Nucleic acid

Contributed by: Edward Johnson, Mary Lou Pardue

Publication year: 2014

An acidic, chainlike biological macromolecule consisting of multiply repeated units of phosphoric acid, sugar, and purine and pyrimidine bases. Nucleic acids as a class are involved in the preservation, replication, and expression of hereditary information in every living cell. There are two types: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

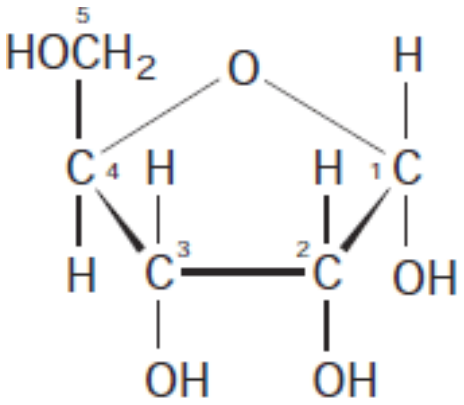
Deoxyribonucleic Acid

Each DNA strand is a long polymeric molecule consisting of many individual nucleotides linked end to end. The great size and complexity of DNAs are indicated by their molecular weights, which commonly range in the hundreds of millions. DNA is the chemical constituent of the genes of an organism, and is thus the ultimate biochemical object of the study of genetics. Information is contained in the DNA in the form of the sequence of nucleotide building blocks in the nucleic acid chain.

Nucleotides

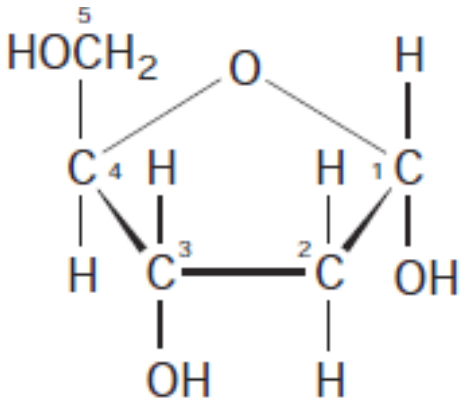
The number of nucleotide building blocks in DNA is relatively small—only four nucleotides constitute the vast majority of DNA polymeric units. These are deoxyadenylic, deoxyguanylic, deoxycytidylic, and deoxythymidylic acids. For purposes of brevity, these nucleotides are symbolized by the letters A, G, C, and T, respectively. Each of these nucleotides consists of three fundamental chemical groups: a phosphoric acid group, a deoxyribose

5-carbon sugar group (**1b**), and a nitrogenous base which is a derivative of either purine (**2**) or pyrimidine (**3**).



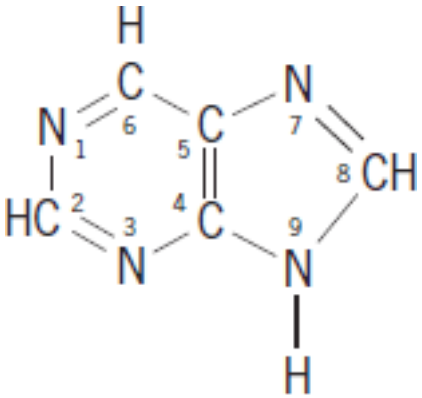
D-Ribose

(1a)

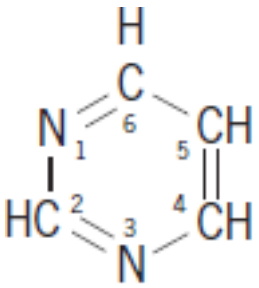


2-Deoxy-D-ribose

(1b)

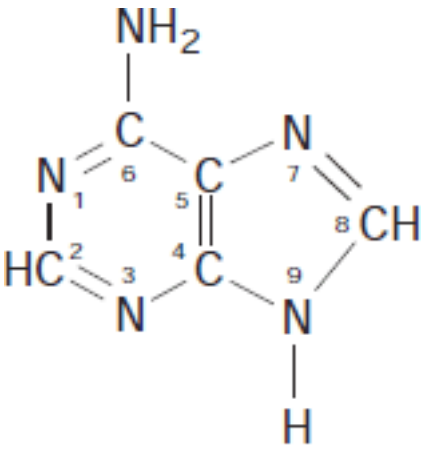


(2)

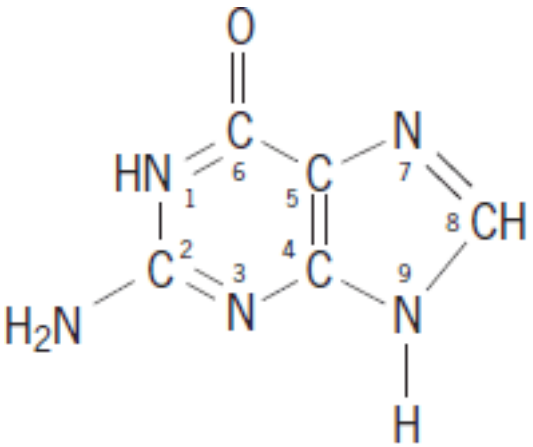


(3)

Some nucleotides contain the purine groups adenine (6-aminopurine; **4**) or guanine (2-amino-6-oxypurine; **5**),

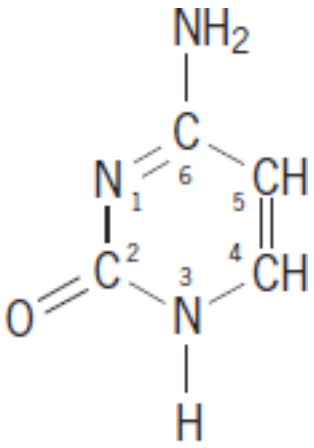


(4)

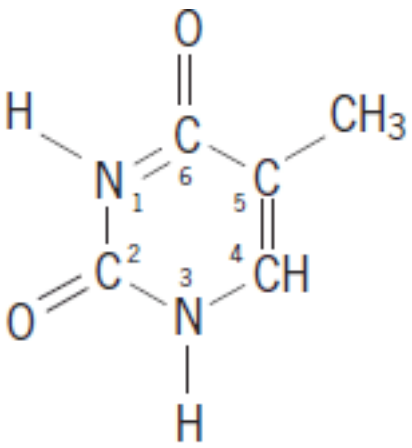


(5)

and some contain the pyrimidine groups cytosine (2-oxy-5-aminopyrimidine; **6**) or thymine (2,6-dioxy-6-methyl-pyrimidine; **7**).

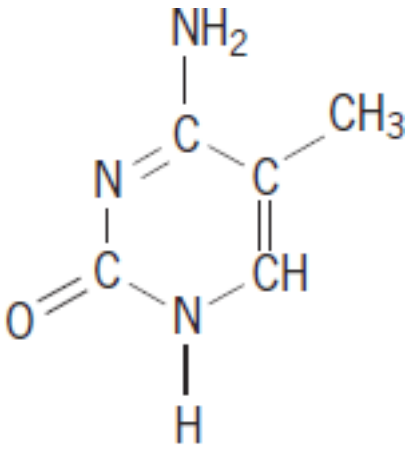


(6)

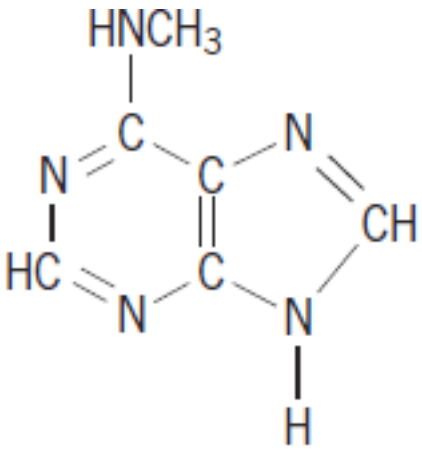


(7)

These are the only major bases found in most DNA, although in specific sequences certain methylated derivatives of these bases, such as 5-methyl cytosine (**8**) or N^6-methyl adenine (**9**),

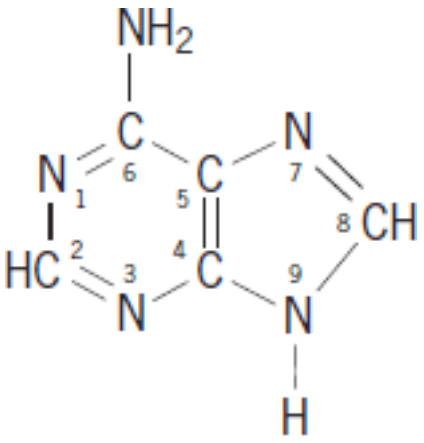


(8)

