VOLUME 12

Franco to Goethals

THE ENCYCLOPEDIA

AMERICANA

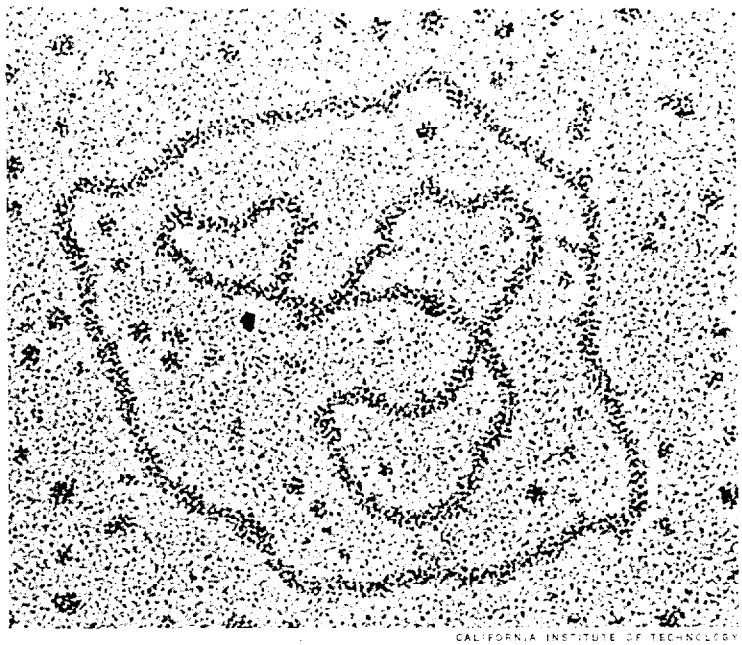
INTERNATIONAL EDITION

COMPLETE IN THIRTY VOLUMES

FIRST PUBLISHED IN 1829

GROLIER INCORPORATED

International Headquarters: Danbury, Connecticut 06816



CALIFORNIA INSTITUTE OF TECHNOLOGY

ELECTRON MICROGRAPH of viral DNA, showing two ring molecules, each of which contains the total genetic information of ox virus particle, equivalent to 15 to 20 genes. These DNA rings were synthesized by Arthur Kornberg and his associates at the Stanford University Medical School.

GENE, jēn, a natural unit of the hereditary material, which is the physical basis for the transmission of the characteristics of living organisms from one generation to the next. At first glance, the hereditary material does not seem to lend itself to rational dissection. The progeny of a pair of cats is a kitten, which develops into a cat; dogs engender pups; humans engender children. These are the most primitive observations concerning heredity, and they give no hint of the possible analysis of the hereditary material into unit components, or genes.

Initial Concept of the Gene. Modern research in genetics is founded on Gregor Mendel's insight that one must study discrete differences between variants within a single species capable of interbreeding. Mendel, a 19th century Austrian botanist-monk, studied the hereditary transmission of characteristics in garden peas and provided the basis of the science of genetics. He foretold the modern concept of the gene by hypothesizing that the choice between alternative manifestations of a character was determined by an invisible "anlage"—a term that might be translated into its modern equivalent, "gene."

One could extrapolate from Mendel's findings to the proposition that every character of a plant or animal is controlled by a corresponding gene. For example, the height of a pea plant (tall or dwarf) is determined by the inheritance of a specific gene. Later it became clear that many different genes could cooperate in determining a given character, perhaps by influencing its development in different ways. The work of Mendel and others thus led to the concept of the gene as some kind of hereditary unit, and for much of the first half of the 20th century the term was often used with no greater discrimination than that. Thus we may speak of "the genes" collectively, meaning the totality of genetic, or hereditary, information in a cell or in an organism.

Questions about the Gene. The concept of the gene, while in itself a great advance, left many unanswered questions. For example, how large is a gene? How many genes are in a single cell? Are all genes alike in the different cells of the body? Can a gene be seen under a light microscope? Do genes change, or mutate, independently of one another? How closely do human genes resemble genes of other species? Can a gene be synthesized in the laboratory? Is a virus a gene? What is the chemical nature of the gene? How do we define the boundaries of one gene that set it off from another gene? What is the connection between a gene and the character it controls? Are all genes in the chromosomes of cells? Do genes vary, and if so, does any rational classification emerge?

Search for Answers. Before the introduction of biochemical analysis, breeding experiments gave brilliant, if tentative, answers to many questions about the gene. Indeed, between 1915 and 1945 many geneticists took pride in how far they could push genetic analysis without direct chemical information. Mapping studies, based on the findings of breeding experiments, showed that genes were generally arranged in linear order on *chromosomes*. The studies failed, however, to reveal very much about the nature of a particular gene, in a narrow sense, and became confusing when applied on a highly refined scale. With the advent of biochemical analysis, tremendous advances in the field of genetics were made, and in the 1950's and 1960's all of the questions listed above were illuminated.

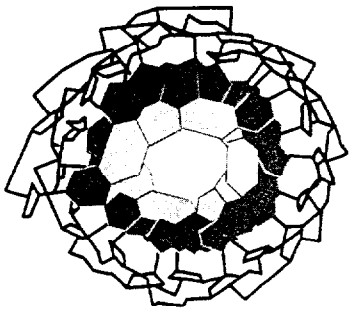
Scope of this Article. The origin and historical development of the concept of the gene are presented fully in the article GENETICS. In this article, we shall proceed directly to contemporary views of the concept of the gene, primarily from a biochemical standpoint, with the aim of presenting a clear view of what is now known about the gene, even at the expense of the exciting perspective of the history of genetics. We will thus take advantage of hindsight in order to redefine "gene" and to try to answer many questions from the viewpoint of contemporary molecular genetics.

CARRIERS OF HEREDITARY INFORMATION

Hereditary, or genetic, information in a cell is usually carried in coded form by a *nucleic acid*. The only known exceptions are a few steady-state systems, such as bacterial cell walls, that are self-propagating and do not rely on nucleic-acid coding for hereditary transmission.

