Molecular Biology
Part I

Genes

Genes are discrete physical entities present in all living organisms that control hereditary characteristics passed from parents to offspring of organisms.

Study of heredity is called **Genetics**. Gregor Mendel an unknown monk in Brno (now in Czechoslovakia, it was in Austria) published a paper in 1886 that pioneered the experimental study of genetics and his famous laws are called **Mendel's laws**. Classical genetics assumed genes are abstract attributes occurring in two variant forms (called *alleles- dominant and recessive*). Each individual inherits two genes, one from each of its parents.

In the 1930, it was recognized that like all particles in human body, genes must be composed of molecules and the field devoted to understanding the chemical nature of genes was termed **molecular biology**.

Mendel and his Genes

- What are genes?

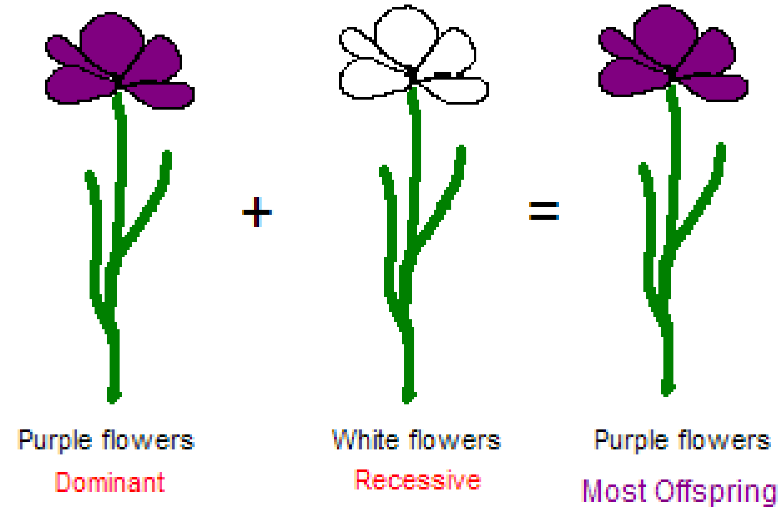
  -physical and functional traits that are passed on from one generation to the next.
- Genes were discovered by Gregor Mendel in the 1860s while he was experimenting with the pea plant. He asked the question:

Do traits come from a blend of both parent's traits or from only one parent?

The Pea Plant Experiments

- Mendel discovered that genes were passed on to offspring by both parents in two forms: dominant and recessive.

  - The dominant form would be the phenotypic characteristic of the offspring



Biological Information

Soon biologists realized that genes are not merely units of inheritance but they are actually units of **biological information** that control all aspects of life –birth, growth, functioning as a living organism and death.

Bioinformatics and Computational Molecular Biology are concerned with the use of computing and mathematical sciences as tools to advance traditional laboratory-based biology.

The need to process an exponentially growing amount of biological information for further scientific advances and to understand its role in heredity, chemical processes within the cell, drug discovery, evolutionary studies etc. have created new problems that are of interdisciplinary nature.

Protein or DNA

- By 1920s, it was established that:
  - Genes reside on chromosomes
  - Chromosomes are made of protein and DNA

  Chromosomes were discovered in 19th century as threadlike structures in the nucleus of a eukaryotic cell that could be observed under microscope as the cells begin to divide.

  Biochemical analysis concluded that chromosomes contain both DNA and protein.
- The question was: What is the genetic material?
  - Protein?
  - DNA?
  - Both?

  The chemical structures of both DNA and protein were still unknown mysteries.

Properties of Genetic Material

- Genetic Material:
  - Must be able to exist in almost infinite variety of forms
  - Protein was believed to be able to form long chains – **macro molecules**. Many proteins were known.
  - DNA was believed to be a small, invariant molecule.

From the point of view of variability in species, proteins as carriers of genetic information seemed to make more sense. But this hypothesis was proved to be false by a famous experiment (by Griffith in 1942 and interpreted by Avery) that discovered a fundamental principle of Biology, called the **Transforming Principle.**

The Transforming Principle

- *Diplococcus pneumoniae* exists in two forms:
  - Both the forms have a coating that surrounds the cell
  - The coating is made of a *polysaccharide* secreted by the bacterium
  - Each form secretes a different polysaccharide, hence a different coating & appearance.
  - The *smooth* (or *S*) form is virulent, whereas the *rough* (or *R*) form is avirulent.

Experiment (Griffith,Avery 1928)

- Mouse injected with S form -> infected with pneumonia
- Mouse injected with R form -> healthy
- Mouse injected with heat-killed S form -> healthy
- Mouse injected with heat-killed S bacteria + live R bacteria -> infected with pneumonia
- Conclusion – a component (the genetic material) of heat-killed S-bacteria was able to enter R cell and transform them into smooth form.

DNA

- Stores all information of life
- 4 “letters” base pairs. AGTC (adenine, guanine, thymine, cytosine) which pair A-T and C-G on complimentary strands.

http://www.lbl.gov/Education/HGP-images/dna-medium.gif

Structure of DNA

DNA stands for Deoxyribonucleic acid.

A DNA is a long ***polymeric molecule***, also called a **macromolecule**.

A **polymer** is a long chain of molecules called **monomers**.

The monomer for DNA is called a nucleotide.

The nucleotides are components of nucleic acid.

Nucleotide is itself a very complex molecule.

Covalent Bonds

- A **covalent bond** is formed when two atoms are held together due to sharing of electrons in their outer shells.

