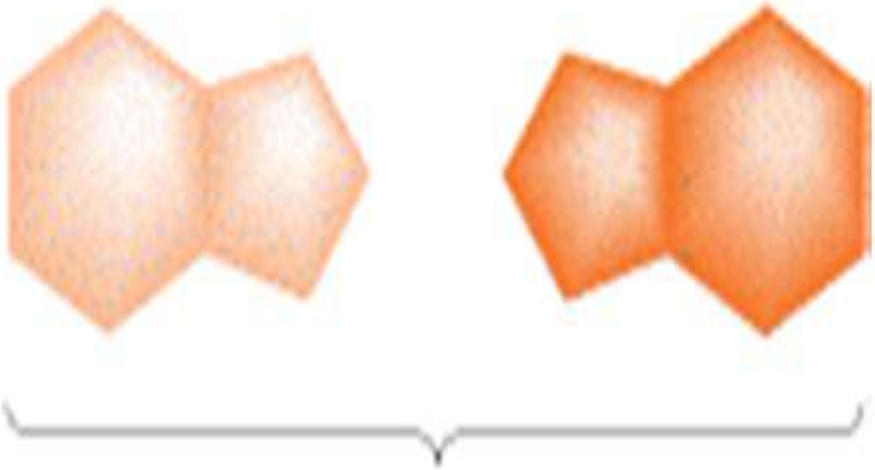



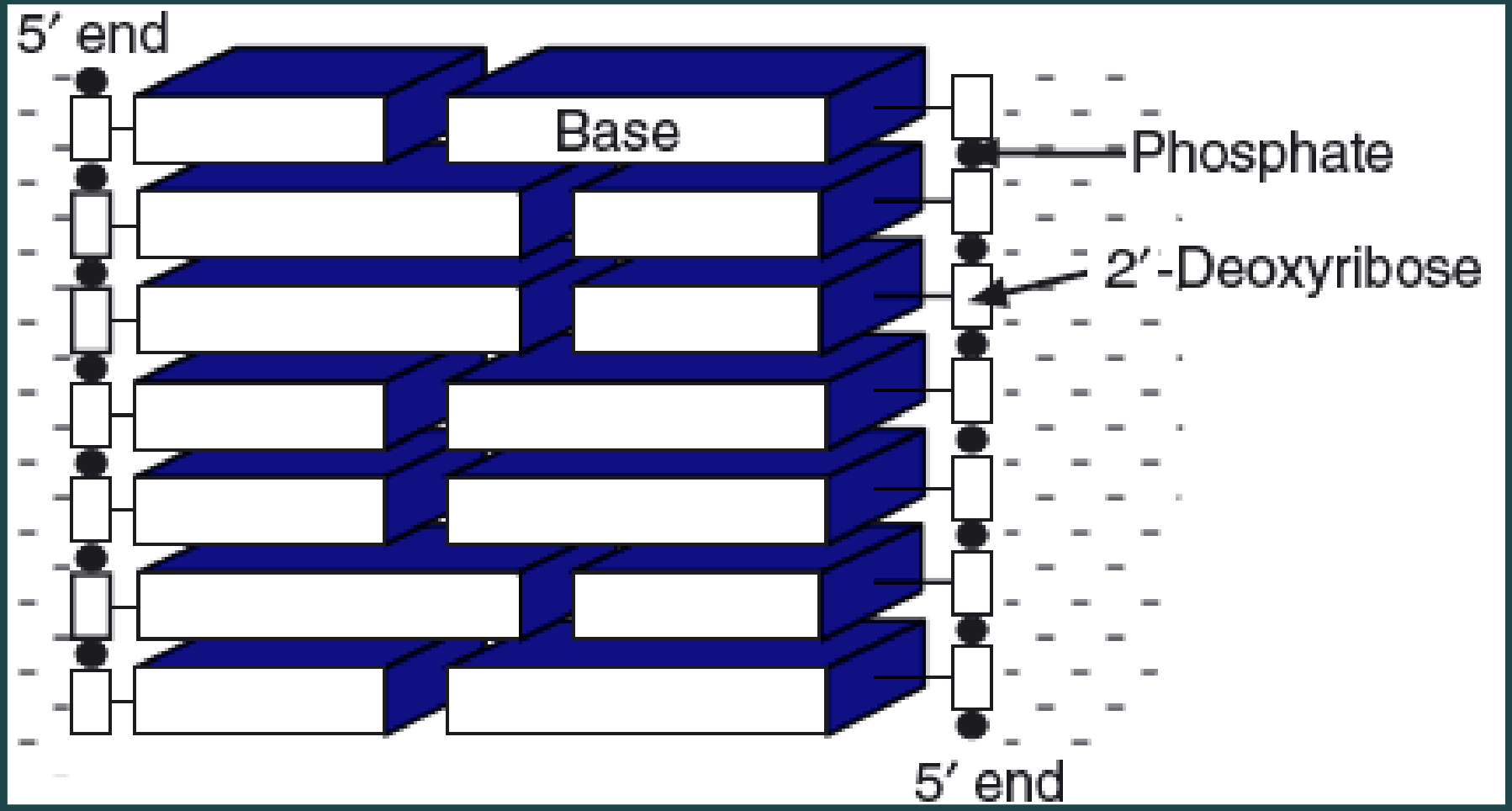

- The A-T and G-C bonds are the most stable of the various base-pairs on energetic grounds.
- These two base pairs have essentially identical dimensions within the double helix structure of DNA and thus avoiding partnership of purine-purine which will be too wide or pyrimidine-pyrimidine which is too narrow. This means that these are also the only two base pairs which will properly fit into the double helix that is the outside diameter of the double helix will be uniform over its length
- Hydrogen bonds are weak they can be broken and rejoined relatively easily

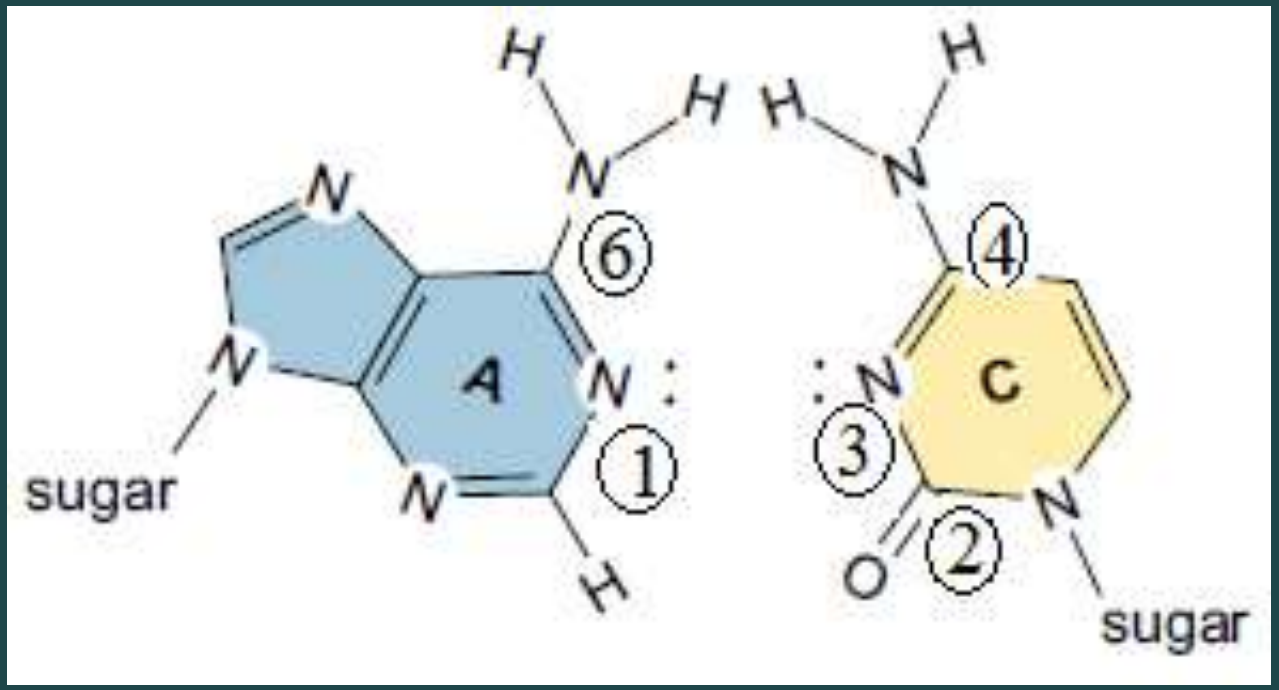
Pyrimidine + pyrimidine: DNA too thin



Purine + purine: DNA too thick




5′ end
Base
Phosphate
2′-Deoxyribose
5′ end




**Suppose we tried to pair an adenine with a cytosine.**

1- Then we would have a hydrogen bond acceptor (N1 of adenine) lying opposite a hydrogen bond acceptor (N3 of cytosine).

2- Likewise, two hydrogen bond donors, the NH2 groups at C6 of adenine and C4 of cytosine, would lie opposite each other.

•Thus, an A:C base pair would be unstable

# The DNA Double Helix Is a Stable Structure

**First**, Hydrogen bond

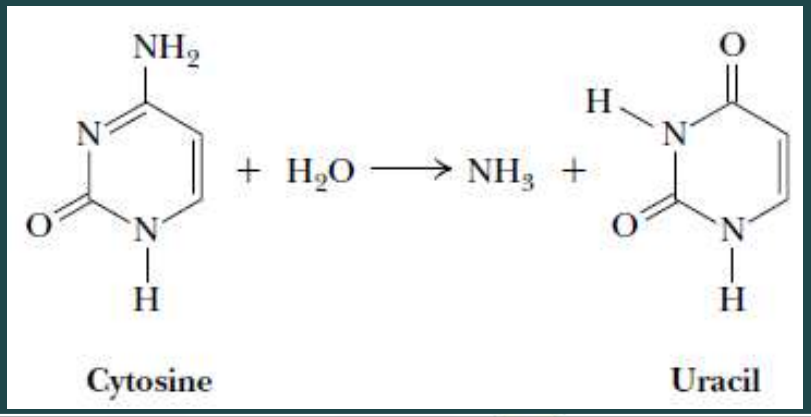
- The two strands of DNA are held together by H-bonds
- Polar atoms in the sugar–phosphate backbone form external H bonds with surrounding water molecules.

