SECTION 1. NUCLEIC ACIDS: THE BIOMOLECULES OF HEREDITY

LEARNING OBJECTIVES:

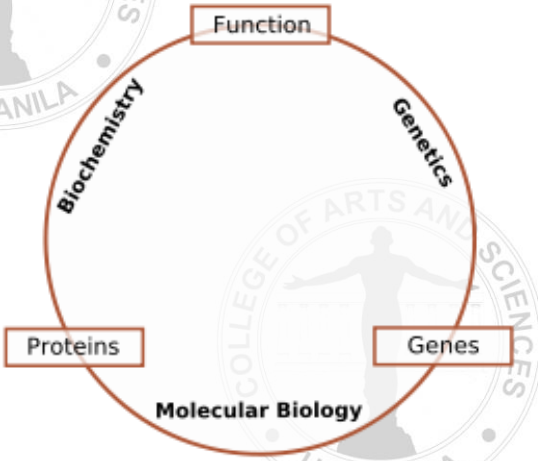
By the end of this section, you will be able to do the following:

- Identify the components of nucleic acids
- Explain the structural differences between DNA and RNA
- Describe DNA structure and how it is packed in the cell
- Explain the functions of different types of RNA

Genomics, genetics and biochemistry: what's the difference?

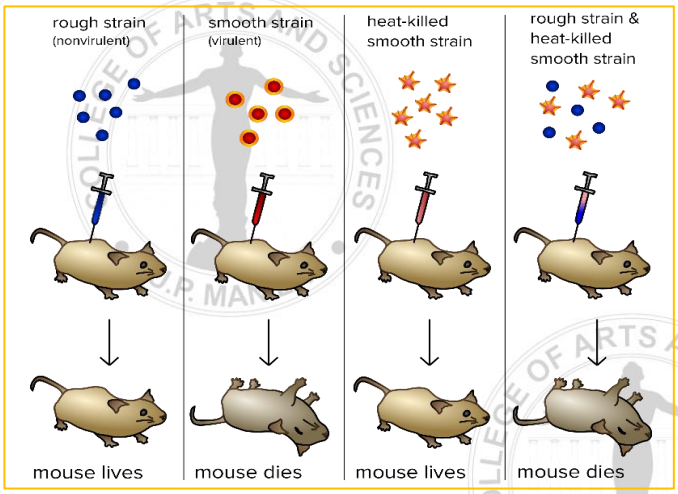
- **Genetics**: inherited phenotypes
- **Biochemistry**: chemistry of living organisms and/or cells
- **Molecular biology**: molecular basis of biological activity
- **Genomics**: study of **genomes**

There is not a hard line between these disciplines as there once was!



Functionally, DNA is the macromolecule which stores the genetic information of individuals. RNA transmits this information to make proteins, the molecule responsible for visible or observable traits. DNA determines the organism's genotype. This molecule then dictates the type and structure of proteins, the molecules responsible for the organism's phenotype.

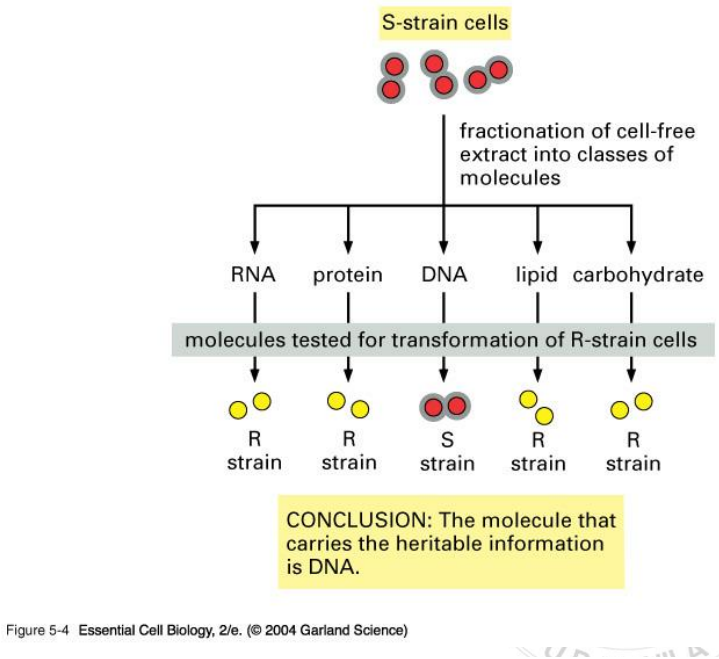
Our current understanding of DNA began with the discovery of nucleic acids followed by the development of the double-helix model. In the 1860s, Friedrich Miescher isolated phosphate-rich chemicals from white blood cells which he called nuclein. The first attempt to identify the actual hereditary substance was made by Frederick Griffith, using bacterial transformation in two strains of *Pneumococcus* bacteria. The R strain is non-pathogenic while the S strain is pathogenic. When Griffith injected living S strain into mice, they died from pneumonia, but live R strain did not. When he injected mice with the heat-killed S strain, they also survived. In a third set of experiments, a mixture of live R strain and heat-killed S strain were injected into mice, and—to his surprise—the mice died. Only the S strain of bacteria was recovered from the dead mouse, and when this isolated S strain was injected into fresh mice, the mice died. Griffith concluded that R-strain bacteria must have taken up a "transforming principle" from the heat-