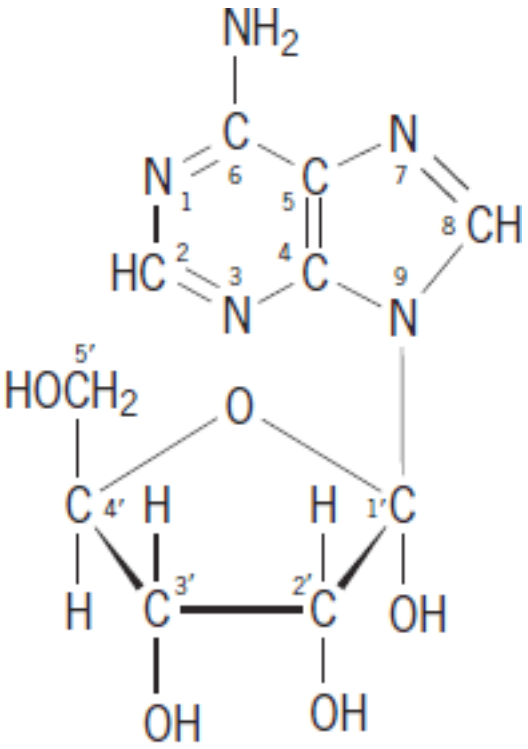


(9)

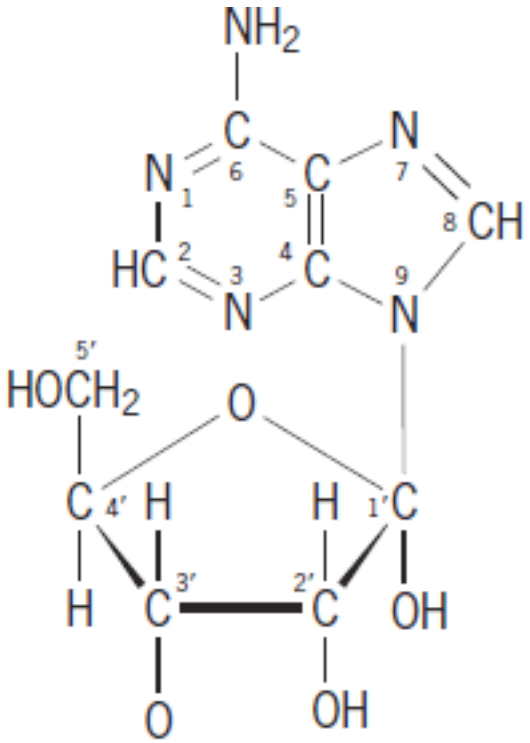
can also be detected. In each nucleotide, these subunits are linked together in the following order: purine or pyrimidine base–ribose sugar–phosphoric acid (**10**)–(**12**).



(10)



(11)



(12)

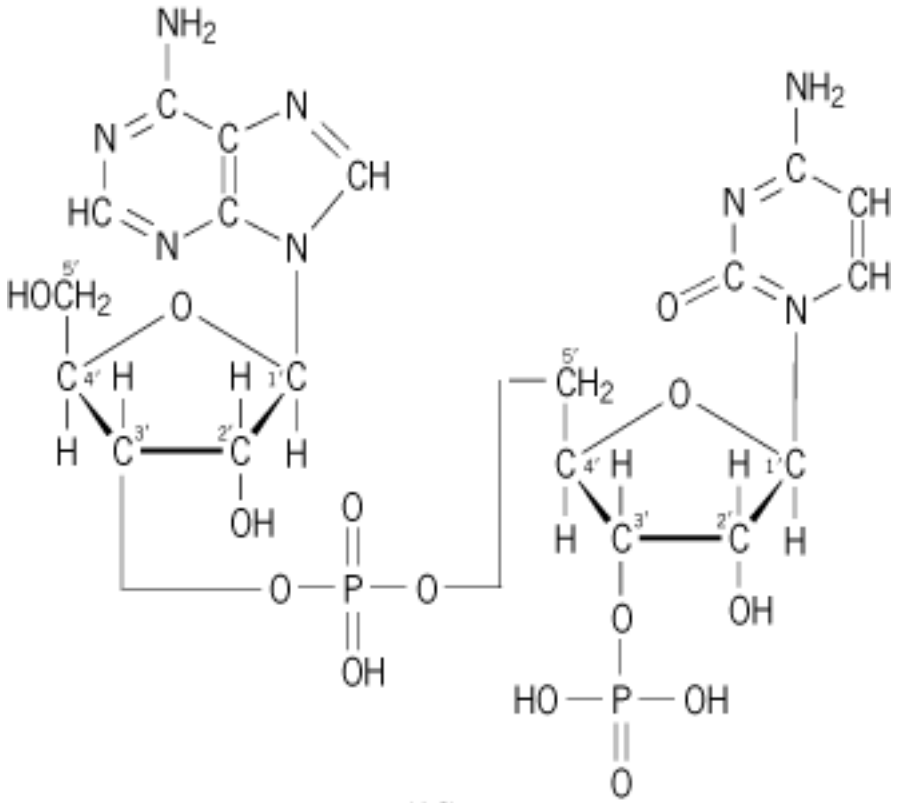
Removal of the phosphoric acid group leaves a base-sugar compound which is called a nucleoside (**11**). In each nucleoside, the base is attached to the sugar through a bond from nitrogen at position 9 (for purines) or 3 (for pyrimidines) to carbon at position 1 of the sugar ring. The nucleosides and nucleotides are named for the base that they contain as follows:

Base	*Nucleoside*	*Nucleotide*
Adenine (**10**)	Adenosine (**11**)	Adenylic acid (**12**)
Guanine	Guanosine	Guanylic acid
Cytosine	Cytidine	Cytidylic acid
Thymine	Thymidine	Thymidylic acid
Uracil	Uridine	Uridylic acid

It is necessary to denote the position of the phosphoric acid residue when describing nucleotides. Nucleotides synthesized by cells for use as building blocks in nucleic acids all have phosphoric acid residues coupled to the 5′ position of the ribose sugar ring, as shown for deoxyadenylic acid (deoxyadenosine-5′-phosphate; **12**). Upon

hydrolysis of DNA, however, nucleotides can be produced which have phosphoric acid coupled to the 3′ position of the sugar ring.

When DNA is hydrolyzed by using the enzyme deoxyribonuclease I, prepared from bovine pancreas, the principal products are oligonucleotides ending with 3′-hydroxyl groups and 5′-phosphoric acid groups. In contrast, when DNA is hydrolyzed by using the enzyme micrococcal nuclease, prepared from *Staphylococcus aureus*, the principal products are nucleotides or oligonucleotides ending with 3′-phosphoric acid groups and 5′-hydroxyl groups. Studies such as these led very early to the conclusion that in intact DNA the nucleotides were linked via phosphoryl groups which join the 3′ position of one sugar group to the 5′ position of the next sugar group (**13**).



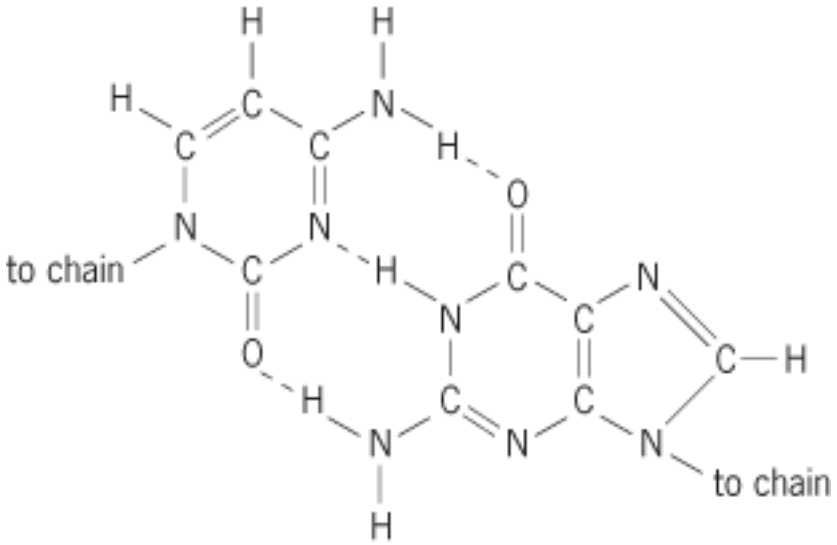
(13)

This 5′-to-3′ linkage of nucleotides imparts a polarity to each DNA strand which is an important factor in the ability of DNA to form the three-dimensional structures necessary for its ability to replicate and to serve as a genetic template. *See also:* NUCLEOTIDE.

