

Nucleic Acid Structure and Base Pairing

A



Objective

To understand how the structure of nucleic acids and base pairing relate to biological function



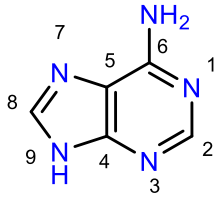
Blackburn, G. M. *et al.* Nucleic Acids in Chemistry and Biology. (The Royal Society of Chemistry, 2022).

Frank-Kamenetskii, M. D. & Liapin, L. Unraveling DNA. (Wiley, 1993).

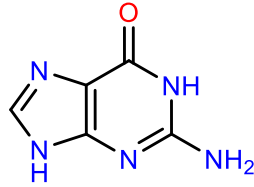
Alberts, B. *et al.* Molecular Biology of the Cell (Seventh Edition). (W. W. Norton, Incorporated, 2022).

Nucleobase Structure

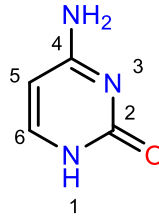
DNA Base



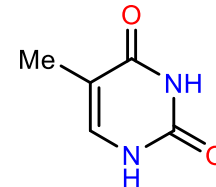
adenine (A)



guanine (G)

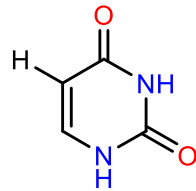


cytosine (C)

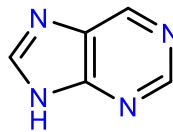


thymine (T)

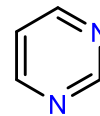
RNA Base



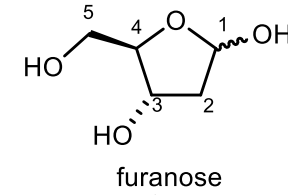
uracil (U)



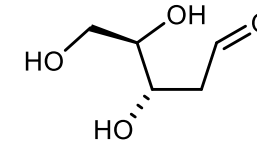
purine (R)



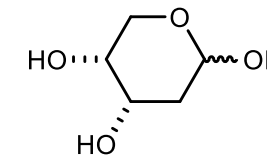
pyrimidine (Y)



furanose



acyclic

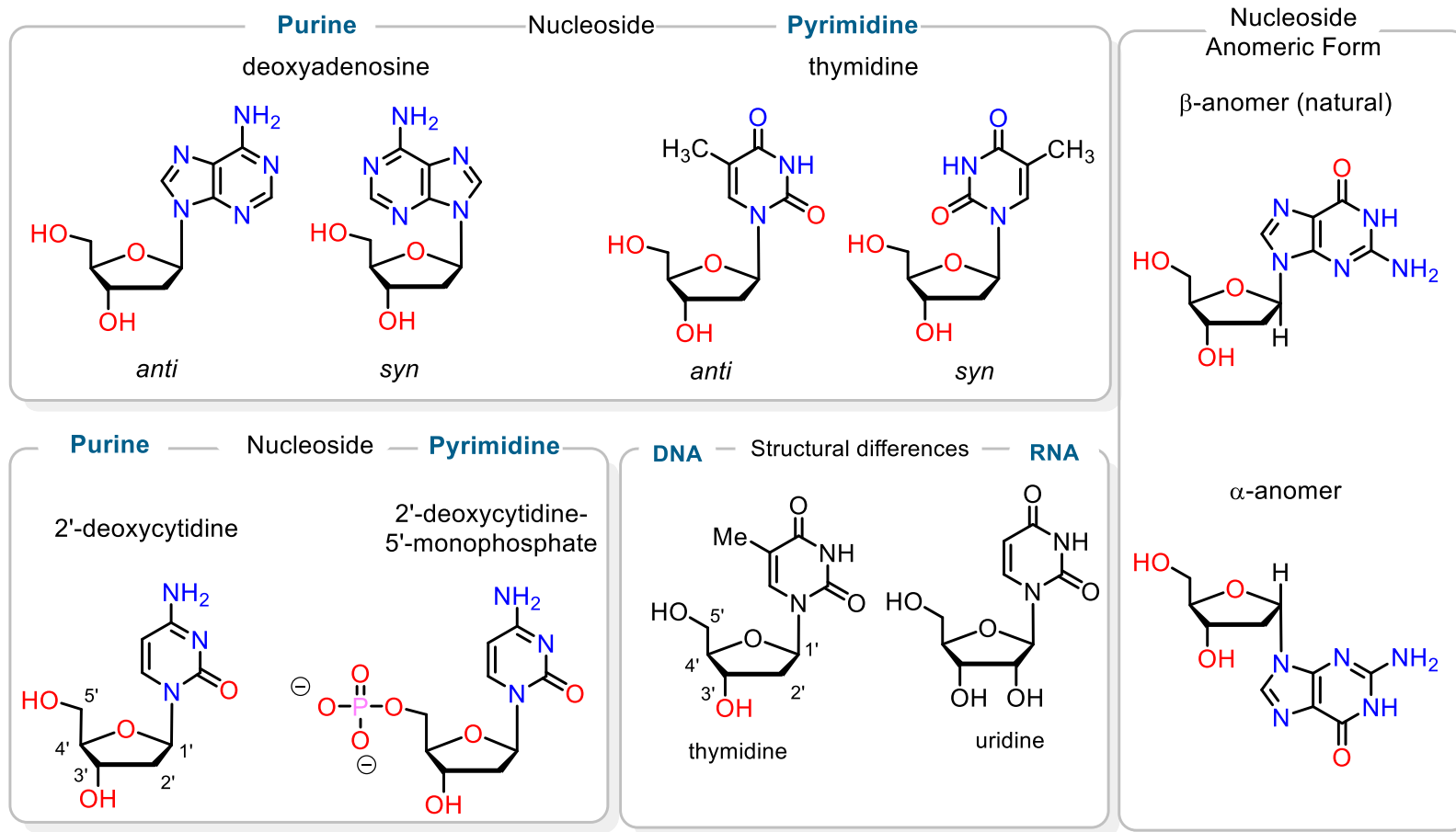


pyranose

The deoxyribose sugar exhibits dynamic mutarotation as a free monosaccharide, interconverting among furanose, acyclic, and pyranose forms; however, incorporation into DNA locks it into the furanoside form, providing the structural rigidity essential for nucleic acid architecture.

