Nucleic Acid: DNA & RNA

Lec. 8&9 | DNA & RNA

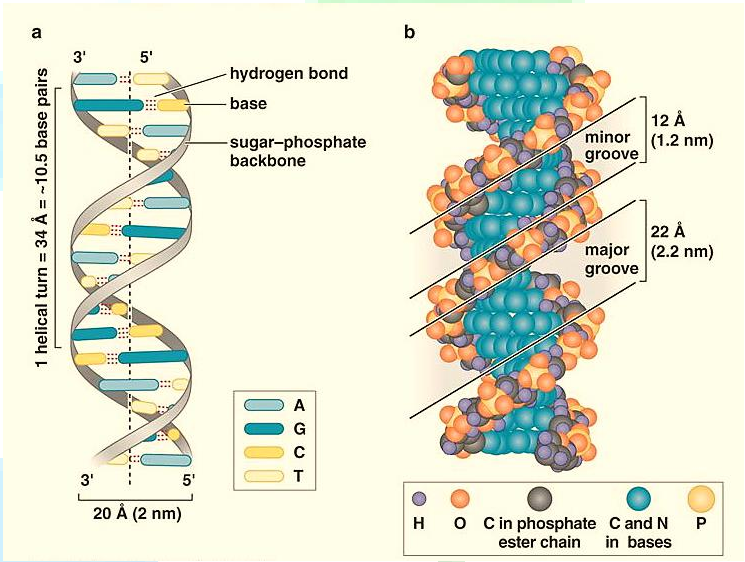
The Structures of DNA

INTRODUCTION

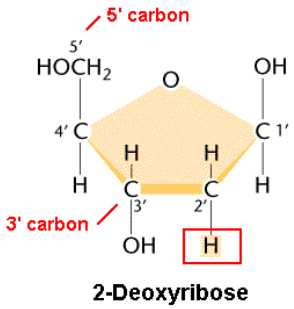
The discovery that DNA is the prime genetic molecule, carrying all the hereditary information within chromosomes, immediately focused attention on its structure. It was hoped that knowledge of the structure would reveal how DNA carries the genetic messages that are replicated when chromosomes divide to produce two identical copies of themselves. During the late 1940s and early 1950s, several research groups in the United States and in Europe engaged in serious efforts both cooperative and rival to understand how the atoms of DNA are linked together by covalent bonds and how the resulting molecules are arranged in three-dimensional space. Not surprisingly, there initially were fears that DNA might have very complicated and perhaps bizarre structures that differed radically from one gene to another. Great relief, if not general elation, was thus expressed when the fundamental DNA structure was found to be the double helix. It told us that all genes have roughly the same three-dimensional form and that the differences between two genes reside in the order and number of their four nucleotide building blocks along the complementary strands.

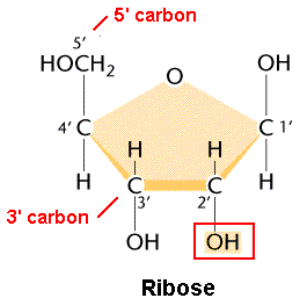
DNA STRUCTURE: DNA Is Composed of Polynucleotide Chains

The most important feature of DNA is that it is usually composed of two **Polynucleotide Chains** twisted around each other in the form of a double helix (Figure on right). The upper part of the figure (**a**) presents the structure of the double helix shown in a schematic form. Note that if inverted 180° (for example, by turning this book upside-down), the double helix looks superficially the same, due to the complementary nature of the two DNA strands. The space-filling model of the double helix, in the lower part of the figure (**b**), shows the components of the DNA molecule and their relative positions in the helical structure. The backbone of each strand of the helix is composed of alternating sugar and phosphate residues; the bases project inward but are accessible through the major and minor grooves.



Let us begin by considering the nature of the nucleotide, the fundamental building block of DNA. The nucleotide consists of a phosphate joined to a sugar, known as **2`-Deoxyribose,** to which a base is attached. The sugar is called *2`*-deoxyribose because there is no hydroxyl at position 2` (just two hydrogens). Note that the positions on the ribose are designated with primes to distinguish them from positions on the bases (see the discussion below). We can think of how the base is joined to