Helix

In 1953, James Watson and Francis Crick proposed a double-helical structure for DNA based largely on x-ray diffraction studies of Maurice Wilkins and Rosalind Franklin. It had been known earlier that in most DNAs the ratios of A to T and of G to C are approximately 1:1. The Watson-Crick model attributes these ratios to the

phenomenon of base pairing, in which each purine base on one strand of DNA is hydrogen-bonded to a complementary pyrimidine base in an opposing DNA strand (**14**).



(14)

The x-ray diffraction studies suggested that DNA can assume a structure called the B form, which is a right-handed helical configuration resembling a coiled spring. In most DNAs, there are two single DNA strands which compose each helix. The strands wind about each other, with their sugar-phosphate chains forming the coil of the helix and with their bases extending inward toward the axis of the helix. The configuration of the bases allows hydrogen bonding between opposing purines and pyrimidines. Each of the base pairs lies in a plane at approximately right angles to the helix axis, forming a stack with the two sugar-phosphate chains coiled around the outside of the stack. In this configuration, a wide (major) groove exists on the surface of the helix, winding parallel to a narrow (minor) groove (**Fig. 1**). In the Watson-Crick model, the two opposing DNA strands have an opposite polarity; that is, they are antiparallel, with their 5′ ends lying at opposite ends of each double-stranded molecule. The right-hand helix completes one turn every 10 nucleotides, and bases are spaced 0.34 nanometer apart. The width of the double-stranded helix in the B form averages about 2 nm. DNA can exist in helical structures other than the B form. One configuration, termed the Z form, is a left-handed helical structure with 12 nucleotides per turn. The Z form can exist in DNA sequences with alternating guanine and cytosine bases and may be functional in localized DNA regions, although the B form is thought to predominate in most biological systems.

Due to the phenomenon of base pairing, the two DNA strands of a helix can be “melted” apart (denatured) or, once denatured, allowed to reanneal. Heating a DNA duplex in solution breaks the hydrogen bonds involved in base pairing and allows the two single strands of the helix to dissociate. This melting is dependent on the base sequence in the DNA, since there are two hydrogen bonds between A-T pairs and three hydrogen bonds between G-C pairs (**14**). In a solution of sheared DNA molecules, the rate of reannealing of denatured DNA depends on

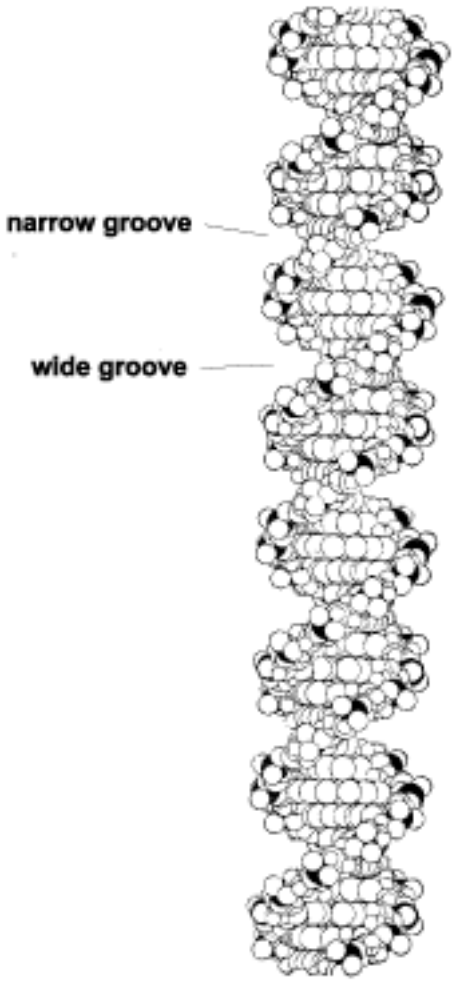


Fig. 1 **Three-dimensional configuration of DNA.**

the extent to which the denatured DNA sequences are reiterated. Studies on the kinetics of renaturation of DNA from various sources have shown that different organisms differ widely in the complexity of sequences in their DNA. In general, cells of higher organisms contain more unique or seldom-repeated DNA sequences than do cells of lower organisms. This reflects in part the fact that higher organisms possess more individual genes required to code for the larger number of proteins they contain. Renaturation studies also show that many cells possess highly reiterated DNA sequences that may constitute a significant percentage of the total genome. When sheared cellular DNA sequences are separated according to their buoyant density, such as by centrifugation in density gradients of cesium chloride, many of these highly reiterated sequences can be detected as satellite DNA bands; that is, they form bands of DNA at positions in the gradient which differ from that of the bulk of cellular DNA. Several of the known satellite DNAs contain genes which are present in multiple copies in cells. For example, the genes coding for RNAs of ribosomal subunits of cells are characteristically located in satellite DNA bands. At this point, however, the functional significance of most highly reiterated DNA sequences in cells is not known.

C value

The amount of DNA in the haploid genome of a given organism (called the C value) is only loosely correlated with the degree of evolutionary advancement of the organism. There are about 2×10^{-16} g of DNA in a bacteriophage, as compared to about 10^{-14} g in the bacterium *Escherichia coli* and about 3×10^{-12} g in rat liver cells. Whereas mammalian cells contain about $2–3 \times 10^{9}$ nucleotide pairs of DNA, amphibian cells vary widely, ranging from less than 2×10^{9} to about 1.5×10^{11} nucleotide pairs. Consideration of these figures leads to what is known as the C value paradox: generally cells of higher organisms contain much more DNA than is necessary to code for the number of proteins they contain. This paradox is not explained simply by different degrees of gene reiteration in different organisms. Instead, it is more likely that only a fraction of the total DNA codes for proteins in higher organisms and that the relative amount of noncoding DNA is highly variable. It is now known that many noncoding DNA sequences exist adjacent to RNA-coding sequences and serve in a regulatory capacity. In addition, noncoding DNA sequences, termed intervening sequences, may exist in the middle of sequences coding for certain RNA species.

Sequences

The sequence of nucleotide pairs in the DNA determines all of the hereditary characteristics of any given organism. The DNA acts as a template which is copied, through the process of transcription, to make RNA. The RNA in turn serves as a template in a process by which its encoded information is translated to determine the amino acid sequences of proteins. Each amino acid in a protein chain is specified by a triplet of nucleotides (in RNA) or nucleotide pairs (in DNA) known as a codon. The set of correlations between the amino acids and their specifying codons is called the genetic code. Each gene which codes for a protein thus contains a sequence of triplet codons which corresponds to the sequence of amino acids in the polypeptide. This sequence of codons may be interrupted by intervening DNA sequences so that the entire coding sequence is not continuous. In addition to coding sequences, there also exist regulatory sequences, which include promoter and operator sequences involved in initiating gene transcription and terminator sequences involved in stopping transcription. Regulatory sequences are not necessarily made up of triplets, as are the codons. In order to study the regulation of a given gene, it is necessary to determine its nucleotide sequence. *See also:* GENETIC CODE.

There are several methods for sequencing DNA. Most of these methods employ radioactive end-labeling of one or both DNA strands, followed by either cleavage or extension of labeled strands to produce end-labeled fragments which terminate at nucleotides with specific bases. For example, one commonly used method involves labeling the 5′ end of a DNA strand with ^{32}P-phosphate by using the enzyme polynucleotide kinase and γ-^{32}P-adenosine triphosphate as a phosphate donor. Procedures are then used to induce base-specific chemical cleavage of the end-labeled strands at each of the four nucleotides in turn. Polyacrylamide gel electrophoresis is then used to size the radioactive fragments, and the sequence of nucleotides from the labeled 5′ end can be deduced.

The complete DNA sequences of several different genes, together with adjacent regulatory sequences, are now known. The first entire genome of any organism to be sequenced was that of the single-strand DNA phage φX174. This sequence of 5386 nucleotides was worked out by Fred Sanger and coworkers. One interesting aspect of this sequence is that it reveals the presence of overlapping genes coding for proteins. A single nucleotide sequence can code for more than one amino acid sequence, depending on the phasing with which the sequence is grouped into triplet codons during the process of protein synthesis. Specific start and stop signal codons are required to specify the phasing.

