Lecture 1. Introduction to molecular biology. Part I.

Learning outcomes:

1. Describe the Chargaff, Griffith, Avery-MacLeod-McCarty, Hershey-Chase experiments and explain their significance.
2. Explain informational properties of macromolecules.
3. Explain the central dogma of molecular biology.
4. Briefly discuss the role of molecular biology in medicine.
5. Describe, identify and and draw the components of nucleosides and nucleotides.
6. Characterise and describe the chains of nucleic acids in DNA and RNA.

Many people believe that American biologist James Watson and English physicist Francis Crick discovered DNA in the 1950s. In reality, this is not the case. Rather, DNA was first identified in the late 1860s by Swiss chemist Friedrich Miescher. Then, in the decades following Miescher's discovery, other scientists--notably, Phoebus Levene and Erwin Chargaff--carried out a series of research efforts that revealed additional details about the DNA molecule, including its primary chemical components and the ways in which they joined with one another. Without the scientific foundation provided by these pioneers, Watson and Crick may never have reached their groundbreaking conclusion of 1953: that the DNA molecule exists in the form of a three-dimensional double helix. Many people believe that American biologist James Watson and English physicist Francis Crick discovered DNA in the 1950s. In reality, this is not the case. Rather, DNA was first identified in the late 1860s by Swiss chemist Friedrich Miescher. Then, in the decades following Miescher's discovery, other scientists--notably, Phoebus Levene and Erwin Chargaff--carried out a series of research efforts that revealed additional details about the DNA molecule, including its primary chemical components and the ways in which they joined with one another. Without the scientific foundation provided by these pioneers, Watson and Crick may never have reached their groundbreaking conclusion of 1953: that the DNA molecule exists in the form of a three-dimensional double helix.

Although few people realize it, 1869 was a landmark year in genetic research, because it was the year in which Swiss physiological chemist Friedrich Miescher first identified what he called "nuclein" inside the nuclei of human white blood cells. More than 50 years passed before the significance of Miescher's discovery of nucleic acids was widely appreciated by the scientific community. Meanwhile, even as Miescher's name fell into obscurity by the twentieth century, other scientists continued to investigate the chemical nature of the molecule formerly known as nuclein. One of these other scientists was Russian biochemist Phoebus Levene. Levene is credited with many firsts. For instance, he was the first to discover the order of the three major components of a single nucleotide (phosphate-sugar-base); the first to discover the carbohydrate component of RNA (ribose); the first to discover the carbohydrate component of DNA (deoxyribose); and the first to correctly identify the way RNA and DNA molecules are put together.

During the early years of Levene's career, neither Levene nor any other scientist of the time knew how the individual nucleotide components of DNA were arranged in space; discovery of the sugar-phosphate backbone of the DNA molecule was still years away. The large number of molecular groups made available for binding by each nucleotide component meant that there were numerous alternate ways that the components could combine. Several scientists put forth suggestions for how this might occur, but it was Levene's "polynucleotide" model that proved to be the correct one.

Erwin Chargaff was one of a handful of scientists who expanded on Levene's work by uncovering additional details of the structure of DNA, thus further paving the way for Watson and Crick. Chargaff, an Austrian biochemist, had read the famous 1944 paper by Oswald Avery and his colleagues at Rockefeller University, which demonstrated that hereditary units, or genes, are composed of DNA. This paper had a profound impact on Chargaff, inspiring him to launch a

research program that revolved around the chemistry of nucleic acids. Of Avery's work, Chargaff (1971) wrote the following:

"This discovery, almost abruptly, appeared to foreshadow a chemistry of heredity and, moreover, made probable the nucleic acid character of the gene... *Avery gave us the first text of a new language, or rather he showed us where to look for it. I resolved to search for this text."*

As his first step in this search, Chargaff set out to see whether there were any differences in DNA among different species. After developing a new paper chromatography method for separating and identifying small amounts of organic material, Chargaff reached two major conclusions (Chargaff, 1950). First, he noted that the nucleotide composition of DNA varies among species. In other words, the same nucleotides do not repeat in the same order, as proposed by Levene. Second, Chargaff concluded that almost all DNA--no matter what organism or tissue type it comes from--maintains certain properties, even as its composition varies. In particular, the amount of adenine (A) is usually similar to the amount of thymine (T), and the amount of guanine (G) usually approximates the amount of cytosine (C). In other words, the total amount of purines (A + G) and the total amount of pyrimidines (C + T) are usually nearly equal. (This second major conclusion is now known as "Chargaff's rule.") Chargaff's research was vital to the later work of Watson and Crick, but Chargaff himself could not imagine the explanation of these relationships--specifically, that A bound to T and C bound to G within the molecular structure of DNA

