

Concept article: Minireviews

DNA Life's Code and Blueprint

RACHEDI Abdelkrim

Laboratory of Biototoxicology, Pharmacognosy and biological valorisation of plants, Faculty of Sciences, Department of Biology, University Dr Tahar Moulay Saida, 20100 Saida, Algeria.

Correspondence: Abdelkrim RACHEDI – E.mail: abdelkrim.rachedi@univ-saida.dz

Abstract

Deoxyribonucleic acid (DNA) is a complex molecule that carries the genetic information of all living organisms. The discovery of its structure in 1953 by Watson and Crick revolutionized the field of molecular biology and led to significant advances in our understanding of genetics. DNA can adopt various conformations, and its sequence determines the sequence of amino acids in the resulting protein. Genes are segments of DNA that contain the instructions for making proteins, and the human genome contains approximately 20,000 to 25,000 genes. The relationship between DNA and genomes is essential to our understanding of genetics. This article aims to provide a comprehensive overview of the structure, function, biology-relevant conformations, genes, and relation to genomes of DNA.

Key words

DNA, Genome, DNA conformations, Genetic Code, Biological Function, 3D-Structure, Bio-Databases.

Introduction

DNA is a fundamental molecule that plays a crucial role in the storage and transmission of genetic information in all living organisms. It is a long polymer composed of four different nucleotide monomers: adenine (A), cytosine (C), guanine (G), and thymine (T). These nucleotides are arranged in a specific sequence to form the genetic code that determines the characteristics of an organism, Figure 1.