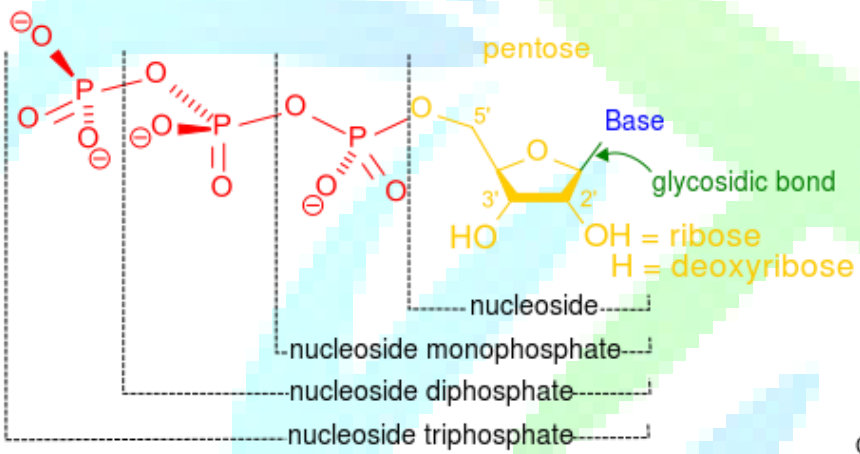


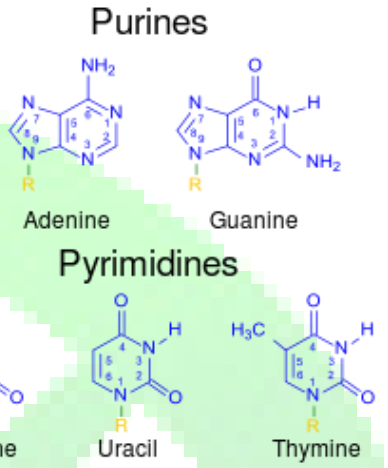


Nucleic Acid: DNA & RNA

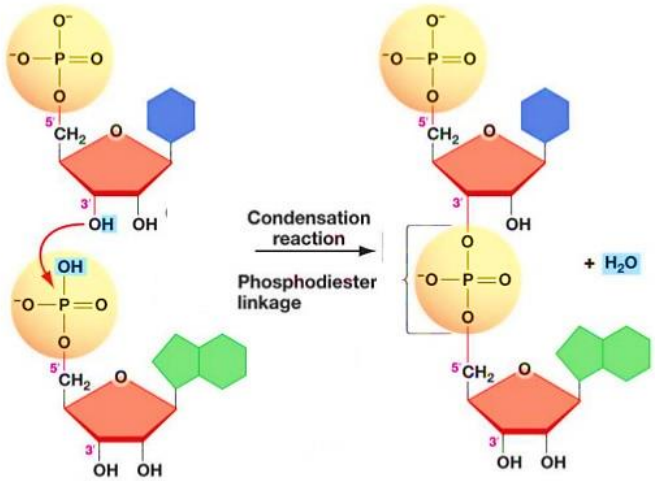
2`deoxyribose by imagining the removal of a molecule of water between the hydroxyl on the 1` carbon of the sugar and the base to form a glycosidic bond. The sugar and base alone are called a **Nucleoside.** Likewise, we can imagine linking the phosphate to 2`deoxyribose by removing a water molecule from between the phosphate and the hydroxyl on the 5` carbon to make a 5` phosphomonoester. Adding a phosphate (or more than one phosphate) to a **nucleoside** creates a **Nucleotide.**

Thus, by making a glycosidic bond between the base and the sugar, and by making a





phosphoester bond between the sugar and the phosphoric acid, we have created a nucleotide. Nucleotides are, in turn, joined to each other in polynucleotide chains through the 3` hydroxyl of 2`deoxyribose of one nucleotide and the phosphate attached to the 5` hydroxyl of another nucleotide. This is a **Phosphodiester linkage** in which the phosphoryl group between the two nucleotides has one sugar esterified to it through a 3` hydroxyl and a second sugar esterified to it through a 5` hydroxyl.