Almost all genetic material in almost all terrestrial (earth) organisms is encoded in the nucleic acid DNA. DNA (*deoxyribonucleic acid*) usually occurs as a giant double-stranded mole-

STRUCTURAL MODEL OF DNA



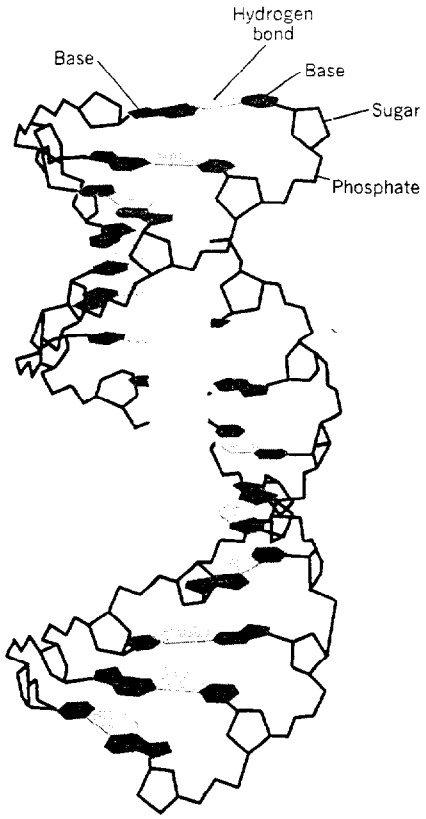


Fig. 1. DNA usually occurs as a double-stranded molecule. Each strand consists of a chain of nucleotides, each of which contains a phosphate, a deoxyribose sugar, and a base. *(Above)* Top view of DNA structural model. *(Below)* DNA model wound around a fiber axis. Bases, which protrude from the sugars in each chain, are linked to bases of opposing strand by hydrogen bonds. These links form horizontal supports, holding the DNA chains together.

cule (see Fig. 1). Each strand consists of a long chain of *nucleotides*, each of which contains a phosphate, deoxyribose sugar, and one of four bases—adenine (abbreviated A), thymine (T), cytosine (C), and guanine (G). In the double-stranded form, the A always pairs with T, and the C always pairs with G. The nucleotide sequence codes specific hereditary information. DNA is found in almost all cells and is capable not only of replicating itself but also of serving as a template, or blueprint, for the production of RNA (*ribonucleic acid*), a closely related substance. RNA is the only other substance known to code genetic information and, in virus infected cells, to be capable of self-replication. See also DNA; Nucleic Acid; RNA.

In theory, substances other than DNA and RNA might be found to play a similar role and to code genetic information. Life may exist beyond the earth, say on Mars or Jupiter, or on the planets of other stars much farther away, and there is no fundamental reason why other genetic polymers (informational, chainlike molecules) might not have evolved elsewhere or even still be found or synthesized on the earth. There is, however, no experimental foundation for these speculations as yet.

DNA most often is found in the nucleus of the cell, but it may also be found in certain cytoplasmic organelles and in other particles such as episomes and viruses. RNA that codes genetic information is found only in some simple viruses.

DNA in Chromosomes. In eukaryons (higher plants and animals whose cells contain nuclei bounded by definite membranes), most of the DNA is found in chromosomes. Chromosomes are components of the cell nucleus, containing structural materials like histones as well as a fundamental allotment of DNA. The detailed structure of the chromosome is still controversial and remains a major challenge to contemporary research in molecular genetics. One question is whether the DNA of a chromosome is, as much evidence suggests, a single immense molecule. This would be many centimeters long if unwound from a human chromosome. Alternatively the DNA might be segmented into meaningful units held together by non-DNA material. See also CHROMOSOME.

In organisms lacking nuclear membranes, such as viruses and bacteria, the DNA is characteristically a single molecule not tightly bound to other substances. However, the DNA content of a whole bacterium is less than 1/100 of the DNA of a single chromosome in a eukaryon, so that the possible scope for a more complex organization of DNA in eukaryons is evident.

DNA in Certain Cytoplasmic Organelles. DNA is also found in structures outside the chromosome. For example, small amounts occur in the mitochrondria of all cells, in the centrioles and basal granules of animal cells, and in the chloroplasts of plant cells.

Extrachromosomal DNA is the probable basis of several examples of extrachromosomal heredity in algae and protozoa. Definite proof of extrachromosomal DNA was obtained from the studies of the American geneticists Tracy Morton Sonneborn and John Preer on the kappa particles in paramecia. Kappa particles are nonnuclear particles of DNA transmitted through the cytoplasm of the paramecium. Those paramecia that contain large amounts of kappa particles are known as "killer" strains because they secrete a substance that kills other "sensitive" paramecia

of the same species that do not contain large amounts of kappa particles. The kappa particles can be thought of equally logically as cytoplasmic, extrachromosomal genes, or as intracellular symbionts.

Extrachromosomal DNA, such as the DNA found in mitochondria and centrioles in animal cells, should behave like the DNA found in chromosomes in being vulnerable to mutation and to natural selection—evolution. As such, extrachromosomal DNA might be relevant to studies of differentiation and cancer. Conventional breeding methods have given no data on this question. Some studies in which tissue cells from different species—some as distant as man and fish—were artificially fused, promise to reopen this search.

DNA in Episomes and Viruses. Some particles, known as *episomes*, are chromosomal fragments of either whole-cell or viral origin that may move in and out of the chromosome, possibly by a kind of crossing-over, or exchange of parts, between chromosomes. Episomes are known to be DNA molecules, but their exact relationship to the rest of the DNA is not yet clear.

In the common colon bacterium *Escherichia coli*, the genes for sexual competence are episomal. The episome for maleness, for example, is infectiously transmitted by contact with female cells. It can be "cured" by exposing male cells to cobalt salt or to some of the same dyes that eliminate mitochondria in yeast. Episomes are also important in the infectious transmission of resistance to some antibiotics (resistance transfer factors) found in some bacteria and in bacterial capacity to produce other antibiotics.

Episomes may also occur in mammalian cells, the most provocative examples being the tumor-inducing viruses SV40 and polyoma. These viruses infect normal tissue cells and transform them to cancerlike cells. In the process the viral particles disappear, leaving strong evidence that they have united with the chromosomes. Many cell generations later, they may reemerge if, for example, the cancerlike cell is fused with a fresh, uninfected normal cell, or if the cell is exposed to radiation.