Figure 2*a* shows an electron micrograph of a single DNA molecule containing the genes coding for ribosomal RNAs of the slime mold *Physarum polycephalum*. The coding sequences for two RNA species, 19S and 26S, are revealed by R-loop hybridization. (This is a process by which RNA is annealed with double-stranded DNA, thus displacing a single DNA strand.) In order to prepare the nucleic acids for visualization, they are first coated with a layer of electron-dense heavy metals—in this case, platinum and palladium. **Figure 3** shows the DNA sequence near the start point of transcription of the gene coding for the enzyme β-galactosidase in the bacterium *E. coli*. This sequence contains short segments which are important for binding and subsequent action of enzymes that polymerize RNA. Regions for binding of proteins (CAP protein, RNA polymerase, and repressor protein) are shown on the DNA map. The *i* gene codes for the repressor protein which binds to the operator region and regulates transcription. The CAP protein must be bound to the DNA in order for the polymerase to initiate properly. The CAP protein binds its site only in the presence of the compound cyclic adenosine monophosphate. RNA polymerase binds a site about 40 nucleotides long and, in the absence of repressor, begins transcription of messenger RNA. The *z* gene codes for the enzyme β-galactosidase. The DNA sequence of one strand of the polymerase binding region is shown at bottom. The entry site and Pribnow's box sequences are important recognition sequences for polymerase binding and initiation. *See also:* OPERON.

Enzymes called restriction endonucleases cleave DNA at specific sequences. It is possible, by using these enzymes and DNA ligases, to splice an exogenous gene segment into an existing DNA sequence. The resulting structure is known as recombinant DNA. When the recombinant includes the genome of a plasmid or phage, this spliced DNA molecule can be propagated in bacteria. This process is known as gene cloning. Through this process, genes which are present only in a few copies in mammalian cells can be grown in vast quantities in bacteria. Cloning is industrially important for the production of certain gene products, such as some hormones. *See also:* GENETIC ENGINEERING.

Replication

Since DNA is the substance containing the chemical code governing the heredity of all cells, it is clear that its biosynthesis and passage from generation to generation must be extremely precise. The way in which a new DNA molecule arises with the same nucleotides arranged in the same sequence as in the parent DNA molecule is one of the most intriguing problems in biochemistry.

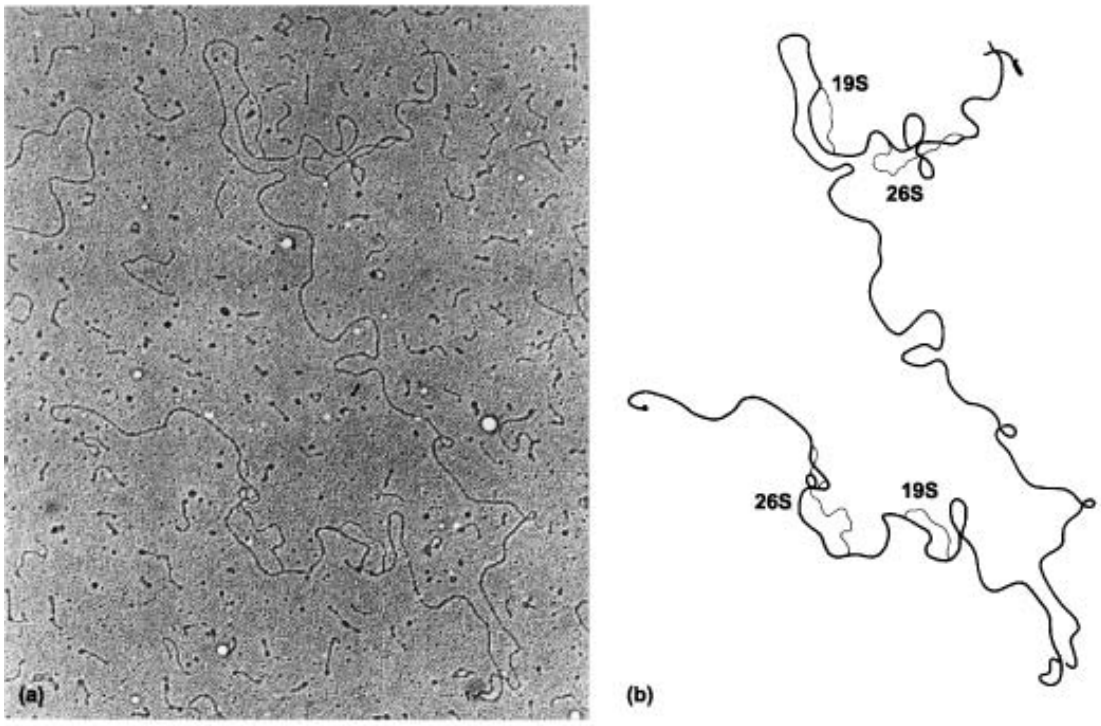


Fig. 2 Ribosomal RNA genes of *Physarum polycephalum*: (*a*) electron micrograph and (*b*) map, showing 19S and 26S RNA coding regions as R-loop hybrids. This DNA molecule is 60,000 base pairs long, or about 20×10^{-6} m. (*Courtesy of G. R. Campbell, V. C. Littau, P. W. Melera, V. G. Allfrey, and E. M. Johnson*)

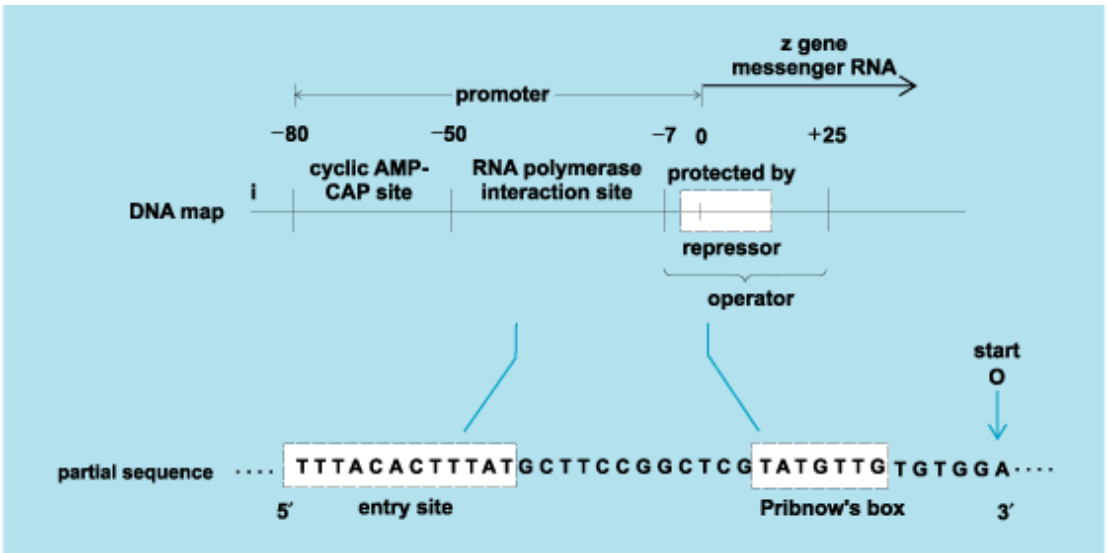
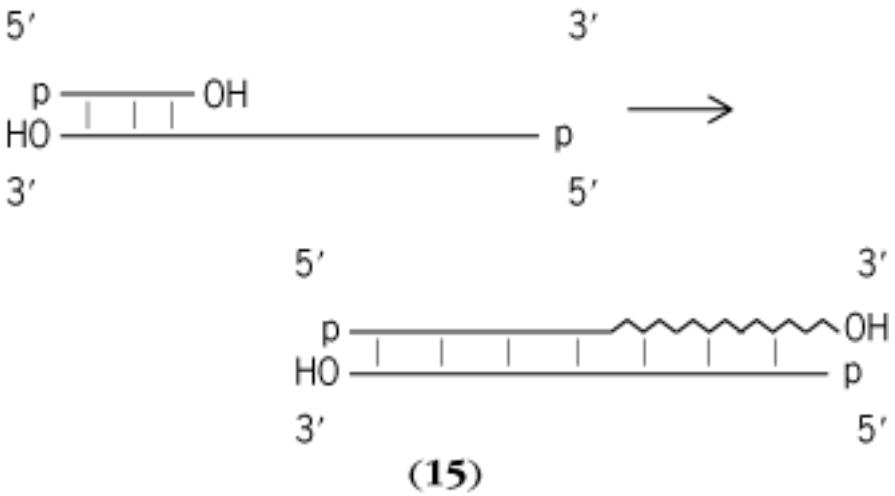


Fig. 3 DNA sequence of the promoter-operator region of the *lac* operon in *Escherichia coli*.