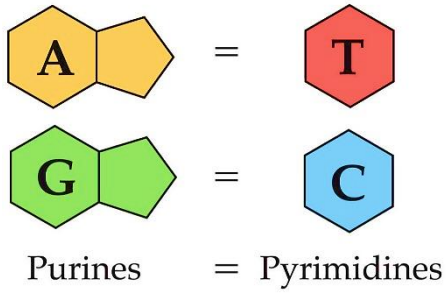


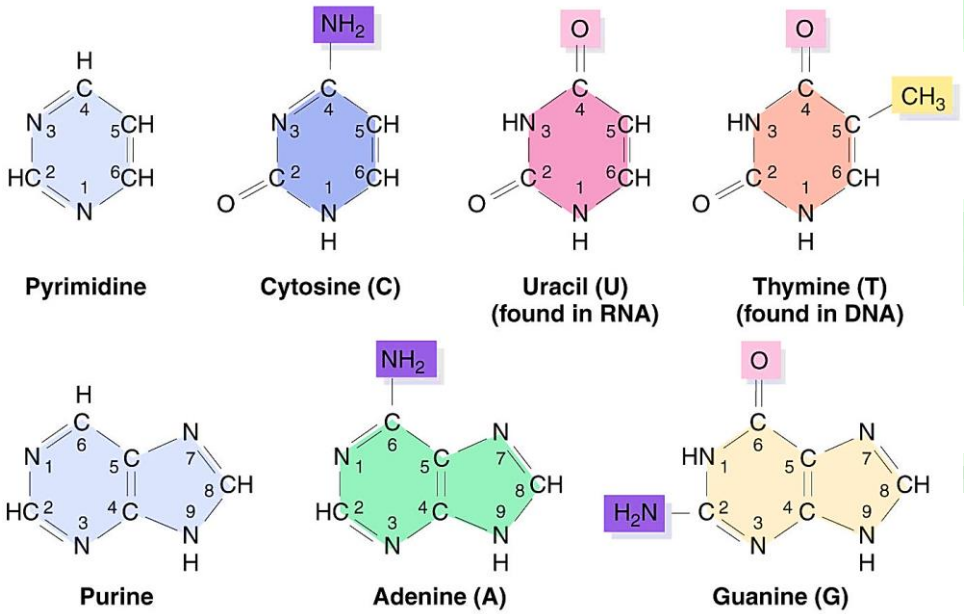
Phosphodiester linkages create the repeating, sugar-phosphate backbone of the polynucleotide chain, which is a regular feature of DNA. In contrast, the order of the bases along the polynucleotide chain is irregular. This irregularity as well as the long length is, as we shall see, the basis for the enormous information content of DNA. The phosphodiester linkages impart an inherent polarity to the DNA chain. This polarity is defined by the asymmetry of the nucleotides and the way they are joined. DNA chains have a free 5` phosphate or 5` hydroxyl at one end and a free 3` phosphate or 3` hydroxyl at the other end. The convention is to write DNA sequences from the 5` end (on the left) to the 3` end, generally with a 5` phosphate and a 3`hydroxyl.

Each Base Has Its Preferred Tautomeric Form

Nucleic Acid: DNA & RNA

The bases in DNA fall into two classes, **Purines** and **Pyrimidines**. The purines are **Adenine** and **Guanine**, and the pyrimidines are **Cytosine** and **Thymine**. The purines are derived from the double-ringed structure shown in Figure below. Adenine and guanine share this essential structure but with different groups attached. Likewise, cytosine and thymine are variations on the single-ringed structure. The figure also shows the numbering of the positions in the purine and pyrimidine rings. The bases are attached to the deoxyribose by glycosidic linkages at N1 of the pyrimidines or at N9 of the purines. Each of the bases exists in two alternative **Tautomeric states**, which are in equilibrium with each other. The equilibrium lies far to the side of the conventional structures are the predominant states and the ones important for base pairing. The nitrogen atoms attached to the purine and pyrimidine rings are in the amino form in the predominant state and only rarely assume the imino configuration. Likewise, the oxygen atoms attached to the guanine and thymine normally have the keto form and only rarely take on the enol configuration. As examples, tautomerization of cytosine into the imino form (a) and guanine into the enol form (b). As we shall see, the capacity to form an alternative tautomer is a frequent source of errors during DNA synthesis.

