Molecular Biology

Done By:
Khader Nassar, Rania Tayseer, Ayah Saffarini, Sajjad Al-Nauimi

Corrected By:
Sajjad Al-Nauimi
Molecular Biology (1)
Structure of nucleic acids

Prof. Mamoun Ahram
Second semester, 2020-2021
Course resources

• Lectures
Outline

- Nucleic acid structure
- Basic techniques
- The human genome
- DNA replication
- DNA mutations
- DNA repair
- Transcription
  - Regulation of transcription in prokaryotes
  - Regulation of transcription in eukaryotes
- Analysis of gene expression
- Translation and its regulation
Resources

• This lecture
• Cooper, Ch. 2, pp. 54-56, Ch. 4, 116-118, Ch. 6, pp.203-208
What is molecular biology?

Central dogma of molecular biology:

- **Replication:** DNA is used to make RNA and RNA is used to make proteins. Replication: DNA can make a copy of itself, and it is catalyzed by DNA Polymerase.

- **Transcription:** You have a DNA molecule, and this DNA molecule is used to make an RNA molecule and it is catalyzed by RNA Polymerase.

- **Translation:** RNA is used to synthesize proteins, and this involves ribosomes.

- Things have gotten more complex as we understand more of how nature functions and works (You can make DNA out of RNA, and you can make RNA out of DNA).

In fact, molecular biology is biochemistry. It talks about the different reactions, the biochemical structures and the information in these molecules. But, it's specifically related to DNA and RNA. It is NOT genetics. (There is an overlap between molecular biology and genetics.)

Central dogma of molecular biology:

- Basically, DNA is used to make RNA and RNA is used to make proteins.

- Replication: DNA can make a copy of itself, and it is catalyzed by DNA Polymerase.

Genetics deals with patterns of inheritance of phenotypes and genotypes. It deals with inheritance, heredity, and chromosomal structures rather than DNA and RNA.

Molecular biology deals with biochemical reactions that take place in DNA and RNA.
Nucleic acids

- There are types
  - Deoxyribonucleic acid (DNA)
  - Ribonucleic acid (RNA)
- The primary structure of nucleic acids is linear polymers of nucleotides (monomers) bound to each other via phosphodiester bonds.
- DNA is coiled and can be associated with proteins forming chromosomes.

These are known as monomers and are connected to each other to form this large polymer (DNA polymer) in this case.
Nucleotide: made of 3 components. A sugar molecule that is linked with a nitrogenous base on one carbon and on a different carbon, its linked with phosphate.

Pay attention to the numbering of carbons, this symbol (the symbol above the numbers) means prime. But why do we have it? To differentiate between the groups in the ribose vs the groups in the nitrogenous bases.

At carbon number 5 we have a phosphate group that is linked to this carbon. The phosphate group carries negative charges so if you have a DNA molecule that contains a lot of phosphate groups becomes acidic molecule. That’s why its known as deoxyribonucleic ACID

Ribose for RNA
Deoxyribose for DNA

What do we mean by deoxy? It means that in carbon number 2 the oxygen is replaced by hydrogen and there is no more oxygen.

Phosphate

Purine or pyrimidine base

Glycosidic bond

Ribose vs. deoxyribose

This whole structure is a nucleotide

This is why they are acidic

• Positively charged ions (Na$^+$ or Mg$^{2+}$) associate with the phosphate groups.
  • Example: histones For more stability and less repulsion.

Ribose for RNA
Deoxyribose for DNA

What do we mean by deoxy? It means that in carbon number 2 the oxygen is replaced by hydrogen and there is no more oxygen.

Chemical composition and bonds

Glycosidic bond

Pay attention to the numbering of carbons, this symbol (the symbol above the numbers) means prime. But why do we have it? To differentiate between the groups in the ribose vs the groups in the nitrogenous bases.

At carbon number 5 we have a phosphate group that is linked to this carbon. The phosphate group carries negative charges so if you have a DNA molecule that contains a lot of phosphate groups becomes acidic molecule. That’s why its known as deoxyribonucleic ACID

Ribose for RNA
Deoxyribose for DNA

What do we mean by deoxy? It means that in carbon number 2 the oxygen is replaced by hydrogen and there is no more oxygen.

Phosphate

Purine or pyrimidine base

Glycosidic bond

Ribose vs. deoxyribose

This whole structure is a nucleotide

This is why they are acidic

• Positively charged ions (Na$^+$ or Mg$^{2+}$) associate with the phosphate groups.
  • Example: histones For more stability and less repulsion.