**Second**, the negatively charged phosphate groups are all situated on the exterior surface of the helix in such a way that they have minimal effect on one another and the phosphates interact with positively charged (magnesium, potassium, or sodium) ions and with positively charged histone proteins.

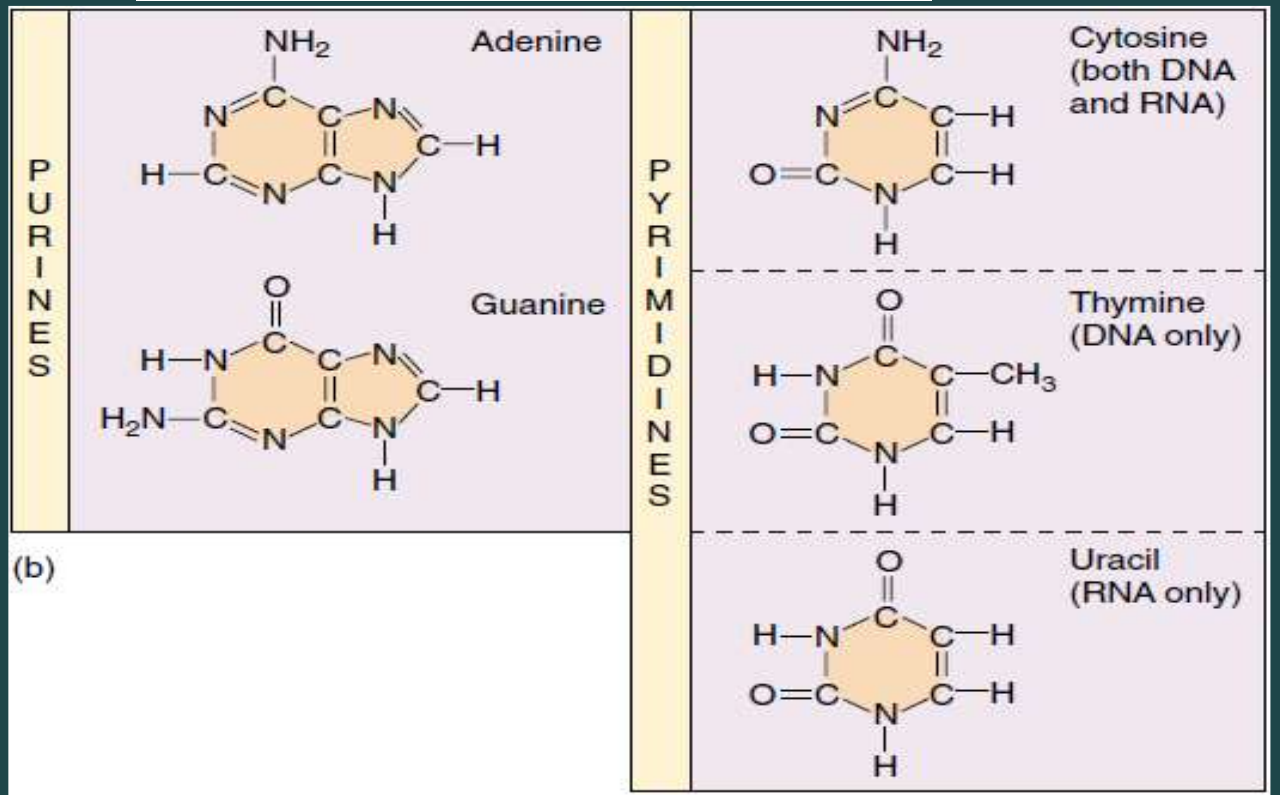
**Third**, Stacking interactions which involve hydrophobic interaction and Van der Waals forces in the core of the helix. Base stacking helps to minimize contact of the bases with water, and base-stacking interactions are very important in stabilizing the three-dimensional structure of nucleic acids.

# Significance of chemical differences between DNA and RNA

## 1- DNA contains thymine instead of uracil



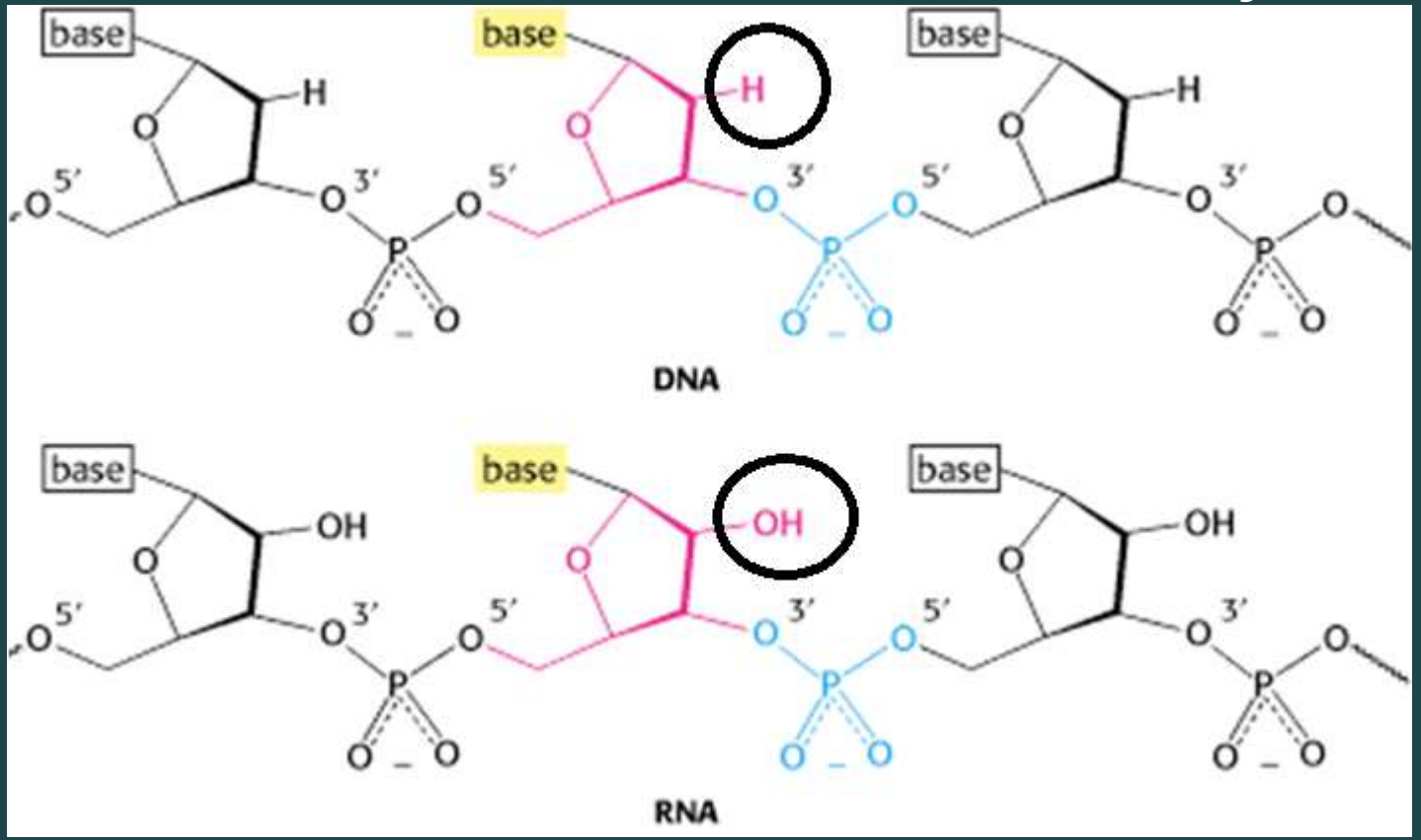

Deamination of cytosine forms uracil




Thymine has methyl group to distinguish itself from uracil

# 2- DNA contains 2-deoxyribose sugar instead of ribose sugar.

- The absence of the 2’-hydroxyl group in DNA further increases its resistance to hydrolysis.



# Heat and alkali effect on nucleic acid

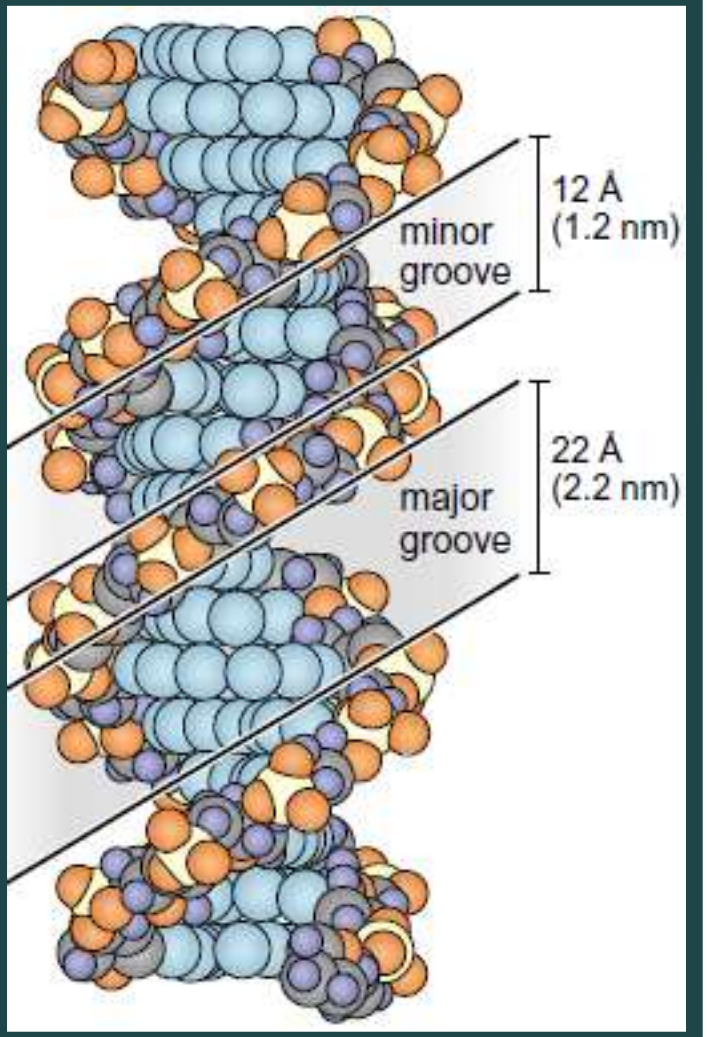
- In DNA both alkali and heat cause
- The two strands of the DNA helix to separate (denature).
- It does not break the phosphodiester bonds.

