DNA and Chromatin

The Discovery of DNA Structure

DNA (deoxyribonucleic acid) is a molecule that encodes the genetic information of living organisms. The structure of this molecule was first described by James Watson and Francis Crick. Using X-ray crystallography data collected by Maurice Wilkins and Rosalind Franklin (see the famous Photo 51 below), Watson and Crick proposed that DNA possessed a double helical structure. This accurate description of DNA’s structure won Watson and Crick, along with Wilkins, the Nobel Prize in physiology or medicine in 1962.

Photo 51

X-ray crystallography
**Nucleic Acids**

Nucleic acids are macromolecules that function in encoding, transmitting, and expressing the genetic material of organisms. There are two types of nucleic acids: **deoxyribonucleic acid (DNA)** and **ribonucleic acid (RNA)**. Our genes are made up of DNA, which encodes the instructions for its own replication as well as for RNA synthesis. RNA, in turn, serves as a template for the synthesis of proteins. This flow of genetic material can be summed up as: **DNA → RNA → Protein**.

**Nucleotides** are the monomers of nucleic acids. The following three components make up a nucleotide:

1. **A nitrogenous base**
   - **Pyrimidines**: One class of nitrogenous bases. A six-membered ring.
     - Cytosine (C), thymine (T), and uracil (U) are pyrimidine bases.
   - **Purines**: The other class of nitrogenous bases. A six-membered ring fused to a five-membered ring.
     - Adenine (A) and guanine (G) are purine bases.

2. **A five-carbon sugar**
   - **Ribose**: The five-carbon sugar in RNA.
   - **Deoxyribose**: The five-carbon sugar in DNA. Similar to ribose but lacks an oxygen atom on the second carbon atom in the ring. As a result, deoxyribose is more chemically stable than ribose.

3. **One or more phosphate groups**

Note: A nucleotide without any phosphate groups is called a **nucleoside**.

---

**Nitrogenous Bases**

- **Purines**: Adenine, Guanine
- **Pyrimidines**: Thymine, Cytosine, Uracil
**Polynucleotides (Nucleic Acids):**

Nucleotides are linked together via a **phosphodiester linkage**, wherein a phosphate group links the five-carbon sugars of two nucleotides together. This forms a polynucleotide, a molecule formed by multiple nucleotides linked together. The alternating sugar and phosphate molecules of the nucleotide chain form, what we call, a **sugar-phosphate backbone**.

There are two ends of a polynucleotide:

- **5' end** = The end with a *phosphate* group attached to carbon 5 of the sugar
- **3' end** = The end with a *hydroxyl* group attached to carbon 3 of the sugar

An example of a textual representation of a nucleotide polymer is: 5'-ATGACCT-3'. This *oligonucleotide* is listed starting from the 5' end, followed by the sequence of nitrogenous bases, ending with the 3' end.

![Sugar-phosphate backbone of polynucleotides]

Both DNA and RNA are polynucleotides, but differ in several respects. As stated previously, DNA contains deoxyribose whereas RNA contains ribose. DNA molecules contain two polynucleotide strands that together form a double helix. On the other hand, RNA is *usually* single stranded, however, RNA does have the ability to form double helices like DNA. Finally, DNA contains the nitrogenous base thymine but not uracil while RNA has uracil but not thymine. The table below summarizes the differences between DNA and RNA.

<table>
<thead>
<tr>
<th></th>
<th>DNA</th>
<th>RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pentose:</td>
<td>Deoxyribose</td>
<td>Ribose</td>
</tr>
<tr>
<td>Structure:</td>
<td>Double-stranded helix</td>
<td>Usually single-stranded</td>
</tr>
</tbody>
</table>
**Base pairing:**
Nitrogenous bases in DNA and RNA pair with each other as follows:
- Adenine (A) always pairs with thymine (T) or uracil (U)
- Cytosine (C) always pairs with guanine (G)

What is the complementary DNA sequence to 5'-ATGACCT-3' ? Well, now that we know about base pairing, it should be easy to see that the answer is 3'-TACTGGA-5'.

The pairing between nitrogenous bases involves hydrogen bonds. In DNA, base pairing is responsible for holding together the double helix.

Also worth mentioning is that the GC base pair is held together by three hydrogen bonds whereas the AT base pair is held together by two hydrogen bonds. As a result, DNA with a higher GC-content have higher melting temperatures (i.e. more heat is required to denature the DNA).

**Adenosine triphosphate (ATP)** is a molecule that provides energy for many metabolic processes that occur within a cell. It is known as the molecular currency within cells. It stores chemical energy in its energy-rich phosphate groups. When a phosphate group is removed via hydrolysis, ATP becomes ADP (adenosine diphosphate) and large amounts of energy is released. ATP is a nucleotide (or, more specifically, a nucleoside triphosphate), containing adenine, three phosphate groups, and a five-carbon sugar that is usually ribose (although deoxyribose is found in certain types of ATP). Below shows the chemical structure of ATP.