killed S bacteria, which allowed them to "transform" into smooth-coated bacteria and become virulent.

Figure 1. Griffith's experiment with pneumococcus bacteria. 1st *Image modified from "Griffith experiment," by Madeleine Price Ball (CC0/public domain)*

The group of Avery, MacLeod and McCarty (1944) aimed to characterize this "transforming principle" by purifying the various biomolecules present in the S strain from the dead mice, and testing each for the ability to transform R strain into S. They found that when DNA was degraded, the resulting mixture was no longer able to transform the bacteria, whereas all of the other combinations were able to transform the bacteria. This led them to conclude that DNA was the transforming principle.

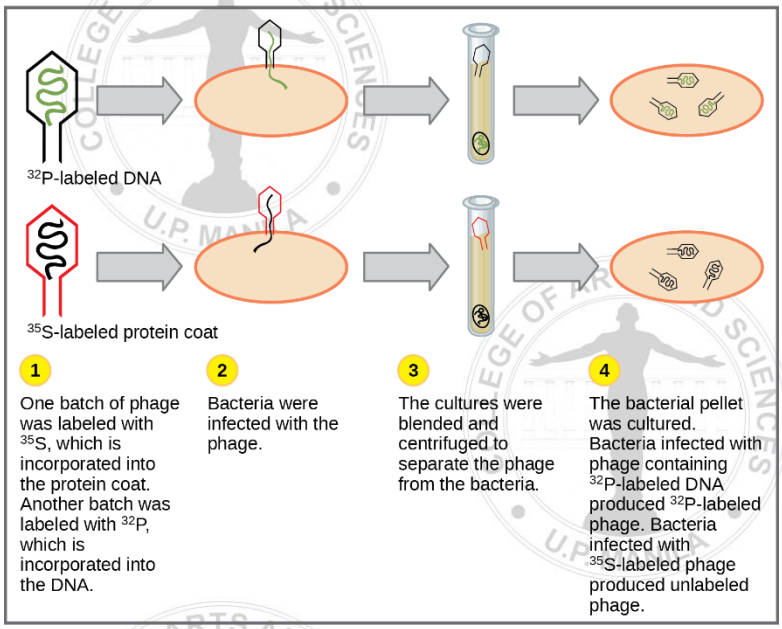


However, even with their results, DNA was still considered by many scientists to be too simple a molecule to carry biological information.

Figure 2. Avery, MacLeod and McCarty's experiment showed that DNA is the transforming principle.

The decisive experiment conducted by Martha Chase and Alfred Hershey in 1952 using bacteriophage provided confirmatory evidence that DNA is the substance that transmits hereditary information. Bacteriophage, or bacterial virus, were grown in one of two isotopic mediums in order to radioactively label a specific viral component

- Viruses grown in radioactive *sulfur* (^{35}S) had radiolabelled *proteins* (sulfur is present in proteins but not DNA)
- Viruses grown in radioactive *phosphorus* (^{32}P) had radiolabeled *DNA* (phosphorus is present in DNA but not proteins)



The viruses were allowed to infect *E. coli*, and then the virus and bacteria were separated via centrifugation. The larger bacteria formed a solid pellet while the smaller viruses remained in the supernatant. The bacterial pellet was found to be radioactive when infected by the ^{32}P–viruses but not the ^{35}S–viruses. This demonstrated that **DNA**, not protein, was the genetic material because DNA was transferred to the bacteria.