	Electrons in outer shell	Covalent bonds
Carbon(C)	4	4
Hydogen(H)	1	1
Nitrogen (N)	5	3,5
Oxygen(O)	6	2
Phosphorus(P)	5	3,5
and Sulfer(S)	6	2 to 6

- The stabilty of a bond is expressed as kilocalorie/per $6.02x10^{23}$ molecules. O-H: 110 kcal/mol., C-O:84 kcal/mol,C=O: 172 kcal/mol.

Hydrogen Bonds

The **hydrogen bond** is a weak electro static bond between a negatively charged atom (viz. oxygen or nitrogen) and a positively charged hydrogen atom that is already covalently bound. The energies of these weaker bonds are in the range of 1 to 6 kcal/mole.

Covalent bonds are indicated by solid lines connecting the molecules. The hydrogen bonds are depicted by drawing a dotted line between the molecules.

Nucleotides are made of three basic components

- **Nitrogenous Base:**
  - N is important for hydrogen bonding between bases
  - A – adenine with T – thymine (double H-bond)
  - C – cytosine with G – guanine (triple H-bond)

- **Sugar:**
  - Ribose (5 carbon)
  - Base covalently bonds with 1’ carbon
  - Phosphate covalently bonds with 5’ carbon
  - Normal ribose (OH on 2’ carbon) – RNA
  - deoxyribose (H on 2’ carbon) – DNA
  - dideoxyribose (H on 2’ & 3’ carbon) – used in DNA sequencing

- **Phosphate:**
  - negatively charged

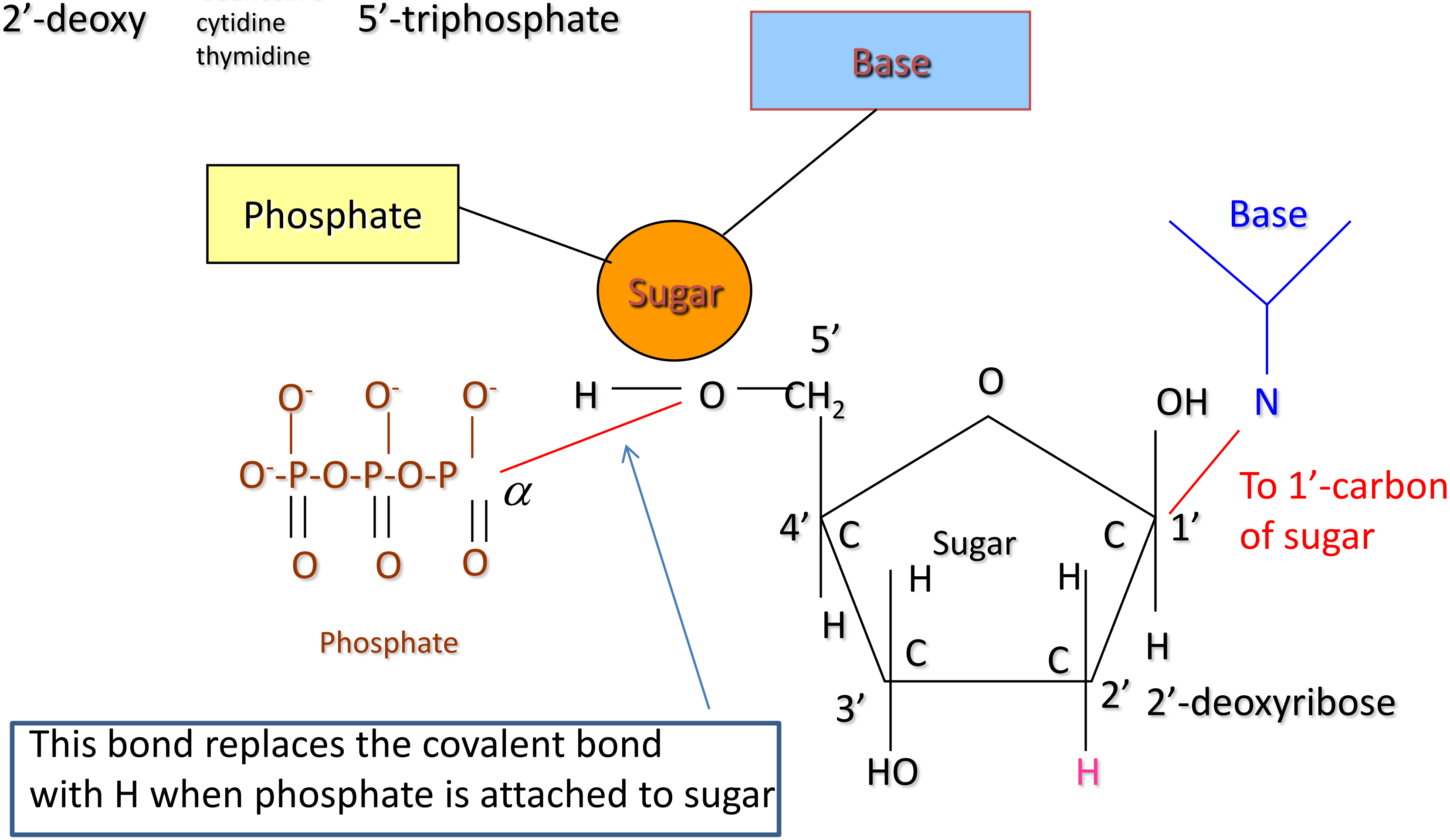
- DNA is a polynucleotide
- Individual nucleotides are bound by a phosphodiester bond, a covalent bond.

