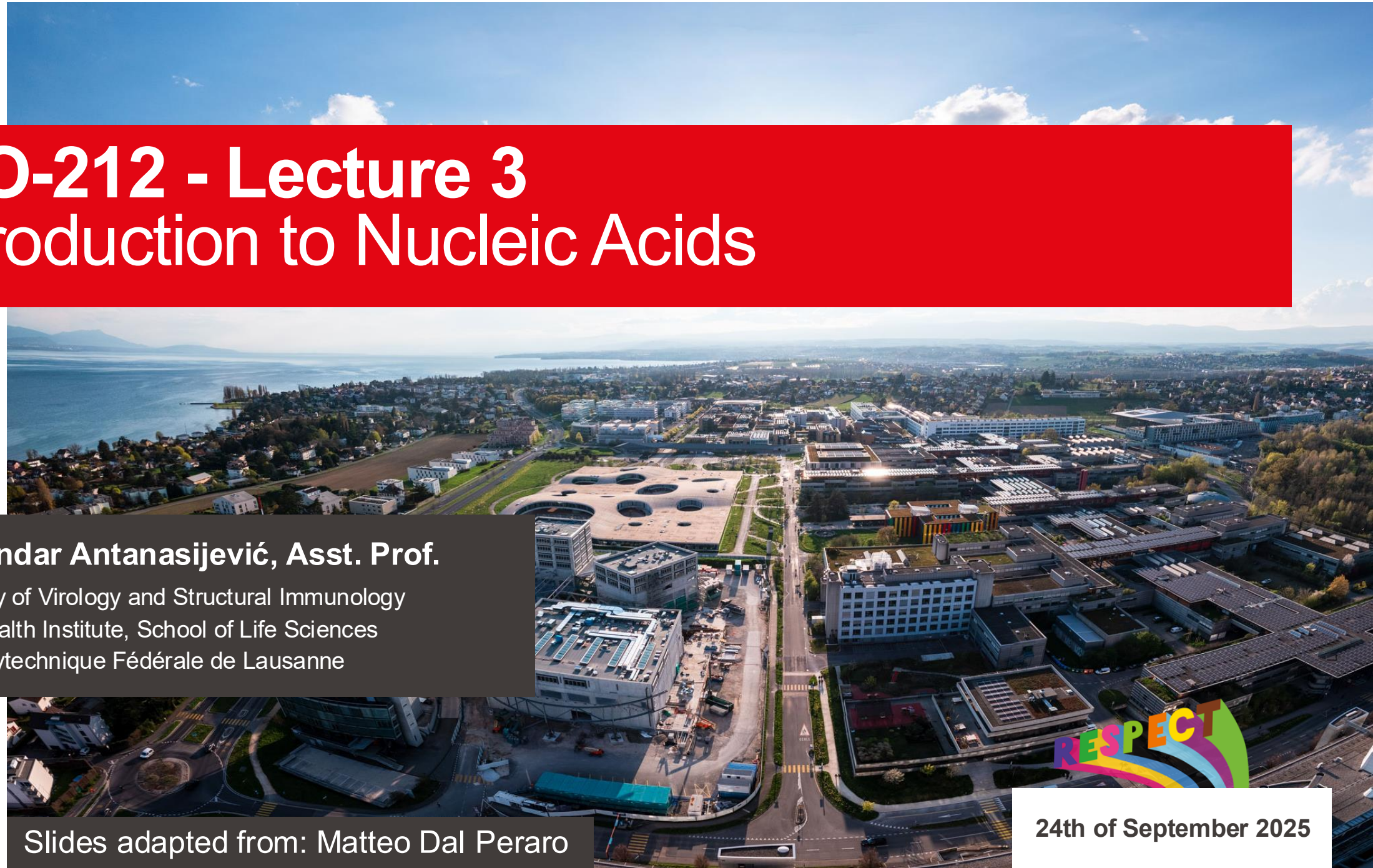
BIO-212 - Lecture 3
Introduction to Nucleic Acids



Aleksandar Antanasijević, Asst. Prof.

Laboratory of Virology and Structural Immunology
Global Health Institute, School of Life Sciences
École Polytechnique Fédérale de Lausanne

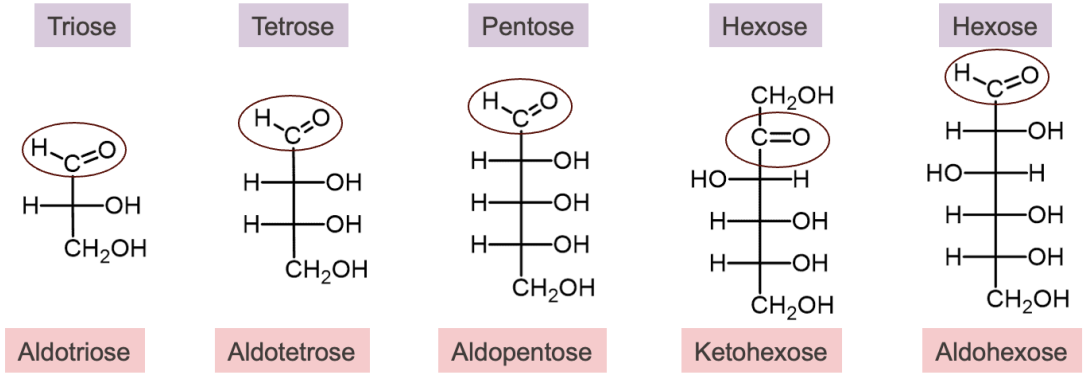
Slides adapted from: Matteo Dal Peraro

24th of September 2025

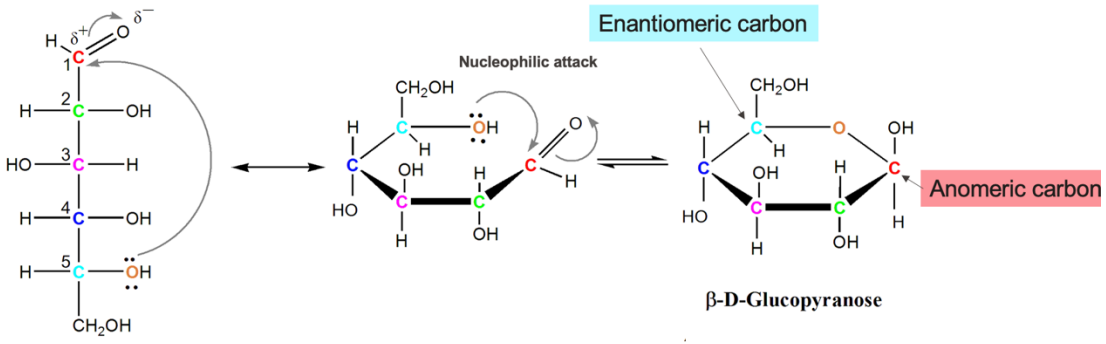
Lecture 2 – Quick Summary

• Carbohydrate building blocks

Monosaccharides:

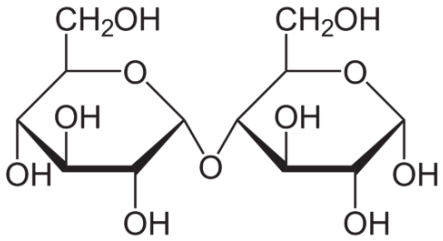


Hemiacetal and hemiketal forms in solution:

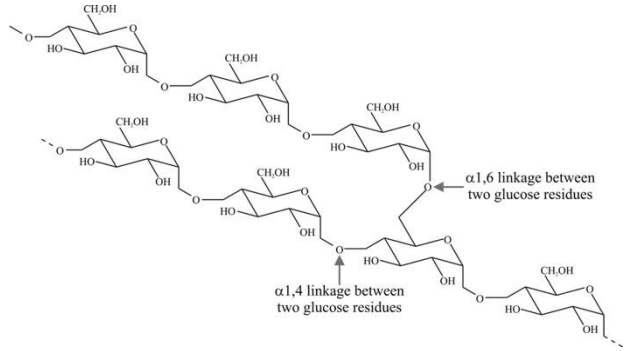


• Carbohydrate polymers and their roles

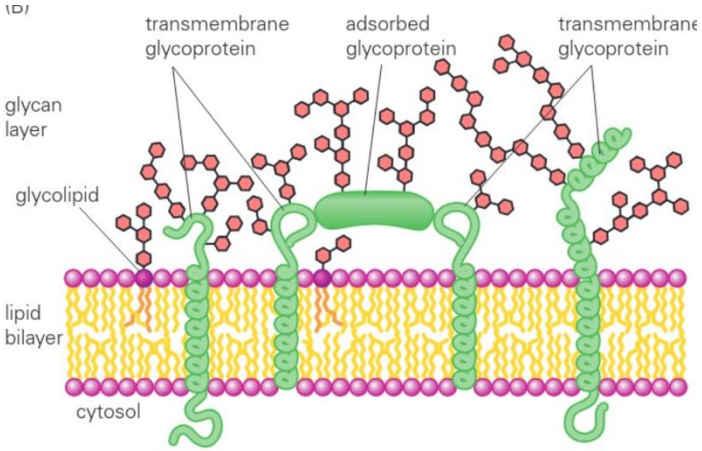
Disaccharide
(Maltose)



Polysaccharide
(Amylopectin)

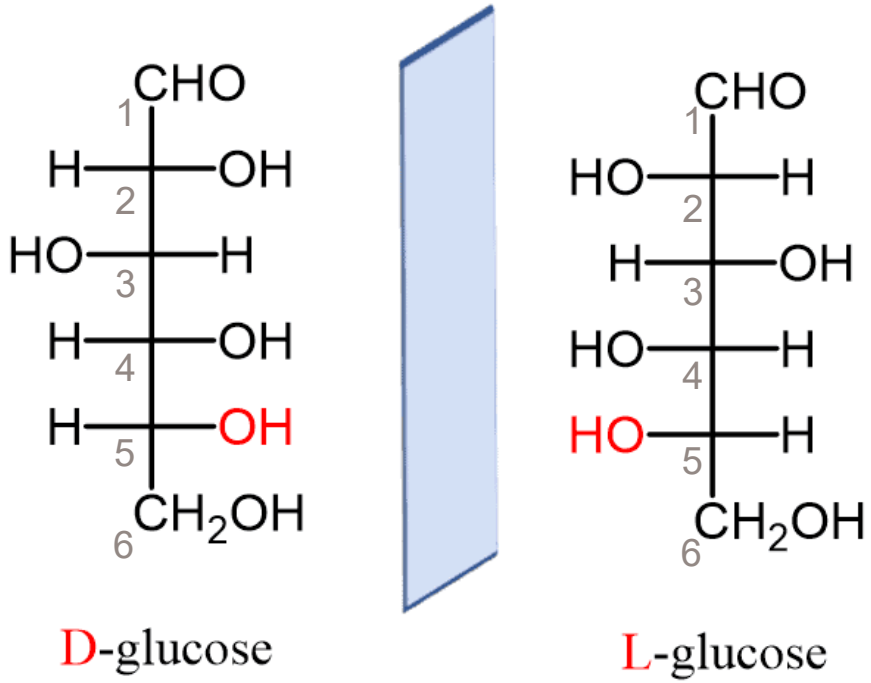


Glycolipids and glycoproteins



Lecture 2 - Extra Clarifications

• Enantiomeric forms of carbohydrates

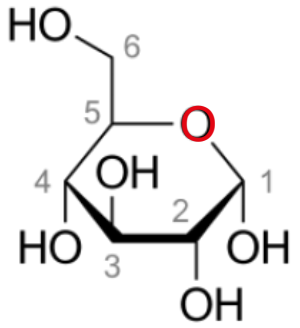


D-glucose L-glucose

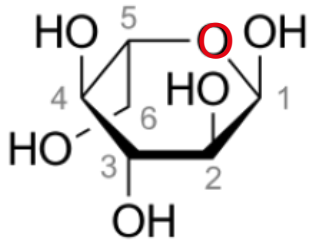
- Enantiomers of monosaccharides are connected by mirror symmetry at every chiral center.
- The main chiral center (C5 in glucose) is just used to assign D- or L- form

• The position of terminal -CH_2OH

- In the cyclic form the position of terminal -CH_2OH group will depend on the D- or L- isomerism:



α-D-glucopyranose



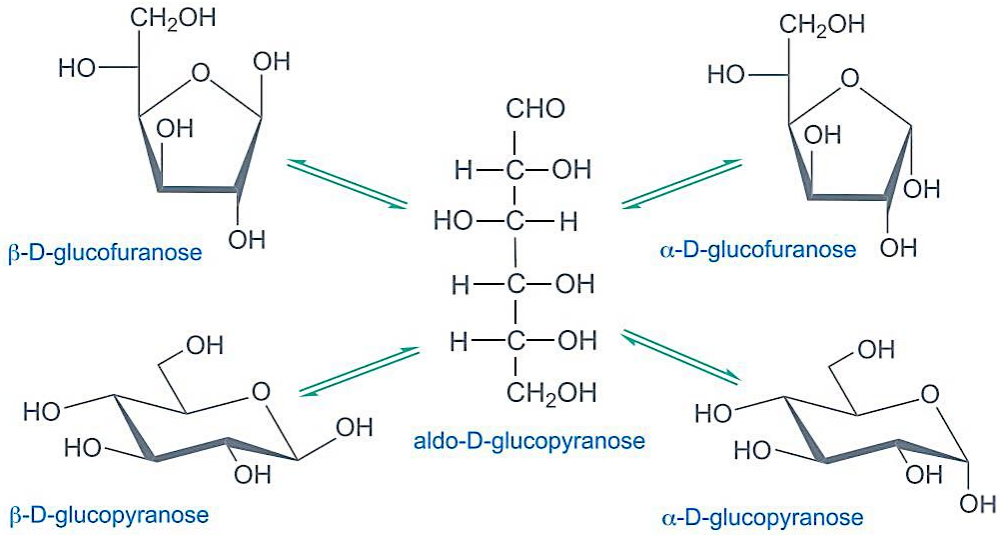
α-L-glucopyranose

- This is counter-inutitive given the orientation of H at position C5 (CH_2OH expected to face down in D-form and up in L-form, but that does not happen)
- The final orientation is influenced by the rotation around the C4-C5 bond during cycle assembly.
- See: https://www.youtube.com/watch?v=kXnkjLAvGwk

Lecture 2 - Extra Clarifications

• Multiple carbon options for cyclic state assembly

- In theory any OH group at or after the 5th carbon can be used for assembly into cyclic form



- In the majority of cases the most stable linkage is via **pen-ultimate carbon** (e.g., C5 in glucose)

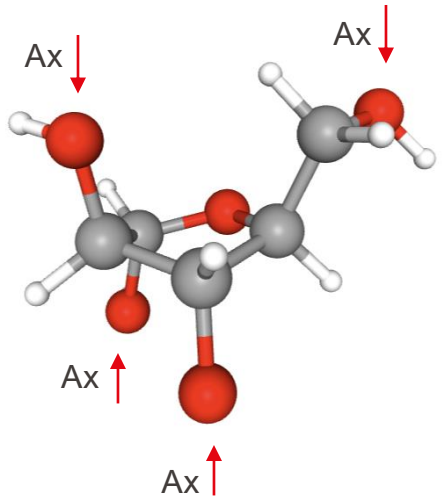
- 5- and 6-membered rings are energetically favored but preference for one or the other depends on the specific monosaccharide and orientation of its OH groups

• Achieving maximum distance of electron-rich groups

- State of minimum energy is the one with maximum spacing between electron-rich groups

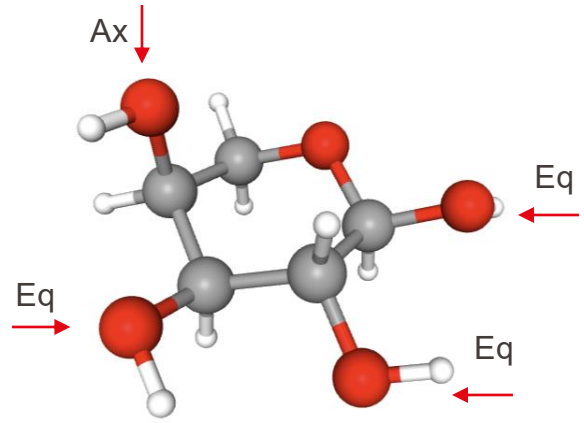
- Example: **L-arabinose** (an aldopentose)

Alpha-L-Arabinofuranose



10-20% of molecules in solution
(combined alpha and beta states)

Alpha-L-Arabinopyranose



80-90% of molecules in solution
(combined alpha and beta states)

- Equatorial orientation (right) provides better spacing between OH groups compared to axial (left)

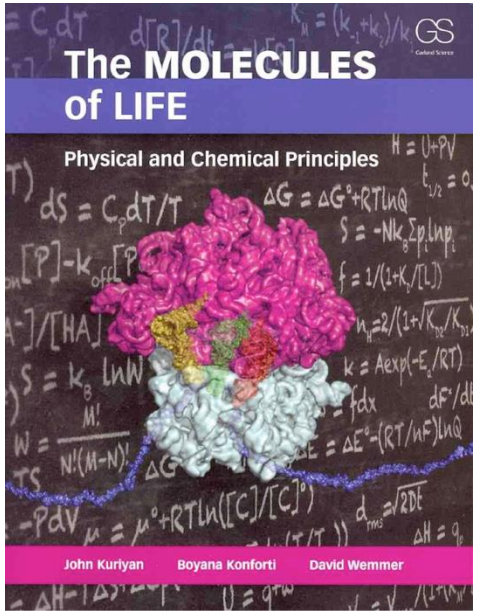
Lecture 3 - Outline

Today:

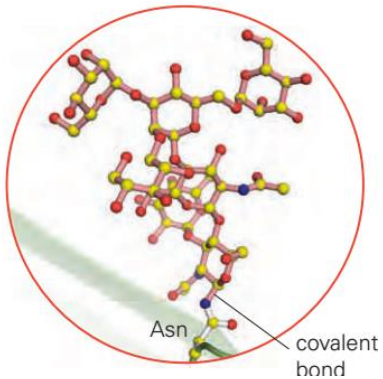
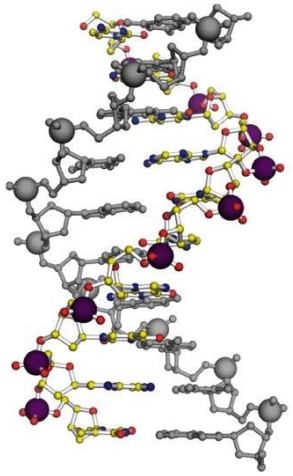
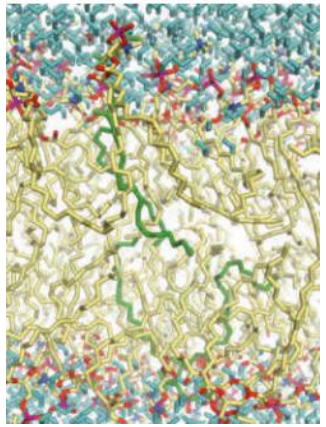
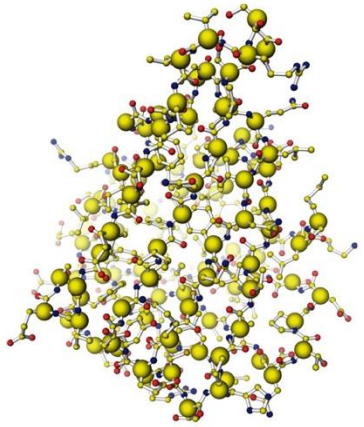
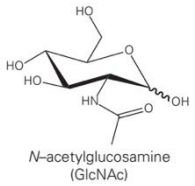
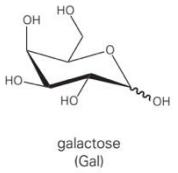
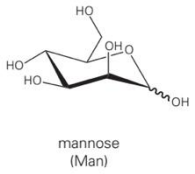
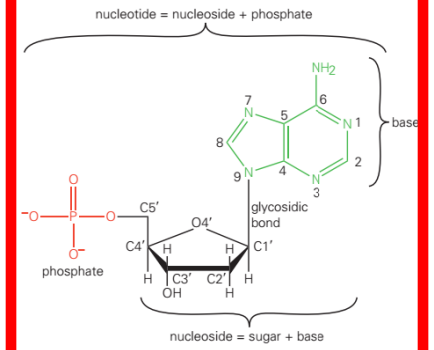
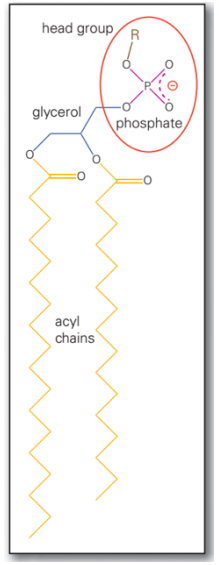
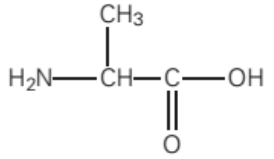
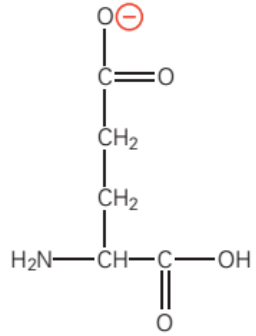
- Nucleotide building blocks
- Assembly into nucleic acids
- Relevant covalent and non-covalent interactions