Fig. 3. In Hershey and Chase's experiments, bacteria were infected with phage radiolabeled with either ^{35}S, which labels protein, or ^{32}P, which labels DNA. Only ^{32}P entered the bacterial cells, indicating that DNA is the genetic material. From https://openstax.org/books/biology-2e/pages/14-1-historical-basis-of-modern-understanding

STRUCTURE OF NUCLEIC ACIDS

Nucleic acids are made up of polymers of nucleotides linked together by phosphodiester bonds. There are two types of nucleic acids:

1. deoxyribonucleic acid (DNA)
2. ribonucleic acid (RNA)

DNA and RNA are similar in that the repeating unit of both is the nucleotide. Each nt contains a sugar, a phosphate and a nitrogen base (Fig. 1). However, there are some functional as well as structural differences between these two types of nucleic acids.

DNA and RNA: Monomeric Units

- DNA differs from RNA in the type of sugar as well as the base.
  - ❑ DNA contains deoxyribose while RNA contains ribose.
    - ➢ The names of deoxyribonucleic acid (DNA) and ribonucleic (RNA) acid come from their sugar component.
  - ❑ DNA bases: A, G, C, T
  - ❑ RNA bases: A, G, C, U
- The nucleotide base ("sidechain") is connected to the C1' of the sugar ("mainchain") by an N-linked glycosidic bond.

- The carbon atoms on the sugar are numbered 1' to 5'. The primes distinguish the atoms on the sugar from those on the base.
- The phosphate group is linked by a phosphodiester bond.
- DNA is also a double stranded molecule while RNA is usually single stranded.