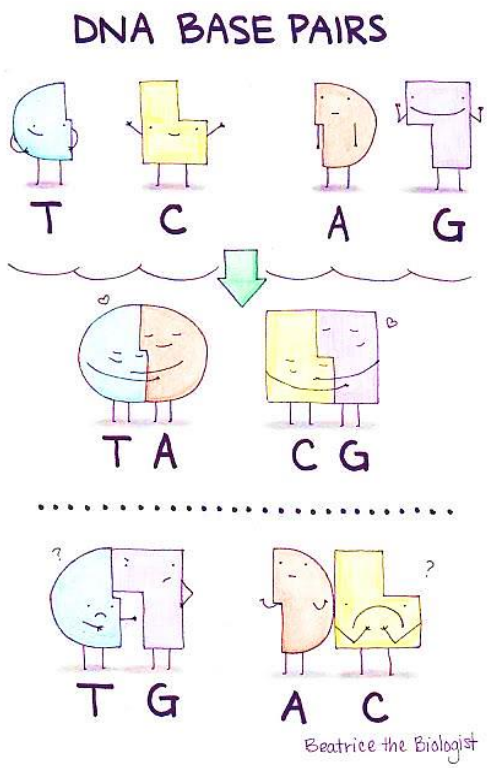


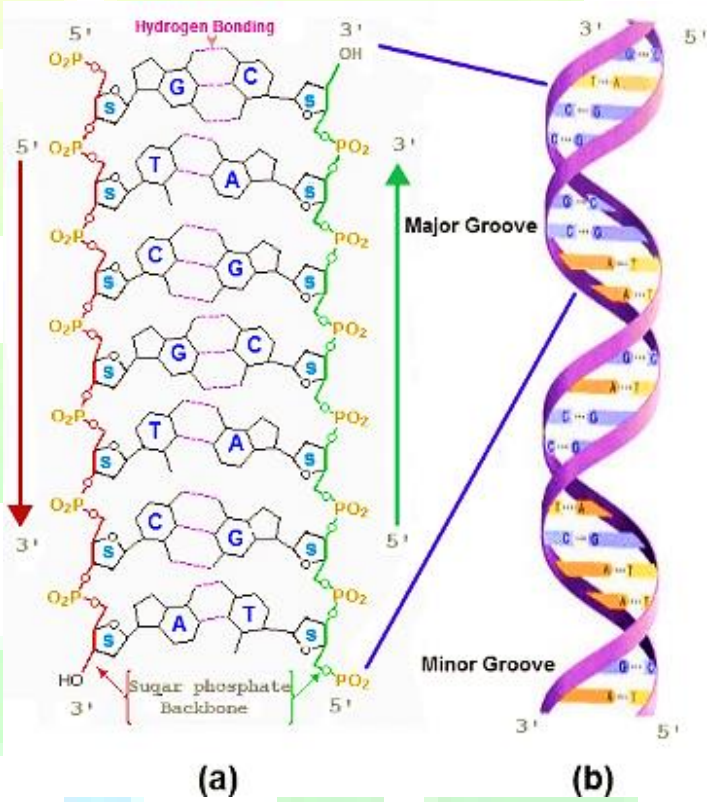


The Two Strands of the Double Helix Are Held Together by Base Pairing in an Anti-Parallel Orientation

The double helix is composed of two polynucleotide chains that are held together by weak, non-covalent bonds between pairs of bases. Adenine on one chain is always paired with thymine on the other chain and, likewise, guanine is always paired with cytosine. The two strands have the same helical geometry but base pairing holds them together with

Nucleic Acid: DNA & RNA

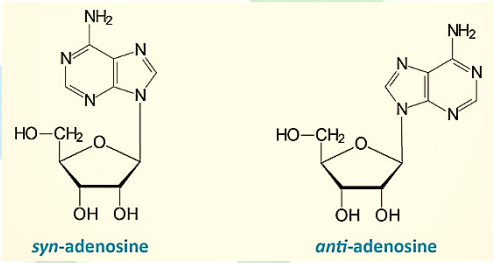




the opposite polarity. That is, the base at the 5` end of one strand is paired with the base at the 3` end of the other strand. The strands are said to have an anti-parallel orientation. This anti-parallel orientation is a stereochemical consequence of the way that adenine and thymine and guanine and cytosine pair with each together.

VIP!!!; ***Conformations of bases in Nucleosides***

- syn and anti-orientations
- Rotation around glycosidic bond has a slight barrier because of the H on the C2' of the ribose.



Brief History.....

1. 1869 - **Miescher** : Isolated nuclein from soiled bandages
2. 1902 - **Garrod**: Studied rare genetic disorder: Alkaptonuria; concluded that specific gene is associated with absence of a specific enzyme.
3. 1903 - **Sutton** : Chromosome structure
4. 1913 - **Morgan** : Gene mapping
5. 1926 - **Sumner** : Purified Urease; identified enzyme to be proteins
6. 1928 - **Griffith** : Transforming Principle : a chemical transferred from dead bacteria to living cells caused genetically converted strains ("transformation")
7. 1944 - **Avery**, **McCarty**, and **Macleod** : Identified Griffith's "transformation principle"
8. as DNA
9. 1947 - **Chargaff** : Base pairing
10. 1950's - **Franklin** : X-ray of DNA
11. 1953 - **Watson** and **Crick** : DNA double helix