In all cells, DNA is synthesized by using parent DNA as a template for a polymerization reaction which requires deoxyribonucleoside 5′-triphosphates as precursors and which is catalyzed by DNA polymerase enzymes. (The deoxyribonucleoside triphosphates are synthesized in cells by addition of three phosphate groups to each nucleoside.) The DNA synthetic reaction can be summarized as reaction (1),

$$\left.\begin{matrix}\text{dATP}\\ \text{dGTP}\\ \text{dCTP}\\ \text{dTTP}\end{matrix}\right\} \xrightarrow[\text{DNA polymerase}]{\text{DNA, } Mg^{2+}} \text{DNA}\cdot(\text{4dXMP})_n + (\text{4pp}_i)_n \qquad (1)$$

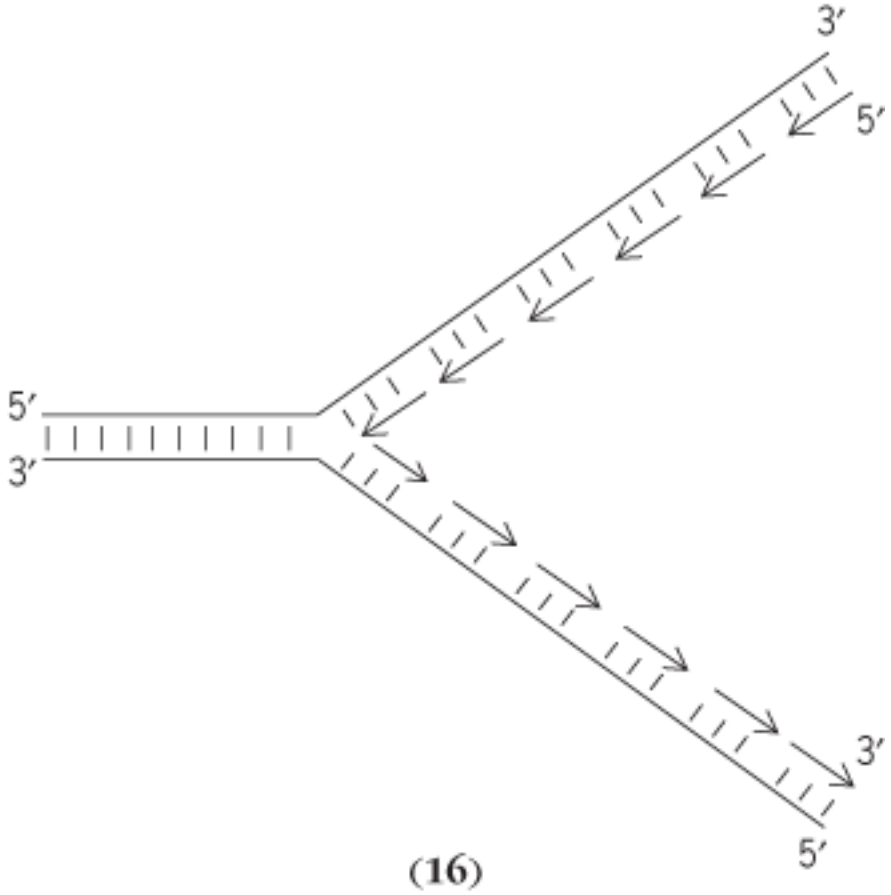
where XMP represents nucleoside monophosphate and pp_i represents inorganic pyrophosphate. The reaction always proceeds in a 5′-to-3′ direction. That is, each deoxyribonucleotide is attached through its 5′-phosphate to the 3′-OH group of an existing DNA strand, with the release of pyrophosphate. This 5′-to-3′ directionality is true for all known nucleic acid biosynthetic reactions. All DNA polymerase enzymes require an existing strand of DNA as a template, with a base-paired segment to serve as a primer. The primer sequences may consist of ribonucleotides, but the template is DNA. No DNA polymerase is capable of synthesizing DNA de novo. The primer must possess a free 3′-OH group for synthesis to begin. Through base pairing, nucleotides complementary to those in the template strand are added (**15**).



(15)

During DNA replication in cells, each strand of the helix serves as a template for synthesis. The resulting coupling of each new DNA strand with a preexisting strand is known as semiconservative replication. At a point along the double helix, the strands unwind and, after attachment of primers, synthesis proceeds bidirectionally at what is termed the replication fork. When viral DNA is replicated in *E. coli*, the newly synthesized DNA, called nascent

DNA, occurs in short fragments, now called Okazaki fragments (**16**).



(16)

The action of enzymes known as DNA ligases is required to join the Okazaki fragments. In fact, it is now known that many enzymes in addition to DNA polymerase are involved in the process of DNA replication in cells. Studies on bacterial mutants defective in DNA synthesis have implicated more than a dozen such enzymes.

There are differences between DNA synthetic mechanisms in bacteria and higher organisms. In lower organisms such as bacteria or viruses, DNA replication begins at a single point in the genome and proceeds bidirectionally to replicate all the DNA. In higher organisms, replication begins at many points, thus forming numerous "eyes." The parental DNA strands are separated and replicated until these eyes meet. The DNA polymerases of bacteria comprise several different enzymes, each of which is involved in a separate aspect of replication. The DNA polymerases of higher organisms comprise a still different set of enzymes.

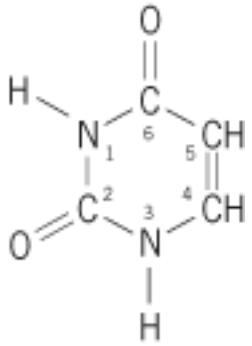
Function

In every living cell, as well as in certain viruses and subcellular organelles, the function of DNA is similar: it encodes genetic information and replicates to pass this information to subsequent generations. The nucleotide

sequence of DNA in each organism determines the nature and number of proteins to be synthesized, as well as the organization of the protein-synthesizing apparatus. The first conclusive evidence that DNA alone can direct cell function was obtained in bacterial cells. In 1944, O. T. Avery, C. M. MacLeod, and M. McCarty showed that purified DNA from a strain of smooth, polysaccharide-encapsulated *Pneumococcus* could transform rough, nonencapsulated strains so that they and their offspring resemble the DNA donor. Subsequent studies on bacterial transformation showed that numerous other hereditary characteristics could be transferred between strains using pure DNA. G. Beadle and E. L. Tatum showed that mutations in DNA of the mold *Neurospora* resulted in alterations in the amino acid sequence of the protein. The entire process of gene expression, by which the flow of information proceeds from DNA to RNA to protein, remains one of the most fertile areas of molecular biological research. *See also:* DEOXYRIBONUCLEIC ACID (DNA).

Ribonucleic Acid

RNAs are long polymeric chains of ribonucleotides joined end to end in a 5′-to-3′ linkage. The primary chemical difference between RNA and DNA is in the structure of the ribose sugar of the individual nucleotide building blocks: RNA nucleotides possess a 2′-OH group (**1a**), whereas DNA nucleotides do not (**1b**). Another major chemical difference between RNA and DNA is the substitution of uridylic acid, which contains the base uracil (2,6-dioxypyrimidine; **17**)



(17)

for thymidylic acid as one of the four nucleotide building blocks. Thus incorporation of radioactive uridine can be used as a specific measure of RNA synthesis in cells, while incorporation of radioactive thymidine can be used as a measure of DNA synthesis. Further modifications of RNA structure exist, such as the attachment of various chemical groups (for example, isopentenyl and sulfhydryl groups) to purine and pyrimidine rings, methylation of the sugars (usually at the 2′ position), and folding and base pairing of sections of a single RNA strand to form regions of secondary structure. Unlike DNA, nearly all RNA in cells is single-stranded (except for regions of secondary structure) and does not consist of double-helical duplex molecules. Another distinguishing characteristic of RNA is its alkaline lability. In basic solutions, RNA is hydrolyzed to form a mixture of nucleoside 2′- and 3′-monophosphates. In contrast, DNA is stable to alkali.

Classes

Cellular RNA consists of classes of molecules widely divergent in size and complexity.