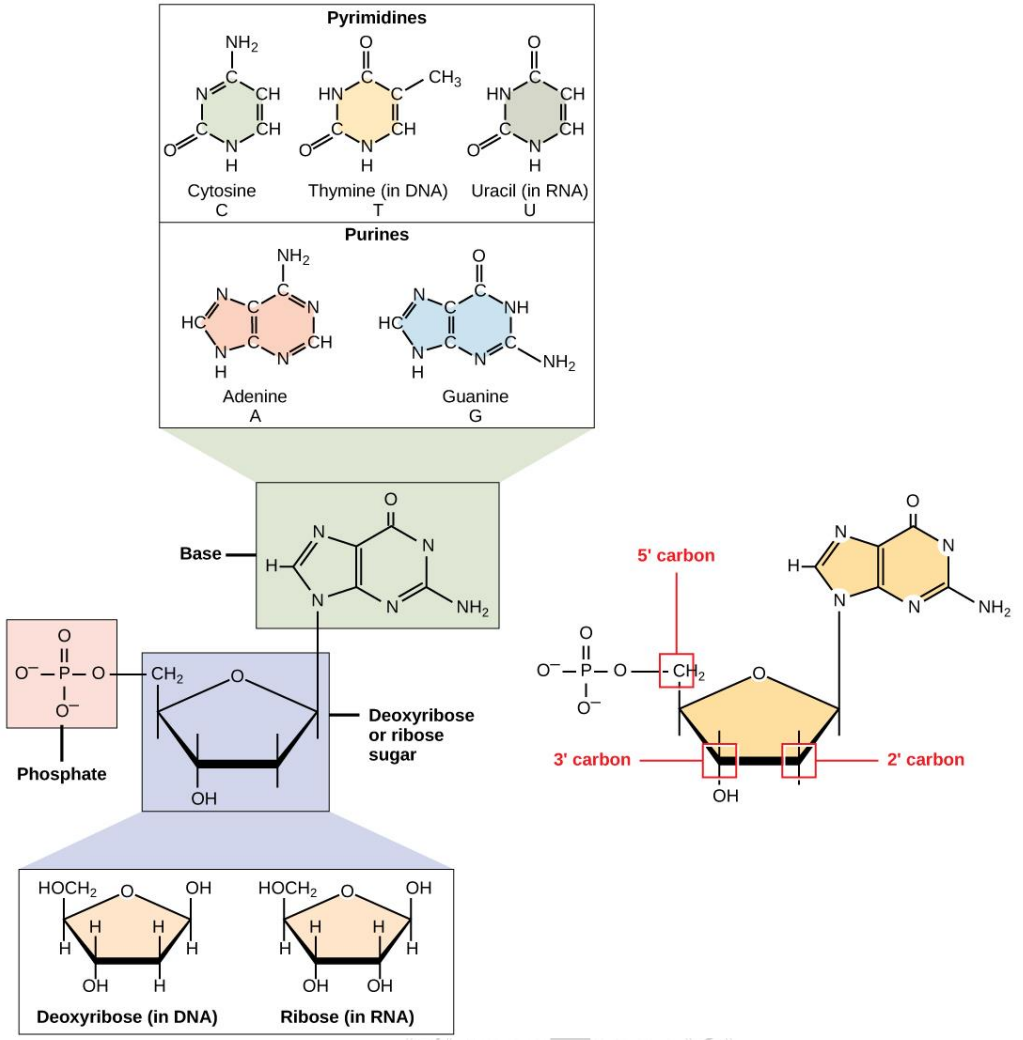


Figure 4. Three components comprise a nucleotide: a nitrogenous base, a pentose sugar, and one or more phosphate groups. Carbon residues in the pentose are numbered 1′ through 5′ (the prime distinguishes these residues from those in the base, which are numbered without using a prime notation). The base is attached to the ribose's 1′ position, and the phosphate is attached to the 5′ position. When a polynucleotide forms, the incoming nucleotide's 5′ phosphate attaches to the 3′ hydroxyl group at the end of the growing chain. Two types of pentose are in nucleotides, deoxyribose (found in DNA) and ribose (found in RNA). Deoxyribose is similar in structure to ribose, but it has an H instead of an OH at the 2′ position. We can divide bases into two categories: purines and pyrimidines. Purines have a double ring structure, and pyrimidines have a single ring.
From https://openstax.org/books/biology-2e/pages/3-5-nucleic-acids

Nucleic acid nomenclature:

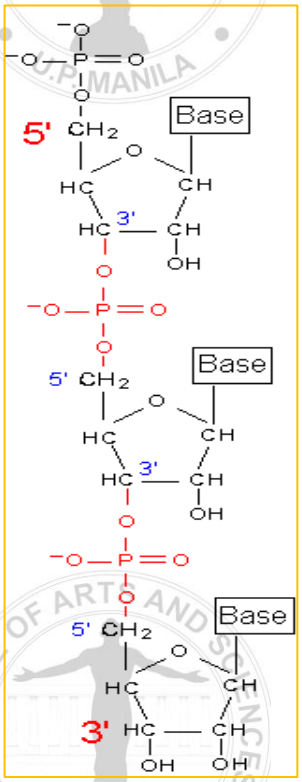
- Base + sugar = nucleoside
- Base + sugar + phosphate = nucleotide

Base	Nucleoside (= base + pentose)		Nucleotide (= nucleoside + phosphate)		
	Ribo-nucleoside	Deoxyribo-nucleoside	NMP dNMP	NDP dNDP	NTP dNTP
Purines					
Adenine	Adenosine	Deoxyadenosine	AMP dAMP	ADP dADP	ATP dATP
Guanine	Guanosine	Deoxyguanosine	GMP dGMP	GDP dGDP	GTP dGTP
Pyrimidines					
Cytosine	Cytidine	Deoxycytidine	CMP dCMP	CDP dCDP	CTP dCTP
Thymine	Thymidine	Deoxythymidine	TMP dTMP	TDP dTDP	TTP dTTP
Uracil	Uridine	Deoxyuridine	UMP dUMP	UDP dUDP	UTP dUTP

Table 1. Cellular nucleosides and nucleosides. From https://www.web-books.com/MoBio/Free/Ch3A4.htm

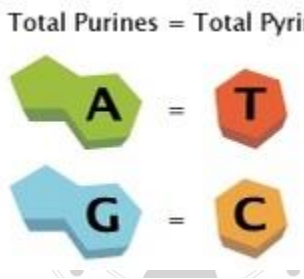
Primary structure: polynucleotide polymers

- DNA and RNA polynucleotides are connected by phosphates between the 3' and 5' positions of the sugars.
- The phosphates are always ionized (pKa~1), therefore nucleic acids are polyanions.
- Sequence of nucleotide bases is written in the 5'-3' direction.



DNA Secondary structure: double helix

In the 1950s, Francis Crick and James Watson worked together to determine the structure of DNA. Through model building, as well as X – ray diffraction data obtained from Rosalind Franklin, they discovered that the DNA molecule is a double helix two polynucleotide strands. They also realized the importance of Chargaff's rules: the two strands are held together by specific base-pairing. Specifically, T forms two H-bonds with A, and G forms three H-bonds with C. Due to specific base pairing, the two strands are



complementary to each other. Hence, the nucleotide sequence of one strand determines the sequence of its complementary strand.