Episomes, then, are those plasmids that can enter the chromosome of the host cell. The existence of episomes emphasizes the ambiguity of the distinction between gene and virus. DNA in a chromosome makes up genes, but the same genetic information may be encapsulated in an infectious particle that would then be called a virus. The ambiguity is further illustrated by the process of transduction in bacteria—a process in which genetic material typical of host cell function can be incorporated into viruses and thereby transmitted to new bacteria with consequent changes in their genotype, or genetic makeup.

Many virologists believe that we must redouble our research on live viruses used for mass vaccination programs, lest we inadvertently spread some risk of long-delayed cancer or even provoke unexpected changes in the genetic information passed from one generation to the next.

RNA in Viruses. RNA, the only other polymer known to code genetic information, is very similar in structure to DNA, one difference being that thymine is replaced by uracil (abbreviated U). Some simple viruses, like the influenza virus, encode their genetic information in RNA, whereas other viruses retain DNA codes.

RNA is abundant in normal cells as a gene product—that is, a substance formed, like all other substances, according to the instructions of a DNA code. The evolutionary origin of viral RNA is unknown. Viral RNA might be the ancestor of DNA; it might have evolved separately but parallel to DNA; or it might have evolved as RNA that somehow escaped from the cell. Whatever its origin, in infections by those viruses whose genetic information is encoded in RNA, and not in DNA, the viruses must induce a replicating enzyme to reproduce RNA or DNA from an RNA template. Such enzymes have been found only in virus-infected cells. It appears that RNA can neither replicate itself, nor influence the replication of its parent DNA, in normal (DNA coding) cells.

Self-Propagating Steady-State Systems. Some simple organisms have self-propagating, steady-state systems that do not rely on the abstract coding afforded by the DNA or RNA mechanism. A few unicellular organisms exhibit the propagation of cell-wall alterations that reflect the continuity of a pattern in the assembly of the wall itself—a continuity that is not dependent on DNA. This is roughly analagous to the process of crystal growth, in which the electronic structure of a preexisting crystalline pattern of atoms influences the further assembly of new atoms into the crystal's characteristic pattern.

The general significance of self-propagating, steady-state systems is controversial, although the principle has been clearly demonstrated in some special cases. Aaron Novick, an American molecular biophysicist, demonstrated that certain bacteria can be poised for many generations in either of two stable self-maintaining states, depending on their previous exposure to a pulse of a high level of an inducing substrate. "Plus" cells concentrate the inducer from a dilute solution and are stabilized in this state by the inducer accumulated within the cells. Conversely, "minus" cells, which do not concentrate the inducing substance, remain uninduced and are stabilized in the minus state. Such systems require a steady metabolic turnover to remain stable. They are notably vulnerable to environmental upset, and this has led to speculation about environmentally induced mutations.

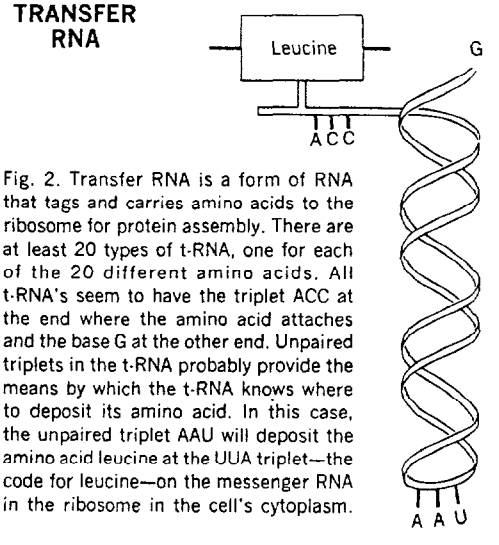


Fig. 2. Transfer RNA is a form of RNA that tags and carries amino acids to the ribosome for protein assembly. There are at least 20 types of t-RNA, one for each of the 20 different amino acids. All t-RNA's seem to have the triplet ACC at the end where the amino acid attaches and the base G at the other end. Unpaired triplets in the t-RNA probably provide the means by which the t-RNA knows where to deposit its amino acid. In this case, the unpaired triplet AAU will deposit the amino acid leucine at the UUA triplet—the code for leucine—on the messenger RNA in the ribosome in the cell's cytoplasm.

Thus, we have determined that, except for a few self-propagating systems in simple organisms, hereditary information is transmitted through DNA in chromosomes, certain cytoplasmic organelles, and episomes and viruses, and through RNA in certain simple viruses. For the vast majority of organisms, however, it is the DNA in chromosomes that carries most hereditary information. In the remainder of the article, references to a gene will be to a *chromosomal gene*.

There are several categories of chromosomal genes. Each type produces a product with a different function in the cell.

CLASSIFICATION OF CHROMOSOMAL GENES

Genes may function in more ways than are now known. However, the following categories, which are not necessarily mutually exclusive, can be recognized.

Message, or *structural*, *genes* code for the amino-acid sequence of a specific protein through a messenger-RNA (mRNA) copy. This carries the coded instructions from the nucleus to the ribosomes, which are the sites of protein assembly in the cell.

Operator genes signal instructions to regulate the readout or transcription of adjacent message genes. The operator genes in turn may be switched on and off in response to repressor substances, which are products of regulatory genes.

Ribosomal RNA genes are transcribed into the structural RNA of the ribosomal particles (rRNA) on which protein assembly occurs.

Transfer RNA genes are transcribed into the diverse transfer RNA (tRNA) molecules that are responsible for tagging and carrying individual amino acids to the ribosomes for protein assembly. There are at least 20 kinds of transfer RNA, one for each of the 20 amino acids. A specific tRNA binds to a specific amino acid and carries it to that part of the mRNA that codes for that particular amino acid in the ribosome. See Fig. 2.

Regulatory genes are structural genes whose products are repressors. These act on operators to control the transcription of other genes.

Suppressor genes reverse the phenotypic, or observable, effects of mutations in other genes by a variety of mechanisms.