Chargaff's realization that A = T and C = G, combined with some crucially important X-ray crystallography work by English researchers Rosalind Franklin and Maurice Wilkins, contributed to Watson and Crick's derivation of the three-dimensional, double-helical model for the structure of DNA. Watson and Crick's discovery was also made possible by recent advances in model building, or the assembly of possible three-dimensional structures based upon known molecular distances and bond angles, a technique advanced by American biochemist Linus Pauling. In fact, Watson and Crick were worried that they would be "scooped" by Pauling, who proposed a different model for the three-dimensional structure of DNA just months before they did. In the end, however, Pauling's prediction was incorrect.

Using cardboard cutouts representing the individual chemical components of the four bases and other nucleotide subunits, Watson and Crick shifted molecules around on their desktops, as though putting together a puzzle. They were misled for a while by an erroneous understanding of how the different elements in thymine and guanine (specifically, the carbon, nitrogen, hydrogen, and oxygen rings) were configured. Only upon the suggestion of American scientist Jerry Donohue did Watson decide to make new cardboard cutouts of the two bases, to see if perhaps a different atomic configuration would make a difference. It did. Not only did the complementary bases now fit together perfectly (i.e., A with T and C with G), with each pair held together by hydrogen bonds, but the structure also reflected Chargaff's rule

Although scientists have made some minor changes to the Watson and Crick model, or have elaborated upon it, since its inception in 1953, the model's four major features remain the same yet today. These features are as follows:

- DNA is a double-stranded helix, with the two strands connected by hydrogen bonds. A bases are always paired with Ts, and Cs are always paired with Gs, which is consistent with and accounts for Chargaff's rule.
- Most DNA double helices are right-handed; that is, if you were to hold your right hand out, with your thumb pointed up and your fingers curled around your thumb, your thumb would represent the axis of the helix and your fingers would represent the sugar-phosphate backbone. Only one type of DNA, called Z-DNA, is left-handed.
- The DNA double helix is anti-parallel, which means that the 5' end of one strand is paired with the 3' end of its complementary strand (and vice versa). As shown in Figure 4, nucleotides are linked to each other by their phosphate groups, which bind the 3' end of one sugar to the 5' end of the next sugar.

• Not only are the DNA base pairs connected via hydrogen bonding, but the outer edges of the nitrogen-containing bases are exposed and available for potential hydrogen bonding as well. These hydrogen bonds provide easy access to the DNA for other molecules, including the proteins that play vital roles in the replication and expression of DNA

Nucleic acid, naturally occurring chemical compound that is capable of being broken down to yield phosphoric acid, sugars, and a mixture of organic bases (purines and pyrimidines). Nucleic acids are the main information-carrying molecules of the cell, and, by directing the process of protein synthesis, they determine the inherited characteristics of every living thing. The two main classes of nucleic acids are deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). DNA is the master blueprint for life and constitutes the genetic material in all free-living organisms and most viruses. RNA is the genetic material of certain viruses, but it is also found in all living cells, where it plays an important role in certain processes such as the making of proteins.

Nucleic acids are polynucleotides—that is, long chainlike molecules composed of a series of nearly identical building blocks called nucleotides. Each nucleotide consists of a nitrogen-containing aromatic base attached to a pentose (five-carbon) sugar, which is in turn attached to a phosphate group. Each nucleic acid contains four of five possible nitrogen-containing bases: adenine (A), guanine (G), cytosine (C), thymine (T), and uracil (U). A and G are categorized as purines, and C, T, and U are collectively called pyrimidines. All nucleic acids contain the bases A, C, and G; T, however, is found only in DNA, while U is found in RNA. The pentose sugar in DNA (2′-deoxyribose) differs from the sugar in RNA (ribose) by the absence of a hydroxyl group (—OH) on the 2′ carbon of the sugar ring. Without an attached phosphate group, the sugar attached to one of the bases is known as a nucleoside. The phosphate group connects successive sugar residues by bridging the 5′-hydroxyl group on one sugar to the 3′-hydroxyl group of the next sugar in the chain. These nucleoside linkages are called phosphodiester bonds and are the same in RNA and DNA.