Ribose for RNA
Deoxyribose for DNA

What do we mean by deoxy? It means that in carbon number 2 the oxygen is replaced by hydrogen and there is no more oxygen.
Nitrogenous bases

There are two types of bases: Pyrimidine and Purine.

YOU DON'T HAVE TO MEMORIZE THE NUMBERING OF THE DIFFERENT GROUPS

To nitrogen number one (Pyrimidine) you have the glycosidic bond between the base and the sugar and to nitrogen number 9 (Purine) you also have the glycosidic bond between the purine and the sugar, notice that pyrimidines are single ringed molecules. And purines are double ringed molecules.

How to differentiate? (that's how the doctor did)

The big word is for the small structure and the small word is for the big structure.

Notice the differences between the 3 different types of pyrimidine:

1. Cytosine: we have a ketone and amino group
2. Thymine: no more amino group and has a methyl group and a ketone group
3. Uracil: Looks like cytosine except that it has two keto groups.

Purines are two types:

1. Adenine: Amino group
2. Guanine: Ketone group (on top)
In prokaryotes and eukaryotes, not viruses.

DNA vs. RNA

- DNA: double-stranded, deoxyribose as the sugar, bases: Thymine (T), Cytosine (C), Adenine (A), Guanine (G)
- RNA: generally single-stranded, ribose as the sugar, bases: Uracil (U), Cytosine (C), Adenine (A), Guanine (G)

There are always exceptions especially in lower organisms like viruses and bacteria.
A nucleotide can have one or two or three phosphate groups, that's why we use mono, di, tri to differentiate between them.

-Nucleotides make up DNA & RNA.
How to differentiate between them?
1. It contains a sugar, nitrogenous base and a phosphate group --> Nucleotide

2. How many phosphate groups?
   1 --> Mono / 2 --> Di / 3 --> Tri

3. Look at the sugar at carbon(2)
   Hydroxyl group --> Ribose --> RNA
   Hydrogen group --> Deoxyribose --> DNA

4. Look at the base (purine or pyrimidine).

5. Look at the name of the base

- If you have 2 phosphate groups the name becomes Diphosphate not monophosphate
Formation of a nucleic acid polymer

The phosphodiester bond links nucleotides to each other's (between Carbon number 3 with a phosphate that is linked to Carbon number 5.)

The 3' end can be used to elongate the nucleic acid
A letter d can be added to indicate a deoxyribonucleotide residue.
- for example, dG is substituted for G.
- The deoxy analogue of a ribooligonucleotide would be d(GACAT).

Does uracil exist in DNA? It does when there is a mutation so it must be repaired.

-What is the sequence for the DNA molecule (order of bases)?
TGCA
Characteristics of DNA molecule:

- A double helix
- Specific base-pairing
  - A = T; G = C; Pur = pyr
- Complementary
- Backbone vs. side chains
- Antiparallel
- Stability vs. flexibility
- Groovings

The DNA structure was revealed by two scientists known as Watson and Crick and they did something known as crystallography, basically they converted DNA molecule into a crystal then they hit it with x-ray and they got this pattern. Based on this they interpreted this image into the structure of DNA and they got the nobel prize in 1961-1962, their discovery was published in 1953.

Double helix is the famous name for the DNA molecule. We mean by helix is that it look like a spring, so it rotates, and we have two strands that intertwine around each other in a helical formation.

This is a perfect helix; it is a really uniform helical structure. DNA is not a perfect helix rather than that, it is a sort of like bend.
There is a Russian scientist known as Chargaff, he noticed that in the 1940s something interesting about DNA molecules and that is the number of A’s is always equal to the number of T’s as nucleotides, and the number of G’s is always equal to the number of C’s and the number of purines is always equal to the number of pyrimidines.

But it is not necessary that the number of As and Ts would be equal to the number of Gs and Cs, so based on this information Watson and Crick came up with their model for DNA structure.
Base pairing

The two strands bind to each other via hydrogen bonding, the hydrogen bonding between the bases. Whenever you have an A it is always hydrogen bonded to a T this T exists on the other strand and number of hydrogen bond is always two. On the other hand, a G is also hydrogen bonded to a C and the number of hydrogen bond is always three.