Kinetic genes are a mixed bag of genes that regulate chromosome movement at cell division. For example, they regulate the centromere of the chromosomes and the initiation of chromosome replication.

Synaptic recognition genes drive the unknown mechanism by which homologous chromosomes synapse (pair off) during the reduction division in higher organisms.

In addition to the genes described above, the chromosomes contain DNA whose function, if any, has not been determined. This DNA may constitute functionless *inert genes*. It may constitute *redundant genes* that duplicate other genes in the chromosome. Or it may be *structural DNA* (not to be confused with *structural gene*, as the *message gene* is sometimes called), which

GENETIC CODE

Table 1. The genetic code is a triplet code, consisting of 64 entries. Each triplet, called a codon, represents a frame of three nucleotides (U = uracil, C = cytosine, A = adenine, G = guanine) on messenger RNA. At least one codon, and usually more than one, codes for each of the 20 amino acids. A few triplets signal the "stop" of protein assembly and at least one—methionine—may signal the "start."

First letter	Second letter: U	Second letter: C	Second letter: A	Second letter: G	Third letter
U	UUU Phenylalanine (Phe) UUC Phenylalanine (Phe)	UCU Serine (Ser) UCC Serine (Ser) UCA Serine (Ser) UCG Serine (Ser)	UAU Tyrosine (Tyr) UAC Tyrosine (Tyr)	UGU Cysteine (Cys) UGC Cysteine (Cys)	U C
	UUA Leucine (Leu) UUG Leucine (Leu)		UAA Stop UAG Stop	UGA Stop UGG Tryptophane (Tryp)	A G
C	CUU Leucine (Leu) CUC Leucine (Leu) CUA Leucine (Leu) CUG Leucine (Leu)	CCU Proline (Pro) CCC Proline (Pro) CCA Proline (Pro) CCG Proline (Pro)	CAU Histidine (His) CAC Histidine (His)	CGU Arginine (Arg) CGC Arginine (Arg) CGA Arginine (Arg) CGG Arginine (Arg)	U C A G
			CAA Glutamine (GluN) CAG Glutamine (GluN)		
A	AUU Isoleucine (Ileu) AUC Isoleucine (Ileu) AUA Isoleucine (Ileu)	ACU Threonine (Thr) ACC Threonine (Thr) ACA Threonine (Thr) ACG Threonine (Thr)	AAU Asparagine (AspN) AAC Asparagine (AspN)	AGU Serine (Ser) AGC Serine (Ser)	U C
	AUG Methionine (Met)		AAA Lysine (Lys) AAG Lysine (Lys)	AGA Arginine (Arg) AGG Arginine (Arg)	A G
G	GUU Valine (Val) GUC Valine (Val) GUA Valine (Val) GUG Valine (Val)	GCU Alanine (Ala) GCC Alanine (Ala) GCA Alanine (Ala) GCG Alanine (Ala)	GAU Aspartic acid (Asp) GAC Aspartic acid (Asp)	GGU Glycine (Gly) GGC Glycine (Gly) GGA Glycine (Gly) GGG Glycine (Gly)	U C A G
			GAA Glutamic acid (Glu) GAG Glutamic acid (Glu)		

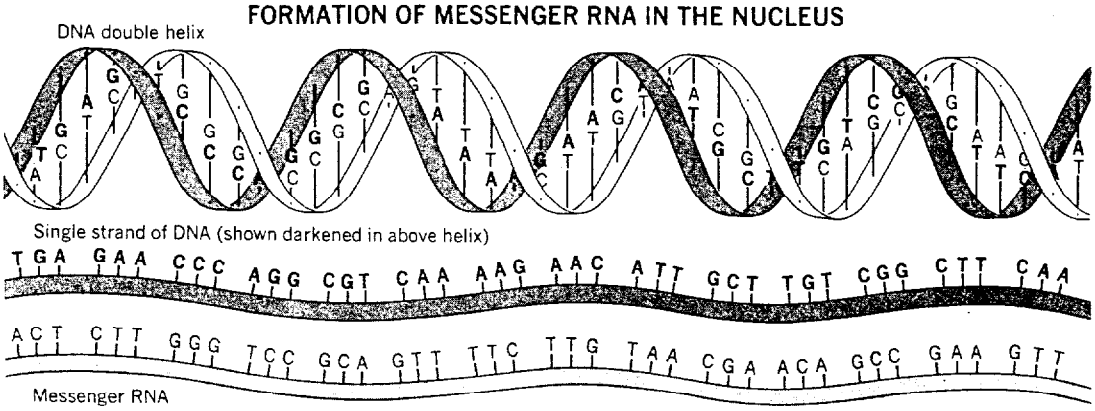


Fig. 3. As the first step in protein synthesis, the DNA helix "unzips," and a single strand of the DNA serves as a template for the formation of a complementary stand of messenger RNA, which then leaves the nucleus and goes to the ribosome to direct the amino acid sequence of a specific protein.

contributes to the structure but not the genetic functioning of the chromosome. However, such DNA may have the function of regulating the pace and efficiency of transcription or of inhibiting the effect of other nearby genes. This DNA may account for such apparent anomalies as the higher DNA content of the cells of amphibians compared to the DNA content of man. One would expect more DNA in the cells of man, the more complex organism. The redundant DNA in amphibians can be identified by its deviant chemical composition and by the reduplication of similar sequences.

Genes that fall into one of the above categories generally have fairly specialized effects. For example, the message genes for the structural proteins of the cell and for the constituents of the chromosomes themselves are central to the organization of cellular structures. Similarly, the message genes for the proteins of the mitotic spindle apparatus and for the regulators thereof are crucially important to pacing the rates of cell division. The message genes for the enzymes of DNA replication, if so altered as to produce defective enzyme products, influence the overall mutation rate of the organisms carrying them.

At this point, we have sufficient information to define a single gene as *a segment of DNA that encodes a definable product or function.* For a message gene, the product is messenger RNA, which in turn directs the synthesis of a protein chain. An operator gene is a segment necessary and sufficient to respond to one repressor molecule. A regulatory gene is a message gene whose protein product happens to be a repressor. The ribosomal RNA and transfer RNA genes form as their products the structural RNA of the ribosomes and the transfer RNA of the cytoplasm.