The Nitrogenous Bases

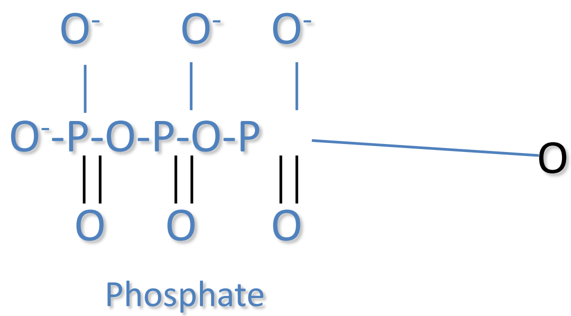
- **Purines**
  - A and G
  - double ring structures attached to 1’ carbon of the sugar
  - are heavier
- **Pyrimidines**
  - C and T
  - Single ring structures attached to the 1’-carbon of the sugar
  - Lighter

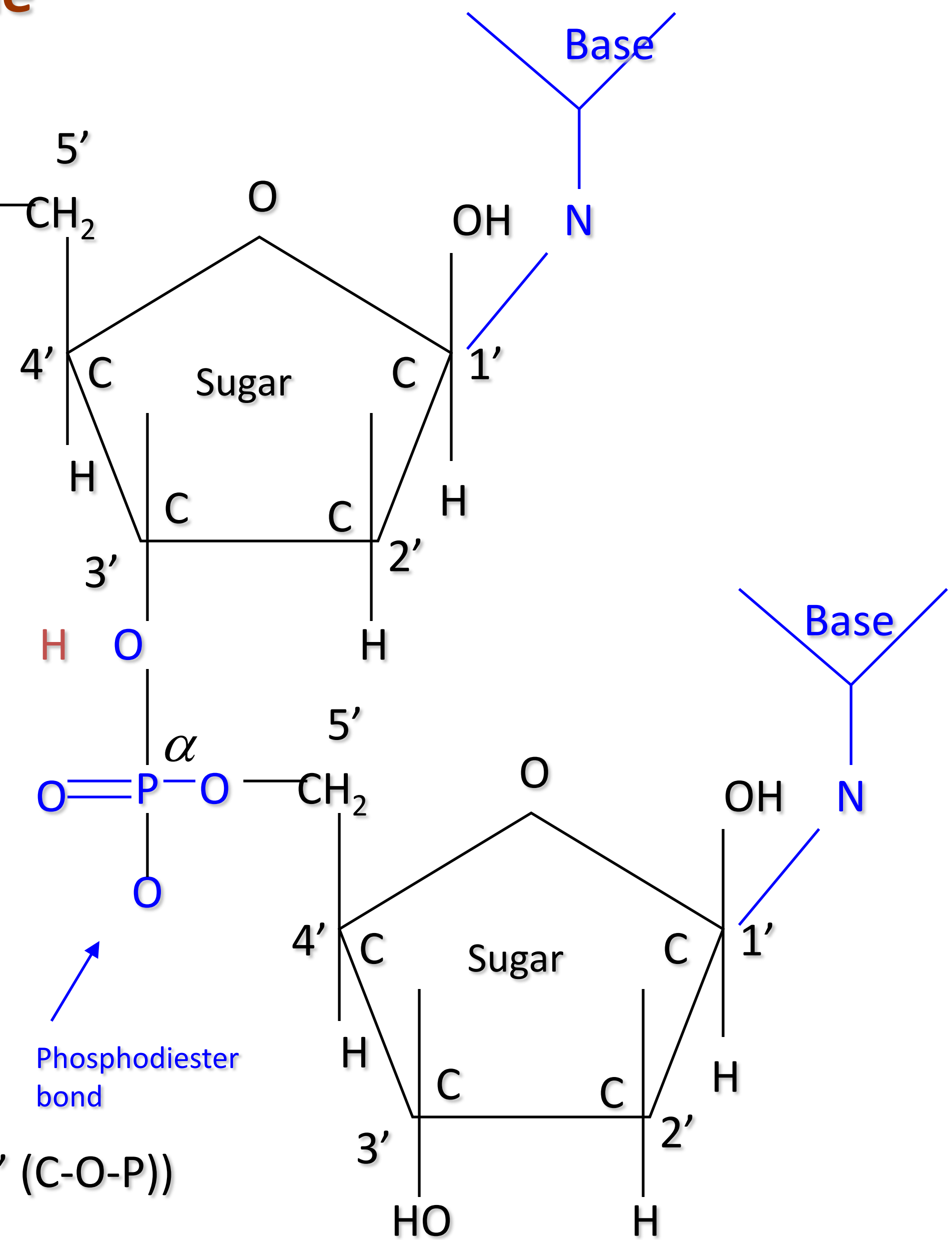
Chemical Structure of a Nucleotide

2'-deoxy adenosine / Guanosin e / cytidine / thymidine 5'-triphosphate