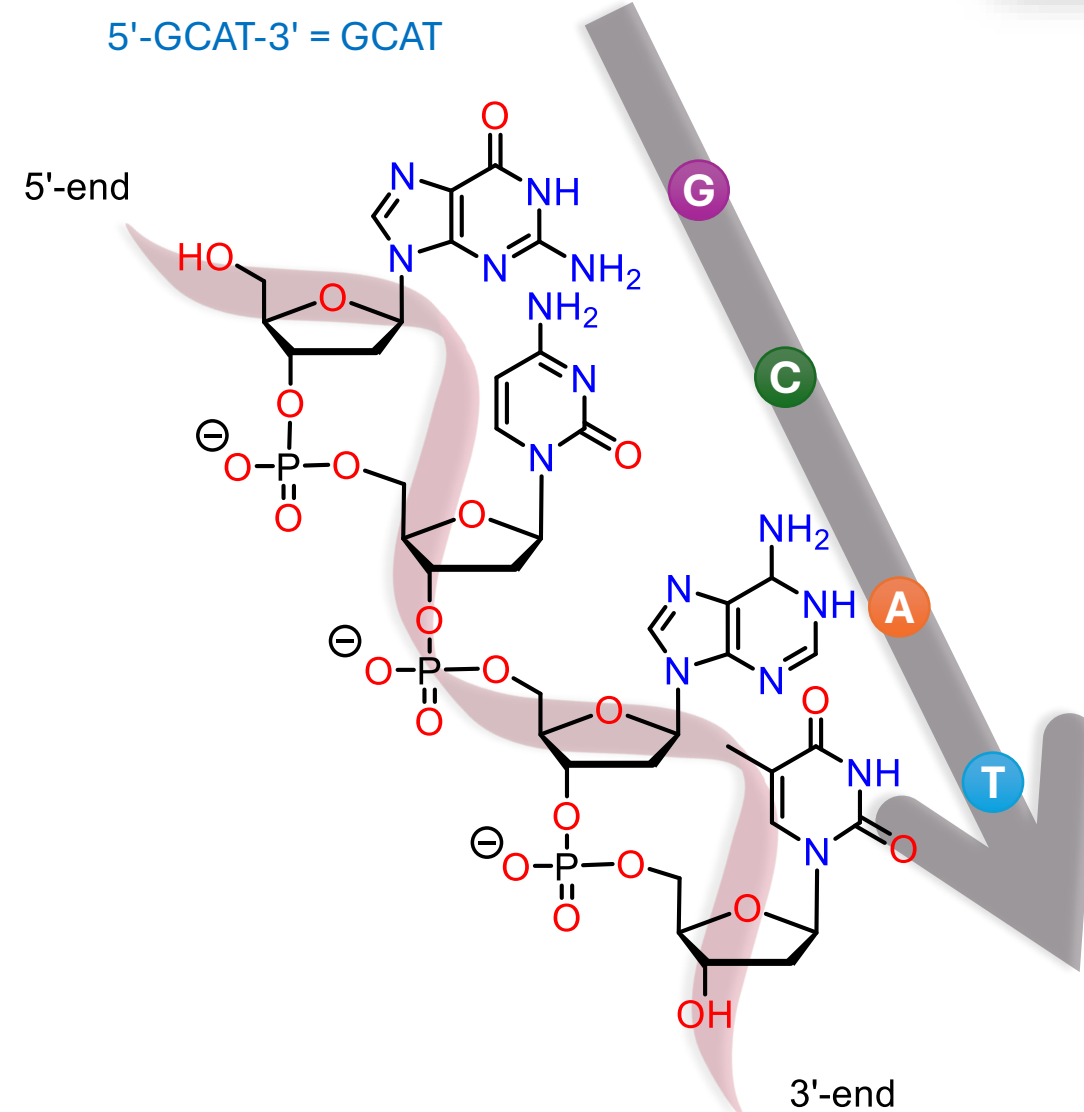
Nucleosides and Nucleotides

Structure



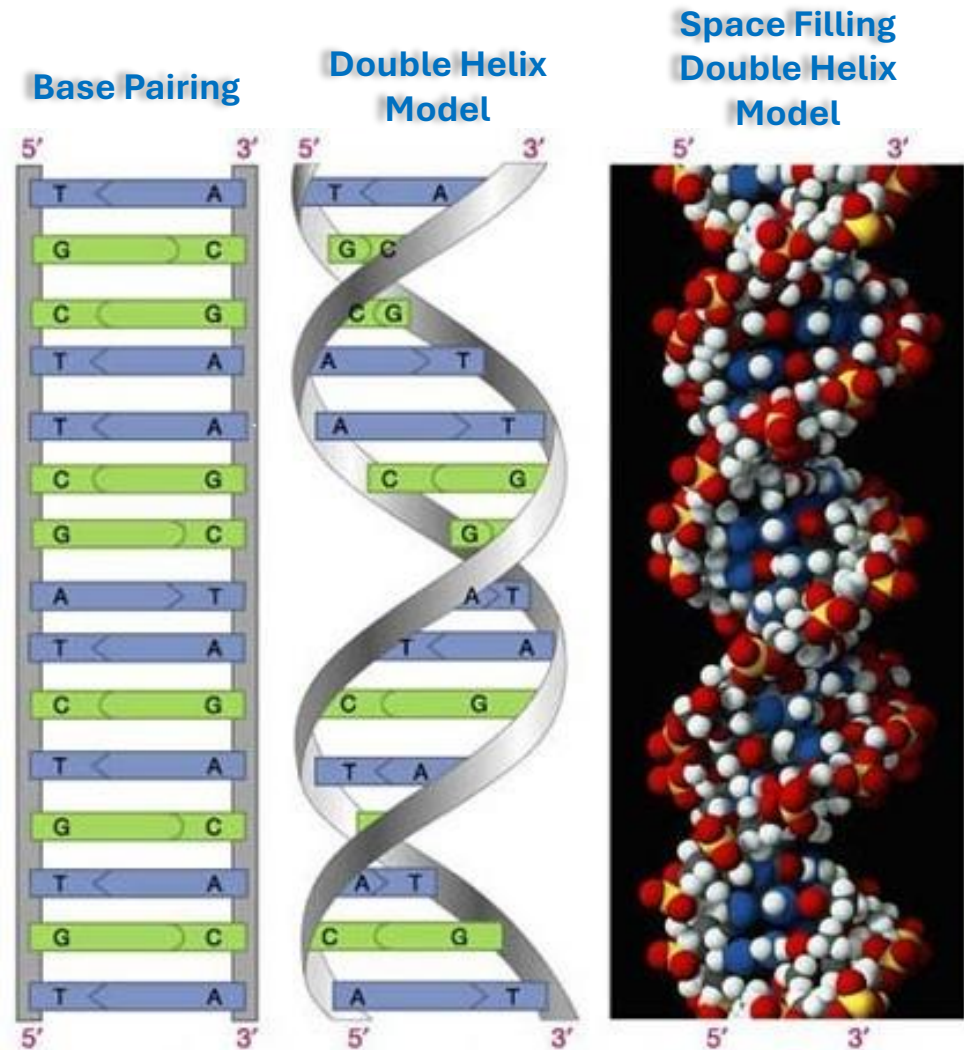
Sequence

- ❑ DNA is a salt of a phosphoric acid ester (a phosphodiester). It typically exists as a sodium or potassium salt and is soluble in aqueous buffers.
- ❑ DNA sequences are written by convention in the 5'- to 3'-direction



Sequence

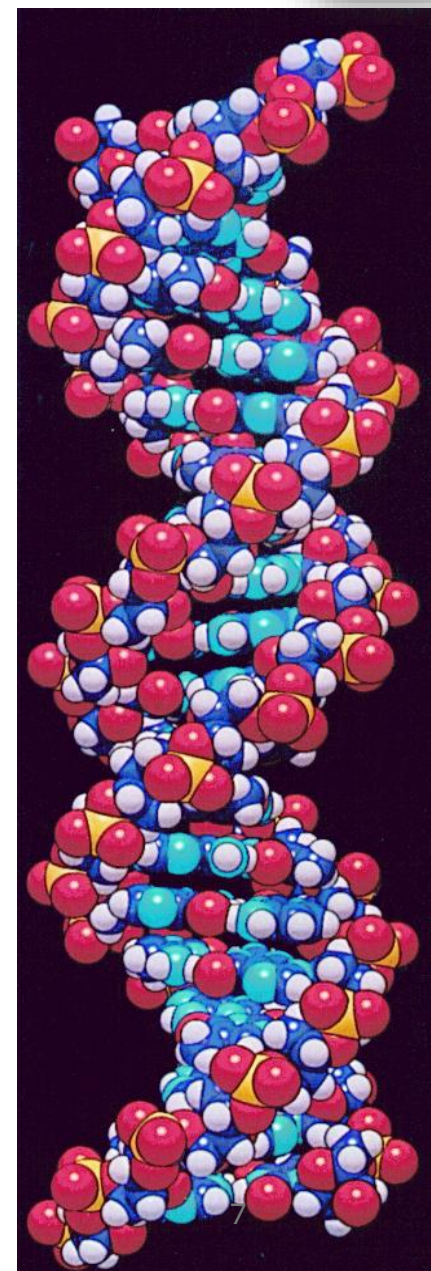
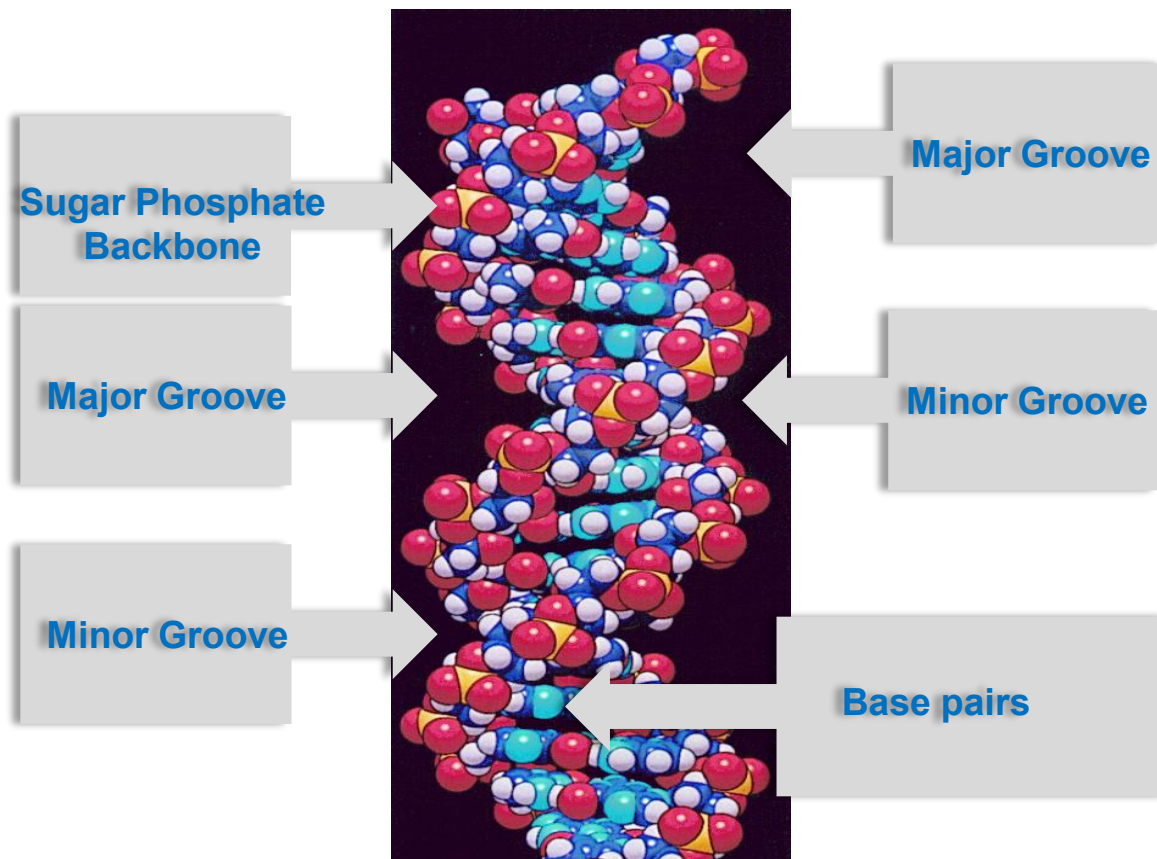
- ☐ B-DNA is right-handed
- ☐ The base pairs stack on each other ~10.5 bp per helical turn
- ☐ The hydrophobic base pairs are on the inside
- ☐ The hydrophilic sugar phosphate backbone is on the outside exposed to water