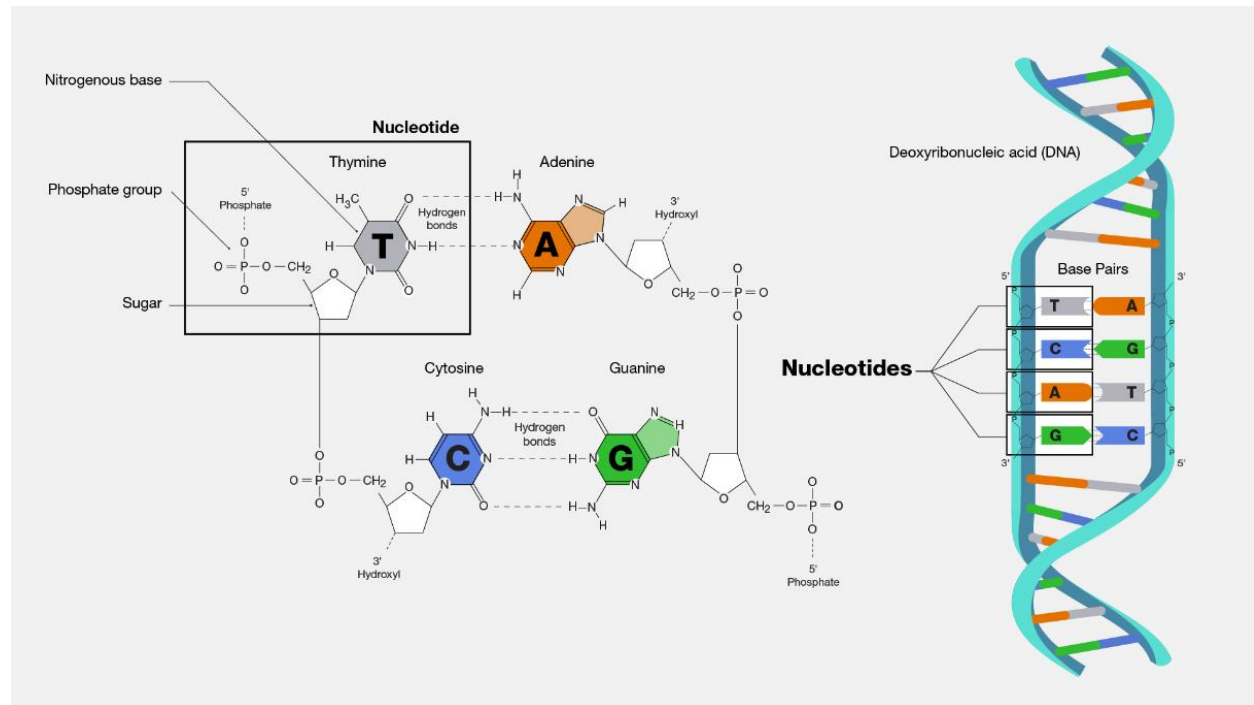


Figure 1. Nucleotides adenine (A), cytosine (C), guanine (G) and thymine (T) in the DNA. A DNA nucleotide consists of a deoxyribose sugar molecule attached to a phosphate group and a nitrogen-containing base.

Image source: <https://www.genome.gov/sites/default/files/media/images/tg/Nucleotide.jpg>

The discovery of the structure of DNA by Watson and Crick in 1953 was a turning point in the history of biology. Their discovery revealed that DNA has a double helix structure, Figure 2, which is composed of two chains, also known as strands, which in turn consist of nucleotides (A, C, G and T) that are connected to the next by covalent bonds. The DNA strands run in opposite directions to each other and are held together by hydrogen bonds between the nitrogenous base pairs, which are located in the center of the helix. The two strands are oriented in an antiparallel fashion, meaning that one strand runs in the 5' to 3' direction, while the other strand runs in the opposite 3' to 5' direction.

This arrangement of the strands has been found crucial for DNA replication and transcription, which are essential processes in the expression of genetic information. The structure of DNA allows for the faithful transmission of genetic information from one generation to the next.

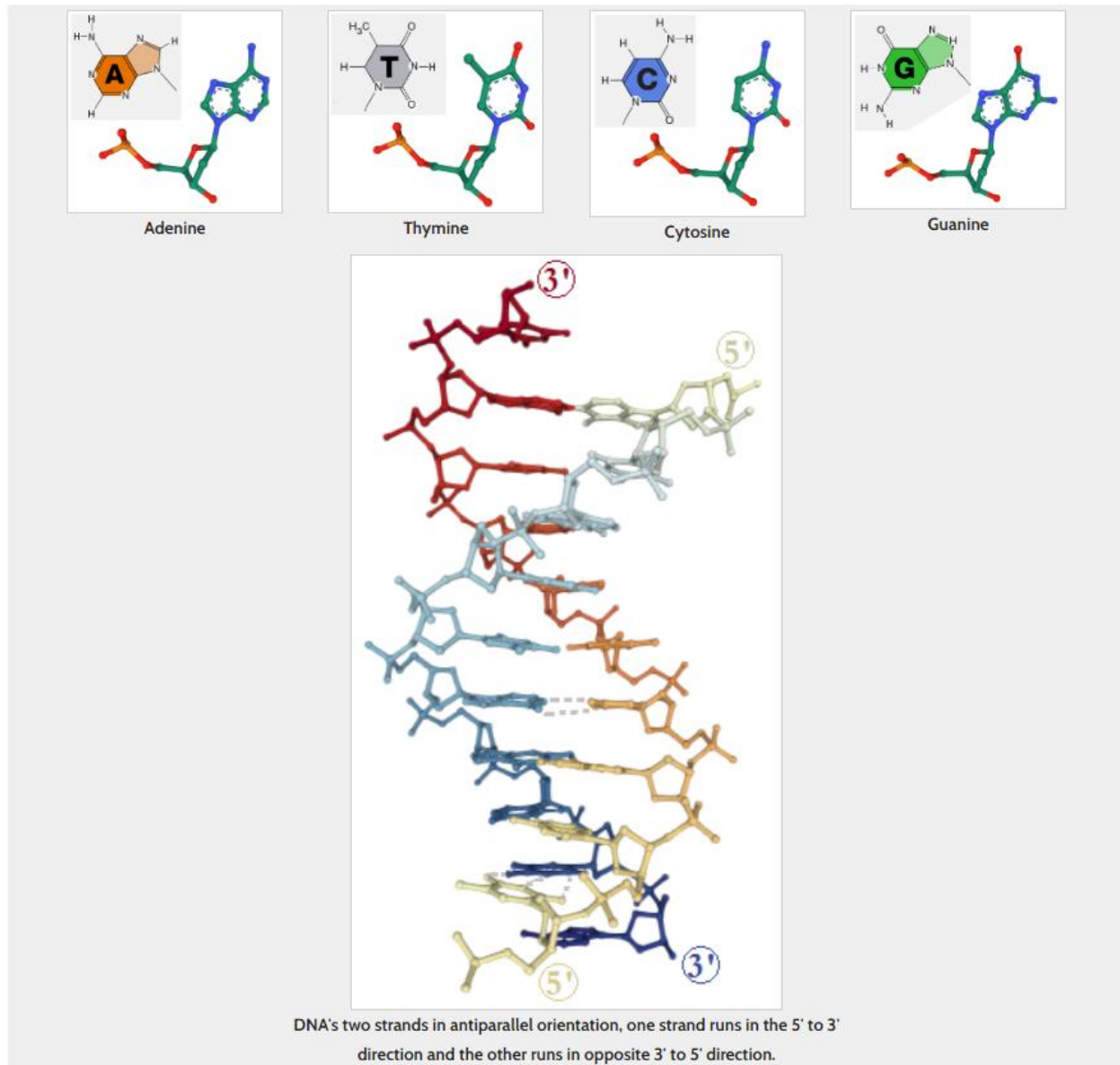


Figure 2. The Watson and Crick model for double helix DNA. The two strands are antiparallel oriented, one strand runs in the 5' to 3' direction and the other runs in opposite 3' to 5' direction. The two strands are held together via hydrogen bonds between pairs of the bases A-T and C-G

DNA can be characterised as the fundamental code and blueprint of life. This is due to the fact that DNA carries the genetic information that governs the characteristics and functions of all living organisms. Composed of nucleotides arranged in a specific sequence, the DNA molecule encodes the instructions for an organism's growth, development, and reproduction.

The sequence of nucleotides in DNA serves as a code for the production of proteins, which are crucial for the structure and function of cells and tissues. The DNA code is passed down from generation to generation, providing a blueprint for the development of new individuals with inherited traits.

X-ray Crystallography and Photograph 51:

The photograph known as Photograph 51, Figure 3, played a crucial role in uncovering the structure of DNA. This X-ray diffraction image was taken by Rosalind Franklin and Raymond Gosling at King's College London in 1952. By providing important details about the helical structure of DNA, the photograph contributed significantly to the discovery of the molecule's overall structure.

Franklin used the photograph to determine the distance between the phosphate groups in the DNA backbone and the distance between the nitrogenous bases. She was able to deduce from this information that the DNA structure was helical and that the bases were stacked on top of one another.

James Watson and Francis Crick, who were also working on the DNA structure problem at the time, were shown the photograph by their colleague, Maurice Wilkins. The photograph was a key piece of evidence that supported Watson and Crick's own model of the DNA double helix. It provided them with the missing piece of the puzzle that confirmed their theory of the structure of DNA.

Initially, Franklin's contributions to the discovery of the structure of DNA were overlooked, but her work, including the photograph, was eventually recognized as a crucial component of the discovery. Today, Photograph 51 remains an iconic symbol of the discovery of the structure of DNA and serves as a reminder of the importance of interdisciplinary scientific research.

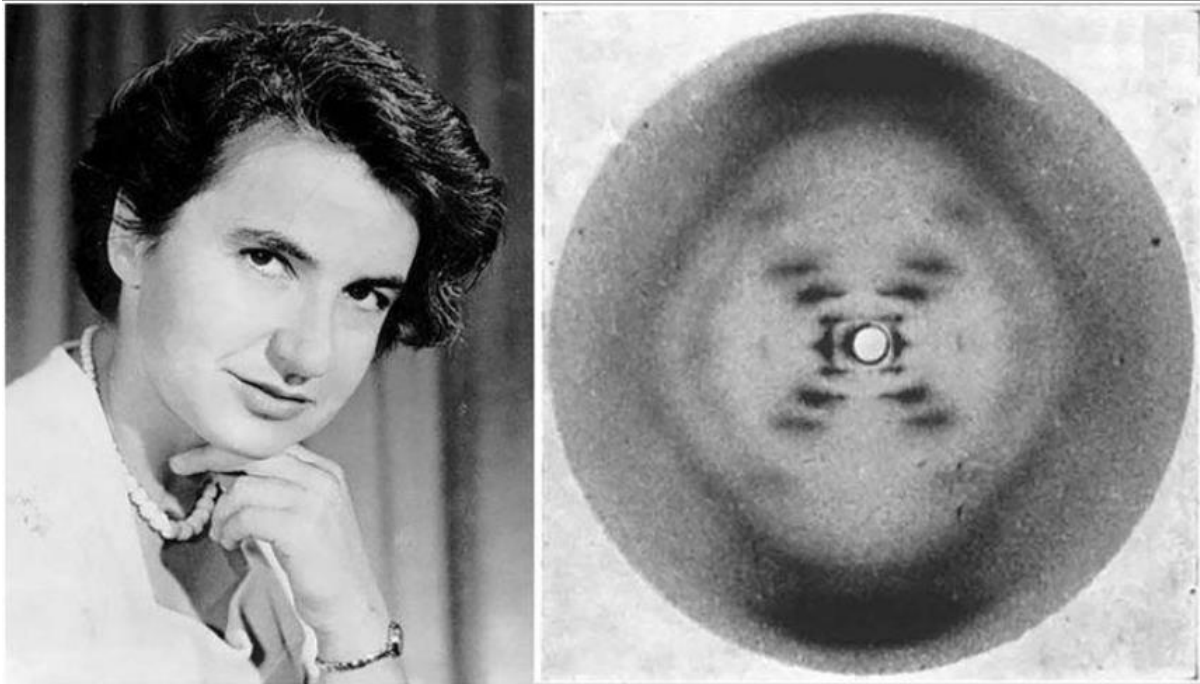


Figure 3. The X pattern, in the renowned Photograph 51 (right-side), indicates a molecule with a helical structure and regular repeats. Photograph taken by Rosalind Franklin, left-side, and her student Raymond Gosling using X-ray crystallography on hydrated fibers of B-form DNA. The image is slightly edited, see source:

<https://www.genengnews.com/topics/omics/reflections-on-the-double-helix-platinum-anniversary/>

Function of DNA:

The primary function of DNA is to carry the genetic information of an organism. The genetic information is encoded in the sequence of nucleotides, which are read and translated into proteins. Proteins are essential for the structure, function, and regulation of cells, tissues, and organs in the body. DNA also plays a crucial role in cell division, as it is replicated during cell division to ensure that each daughter cell receives a complete copy of the genetic information.

Genetic code:

The genetic code refers to a system of rules that governs the translation of genetic information encoded within DNA and RNA sequences into proteins. This code functions as a language, consisting of four nucleotide "letters" in DNA and RNA, and 20 amino acid "words" in proteins.

Every three-nucleotide set, known as a codon, specifies a particular amino acid or a stop signal during protein synthesis. There are 64 possible codons, encoding for 20 amino acids and three stop signals. The redundancy in the genetic code provides some flexibility, as multiple codons can code for the same amino acid.

The genetic code is nearly universal among living organisms, with some minor variations in certain organisms such as mitochondria and some bacteria. This universality of the genetic code allows for the transfer of genetic information between different organisms, as well as the development of techniques such as recombinant DNA technology.

Understanding the genetic code and its variations is essential in the study of genetics and molecular biology. The discovery of the genetic code was a major milestone in molecular biology, opening the door for the exploration of gene and protein structure and function.

Genes:

Genes are segments of DNA that contain the instructions for making proteins. Each gene has a specific sequence of nucleotides that determines the sequence of amino acids in the resulting protein. The human genome contains approximately 20,000 to 25,000 genes. Genes are regulated by various mechanisms, such as transcription factors, which control the expression of genes in response to various signals.

Biology-relevant Conformations:

DNA can adopt various conformations that are essential for its function. The most common conformation of DNA is the B-form, which is the double helix structure first described by Watson and Crick. However, DNA can also adopt other conformations, such as the A-form, the Z-form, and the cruciform structure, Figure 4. These conformations are involved in various biological processes, such as transcription, replication, and DNA repair.

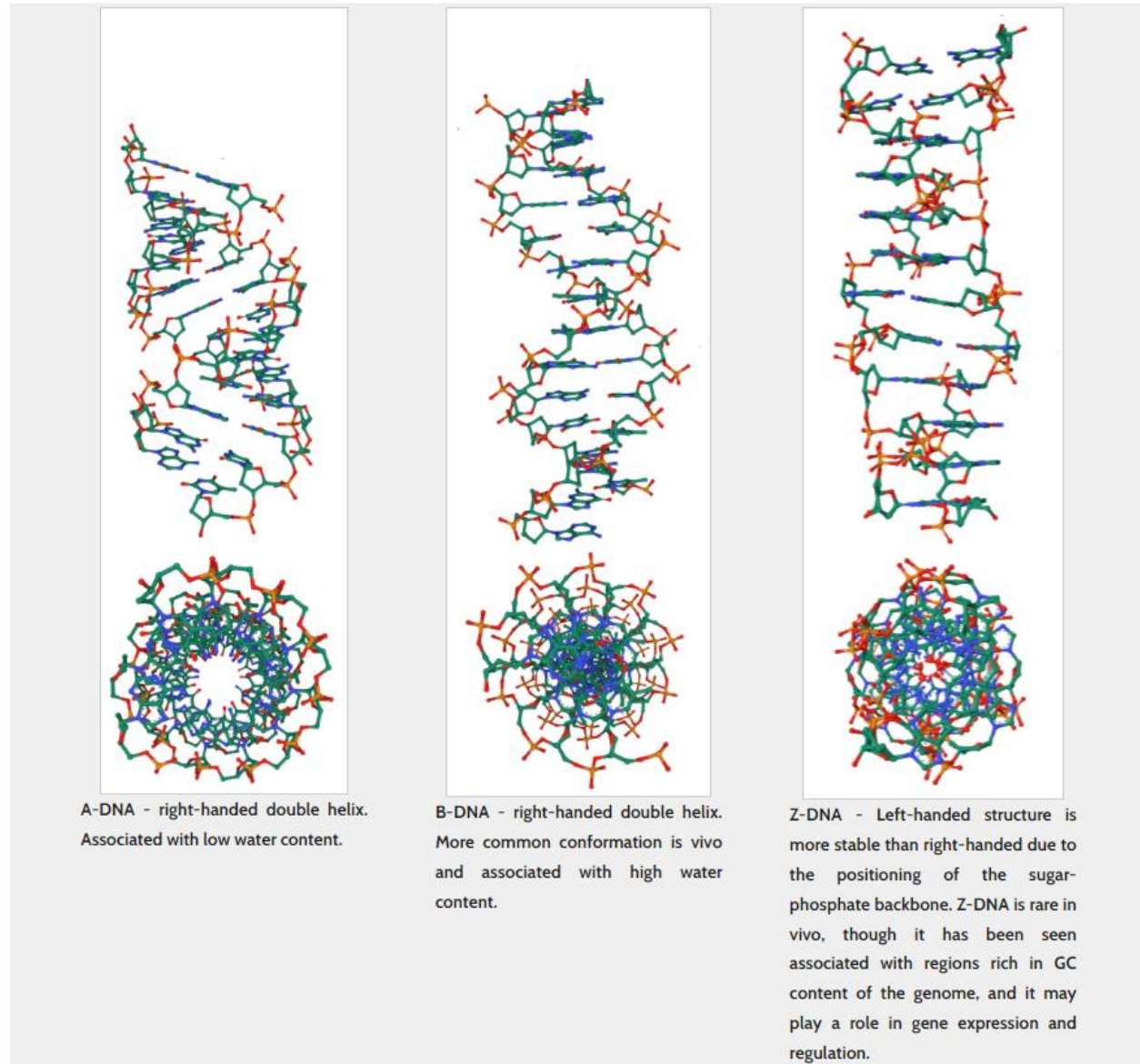


Figure 4. The three biology-relevant conformations of DNA.

DNA relation to genome:

A genome is the complete set of genetic information of an organism. The human genome, for example, consists of approximately 3 gigabase pairs (Gbp - billion base pairs) arranged in 23 pairs of chromosomes. However, the largest known genome to date is that of the rare plant *Paris Japonica* which is 149 Gbp; around 50 times larger than a human genome. Otherwise, the smallest known genome is that of *Carsonella ruddii* which is only 160 kilobase pairs in size.

It's important to note that genome size can vary widely even within a single taxonomic group, and new discoveries could potentially shift the rankings of the smallest and largest genomes.

The relationship between DNA and genomes is essential to our understanding of genetics. The study of genomes has led to significant advances in our understanding of human biology, evolution, and disease. Publicly available databases, such as NCBI and EMBL-EBI, provide access to vast amounts of data on DNA sequences, structures, and genomes and have been instrumental in advancing our understanding of genetics.

DNA data and available databases:

High-throughput sequencing technologies lead to the generation of massive amounts of genomic and DNA sequences data. Implementation of bioinformatics based analysis of sequence data and annotation resulted in important databases that contain information on entire genomes. These databases contain not only the DNA sequence of genomes, but also annotations of genes, information on motifs, transcription factors and many other DNA elements and other genomic features, such as regulatory elements and chromatin domains.

The availability of these databases has greatly facilitated the study of genetics and genomics, allowing researchers to compare and analyze the DNA sequences and structures of different organisms. Publicly available databases, such as NCBI and EMBL-EBI, provide access to vast amounts of data on DNA sequences, structures, and genomes. GenBank is a comprehensive database of DNA sequences, maintained by the National Center for Biotechnology Information (NCBI). It provides free access to over 250 million nucleotide sequences, including complete genomes, plasmids, and organelle genomes. Other databases, such as the Protein Data Bank (PDB), contain information on the 3D structure of DNA targeted sequences in addition to protein structures.

Motif databases such as TRANSFAC and JASPAR are examples of resources that contain information on DNA motifs, which are used to identify and annotate regulatory regions of genes and to predict the function of non-coding regions of DNA. These databases are also used to develop models of transcription factor binding sites, which are essential for understanding gene regulation. The database such as the Nucleic Acids and Ligands Database (NALD) provide DNA structural motifs and ligands binding data that would be relevant to novel drugs design.

The Human Genome Project, completed in 2003, was a collaborative effort to sequence and map the entire human genome. The project provided valuable insights into the structure and function of the human genome and paved the way for personalized medicine and genomic medicine.

Methods

The study of DNA has a long and storied history, starting with its initial discovery in the late 1800s. In 1869, Friedrich Miescher, a Swiss physician, isolated a substance from the nuclei of white blood cells that he called "nuclein." This substance was later identified as DNA, and subsequent research into its properties revealed its central role in the transmission of genetic information.

In the early 1950s, James Watson and Francis Crick made a breakthrough in the study of DNA by proposing its structure as a double helix. Their discovery was based on a combination of X-ray crystallography data collected by Rosalind Franklin and Maurice Wilkins, as well as their own theoretical work. This discovery paved the way for a better understanding of the function and biology of DNA, as well as its relationship to the transmission of genetic information.

Since then, a great deal of research has been conducted on DNA, including investigations into its structure, function, biology-relevant conformations, genes, and its relationship to genomes. This research has been facilitated by the development of numerous experimental techniques and technologies, including X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and next-generation sequencing (NGS).

The study of genomes has been revolutionized by the development of bioinformatics analysis and computational methods. One of the most important applications of bioinformatics is in the analysis of DNA sequences, which has led to the development of databases that contain information on genomes, DNA sequences of verity of sources and objectives, motifs, genes, and other DNA elements.

Discussion

The discovery of the structure of DNA by Watson and Crick in 1953 was a significant milestone in the history of biology. It revolutionized the field of molecular biology and led to significant advances in our understanding of genetics. The study of DNA has allowed us to unravel the mysteries of human biology, including the basis of diseases and the mechanisms of evolution.

The structure of DNA is crucial to its function. The double helix structure allows for the storage and transmission of genetic information, as well as the replication of DNA during cell division. DNA also plays a critical role in the regulation of gene expression. DNA is transcribed into messenger RNA (mRNA), which is then translated into protein. The sequence of nucleotides in DNA determines the sequence of amino acids in the resulting protein.

DNA can adopt different conformations, such as A-DNA, B-DNA, and Z-DNA. A-DNA and B-DNA are the most common forms found in nature, while Z-DNA is less common but has been implicated in gene regulation. A-DNA is a right-handed double helix that is shorter and wider than B-DNA. It is found in RNA-DNA hybrids and in DNA-protein complexes. B-DNA is a right-handed double helix with a uniform diameter and is the most stable conformation of DNA in cells. It is found in the majority of DNA structures, including the human genome. Z-DNA is a left-handed double helix that is longer and thinner than B-DNA. It is formed when DNA is under torsional strain, and its formation has been implicated in gene regulation.

Genes are segments of DNA that contain the instructions for making proteins. Genes can be turned on or off, depending on the needs of the cell. The Human Genome Project, which was completed in 2003, identified all the genes in the human genome. The human genome contains approximately 20,000 to 25,000 genes, which make up less than 2% of the total genome. The remaining DNA is non-coding and has been found to play important roles in gene regulation and other cellular processes.

The relationship between DNA and genomes is essential to our understanding of genetics. A genome is the complete set of genetic instructions for an organism. The genome of an organism is stored in its DNA, which is organized into chromosomes. Humans have 23 pairs of chromosomes, which contain approximately 3 billion base pairs of DNA. The sequence of DNA in an organism's genome determines its traits and characteristics.

Publicly available databases provide access to DNA sequences, structures, and genomes. Some of the most popular databases include the National Center for Biotechnology Information (NCBI) and the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI). These databases contain vast amounts of data on DNA sequences, structures, and genomes from various organisms, including humans.

The NCBI provides access to several databases, including GenBank, which is a comprehensive database of DNA sequences from over 400,000 organisms. The NCBI also provides access to the RefSeq database, which contains annotated and curated DNA sequences for over 130,000 organisms. In addition, the NCBI provides access to several databases that contain information on genetic variation, such as dbSNP and dbVar.

The EMBL-EBI provides access to several databases, including Ensembl, which is a comprehensive database of annotated genomes for over 80 species, including humans. Ensembl provides access to genomic data, including DNA sequences, gene annotations, and genetic variation data. The EMBL-EBI also provides access to several other databases, including UniProt, which is a comprehensive database of protein sequences and annotations.

Conclusion

DNA is a complex molecule that carries the genetic information of all living organisms. The discovery of its structure in 1953 by Watson and Crick revolutionized the field of molecular biology and led to significant advances in our understanding of genetics. DNA can adopt various conformations that are essential for its function, and genes are segments of DNA that contain the instructions for making proteins. The study of DNA and genomes has provided valuable insights into the structure and function of living organisms, and publicly available databases have facilitated the discovery of new genes and the development of novel therapies.

In conclusion, the study of DNA is a critical component of modern biology, and its impact on fields such as medicine and biotechnology cannot be overstated. The discovery of its structure, function, and biology-relevant conformations has provided us with an incredible understanding of the molecular basis of life. The identification of genes and the relationship between DNA and genomes have paved the way for personalized medicine and genomic medicine, and the availability of publicly accessible databases has enabled researchers worldwide to advance their research on DNA.

It is important to recognize that the study of DNA is an ongoing process, and there is still much to learn. New techniques and technologies continue to be developed, and ongoing research will undoubtedly lead to new discoveries and insights into the molecular basis of life. Nonetheless, the field of molecular biology and genetics owes a great debt to the pioneering work of Watson and Crick, who unlocked the secrets of DNA over half a century ago.

References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. DNA Structure and Replication. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK26821>

Bailey TL, Boden M, Buske FA, et al. (2009). "MEME Suite: tools for motif discovery and searching". Nucleic Acids Research. 37 (Web Server issue): W202–8.

Collins FS, Morgan M, Patrinos A. The Human Genome Project: Lessons from Large-Scale Biology. Science. 2003 Apr 25;300(5617):286-90. Available at: <https://pubmed.ncbi.nlm.nih.gov/12690187>

Kim EJ, Lee YS. DNA Conformation: Z-DNA, an Unusual DNA Structure. Molecules. 2019 Apr 26;24(9):1636.

Mathelier A, Fornes O, Arenillas DJ, et al. (2016). "JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles". Nucleic Acids Research. 44 (D1): D110–5.

Nakabachi, A., Yamashita, A., Toh, H., Ishikawa, H., Dunbar, H. E., Moran, N. A., & Hattori, M. (2006). The 160-kilobase genome of the bacterial endosymbiont Carsonella. Science, 314(5797), 267.

Available at: <https://www.science.org/doi/10.1126/science.1134196>

Pennisi, E. (2010). ScienceShot: Biggest Genome Ever Japanese flower has 50 times more DNA than humans do. Science 2010 Available at: [scienceshot-biggest-genome-ever](https://www.scienceshot.org/biggest-genome-ever)

Sinden RR. DNA Structure and Function. Academic Press; 2013.

Watson JD, Crick FHC. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. Nature. 1953 Apr 25;171(4356):737-8.

Available at: <https://www.nature.com/articles/171737a0>

Wingender E, Schoeps T, Haubrock M, et al. (2015). "TFClass: expanding the classification of human transcription factors to their mammalian orthologs". Nucleic Acids Research. 44 (D1): D300–4.

Wilkins, M. H. F., Stokes, A. R., & Wilson, H. R. (1953). Molecular structure of deoxypentose nucleic acids. Nature, 171(4356), 737-738.

Available at: <https://www.nature.com/articles/171738a0>

Abdelkrim, R. (2023). Bioinformatics: An Exciting Field of Science - Importance and Applications. Journal of Concepts in Structural Biology & Bioinformatics (JSBB), 1(4). Available at: [Bioinformatics An Exciting Field of Science](#)

Abdelkrim, R. (2023). Databases in Biology. Journal of Concepts in Structural Biology & Bioinformatics (JSBB), 1(4).

Available at: [Dbases In Biology](#)

Abdelkrim R., Khuphukile M. NALD: Nucleic Acids and Ligands Database, Modeling Approaches and Algorithms for Advanced Computer Applications (Springer) 2013, 488: 329-336 Available at: http://www.springerlink.com/openurl.asp?id=doi:10.1007/978-3-319-00560-7_36

NCBI. National Center for Biotechnology Information. [Online] Available at: <https://www.ncbi.nlm.nih.gov/>

NALD. Nucleic Acids and Ligands Database. [Online] Available at: <https://bioinformatics.univ-saida.dz/bit2/?arg=SB1>

PDB. Protein Data Bank. [Online]
Available at: <https://www.rcsb.org/>

The Genome Reference Consortium.
Available at: <https://www.ncbi.nlm.nih.gov/grc>

Ensembl Genome Browser.
Available at: <https://www.ensembl.org>

The Human Genome Variation Database.
Available at: <https://www.hgvd.genome.med.kyoto-u.ac.jp>

The Exome Aggregation Consortium.
Available at: <https://gnomad.broadinstitute.org/>