Ribosomal RNA. The most abundant class of RNA in cells is ribosomal RNA. This class comprises those molecular species that form part of the structure of ribosomes, which are components of the protein-synthesizing machinery in the cell cytoplasm. The predominant RNA molecules are of size 16S and 23S in bacteria (the S value denotes the sedimentation velocity of the RNA upon ultracentrifugation in water), and 18S and 28S in most mammalian cells. In most mammalian cells, the 18S RNA is approximately 2100 nucleotides long, and the 28S RNA approximately 4200 nucleotides. In the large subunit of ribosomes, the 28S RNA is associated with a smaller 5S RNA species which is about 120 nucleotides long. In eukaryotic cells, another small RNA species, 5.8S RNA, is also associated with 28S RNA. All of the ribosomal RNA species are relatively rich in G and C residues. For example, the 28S RNA of the newt *Triturus* contains 60.9% G and C residues. As they exist in ribosomes, the ribosomal RNAs contain considerable intrachain base pairing which helps maintain a folded three-dimensional structure. *See also:* RIBOSOMES.

Messenger RNA. Another prominent class of RNAs consists of the messenger RNA molecules and their synthetic precursors. Messenger RNAs are those species that code for proteins. They are transcribed from specific genes in the cell nucleus, and carry the genetic information to the cytoplasm, where their sequences are translated to determine amino acid sequences during the process of protein synthesis. The messenger RNAs thus consist primarily of triplet codons. Most messenger RNAs are derived from longer precursor molecules which are the primary products of transcription and which are found in the nucleus. These precursors undergo several steps known as RNA processing which result in production of cytoplasmic messenger molecules ready for translation. Processing steps frequently include addition of a 7-methyl guanosine residue to the 5′ terminal triphosphate group, splicing-out of intervening sequences in the coding regions, and addition of multiple adenylic acid residues (often as many as 200) to the 3′ terminus of the polynucleotide chain. Certain messengers, such as those coding for histones, do not possess a polyadenylic acid tail. The messenger RNAs as a class possess a high degree of sequence complexity and size diversity, stemming from the fact that they are largely the products of uniquely represented structural genes. Most messenger RNAs range from 1500 to 3000 nucleotides in length. The relative size of their synthetic precursors varies widely, ranging from only slightly longer to more than twice as long as the final messenger RNA product.

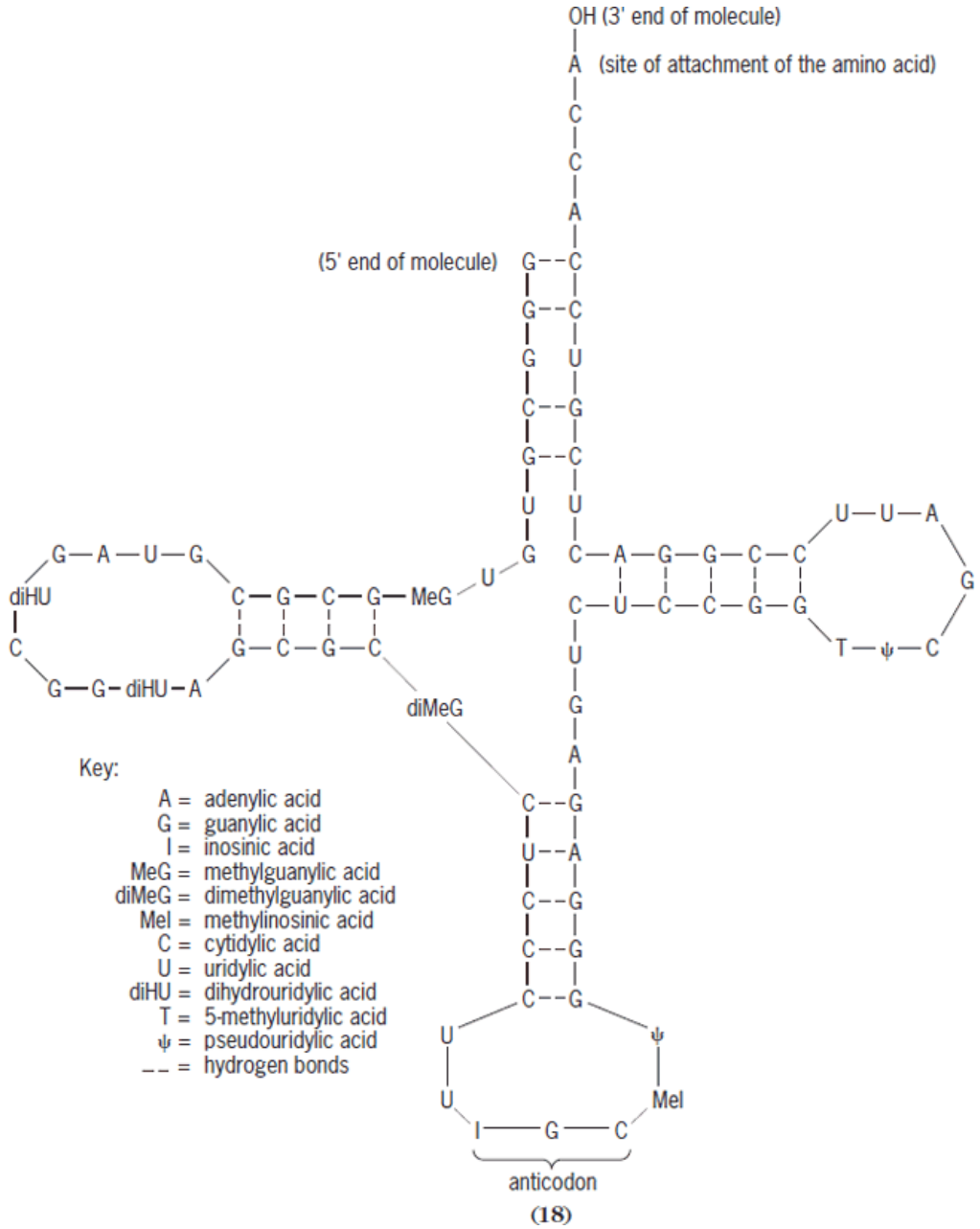
Transfer RNA. The third major RNA class is transfer RNA. There are approximately 80 transfer RNA molecular species in the bacterium *E. coli*. These are small RNA molecules 73–93 nucleotides long which possess a relatively high proportion of modified and unusual bases, such as methylinosine or pseudouridine. Each transfer RNA molecule possesses an anticodon and an amino acid binding site. The anticodon is a triplet complementary to the messenger RNA codon for a particular amino acid. A transfer RNA molecule bound to its particular amino acid is termed charged. The charged transfer RNAs participate in protein synthesis: through base pairing, they

bind to each appropriate codon in a messenger RNA molecule and thus order the sequence of attached amino acids for polymerization.

In addition to the three major RNA classes, numerous small RNA species are found in the cell nucleus. These RNAs frequently have a high content of unusual bases. Some of them are synthesized in the cell nucleolus. Their function is unknown. Some of the small nuclear RNAs arise from the splicing out of intervening sequences from the transcripts of large interrupted genes.

Sequences

Alanyl transfer RNA (a transfer RNA recognizing the alanine codon) was the first RNA molecule to be sequenced in its entirety (**18**).



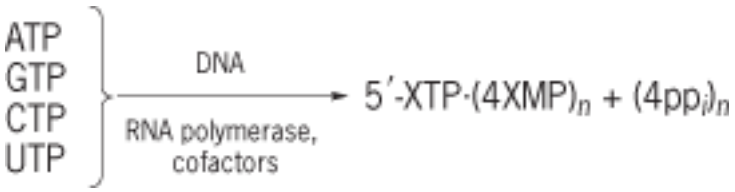
(18)

The sequences of many other RNA molecules and segments of molecules are now known. Of particular interest are sequences near RNA transcription initiation or termination sites, since these provide information about gene-regulatory regions. Comparison of sequences of messenger RNAs and ribosomal RNAs has revealed that base pairing is important in attachment of messenger RNA to the ribosomes during initiation of protein synthesis. In

general, RNA sequencing is carried out by cleavage of radioactively labeled RNA with nucleotide-specific nuclease enzymes, followed by two-dimensional separation and further analysis of the labeled cleavage products.