Nucleic Acid: DNA & RNA

Properties of pyrimidines & purines

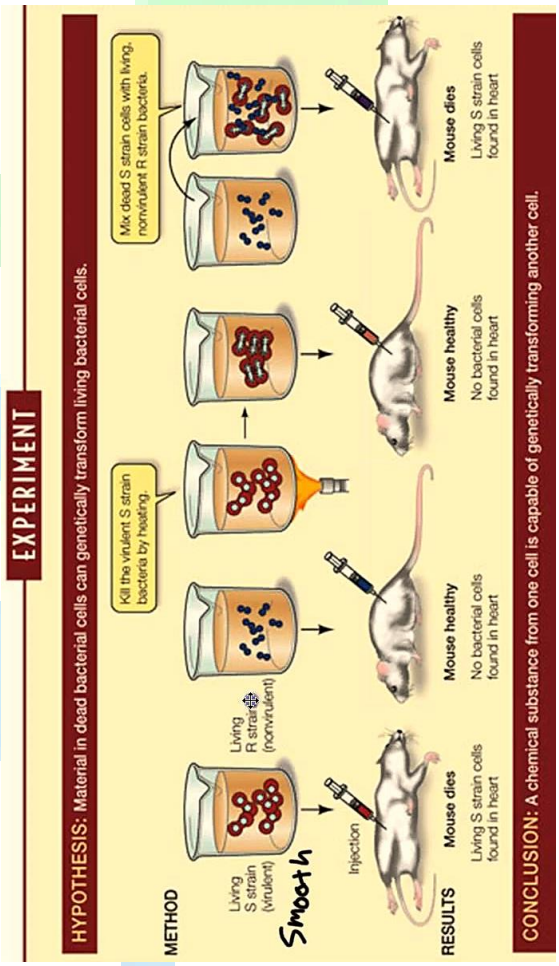
- Pyrimidine conformations are planar; purines are somewhat puckered
- Plain pyrimidines and purines have low solubility (not many polar bonds)
- Cytosine, Thymine, Uracil, Guanine, and Adenine are more soluble because they have many polar groups that are available for hydrogen bonding.
- Because they are aromatic, pyrimidines and purines can all absorb UV light.
- DNA & RNA concentration in a sample can be found by measuring UV absorbance.

Genes are DNA (Nucleic Acid)

Mendle's experiments in the late 19th century the showed that a gene is a discrete chemical entity (unit of heredity) that is capable of changing (mutable). At the beginning of the 20th century Sutton and Boveri realized that a gene is part of a chromosome. Subsequent experiments in the early to middle of the 20th century showed that chemical entity is a nucleic acid, most commonly DNA.

Transformation experiments

Griffith (1928) was a microbiologist working with **Avirulent** strains of Pneumococcus; infection of mice with such strains does not kill the mice. He showed that these avirulent strains could be **Transformed** into **Virulent Strains**, that is, infection with the transformed bacteria kills mice. **Smooth** (S) strains produce a capsular polysaccharide on their surface, which allow the Pneumococci to escape destruction by the mouse, and the infection proceeds, i.e. they are virulent. This polysaccaride can be **Type I, II, or III**. Virulent S strains can be killed by heat (i.e., sterilization) and, of course, the dead bacteria can no longer infect the mouse. The smooth strains can give rise to variants that do not produce the polysaccharide. Colonies of these bacteria have a **Rough** (**R**) appearance, but more importantly they are not immune to the mouse's defenses, and cannot mount a lethal infection, i.e. they are avirulent. When **Heat-killed S bacteria of type III are co-inoculated with live R (avirulent) bacteria derived from** type II, the mouse **dies** from the productive infection. This shows that the live R bacteria had **Acquired** something from the dead S bacteria that allowed the R bacteria to become virulent! The virulent bacteria recovered from the mixed infection now had a smooth phenotype, and made type III capsular polysaccharide. They had been **Transformed** from rough to smooth, from type II to type III. Transformation simply means that a character had been changed by some treatment of the organism. In 1944,