The Double helix - B-DNA

Right-handed helix

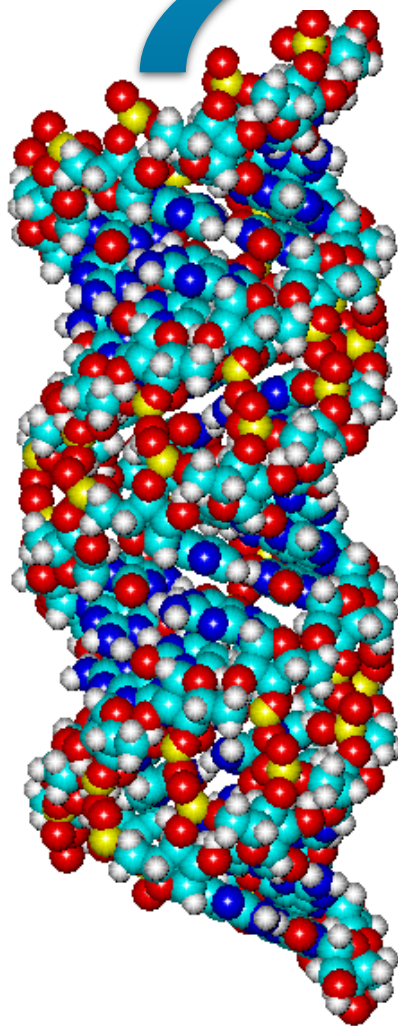
Each base pair is separated by 0.34 nm



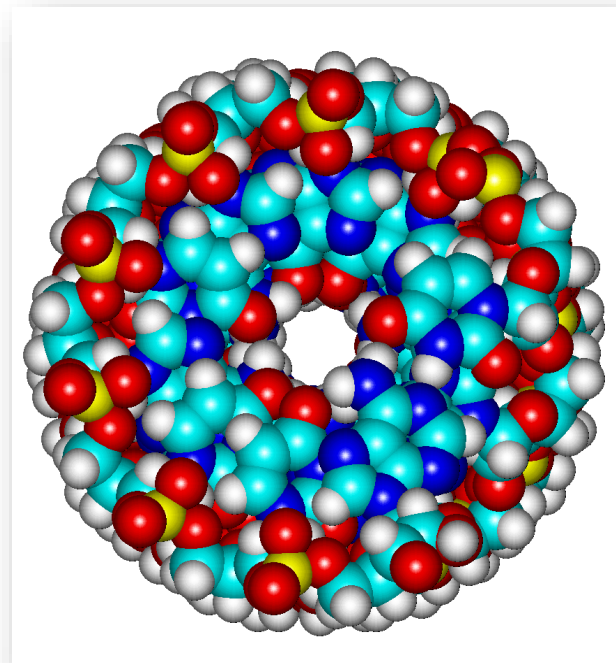
Nucleic Acids Structure

RNA Duplexes Exist in the A Conformation

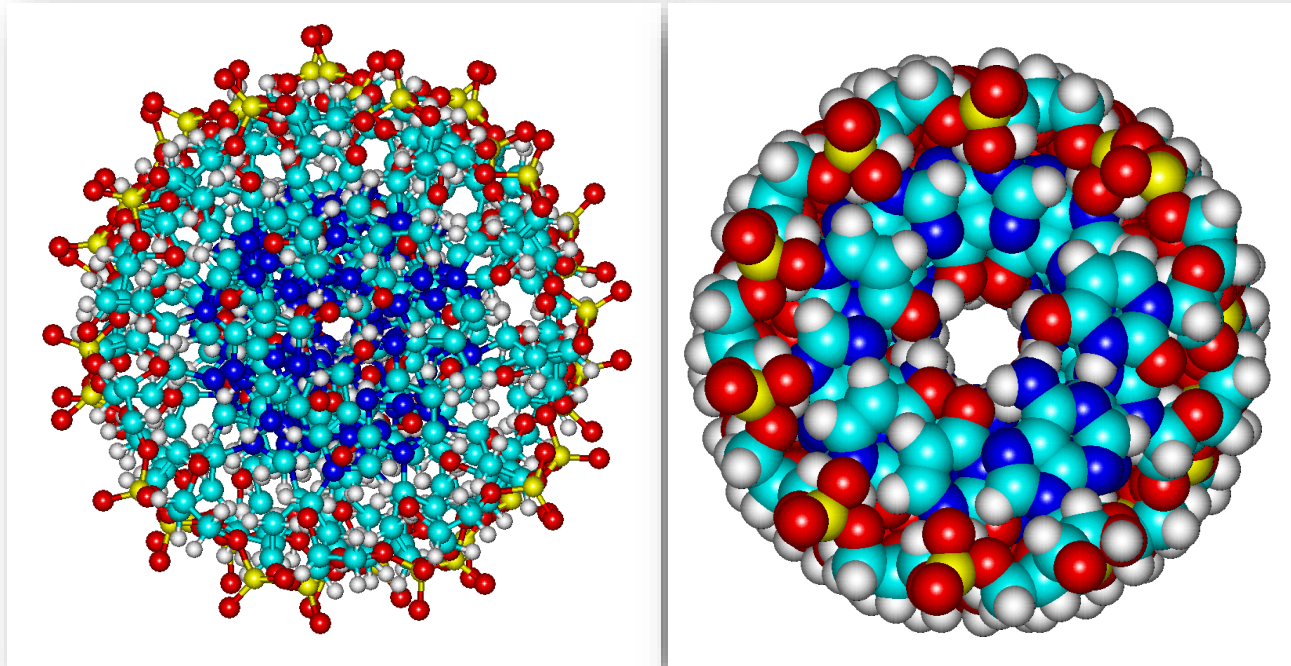
- ☐ The major groove is very deep
- ☐ The minor groove is very shallow
- ☐ The base pairs are pushed towards the outside of the helix



Right-handed antiparallel
double helix



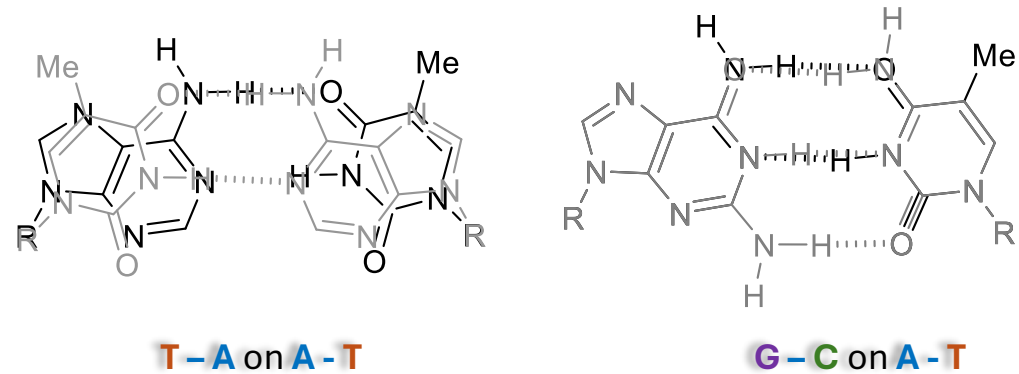
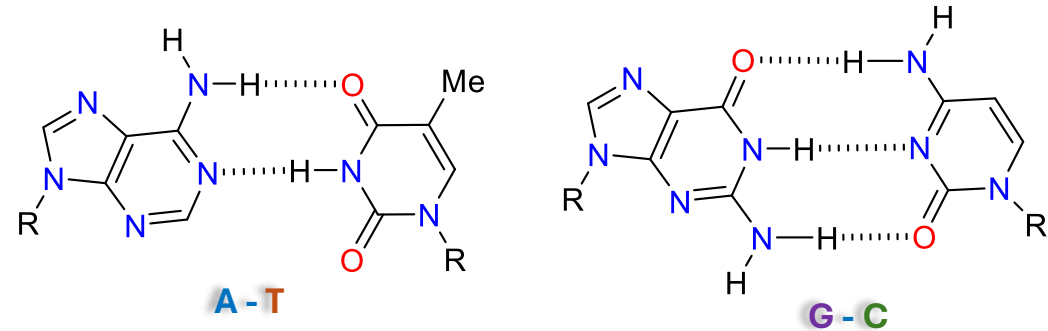
In the B form, the base pairs are closer to the centre of the helix compared to the A form



Nucleic Acids – Duplex Structure

A

Base pair

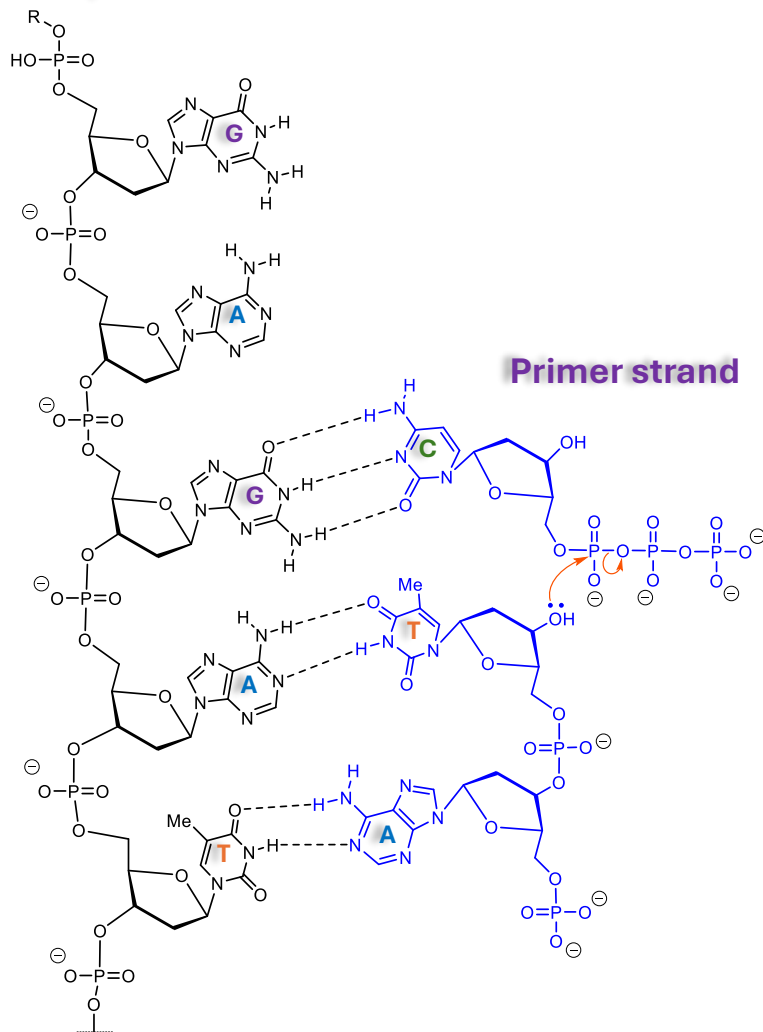


Watson-Crick base pairs are pseudo-symmetric
and produce a regular double helix

All other combinations of DNA bases have different shapes (later)