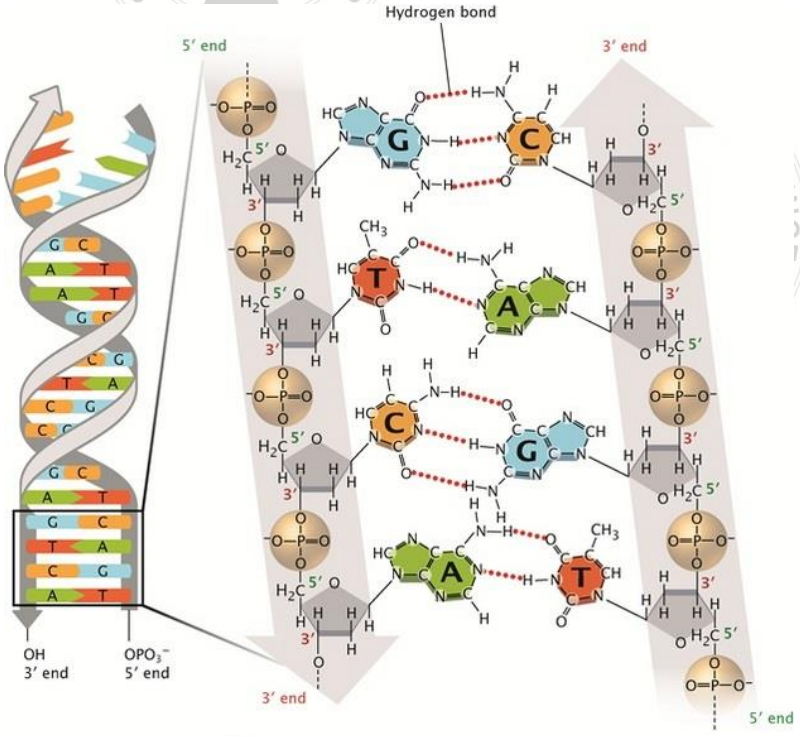


Figure 5. Base pairing in DNA. Two hydrogen bonds connect T to A; three hydrogen bonds connect G to C. The sugar-phosphate backbones (grey) run anti-parallel to each other, so that the 3′ and 5′ ends of the two strands are aligned. © 2013 Nature Education

The features of the double helical structure of DNA reported by Watson and Crick, which is now referred to as B-DNA, and is the most common conformation in most living cells, are as follows:

- ❑ The helix is right-handed; the chains are antiparallel.
- ❑ The bases are almost perpendicular to the helix axis.
- ❑ Rise 3.4 Å/base pair; 10 bp/turn
- ❑ Spacing between the two phosphate chains in the direction of the helix not uniform. There is a wider groove (major groove) and a narrower groove (minor groove).

I1962, James Watson, Francis Crick, and Maurice Wilkins, in whose laboratory Franklin worked, were awarded the Nobel Prize in Medicine. Unfortunately, by then Franklin had died, and Nobel prizes are not awarded posthumously.

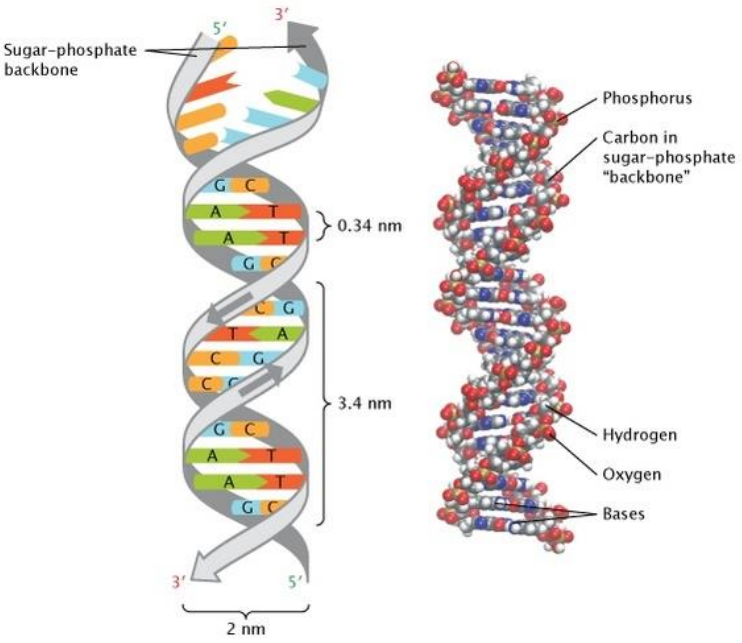


Figure 6: The double-helical structure of DNA. The 3-dimensional double helix structure of DNA, correctly elucidated by James Watson and Francis Crick. Complementary bases are held together as a pair by hydrogen bonds. **© 2013 Nature Education**

Other DNA forms exist

- **A form**
  - ❑ Right-handed
  - ❑ favored in solutions relatively devoid of water; *eg,* high salt
- **Z form**
  - ❑ Left-handed
  - ❑ So-called because its bases seem to zigzag
  - ❑ Found in DNA molecules with alternating G-C sequences in alcohol or high salt solution