Nucleic Acid: DNA & RNA

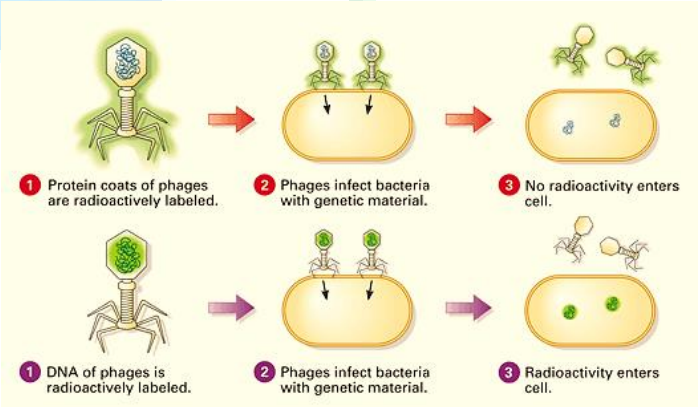
Avery, McCarty and Macleod showed that the **Transforming principle** is **DNA**. Earlier work from Friedrich Meicher (around 1890 to 1900) showed that chromosomes are nucleic acid and protein. Avery, McCarty and Macleod used biochemical fractionation of the bacteria to find out what chemical entity was capable of transforming avirulent R into virulent S bacteria, using the pneumococcus transfomation assay of Griffith. Given the chromosomal theory of inheritance, it was thought most likely that it would be protein or nucleic acid. At this time, nucleic acids like DNA were thought to be short oligonucleotides (four or five nucleotides long), functioning primarily in phosphate storage. Thus proteins, with their greater complexity, were the favored candidate for the transforming entity, at least before the experiment was done. Different biochemical fractions of the dead S bacteria were added to the live R bacteria before infection, testing to see which fraction transformed avirulent R into virulent S bacteria. The surprising result was that

DNA, not protein, was capable of transforming the bacteria.

The carbohydrate fraction did not transform, even though it is a polysaccharide that makes the bacteria smooth, or S. Neither did the protein fraction, even though most enzymes are proteins, and proteins are a major component of chromosomes. But the DNA fraction did transform, showing that it is the "transforming principle" or the chemical entity capable of changing the bacteria from rough to smooth. At the time it was thought that DNA did not have sufficient complexity to be the genetic material. However, we now know that native DNA is a very long polymer and these earlier ideas about DNA being very short were derived from work with highly degraded samples.

DNA, not protein, is passed on to progeny

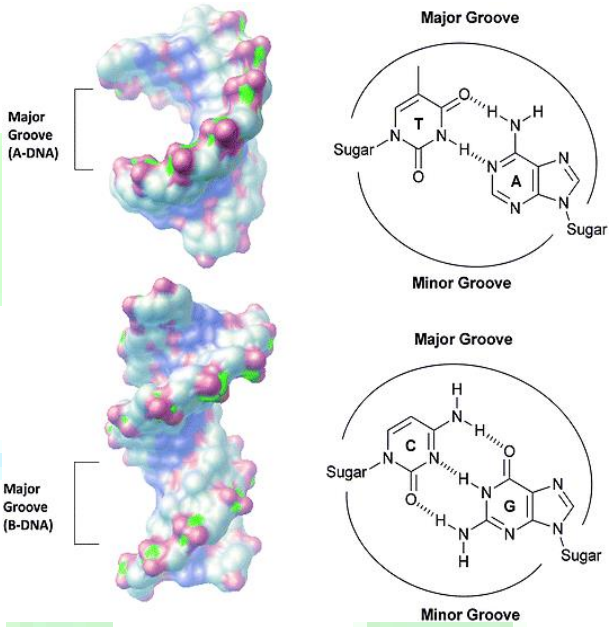
Hershey and Chase (1952) realized that they could use two new developments (at the time) to rigorously test the notion that DNA was the genetic material. Bacteriophage (or phage, or viruses that infect bacteria) had been isolated that would infect bacteria and lyse them, producing progeny phage. By introducing different radioactive elements into the protein and the DNA of the phage, they could determine which of these components was passed on to the progeny. Only genetic, inheritable material should have this property. (This was one of the earliest uses of radioactive. The proteins of T2 phage were labeled with 35S (e.g. in methionine and cysteine) and the DNA was labeled with 32P (in the sugar-phosphate backbone, as will be presented in the next section). The bacterium *E. coli* was then infected with the rabiolabeled phage. Shortly after the infection, Hershey and Chase knocked the phage coats off the bacteria by mechanical disruption in the Waring Blender, and monitored where the radioactivity went. Most of the 35S (80%) stayed with the phage coats, and most of the 32P (70%) stayed with the infected bacteria. After the bacteria lysed from the infection, the progeny phage were found to