Reading suggestions:

- The Molecules of Life (Chapter 1, 19)

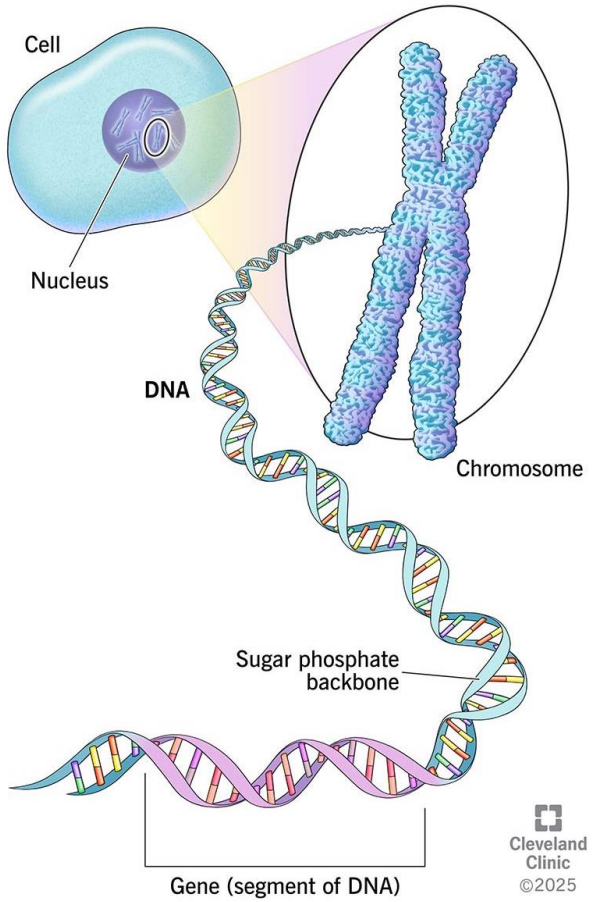


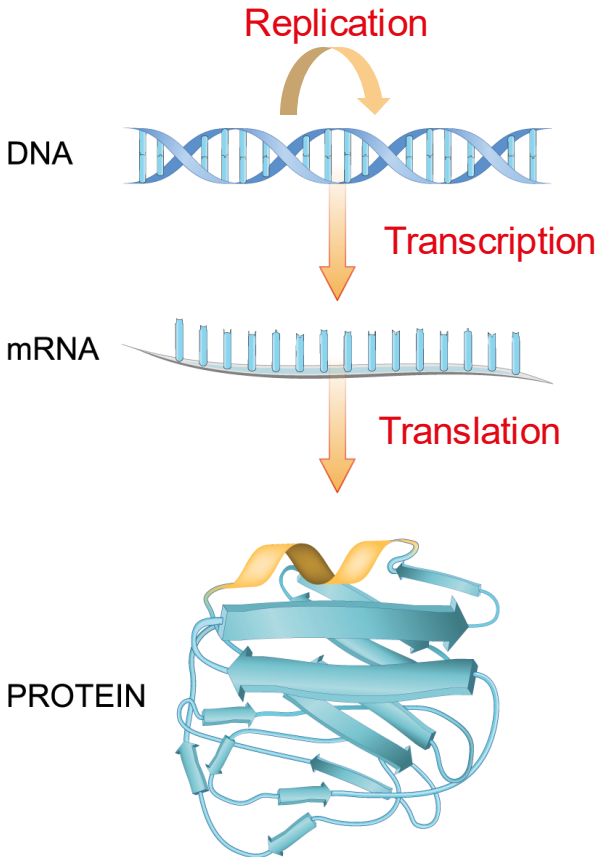
The molecules of Life

	Carbohydrates	Nucleic Acids	Lipids	Proteins
Macromolecular Structure				
Building Block	  			 

Nucleic acids in cells

- The main biological functions are **storage**, **transmission** and **expression** of genetic information





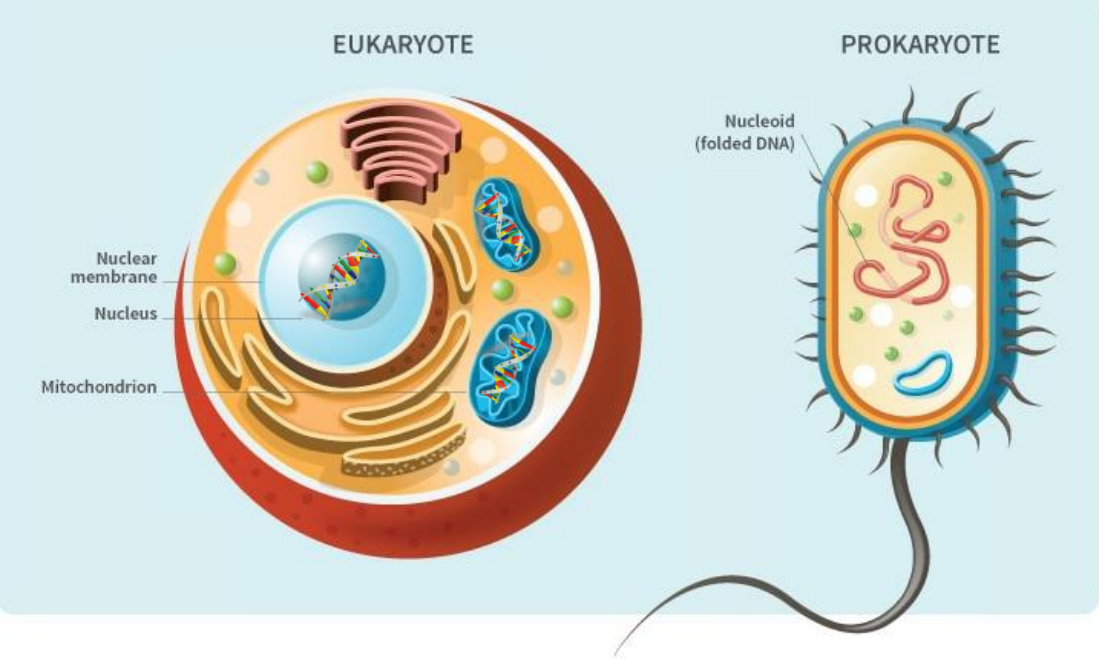
Johannes Friedrich Miescher
(1844-1895)

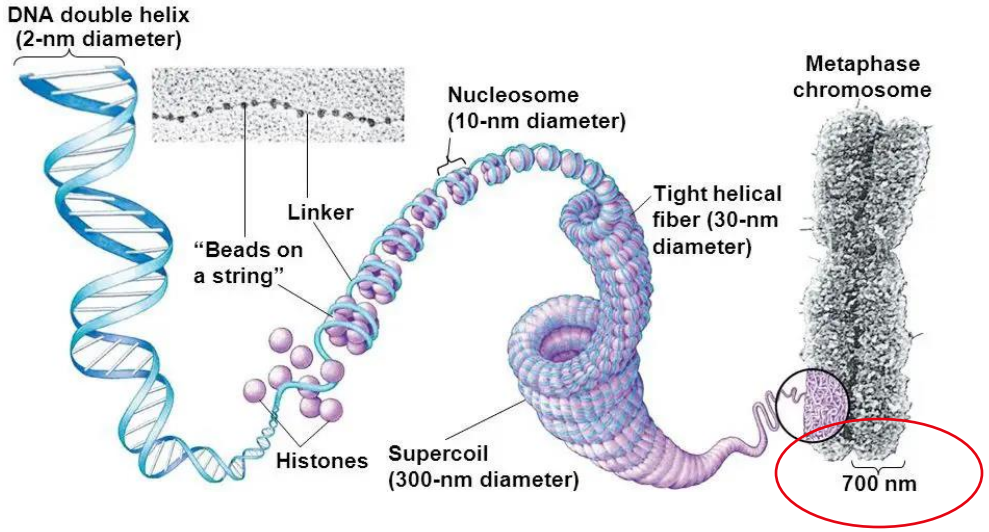
First person to isolate nucleic acids and suggest a role in heredity (“nuclein”)

- Nucleotides also serve as **energy carriers** and **signaling molecules**, while RNA has many **regulatory**, **catalytic**, and **sensing** roles (particularly important for translation).

Nucleic acids in cells

• In eukaryotes most **DNA** is stored as chromatin within the nuclear envelope, and in organelles such as mitochondria and chloroplasts (plants/algae)
• **RNA** exists in the nucleus, cytosol and inside certain organelles (e.g., mitochondria, exosomes, chloroplasts etc.)





The length of human DNA in extended state = **2.2 m**
Diameter of the cell nucleus: **5-10 µm**

• Human haploid genome is composed of $3.2*10^9$ nucleotide pairs arranged into ~10^7 nucleosomes which assemble into 23 chromosomes
• Helical organization and packaging with histones reduces the effective size by 5-6 orders of mag.

Nucleotides – The building blocks of nucleic acids

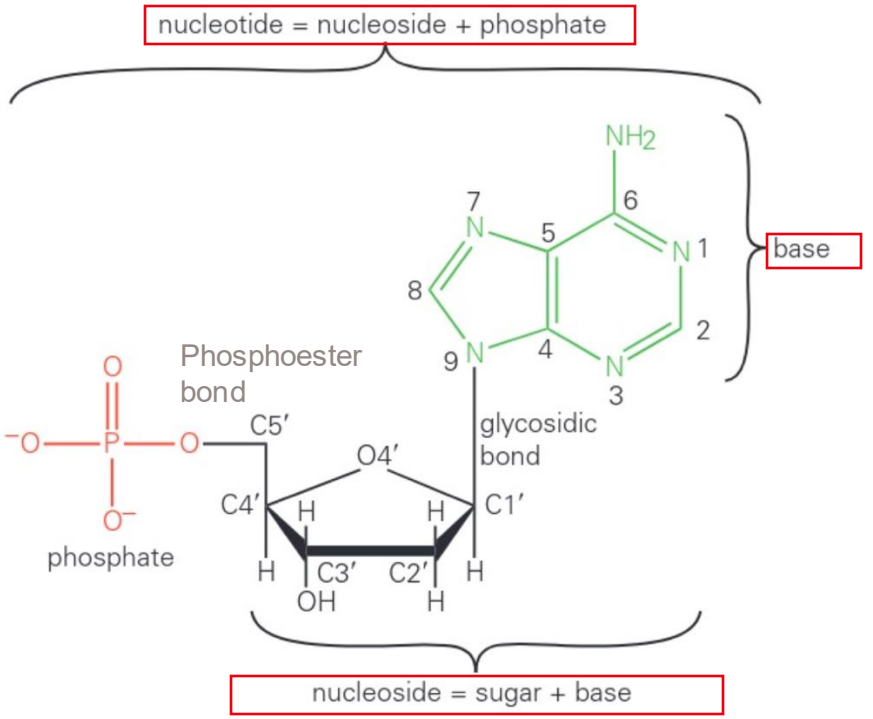
- DNA and RNA are both polymers of **nucleotides**.
- They can be synthesized endogenously and are therefore non-essential nutrients

Linkage between the base and carbohydrate:

- **N-glycosidic** (similar to O-glycosidic but via Nitrogen)

Linkage between the phosphate and carbohydrate:

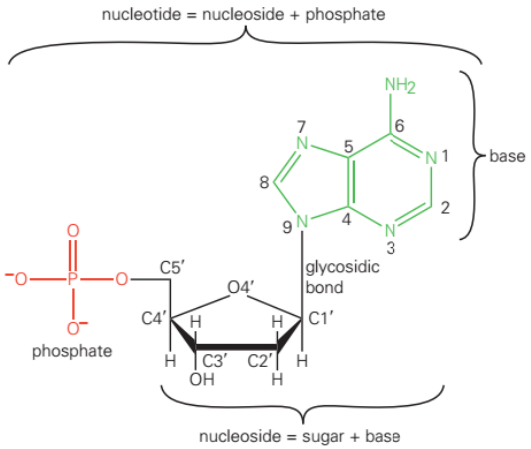
- **Phosphoester** (hydroxyl group to phosphate)



- Each nucleotide consists of the following functional groups:
  1) **Aldopentose** monosaccharide (black)
  2) Nitrogen-containing aromatic ring system, i.e. **Lewis base** (green)
  3) **Phosphate** group (red)

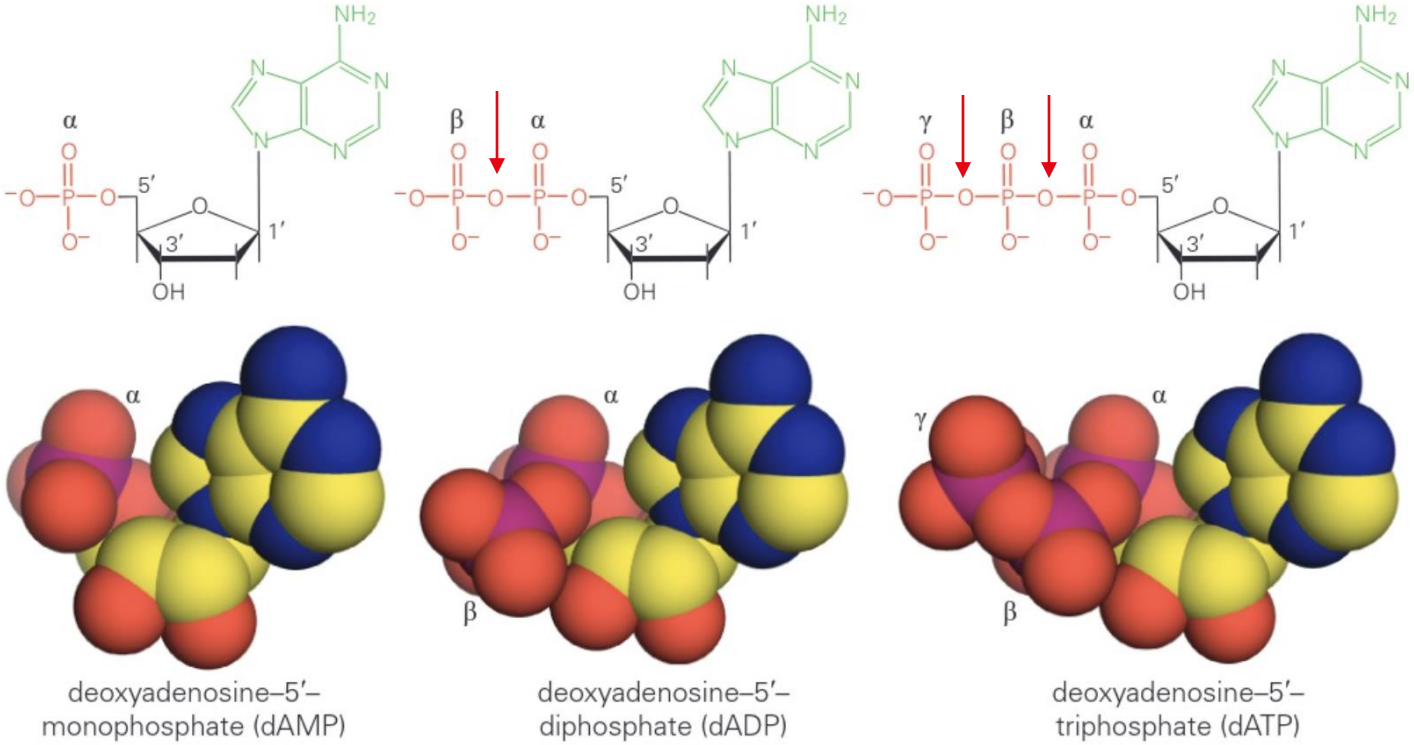
Note the difference between these chemical entities

Nucleotides – The phosphate group(s)



- Nucleotides with one, two or three phosphate groups are referred to as nucleotide **mono-**, **di-** or **tri-phosphate**.

- The three phosphate groups are called **alpha**, **beta** and **gamma**



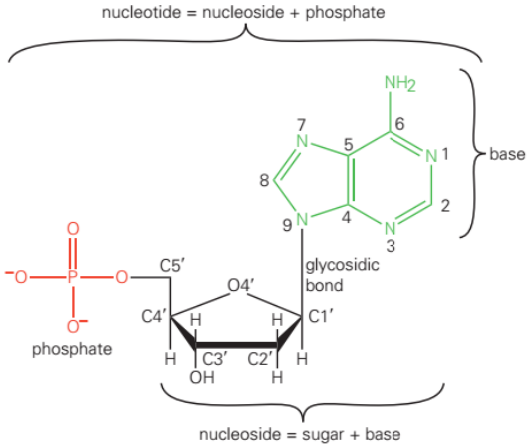
Phosphate groups are linked via **phosphoanhydride** bonds

Breakage of these bonds (i.e., in ATP) produces **~30-40 kJ/mol** of energy.

This is a fundamental process in **cell metabolism**.

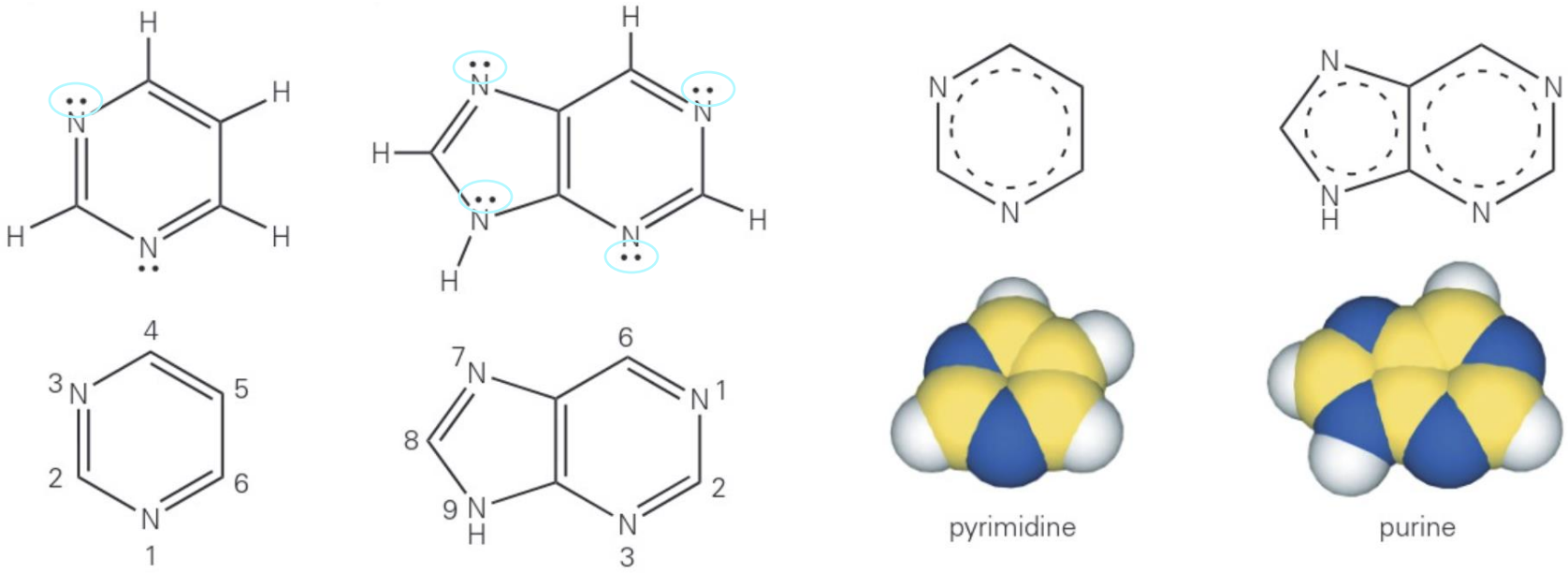
3D assembly of nucleotides (sphere representations of heavy atoms)

Nucleotides – The Bases



- The nucleotides comprising DNA and RNA are built with **5 different bases** (4 per nucleic acid type)

- The name "**base**" comes from its chemical composition - The ring systems contain **lone pairs of electrons in nitrogen atoms** being able to act as electron pair donors – so called **Lewis bases**

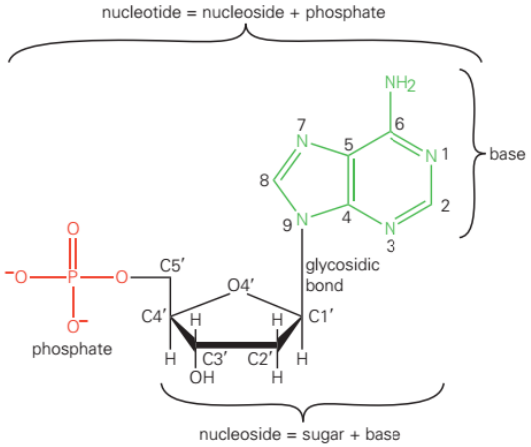


The rings are sp2 hybridized making them planar.