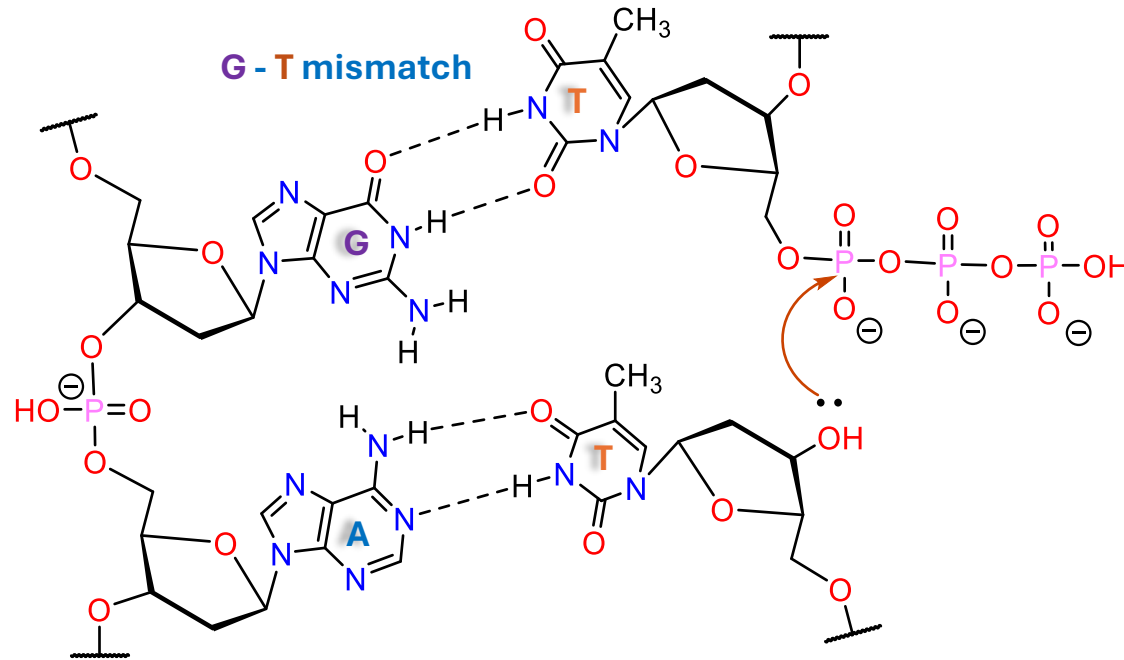
Replication

Template strand



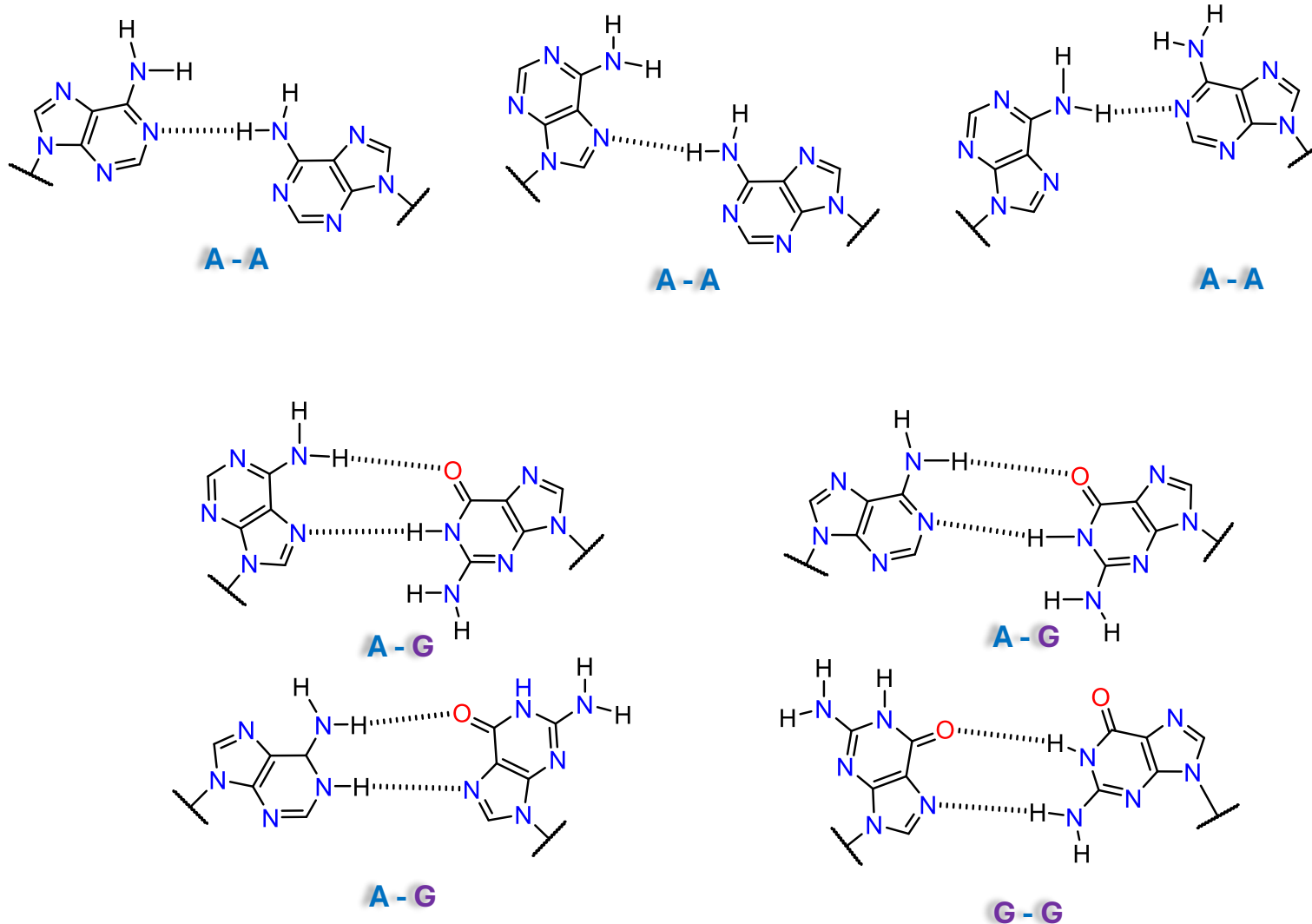
This process is controlled by Watson-Crick base pairing and by DNA polymerases. It does not work without DNA polymerase.

Transcription: DNA codes for RNA in a similar process catalysed by RNA polymerase enzymes



Mis-paired bases can also be incorporated at low level during the replication. G - T and G - A mismatches are fairly stable base pairs.

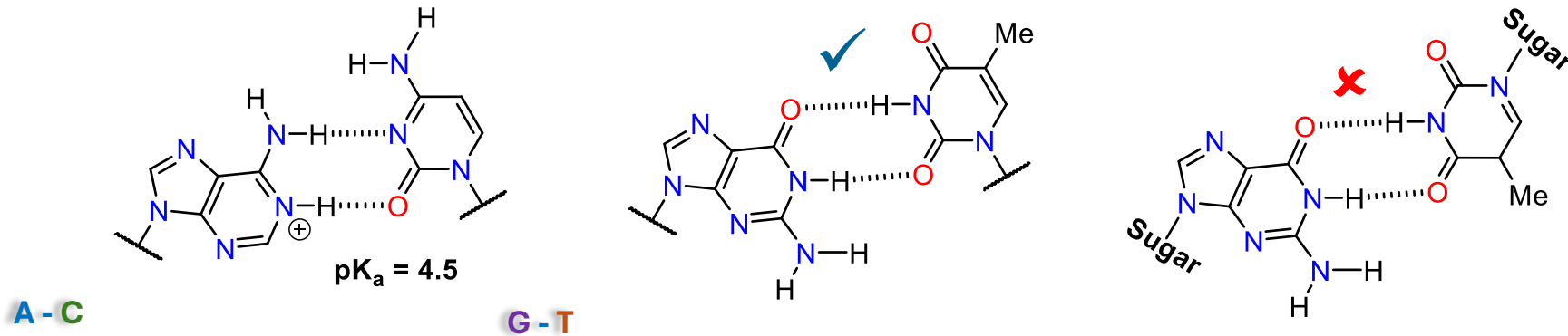
Purine:purine mismatched base pairs



They rarely occur during replication as they are misshaped and some are very unstable, so the integrity of the genome is not significantly compromised.

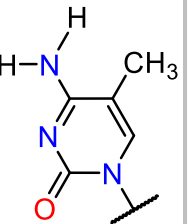
Purine:pyrimidine mismatched base pairs

A

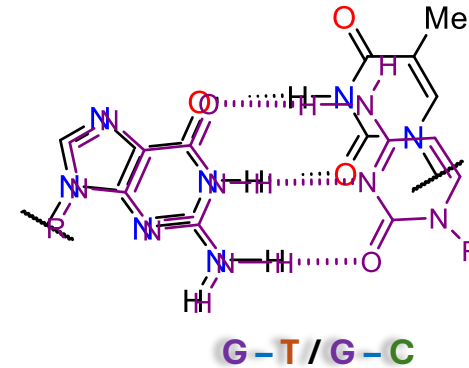
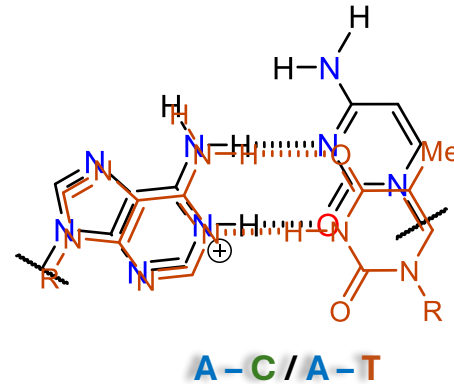


Sugars on opposite sides of helix, not possible in a DNA duplex.