- In RNA alkali causes the cleavage of the phosphodiester bonds.

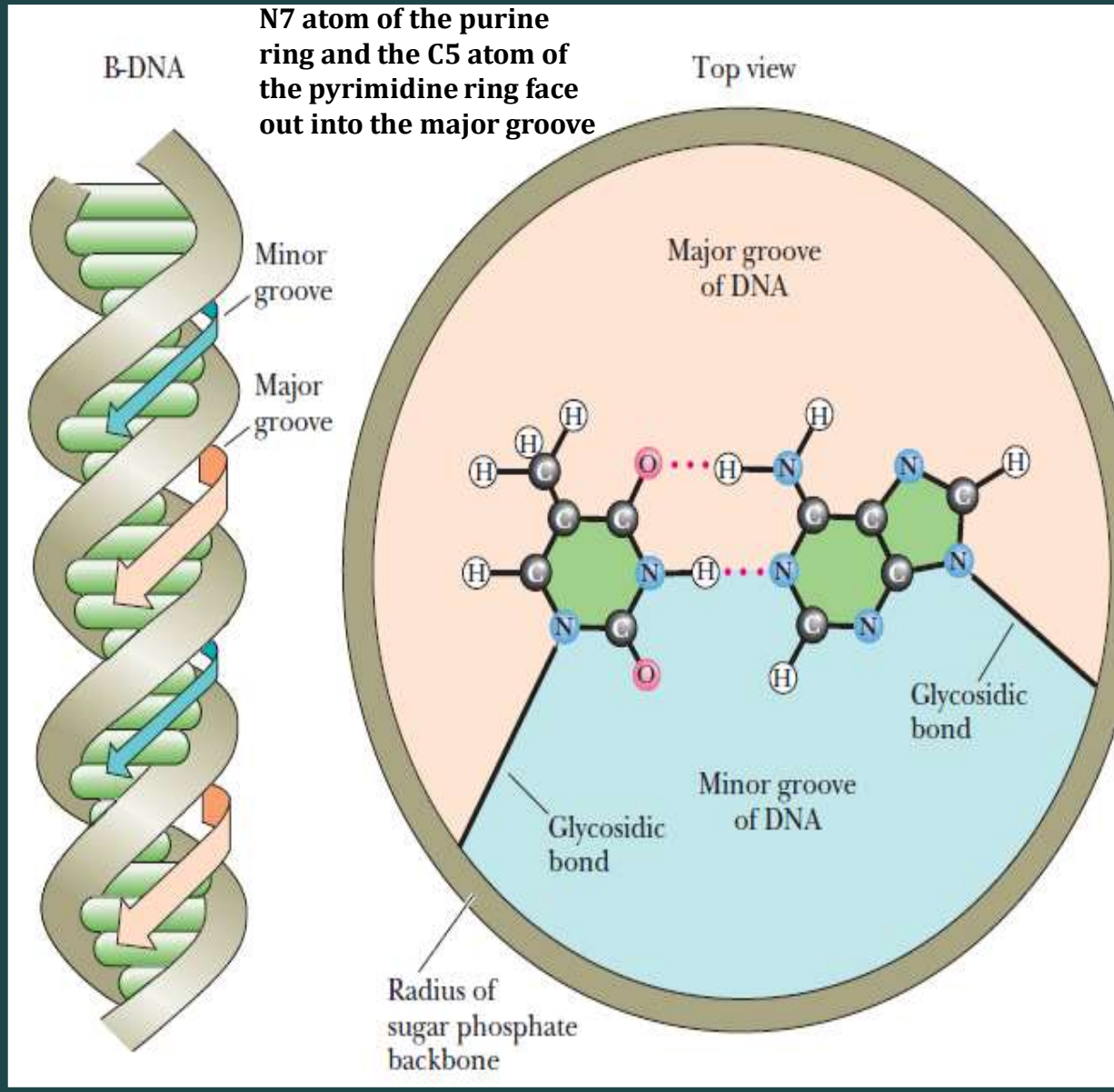
- Therefore, alkali is used to remove RNA from DNA and to separate DNA strands before, or after, electrophoresis on polyacrylamide or agarose gels.

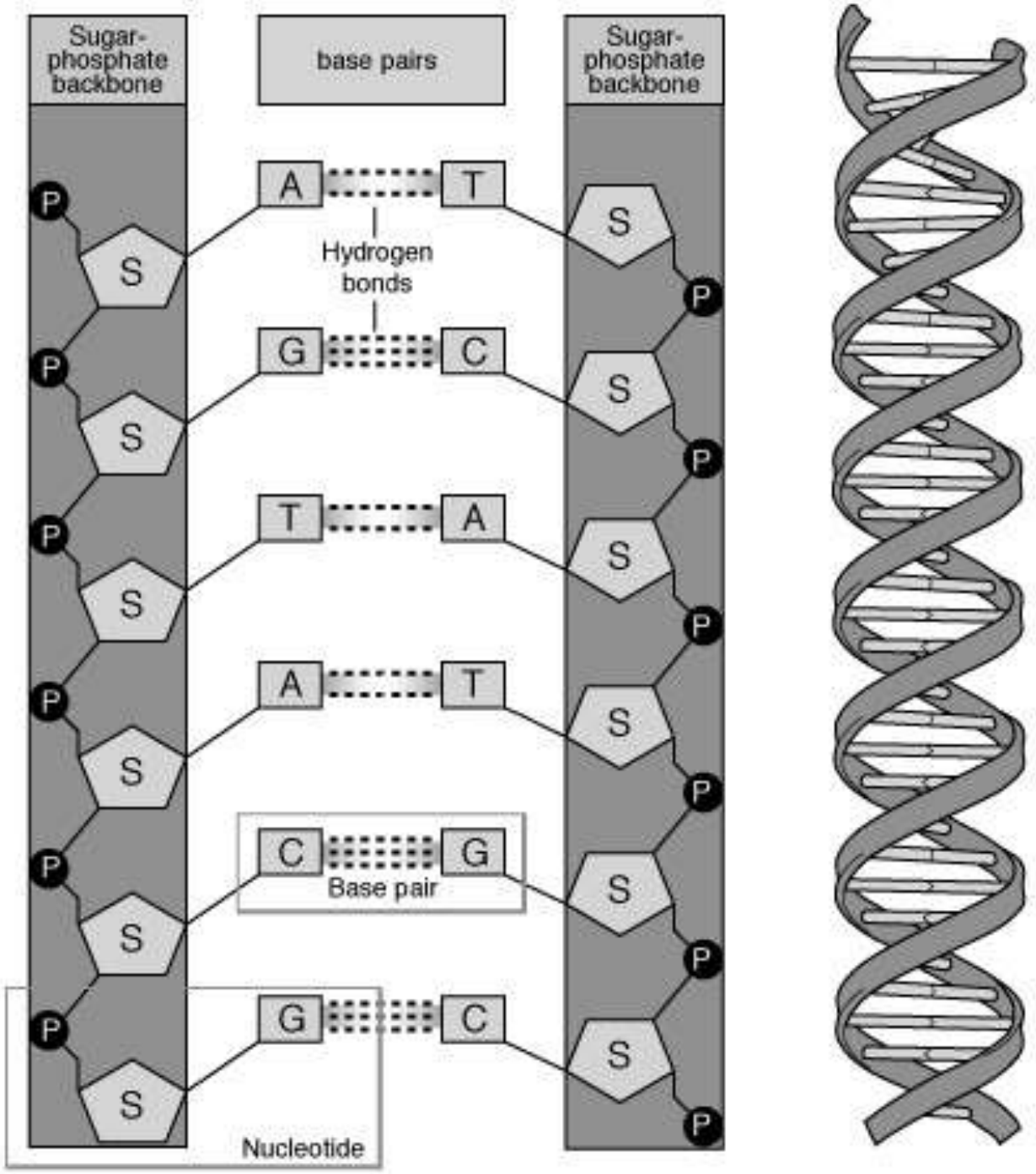
- The separation of DNA strands is called melting $\mathbf{T_m}$.
- The $T_m$ is defined as the temperature in degrees Celsius, at which 50% of all molecules of a given DNA sequence are hybridized into a double strand, and 50% are present as single strands.
- If the temperature is slowly decreased, complementary single strands can realign and base-pair, re-forming a double helix essentially identical to the original DNA. This process is known as renaturation, reannealing, or hybridization.
- Hybridization is used extensively in research and clinical testing.