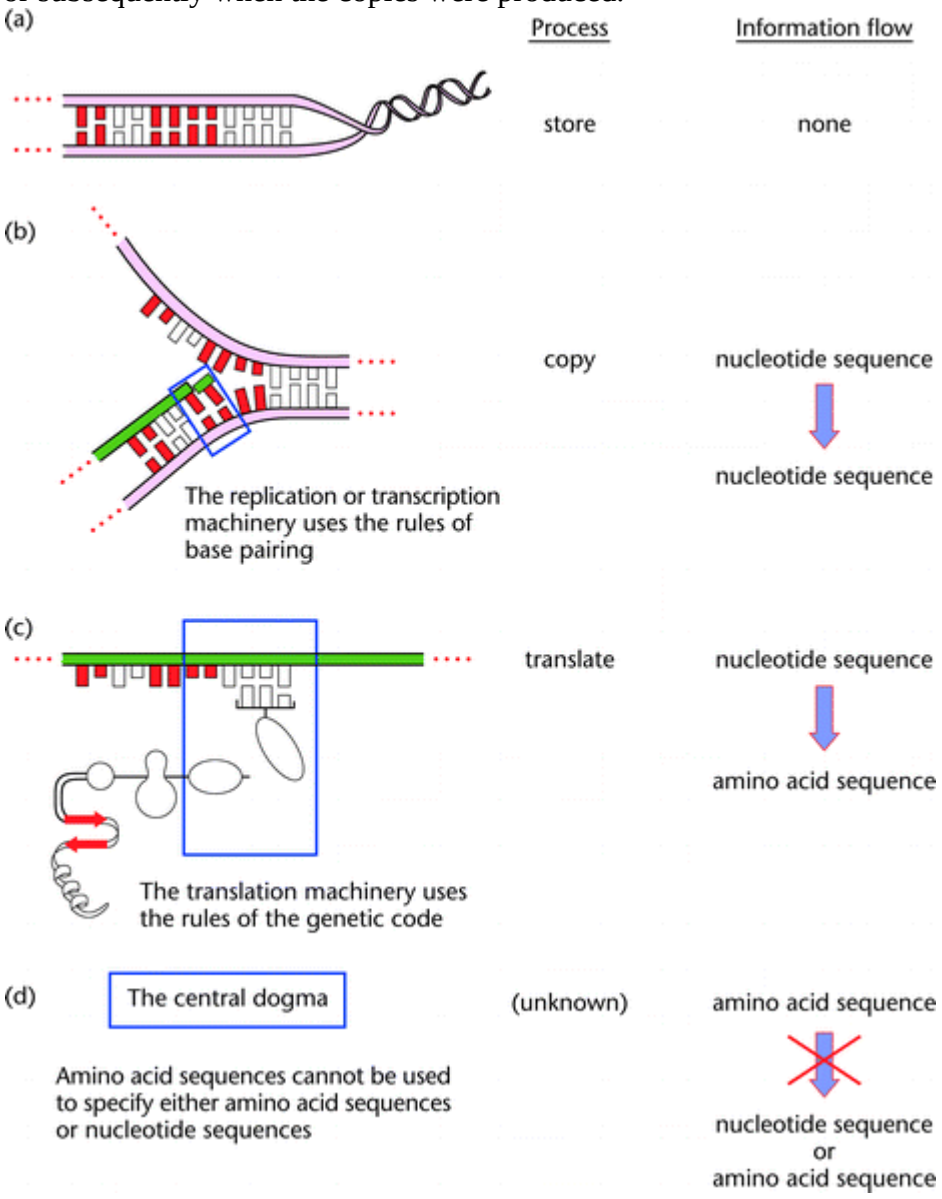
DNA is a polymer of the four nucleotides A, C, G, and T, which are joined through a backbone of alternating phosphate and deoxyribose sugar residues. These nitrogen-containing bases occur in complementary pairs as determined by their ability to form hydrogen bonds between them. A always pairs with T through two hydrogen bonds, and G always pairs with C through three hydrogen bonds. The spans of A:T and G:C hydrogen-bonded pairs are nearly identical, allowing them to bridge the sugar-phosphate chains uniformly. This structure, along with the molecule's chemical stability, makes DNA the ideal genetic material. The bonding between complementary bases also provides a mechanism for the replication of DNA and the transmission of genetic information. The double helical structure of normal DNA takes a right-handed form called the B-helix. The helix makes one complete turn approximately every 10 base pairs. B-DNA has two principal grooves, a wide major groove and a narrow minor groove. Many proteins interact in the space of the major groove, where they make sequence-specific contacts with the bases. In addition, a few proteins are known to make contacts via the minor groove.