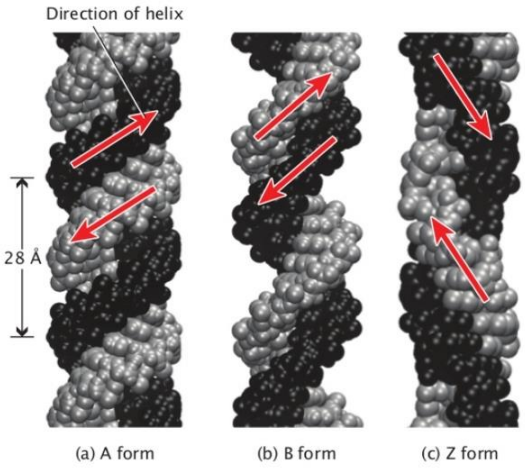


Figure 7. Three different conformations of the DNA double helix. (A) A-DNA is a short, wide, right-handed helix. (B) B-DNA, the structure proposed by Watson and Crick, is the most common conformation in most living cells. (C) Z-DNA, unlike A- and B-DNA, is a left-handed helix. © 2014 Nature Education Adapted from Pierce, Benjamin. *Genetics: A Conceptual Approach*, 2nd ed.

EXERCISE 1. What is the base % composition of a DNA molecule which is 18% A?
ANSWER: Since normally, A is complementary to T, it follows that T is also 18%. The remaining 64% is G + C. But C is also complementary to G in the DNA double helix, so that each of C and G would constitute 32 % of the composition.

Higher levels of DNA structure in eukaryotes

The human genome consists of

1. nuclear genome
   - comprises ~3 x10^9 nts, divided into 23 pairs of chromosomes

2. mitochondrial genome
   - circular DNA molecule of ~16,000 nts
   - multiple copies present in mitochondria

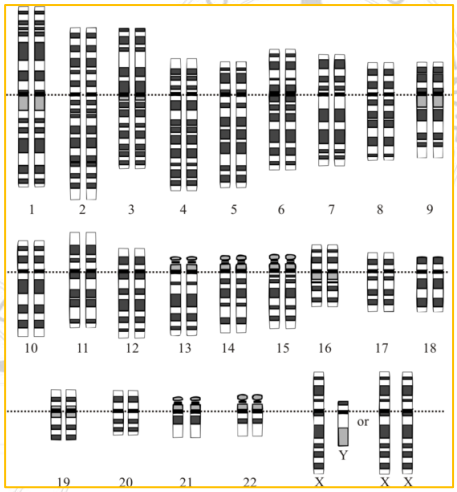


Fig. 8. The human nuclear genome is found in 23 pairs of chromosomes. From Wikimedia Commons, the free media repository

The nuclear genome of eukaryotic organisms is packaged with histone proteins to create a compacted structure called a nucleosome. Nucleosomes help to supercoil the DNA, resulting in a greatly compacted structure that allows chromosomes to fit in the nucleus. Supercoiling also helps to protect the DNA from damage and also allows chromosomes to be mobile during mitosis and meiosis.

Histones and chromatin

- Eukaryotic double stranded is complexed with histone proteins to form a complex called a *nucleosome*
  - *histone octamer* forms a protein core around which ds DNA is wound
  - nucleosome core particle = eight histone proteins (2 each of H2A, H2B, H3, and H4) and and 146 nt ds DNA
  - H**istones** have a high proportion of **positively charged amino** acids, which allows tight binding to negatively charged DNA

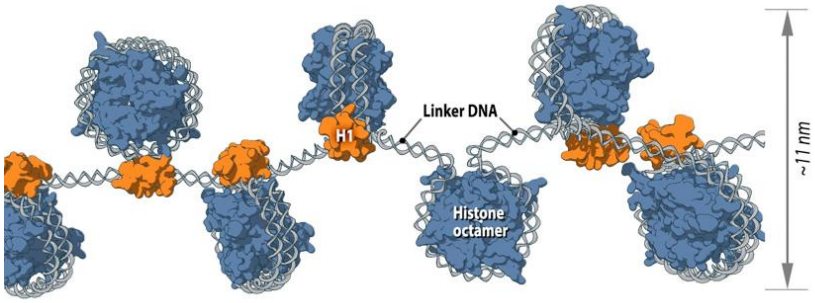


Figure 9. Each nucleosome consists of histone octamer core, assembled from the histones H2A, H2B, H3 and H4 (or other histone variants in some cases) and a segment of DNA that wraps around the histone core. Adjacent nucleosomes are connected via "linker DNA". From https://www.mechanobio.info/genome-regulation/what-are-nucleosomes/