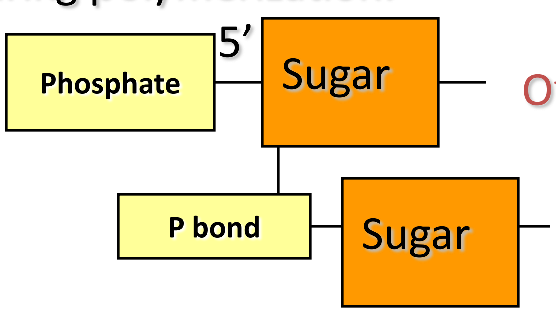


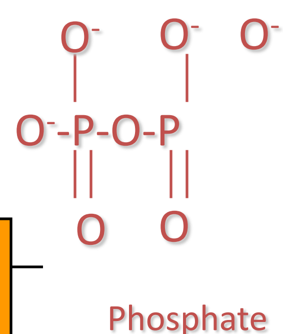
Sugar-phosphate Backbone





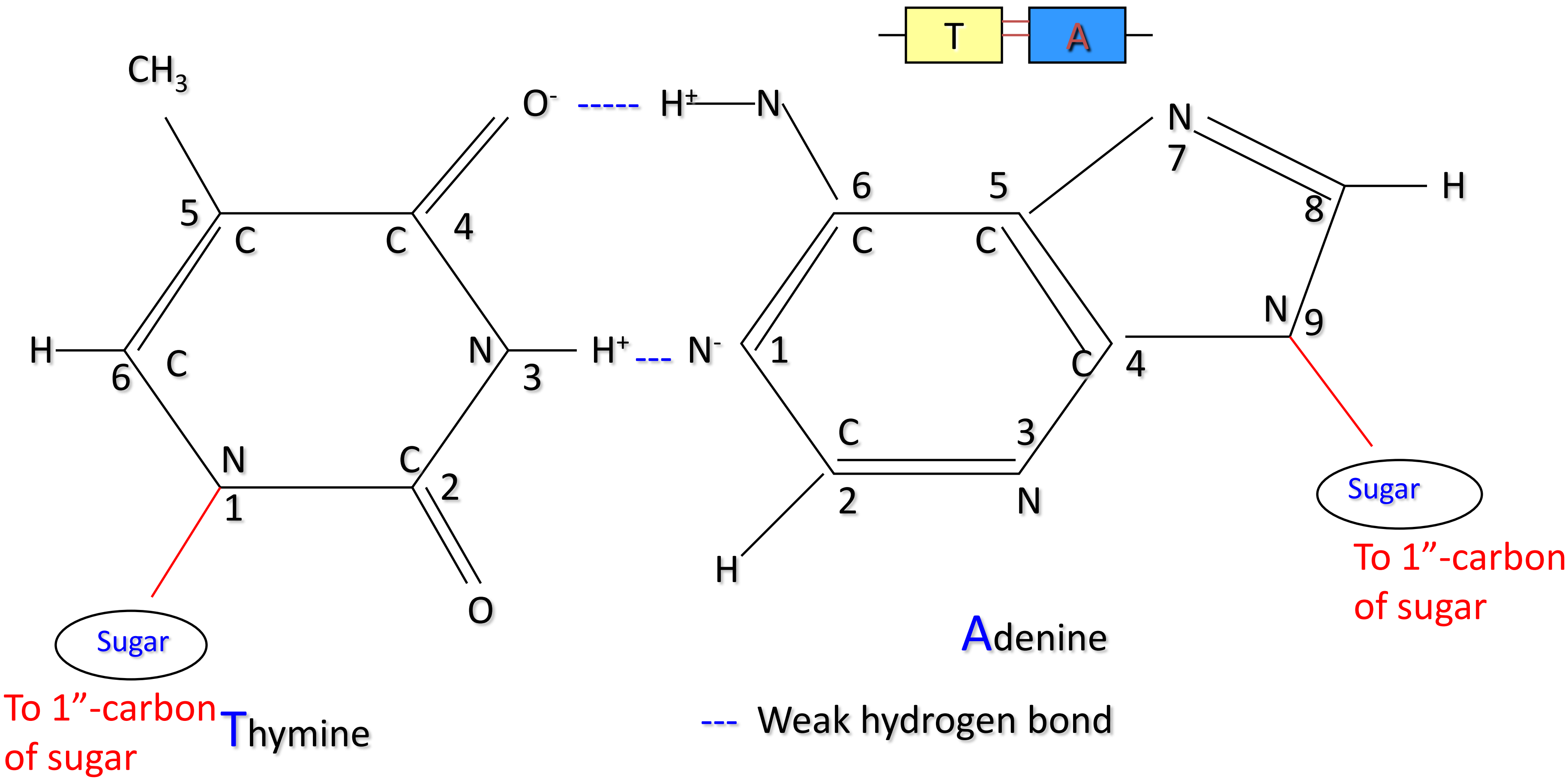
The nucleotides are linked together by joining the α-phosphate group attached to 5'carbon of one to 3'-carbon of the next in chain. The beta and gamma phosphates and the hydroxyl group of 3'-carbon are cleaved off during polymerization.