RNA is a single-stranded nucleic acid polymer of the four nucleotides A, C, G, and U joined through a backbone of alternating phosphate and ribose sugar residues. It is the first intermediate in converting the information from DNA into proteins essential for the working of a cell. Some RNAs also serve direct roles in cellular metabolism. RNA is made by copying the base sequence of a section of double-stranded DNA, called a gene, into a piece of single-stranded nucleic acid. This process, called transcription

The Concept of Information Flow

The DNA macromolecule is an elegant structure that accommodates the need to store and transmit genetic information. A chromosomal DNA molecule is, in fact, two single chains of nucleotide subunits that are complementary to each other (Figure **1**a). Thus there is present both the primary information, and the template necessary to make a new copy of that information within this chemically inert macromolecule. The redundancy helps to ensure that essential

information is not lost, because if one copy is damaged, the other can serve as a template for repair. The dangers inherent in transmitting genetic information are also reduced. If only a single chain were present, essential information might be lost either when the necessary template was assembled, or subsequently when the copies were produced.



Information transfer between DNA, RNA and protein macromolecules. (a) Complementary base pairing in a short segment of DNA. The four types of nucleotides are represented by the bars (solid or open, long or short), and each nucleotide can pair with only one complement. (b) Information transfer between DNA (drawn as in part a) and RNA (green). RNA can also be used to specify DNA by an analogous mechanism. (c) Information transfer from RNA to protein. The 20 types of amino acids (the four oval shapes represent four of these) are linked according to the sequence of nucleotides in the messenger RNA, which in turn was copied from the gene in part b. The next amino acid which will be added is shown with its adaptor transfer RNA (tRNA) attached. (d) The central dogma of molecular biology specifies the forbidden information transfers.

The central dogma of molecular biology predicts that a particular sequence of amino acids (a protein) cannot be used to specify or even alter a particular sequence of nucleotides (a gene). Instead, information flows from nucleic acids to proteins, in that an elaborate machinery exists to ‘translate’ the nucleic acid ‘alphabet’ to the amino acid ‘alphabet’ according to the rules of the genetic code. Cells exhibit no trace of a ‘back-translation’ machinery, and organisms can

transmit only their genes to their offspring. Even though the genetic material is not entirely constant, advantageous mutations do not arise in a directed manner. The predictions of the central dogma have withstood every challenge, and are likely to remain as the central organizing principles of molecular biology.

Modern molecular medicine encompasses the utilization of many molecular biological techniques in the analysis of disease, disease genes and disease gene function. The study of disease genes and their function in an unaffected individual has been possible by the development of recombinant DNA and cloning techniques. The basis of the term recombinant DNA refers to the recombining of different segments of DNA. Cloning refers to the process of preparing multiple copies of an individual type of recombinant DNA molecule. The classical mechanisms for producing recombinant molecules involves the insertion of exogenous fragments of DNA into either bacterially derived plasmid (circular double stranded autonomously replicating DNAs found in bacteria) vectors or bacteriophage (viruses that infect bacteria) based vectors.

The questions for self - control:

1. What is the subject of molecular biology? What are its connections with other sciences?
2. What was the history of molecular biology? What is the main carrier of genetic information, how it was discovered and studied?
3. What are the nucleosides, nucleotides and nucleic acids? What are their chemical structure and properties?

Recommended readings:

1. Russell P.J. iGenetics. A molecular approach. 3rd ed. 2009. Pearson. pp. 9-14.
2. Hartwell L. et al. Genetics. From genes to genomes. 4th ed. 2011. McGraw Hill. pp. 163-168.
3. Watson, James D. Molecular biology of the gene. 7th ed. Pearson. pp. 21-33.
4. John McMurry, et al. Fundamentals of General, Organic, and Biological Chemistry, 8th Edition. 2018. Pearson Education Limited. pp. 814-823.