- ❑ Mismatched bases (e.g. A-C or G-T) are recognised by DNA polymerases and repair enzymes. They do not fit into a regular smooth double helix.
- ❑ The base from the daughter (unmethylated) DNA strand is removed by DNA repair enzymes. The parent strand has methylation at some cytosine residues, allowing it to be distinguished from the daughter strand.

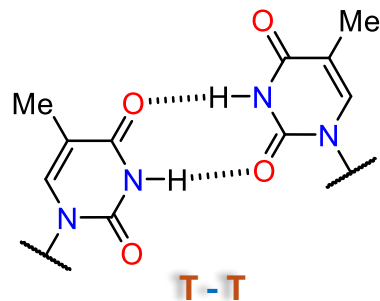
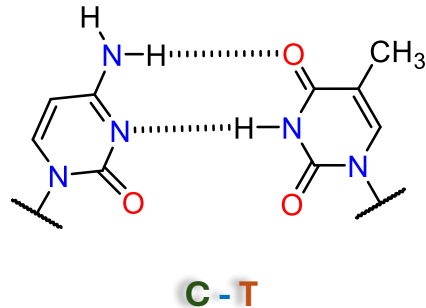
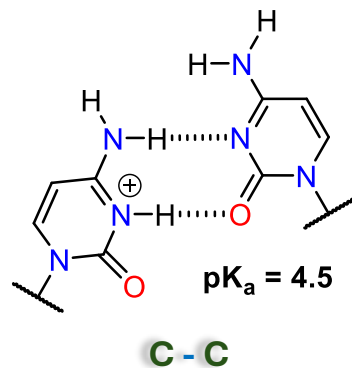


Purine:pyrimidine mismatched base pairs



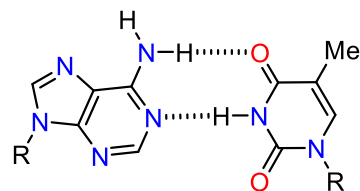
- ❑ Mismatched bases (e.g. A-C or G-T) are recognised by DNA polymerases and repair enzymes. They do not fit into a regular smooth double helix. Here we illustrate that the mismatched base pairs do not overlap well with Watson Crick base pair.

Pyrimidine – pyrimidine mismatched base pairs

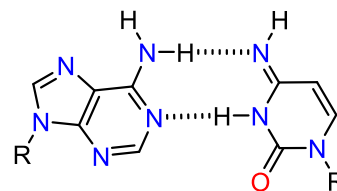


- ❑ Pyrimidine – pyrimidine base pairs are smaller than the Watson-Crick base pairs. They have poor base stacking and are therefore very unstable.

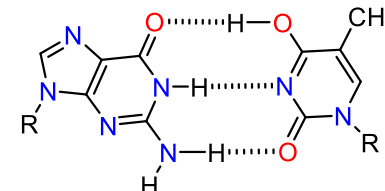
The tautomer hypothesis



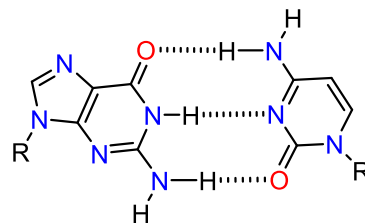
A-T



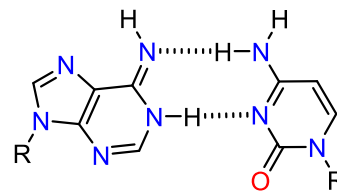
A-C*



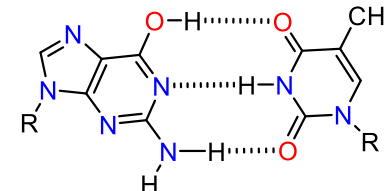
G-T*



G-C



A*-C

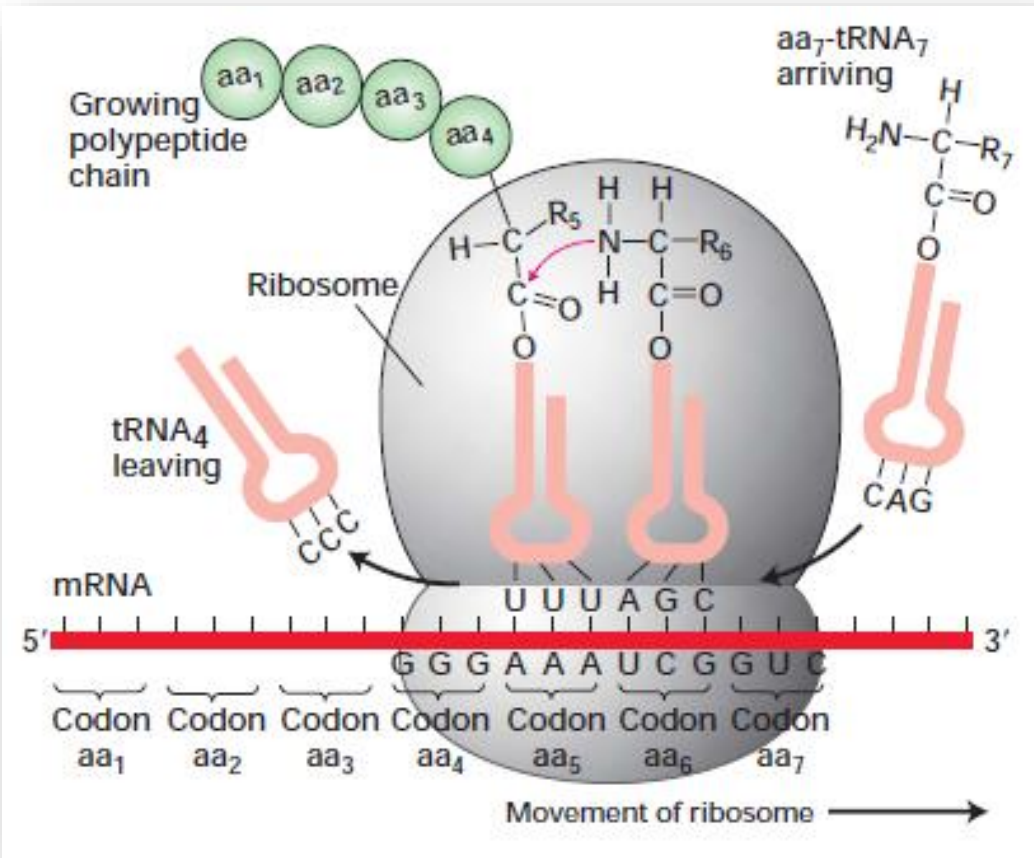


G*-T

* = minor 'imino or enol' - tautomer

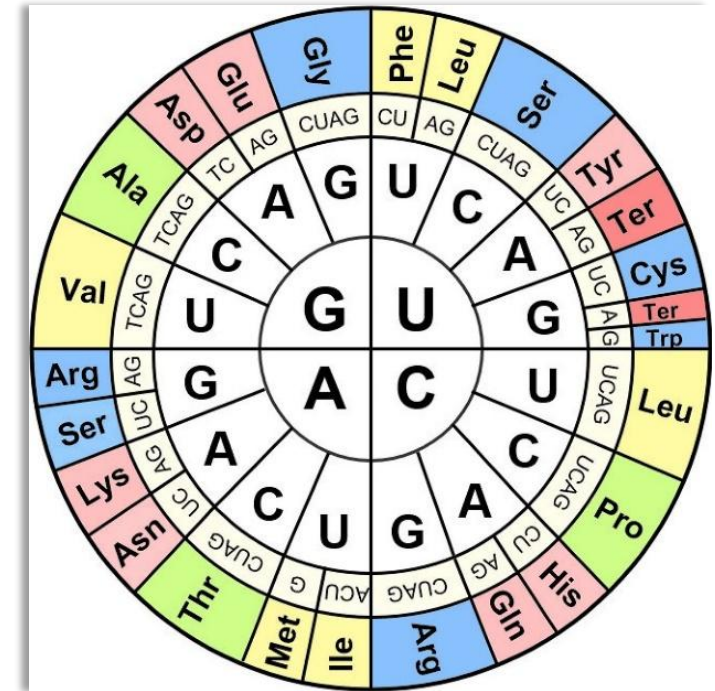
- ❑ All base pairs in this figure have the same shape. So, the A-C and G-T mismatches are excellent mimics of Watson-Crick base pairs.
- ❑ Around 1 in 10^4 molecules of the canonical bases exists as minor tautomers at any time
- ❑ The probability of a nucleobase being a minor tautomer during replication and also during proofreading is $10^{-4} \times 10^{-4} = 10^{-8}$
- ❑ This is the error rate during copying of the human genome
- ❑ Evolution has 'chosen' DNA bases that have one tautomer that predominates