This has very important implication about the stability of DNA molecule as a double stranded molecule.
DNA is complementary meaning that the bases complement each other so that whenever you have a C you have a G on the opposite strand, and whenever you have an A you have T on the opposite strand, so we say the DNA molecule is complementary, the two strands are complementary to each other.
If you look at the structure of DNA there is a backbone and side chains, so the backbone basically is the straight line of phosphate, sugar, phosphate, sugar, phosphate, sugar and so on. The side chains are the bases themselves; they extend almost perpendicular to the backbone.

This is an important concept in organic chemistry and in biochemistry, especially in polymers. In polymers you always have backbone and sometime have side chains.
Remember the DNA is double stranded molecule and each one of these strands has a five prime end and three prime end, but the two strands are not parallel to each other, they are anti-parallel, so in one strand on top there is five prime end and on the other end there is three prime end, so if you look at the opposite strand you would notice that on top you have the three prime end and, in the bottom, you have the five-prime end.
Writing the sequence of nucleic acids

DNA  5’ ...ATGGCCCTGGGACTTCA... 3’
    3’ ...TACCGGACCTGAAAGT... 5’

OR  A T G G C C T G G A C T T C A

RNA  5’ ...AUGGCCGUUGGACUUCA... 3’

If the doctor, ask you to read or write the sequence of DNA or RNA you do not have to
tell him two things:
1) you do not have to say 5 prime and 3 prime because once you start it is known that
you start at the 5-prime end, and you end at the 3-prime end.
2) you do not have to say what the complementary strand is, in DNA.

If I want to say the complementary strand in DNA, I will have to do it from 5 prime
end unless I say 3 prime TACCGG.........
If I do not say 5 prime and start reading the bases you should know that I am starting
at the 5-prime end.
DNA is flexible, yet stable, basically the idea that it is look like an electric wire so if you try to stretch it or rip it off you would not be able to because it is a strong structure, but it can be bent.

This is an important characteristic in DNA, this is how it is packed in the nucleus of every cell in our body, this is how it interact with proteins as well, because it is flexible it can bend.
DNA grooves

Since DNA molecule is not a perfect helical molecule, it contains grooves, so there are two grooves, one is known as a minor groove and the other is a major groove. The space in the major groove is larger than the space in minor groove, and they are alternate, so you have minor groove, major groove, minor groove, major groove, and so on.

Opposite to a minor groove there is a major groove, and opposite to a major groove there is a minor groove, these groove are important for facilitating the interaction between DNA and proteins. Proteins need space, so that they insert themselves within the major groove, and the amino acids of proteins can interact with the exposed groups of the nucleotides.

Once the proteins insert themselves into the major groove, they would be closer to the nucleotides and that is why proteins can interact specifically with specific sequences of DNA because we have specific amino acids within the protein that exist in a certain space that can interact with specific sequences or specific nucleotides within the major groove and this specific interaction is very important as well when we start talking about replication and transcription.
DNA-protein interaction
Prokaryotes versus eukaryotes

DNA is organized into a large structure called a chromosome.

There is a difference in chromosome structures in Prokaryote & Eukaryote:

1- Nucleus:
- Eu—< Indicates that this cell contains true nucleus → has nuclear membrane that surrounds DNA
- Pro—< Indicates that this cell doesn’t contain true nucleus (doesn’t have membrane), So that DNA is swimming inside the cytosol.

2- DNA shape:
- Prokaryote → circular DNA (single DNA)
- Eukaryote → Linear DNA (Multiple—< different strands)

No nucleus
Single loop of DNA

Has a Nucleus with DNA in non-looped chromosomes
In eukaryotes...

In eukaryotes, DNA is coiled to package the large DNA.

Eukaryotic DNA is complexed with a number of proteins, principally histones, which package DNA.

- Chromatin = DNA molecule + proteins. —<Histones

- The basic structural unit of chromatin is known as a nucleosome.

Histones are positively charged proteins that bind with DNA and then DNA wraps around it.
Nucleosomes

- A nucleosome consists of DNA wrapped around a nucleosome core particle, linker DNA, and histone H1.
- The histone core particle is an octamer (two molecules of histones H2A, H2B, H3, and H4) and the DNA wrapped around it.
- A linker DNA connects two nucleosome core particles.
- Histone H1 is bound to the octamer and wrapped DNA (a chromatosome).
- Histones are positively charged facilitating DNA interaction and charge neutralization.

The histone protein core is an octamer (8 molecules: two molecules of histones H2A, H2B, H3, and H4).