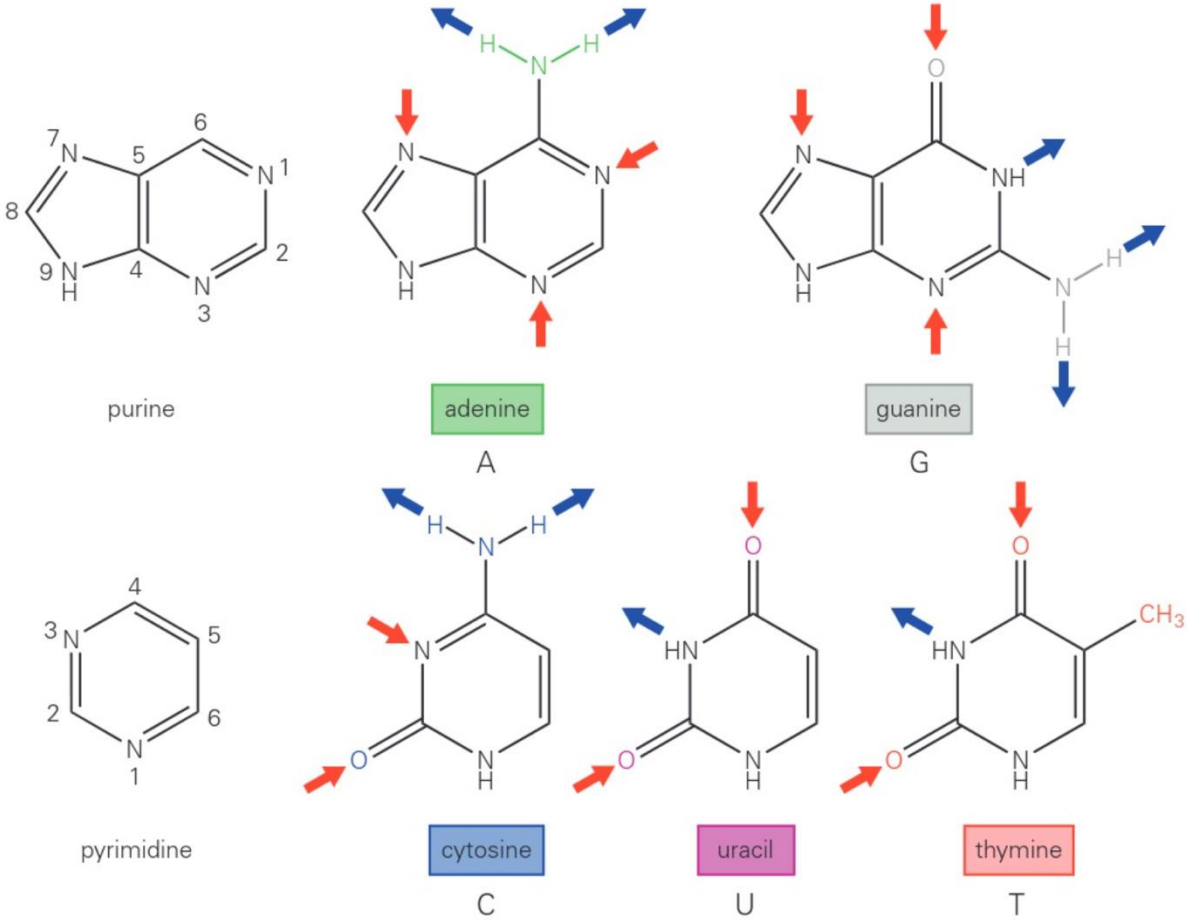
Free electrons in N atoms make them potential hydrogen bond acceptors

- Nucleotide bases in RNA and DNA are substituted forms of two heterocyclic molecules known as **pyrimidine** and **purine**

Nucleotides – The Bases



- DNA contains two substituted purines (**adenine** and **guanine**)
- DNA contains two substituted pyrimidines (**cytosine** and **thymine**)
- In RNA **thymine** is replaced by **uracil**

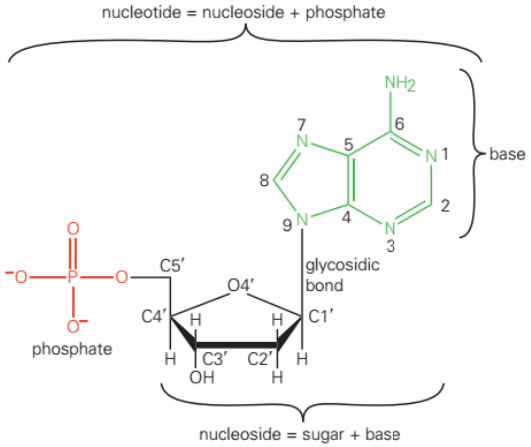


Hydrogen bond acceptors

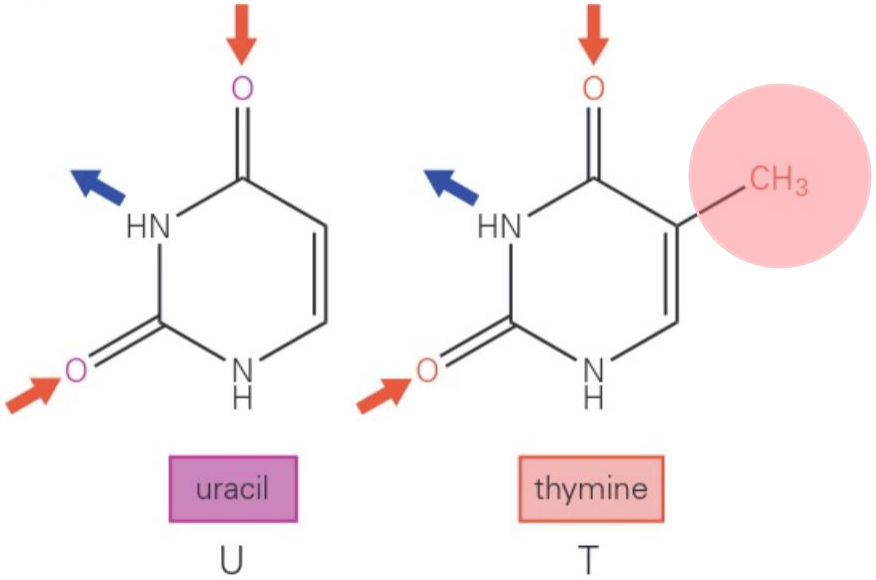
Hydrogen bond donors

The unique distribution of hydrogen bond donors and acceptors in each base are key to DNA replication and transcription

Nucleotides – Thymine vs Uracil



- **Thymine** and **Uracil** feature the same hydrogen bond donor/acceptor organization, but thymine has one extra methyl (CH_3) group
- They replace each other in **DNA (T)** versus **RNA (U)** but are mutually exclusive

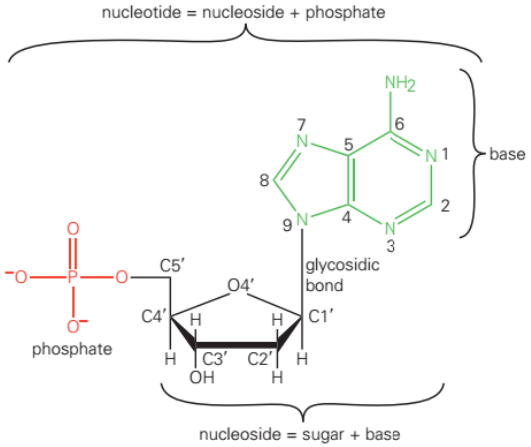


The extra **methyl group** present in thymine plays several important roles:

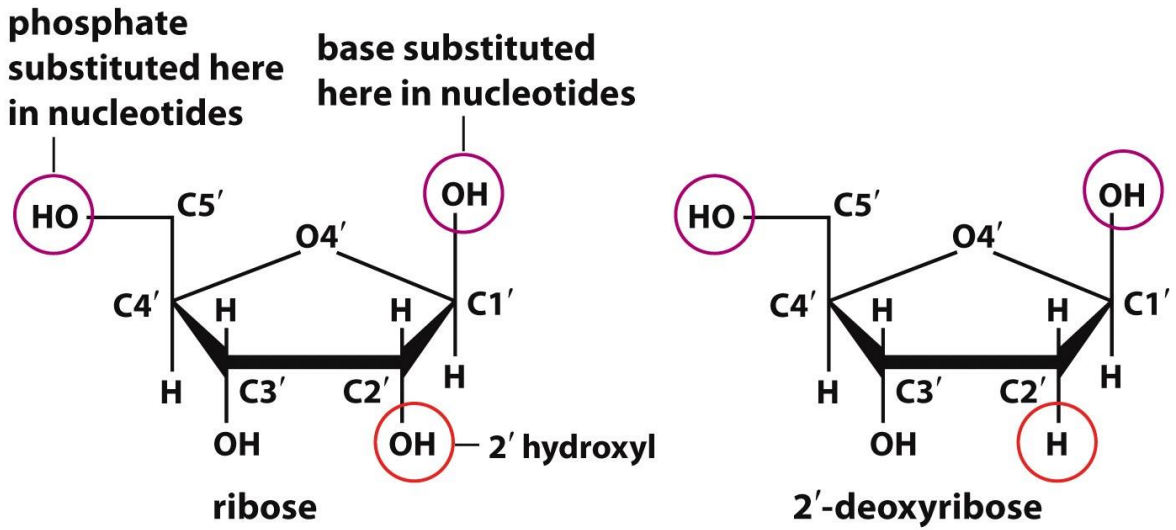
- Differentiation of DNA vs RNA by proteins
- Better packing of DNA helix due to hydrophobicity
- Detection of C->U conversion in DNA which reduces the mutation rate (improved genomic stability)
- RNA is meant to be short-lived intermediate so it uses "cheaper" uracil requiring fewer reactions to produce

- Interestingly, **other bases (e.g., cytosine) can get methylated** by certain cellular enzymes and such modifications are the basis of **epigenetic** and **gene expression regulation**.

Nucleotides – The pentose sugar



- **Ribose** is the carbohydrate used in **RNA**.
- **2'-deoxy-ribose** is the carbohydrate used in **RNA**.



Ribonucleotides

Deoxyribonucleotides

RNA = Ribonucleic Acid.

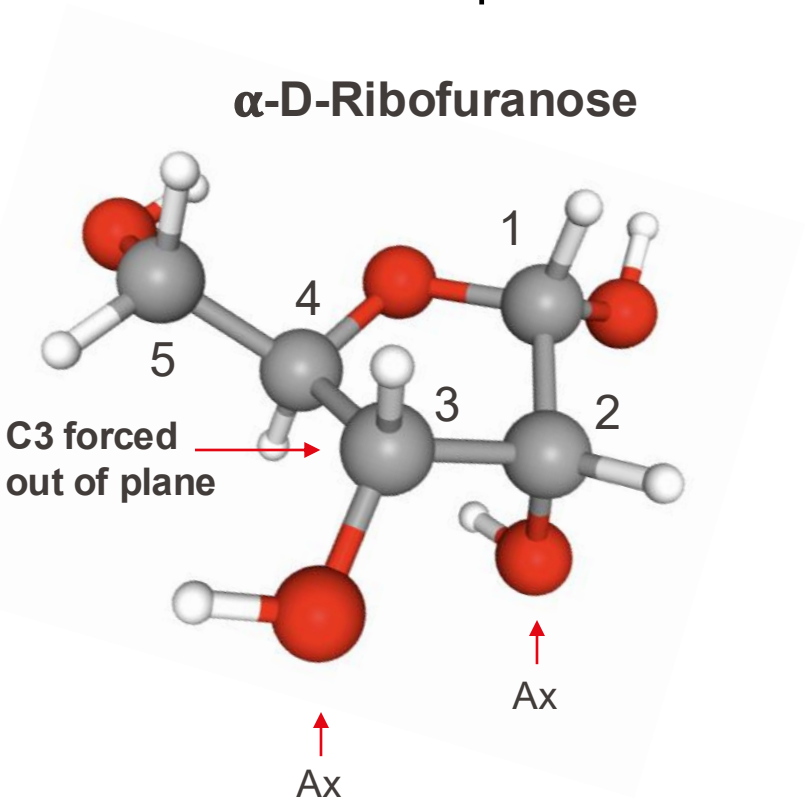
DNA = Deoxyribonucleic Acid

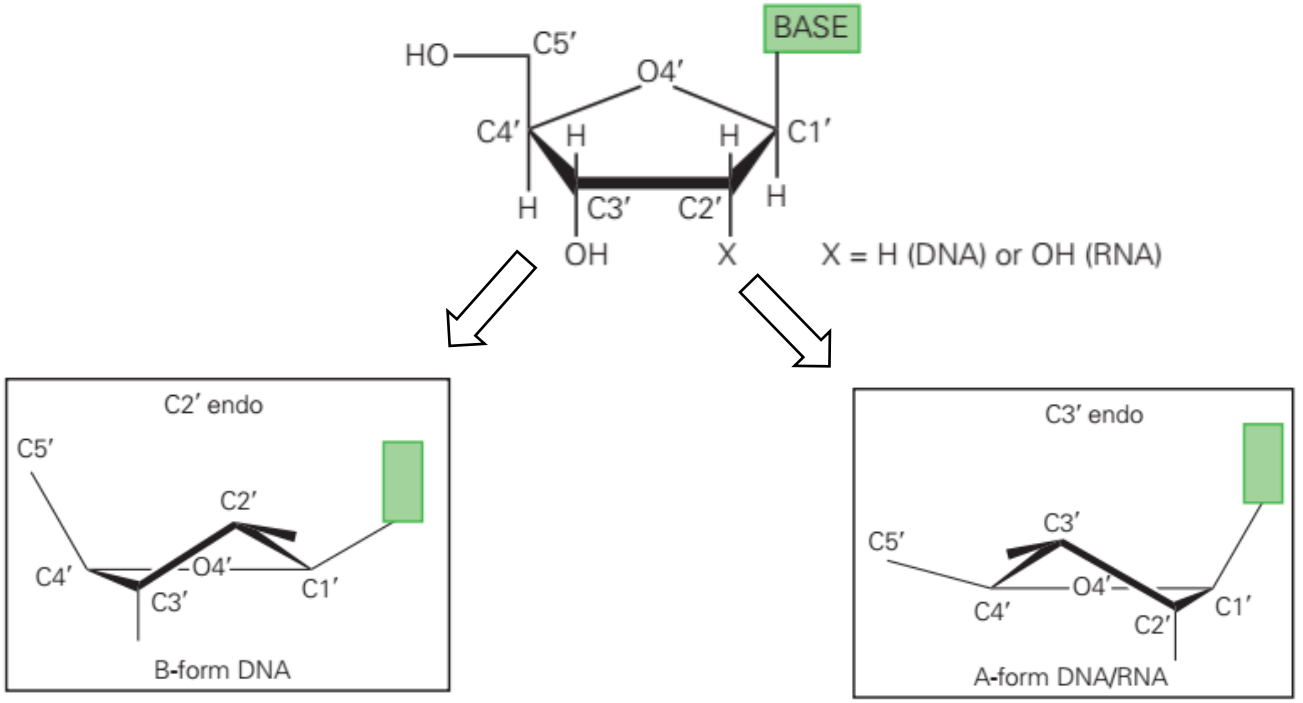
The **2' hydroxyl (OH)** group in ribose causes several important differences:

- OH can attack adjacent phosphate group in RNA causing spontaneous cleavage
- The repulsion with 3'OH causes forces C3 carbon out of the plane (C3'-pucker)
- Provides extra hydrogen bond donor and acceptor group useful for RNA recognition

Nucleotides – The pentose sugar

- **Furanose** sugars exist in the so-called "**envelope**" conformation which causes one group to be displaced out of the plane ("**pucker**")
- In energetically favorable conformations four of the atoms of the pentose ring are roughly coplanar and one is out of the plane



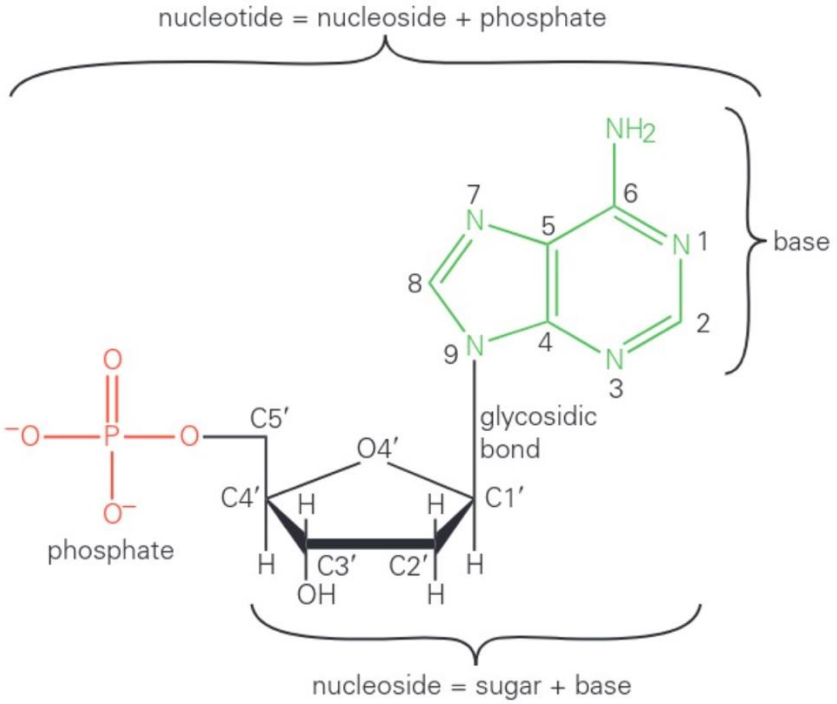


- In **DNA** and **RNA** molecules the pentose adopts a different sugar pucker conformation

Nucleotide naming

- Please note the name distinctions between **bases**, **nucleosides** and **nucleotides**

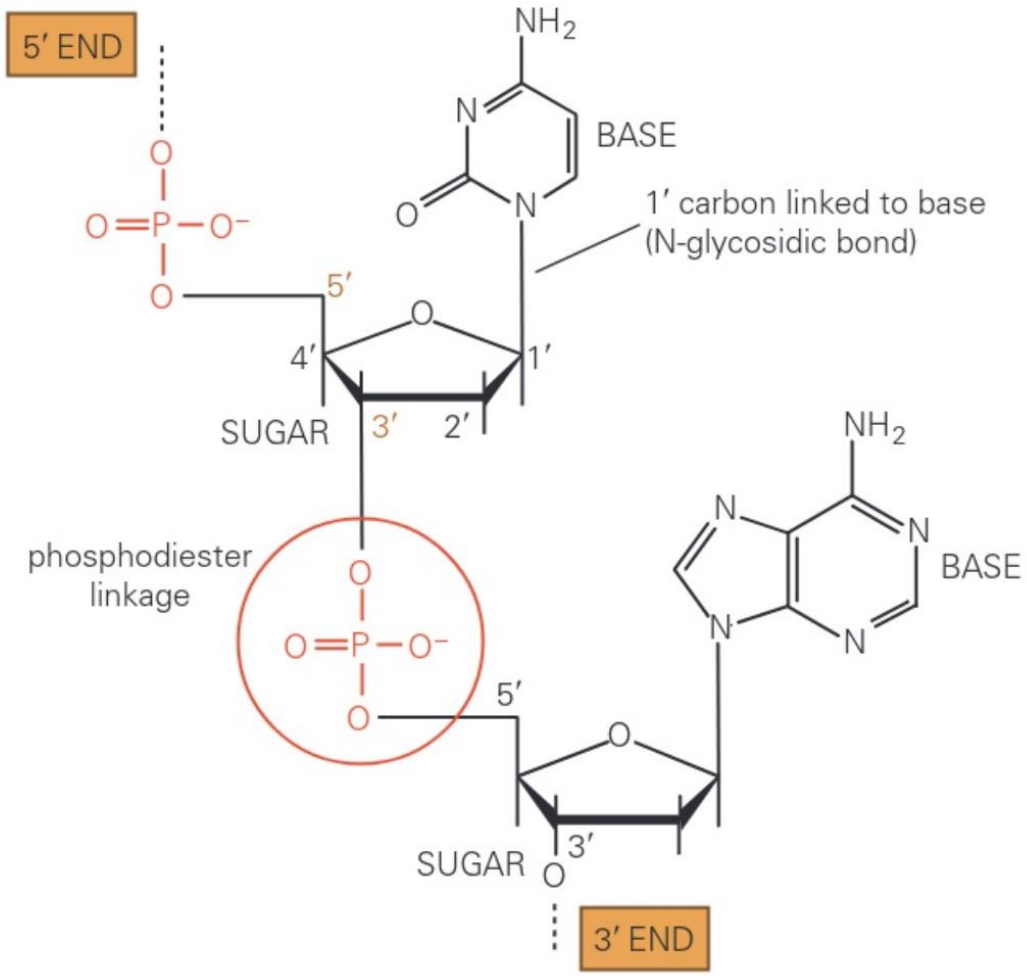
Base	Nucleosides	Nucleotides
RNA		
Adenine (A)	Adenosine (A)	Adenosine 5'-monophosphate (AMP)
Guanine (G)	Guanosine (G)	Guanosine 5'-monophosphate (GMP)
Cytosine (C)	Cytidine (C)	Cytidine 5'-monophosphate (CMP)
Uracil (U)	Uridine (U)	Uridine 5'-monophosphate (UMP)
DNA		
Adenine (A)	Deoxyadenosine (A)	Deoxyadenosine 5'-monophosphate (dAMP)
Guanine (G)	Deoxyguanosine (G)	Deoxyguanosine 5'-monophosphate (dGMP)
Cytosine (C)	Deoxycytidine (C)	Deoxycytidine 5'-monophosphate (dCMP)
Thymine (T)	Deoxythymidine (T)	Deoxythymidine 5'-monophosphate (dTMP)



It is very important that you know the structures of all 5 bases, nucleotides and nucleosides!!!