Translation at the ribosome – 4 letter code



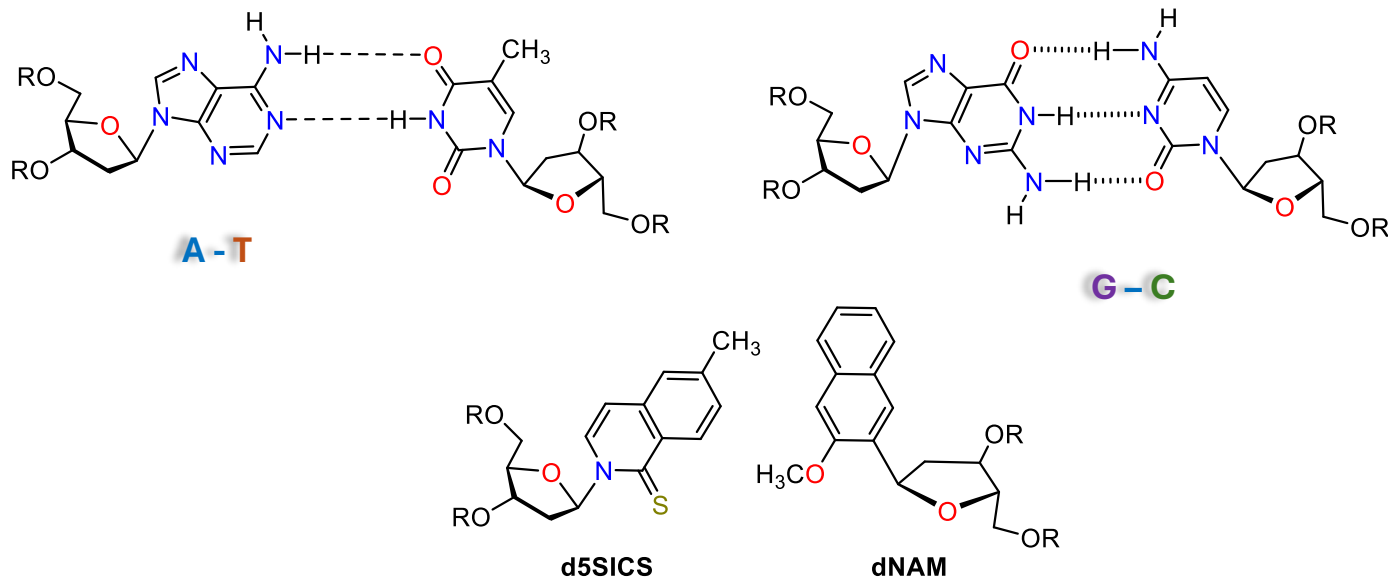
Each 3-letter codon originating from 4 letters (A, G, C, U) corresponds to a specific amino acid. e.g. CCC = proline, UUU = Phenylalanine, AGC = serine, CAG = glutamine

The codon wheel



Codon position 1 is in the center of the wheel, codon position 2 is in the middle of the wheel, and codon position 3 is near the periphery of the wheel, next to the three-letter amino acid designation. TER, a polypeptide chain termination codon.

Expanding the genetic code



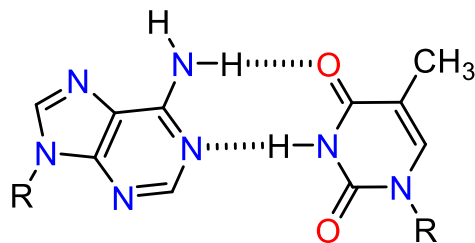
methylisoquinolinethione: methoxynaphthyl

- ❑ The natural four-letter genetic alphabet AT and GC is conserved throughout all life.
- ❑ Unnatural base pairs have been developed. **d5SICS-dNaM** is equivalent to an additional base pair in DNA. When combined with AT and GC it provides a six-letter genetic alphabet.
- ❑ This modified DNA may be amplified by DNA polymerases with high efficiency and <99.9% fidelity.
- ❑ The first organism that can grow and replicate with an unnatural base pair in its DNA – giving the cell six nucleotides instead of the usual four – has been created.
- ❑ **NB: Shape recognition of base pairs by DNA polymerases. H-bonding not necessary**

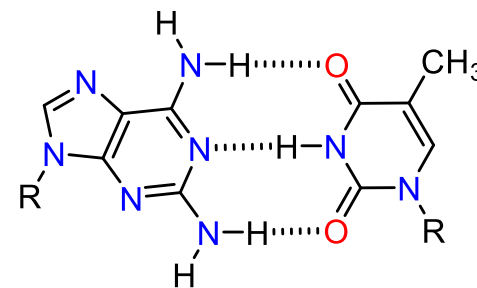
Increasing the stability of duplexes

- ☐ Oligonucleotides are used extensively in diagnostic and therapeutic applications.
- ☐ They are required to recognise short complementary DNA or RNA sequences embedded in longer nucleic acid stretches.
- ☐ High selectivity for the correct target and tight binding are important.
- ☐ This cannot always be achieved by unmodified oligonucleotides.

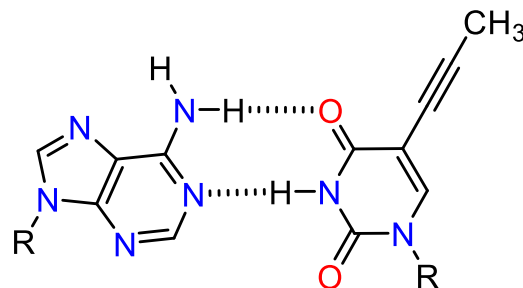
Increasing duplex stability via non-canonical A or T nucleobases



A - T



2-aminoadenine - thymine base pair
(3 hydrogen bonds)

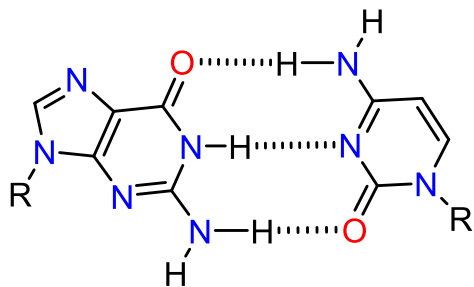


A - 5-propynyl T base pair
(extra base stacking)

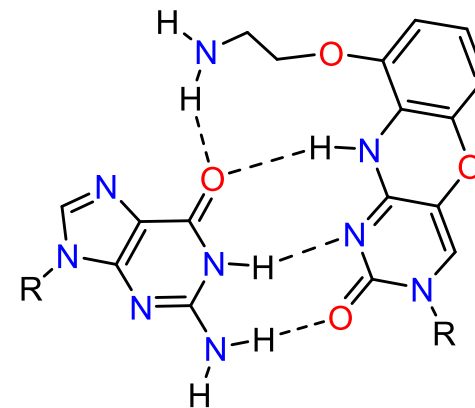
Duplex stability can be increased by

- ☐ Increasing the number of the inter base H bonds
- ☐ Increasing base stacking
- ☐ Introducing positive charges into the nucleotides

Increasing duplex stability via non-canonical C nucleobases

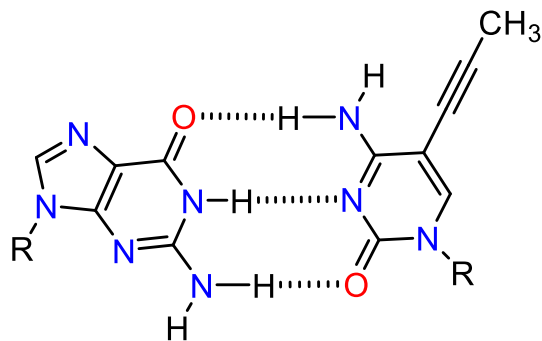


G - C



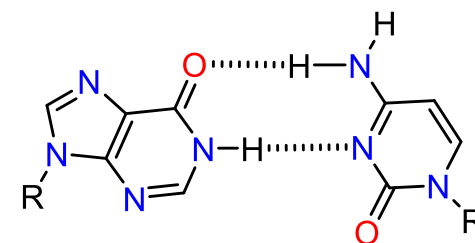
G - clamp base

(4 H-bonds and extra base stacking)



G - 5-propynyl C

(3 H-bonds and extra base stacking)



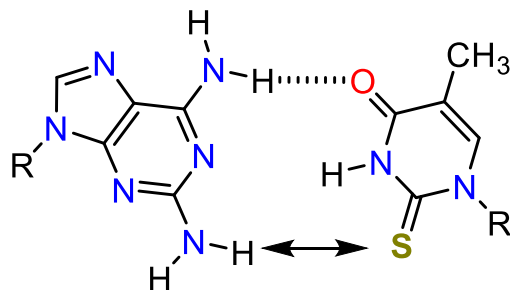
Deoxyinosine - C*

(2 H-bonds)

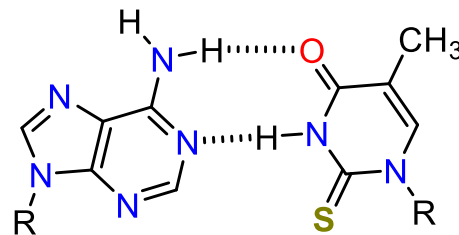
destabilising

*Sometimes it is desirable to decrease duplex stability. These can be achieved by reducing the number of H bonds.