- Nucleosomes are linked by an additional histone protein (H1 histone) to form a string of chromatosomes
- These then coil to form a solenoid structure (~6 chromatosomes per turn) which is condensed to form a 30 nm fiber
- The fibers form loops, which are compressed and folded around a protein scaffold to form chromatin
- Chromatin will then supercoil during cell division to form chromosomes that are visible when stained under microscope

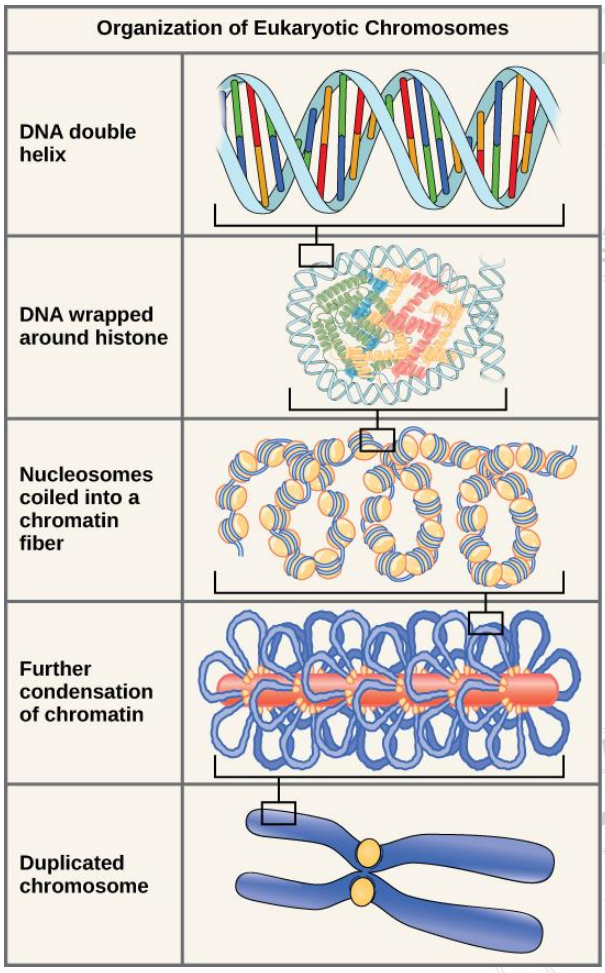


Figure 10. These figures illustrate the compaction of the eukaryotic chromosome. From https://openstax.org/books/biology-2e/pages/14-2-dna-structure-and-sequencing

Cellular functions of RNA

Cells access the information stored in DNA to direct the synthesis of proteins through the process of **translation**. The three main types of RNA, **messenger RNA (mRNA)**, **ribosomal RNA (rRNA)**, and **transfer RNA (tRNA),** are directly involved in protein synthesis.

Table 1. Structure and Function of the main types of RNA			
	mRNA	**rRNA**	**tRNA**
Structure	Short, unstable, single-stranded **RNA** correspon ding to a gene encoded within DNA	Longer, stable RNA molecules composing 60% of ribosome's mass	Short (70-90 nucleotides), stable RNA with extensive intramolecular base pairing; contains an amino acid binding site and an mRNA binding site
Function	Serves as intermediary between **DNA** and protein; used by ribosome to direct synthesis of protein it encodes	Ensures the proper alignment of mRNA, tRNA, and ribosome during protein synthesis; catalyzes **peptide bond** formation between amino acids	Carries the correct amino acid to the site of protein synthesis in the ribosome

From https://courses.lumenlearning.com/microbiology/chapter/structure-and-function-of-rna/

- Messenger RNA (**mRNA**) serves as the intermediary between DNA and the synthesis of protein products during translation.
- Ribosomal RNA (**rRNA**) is a major constituent of ribosomes. It ensures the proper alignment of the mRNA and the ribosomes during protein synthesis and catalyzes the formation of the peptide bonds between two amino acids during protein synthesis.
- Transfer RNA (**tRNA**) is a small type of stable RNA that carries an amino acid to the corresponding site of protein synthesis in the ribosome. The base pairing between the tRNA and mRNA allows for the correct amino acid to be inserted in the polypeptide chain being synthesized.