# Grooves (space between strands)

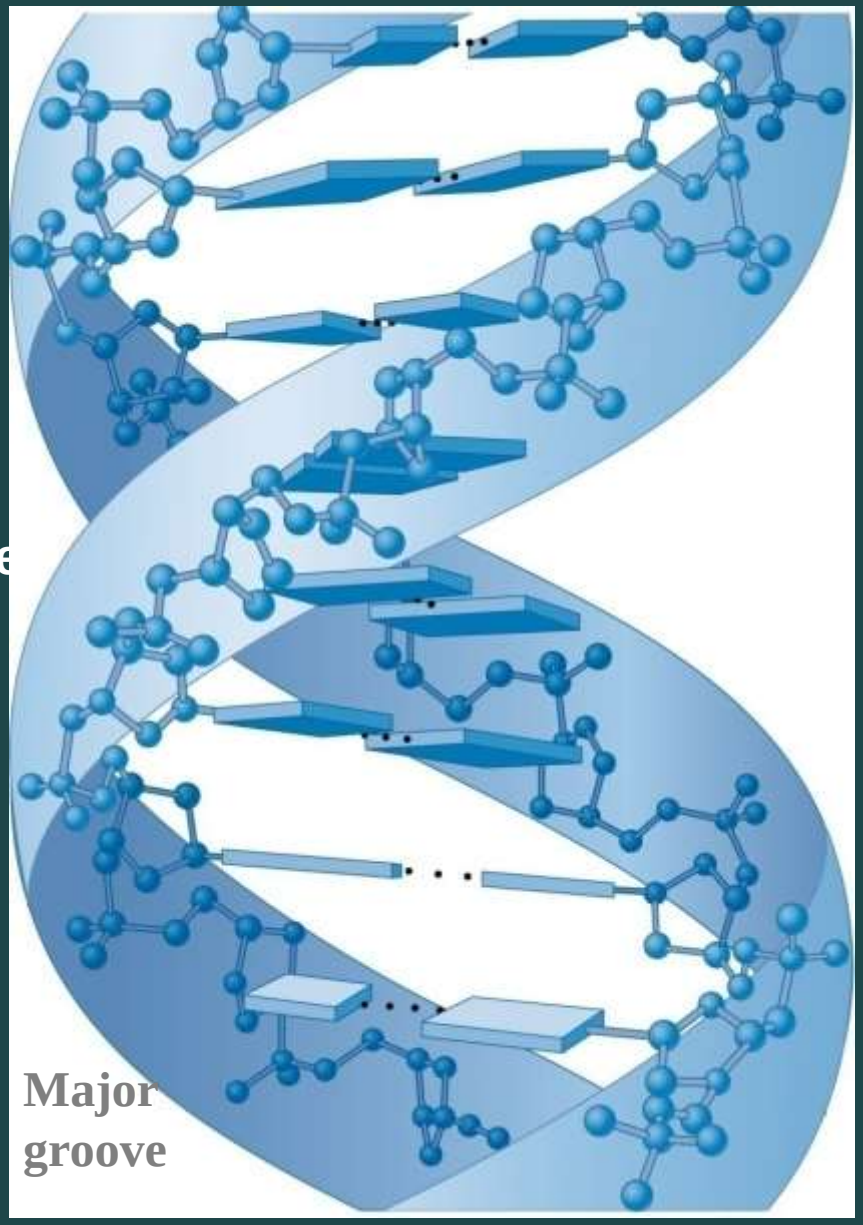


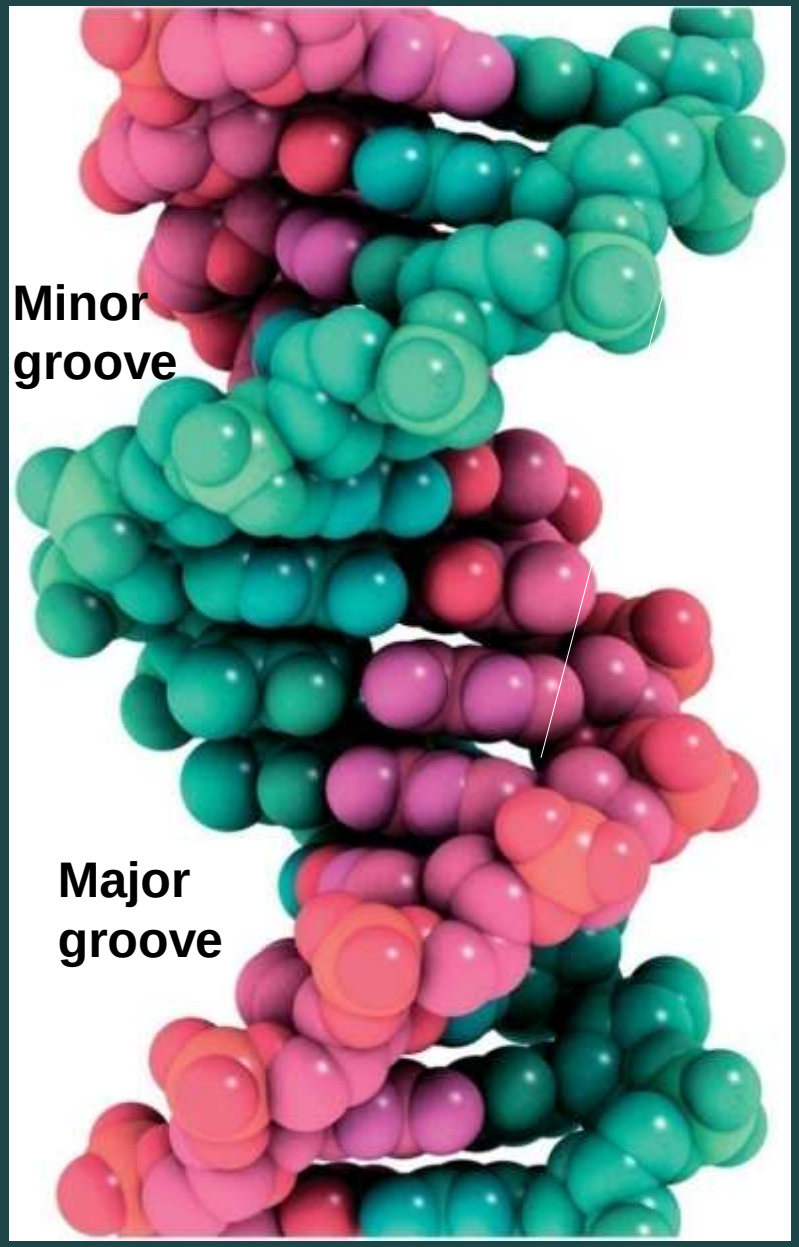

**N7 atom of the purine ring and the C5 atom of the pyrimidine ring face out into the major groove**

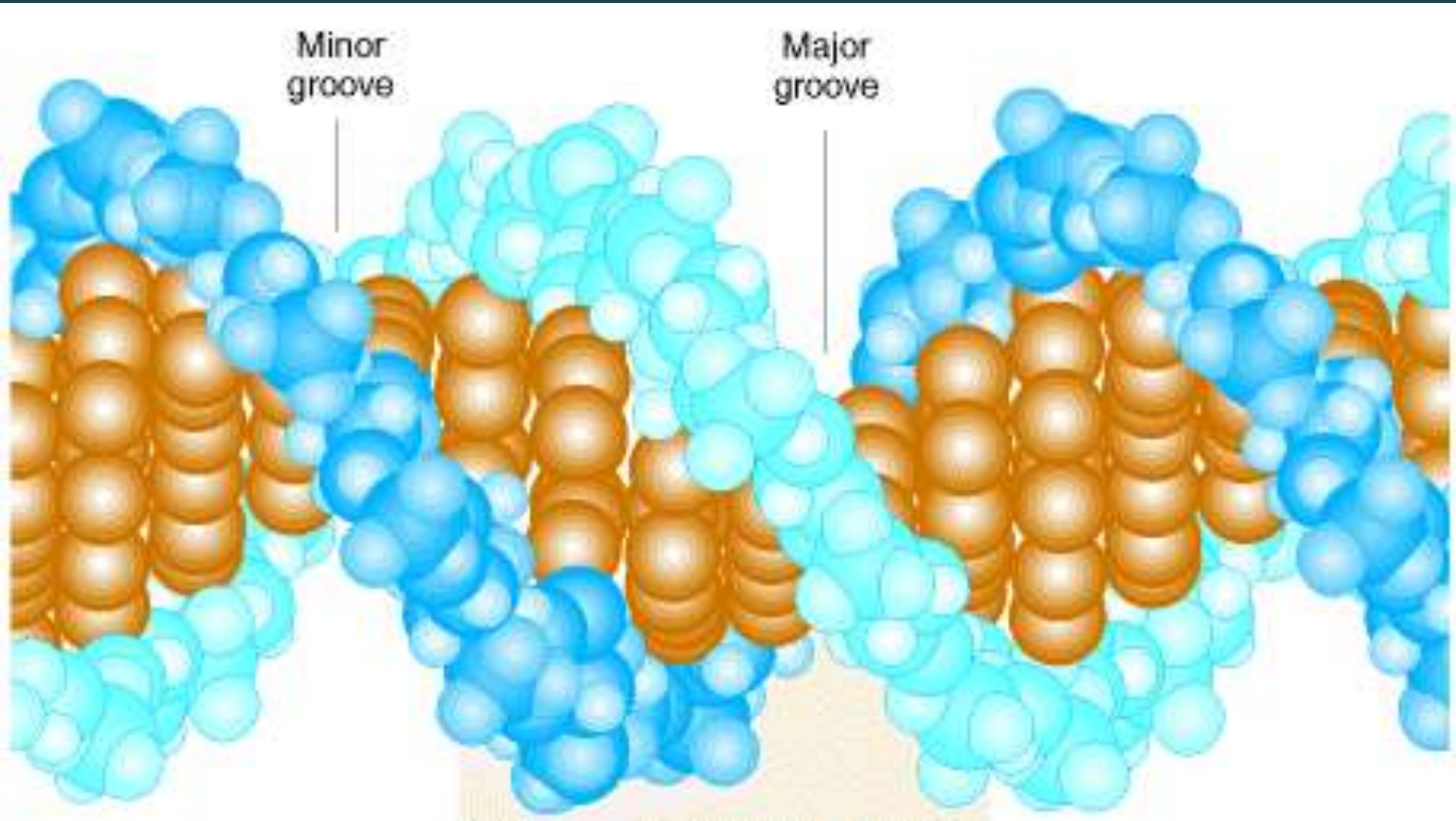


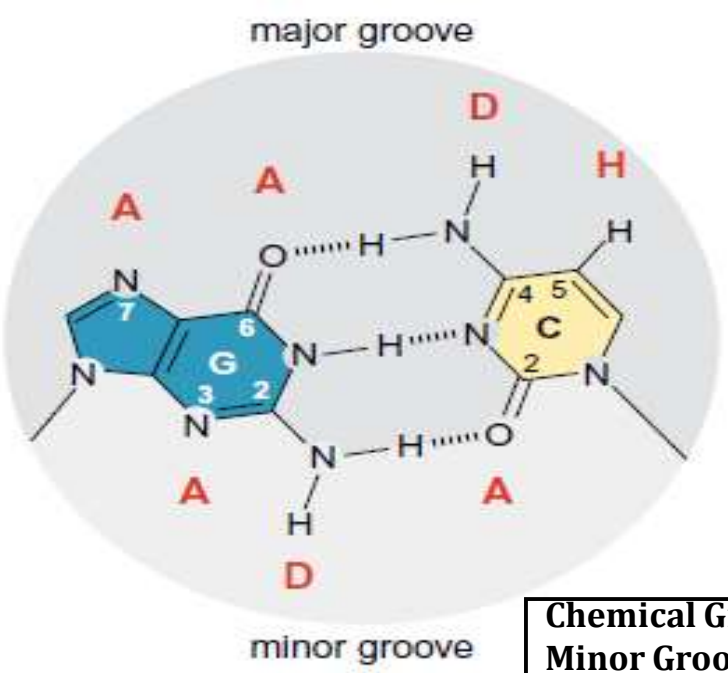

Sugar-phosphate backbone
base pairs
Sugar-phosphate backbone
A
T
Hydrogen bonds
G
C
T
A
A
T
C
G
Base pair
G
C
P
S
Nucleotide