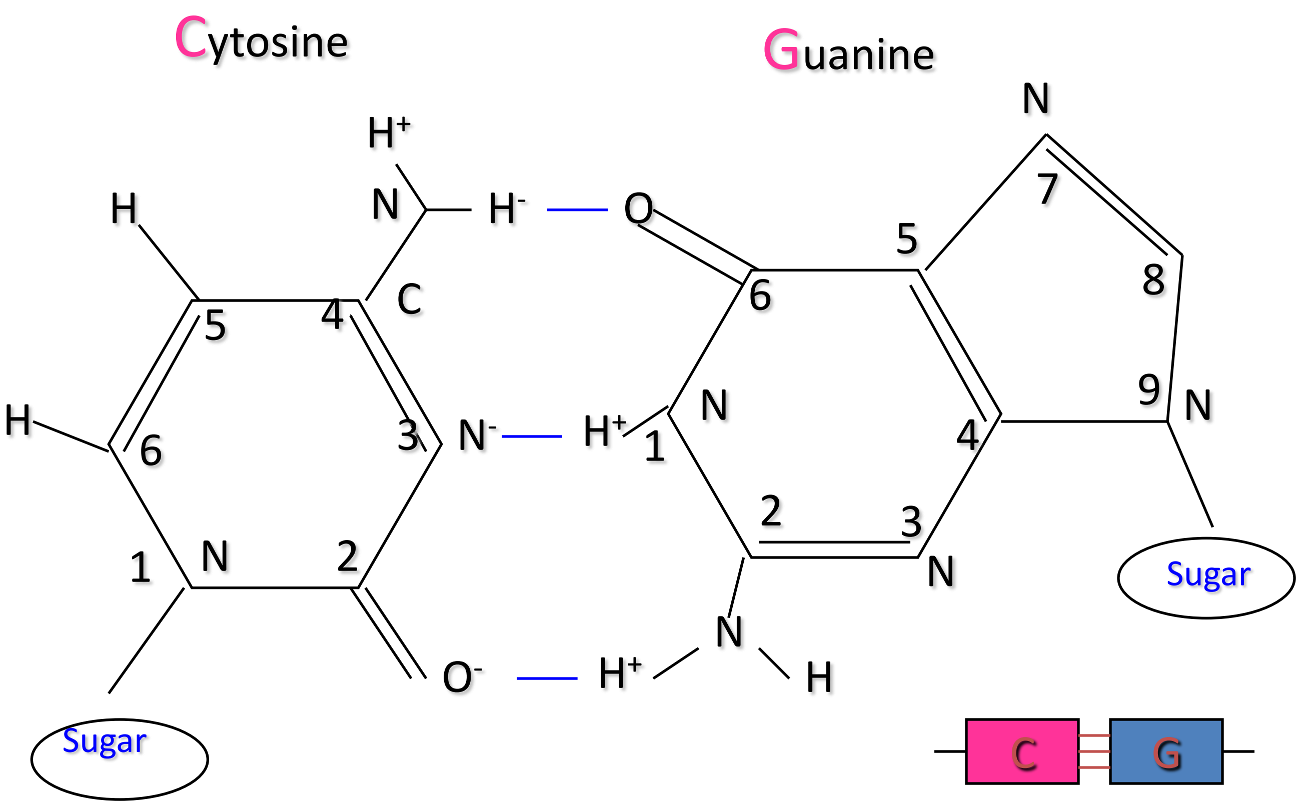


(3'-5' 'phospho' 'di'(two) – 'ester' (C-O-P))

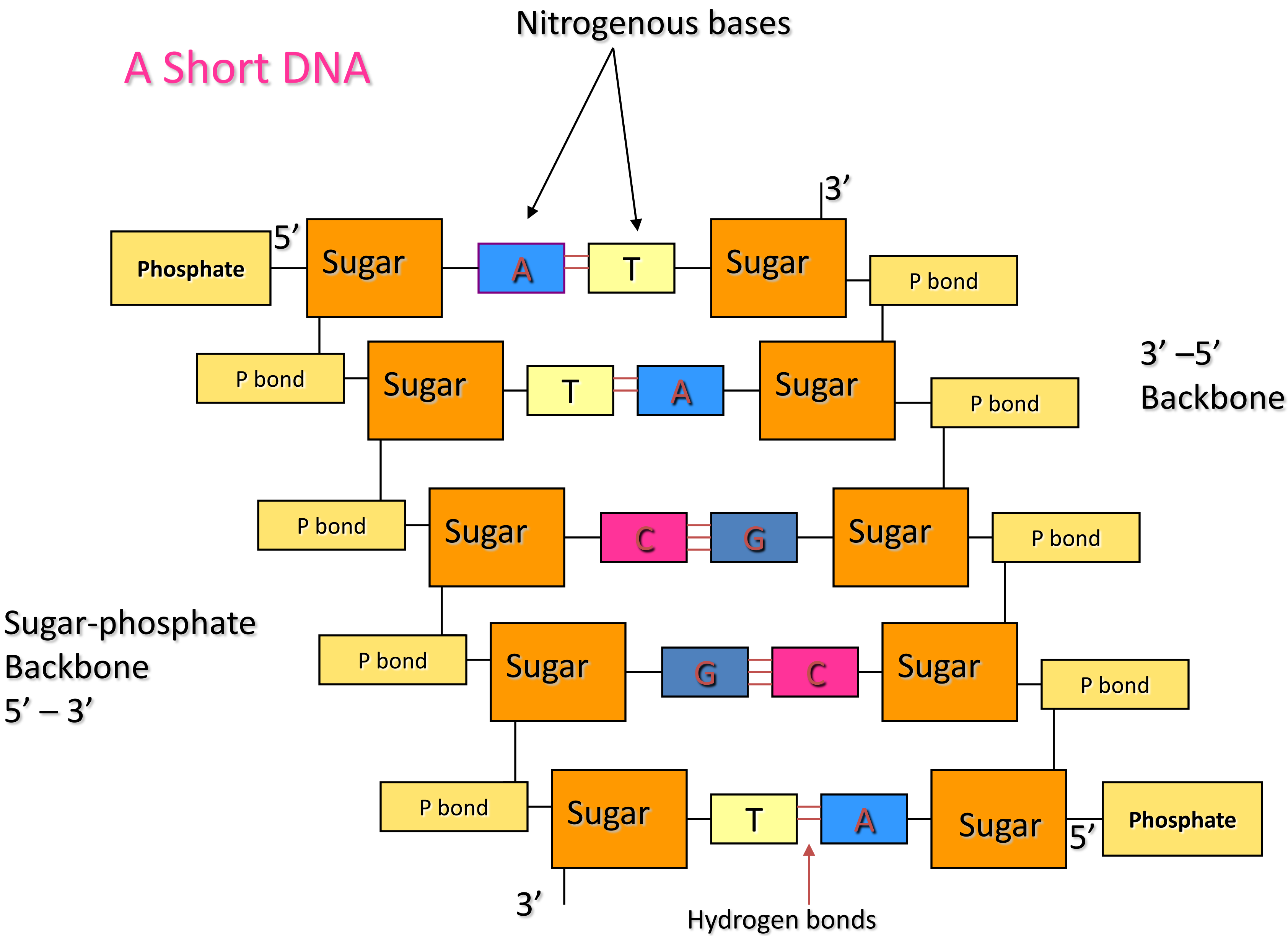
Chemical Structures of the Bases and Base Pairing Between A and T