Nucleic acids are polymers of nucleotides

• Nucleotides are joined together in DNA and RNA by the formation of a **phosphodiester bond** between the **3' carbon of one nucleotide and the 5' of the subsequent nucleotide**



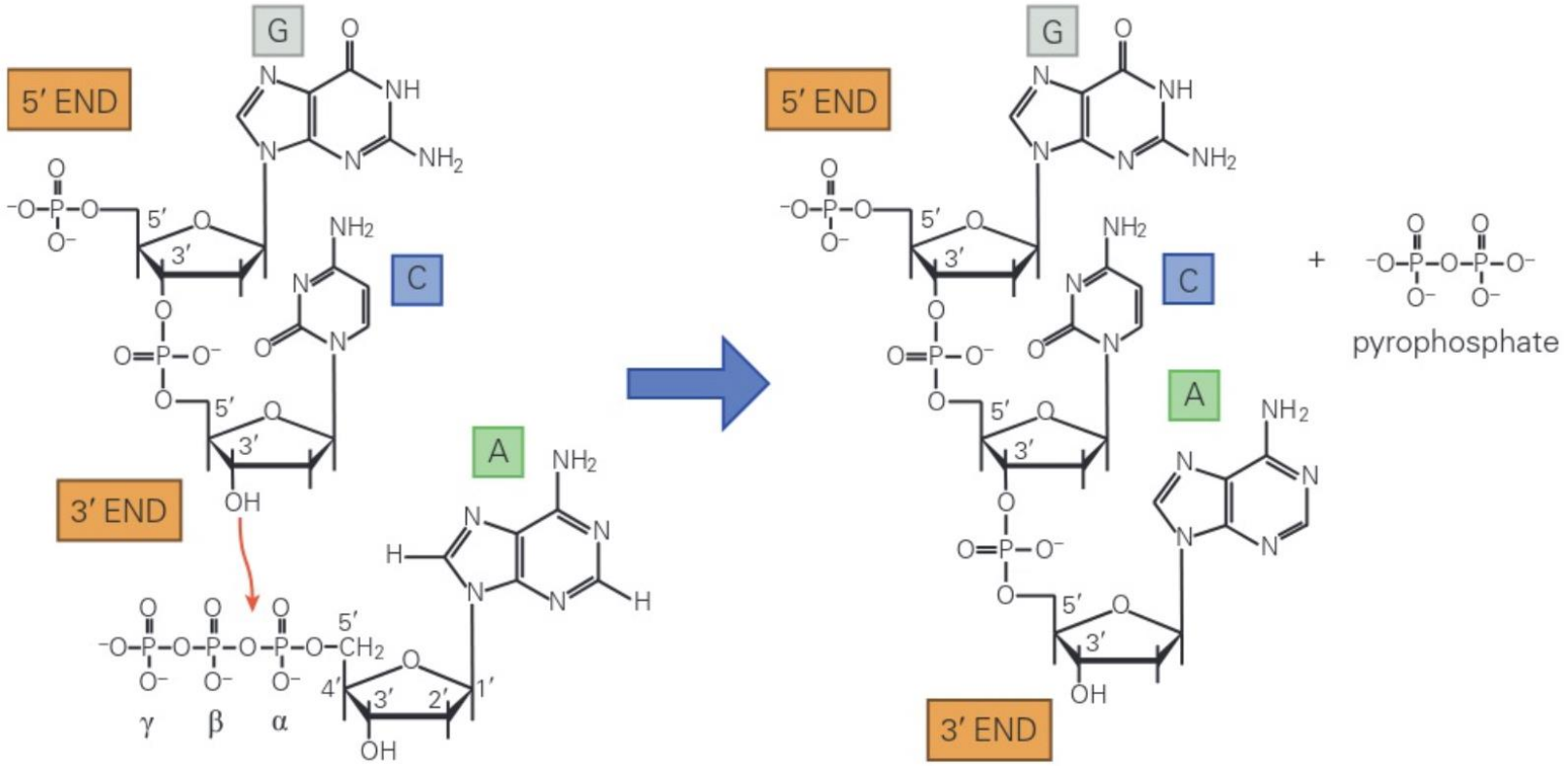
• The phosphate groups are **negatively charged** (anion nature) which represent an important determinant of the 3D structure of nucleic acids

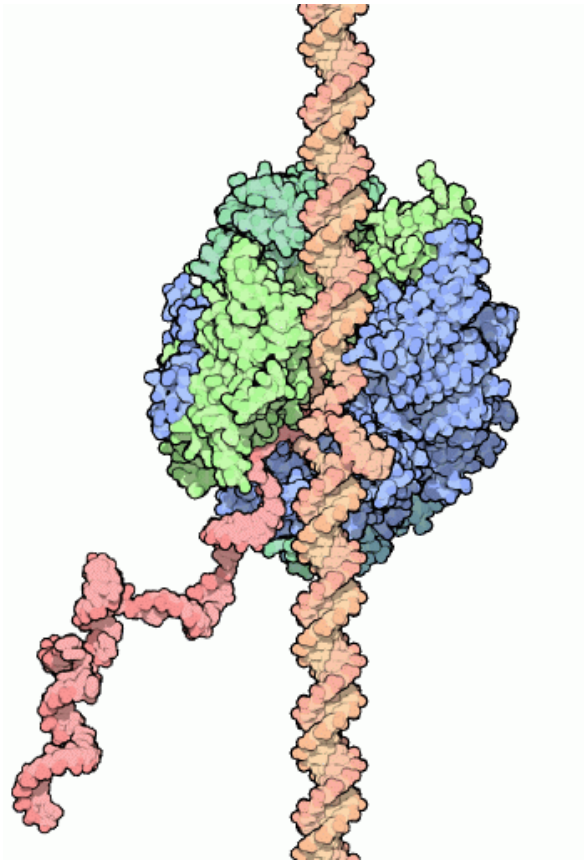
• **Phosphate-Sugar** linkages represent the backbone of nucleic acid polymers (repeating pattern of identical 2 units)

• The **bases remain accessible** with the potential hydrogen bond donors/acceptors being relatively unhindered by the backbone (rotation possible around glycosidic bond).

Stepwise addition of nucleotides

- The synthesis of new molecules of DNA and RNA involves the stepwise addition of nucleotides to one end of the chain. The enzymes that catalyze this reaction are called **polymerases**.
- The **triphosphate group** is high in energy and the hydrolysis of **phosphoanhydride bonds** drives the reaction



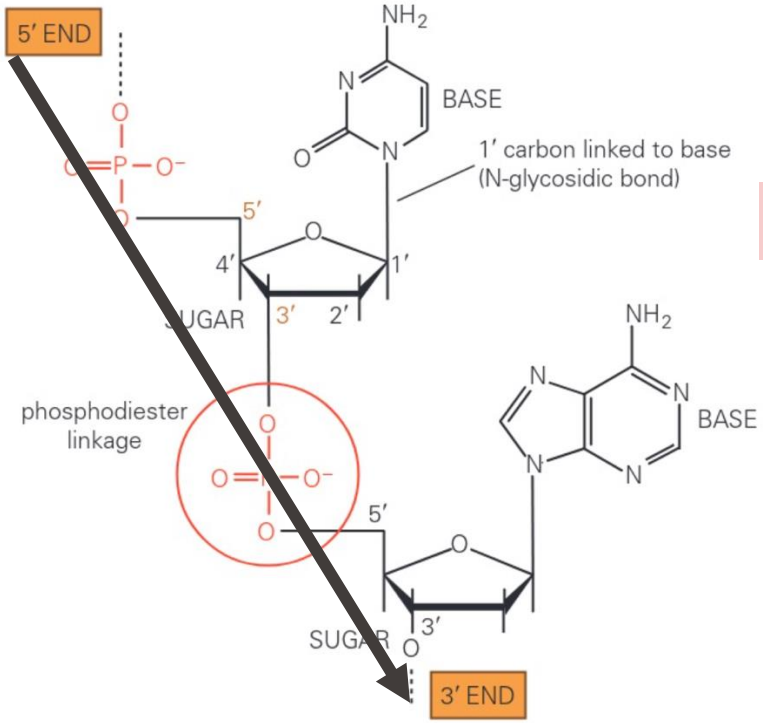


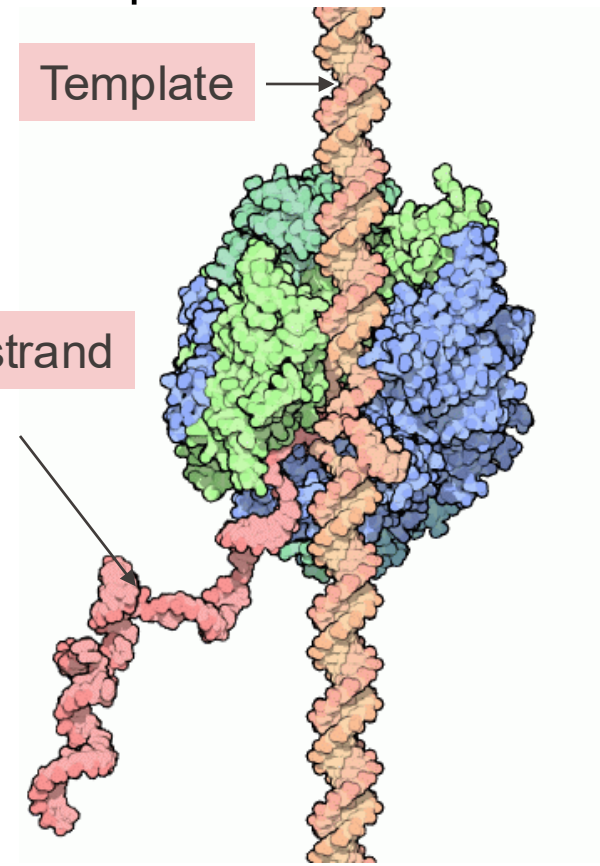
Model of RNA polymerase in action (adapted from rcsb.org).

Stepwise addition of nucleotides

- DNA and RNA synthesis are **template directed** – The polymerases use a template strand to select each nucleotide to be added to the newly synthesized strand.
- This template-based assembly represents the essence of genetic data transfer during cell division as well as protein synthesis based on the corresponding DNA gene sequence

- **3' -> 5' phosphodiester** linkage allows to impose directionality to the growing strand
- By convention DNA sequences are written from **5' to the 3' end.**
- All known polymerases assemble nucleic acids in 5' -> 3' direction

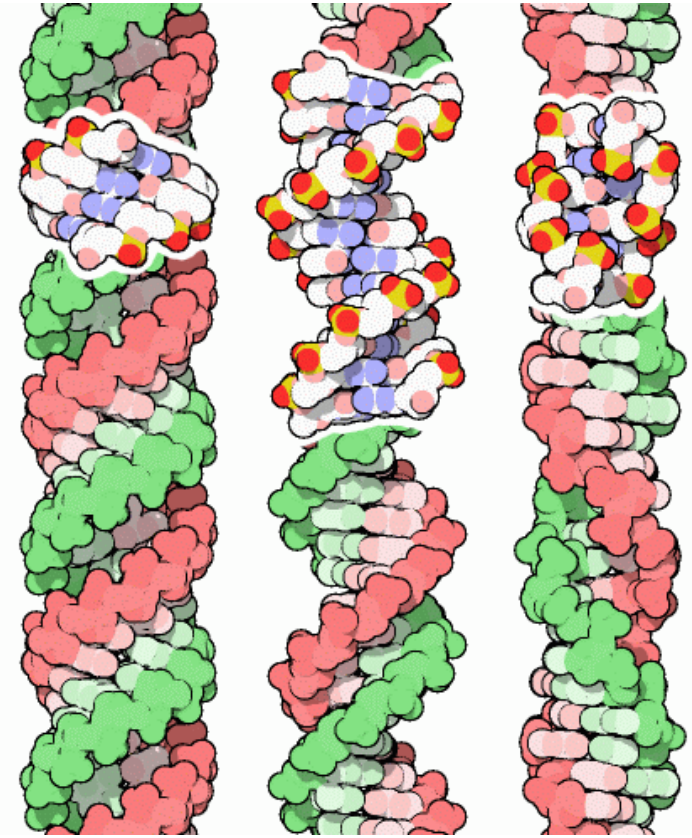




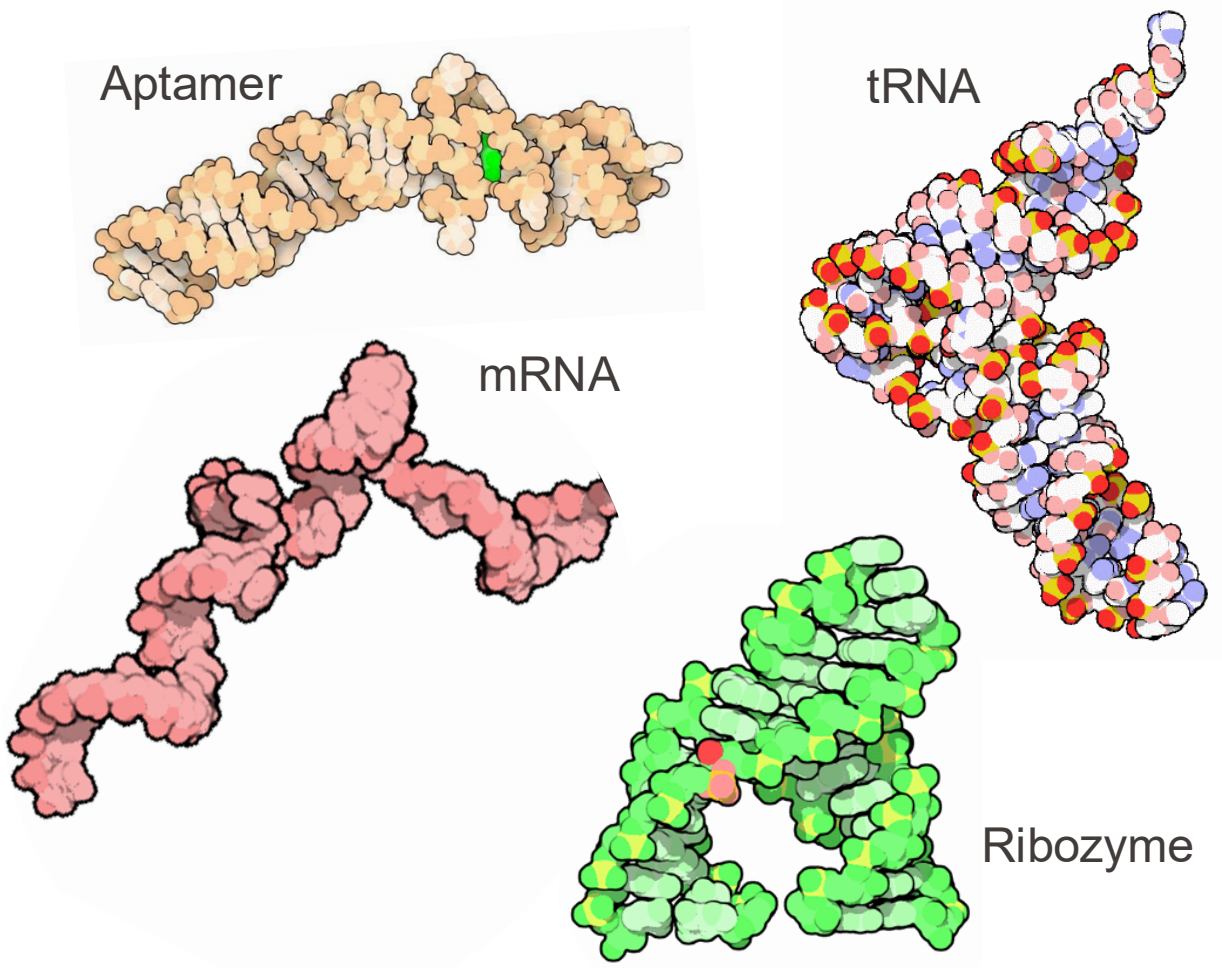
Model of RNA polymerase in action (adapted from rcsb.org).

3D assembly of DNA and RNA

- DNA assembles into double-stranded helices
- RNA assembly varies but is usually single-stranded.



A-form B-form Z-form



- 3D structure is formed through a combination of covalent and non-covalent interactions

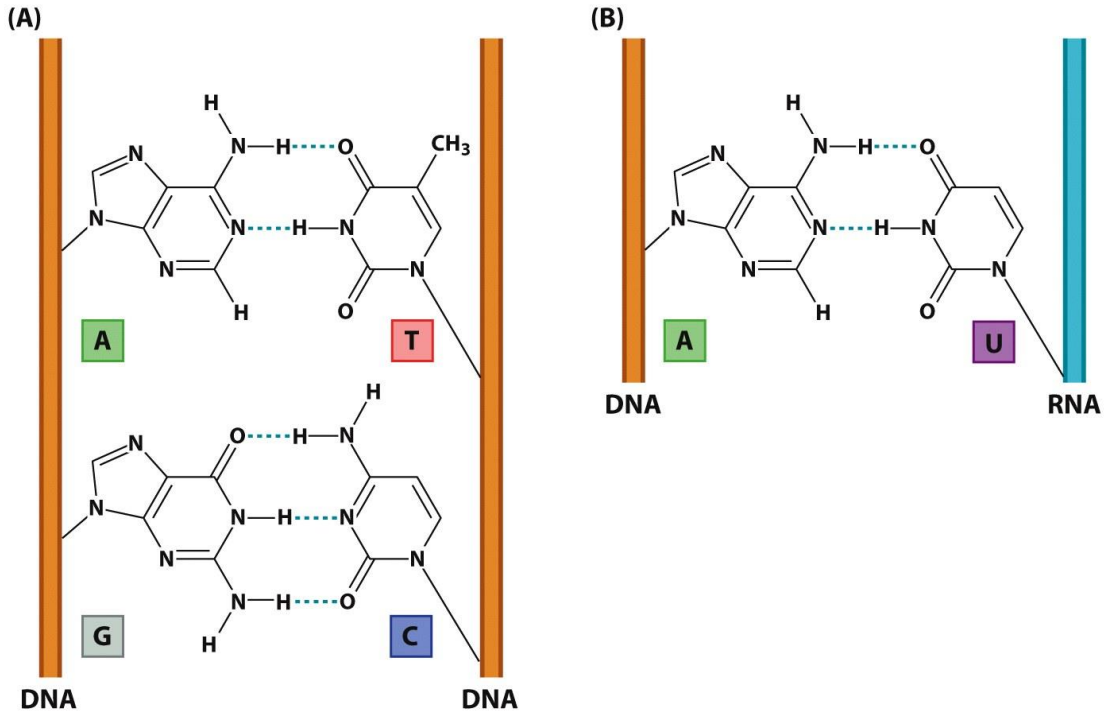
The structure of DNA

- DNA forms a **double helix** with **antiparallel strands**
- Bases face towards the inside of the helix and the phosphate backbone groups are on the outside.
- The pairing is achieved through **hydrogen bonds created between complementary bases**

2 hydrogen bonds

GC pair is ~5-10 kJ/mol more stable than AT

3 hydrogen bonds



The Watson-Crick base pairs: A-T, G-C and A-U

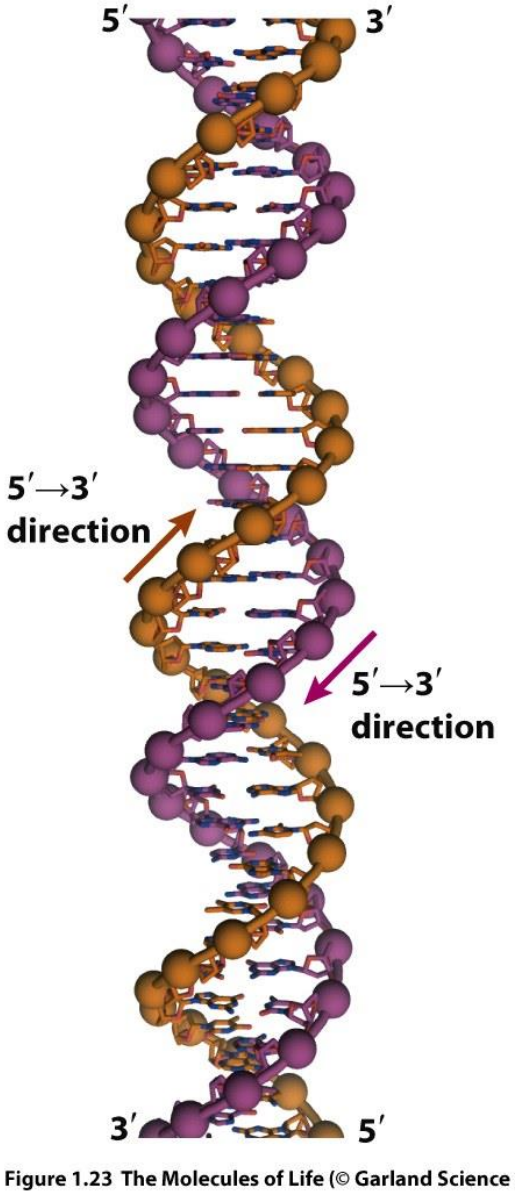


Figure 1.23 The Molecules of Life (© Garland Science 2

Phosphate groups in spheres

The structure of DNA - Non-covalent Interactions

- In addition to hydrogen bonding the double helix is stabilized by **stacking of base pairs**

Beautiful example on van der Waals interactions given that the radius of **carbon and nitrogen are 1.7 and 1.6 Å** respectively – the observed **rise per base-pair is 3.3-3.4 Å**.