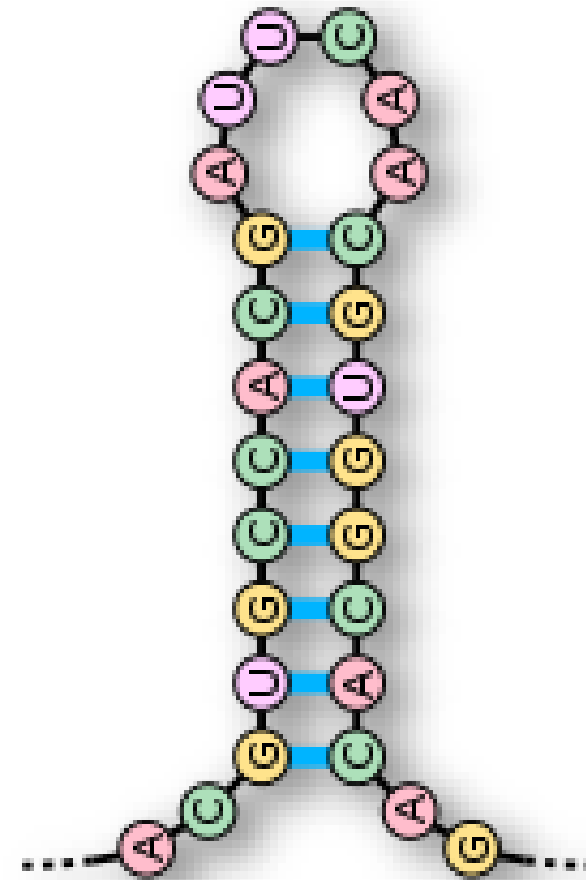
Reducing intramolecular structure via non-canonical A and T nucleobases



2-Aminoadenine - **2-thiouracil** base pair
steric clash, 1 hydrogen bond (unstable)

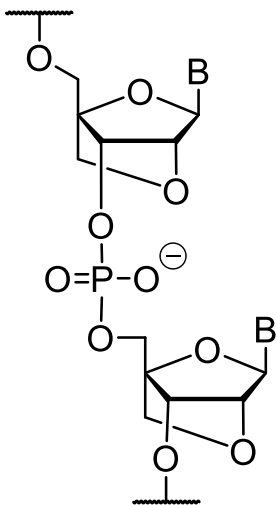


Adenine - **2-thiouracil** base pair
2 hydrogen bonds (stable)



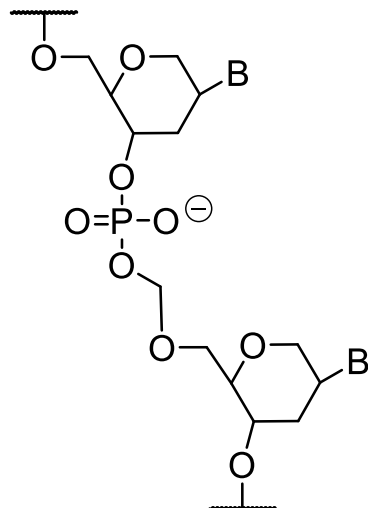
Applications: Oligonucleotides containing 2-thio-T and 2-aminoadenine form stable duplexes with complementary unmodified DNA but do not form intramolecular structures that might adversely affect their performance

Analogue of natural nucleic acids with modified sugars and/or backbones

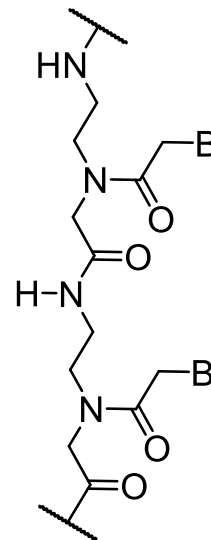


LNA

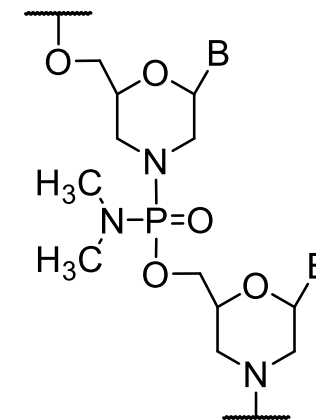
(conformationally locked in A-form)



HNA



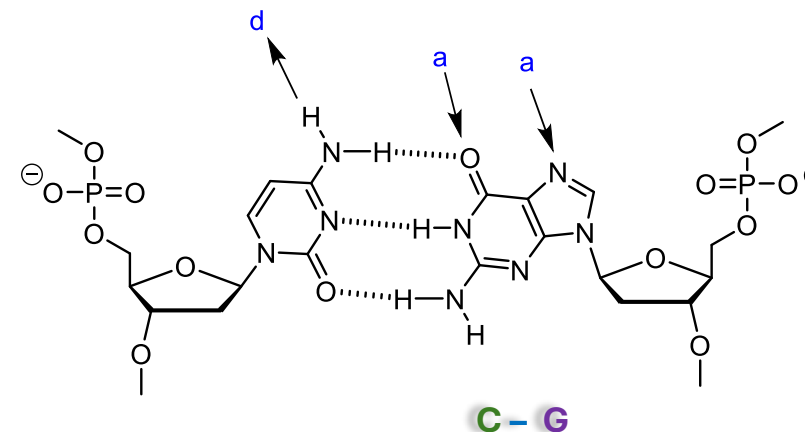
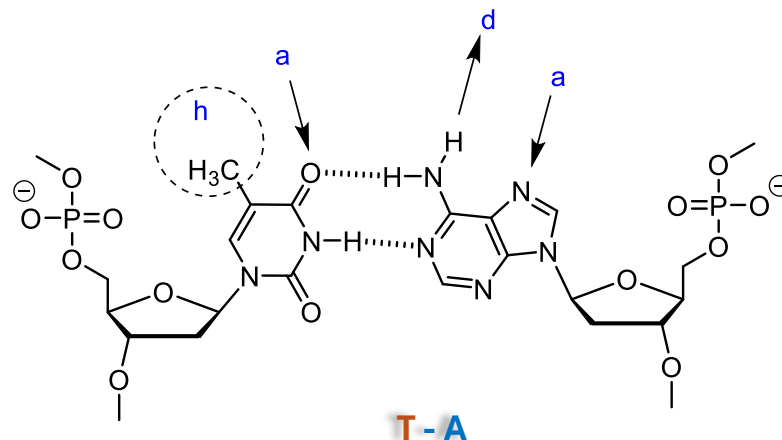
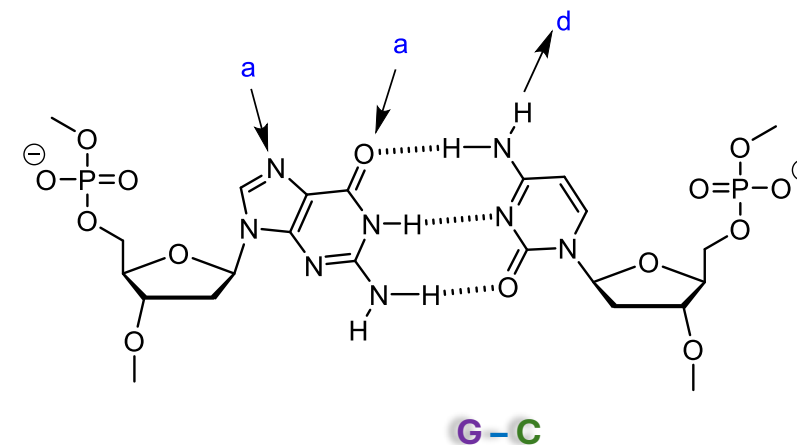
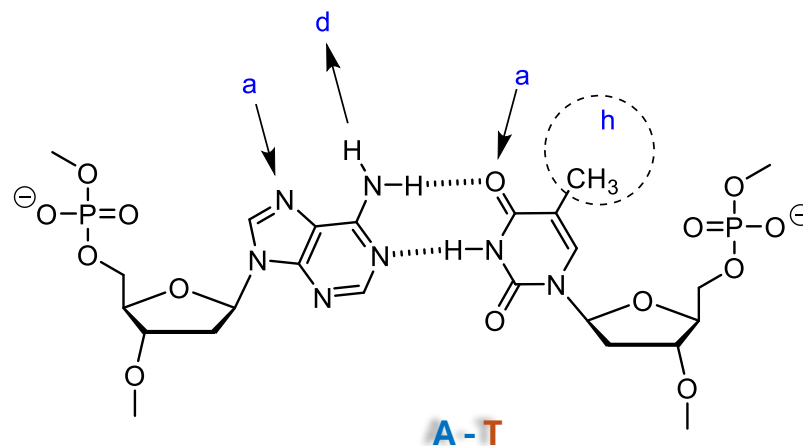
PNA
uncharged



Morpholino uncharged/cationic

- ❑ These analogues are designed to bind to complementary DNA or RNA with high affinity.
- ❑ Examples shown here are locked nucleic acid (LNA), hexitol nucleic (HNA), peptide nucleic acid (PNA) and morpholino nucleic acid analogues

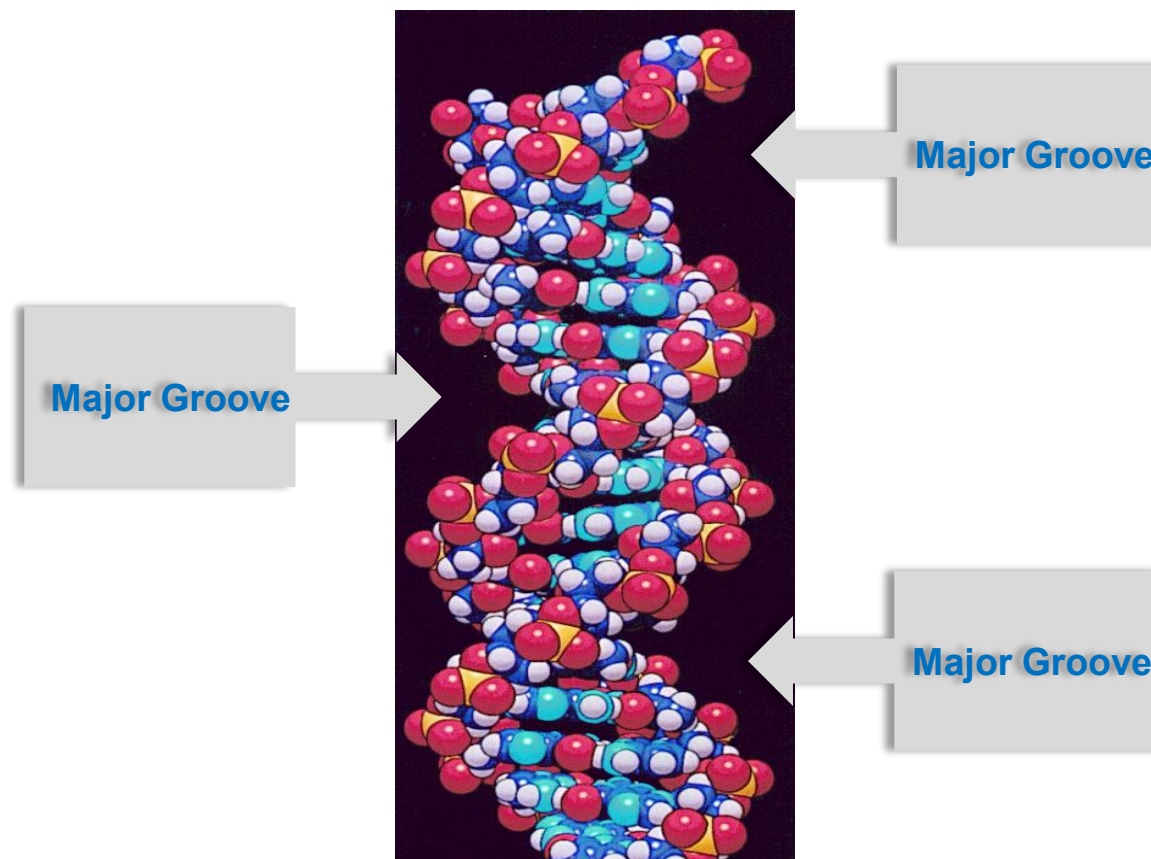
A - T, T - A, C - G and G - C are different in the major groove of DNA