Transcription

The process of RNA biosynthesis from a DNA template is called transcription. Transcription requires nucleoside 5′-triphosphates as precursors, and is catalyzed by enzymes called RNA polymerases. Unlike DNA replication, RNA transcription is not semiconservative: only one DNA strand (the "sense" strand) is transcribed for any given gene, and the resulting RNA transcript is dissociated from the parental DNA strand. The RNA synthetic reaction can be summarized as reaction (2),



(2)

where 5′-XTP represents nucleoside 5′-triphosphate, XMP represents nucleoside monophosphate, and pp_i represents inorganic pyrophosphate. As is true of DNA synthesis, RNA synthesis always proceeds in a 5′-to-3′ direction. Since RNA synthesis requires no primer, as does DNA synthesis, the first nucleotide in any primary transcript retains its 5′-triphosphate group. Thus an important test to determine whether any RNA molecule is a primary transcript is to see whether it possesses an intact 5′-triphosphate terminus. As in DNA replication, base pairing orders the sequence of nucleotides during transcription. In RNA synthesis, uridine, rather than thymidine, base-pairs with adenine.

There are profound differences between bacteria and higher organisms with regard to mechanisms of RNA synthesis and processing. In bacteria, RNA synthesis and translation are coupled; that is, nascent messenger RNA chains are attached directly to ribosomes for protein synthesis. In higher organisms, which possess a distinct cell nucleus, several processing and transport steps intervene between RNA synthesis (in the nucleus) and translation (in the cytoplasm). Bacterial RNA polymerization employs essentially one RNA polymerase enzyme. In contrast, eukaryotic RNA synthesis employs predominantly three RNA polymerases. Messenger RNA synthesis is catalyzed by one enzyme, 5S and transfer RNA synthesis by another, and ribosomal RNA synthesis by still another. All three of these eukaryotic RNA polymerases carry out the same 5′-to-3′ phosphodiester linkage of nucleotides. They differ in their protein subunit composition and in the regulatory DNA sequences they recognize. The recognition of different regulatory sequences by different RNA polymerases under different cellular conditions is one of the most interesting problems in molecular biology. It is this sequence-specific recognition that lies at the heart of the problem of differential gene activity and thus of cell differentiation.

It is known that many proteins cofunction with RNA polymerases to regulate DNA binding and RNA chain initiation. Among these regulatory proteins are several whose activities are influenced by hormones. For

example, an estrogen-binding protein regulates transcription of the chick ovalbumin gene, and a cyclic adenosine monophosphate–binding protein regulates RNA polymerase activity at the *lac* operon in *E. coli* (**16**). In addition to these transcriptional regulatory factors, chromosomal proteins called histones exist in complexes with the DNA that can influence the rate of passage of polymerases. These histone-DNA complexes, or nucleosomes, are present on most eukaryotic cellular DNA, and are subject to several biochemical modifications which are influenced by hormones. The interaction of RNA polymerases with DNA in nucleosomes is presently an active area of study. *See also:* HISTONE; TRANSCRIPTION.

Functions

The primary biological role of RNA is to direct the process of protein synthesis. The three major RNA classes perform different specialized functions toward this end. The 18S and 28S ribosomal RNAs of eukaryotes are organized with proteins and other smaller RNAs into the 45S and 60S ribosomal subunits, respectively. The completed ribosome serves as a minifactory at which all the components of protein synthesis are brought together during translation of the messenger RNA. The messenger RNA binds to the ribosome at a point near the initiation codon for protein synthesis. Through codon-anticodon base pairing between messenger and transfer RNA sequences, the transfer RNA molecules bearing amino acids are juxtaposed to allow formation of the first peptide bond between amino acids. The ribosome then, in a presently unknown fashion, moves along the messenger RNA strand as more amino acids are added to the peptide chain. *See also:* PROTEIN.

RNA of certain bacterial viruses serves a dual function. In bacteriophages such as f2 and Q-beta, the RNA serves as a message to direct synthesis of viral-coat proteins and of enzymes needed for viral replication. The RNA also serves as a template for viral replication. Viral RNA polymerases which copy RNA rather than DNA are made after infection. These enzymes first produce an intermediate replicative form of the viral RNA which consists of complementary RNA strands. One of these strands then serves as the sense strand for synthesis of multiple copies of the original viral RNA. *See also:* BACTERIOPHAGE.

RNA also serves as the actively transmitted genomic agent of certain viruses which infect cells of higher organisms. For example, Rous sarcoma virus, which is an avian tumor virus, contains RNA as its nucleic acid component. In this case, the RNA is copied to make DNA by an enzyme called reverse transcriptase. The viral DNA is then incorporated into the host cell genome, where it codes for enzymes which are involved in altering normal cell processes. These enzymes, as well as the site at which the virus integrates, regulate the drastic transformation of cell functions, which induces cell division and the ultimate formation of a tumor. Transcription of the viral DNA results in replication of the original viral RNA. *See also:* CYCLIC NUCLEOTIDES; REVERSE TRANSCRIPTASE; RIBONUCLEIC ACID (RNA); TUMOR VIRUSES.

Edward Johnson

In Situ Hybridization

In situ hybridization is a technique that permits identification of particular DNA or RNA sequences while these sequences remain in their original location in the cell. Single-stranded nucleic acids, both DNA and RNA, can bind to single strands of complementary nucleotide sequences to produce double-stranded molecules (nucleic acid hybrids). The stability of such hybrids is primarily dependent on the number of base pairs formed between the two strands. Under stringent hybridization conditions, only completely complementary sequences will form enough base pairs to remain associated. Thus, a single-stranded nucleic acid probe can, by hybridization, identify a gene (DNA) or RNA transcribed from that gene.

In situ hybridization can be used to localize DNA sequences within chromosomes or nuclei and also to localize RNA within the nucleus or cytoplasm of the cell. Although most in situ hybridization studies are analyzed with the light microscope, it is also possible to perform in situ hybridization experiments on chromosomes prepared for analysis in the electron microscope.

The hybridization probe is the piece of single-stranded nucleic acid whose complementary sequence is to be localized within the cell. For example, the probe might be a recombinant DNA molecule carrying the gene that codes for a known protein. Nucleic acid probes can be labeled with the radioisotope tritium (^{3}H) by several different techniques. Sulfur-35 can also be used to label probes. At the end of an experiment, the location of the hybridized radioactive probe is detected by autoradiography of the cell preparation, thus demonstrating the location of the complementary sequence or sequences in the cell or tissue (**Fig. 4**). It is also possible to use nonradioactive probes made with modified nucleotides. For those probes, the hybrid is detected by an antibody against the modified nucleotides. The antibodies used have been joined to fluorescent molecules or to enzymes to allow cytological detection at the end of the experiment. *See also:* AUTORADIOGRAPHY.

The cell preparation is made by treating the tissue to be studied with a fixative. Fixatives are agents that preserve the morphology of the cell as well as possible during subsequent steps. Fixatives usually act by cross-linking or denaturing proteins. After cells have been fixed, they are sectioned or squashed to yield preparations that are thin enough (a few micrometers) to be mounted on slides and analyzed in the microscope. If the experiment is designed to localize DNA in the cells, the DNA is denatured by treating the cells with a denaturing agent, such as alkali. The denaturation step also removes some protein from the DNA, but enough protein is left to maintain cellular morphology. Since RNA is, for the most part, single-stranded, no denaturing step is used for hybridization to RNA in cells.

The probe is placed on the cytological preparation and held there under stringent conditions until hybrids have had sufficient time to form. The nonhybridized probe is then washed off the preparation. If the probe is radioactive, the preparation is covered with autoradiographic emulsion and exposed in the dark until tritium decay has produced grains in the emulsion over the labeled hybrid. The autoradiograph is then developed, and the location of grains over the cells is analyzed. When a nonradioactive probe is used, the hybrid detection is

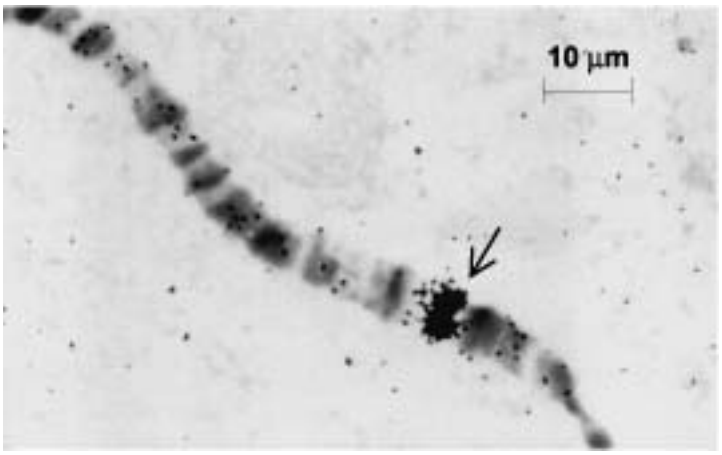


Fig. 4 **Autoradiograph of part of a chromosome from *Drosophila melanogaster* hybridized with ^{3}H-DNA carrying the genes for muscle tropomyosin. Autoradiographic grains (arrow) are over the region that has the genes for both tropomyosin I and tropomyosin II. (*From V. L. Bautch et al., Organization and expression of Drosophila tropomyosin genes, J. Mol. Biol., 162:231–250, 1982*)**

accomplished by incubating the preparation with an antibody that recognizes the modified nucleotide probe. If the antibody is coupled to a fluorescent molecule, the preparation is viewed in a fluorescence microscope to localize sites of hybridization. If the antibody is tagged with an enzyme, the preparation is incubated with a substrate from which the enzyme will produce a colored product, thus marking the site of the hybrid. *See also:* IMMUNOFLUORESCENCE.