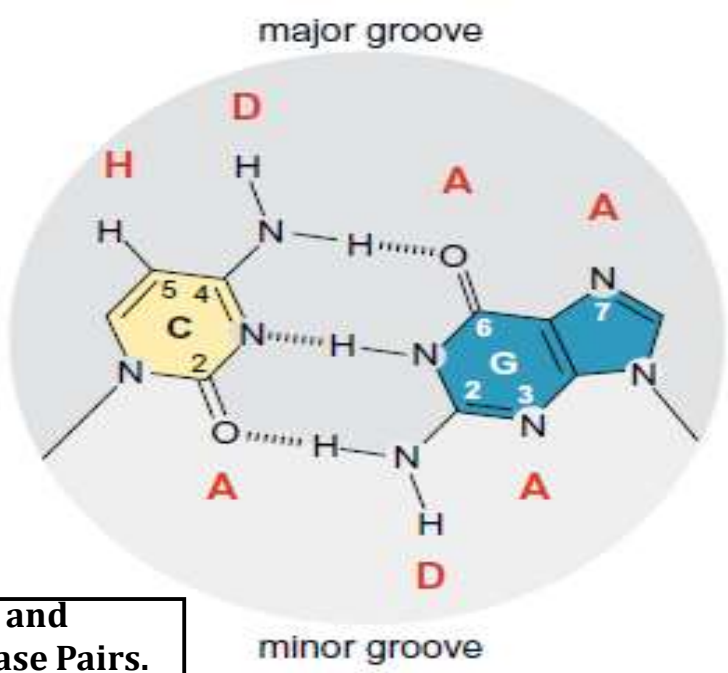
Minor groove



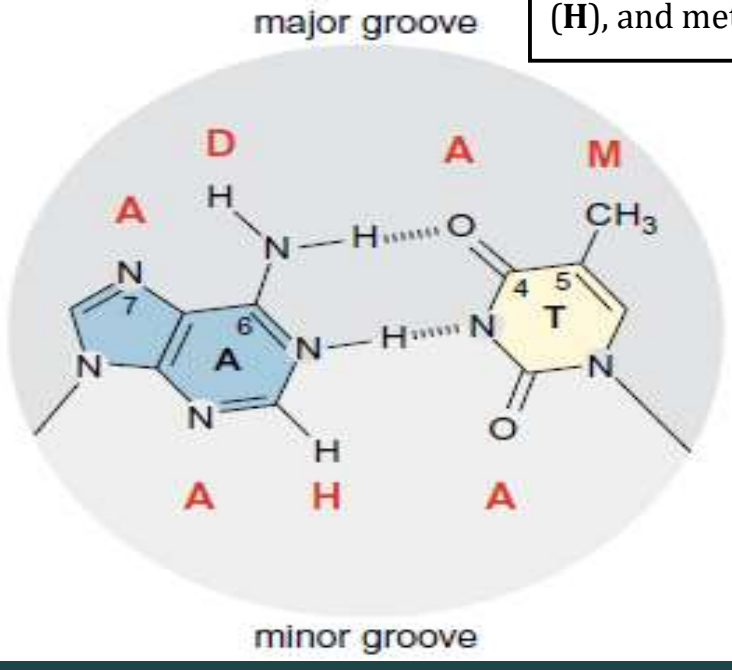



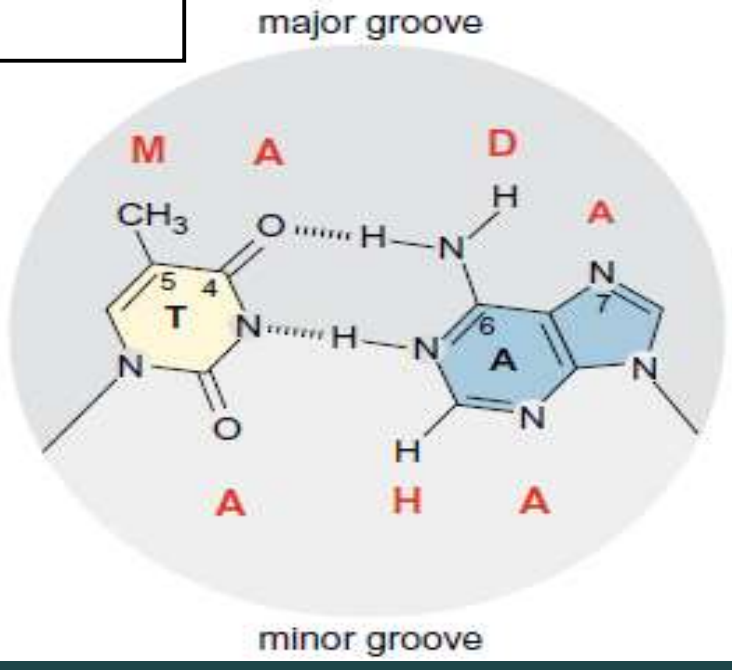

Minor
groove
Major
groove

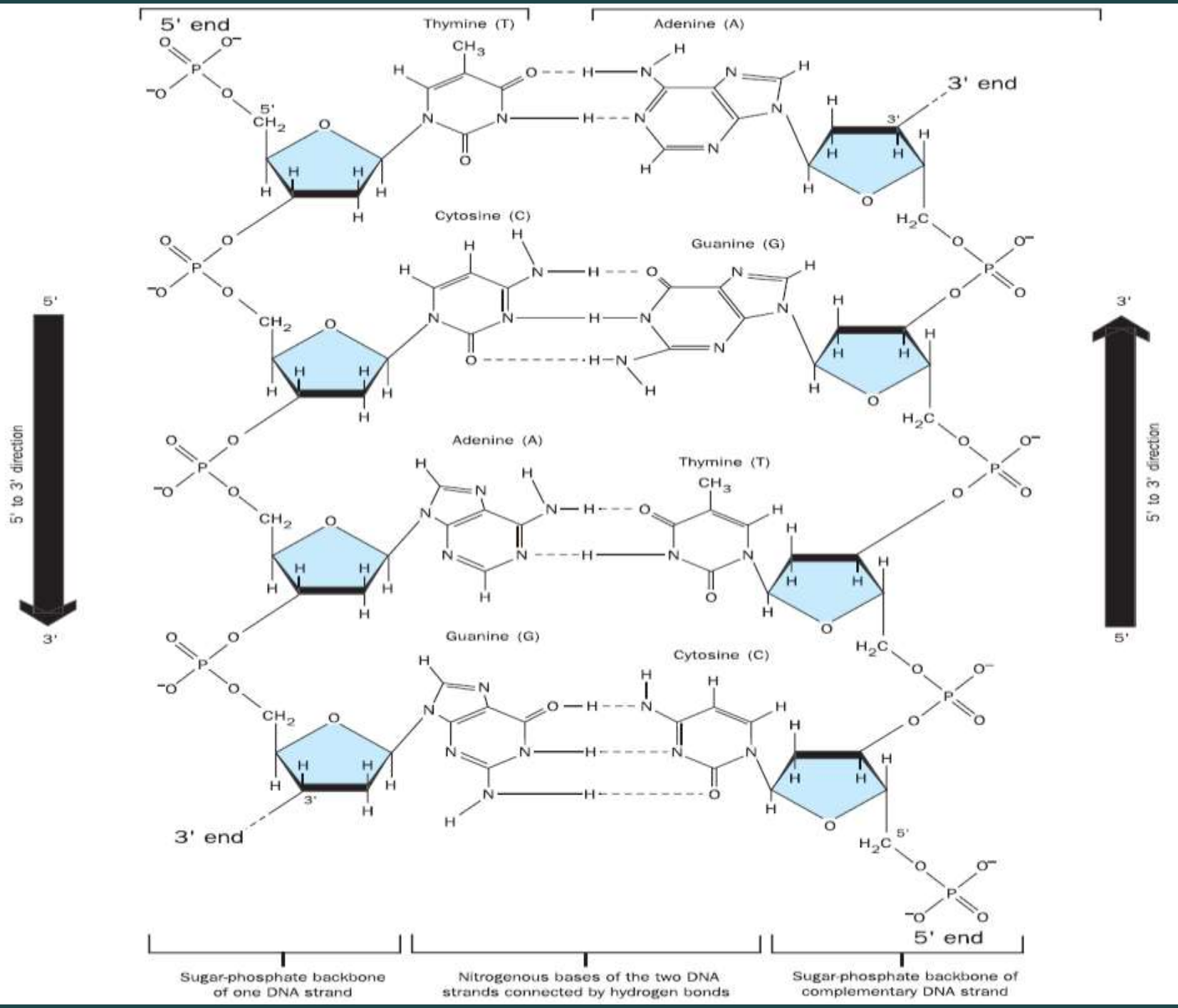



**Chemical Groups Exposed in the Major and Minor Grooves from the Edges of the Base Pairs.** The letters in red identify hydrogen bond acceptors (**A**), hydrogen bond donors (**D**), nonpolar hydrogens (**H**), and methyl groups (**M**).