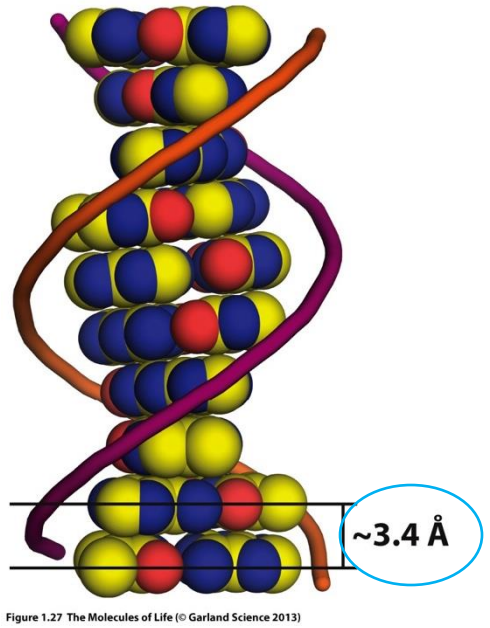
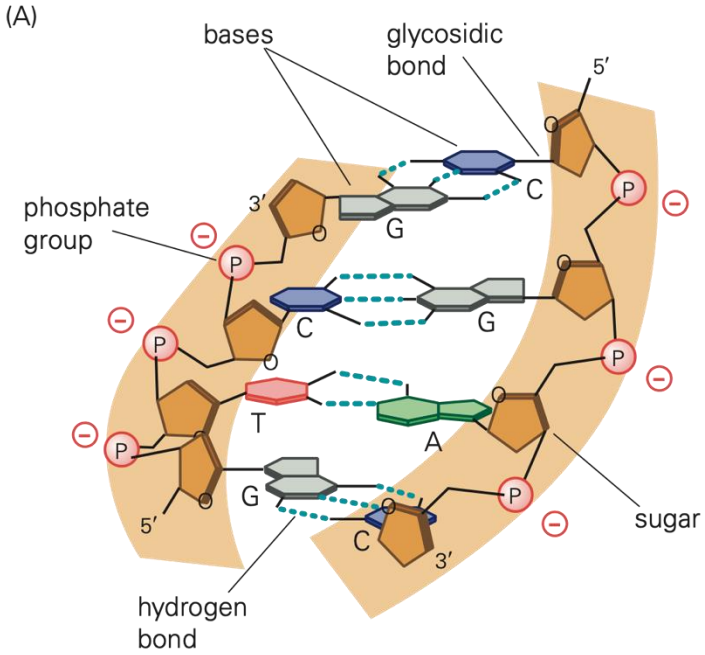


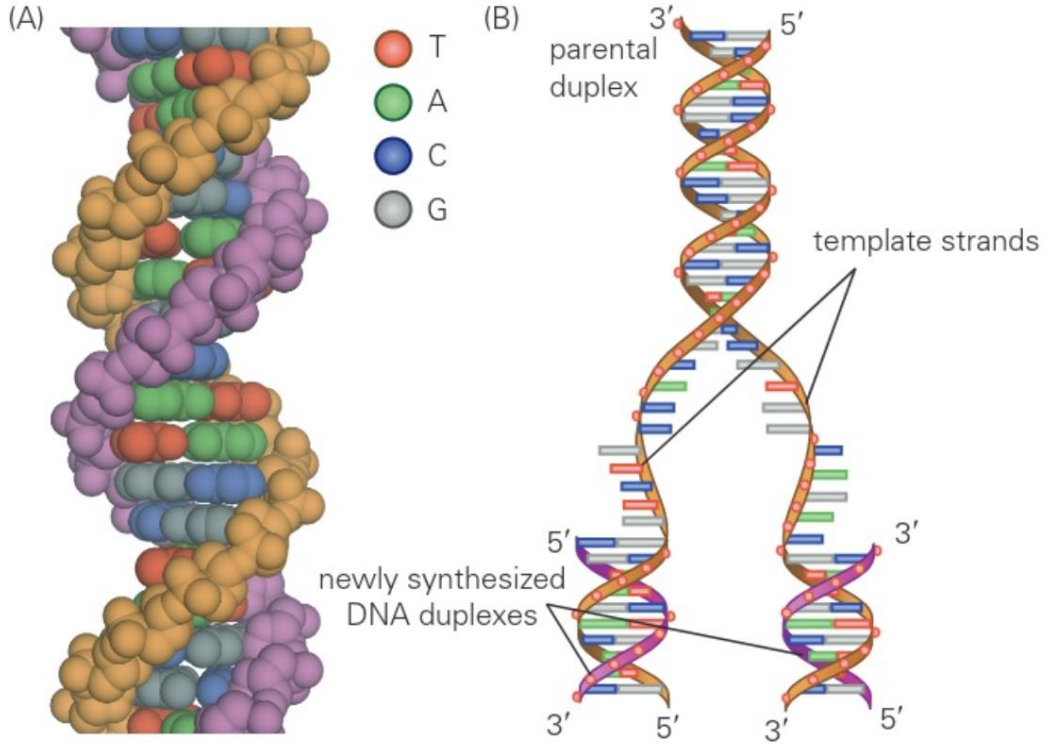
Figure 1.27 The Molecules of Life (© Garland Science 2013)



- Electrostatic interactions assure proper backbone geometry. Allowing for interactions with ions and water and minimizing repulsion between phosphates
- Base stacking is supported by solvent exclusion and **van der Waals interactions**

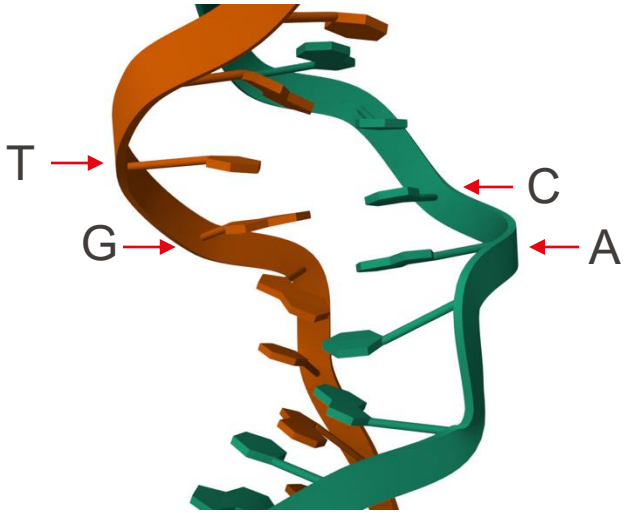
DNA Replication relies on base complementarity

- **DNA polymerase** is an enzyme that binds to exposed template DNA strand and adds complementary bases in 5’ -> 3’ direction
- The end result are two DNA strands that are **reverse complements** of each other



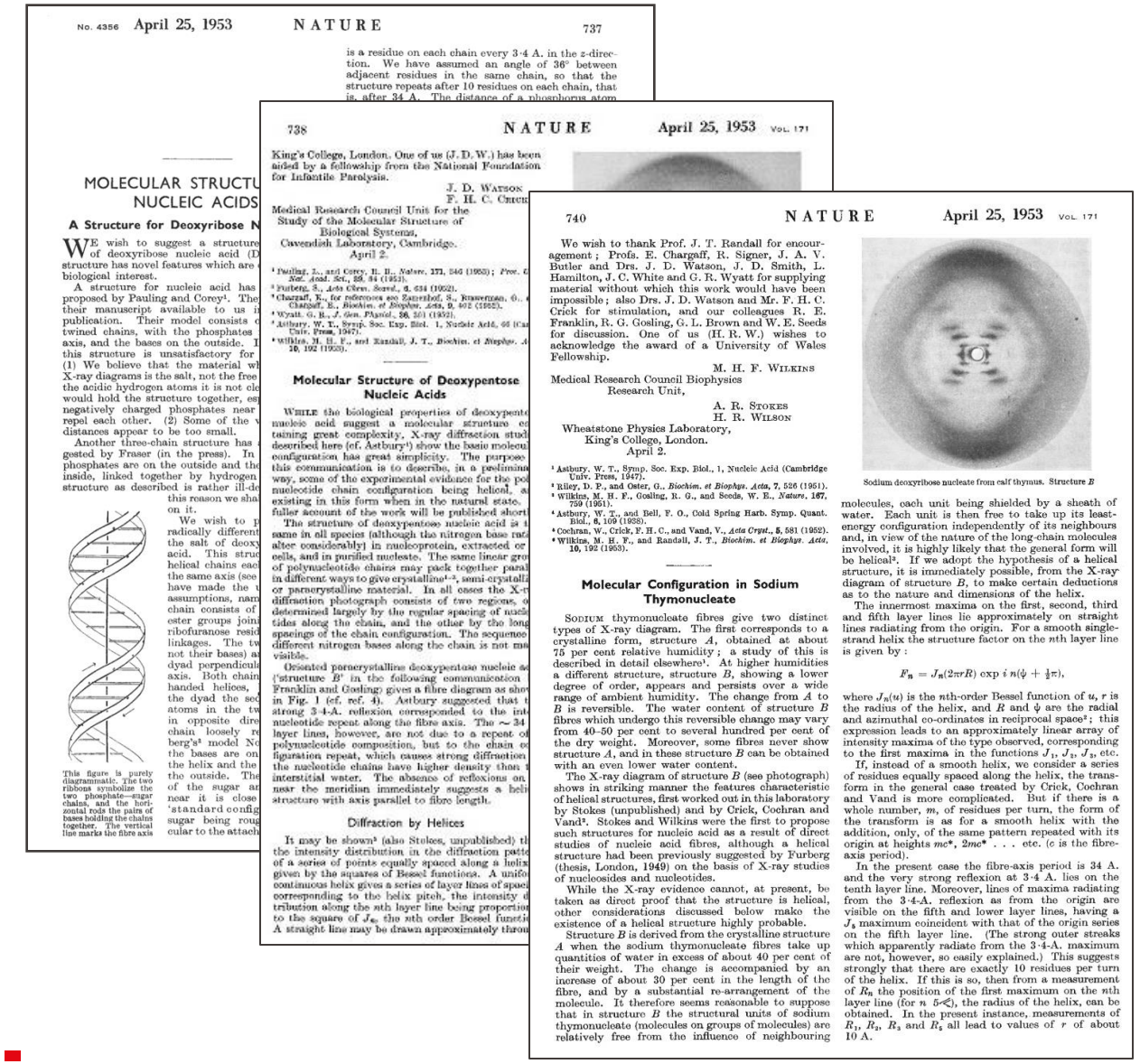
Complementary strands

2 consecutive missmatches



- Mismatching base pairs result in disruption of local conformation which is detected by the DNA polymerase causing it to stall and excise the problematic base (“**proofreading**”)

Historical detour - How was DNA structure solved?



No. 4356 April 25, 1953 NATURE 737

is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom

MOLECULAR STRUCTU
NUCLEIC ACIDS

A Structure for Deoxyribose N

WE wish to suggest a structure of deoxyribose nucleic acid (D structure has novel features which are biological interest.

A structure for nucleic acid has proposed by Pauling and Corey[1]. The their manuscript available to us i publication. Their model consists twined chains, with the phosphates axis, and the bases on the outside. this structure is unsatisfactory for (1) We believe that the material wh X-ray diagrams is the salt, not the free the acidic hydrogen atoms it is not cle would hold the structure together, es negatively charged phosphates near repel each other. (2) Some of the distances appear to be too small.

Another three-chain structure has gested by Fraser (in the press). In phosphates are on the outside and the inside, linked together by hydrogen structure as described is rather ill-de this reason we sha on it.

We wish to radically different the salt of deoxy acid. This struc helical chains eac the same axis (see have made the assumptions, nam chain consists of ester groups join ribofuranose resid linkages. The t not their bases) a dyad perpendicul axis. Both chain handed helices, the dyad the se atoms in the t in opposite dire chain loosely berg's[8] model N the bases are on the helix and the the outside. The of the sugar a near it is close 'standard config sugar being roug cular to the attach

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

738 NATURE April 25, 1953 Vol. 171

King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. WATSON
F. H. C. CRICK

Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

[1] Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. Nat. Acad. Sci.*, **39**, 84 (1953).
[2] Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).
[3] Chargaff, E., for references see Zamenhof, S., Brawerman, G., Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).
[4] Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).
[5] Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Cambridge Univ. Press, 1947).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

WHILE the biological properties of deoxypento nucleic acid suggest a molecular structure co taining great complexity, X-ray diffraction studi described here (cf. Astbury[4]) show the basic molecu configuration has great simplicity. The purpose this communication is to describe, in a preliminar way, some of the experimental evidence for the pol nucleotide chain configuration being helical, a existing in this form when in the natural state. fuller account of the work will be published short

The structure of deoxypentose nucleic acid is same in all species (although the nitrogen base ra alter considerably) in nucleoprotein, extracted or cells, and in purified nucleate. The same linear gro of polynucleotide chains may pack together paral in different ways to give crystalline[1-3], semi-crystall or paracrystalline material. In all cases the X-r diffraction photograph consists of two regions, o determined largely by the regular spacing of nucl tides along the chain, and the other by the long spacings of the chain configuration. The sequence different nitrogen bases along the chain is not mu visible.

Oriented paracrystalline deoxypentose nucleic a ('structure B' in the following communication Franklin and Gosling) gives a fibre diagram as sho in Fig. 1 (cf. ref. 4). Astbury suggested that t strong 3·4-A. reflexion corresponded to the int nucleotide repeat along the fibre axis. The ~ 34 layer lines, however, are not due to a repeat o polynucleotide composition, but to the chain c figuration repeat, which causes strong diffraction the nucleotide chains have higher density than t interstitial water. The absence of reflexions on near the meridian immediately suggests a heli structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[5] (also Stokes, unpublished) th the intensity distribution in the diffraction patte of a series of points equally spaced along a helix given by the squares of Bessel functions. A unifo continuous helix gives a series of layer lines of spaci corresponding to the helix pitch, the intensity d tribution along the nth layer line being proportion to the square of J_n, the nth order Bessel functi A straight line may be drawn approximately throu

740 NATURE April 25, 1953 Vol. 171

We wish to thank Prof. J. T. Randall for encouragement; Profs. E. Chargaff, R. Signer, J. A. V. Butler and Drs. J. D. Watson, J. D. Smith, L. Hamilton, J. C. White and G. R. Wyatt for supplying material without which this work would have been impossible; also Drs. J. D. Watson and Mr. F. H. C. Crick for stimulation, and our colleagues R. E. Franklin, R. G. Gosling, G. L. Brown and W. E. Seeds for discussion. One of us (H. R. W.) wishes to acknowledge the award of a University of Wales Fellowship.

M. H. F. WILKINS
Medical Research Council Biophysics Research Unit,

A. R. STOKES
H. R. WILSON
Wheatstone Physics Laboratory, King's College, London.
April 2.

[1] Astbury, W. T., Symp. Soc. Exp. Biol., 1, Nucleic Acid (Cambridge Univ. Press, 1947).
[2] Riley, D. P., and Oster, G., *Biochim. et Biophys. Acta*, **7**, 526 (1951).
[3] Wilkins, M. H. F., Gosling, R. G., and Seeds, W. E., *Nature*, **167**, 759 (1951).
[4] Astbury, W. T., and Bell, F. O., Cold Spring Harb. Symp. Quant. Biol., **6**, 109 (1938).
[5] Cochran, W., Crick, F. H. C., and Vand, V., *Acta Cryst.*, **5**, 581 (1952).
[6] Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Sodium deoxyribose nucleate from calf thymus. Structure *B*

Molecular Configuration in Sodium Thymonucleate

SODIUM thymonucleate fibres give two distinct types of X-ray diagram. The first corresponds to a crystalline form, structure *A*, obtained at about 75 per cent relative humidity; a study of this is described in detail elsewhere[1]. At higher humidities a different structure, structure *B*, showing a lower degree of order, appears and persists over a wide range of ambient humidity. The change from *A* to *B* is reversible. The water content of structure *B* fibres which undergo this reversible change may vary from 40–50 per cent to several hundred per cent of the dry weight. Moreover, some fibres never show structure *A*, and in these structure *B* can be obtained with an even lower water content.

The X-ray diagram of structure *B* (see photograph) shows in striking manner the features characteristic of helical structures, first worked out in this laboratory by Stokes (unpublished) and by Crick, Cochran and Vand[2]. Stokes and Wilkins were the first to propose such structures for nucleic acid as a result of direct studies of nucleic acid fibres, although a helical structure had been previously suggested by Furberg (thesis, London, 1949) on the basis of X-ray studies of nucleosides and nucleotides.

While the X-ray evidence cannot, at present, be taken as direct proof that the structure is helical, other considerations discussed below make the existence of a helical structure highly probable.

Structure *B* is derived from the crystalline structure *A* when the sodium thymonucleate fibres take up quantities of water in excess of about 40 per cent of their weight. The change is accompanied by an increase of about 30 per cent in the length of the fibre, and by a substantial re-arrangement of the molecule. It therefore seems reasonable to suppose that in structure *B* the structural units of sodium thymonucleate (molecules on groups of molecules) are relatively free from the influence of neighbouring molecules, each unit being shielded by a sheath of water. Each unit is then free to take up its least-energy configuration independently of its neighbours and, in view of the nature of the long-chain molecules involved, it is highly likely that the general form will be helical[3]. If we adopt the hypothesis of a helical structure, it is immediately possible, from the X-ray diagram of structure *B*, to make certain deductions as to the nature and dimensions of the helix.

The innermost maxima on the first, second, third and fifth layer lines lie approximately on straight lines radiating from the origin. For a smooth single-strand helix the structure factor on the nth layer line is given by:

$$F_n = J_n(2\pi rR) \exp i\, n(\psi + \tfrac{1}{2}\pi),$$

where $J_n(u)$ is the nth-order Bessel function of u, r is the radius of the helix, and R and ψ are the radial and azimuthal co-ordinates in reciprocal space[4]; this expression leads to an approximately linear array of intensity maxima of the type observed, corresponding to the first maxima in the functions J_1, J_4, J_7, etc.

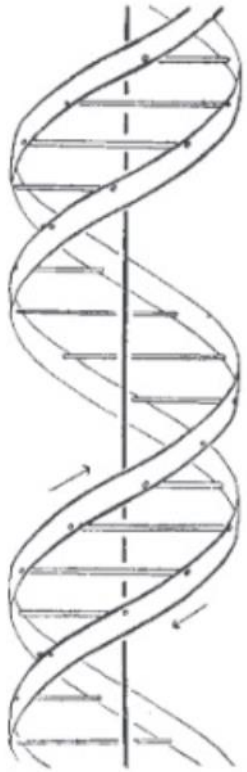
If, instead of a smooth helix, we consider a series of residues equally spaced along the helix, the transform in the general case treated by Crick, Cochran and Vand is more complicated. But if there is a whole number, m, of residues per turn, the form of the transform is as for a smooth helix with the addition, only, of the same pattern repeated with its origin at heights mc^*, $2mc^*$ . . . etc. (c is the fibre-axis period).