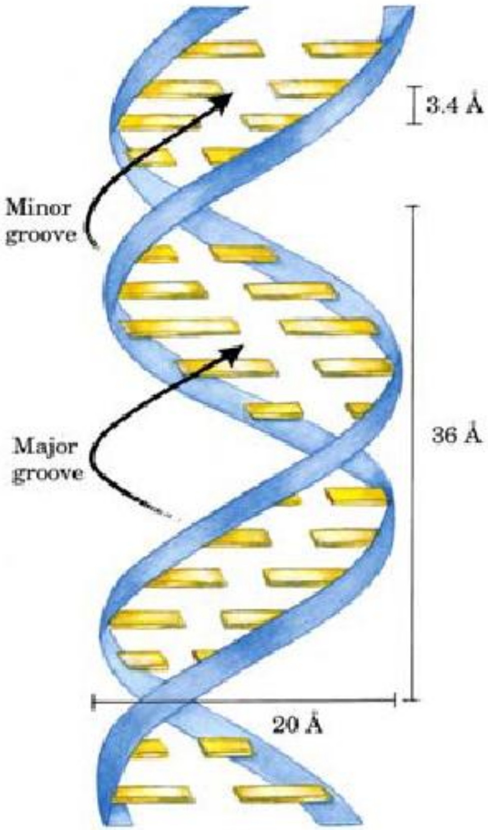
Chemical Structures and base pairing between C and G



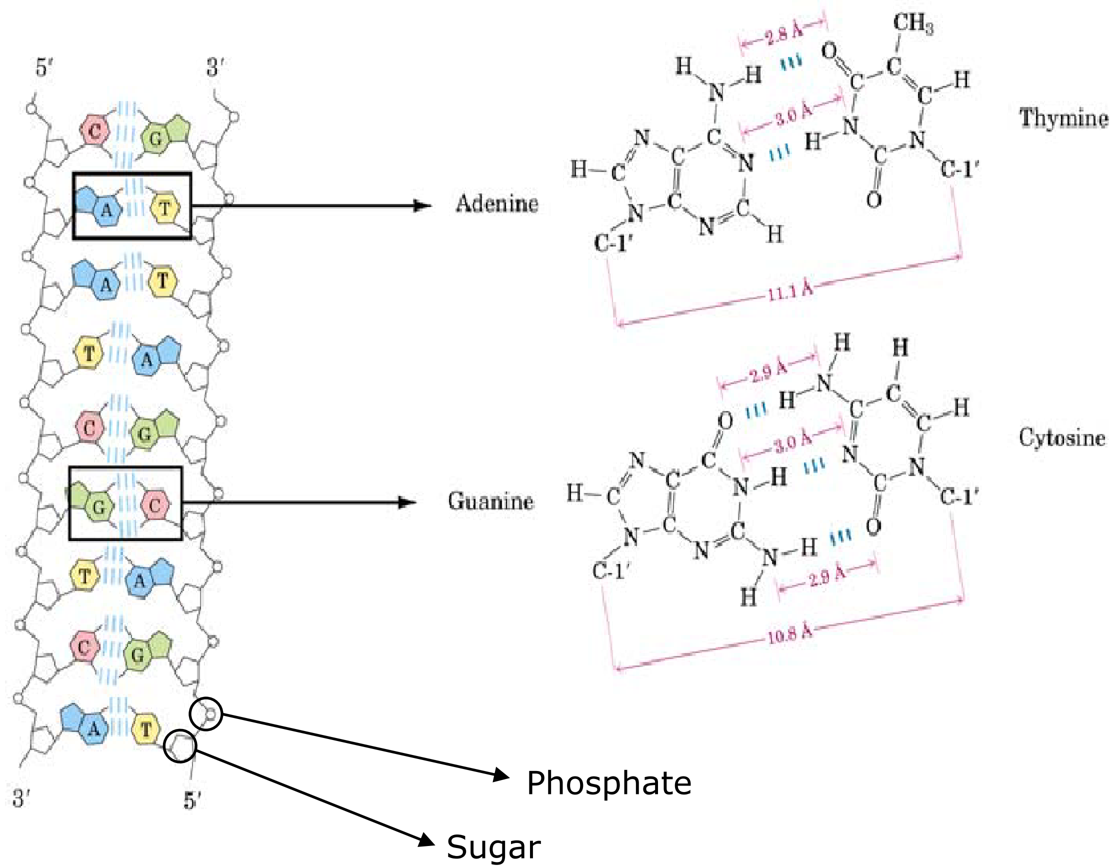
A ‘flattened’ version of the DNA structure



Basic Structure



Watson-Crick base pair structures



Orientation

The DNA or the polynucleotide has two distinct ends. The top nucleotide has the triphosphate group that does not participate in forming the phosphodiester bond. This is called the 5'-terminus. At the other end the 3'-hydroxyl group does not participate in the bond formation. This is called the 3'-terminus. These two distinct ends imply that the polynucleotides have a direction: the 5'-3' direction called the 'downstream' and the 3'-5' direction called the 'upstream'. These directions are very important as we will see later.