MESSAGE GENE, GENETIC CODE, AND PROTEIN SYNTHESIS

The message gene is the best understood and most simply explained of all genes. To most people, "gene" means "message gene." It was only after the discovery of other types of genes that the term "message gene" was adopted.

A message gene is the portion of a single strand of DNA that serves as a template for the formation of a segment of messenger RNA (mRNA) of identical information content. To transmit its coded information, the DNA "unzips" and one strand, encompassing the message gene, serves as the template. A message gene with a set of bases ACTGCTA will serve as a blueprint for the formation of a strand of mRNA with the complementary base sequence UGACGAU. (U in RNA replaces T in DNA.) The sequence of bases encodes an amino-acid sequence for a specific protein. The molecule of mRNA then serves as a template for constructing in the ribosome the amino-acid sequence instructed by the message gene. Thus the mRNA serves to carry out in the cytoplasm the instructions of the nuclear DNA. See Fig. 3.

Genetic Code. There are more than 20 different amino acids and only 4 bases, so the code cannot be based on a one-to-one correspondence between amino acids and bases. Biochemical investigations have revealed that the mRNA code is a triplet code—that is, each successive frame of three nucleotides, sometimes called a *codon*, of the mRNA corresponds to one amino acid of the protein. This rule of correspondence is the genetic code. The genetic code consists of 64 entries—the 64 triplets possible when there are 4 possible nucleotides, each of which can be at any of three places ($4x4x4 = 64$).

Breaking the Code. The concept of the genetic code was anticipated by the American biologist James D. Watson and the British biologist Francis H. C. Crick as an extrapolation of the structural model of DNA that they proposed in 1953. For this work they, along with the British biophysicist Maurice Wilkins, were awarded the 1963 Nobel Prize in physiology or medicine.

The concept was bolstered by the indirect evidence for a messenger RNA put forward by two French Nobel laureates, Jacques Monod and François Jacob, in 1961. The idea was almost immediately corroborated by the demonstration of a message functioning RNA by Sydney Brenner, François Jacob, and Matthew S. Meselson. Within the same year, Marshall Nirenberg (1968 Nobel Prize winner) opened the door to a direct attack on the code by showing that a synthetic RNA containing only the nucleotide chain UUUUU . . . could function as a message for the biosynthesis of the amino acid phenylalanine.

Subsequent work by many other biochemists, especially the Americans Severo Ochoa (1959

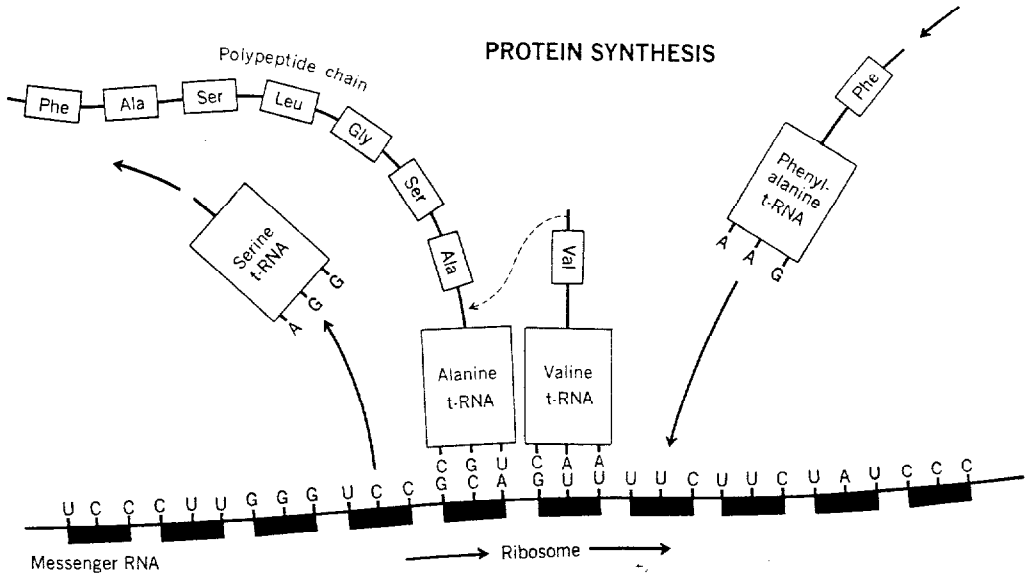


Fig. 4. Ribosomes are the sites of protein synthesis in the cell. Each sequence of three bases—a codon—on the mRNA specifies a particular amino acid. Each amino acid is brought to the ribosome by a specific tRNA whose anticodon forms a bond with the appropriate codon. The tRNA then releases its amino acid and leaves the ribosome. Above, the ribosome is shown moving along strand of mRNA "reading off" the codons. The ribosome has two binding sites for tRNA, one holding a newly arrived tRNA, the other, the growing peptide chain.

Nobel Prize winner) and Gobind Khorana (1968 Nobel Prize winner), using more complex synthetic messages, pinned down the code as a triplet code. Using the fact that the triplet UUU coded for phenylalanine as a starting point, many investigators went on to determine the code, and by 1968 the entire code was broken. See Table 1.

The successful work on the genetic code led to research on the artificial synthesis of RNA molecules that would code for specific protein synthesis. The American biologist Charles Yanofsky supported the code breakers' work from a completely different line of research. He showed that the code was corroborated by the patterns of mutation and of recombination of mutants affecting single amino acids of the protein tryptophane synthetase in the bacterium *E. coli*.

How Coded mRNA Directs Protein Synthesis. The mRNA leaves the nucleus of the cell and moves out into the cytoplasm, where it becomes associated with the ribosomes. Meanwhile, molecules of transfer RNA (tRNA) attach themselves to specific amino acids in the cytoplasm and carry these amino acids to the ribosomes. Each tRNA has a set of three unpaired nucleotides, such as CAA; the unpaired bases in the tRNA are often called *anticodons* (see Fig. 2). At the ribosome, the tRNA recognizes where to deposit its amino acid by pairing its bases with complementary ones of the mRNA at the ribosome. Thus, tRNA with the unpaired set of bases CAA will deposit the amino acid valine at the site on the ribosome where the mRNA with the code GUU is positioned. The amino acids then join by means of peptide bonds to form a peptide chain of a protein. See Fig. 4.