In the present case the fibre-axis period is 34 A. and the very strong reflexion at 3·4 A. lies on the tenth layer line. Moreover, lines of maxima radiating from the 3·4-A. reflexion as from the origin are visible on the fifth and lower layer lines, having a J_5 maximum coincident with that of the origin series on the fifth layer line. (The strong outer streaks which apparently radiate from the 3·4-A. maximum are not, however, so easily explained.) This suggests strongly that there are exactly 10 residues per turn of the helix. If this is so, then from a measurement of R_n the position of the first maximum on the nth layer line (for n 5<), the radius of the helix, can be obtained. In the present instance, measurements of R_1, R_2, R_3 and R_5 all lead to values of r of about 10 A.

3 papers published in Nature in **1953**:

- James Watson & Francis Crick
- Maurice Wilkins
- Rosalind Franklin & Raymond Gosling

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate—sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis



Historical detour - What was known at the time?

• Entering 1950s there was a pretty good understanding of the chemical nature of nucleotides and resulting DNA, including all relevant covalent bonds between building blocks.
• Additionally, Erwin Chargaff in 1950 published his findings on the ratio of nucleotides, that ultimately led to the establishment of Chragaff rules.

• **Chargaff rules:**
  %A ≈ %T
  %G ≈ %C
  %(A+T) ≠ %(G+C)

University of Bern

Rudolf Signer

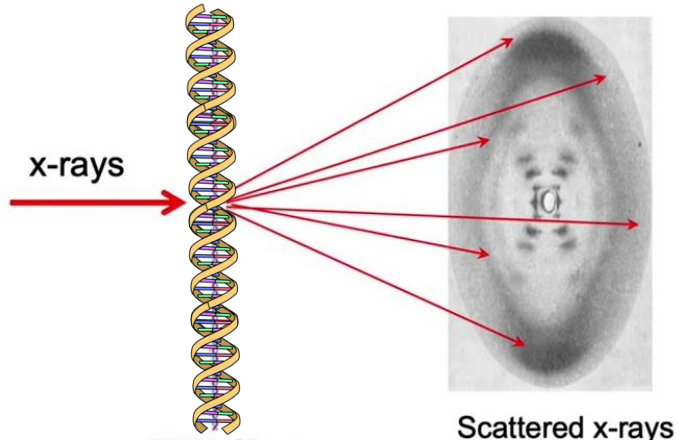
Isolation of calf DNA

Kings College London

Rosalind Franklin

Raymond Gosling

Maurice Wilkins



Historical detour - The Secret of Image 51

- Spots in this diffraction image represent a plane in DNA molecule that regularly repeats leading to constructive interference
- The X and Y axis have the units of 1/distance and can be connected to the DNA planes by Bragg's law

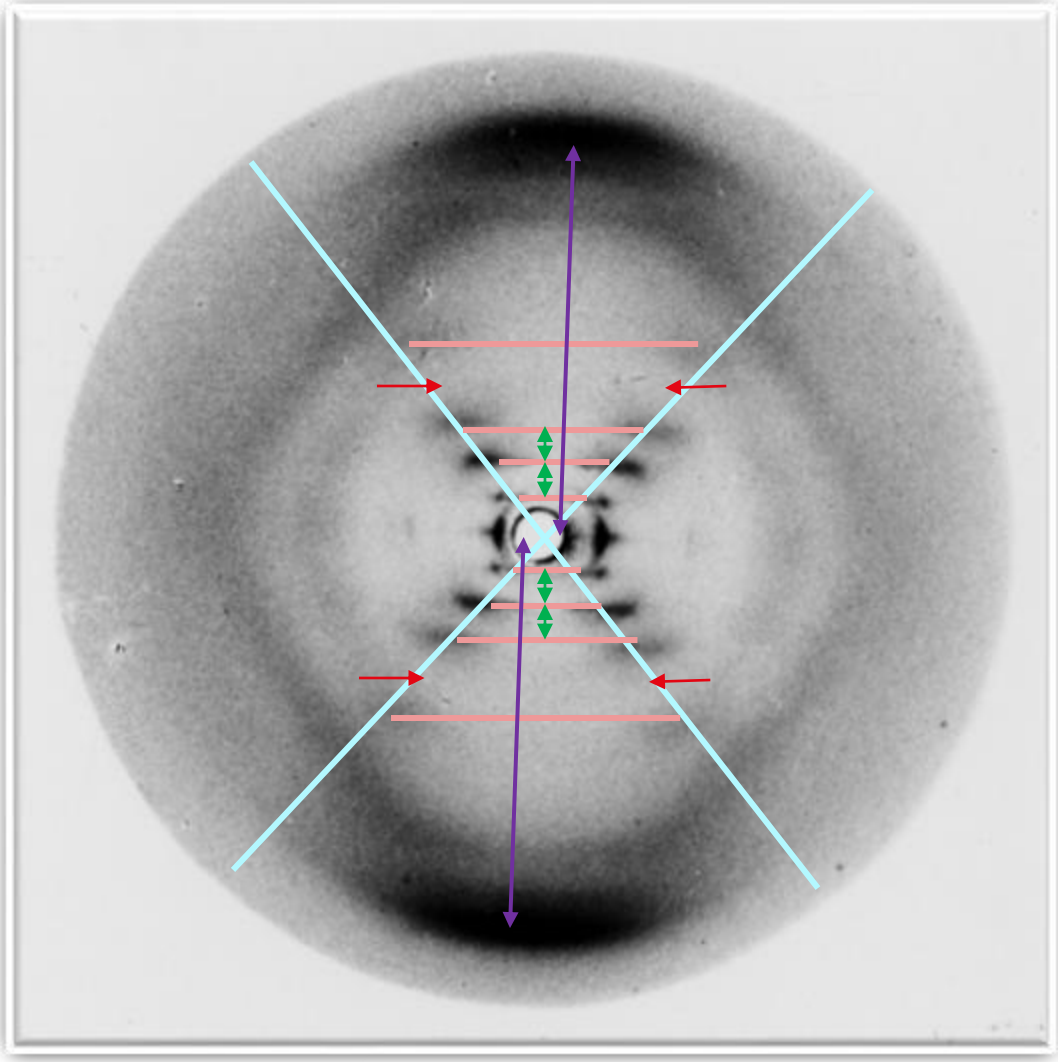
- The X-shape indicates a helix - known based on symmetry operators for helical assemblies

- The horizontal distances can be used to calculate helix width - Resulting in **~20Å**

- The spacing of vertical spots corresponds to one rise of the helix ("helical pitch") - Resulting in **~34Å**

- The position of the large peak corresponds to distance between large chemical groups (i.e., bases) - Resulting in **~3.4Å**

- Missing 4th layer indicates absence of perfect regularity expected for single helix - Must be more than one strand (but even number)



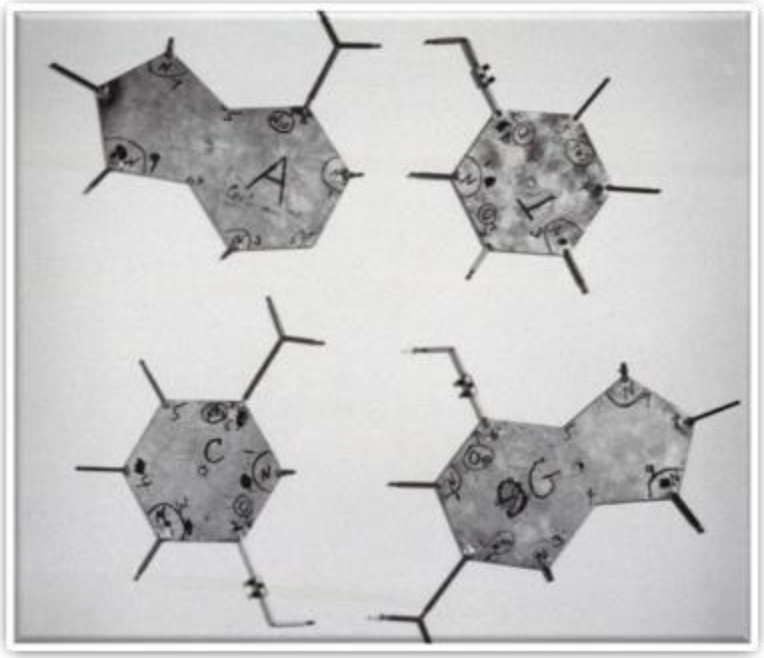
Rosalind Franklin's X-ray image of DNA ("Photo 51")

Historical detour - Crick and Watson

• James Watson and Fracis Crick were junior scientists (PhD student and a postdoc) specializing in molecular modeling at Cambridge, interested in the DNA structure problem
• Controversially, Franklin decided to leave the King's College and her data was shared to Watson and Crick without her permission

DNA model designed by Crick and Watson

Crick and Watson integrated Chargaff's rules into model building assuming that the bases must pair in certain way:



Historical detour - James Watson and base pairing

Historical detour - Nobel Prize in 1962

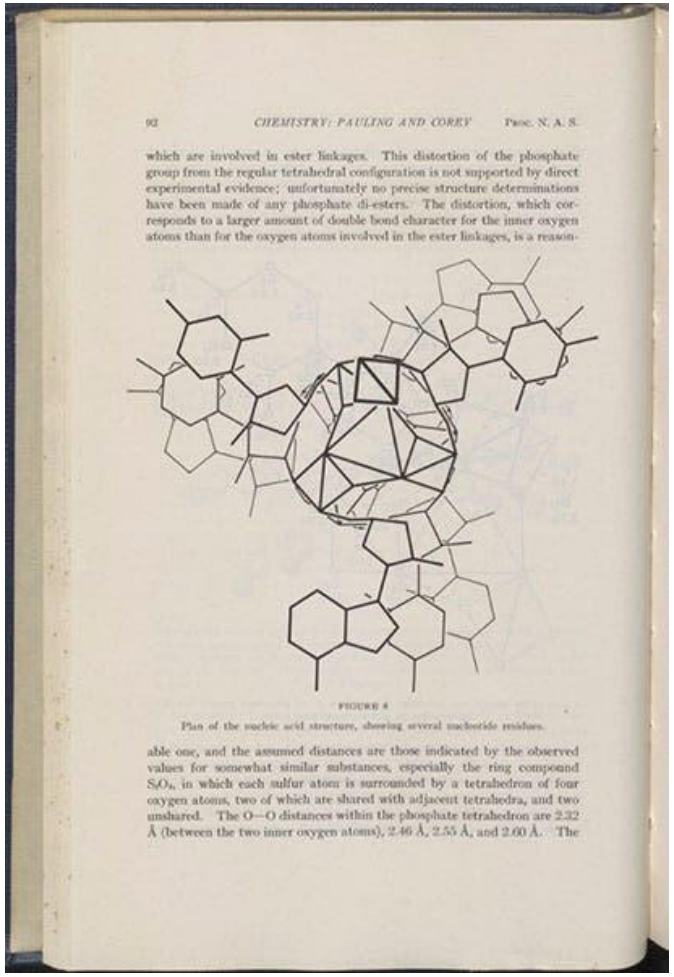
- James Watson, Francis Crick and Maurice Wilkins received the Nobel Prize in Physiology and Medicine in 1962

- Interestingly, Max Perutz and John Kendrew, who were the mentors of Watson and Crick, also won a Nobel Prize in the same year (Chemistry)

- Rosalind Franklin passed away in 1958 and her contribution was not acknowledged until several decades later

Historical detour - Linus Pauling model

• Relying on incomplete X-ray data, Linus Pauling (2-times Nobel Laureate) proposed a structure of DNA as a triple helix

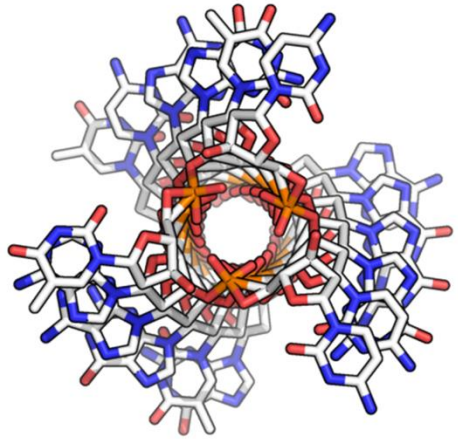


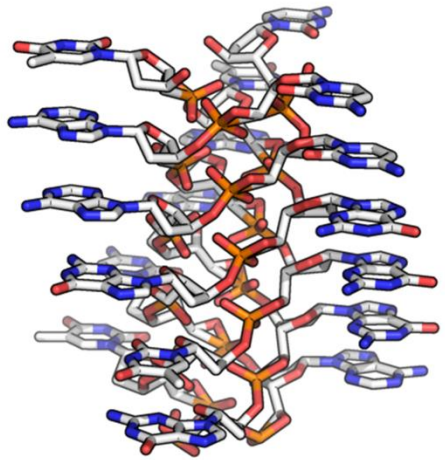
92 CHEMISTRY: PAULING AND COREY PROC. N. A. S.

which are involved in ester linkages. This distortion of the phosphate group from the regular tetrahedral configuration is not supported by direct experimental evidence; unfortunately no precise structure determinations have been made of any phosphate di-esters. The distortion, which corresponds to a larger amount of double bond character for the inner oxygen atoms than for the oxygen atoms involved in the ester linkages, is a reasonable one, and the assumed distances are those indicated by the observed values for somewhat similar substances, especially the ring compound S_3O_9, in which each sulfur atom is surrounded by a tetrahedron of four oxygen atoms, two of which are shared with adjacent tetrahedra, and two unshared. The O—O distances within the phosphate tetrahedron are 2.32 Å (between the two inner oxygen atoms), 2.46 Å, 2.55 Å, and 2.60 Å. The

FIGURE 8

Plan of the nucleic acid structure, showing several nucleotide residues.





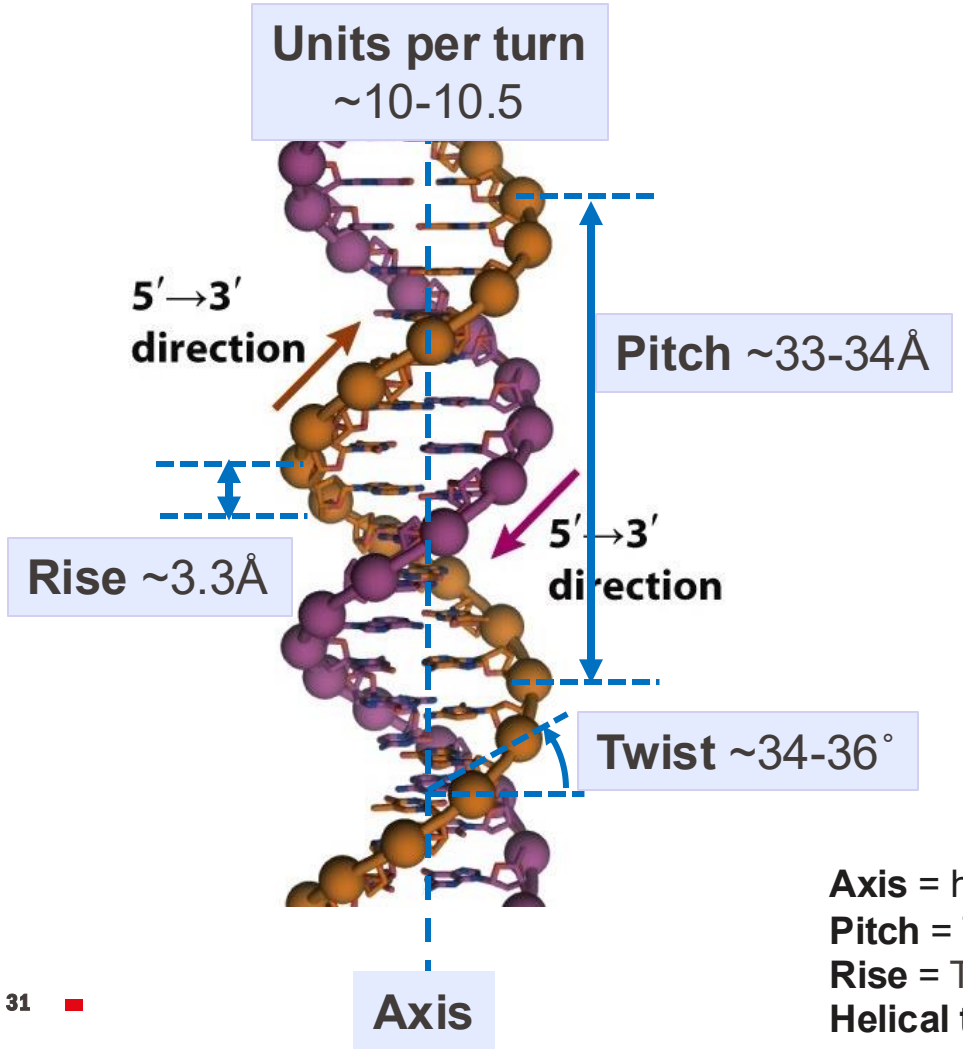
• The Pauling model proposed that all phosphate groups are in the center, while the bases face outwards

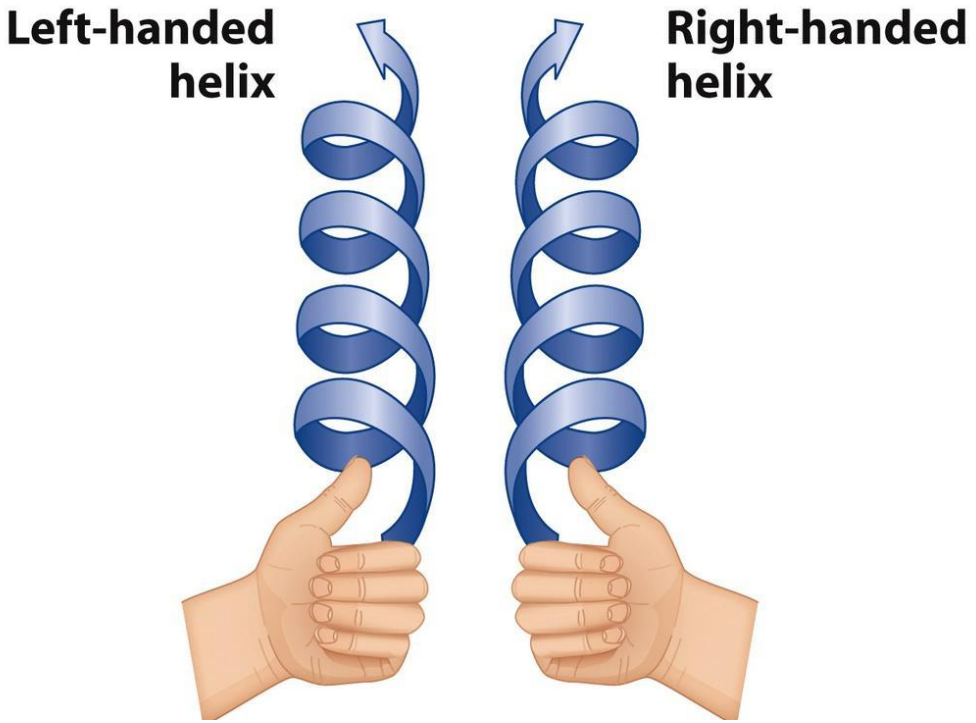
• But there was no proposed solution for resolving the negative charge build-up in the middle

• By his own admission, this proposed structure was his “greatest scientific misstep"

The structure of DNA - Helix geometry

- DNA helix is a repeating unit defined by a **helical axis**, **pitch**, **twist**, and **rise**
- Two strands together wind up to form a **right-handed** double-helix





Axis = helical axis is an imaginary line around which the two strands wind.
Pitch = The distance along the helical axis required for one complete turn of the DNA double helix.
Rise = The distance the helix extends along its axis when you go from one base pair to the next.
Helical twist = The angle of rotation between successive base pairs around the helix axis.