DNA, continued

- DNA has a double helix structure. However, it is not symmetric. It has a “forward” and “backward” direction. The ends are labeled 5’ and 3’ after the Carbon atoms in the sugar component.

5’ AATCGCAAT 3’

3’ TTAGCGTTA 5’

DNA always reads 5’ to 3’ for transcription replication

A String of Four Letters A, T, C, G.

Apparently, there is no limit on the number of nucleotides that can be joined together or any specific ordering of these nucleotides. Thus, the number of possible ‘strings’ that can be formed using this four letter alphabet grows exponentially.

A polynucleotide of length 10 could be any one of 4^{10}=1,048,576 possible different sequences such

as

ATCGAGGTCT

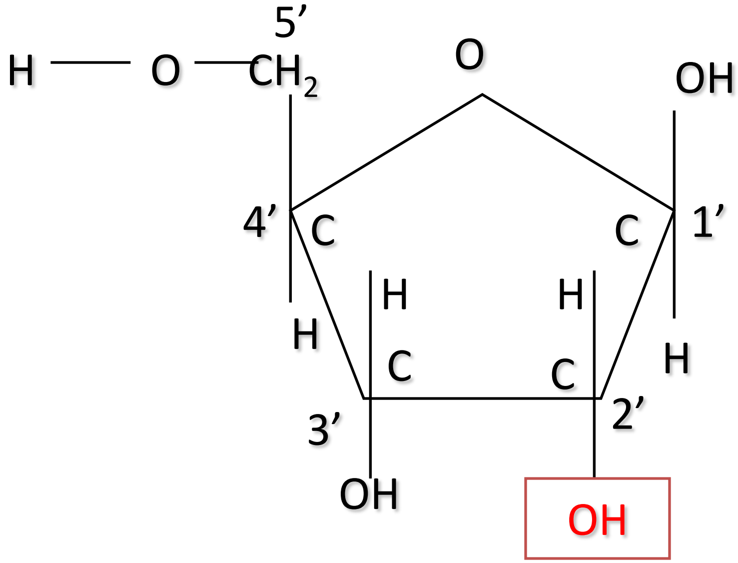
GTATCCGATA

This provides potentially a very large number of variations of the genetic material if DNA length is thousands or millions.

RNA (Ribonucleic acid)

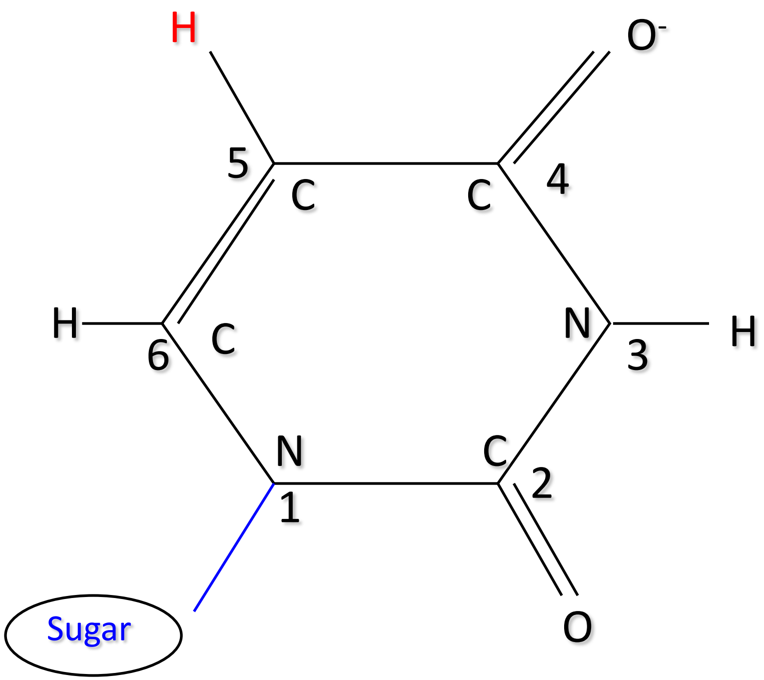
RNA, like DNA, is a polynucleotide.

1)The 'sugar' component in DNA (deoxyribose) is a ribose, that is, the hydroxyl group has not been 'deoxygenated' and (see next slide)