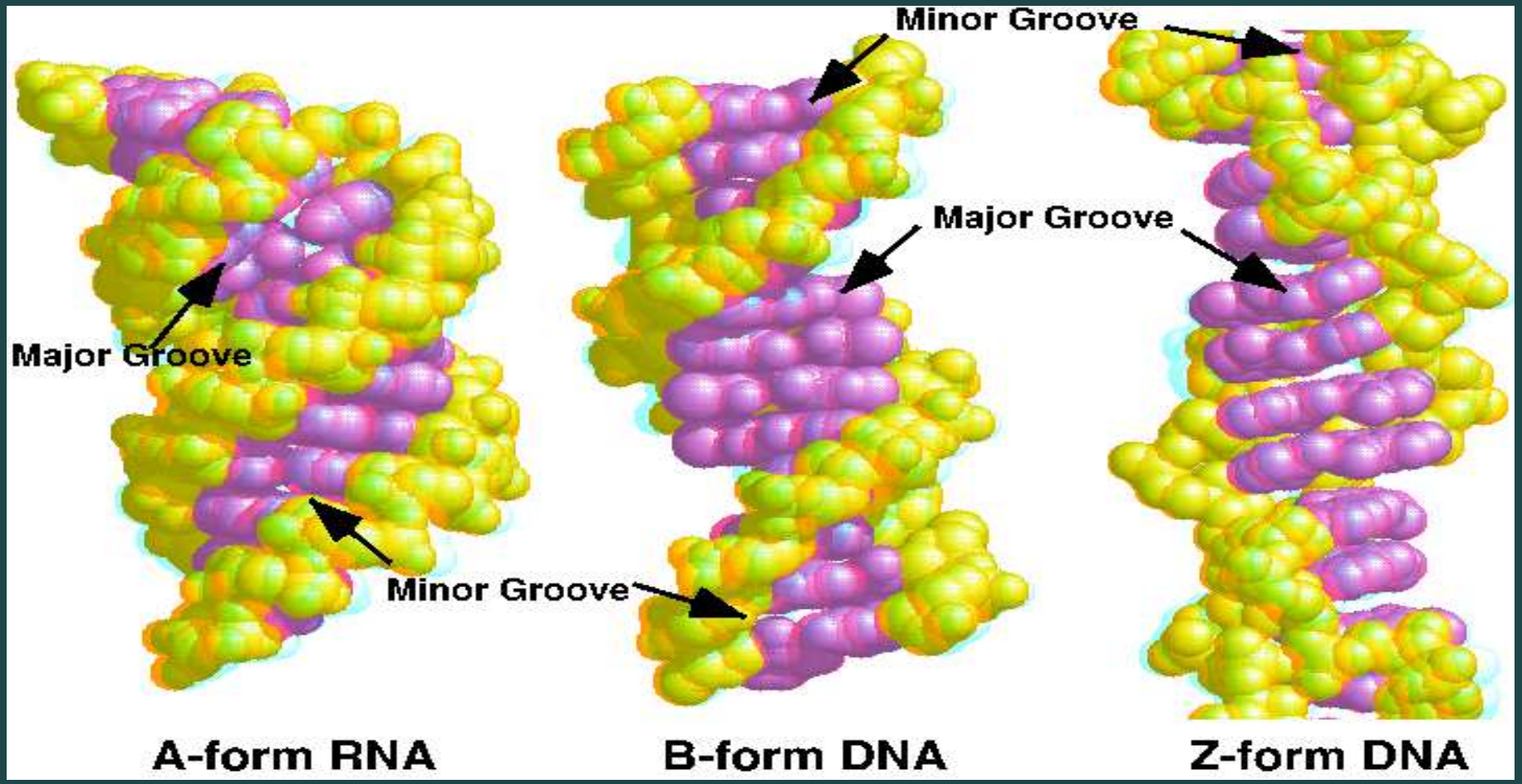



5' end
Thymine (T)
Adenine (A)
3' end
Cytosine (C)
Guanine (G)
Adenine (A)
Thymine (T)
Guanine (G)
Cytosine (C)
3' end
5' end
5' to 3' direction
5' to 3' direction
Sugar-phosphate backbone of one DNA strand
Nitrogenous bases of the two DNA strands connected by hydrogen bonds
Sugar-phosphate backbone of complementary DNA strand

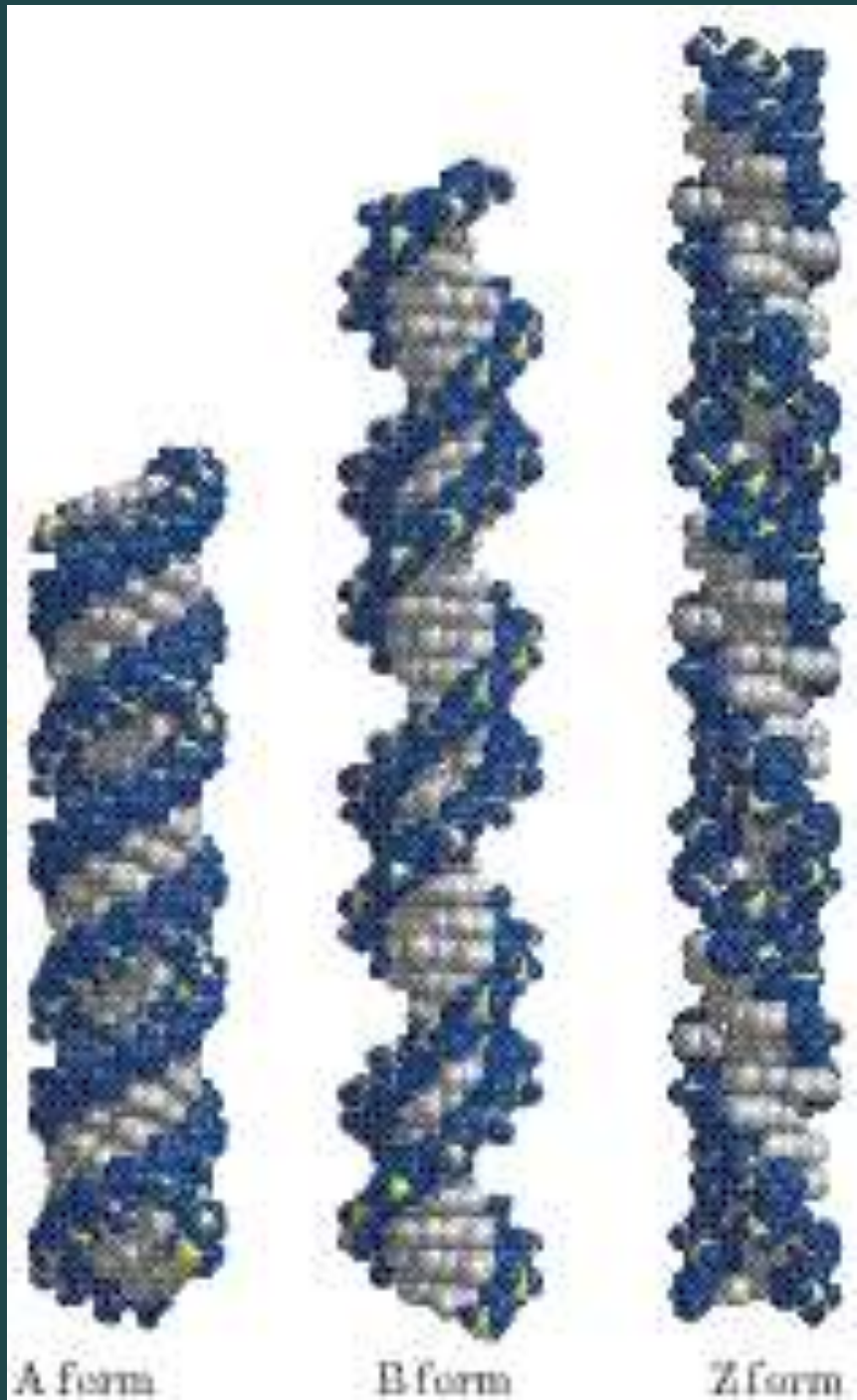
# different types of helical structure

- DNA can adopt 3 forms including,
- A DNA  right-handed has 11 base pairs per turn
- B DNA  right-handed has 10 base pairs per turn
- Z DNA  left-handed has 12 base pairs per turn
- Under physiological conditions, most of the DNA in a bacterial or eukaryotic genome is of the B-DNA.




| | A DNA | B DNA | Z DNA |
|---|---|---|---|
| **Major groove proportions** | **Extremely narrow but very deep** | **Wide and with intermediate depth** | **Flattened out on helix surface** |
| | **Exist** when less water is present. | Exist in normal physiological conditions | Exist when particular base sequences are present, such as stretches of alternating C and G sequences Z-DNA may play a role in regulating gene transcription |
| Helical sense | Right handed | Right handed | Left handed |
| **Diameter** | **26 $A^0$** | **20$A^0$** | **18$A^0$** |

Each structure shown here has 36 base pairs



# DNA-protein interaction and binding

**Regulatory proteins**

(1) High affinity binding to a specific site.

(2) Only small regions of the protein make direct contact with DNA.

(3) The protein-DNA interactions are maintained by hydrogen bonds and van der Waals forces.

(4) Proteins with the helix-turn-helix or leucine zipper motifs form symmetric dimers, and their respective DNA binding sites are symmetric palindromes (sequence of units that can be read the same way in either direction).

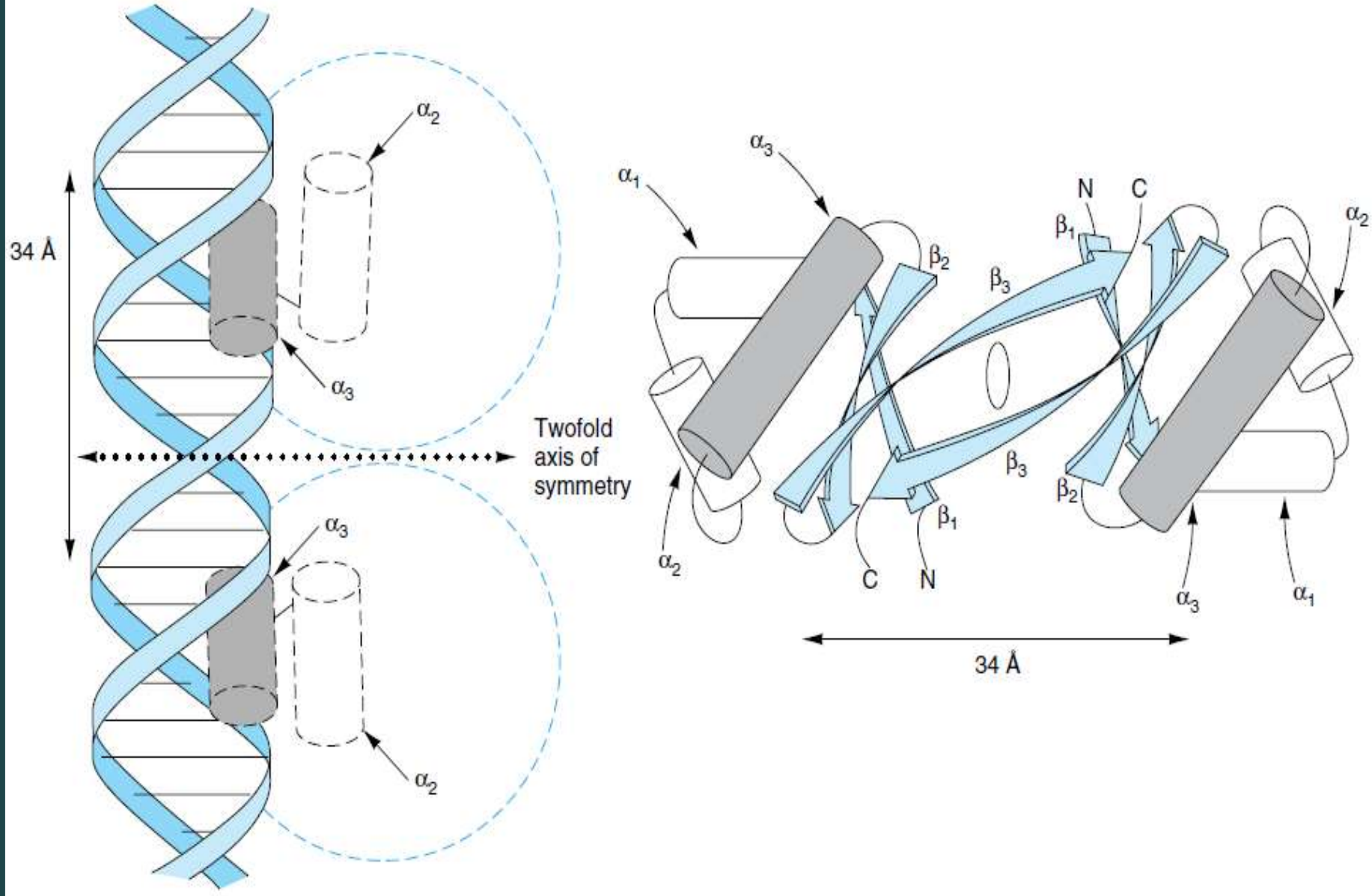
In proteins with the zinc finger motif, the binding site is repeated two to nine times.