Different DNA assembly forms

- DNA can assemble into 3 major forms: **A, B** and **Z**
- Their geometries are very different (e.g., **Z-helix is left-handed**)

Rosalind Franklin

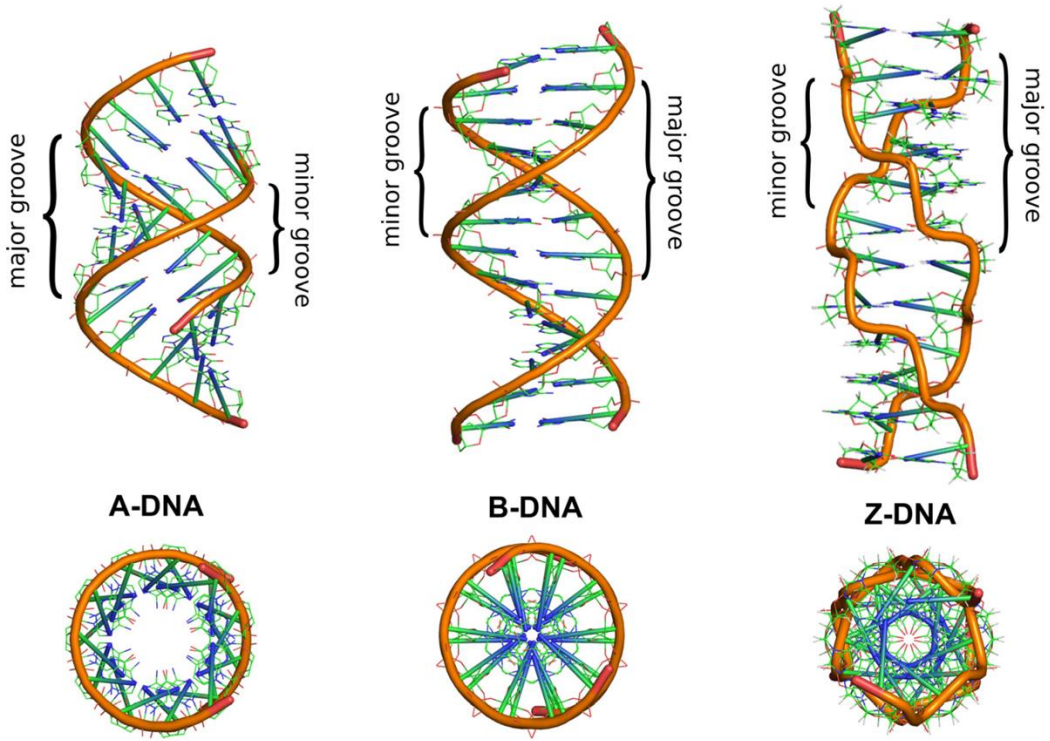
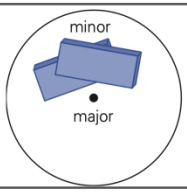
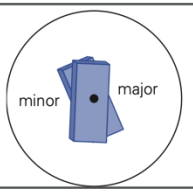
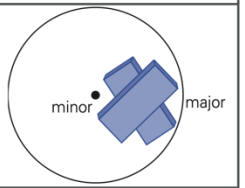


Table 2.1 Structural features of A-, B-, and Z-form helices.

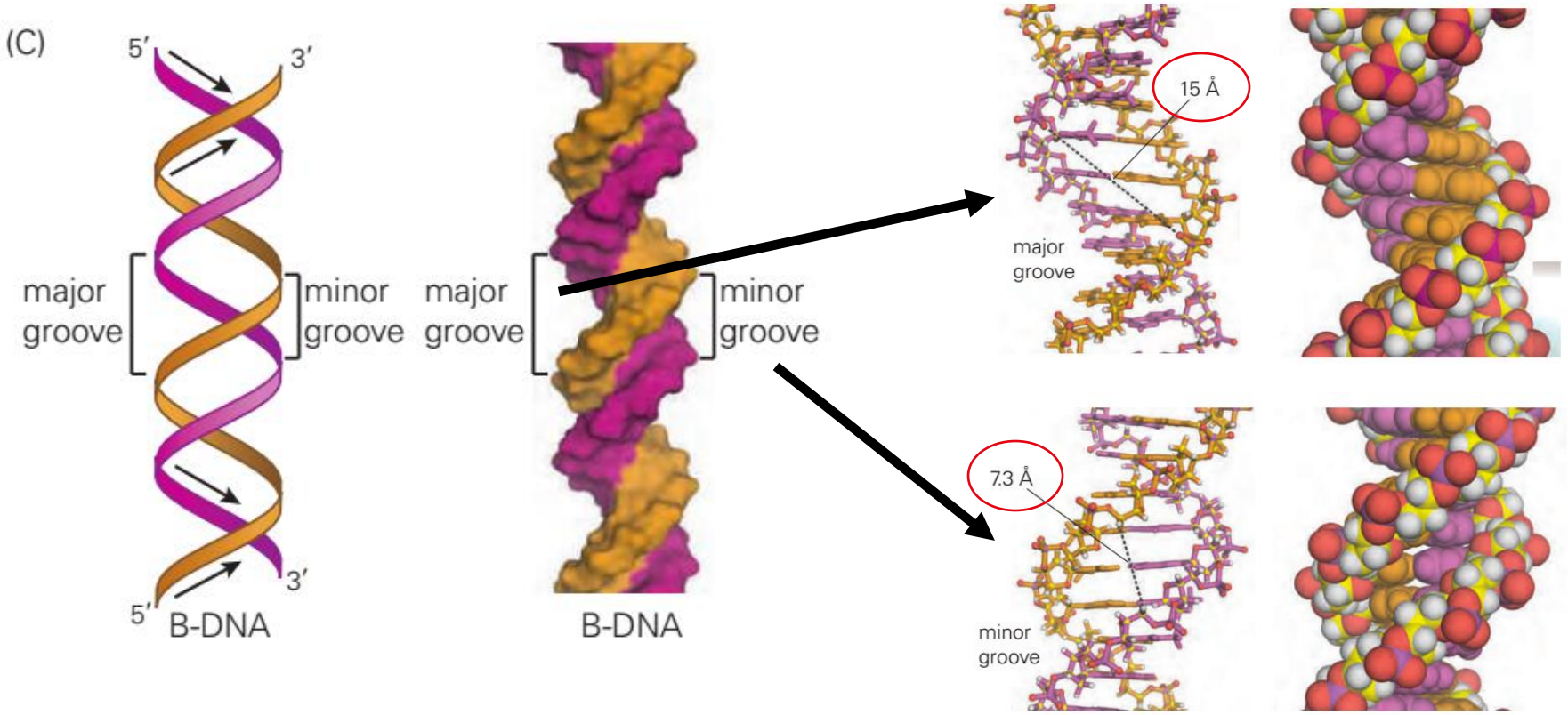
Helical form	A	B	Z
Helical sense	Right	Right	Left
Diameter	~ 26 Å	~ 20 Å	~ 18 Å
Base pairs per turn	~ 11	~ 10	~ 12
Helical twist (rotation per base pair for A and B, per two-base repeat for Z)	~ 34°	~ 36°	~ 60° (CpGp)
Helix pitch (rise per helical turn)	~ 25 Å	~ 33 Å	~ 46 Å
Helix rise (along helix axis; per base pair for A and B, per two-base repeat for Z)	~ 2.3 Å	~ 3.3 Å	~ 7.4 Å (CpGp)
Base tilt (with respect to helix axis)	~ 20°	~ 0°	~ – 9°
Base orientation (with respect to sugar)	Anti	Anti	C anti/G syn
Base pair positions (helix axis indicated by black dot)			
Features of base pair positions	Base pairs displaced from axis; deep major groove, less accessible	Base pairs on axis; both major and minor grooves accessible	Base pairs stick out into the major groove, the minor groove is deep and narrow

(Adapted from R.E. Dickerson et al., and M.L. Kopka, *Science* 216: 475–482, 1982. With permission from AAAS.)

- **The major form in solution is B**, while A and Z are less prevalent (and require special conditions)
- Relative locations of C2' and C3' carbons in deoxyribose are responsible for A and B forms

Major and minor grooves on DNA

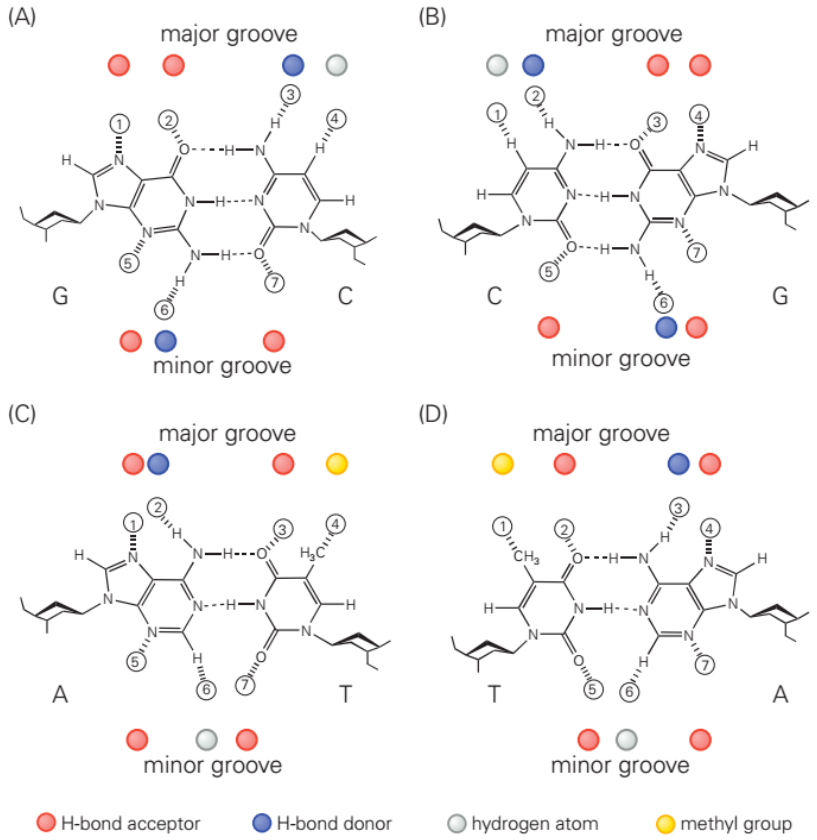
- In addition to sequence, one of the most important features of DNA double helices are the **grooves**.

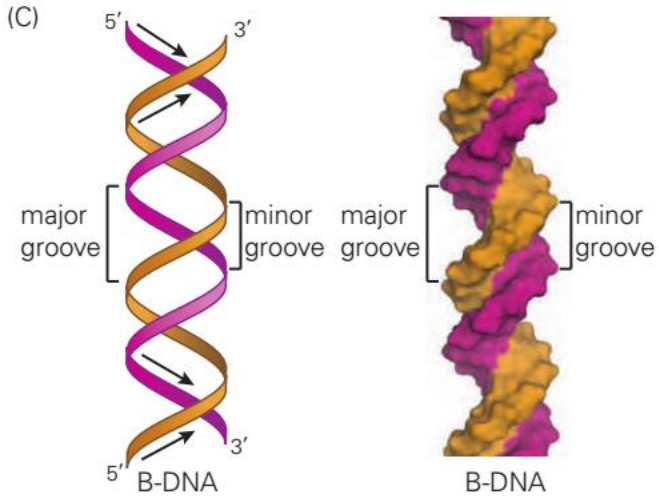


- The **major** and **minor** groove present very distinct structural features
- Very important for the **recognition of DNA by proteins**.

Major and minor grooves on DNA

- Potential **secondary interaction sites** at the edges of Watson-crick base pairs

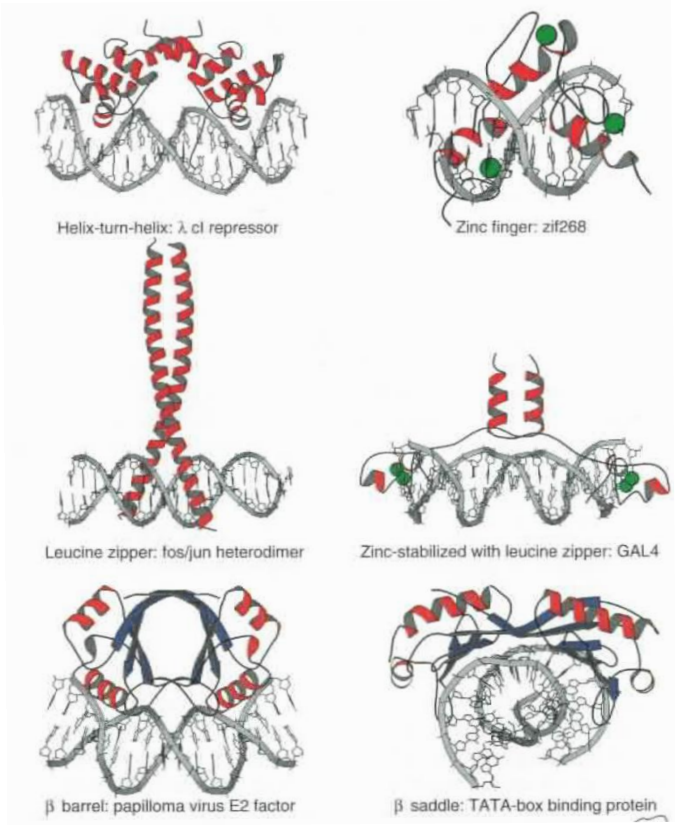




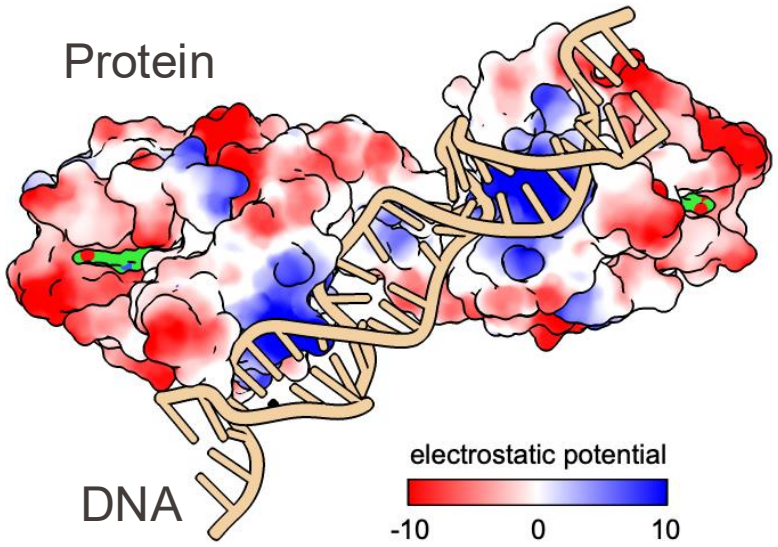
- The major and minor grooves can be identified by looking at the connections of the base pairs with the sugars.
- Major groove on the **convex edge** and the **minor groove** in on the concave edge.

DNA-binding proteins

- DNA-binding proteins primarily target the major groove (Mgr) or a combination of major and minor (Mm).
- Understanding how networks of DNA binding proteins impact cell processes is an essential component of immunology, developmental, cancer and systems biology.



Complex	Binding Motif[a]	Binding Groove[a]	Details of Complex
λ repressor	HTH	Mgr	Canonical HTH; homodimers; 2 helices of Cro dimer cradle Mgr, stabilized by direct H-bond and vdW contacts; little DNA distortion.
CAP repressor	HTH	Mgr	About 90° bend.
trp repressor	HTH	Mgr	Indirect, water-mediated base contacts.
Purine rep.	HTH	Mm	α-helices inserted in mgr.
Yeast MATα2	HTH	Mgr	Homeobox domains bind as monomers.
Zif268	Zn	Mgr	Zinc finger subfamily; each Zn finger recognizes 3 bps.
GATA-1	Zn	Mm	Transcription factors subfamily; single domain coordinated by 4 cysteines.
GAL4	Zn	Mgr	Metal binding subfamily; each of two Zn ions, coordinated by 6 cysteines, recognizes 3 bps.
GCN4	Leu/Zip	Mgr	Canonical; basic region/leucine zipper (α-helices) motif; slight DNA bending.
fos/jun	Leu/Zip	Mgr	α-helices resemble GCN4; unstructured basic region folds upon DNA binding.
fos/jun/NFAT	Leu/Zip	Mgr	α-helices bend to interact with NFAT.
MetJ	β-ribbon	Mgr	Two anti-parallel β-strands in Mgr; bends each DNA end by 25°.
papillomavirus E2 DNA target	β-barrel	Mgr	Domed β-sheets form an 8-strand β-barrel dimer interface with 2 α-helices in Mgr; strong tailored fit for every base of the recognition element; bent DNA; compressed mgr; DNA target crystallized without protein.
TBP	β-saddle	mgr	Ten-β-strand saddle binds in Mgr; significant distortion, ≈ 90° bend.
p53 tumor supp.	Loop/other	Mm	Binds to DNA via protruding loop and helix anchored to anti-parallel β-barrel.
SRY	Loop/other	mgr	Isoleucine intercalated into mgr.
NFAT	Loop/other	Mm	Flexible binding loop stabilized by DNA.
histones	Loop/other	Mm	Nonspecific PO_4 interactions.
distamycin (drug)		mgr	Selective to AT bps; binds in mgr without distortion.

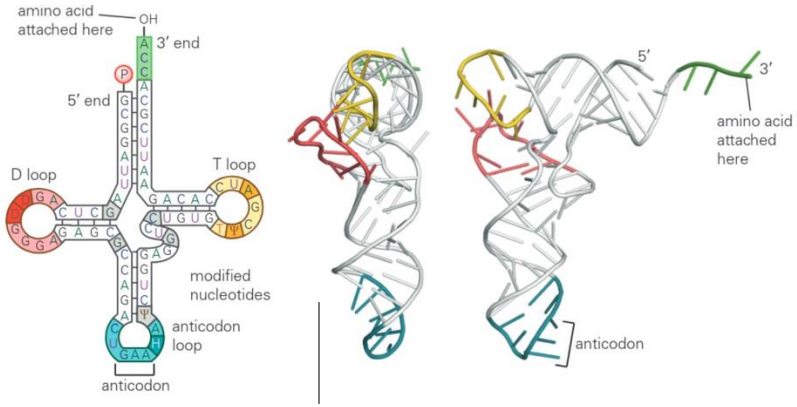


DNA-binding proteins typically engage the DNA backbone using positively charged amino acids