Insofar as the mRNA molecules typically are much longer than the gene whose information they carry, the "start" and "stop" instructions marking the beginning and end of the protein chain must also be conveyed by the information within the mRNA molecule. They do not just correspond to the ends of the mRNA molecule.

Three of the 64 triplet codes have already been found to signal a "stop" to the protein assembly. Similarly, at least in microbes, the triplet AUG, which codes for formylated methionine, may function as the usual start signal. Many microbial proteins are initiated with this amino acid. This process almost certainly is not the one that operates in mammalian cells, but the entire picture is momentarily clouded.

Size and Number of Message Genes and Other Genes. A message gene is the segment of DNA between a "start" code and a "stop" code—that is, the segment that specifies one particular protein. If a protein chain contains 100 amino-acid units, the messenger RNA coding for that protein chain would have to be 300 nucleotide pairs long. Such a nucleotide chain would have a molecular weight of 200,000. Based on the amount of DNA known to be in the nucleus of the small virus ϕ^x, and assuming that all the DNA is typical message genes, it can be estimated that there are 18 genes in the virus. Similarly it can be estimated that there are 15 million genes in a human cell.

Some genes, however, are undoubtedly smaller than 300 nucleotide pairs, and some are longer. Molecules of tRNA, for example, are only about 80 nucleotides long, whereas molecules of rRNA may be more than 5,000 nucleotides long. Operator genes may be expected to resemble tRNA in size, since both share the property of recognizing and coupling a specific amino acid. According to this functional concept, the size of a gene is not an inherent structural property and may vary over more than a fifty-fold range.

Gene Mutations. The length and information content of a message gene may be drastically modified by mutation. For example, the replacement of a single nucleotide in the middle of a gene may change a triplet from one that codes for an amino acid to one for a stop, thus cutting the gene's length. If the replacement results in a triplet that codes for another amino acid, the length of the gene will not be affected, but the protein chain for which it codes will differ from the original in one amino acid.

The insertion or deletion of a single nucleotide can set into play a drastic frame-shift mu-

tation that affects every succeeding frame, or triplet, on the gene. In a deletion, for example, each new frame would encompass two nucleotides from one of the former triplets and one nucleotide from the following triplet. Such mutations would be most disabling from the viewpoint of the effective functioning of the gene to produce a viable product. The paradoxical result that three frame-shifts can cancel each other out so that the nucleotides are again back in frame has been verified experimentally by the British biologist Sydney Brenner.

THE OPERATOR GENE AND OTHER REGULATORY GENES

The concept of an operator came from studies of mating bacteria, which indicated that protein synthesis was under negative regulation—that is, the genetic information of RNA was normally read out unless it was specifically inhibited by hypothetical cytoplasmic substances called *repressors*. Repressors are now thought to be protein products of regulatory genes.

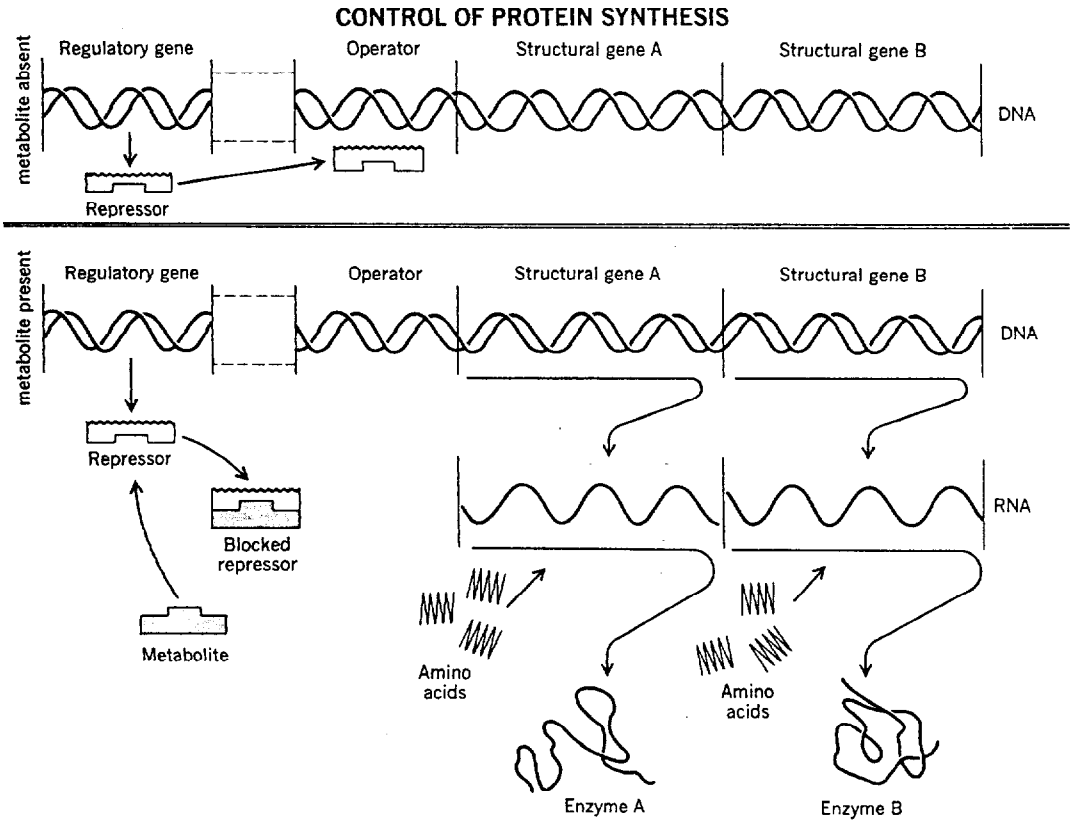
Bacterial mutations that are insensitive to inhibition by the repressor and are capable of releasing a considerable number of nearby genes from repressor control were discovered, giving more direct experimental evidence for the existence of operator genes. In bacterial cells, transcription of DNA information into RNA messages is the principal target of regulation. In higher organisms, the translation of mRNA, already formed, is also subject to regulation. The gene that responds to the repressor by initiating transcription unless inhibited by it is known as an *operator*. See Fig. 5.