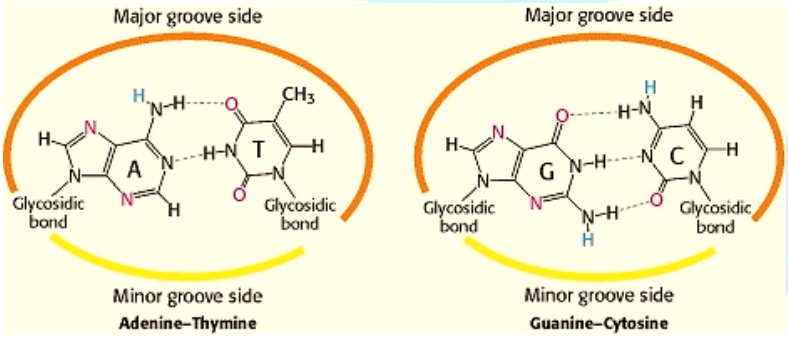
Nucleic Acid: DNA & RNA

carry about 30% of the input 32P but almost none (<1%) of the 35S. Thus the **DNA (32P) behaved like the genetic material** - it went into the infected cell and was found in the progeny phage. The protein (35S) largely stayed behind with the empty phage coats, and almost none appeared in the progeny.

The Major Groove is Rich in Chemical Information



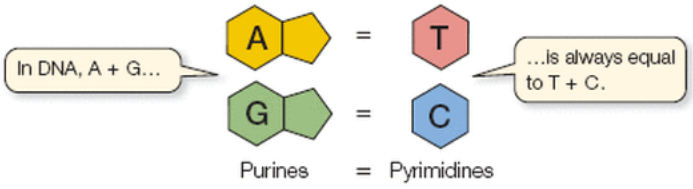
The edges of each base pair are exposed in the major and minor grooves, creating a pattern of hydrogen bond donors and acceptors and of van der Waals surfaces that identifies the base pair. The edge of an A:T base pair displays the following chemical groups in the following order in the major groove: a hydrogen bond acceptor (the N7 of adenine), a hydrogen bond donor (the exocyclic amino group on C6 of adenine), a hydrogen bond acceptor (the carbonyl group on C4 of thymine) and a bulky hydrophobic surface (the methyl group on C5 of thymine). Similarly, the edge of a G:C base pair displays the following groups in the major groove: a hydrogen bond acceptor (at N7 of guanine), a hydrogen bond acceptor (the carbonyl on C6 of guanine), a hydrogen bond donor (the exocyclic amino group on C4 of cytosine), a small non-polar hydrogen (the hydrogen at C5 of cytosine). Thus, there are characteristic patterns of hydrogen bonding and of overall shape that are exposed in the major groove that distinguish an A:T base pair from a G:C base pair, and, for that matter, A:T from T:A, and G:C from C:G. We can think of these features as a code in which **A** represents a **hydrogen bond acceptor, D** a **hydrogen bond donor, M** a **methyl group,** and **H** a **nonpolar hydrogen.** In such a code, **A D A M** in the major groove signifies an A:T base pair, and **A A D H** stands for a G:C base pair. Likewise, **M A D A** stands for a T:A base pair and **H D A A** is characteristic of a C:G base pair.



Chargaffs Rules

A pairs with T (ratio 1:1); G pairs with C (ratio 1:1), the ratios are random in RNA.

Stability of DNA



Nucleic Acid: DNA & RNA

DNA must be stable in order to store genetic information. What accounts for DNA's stability?

- **Aromatic Stacking**

Weak non covalent force caused by overlapping of p-orbitals; also called pi stacking. In DNA, aromatic stacking between the nucleotides contributes to its stability. The pyrimidine and purine bases, which are parallel to each other in DNA, participate in aromatic stacking due to the overlap of their p-orbitals.

- **Hydrogen Bonding**

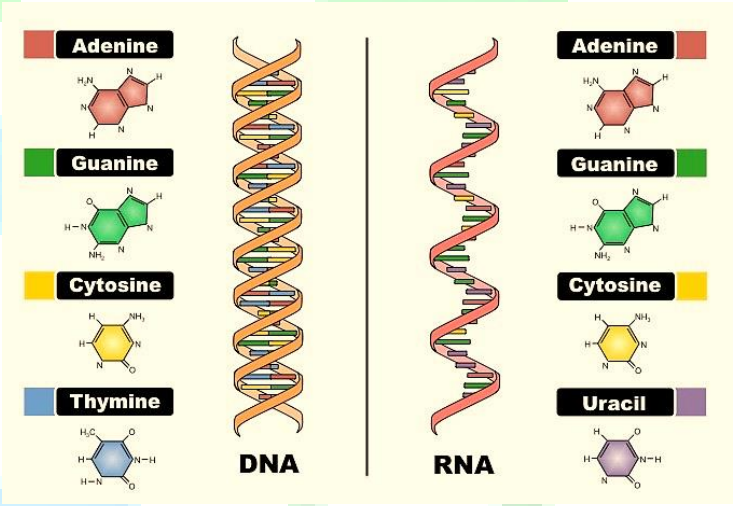
Millions of hydrogen bonds in DNA is the main structural feature that explains why DNA is stable. Hydrogen bonding is strong, but can easily be broken for DNA replication.