a = H-bond acceptor, d = H-bond donor, h = hydrophobic

- ❑ The major groove of DNA provides an accessible surface for protein binding and recognition.
- ❑ This allows enzymes and other DNA-binding proteins to recognize specific DNA sequences through the major groove.

DNA duplex can bind a third strand of DNA

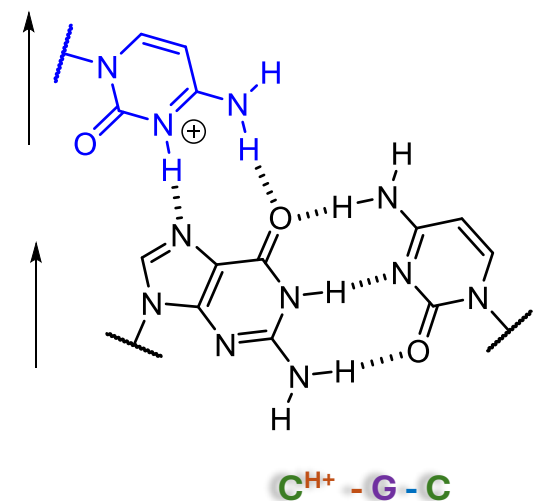
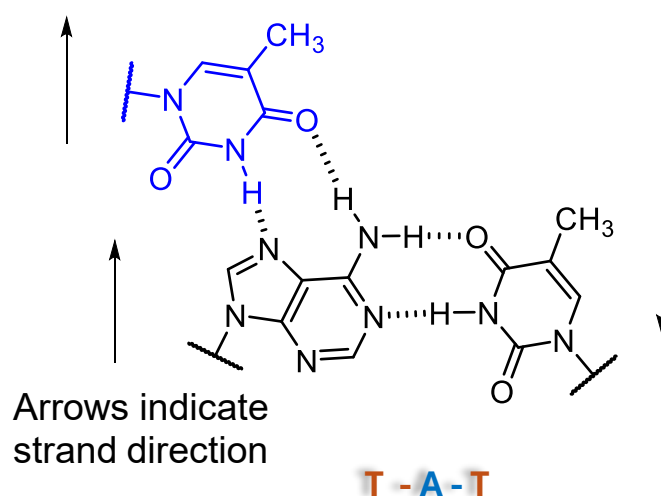


3rd strand binding is in the major groove

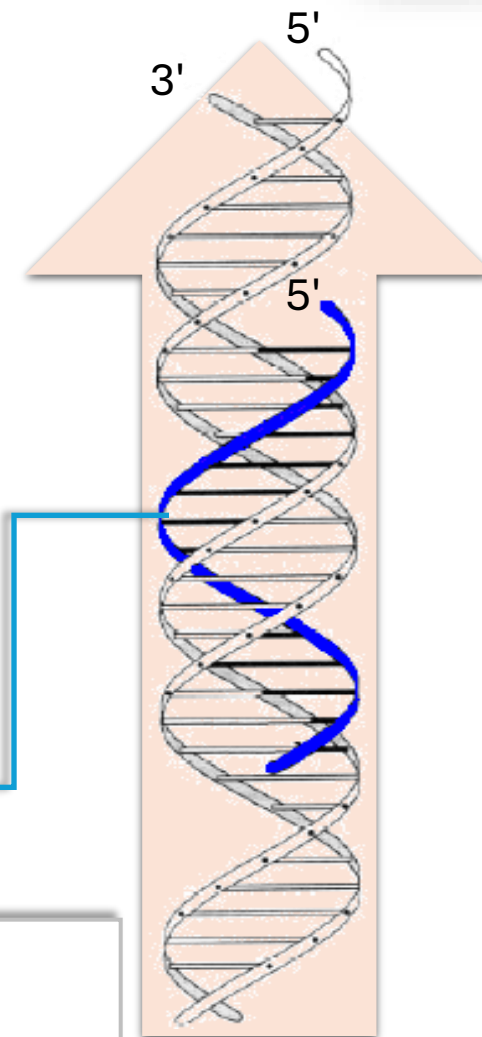
DNA Triple Helix Formation (TFO)

Parallel triplexes

5' -TCCTTGT TFO
 5' -AGGAAGA Duplex
 3' -TCCTTCT



The third strand binds in the major groove of the DNA duplex, forming specific interactions with one strand of the duplex.



Advantages

- ❑ Recognises duplex sequence without denaturation
- ❑ Means of recognising double stranded genomic DNA

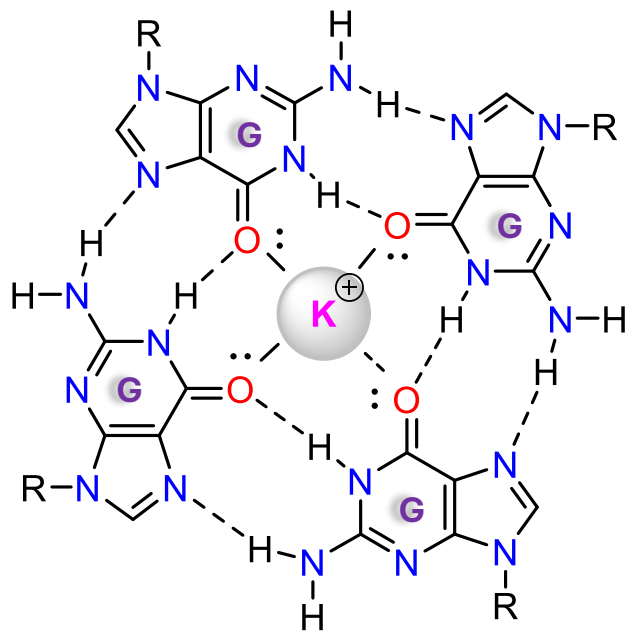
Disadvantages

pH-dependent (C^{H+} - G - C), low affinity, limited sequence recognition

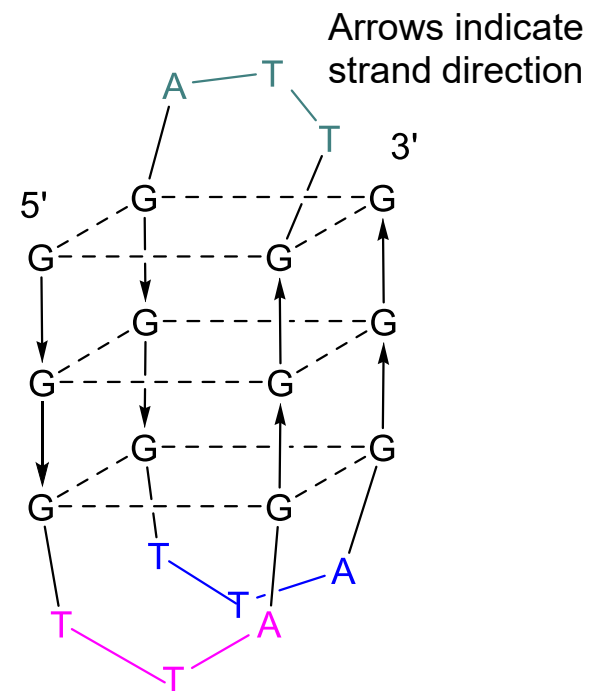
DNA Quadruplex Formation

5' -...GGGTTAGGGTTAGGGTTAGGG...3'

The metal ion is essential.
K⁺ is the best fit and the
most stabilising, but Na⁺
is also possible.



**G-quartet with H-bonding
and base stacking.**



**A G-quadruplex - in a single
strand of DNA.**

- Single strands of DNA can fold into G-quadruplexes which are stabilised by G-quartets. In the human genome these structures form and stabilise the ends of chromosomes. They also block gene expression in a reversible manner.
- G quadruplexes are very stable structures in the presence of K⁺.
- Quadruplexes can be very insoluble in aq. Media.

Jonchhe, S., Lahiri, S. & Rothenberg, E. DNA G-quadruplexes: Structural and functional insights. *DNA Repair* **156**, 103910 (2025).

Summary & Key Takeaways

- ❑ Nucleic acids play a central role in modern diagnostics and therapeutics; therefore, understanding the basic structures of DNA and RNA is essential.
- ❑ The structure of a nucleic acid determines its chemical properties, stability and biological function.
- ❑ Understanding the basic components and structural features of DNA and RNA provides a foundation for studying more advanced nucleic-acid technologies.
- ❑ Small changes in nucleic-acid structure can have significant effects on their biological properties and applications.
- ❑ A strong understanding of nucleic-acid chemistry is therefore crucial for the rational design and development of nucleic-acid-based technologies.