Chromatosome: Octameric Histone Core + DNA + Histone H1 (without Linker DNA).
Histones package chromosomes

Chromosome: consists of 2 chromatins

Chromatin: is the combination of DNA and protein (A group of nucleosome was stacked and assembled)

What determines if the chromatin is Loose or packed? The presence of Histamine H1
Loose: without Histamine H1
Packed: contains Histamine H1
Terms to know
Chromosome consists of sister chromatids (two identical chromatids) like x-shaped

Gene: a region in the DNA molecule, it is used to manufacture RNA Molecules

Chromosome (sister chromatid) → chromatin (Histone +DNA) → Nucleosome → (Double Stranded Molecule) DNA → Monomers → nucleotide (anti parallel)
Remember...we are diploid

**Haploid (n)**
- One copy of each chromosome
  - Three non-homologous chromosomes

**Diploid (2n)**
- Two copies of each chromosome
  - Three pairs of homologous chromosomes (of maternal and paternal origin)

This means that we have two copies of each chromosome. One from the mother and the other from the father.

**Diploid (2n):** Has two copies of each chromosomes. These pairs of homologous chromosomes (maternal and paternal) found in a body cell

**Haploid (1n):** non-homologous chromosomes found in reproductive cells

*In meiosis two each sister chromatid goes to a different cell*
RNA

- It consists of long, unbranched chains of nucleotides joined by phosphodiester bonds between the 3’-OH of one pentose and the 5’-PO$_4^-$ of the next.
- The pentose unit is a ribose (it is 2-deoxyribose in DNA).
- The pyrimidine bases include uracil and cytosine (thymine and cytosine in DNA).
- In general, RNA is single stranded (DNA is double stranded).

RNA does not have a precise structure, but it can fold on itself forming hydrogen bonds within the same molecule.

There are some exception: the genetic material of viruses might have double stranded RNA.

RNA does not have a particular/Define structure like DNA within the same structure. Hydrogen bonds may occur when we have bases complementary to each other.
### Types of RNA

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Non-Coding RNAs</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>* tRNA</td>
<td>Transfer RNA</td>
<td>mRNA translation (structural)</td>
</tr>
<tr>
<td>* rRNA</td>
<td>Ribosomal RNA</td>
<td>mRNA translation (structural)</td>
</tr>
<tr>
<td>* miRNA</td>
<td>micro RNAs</td>
<td>Post-transcriptional transposon repression</td>
</tr>
<tr>
<td>piRNA</td>
<td>Piwi-interacting RNA</td>
<td>DNA methylation, transposon repression</td>
</tr>
<tr>
<td>* siRNA</td>
<td>Short interfering RNA</td>
<td>RNA interference</td>
</tr>
<tr>
<td>snoRNA</td>
<td>Small nucleolar RNAs</td>
<td>RNA modification, rRNA processing</td>
</tr>
<tr>
<td>PROMPT’s</td>
<td>Promoter upstream transcripts</td>
<td>Associated with chromatin changes</td>
</tr>
<tr>
<td>tiRNAs</td>
<td>Transcription initiation RNAs</td>
<td>Epigenetic regulation</td>
</tr>
<tr>
<td>lincRNAs</td>
<td>Long intergenic ncRNA</td>
<td>Epigenetic regulators of transcription</td>
</tr>
<tr>
<td>rasiRNA</td>
<td>Repeat associated small interfering RNA</td>
<td>Involved in the RNA interference (RNAi) pathway</td>
</tr>
<tr>
<td>eRNA</td>
<td>Enhancer-like ncRNA</td>
<td>Transcriptional gene activation</td>
</tr>
<tr>
<td>T-UCRs</td>
<td>Transcribed ultraconserved regions</td>
<td>Regulation of miRNA and mRNA levels</td>
</tr>
<tr>
<td>NATs</td>
<td>Natural antisense transcripts</td>
<td>mRNA stability</td>
</tr>
<tr>
<td>PALRs</td>
<td>Promoter-associated long RNAs</td>
<td>Chromatin changes</td>
</tr>
<tr>
<td>tasiRNA</td>
<td>Trans-acting siRNA</td>
<td>Represses gene expression</td>
</tr>
<tr>
<td>* lncRNA</td>
<td>Long noncoding RNA</td>
<td>Regulation of gene transcription</td>
</tr>
</tbody>
</table>

**tRNA and rRNA are RNA molecules that are important to mRNA translation (proteins synthesis)**

Just Know which types have a star and we will study them in detail later.