In situ hybridization can be used to answer a number of biological questions. Hybridization to the DNA of condensed chromosomes can be used for gene mapping (**Fig. 5***a*). Hybridization to DNA of interphase nuclei allows study of the functional organization of sequences in the diffuse chromatin of this stage of the cell cycle (Fig. 5*b*). Hybridization to cellular RNA allows a precise analysis of the tissue distribution of any RNA (**Fig. 6**).

Mary Lou Pardue

Bibliography

R. L. P. Adams, J. T. Knowler, and D. P. Leader, *The Biochemistry of the Nucleic Acids*, 11th ed., 1992

F. Eckstein and D. M. Lilley, *Nucleic Acids and Molecular Biology*, vol. 8, 1994
DOI: http://doi.org/10.1007/978-3-642-78666-2

F. Eckstein and D. M. Lilley (eds.), *Nucleic Acids and Molecular Biology*, vol. 2, 1988
DOI: http://doi.org/10.1007/978-3-642-83384-7

A. R. Leitch, *In Situ Hybridization*, 1994

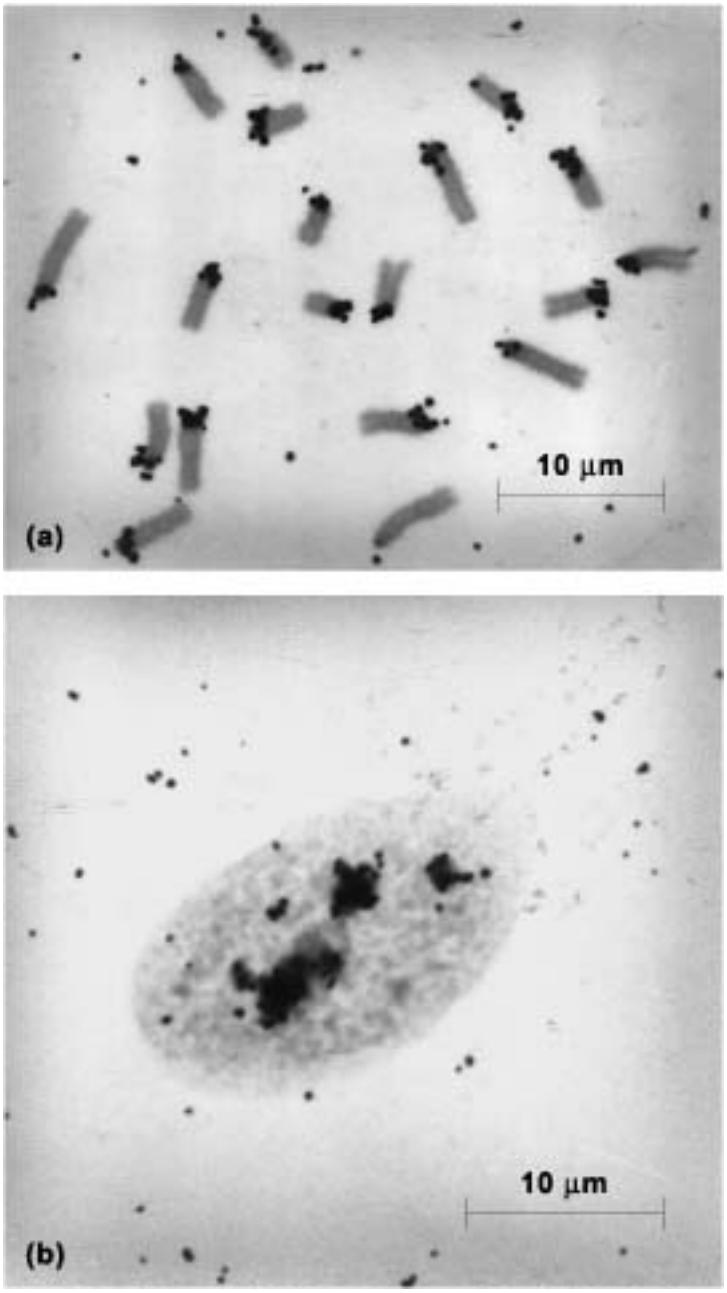


Fig. 5 **Autoradiographs of mouse cytological preparations hybridized with ^{3}H-RNA copied from mouse satellite DNA. (*a*) Metaphase chromosomes, each with the centromere at one end. The black autoradiographic grains indicate that the highly repeated simple sequences of satellite DNA are adjacent to the centromeres in most of the chromosomes. (*b*) Interphase nucleus of Sertoli cell of testis. The autoradiographic grains are found in only a few clusters, indicating that in nuclei of this particular cell type the centromeric regions of the chromosomes are tightly associated into a few small groups. (*From C. D. Darlington and K. R. Lewis, eds*., *Chromosomes Today, vol. 3, Hafner Publishing, 1972*)**

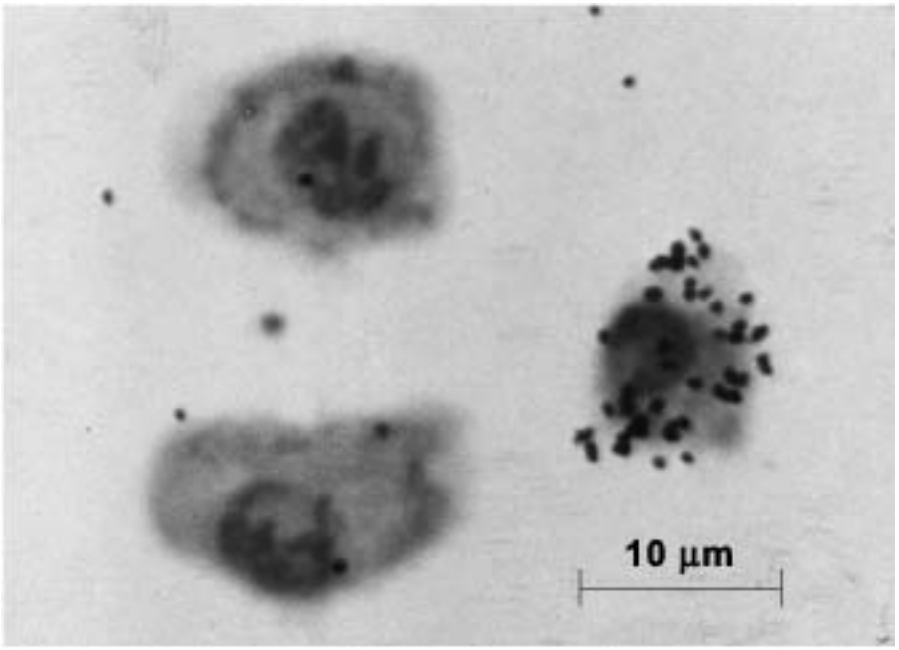


Fig. 6 **Autoradiograph of *Drosophila* cells showing use of in situ hybridization to detect histone messenger RNA. Cells from a nonsynchronized population were hybridized with a ^{3}H-labeled probe complementary to histone messenger RNA, which is found only in cells that are synthesizing DNA. Thus, only the cell on the right, with autoradiographic grains indicating histone messenger RNA in the cytoplasm, appears to be actively replicating its DNA.**

W. Saenger, *Principles of Nucleic Acid Structure*, 1993

P. Tijssen, *Hybridization with Nucleic Acid Probes*, 2 vols., 1993

Additional Readings

G. Mayer (ed.), *The Chemical Biology of Nucleic Acids*, John Wiley & Sons, Chichester, West Sussex, UK, 2010

S. Neidle, *Principles of Nucleic Acid Structure*, Academic Press, London, UK, 2008