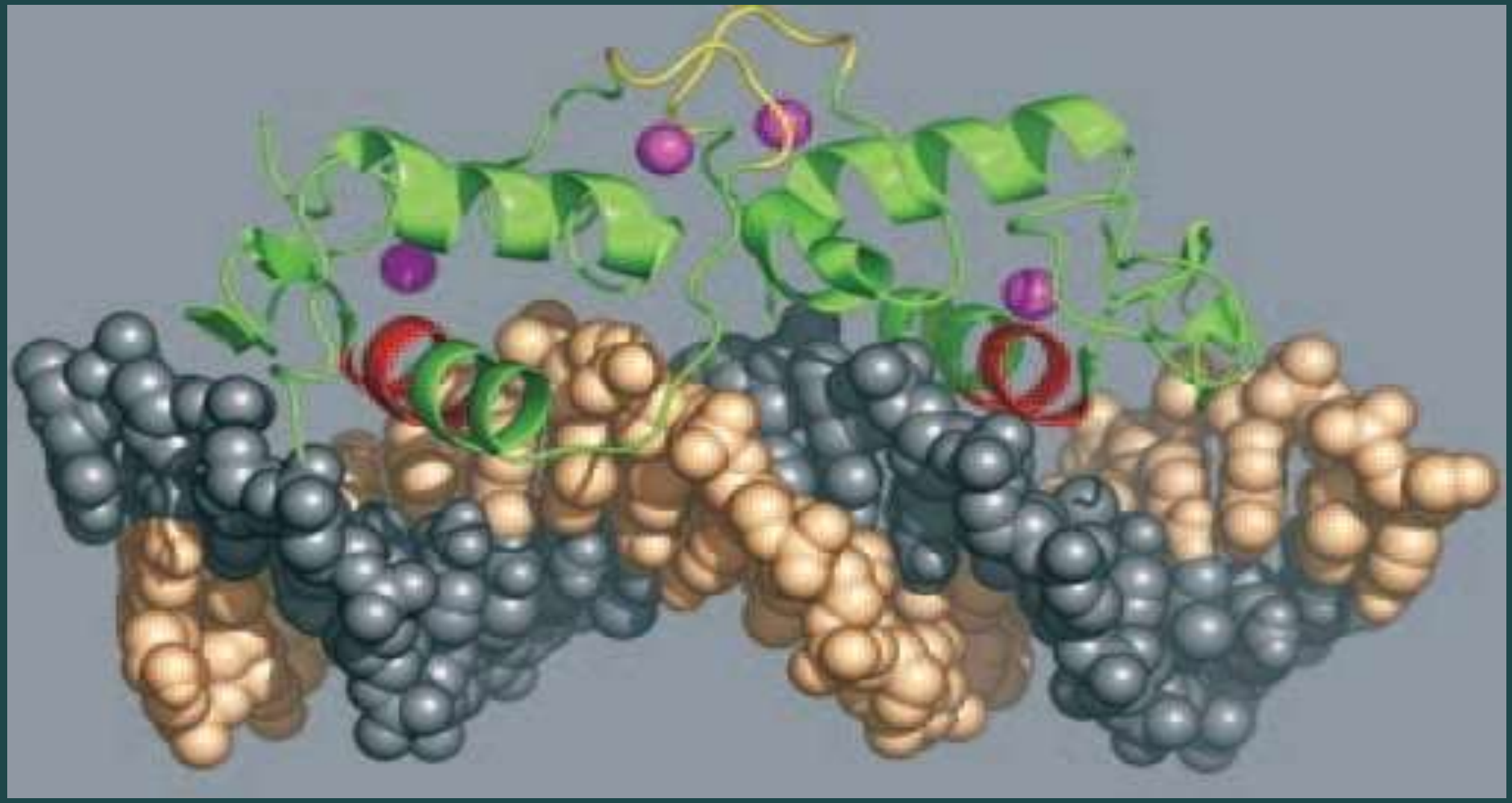
→ ACCTAGGT palindromic

TGGATCCA ←

# The Helix-Turn-Helix Motif

- Consists of three antiparallel β sheets (β1–β3) and three α helices (α1–α3).
- The helix-turn-helix motif is formed because the α3 and α2 helices are held at about 90 degrees to each other by a turn of four amino acids.
- The α3 helices form the DNA recognition surface of many proteins, and the rest stabilize structures.
- The DNA recognition domain interacts with 5 bp.
- Two monomers associate through the antiparallel β3 sheets to form a dimmer that has a twofold axis of symmetry.
- The average diameter of an α helix is 1.2 nm, which fit in the major groove in the B form of DNA.

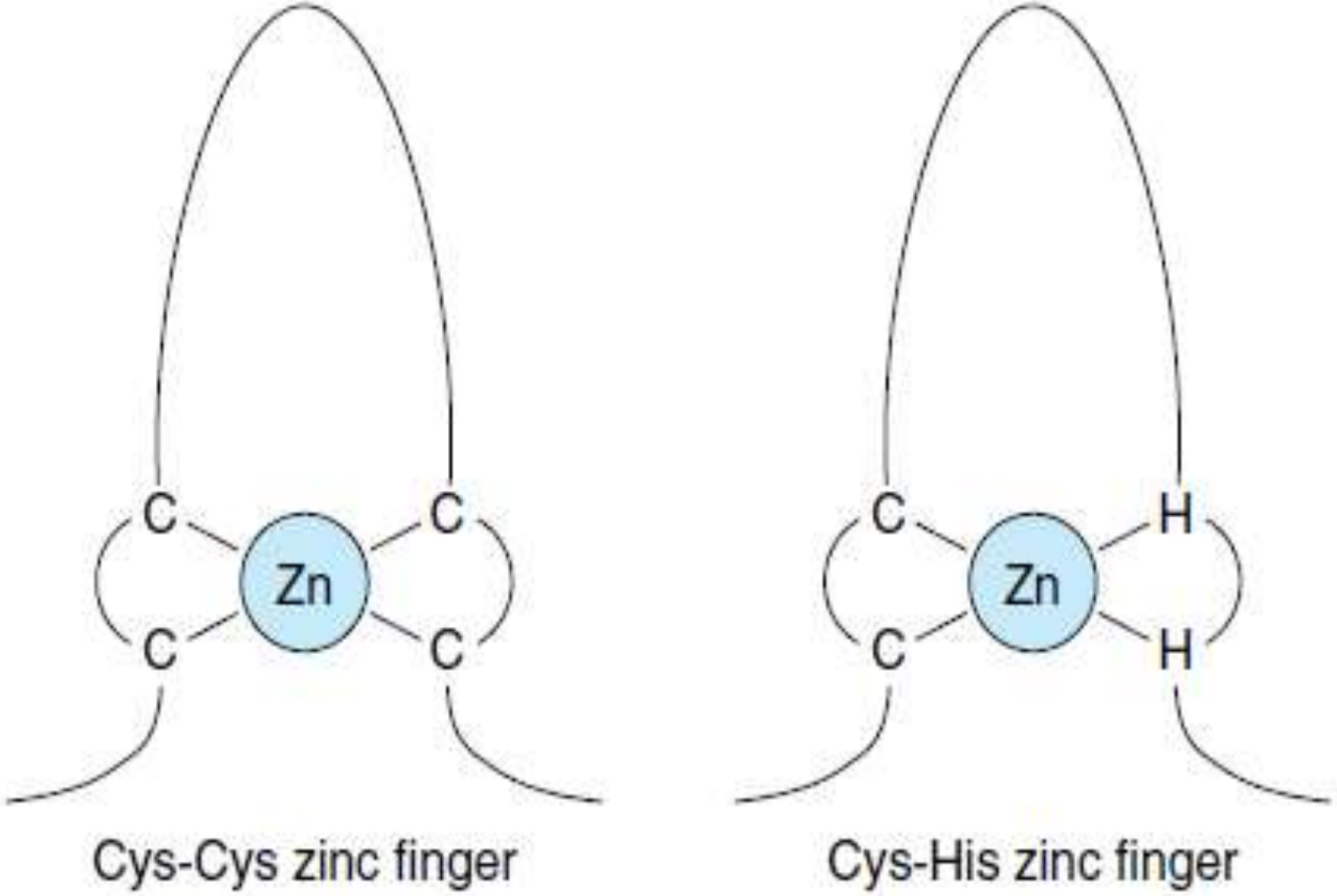

34 Å
α2
α3
Twofold
axis of
symmetry
α3
α2
α1
α3
β2
β3
N
C
β1
α2
β3
β1
β2
C
N
α3
α1
34 Å