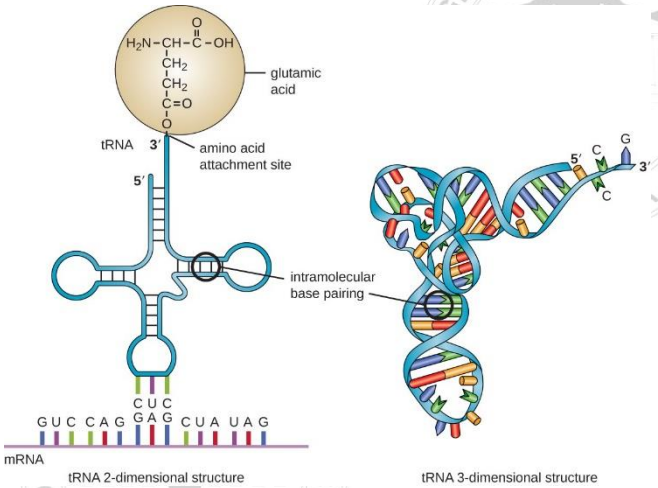


Fig. 11. A generalized illustration of how mRNA and tRNA are used in protein synthesis within a cell.

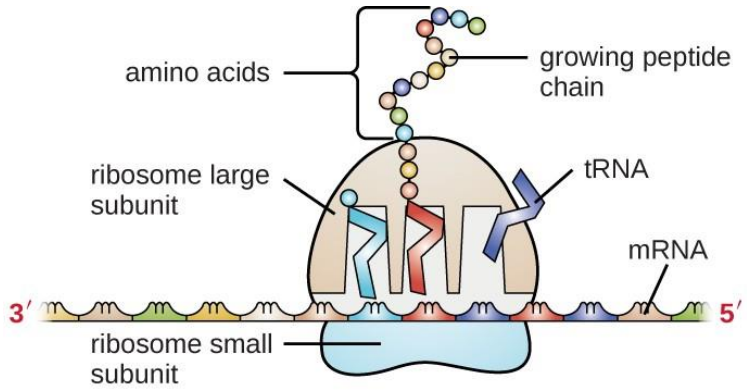


Fig. 12. A tRNA molecule exhibits significant intracellular basepairing, giving it a characteristic three – dimensional shape.
From

https://courses.lumenlearning.com/microbiology/chapter/structure-and-function-of-rna/

Aside from these three major types of RNA, several others have been discovered in recent years. These include:

- *Small nuclear RNAs (snRNA; 150 nt):*
  snRNAs are associated with a group of specific proteins to form the complexes referred to as "small nuclear ribonucleoproteins (snRNP)" in the nucleus. Their primary function is to process the precursor mRNA (pre-mRNA).

- *Small nucleolar RNAs (snoRNA; 60-300 nt):*
  snoRNAs are components of small nucleolar ribonucleoproteins (snoRNPs), which are complexes that are responsible for sequence-specific nucleotide modification.

- *Piwi-interacting RNAs (piRNA; 24-30 nt):*
  Piwi-interacting RNAs bind the PIWI subfamily proteins that are involved in maintaining genome stability in germline cells. They also play a role in gametogenesis.

- *MicroRNAs (miRNA; 21-22 nt):*
  miRNAs are small ncRNAs of ~22 nucleotides (nt) and the most widely studied class of ncRNAs. These RNA species mediate post-transcriptional gene silencing through RNA interference (RNAi), where an effector complex of miRNA and enzymes can target complementary mRNA by blocking the mRNA from being translated or accelerating its degradation. In human, miRNAs are estimated to regulate the translation of >60% of protein-coding genes.

- *Long noncoding RNAs (lncRNA):*
  Long noncoding RNAs are a heterogeneous group of non-coding transcripts larger than 200 nt in size and make up the largest portion of the mammalian non-coding transcriptome. It is estimated that more than 8,000 lncRNAs encoded in the human genome. lncRNAs are essential in many physiological processes. To date, various mechanisms of gene regulation by some lncRNAs have been reported. Most are still of unknown function.

Although RNA is not used for long – term genetic information in prokaryotic and eukaryotic cells, many viruses use RNA as their genetic material. An example is SARS-COV2, the virus which causes COVID-19.

REFERENCES

Stryer 8th ed.

https://www.futurelearn.com/courses/translational-research/0/steps/14201
https://courses.lumenlearning.com/microbiology/chapter/structure-and-function-of-rna/
https://openstax.org/books/biology-2e/pages/14-introduction
https://microbenotes.com/dna-structure-properties-types-and-functions/

ADDITIONAL VIDEO RESOURCES

- The twisting tale of DNA - Judith Hauck
  https://www.youtube.com/watch?v=0_b80fHmuWw
- From DNA to protein - 3D
  https://www.youtube.com/watch?v=gG7uCskUOrA
- The Structure of DNA
  https://www.youtube.com/watch?v=o_-6JXLYS-k