![DNA's double helix:](image1)

![Base pairings (A=T and C≡G):](image2)
More About DNA Structure

Right handed helix vs. Left handed helix:

Before we discuss the structure of DNA any further, it is important that we distinguish between right handed helices and left handed helices. Imagine a three dimensional DNA model: If you screw the DNA clockwise and it moves away from you, the helix is right-handed. If you screw the DNA clockwise and it moves towards you, the helix is left-handed. The figure below contrasts a left-handed helix with a right-handed helix.

Different Forms of DNA:

Three forms of DNA are known to exist: A form, B form, and Z form. See figure below.

- **B form**: The most common form of DNA (and the form captured in photo 51). In this form, DNA exists as a right-handed helix with about 10 nucleotides per turn.
- **A form**: This form prevails in dehydrated samples of DNA. DNA-RNA duplexes and RNA-RNA duplexes also assume a form rather similar to the A form. In the A form, DNA is still a right-handed helix, however, unlike B-DNA, the base pairs are not perpendicular to the helix axis. Also, the helix of A-DNA is thicker and more compact than that of B-DNA.
- **Z form**: This form of DNA is a left-handed helix and its sugar-phosphate backbone assumes a zig-zag pattern. This form of DNA is often formed by alternating purine-pyrimidine sequence (e.g. GCGCGCGCGCGC).
The B Form of DNA:

The figure below shows some of the dimensions of the B form of DNA. As mentioned previously, there are around 10 nucleotides per turn. Further, B-DNA contains major grooves (~2.2 nm wide) and minor grooves (~1.2 nm wide).

Chromatin

In this section, we will discuss how DNA is organized in the cell. We will restrict our discussion to eukaryotic cells (prokaryotic cells, e.g. bacteria, have a different organization of their DNA).

DNA is found in the nucleus. But all the DNA in a single human cell, if stretched end-to-end, would measure about 2 meters. The nucleus of a human cell is only around 6 micrometers. So DNA must be packaged somehow! How does this packaging work? Proteins fold and package DNA into compact structures. The complex of DNA and its associated proteins is called chromatin.

Nucleosomes:

DNA wraps around proteins called histones, which contain many basic (positively charged) amino acid residues. Two of each of the core histones: H2A, H2B, H3, and H4, assemble into an octamer. 146 base pairs of DNA is wrapped ~1.7 times around a histone octamer. There are approximately 50 base pairs of DNA linking one histone octamer to another; such DNA is known as linker DNA. A nucleosome, which consists of a histone octamer with its associated DNA plus an adjacent linker DNA, is the basic unit of chromatin organization. Multiple nucleosomes can form a basic “beads-on-a-string” structure.
The 30-nm fiber:

In actuality, chromatin rarely adopts the extended “beads-on-a-string” form within a living cell. Instead, nucleosomes are packed on top of one another to generate a fiber with a diameter around 30 nanometers. One mechanism that contributes to the formation of the 30-nm fiber form is the presence of action of histone H1. Histone H1 is not a core histone (i.e. it is not part of the histone octamer) but rather, it is a linker histone. A single H1 molecule binds to each nucleosome, changing the path of DNA as it exits from the nucleosome.

Higher order chromatin structure:

Chromatin can be condensed even further through a variety of mechanisms. Remodeling chromatin to make it more condensed or less condensed is one method or epigenetically regulating the expression of genes (epigenetics is the heritable change in gene expression without a change in DNA sequence). Chromatin that is lightly packed is known as euchromatin. Chromatin that is tightly packed is known
as heterochromatin. Euchromatin, due to its looser structure, is more likely to be under active transcription than heterochromatin.

During the early stages (i.e. prophase) of mitosis or meiosis, chromatin condenses a lot to form very compact structures, called chromosomes. Because mitotic chromatin is so condensed, it is transcriptionally inactive.
Regulation of chromatin structure:

The tail region of histones is subject to many modifications. One such modification is the acetylation or deacetylation of lysine residues on histone tails.

The addition of an acetyl group to a lysine amino acid removes the positive charge from the lysine. As a result, DNA becomes less tightly bound to the histone proteins (since DNA is negatively charged) and the chromatin structure becomes looser (euchromatin).

Amino acids on histone tails can also be methylated or demethylated.

As shown above, adding methyl moieties to lysine does not change the charge of lysine. However, it can still result in a change of chromatin architecture. For instance, H3K4me3 (three methyl groups on the fourth lysine of histone H3) is a hallmark of transcriptional activity (loose chromatin structure) and
demethylating methylated H3K4 can result in transcriptional repression. According to the **histone code** hypothesis, “code reader” protein complexes recognize specific histone tail modifications (e.g. H3K4me3) and recruit other proteins to initiate downstream effects (e.g. gene expression, gene silencing, etc.).

Credit: