

Deoxyribonucleic acid (DNA)

Article by:

Wold, Marc S. Department of Biochemistry, University of Iowa College of Medicine, Iowa City, Iowa.

Rich, Alexander Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Weeks, Daniel Department of Biochemistry, University of Iowa, Iowa City, Iowa.

Lutter, Leonard C. Molecular Biology Research Program, Henry Ford Hospital, Detroit, Michigan.

Publication year: 2014

DOI: <http://dx.doi.org/10.1036/1097-8542.186500> (<http://dx.doi.org/10.1036/1097-8542.186500>)

Content

- [Structure](#)
- [Function](#)
- [Genetic variation](#)
- [Recombinant technology](#)
- [In the cell](#)
- [Bibliography](#)
- [Additional Readings](#)

A linear polymer made up of a specific sequence of deoxyribonucleotide repeating units linked by 3',5'-phosphodiester bonds; it is the carrier of genetic information. The set of DNA molecules that contains all genetic information for an organism is called its genome. DNA is found primarily in the nuclei of eukaryotic cells and in the nucleoid of bacteria. Small amounts of DNA are also found in organelles (such as mitochondria and chloroplasts) that contain their own genomes, in autonomously maintained DNAs called plasmids, and in viruses. *See also:* [Nucleic acid \(/content/nucleic-acid/460600\)](#)

DNA is composed of two long polymer strands of the sugar 2-deoxyribose, phosphate, and purine and pyrimidine bases. The backbone of each strand is composed of alternating 2-deoxyribose and phosphate linked together through phosphodiester bonds (**Fig. 1**). A DNA strand has directionality: each phosphate is linked to the 3' position of the preceding deoxyribose and to the 5' position of the following deoxyribose. The four bases found in DNA are adenine, thymine, guanine, and cytosine. Each 2-deoxyribose is linked to one of four bases in the 1' position via a covalent glycosidic bond. A base linked to a sugar is called a nucleoside, and a base linked to a sugar phosphate is called a nucleotide. The numbering of atom positions of the sugar is accompanied by a prime sign to distinguish it from the positional numbering of the bases. The sequence of these four bases allows DNA to carry genetic information. Bases can form hydrogen bonds with each other. Adenine forms two bonds with thymine, and cytosine forms three bonds with guanine. These two sets of base pairs have the same geometry, allowing DNA to maintain the same structure regardless of the specific sequence of base pairs (**Fig. 1**). *See also:* [Purine \(/content/purine/557800\)](#); [Pyrimidine \(/content/pyrimidine/558900\)](#)

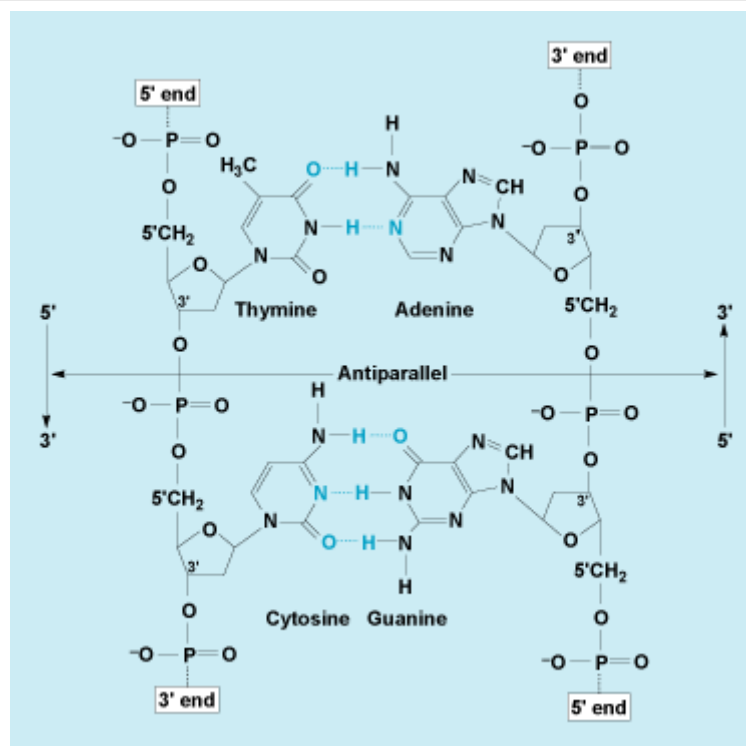


Fig. 1 Diagram of the DNA duplex. Each of the two strands comprises repeating units of a sugar–phosphate polymer. Each nucleotide on one strand is connected to the next by a phosphodiester bond. The linkage is always made from the 3' position of one sugar to the 5' position of the next. The two strands are aligned antiparallel to each other, with deoxyribose on the outside and purines and pyrimidines on the inside of the duplex. Purines and pyrimidines share hydrogen bonds, with adenine binding to thymine and cytosine binding to guanine. (*Reprinted with permission of John Wiley & Sons, Inc., from T. M. Devlin, Textbook of Biochemistry with Clinical Correlations, 4th ed., 2002. Copyright © Wiley-Liss, Inc.*)

Structure

DNA can be single-stranded or aligned to form two-, three-, or even four-stranded structures. It is commonly found in its double-stranded form in chromosomes. In double-stranded DNA, the two strands wrap around each other to form a double helix (**Fig. 2**). The two strands are held together by base pairing and are antiparallel. Thus, if one strand is oriented in the 5' to 3' direction, the other strand will be 3' to 5'. This double-helical structure of DNA was first proposed in 1954 by J. D. Watson and F. H. C. Crick.

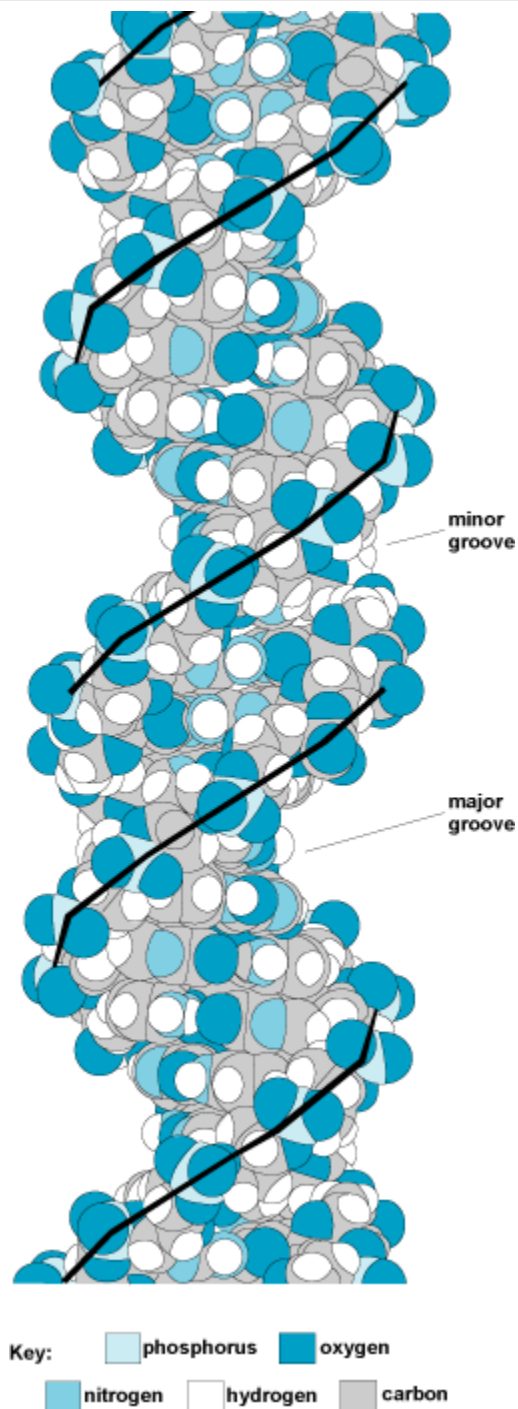
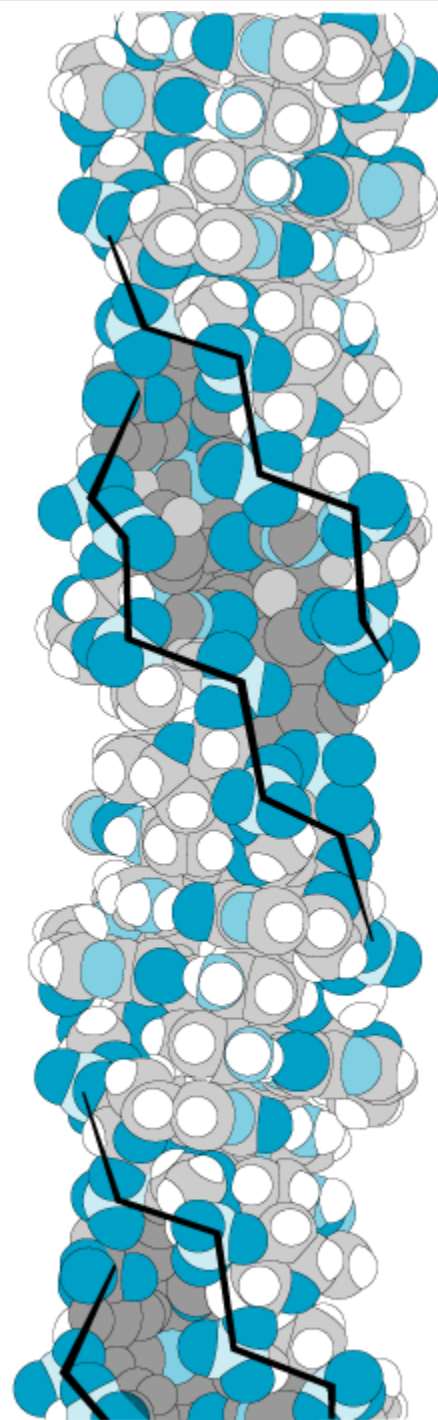


Fig. 2 B-DNA structure. The helical axis is vertical, and a solid line is drawn from phosphate to phosphate along the backbone.

The most common form of DNA is the B-form, which is a right-handed double helix with 10.4 base pairs per turn ([Fig. 2](#)). In a B-form double helix, the ribose–phosphate backbones run along the outside of the helix, and the base pairs are roughly perpendicular to the axis of the helix and lie just offset from the center axis of the helix. This offset causes one of the two grooves that run along the helix to be larger than the other—thus the designations, major groove and minor groove. Proteins that bind to DNA usually bind specifically to one of the two grooves. Less common forms of DNA include the A-form, which is a right-handed double helix that has 11 base pairs per turn and has a wider diameter than the B-form, and the Z-form, which is a narrow, irregular left-handed double helix ([Fig. 3](#)).



Key:

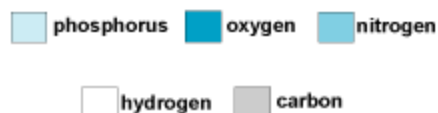


Fig. 3 Z-DNA structure. Darker gray area is the deep, narrow minor groove. A solid line is drawn from phosphate to phosphate along the backbone.

Right-handed conformation

Following Watson and Crick's proposal in 1954, experiments supported a right-handed form of the double helix in which there were approximately 10 base pairs per turn. In this form, each base pair is similar to the neighboring base pair, except that there is a rotation of approximately 36° between the base pairs. The conformation of DNA was determined by x-ray diffraction studies of fibers. A dried DNA fiber produces a type A pattern, in which the base pairs are tilted somewhat away from the

helix axis. However, a wet or hydrated DNA fiber produces a B-type pattern, in which the bases are perpendicular to the helix axis ([Fig. 2](#)).

Left-handed conformation (Z-DNA)

In 1979, it was discovered that DNA can also form a left-handed double helix with the same hydrogen bonding between the two bases ([Fig. 1](#)). This was discovered by solving the structure of crystals of short DNA fragments; crystals can yield a more detailed view of the molecule. A solid line drawn between the phosphates in the left-handed conformation follows a zigzag path ([Fig. 3](#)); hence it is called Z-DNA. In contrast to B-DNA, Z-DNA has only one deep helical groove extending almost to the axis of the helix. The base pairs in left-handed Z-DNA are located near the outside of the helix. B-DNA has a one-base-pair repeating unit, whereas Z-DNA has a dinucleotide repeating pattern. Both forms of DNA have the two sugar–phosphate backbones oriented in opposite directions ([Fig. 1](#)). Z-DNA has been identified in DNA fibers as well as inside living cells.

Conformational changes

In biological systems, DNA is normally subjected to a torsional strain that tends to unwind the double helix. This torsional strain tends to untwist right-handed B-DNA and sometimes is great enough to stabilize certain regions of DNA in the left-handed Z-DNA conformation. The conversion from right-handed B-DNA to left-handed Z-DNA is associated with changes in the relationship of the bases to the sugar backbone. Every other base in Z-DNA is rotated about the bond connecting it to the sugar. This altered form is common in adenine or guanine residues (purines), but less common in thymine or cytosine residues (pyrimidines); thus, sections that form Z-DNA often have alternations of purines and pyrimidines.

Although right-handed B-DNA is the most common form of DNA, left-handed Z-DNA occurs in a number of places and may play an important role in the regulation of gene expression. In addition, Z-DNA appears to be an important component of the process of crossing-over or recombination of DNA molecules. Z-DNA is important in carrying out special functions that are in general associated with special proteins that bind to Z-DNA and not to B-DNA. See also: [Crossing-over \(genetics\)](#) ([/content/crossing-over-genetics/168850](#)); [Gene](#) ([/content/gene/284400](#))

Marc S. Wold
Alexander Rich
Daniel L. Weeks

Function

For cells to live and grow, the genetic information in DNA must be (1) propagated and maintained from generation to generation and (2) expressed to synthesize the components of a cell. These two functions are carried out by the processes of DNA replication and transcription, respectively. See also: [Genetic code](#) ([/content/genetic-code/284900](#))

Replication

Each of the two strands of a DNA double helix contains all of the information necessary to make a new double-stranded molecule ([Fig. 4](#)). During replication, the two parental strands are separated, and each is used as a template for the synthesis of a new strand of DNA. Base incorporation is directed by the existing DNA strand; nucleotides that base-pair with the template are added to the nascent DNA strand. The product of replication is two complete double-stranded helices, each of which contains all of the genetic information (has the identical base sequence) of the parental DNA. Each progeny double helix is composed of one parental and one nascent strand. DNA replication is very accurate. In bacteria, the mutation rate is about 1 error per 1000 bacteria per generation, or about 1 error in 10^9 base pairs replicated. This low error rate is due to a combination of the high accuracy of the replication process and cellular pathways which repair misincorporated bases. See also: [DNA repair](#) ([/content/dna-repair/900107](#))

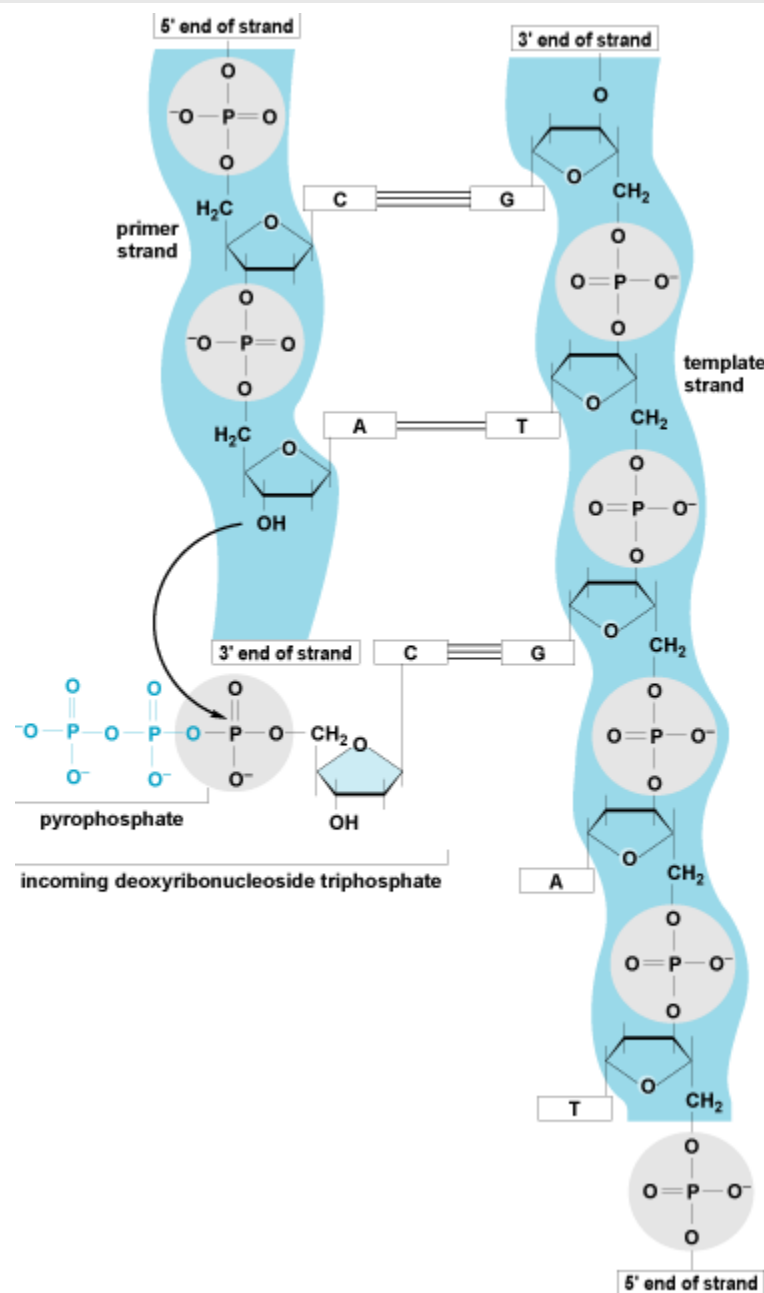


Fig. 4 Replication of DNA. The replication shown here highlights the selection of new base insertion by using Watson-Crick (A-to-T and G-to-C) base pairing, the addition of new bases to the 3' end of an existing chain, and the attack of the 3'OH on the alpha phosphate of a nucleoside triphosphate to add the newest base. (Copyright © 2002. From *Molecular Biology of the Cell*, 4th ed., by B. Alberts et al. Reproduced with permission of Garland Science/Taylor & Francis LLC)

Synthesis of the nascent DNA strands is carried out by a family of enzymes called DNA polymerases. These enzymes synthesize a new phosphodiester bond between the 3' hydroxyl position of the terminal deoxyribose of an existing DNA (or ribonucleic acid, RNA) chain and a deoxynucleotide triphosphate. All DNA polymerases (1) synthesize new DNA in a 5' to 3' direction, (2) require a primer (they must add on to the 3' hydroxyl position of DNA or RNA), and (3) require a DNA molecule as a template. DNA polymerases, in conjunction with multiple accessory proteins, can rapidly synthesize new DNA strands. (The mandatory participation of the 3' hydroxyl forms the rationale for the use of nucleoside analogues that lack a 3' hydroxyl as antiviral and anticancer drugs, because they lead to premature chain termination.) A bacterial replication fork moves at a rate of approximately 1000 nucleotides per second; in eukaryotic cells, forks move at a rate of about 100 nucleotides per second. See also: [Mutation \(/content/mutation/441200\)](#)

Transcription

In transcription, DNA acts as a template directing the synthesis of RNA. RNA is a single-stranded polymer similar to DNA, except that it contains the sugar ribose instead of 2-deoxyribose, and the base uracil instead of thymine (**Fig. 5**). The two strands of DNA separate transiently, and one of the two single-stranded regions is used as a template to direct the synthesis of an RNA strand. As in DNA replication, base pairing between the incoming ribonucleotide and the template strand determines the sequence of bases incorporated into the nascent RNA. Thus, genetic information in the form of a specific sequence of bases is directly transferred from DNA to RNA in transcription. After the RNA is synthesized, the DNA reverts to a double-stranded form. Transcription is carried out by a family of enzymes called RNA polymerases. In prokaryotes, RNA polymerase interacts directly with the sites at which RNA transcription starts (promoters); in contrast, in eukaryotes, large complexes of proteins bind to promoters and then direct RNA polymerase to start RNA synthesis. Newly synthesized RNA is ultimately used to direct protein synthesis by ribosomes in a process called translation. See also: **Protein (/content/protein/550200)**; **Ribonucleic acid (RNA) (/content/ribonucleic-acid-rna/589000)**; **Ribosomes (/content/ribosomes/589200)**; **Transcription (/content/transcription/800940)**

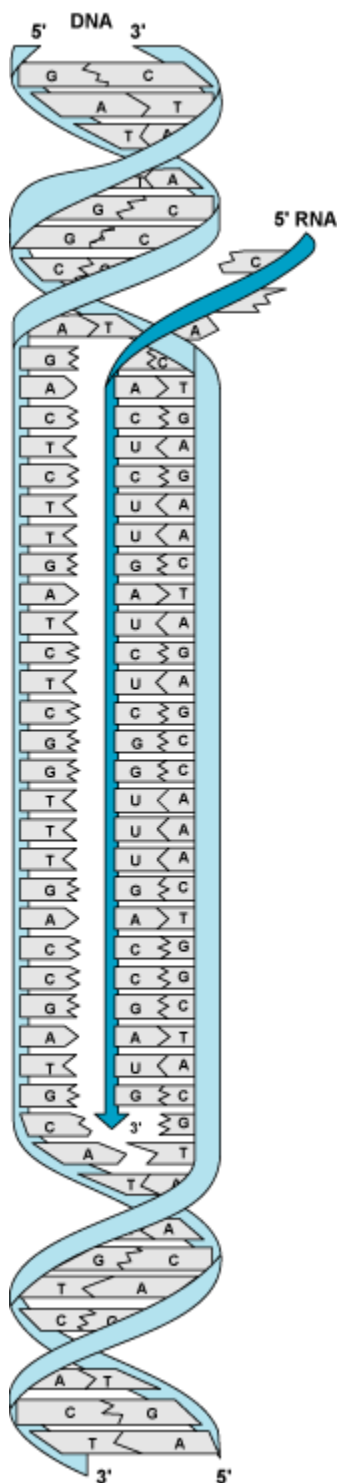


Fig. 5 Transcription of DNA. The double-stranded DNA is transiently separated to allow the synthesis of RNA. Using a mechanism similar to that discussed for DNA replication (Fig. 4), ribonucleotides are added to the 3' end of a growing chain, and the selection of bases to insert is guided by Watson-Crick base pairing. (Reprinted with permission from G. M. Cooper, *The Cell: A Molecular Approach*, 2d ed., 2000. Copyright © Sinauer Associates, Inc.)

Genetic variation

There is a great deal of variation in the DNA content and sequences in different organisms. Because of base pairing, the ratios of adenine to thymine and cytosine to guanine are always the same. However, the ratio of adenine and thymine to guanine and cytosine in different organisms ranges from 1:4 to 3:4. There is also large variation in the amount of DNA in the genome of various organisms. The simplest viruses have genomes of only a few thousand base pairs, while complex eukaryotic organisms have genomes of billions of base pairs. (The lungfish has one of the largest genomes known, about 100

billion base pairs.) This variation partially reflects the increasing number of genes necessary to encode more complex organisms, but mainly reflects an increase in the amount of DNA that does not encode proteins. A large percentage of the DNA in multicellular eukaryotes is noncoding or is repetitive DNA (sequences that are repeated many times). This DNA can serve structural or regulatory functions, it can be related to transposable elements (DNA sequences that can insert into other unrelated sequences), or in many cases it has no known function. Rarely, genes are present in multiple copies to allow high levels of expression of certain cellular components. For example, most cells have hundreds or thousands of copies of the genes encoding the components needed for protein synthesis to ensure sufficient levels of expression of these components.

In most eukaryotes, the DNA sequences that encode proteins are not continuous but have other sequences interspersed within them. The sequences that are expressed in protein are called exons, and the interspersed sequences are called introns. The initial transcript synthesized by RNA polymerase contains both exons and introns and can be many times the length of the actual coding sequence. The RNA is then processed, and the introns are removed through a mechanism called RNA splicing to yield messenger RNA (mRNA), which is translated to make protein. See also: [Exon \(/content/exon/248350\)](#); [Intron \(/content/intron/350450\)](#)

Recombinant technology

Techniques have been developed to allow DNA to be manipulated in the laboratory. These techniques, collectively called molecular biology or molecular genetics, have led to a revolution in biotechnology. This revolution began when methods were developed to cleave DNA at specific sequences and to join pieces of DNA together. Another major component of this technology is the ability to determine the sequence of the bases in DNA. There are two general approaches for determining DNA sequence: either chemical reactions are carried out that specifically cleave the sugar–phosphate bond at sites which contain a certain base, or DNA is synthesized in the presence of modified bases that cause termination of DNA synthesis after the incorporation of a certain base. A series of four reactions, one for each base, are carried out, and the lengths of the DNA fragments produced identify the sequence of the bases. These methods can now be automated so that it is practical to determine the DNA sequences of the entire genome of an organism. The complete sequences of humans and a number of experimentally important organisms have been determined. See also: [Genetic mapping \(/content/genetic-mapping/285200\)](#); [Human genome \(/content/human-genome/757575\)](#)

Another example of recombinant DNA technology is copy DNAs (cDNAs), which are made when an mRNA is converted to DNA using reverse transcriptase (a specialized DNA polymerase that can use RNA as a template). The resulting DNA can then be cloned and made in large quantity, and its sequence determined. Because it is much easier to manipulate DNA than RNA, cDNAs allow any RNA-based species to be characterized. See also: [Biotechnology \(/content/biotechnology/084350\)](#); [Genetic engineering \(/content/genetic-engineering/285000\)](#); [Molecular biology \(/content/molecular-biology/430300\)](#); [Reverse transcriptase \(/content/reverse-transcriptase/585000\)](#)

Marc S. Wold

Daniel L. Weeks

In the cell

The full genome of DNA must be substantially compacted to fit into a cell. For example, virtually every cell in the human body contains a full genome of DNA, which in turn has a total length of about 3 m (10 ft). However, this DNA must fit into a nucleus with a diameter of 10^{-5} m. This immense reduction in length is accomplished in eukaryotes via multiple levels of compaction in a nucleoprotein structure termed chromatin. To summarize, the first level involves spooling about 200 base pairs of DNA onto a complex of basic proteins called histones to form a nucleosome. Nucleosomes are connected like beads on a string ([Fig. 6](#)) to form a 10-nanometer-diameter fiber, and this is further coiled to form a 30-nm fiber. The 30-nm fibers are further

coiled and organized into loops formed by periodic attachments to a protein scaffold. This scaffold organizes the complex into the shape of the metaphase chromosome seen at mitosis.

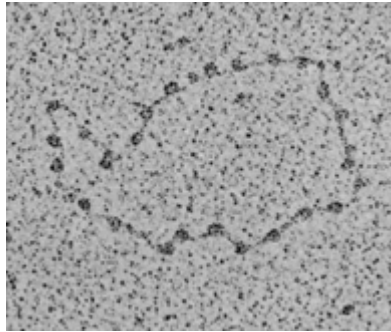


Fig. 6 Electron micrograph of the minichromosome of a virus that infects monkey cells. The circular viral DNA is organized in nucleosome “beads.” In order to improve the ability to visualize individual nucleosomes, the conditions of microscopy in this figure have been adjusted to stretch the minichromosome slightly. In the cell, successive nucleosomes are in direct contact with one another, resulting in a circle of a relatively uniform fiber of about 11 nm diameter. (From L. Lutter et al., *Mol. Cell. Biol.*, 12:5004–5014, 1992)

The nucleosome is the fundamental structural unit of DNA in all eukaryotes. Virtually all of the DNA in each cell of an organism is packaged into nucleosomes. Nucleosomes are squat cylinders (6 nm × 11 nm × 11 nm) arranged as repeating subunits along the DNA at approximately 200-base-pair intervals. Pairs of four different histone proteins (H2A, H2B, H3, and H4) constitute an octamer complex, on the outside of which is wrapped 146 base pairs of DNA to form the nucleosome core. The DNA in the core is wrapped in 1.75 left-handed coils, resulting in approximately sixfold compaction of the DNA. The detailed structure of the nucleosome core (**Fig. 7**) has been determined by x-ray crystallography. Nucleosome cores are connected to one another by about 55 base pairs of linker DNA. The 146-base-pair length of DNA in the nucleosome core is invariant among all species studied, but the length of linker DNA can range from 20 to 100 base pairs among different species and different tissues. The reason for this variation is unknown. The linker DNA binds a fifth histone, H1, that plays a role in condensation of the 10-nm fiber into the higher-order structures.