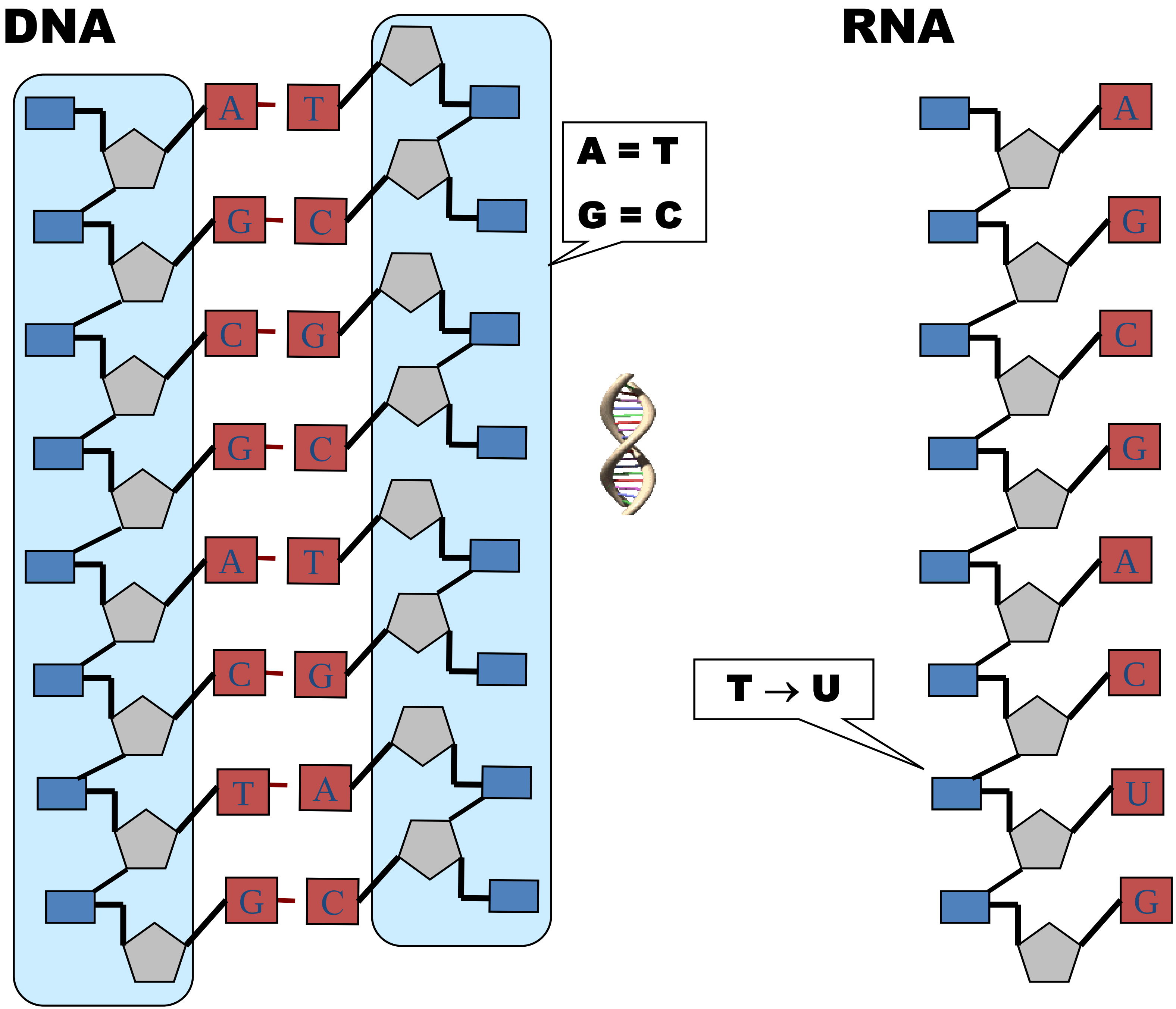


RNA (Ribonucleic acid)

2)The base Thyamine is replaced by Uracil



Uracil (CH_3 is replaced by H)



DNA
A T
G C
C G
G C
A T
C G
T A
G C
A = T
G = C
RNA
A
G
C
G
A
C
U
G
T → U

The difference between RNA and DNA

Except for the differences in ribose and uracil, the structure of RNA as a polynucleotide is the same as that of DNA. The same 3’-5’ phosphodiester bond links the nucleotides in the chain and also there could be any arbitrary ordering of the four elements A,U,C and G and there is no restriction on the length of the sequence.

But , there is one very important difference. RNA in its natural form inside the cell usually exists as a single chain, but DNA exists in the form of two chains wrapped around each other in the form of a **Double Helix, a fundamental and the most important discovery in molecular biology in the early 50’s.**

Basic Structure Implications

- **DNA is (-) charged due to phosphate:**
  gel electrophoresis, DNA sequencing (Sanger method)

- **H-bonds form between specific bases:**
  hybridization – replication, transcription, translation
  DNA microarrays, hybridization blots, PCR
  C-G bound tighter than A-T due to triple H-bond

- **DNA-protein interactions (via major & minor grooves):**
  transcriptional regulation

- **DNA polymerization:**
  5’ to 3’ – phosphodiester bond formed between 5’ phosphate and 3’ OH

Double helix of DNA

- The double helix of DNA has these features:
  - Concentration of adenine (A) is equal to thymine (T)
  - Concentration of cytidine (C) is equal to guanine (G).
  - Watson-Crick base-pairing :A will only base-pair with T, and C with G
    - base-pairs of G and C contain three H-bonds,
    - Base-pairs of A and T contain two H-bonds.
    - G-C base-pairs are more stable than A-T base-pairs
  - Two polynucleotide strands wound around each other.
  - The backbone of each consists of alternating deoxyribose and phosphate groups

Double helix of DNA

- The DNA strands are assembled in the 5' to 3' direction
  - by convention, we "read" them the same way.
- The phosphate group bonded to the 5' carbon atom of one deoxyribose is covalently bonded to the 3' carbon of the next.
- The purine or pyrimidine attached to each deoxyribose projects in toward the axis of the helix.
- Each base forms hydrogen bonds with the one directly opposite it, forming base pairs (also called nucleotide pairs).

Double helix of DNA

- James Watson and Francis Crick proposed a model for the structure of DNA.
  - Utilizing X-ray diffraction data, obtained from crystals of DNA)
- This model predicted that DNA
  - as a helix of two complementary anti-parallel strands,
  - wound around each other in a rightward direction
  - stabilized by H-bonding between bases in adjacent strands.
  - The bases are in the interior of the helix
    - <u>Purine bases</u> form hydrogen bonds with <u>pyrimidine.</u>