- Structure of receptor (dimer) binding to DNA

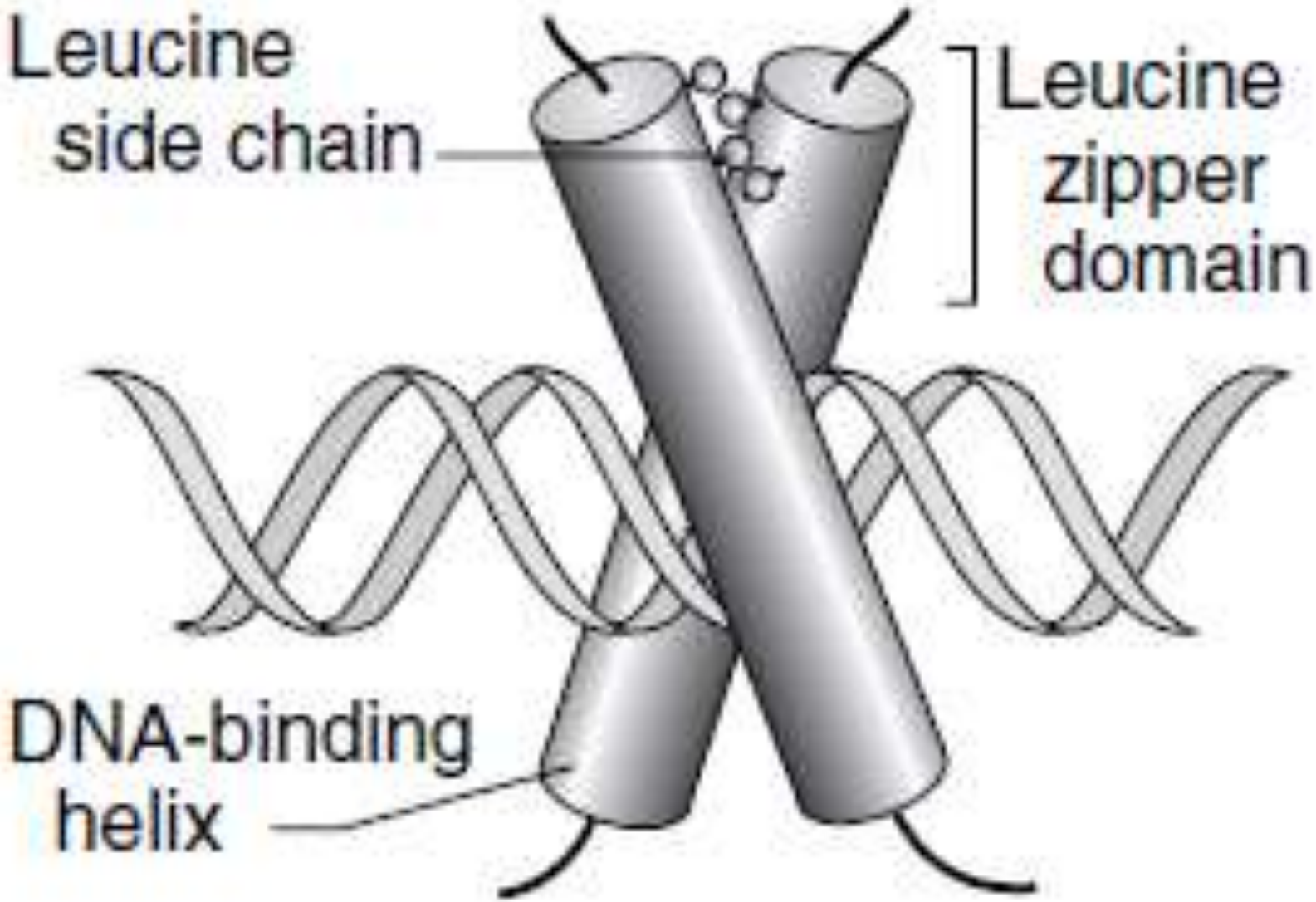
# The Zinc Finger Motif

- Zinc finger motifs: zinc bound at four positions with either four cysteine, or two cysteine and two histidine in a sequence of approximately 20 amino acids.
- The zinc is required to maintain the tertiary structure of this domain.
- Each zinc finger contacts about 5 bp of DNA.
- Zinc finger motifs consist of an α-helix and a β- sheet. The nucleotide recognition signal (contained within the α-helix) of at least one zinc finger binds to a specific sequence of bases in the major groove of DNA.


C
C
Zn
C
C
Cys-Cys zinc finger
C
H
Zn
C
H
Cys-His zinc finger

# The Leucine Zipper (LZ) Motif

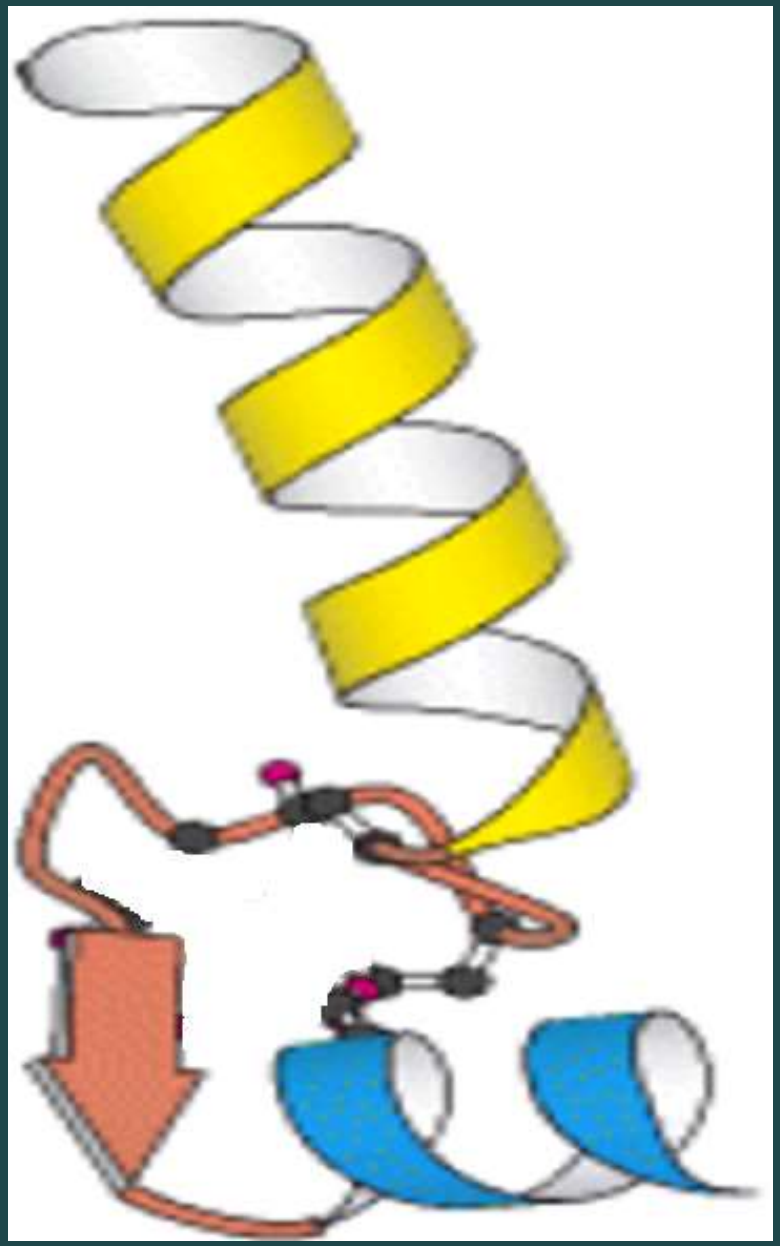
- Leucine zippers also function as dimers to regulate gene transcription.
- The leucine zipper motif is an **α-helix** of 30 to 40 amino acid residues
- LZ contains a **leucine** (nonpolar hydrophobic aa) every **seven** amino acids, positioned so that they align on the same side of the helix.
- Two helices dimerize so that the leucines of one helix align with the other helix through hydrophobic interactions to form a coiled coil.
- The portions of the dimer adjacent to the zipper bind the DNA through basic amino acid residues (arginine and lysine) that bind to the negatively charged phosphate groups. This DNA binding portion of the molecule also contains a nucleotide recognition signal.




**Alpha helix makes one full turn every 3.6 residues**

- **Helix-loop-helix**
- It consists of a short a helix connected by a flexible loop of amino acids to a second longer a helix. (loop is a number of aa their function is to connect two strands and change direction)
- A highly basic set of amino acids in one of the helices binds to the DNA.
- They also function as dimers that is similar to leucine zipper proteins.
- The dimerization region consists of a portion of the DNA-gripping helix and a loop to another helix.

# Helix loop helix



# Mitochondria DNA (mtDNA)

- Mitochondrial DNA is a double strand and circular that is not covered with histone and lacks introns. Exons code for polypeptide, whereas introns do not
- The mtDNA contains 37 genes, 13 of these genes provide instructions for making enzymes involved in oxidative phosphorylation. The remaining genes provide instructions for making transfer RNA (tRNA) and ribosomal RNA (rRNA).
- mtDNA is only inherited from our mother. The mitochondria in mammalian sperm are usually destroyed by the egg cell after fertilization and are present at the base of the sperm's tail, which is usually lost during fertilization.
- mtDNA also does not recombine; there is no shuffling of genes from one generation to the other, as there is with nuclear gen.

- **Why Study mtDNA?**
- There are many diseases caused by mutations in **mtDNA**. Mutations in mtDNA increase the production of potentially harmful molecules called reactive oxygen species (ROS) which is the cause of many diseases including cancer. Tumour development in most of cases is associated with mutation to mitochondrial DNA.
- Because the mitochondria produce energy in cells, symptoms of mitochondrial diseases often involve degeneration or functional failure of tissue. For example, mtDNA mutations have been identified in some forms of diabetes, deafness, and certain inherited heart diseases. In addition, mutations in mtDNA are able to accumulate throughout an individual's lifetime due to lack of repair mechanism. Evidence suggests that the mtDNA mutations contribute to the progression of Parkinson's and Alzheimer's disease.
- In addition to the critical cellular energy-related functions, mitochondrial genes are useful to evolutionary biologists because of their maternal inheritance. By studying patterns of mutations, scientists are able to reconstruct patterns of migration and evolution within and between species. For example, mtDNA analysis has been used to trace the migration of people from Asia to North and South America.

# Three-parents baby

- Mitochondrial replacement IVF (mtIVF)
- Used to avoid transmission of mitochondrial disease.
- **The technique**

1. Take the nucleus of mother egg
2. Insert it into a donor egg that its nucleus removed but still contains mitochondrial DNA
3. Fertilize this hybrid egg with father sperm.

- Religious and ethical objections

# Mitochondrial replacement

Repair is done before fertilisation

## Maternal spindle transfer