Two biochemists at Harvard University, Mark Ptashne and Walter Gilbert, succeeded in isolating two specific repressors from virus-infected *E. coli* cells. One of the repressors inhibits the proliferation of the lambda virus and can be identified as a protein that has a specific affinity for a specific segment of the DNA (the operator gene) of the lambda virus. The other repressor inhibits the transcription of DNA information for the synthesis of the enzyme β galactosidase (lactase) and likewise has a specific affinity for the corresponding DNA.

There is little evidence directly supporting the operator concept in the metabolism of higher organisms. However, analogous functions of regulating gene activity have been postulated for the basic proteins—histones—often found in close association with the DNA in chromosomes. The histones thus may be regarded as at least roughly analogous to the repressor proteins that have been elucidated in microbes.

The clues available to a protein, such as a histone, for recognizing specific kinds of DNA are not yet verified, but the phenomenon has been empirically demonstrated, and its implications can hardly be evaded. The occurrence of such

Fig. 5. Structural genes A and B produce their protein products under the influence of an *operator*, unless the operator is inhibited by a repressor, which is a product of a regulatory gene. In the system shown, the absence of the metabolite (top) allows the repressor to bind the operator, thus "shutting off" protein synthesis. If the metabolite is present (bottom), it blocks the repressor from inhibiting the operator, and protein synthesis occurs.



mechanisms whereby proteins can recognize specific segments of DNA, or genes, opens the possibility that other repressorlike proteins that can influence the mutability of specific genes may be found. Such a phenomenon could help account for the variety of proteins produced as antibodies in the immune process in higher organisms.

SUPPRESSOR GENES AND tRNA

The terms "suppressor genes" and "supersuppressor genes" are perhaps the most confusing in contemporary biology. The terminology resulted when a complicated genetic situation was analyzed in formal terms before corresponding biochemical information was available.

There are many examples in which the phenotypic, or observable, effects of one mutation in a gene is reversed by a second mutation. In some instances, the second change occurs in the same gene as did the first mutation and perhaps involves the same nucleotide as the original mutation. In this case, the second mutation is properly called a *reversion*.

In other cases, the second mutation involves a change in a quite distinct gene. Historically, this second change was described as "suppressing" the effects of the first mutation. One might naively have expected that most examples of genetic suppression would be compensatory changes in which the missing function of the first mutated gene is replaced by a terminal product that has a similar functional effect, even though it may only superficially resemble the normal product of the first gene. Such compensations are known. For example, one pigment can be replaced by another. However, a surprising number of suppressor mutations have been found to restore the original normal protein.

Before the complete mechanism of protein synthesis was elucidated, this process was not understood: how could a mutation at one gene influence the way in which a mutated triplet in another gene was translated into protein? For a time, the matter was even further confused by the discovery that suppressed mutant strains often produced both the mutant and the normal protein simultaneously. Furthermore, suppressor mutations often modified the defects exhibited by several different mutant genes.

In 1961, Yanofsky pointed out that suppressor effects could be explained by changes in the recognition of tRNA. The alterations might involve either a tRNA or the amino-acid-activating enzyme that charges the tRNA with a specified amino acid. Later investigators furnished direct proof that the suppressor genes act by producing modified tRNA's, whose confused specificity results, at least occasionally, in restoring the original amino-acid sequence to the product of a gene whose code had been altered by mutation. Other workers obtained similar results by the artificial chemical modifications of tRNA's. Consequently, the suppressor phenomenon, once very mysterious, has been virtually entirely duplicated in experimental systems. It has also been found that the locus of the suppressor mutation is very likely the gene that specifies the informational sequence of the corresponding tRNA. See Fig. 6.

ISOLATION AND CHEMICAL SYNTHESIS OF GENES

As the previous discussion has elaborated, the boundaries of a gene are somewhat arbitrary, being defined by how a segment of DNA functions rather than by any obvious structural demarcation. In this light, the challenge of "isolating" a gene is a rather subtle one.

In a certain sense, the simple viruses provide a ready-made answer to that challenge. Polio virus and tobacco mosaic virus were each crystallized sometime ago, assuring us of the isolation of a homogeneous set of particles, each containing identical genetic information to the extent of some 7,500 nucleotide units of RNA. This would correspond to about 25 message genes. Attempts to fractionate DNA from bacterial cells in order to separate segments with specific genetic activity have also been made. This has been possible to some extent on the basis of variations in average base composition, which lead to differences in the molecular density, melting temperature, and other properties of different molecules.

More recently, the technique of molecular hybridization—the artificial formation of DNA double helices from the combination of different single strands in solution—has been used to facilitate the isolation of specific segments. A team of Harvard Medical School scientists, headed by Jonathan Beckwith and James Shapiro, in 1969, made molecular hybrids of DNA strands from two different viruses carrying overlapping parts of the bacterial DNA. The overlapping segment, which included a few message genes and a few operator genes for the *lactase* function, was then separated.

In another advance, a group including D. J. Brenner, M. J. Fournier, and B. P. Doctor at the Walter Reed Medical Research Center in Washington used a purfied transfer tRNA to isolate the corresponding DNA "genes" from a bacterium. This is the nearest approach to the authentic "isolation" of a specific single gene. Much more work is needed to establish the purity and chemical homogeneity of these preparations, which may eventually be indispensable for further studies of genetic specificity.

Meanwhile, Gobind Khorana has achieved the chemical synthesis of a string of DNA calculated to have the chemical composition of the gene for alanine tRNA. The chemical structure of this RNA, which is 77 nucleotides long, had previously been worked out by Robert W. Holley (1968 Nobel Prize winner), and the structure of the corresponding DNA gene could be inferred by applying coding rules. Here again, biological studies are still needed to corroborate the precision with which these chemically identified fractions may properly be identified as "synthetic genes."