Differences in assembly of DNA and RNA

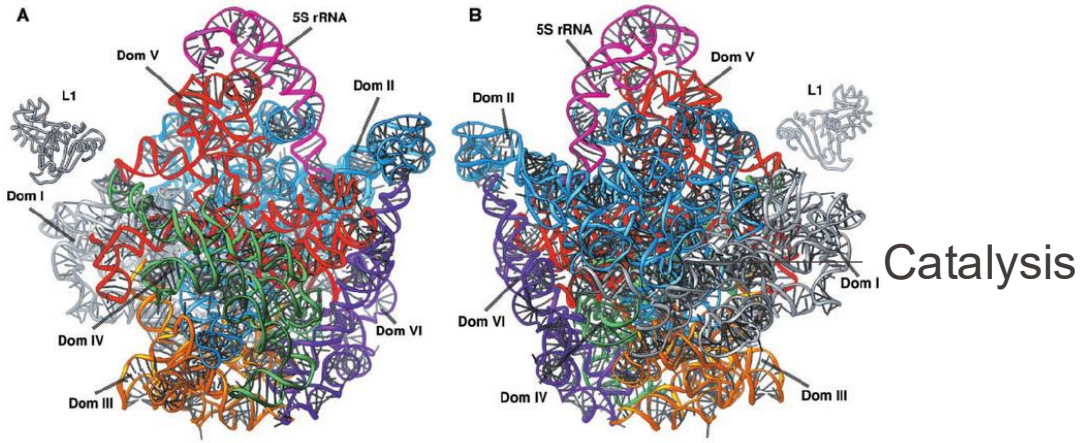
• Due to the unique chemical restraints of the (i) ribose sugar and the (ii) single-stranded nature in most cases, RNA can take on many diverse forms and, consequently, has more biological functions compared to DNA

transfer (t) RNA



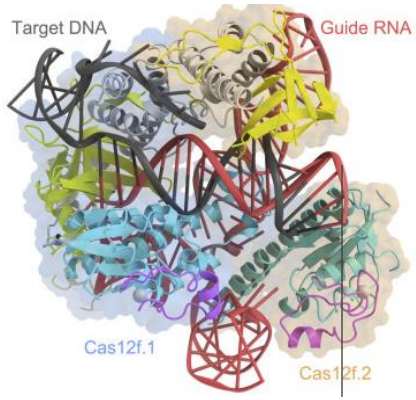
Amino-acid carrier

Ribosomal RNA



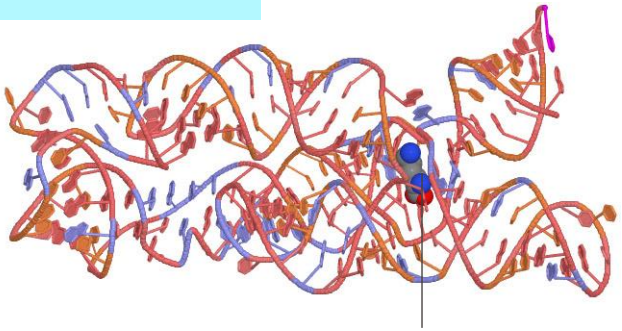
Catalysis

Guide RNA



Defending

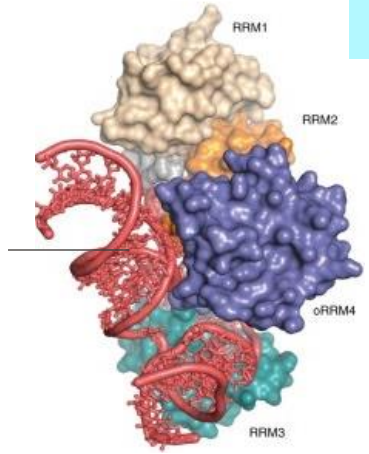
Riboswitch



Ligand sensing

snRNA

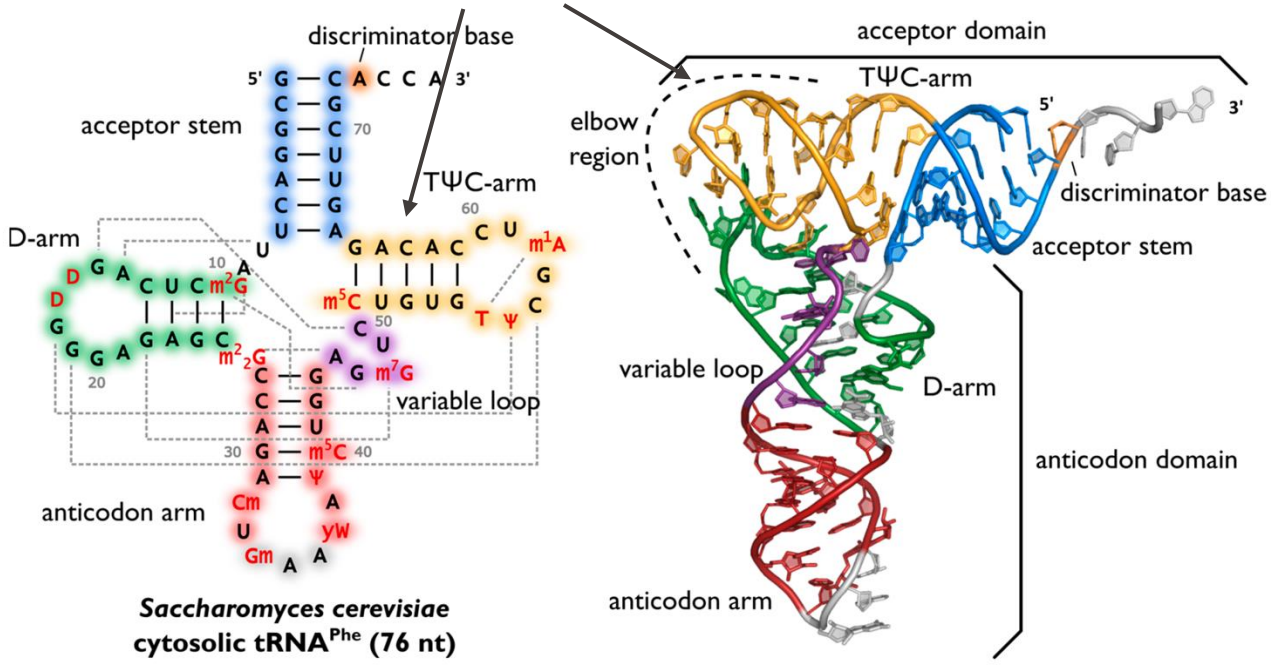
Scaffolding
(and catalysis)



RNA can also form partial double-helices

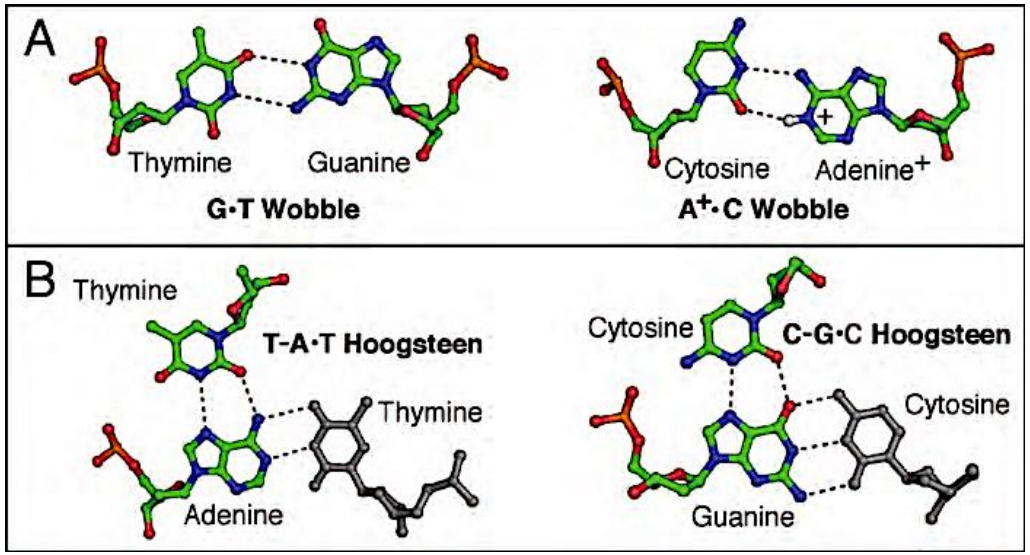
- Partial complementary regions in **RNA** drive the assembly of **shorter double helices** which ultimately dictate the 3D assembly of the entire RNA molecule and its function.
- The conformation is further stabilized by **non-canonical base pairing** interactions. These can also occur in DNA but are more common in RNA.

Complementary regions assemble into short double-stranded helices



Saccharomyces cerevisiae cytosolic $tRNA^{Phe}$ (76 nt)

Non-Watson-Crick base pairing improves stability of non-complementary regions



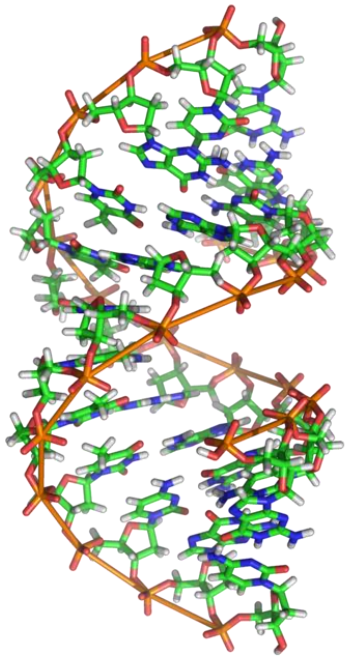
A few examples, but many more exist...

No need to remember these for exam. You just need to be aware of their existence.

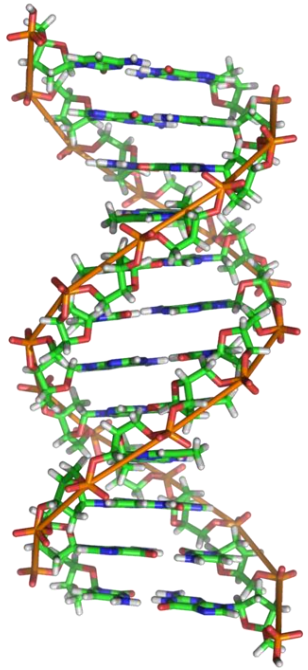
Differences in assembly of dsDNA and dsRNA

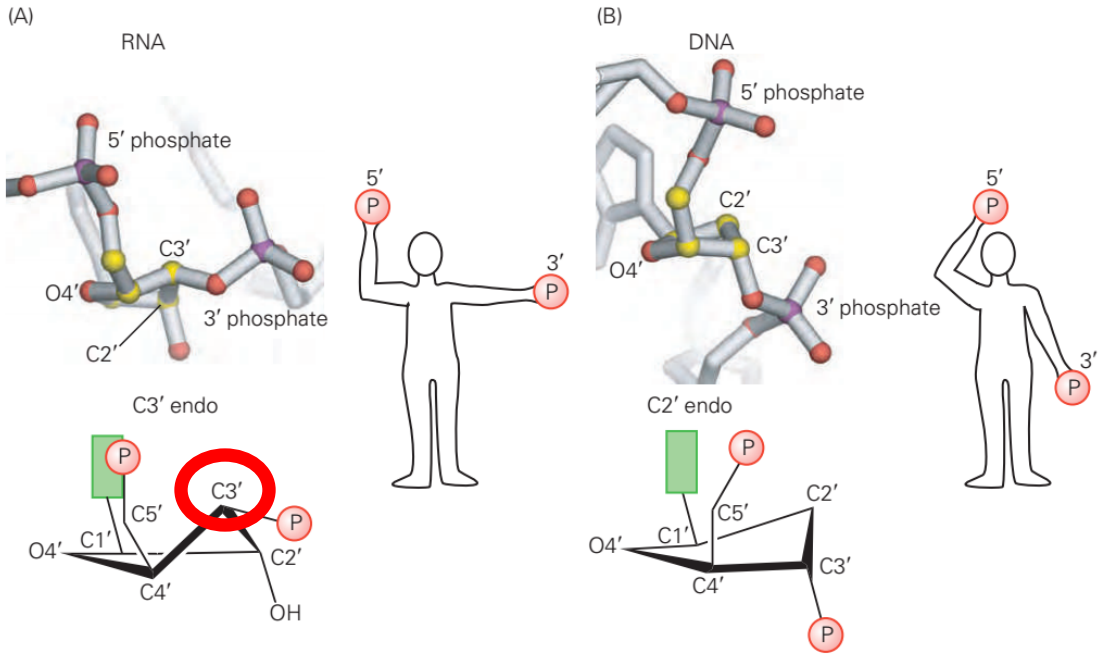
• Double-stranded helices of RNA can be formed locally in molecules like tRNA, but they can also be found in some viruses (e.g., rotavirus) where dsRNA represents the viral genome

dsRNA



dsDNA (B)

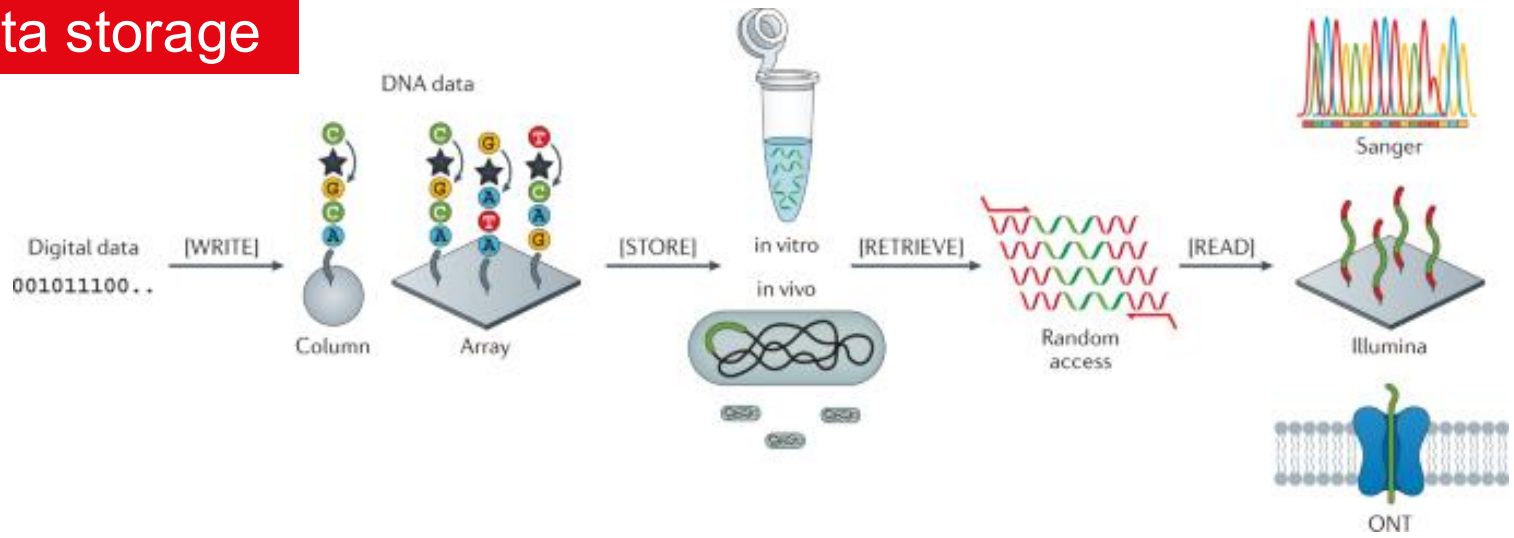




• For RNA the hydroxyl group influences the preferred conformational state of the sugar (**C3'-up state**) which impacts how it packs into a helix. In that sense it is similar to the **A-form** of DNA

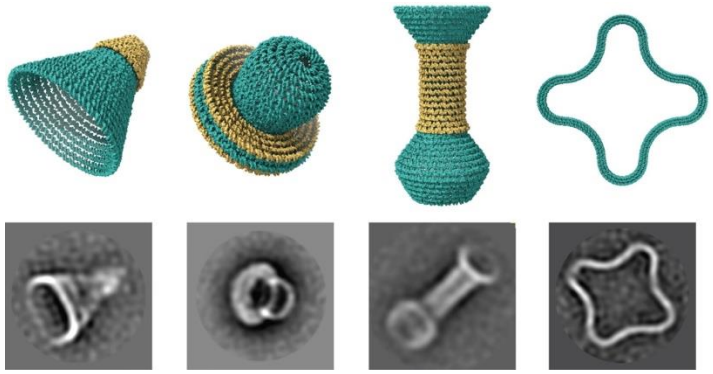
DNA in biotechnology

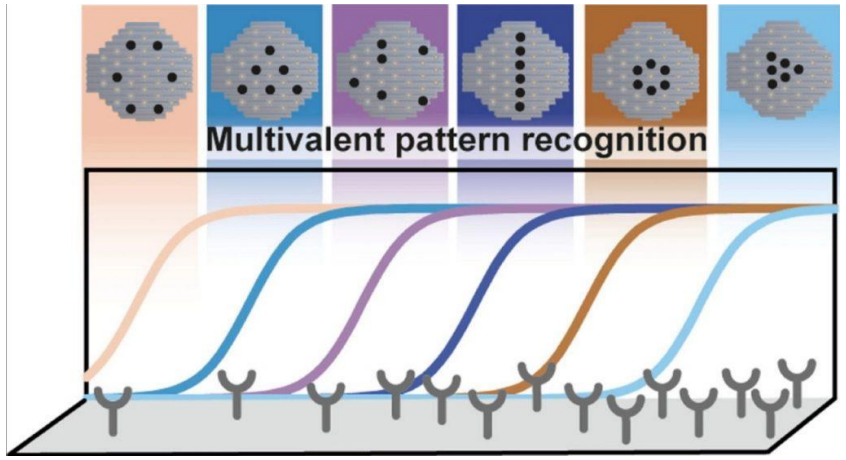
Data storage



- Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA. *Nat Rev Genet* **20**, 456–466 (2019). https://doi.org/10.1038/s41576-019-0125-3

DNA origami



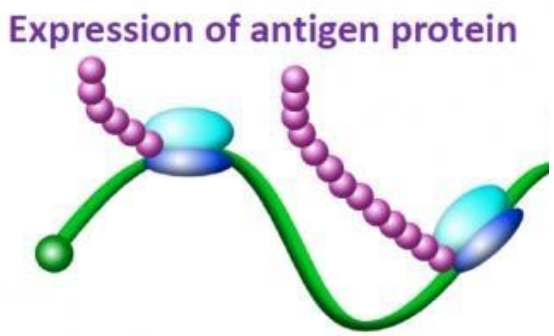


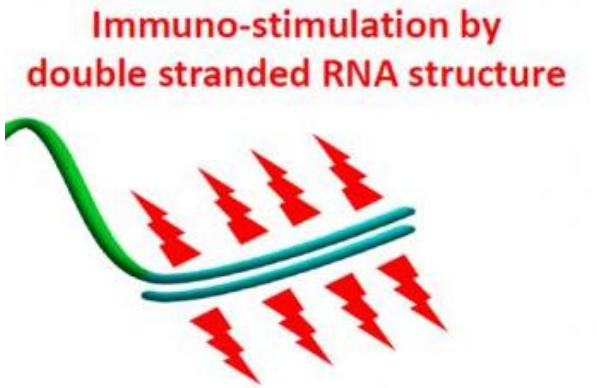
- Hale Bila, Kaltrina Paloja, Vincenzo Caroprese, Artem Kononenko, Maartje M.C. Bastings, *American Chemical Society, (*2022)
- https://pubs.acs.org/doi/10.1021/jacs.2c08529#

https://www.epfl.ch/labs/pbl/

RNA in biotechnology

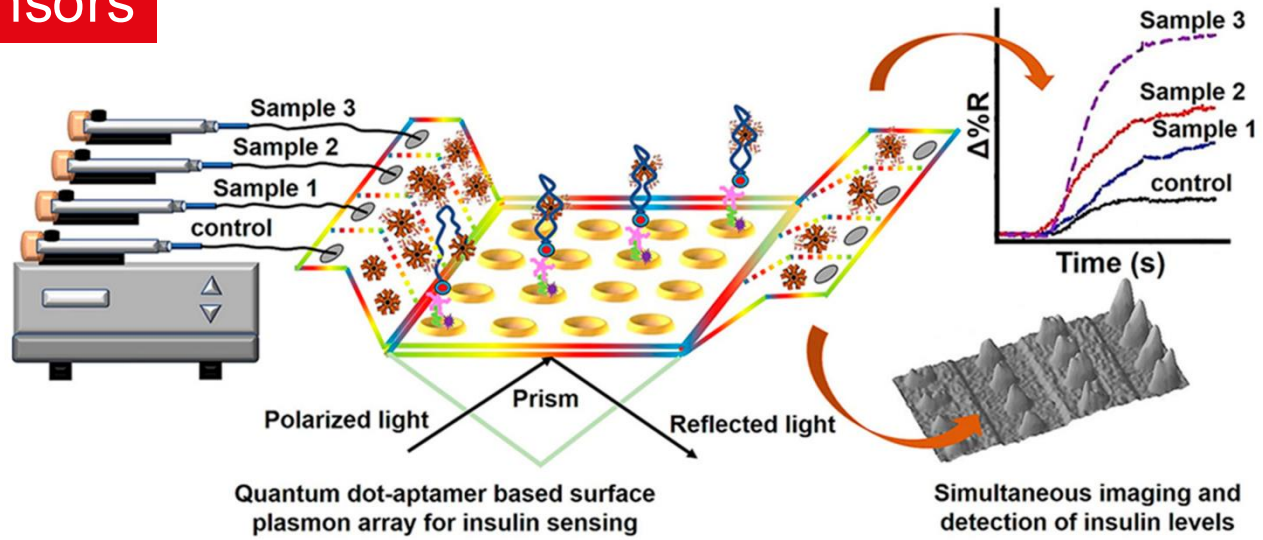
Vaccines and adjuvants





- Abhijeet Girish Lokras, Thomas Rønnemoes Bobak, Saahil Sandeep Baghel, Federica Sebastiani, Camilla F oged, Advanced Drug Delivery Reviews, Vol 213, 2024.
- https://www.sciencedirect.com/science/article/pii/S0169409X24002412

Biosensors



- Nako Nakatsuka, Kelly J. Heard, Alix Faillétaz, Dmitry Momotenko, János Vörös, Fred H. Gage & Krishna C. Vadodaria,
- Molecular Psychiatry volume 26, pages2753–2763, 2021.
- https://www.nature.com/articles/s41380-021-01066-5

https://www.epfl.ch/labs/chemina/

Nucleic acids – Take Home messages

- **DNA and RNA are the informational polymers** in the cell – encode genetic information in a way that can be read by macromolecular machines, to direct the synthesis of other molecules.
- Nucleotides have **pentose sugars** attached to **nitrogenous bases** and **phosphate** groups.
- The nucleotide bases in DNA and RNA are substituted **pyrimidines** or **purines**.
- **4 deoxyribonucleotides** in DNA (A,T,G,C) and **4 ribonucleotides** in RNA (A,U,C,G)
- **DNA and RNA are synthesized in 5' to 3' direction** by sequential reactions that are driven by hydrolysis of nucleotide triphosphates
- DNA forms a **double helix** with antiparallel strands (**direct** and **reverse complement** strands)
- Double helix involves complementary base pairing (A-T and C-G) and is stabilized by, **hydrogen bonds, base pair stacking and electrostatic interactions**
- B-form DNA allows sequence specific **recognition of the major groove by proteins**. Each base pair has a unique set of interacting elements in the major groove but not in the minor groove.
- RNA is typically single-stranded, but it can also assemble into shorter or longer helical structures.

Next week's lecture

Introduction to Lipids

Milena Schuhmacher
ELISIR Scholar
Group Leader at EPFL