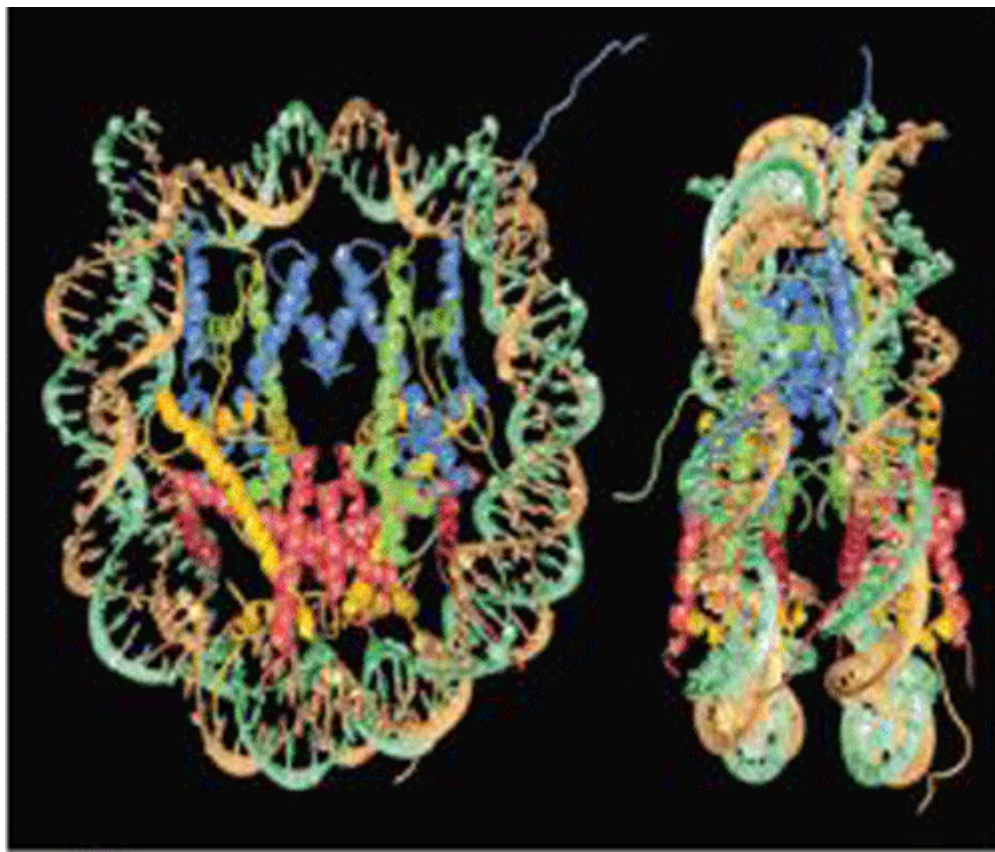


Fig. 7 Structure of the nucleosome core particle as determined by x-ray crystallography. On the left is a view down the superhelix axis, showing the DNA wrapped on the outside of the histone octamer. On the right is a view from the side, showing the two gyres of superhelix of 146 base pairs of DNA. (From K. Luger et al., *Nature*, 389:251–260, 1997)

Nucleosomes reduce the accessibility of DNA to DNA-binding proteins such as polymerases and other protein factors essential for transcription and replication. One level of regulation of gene expression occurs via posttranslational modification of histones. For example, specific protein factors activate transcription by adding acetyl groups to specific amino acid residues in the histones, while other factors with the opposite transcription effect remove those residues. Malfunctions in these regulatory processes have been associated with human diseases. There are several other types of histone posttranslational modifications with regulatory effects. Thus, chromatin structure not only compacts the genome but also affords an additional level of control of gene expression. See also: [Chromosome \(/content/chromosome/134900\)](#); [Histone \(/content/histone/461000\)](#); [Nucleosome \(/content/nucleosome/461100\)](#)

Leonard C. Lutter

Bibliography

- B. Alberts et al., *Molecular Biology of the Cell*, 5th ed., Garland, New York, 2008
- G. M. Cooper and R. E. Hausman, *The Cell: A Molecular Approach*, 3d ed., Sinauer Associates, Sunderland, MA, 2004
- T. M. Devlin, *Textbook of Biochemistry with Clinical Correlations*, 7th ed., Wiley-Liss, New York, 2010
- H. F. Judson, *The Eighth Day of Creation: Makers of the Revolution in Biology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996
- A. Kornberg and T. A. Baker, *DNA Replication*, 2d ed., Freeman, New York, 1992
- H. Lodish et al., *Molecular Cell Biology*, 6th ed., Freeman, New York, 2008
- J. D. Watson, *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*, Touchstone, New York, 1996

J. D. Watson et al., *Recombinant DNA*, 3d ed., Freeman, New York, 2007

Additional Readings

P. Berg and J. E. Mertz, Personal reflections on the origins and emergence of recombinant DNA technology, *Genetics*, 184(1):9–17, 2010 DOI: [10.1534/genetics.109.112144](https://doi.org/10.1534/genetics.109.112144) (<http://dx.doi.org/10.1534/genetics.109.112144>)

S. B. Hake, A. Xiao, and C. D. Allis, Linking the epigenetic “language” of covalent histone modifications to cancer, *Brit. J. Cancer*, 90(4):761–769, 2004 DOI: [10.1038/sj.bjc.6601575](https://doi.org/10.1038/sj.bjc.6601575) (<http://dx.doi.org/10.1038/sj.bjc.6601575>)

K. Luger, Structure and dynamic behavior of nucleosomes, *Curr. Opin. Genet. Develop.*, 13(2):127–135, 2003 DOI: [10.1016/S0959-437X\(03\)00026-1](https://doi.org/10.1016/S0959-437X(03)00026-1) ([http://dx.doi.org/10.1016/S0959-437X\(03\)00026-1](http://dx.doi.org/10.1016/S0959-437X(03)00026-1))

J. Michael and S. Sircar (eds.), *Fundamentals of Medical Physiology*, Thieme Medical Publishers, New York, 2011

H. S. Stoker, *General, Organic, and Biological Chemistry*, 5th ed., Cengage Learning, Belmont, CA, 2011

H. S. Stoker, *Organic and Biological Chemistry*, 6th ed., Cengage Learning, Belmont, CA, 2013

[National Center for Biotechnology Information \(http://www.ncbi.nlm.nih.gov/\)](http://www.ncbi.nlm.nih.gov/)