TEST TUBE SYNTHESIS OF GENES

One of the capstones of molecular genetics was the enzymatic synthesis of a virus DNA, reported in December 1967 by the 1959 Nobel laureate Arthur Kornberg and his associates. Kornberg and his colleagues had studied the enzymes of DNA synthesis since 1955. They found that the products of DNA polymerase (an enzyme needed for DNA replication) action in the test tube consistently lost their biological activity. However, chemical assays revealed that the polymerase faithfully copied the grosser aspects of the DNA used as a starting template. The discrepancy suggested that lesions were introduced into the DNA during enzymatic replication, perhaps by other DNA-degrading enzymes that tend to contaminate polymerase preparations.

SUPPRESSOR PHENOMENON

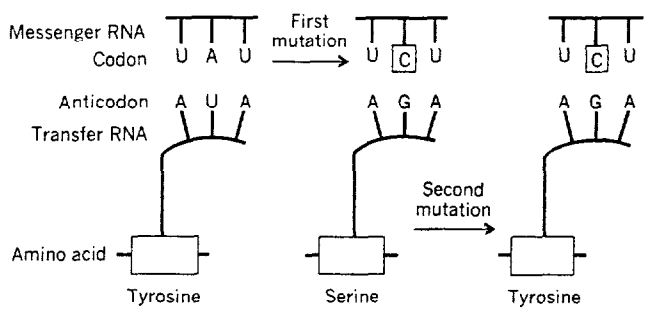


Fig 6. Suppressor genes reverse the effects of mutations in other genes by changing the specificity of t-RNA. The triplet UAU normally codes for tyrosine, most likely by binding the anticodon AUA of tyrosine t-RNA. Mutated to UCU, it codes for serine, but through a second mutation, the serine t-RNA carries tyrosine instead of serine, thus restoring the original amino acid.

Further work in several laboratories then uncovered another enzyme, DNA-ligase, which heals nicks that may interrupt the continuity of a DNA strand. The ligase proved to be the essential key to the successful replication of the DNA of the bacterial virus ϕ^x-174. This virus has a remarkable DNA code, a single molecule consisting of a single strand of DNA just 5,500 nucleotides long and closed into a continuous circle. These features are important technical advantages for experiments on DNA replication.

For copying the virus, two cycles of replication were needed. The first makes a complementary strand of DNA, closed by the application of the ligase. A repeated cycle recomplements the DNA to produce a new assembly that is an exact copy of the original virus information. This is corroborated by demonstrating that the copy can infect cells in the same manner as the original so as to generate indistinguishable harvests of the ϕ^x-174 virus.

Following this work, A. T. Ganesan reported in November 1968 that the typically double-stranded DNA of the bacterium *Bacillus subtilis* could also be replicated by an analogous process. In the process, a DNA-polymerase enzyme preparation was extracted from the bacterial cell membrane that appears to be the natural site of DNA synthesis.

These reports have been described as the creation of living DNA in a test tube. In a deeper sense, they are a demonstration of the process by which DNA is copied within the cell. Thus they open the door to more profound studies of alterations in DNA than could be done in the complex milieu of the intact cell. The techniques described also mesh closely with the methods of organic chemical synthesis developed by Gobind Khorana for the assembly of predesigned DNA sequences.

GENE DAMAGE, MUTATION, AND REPAIR

The remarkable properties of DNA come from the evolution of cellular mechanisms for its accurate replication and transcription into a protein molecule. Errors in the replication or transcription of DNA do occasionally occur, however, and these are known as mutations.

Like the letters of an alphabet, the single units of the DNA molecule are much less specific or interesting than the words they can compose. This is consonant with the lack of any reliable evidence that a specific gene ever "knows how" to mutate in response to the needs of the organism in a new environment. Genes do not mutate to adapt to new environmental conditions. Instead, mutations occur as random errors, and evolution operates retrospectively on the results of the random errors, which are usually disastrous. A previously mutated gene may prove beneficial in a new environment, and the organisms carrying such a gene will then successfully reproduce.

Many mutations are introduced into DNA by reactions with radiation, by artificial or natural chemicals, and by occasional mistakes in DNA replication. Living cells have evolved mechanisms for restoring to normal a large proportion of such mistakes provided they occur in one strand of the double-stranded DNA. Evidently, the strands can be compared with each other. Also, enzymes are available that can excise discrepant regions and resynthesize the correct sequence from the complementary information carried on the other strand and close the gap with the activity of DNA-ligase (an enzyme that heals nicks in DNA strands). Lesions that involve both strands are less likely to be repaired, if they can be repaired at all.

The human genetic disease xeroderma pigmentosum has been described as a defect in the repair mechanism, possibly a miscoding in the message gene for one of the repair enzymes. Afflicted patients are extremely sensitive to sunlight, suffer unrepaired damage to DNA in their cells, and are prone to develop multiple tumors.

In normal individuals, a certain number of DNA errors still escape the repair mechanisms. This fact imposes on us the obligation to protect our genetic endowment from unnecessary exposure to possible mutagens, such as ionizing radiation. In addition to the inherited defects produced when the DNA of the germ cells is injured, mutations in tissue cells are probably a part of the normal aging process and one of the mechanisms for the initiation of cancer. See also GENETICS, HUMAN.

JOSHUA LEDERBERG
Nobel Prize Winner in Physiology or Medicine
Stanford University School of Medicine

Bibliography

Cold Spring Harbor Symposia on Quantitative Biology, *The Genetic Code*, vol. 31.
Lederberg, Joshua, "A View of Genetics," *Science*, vol. 131, pp. 269–276, 1960.
Lederberg, Joshua, "Orthobiosis—The Perfection of Man," *Nobel Symposium 14* (Stockholm 1970).
Muller, Herman J., "The Gene," *Proceedings, Royal Society of London*, series B, vol. 131, pp. 1–37.
Scientific American, *The Molecular Basis of Life—An Introduction to Molecular Biology* (San Francisco 1968).
Scientific American, *Facets of Genetics* (San Francisco 1970).
Watson, James D., *Molecular Biology of the Gene* (New York 1965).
Whitehouse, Harold L. K., *Towards an Understanding of the Mechanism of Heredity* (New York 1967).