The Structures of RNA

RNA; base composition and structure

Base composition

Purified RNA molecules can be degraded by chemical means (alkali) or by enzymes (RNase) and the base composition determined in a manner similar to that described for DNA. One major difference between DNA and RNA is that in RNA there are a fairly large number of minor bases. Also, in the majority of cases there is no base equivalence, signifying that RNA molecules are usually single-stranded.



Structure

There are three major forms of RNA: **Ribosomal RNA** (**rRNA**), **transfer RNA** (**tRNA**) and **messenger RNA** (**mRNA**) and these exist in all the major life forms, together with other distinct RNA molecules which are not universal.

rRNA:

Accounts for about 80% of the total cellular RNA and is associated with protein to form the cytoplasmic particles known as ribosomes. As shown below the ribosome itself can be considered as two subunits, a large and a small, both of which are RNA protein associations with about 65% RNA: 35% protein. The RNA has high molecular mass and is metabolically stable. Transfer RNA is the next most abundant species and accounts for about 15% of the total RNA. These molecules are of much lower molecular mass than the rRNA and are also referred to as 4S RNA. These molecules function as adaptors for amino acids in the course of protein synthesis and many different tRNAs exist, each being specific for one amino acid.

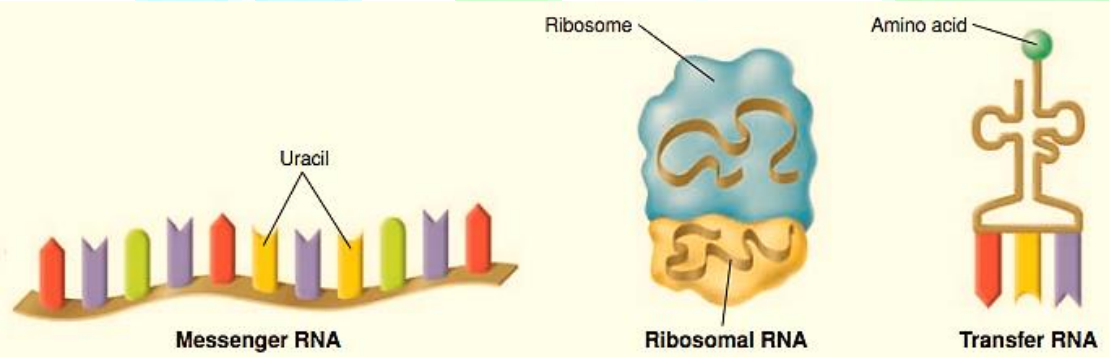
tRNA:

Nucleic Acid: DNA & RNA

The tRNA molecule is single-stranded but the chain folds back on itself in a very distinctive way (see diagram below) such that about 50–60% of the structure is base-paired.

mRNA:

Most of the remainder of the cell RNA (less than 5%) is accounted for by mRNA which, in eukaryotes, originates in the nucleus and migrates to the cytoplasm during protein synthesis. It is of high molecular mass and is metabolically very labile, i.e. easily broken down. mRNA is centrally involved in the transfer and expression of the genetic information and is responsible for the sequence of amino acids in each of the different proteins in the cell. Because the size of the messenger is variable, no S-value is associated with it. Note: the S-values refer to how fast a particular molecule sediments in an ultracentrifuge and the S refers to Svedberg units, the unit of measurement in such studies. Clearly this relates both to the mass and shape of a molecule in that an object of greater mass will move faster than one with lower mass but the same shape, and one with a small volume will move faster than one with a similar mass but larger volume (i.e. less dense).



Types of RNA The three main types of RNA are messenger RNA, ribosomal RNA, and transfer RNA. Ribosomal RNA is combined with proteins to form ribosomes.

RNAs which act as enzymes

In the late 1980s it was shown that certain RNA molecules could act as enzymes, capable of splicing out specific sequences of RNA either on itself or other RNA strands. The name given to such molecules was ribozyme and this work earned a Nobel Prize for one of the researchers (Tom Cech of the University of Colorado).Two of these ribozymes are now being clinically tested as potential treatment against HIV.