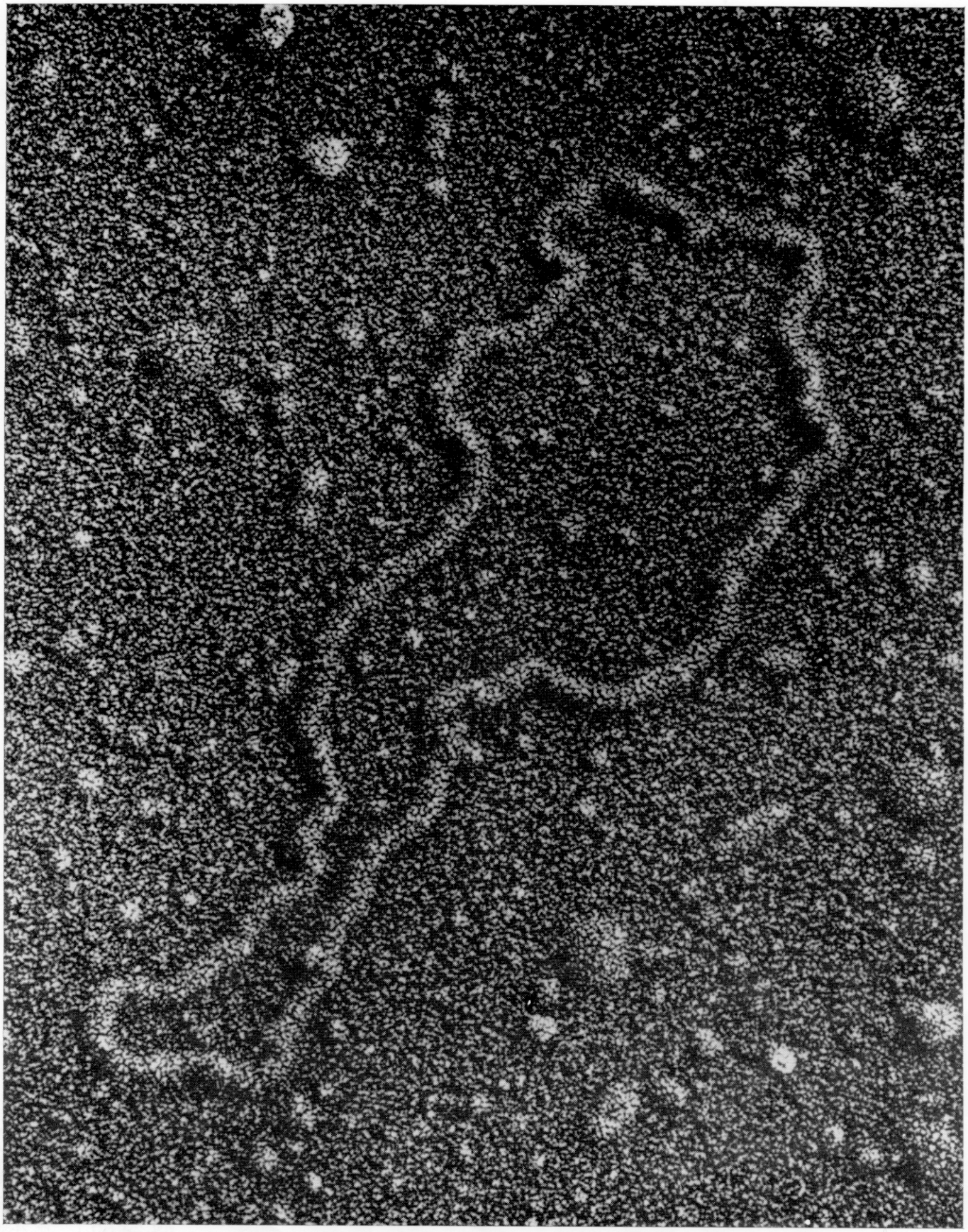
1324
SCIENTIFIC AMERICAN OFFPRINTS

THE MANIPULATION OF GENES

by Stanley N. Cohen

SCIENTIFIC AMERICAN
JULY 1975
VOL. 233, NO. 1 PP. 24-33

PUBLISHED BY **W. H. FREEMAN AND COMPANY** 660 MARKET STREET, SAN FRANCISCO, CALIFORNIA 94104



PLASMID *pSC*101 is shadowed with platinum-palladium and enlarged 230,000 diameters in an electron micrograph made by the author. A plasmid is a molecule of DNA that exists apart from the chromosome in a bacterium and replicates on its own, often carrying the genes for some supplementary activity such as resistance to antibiotics. This plasmid, a small one made by shearing a larger plasmid native to the bacterium *Escherichia coli,* is a circular, or closed-loop, molecule of DNA about three micrometers in circumference that carries the genetic information for replicating itself in *E. coli* and for conferring resistance to the antibiotic tetracycline. It was the "vehicle" for the first gene-manipulation experiments by the author and his colleagues. Foreign DNA was spliced to it and the plasmid was introduced into *E. coli,* where it replicated and expressed both its own and the foreign DNA's genetic information.

THE MANIPULATION OF GENES

Techniques for cleaving DNA and splicing it into a carrier molecule make it possible to transfer genetic information from one organism to an unrelated one. There the DNA replicates and expresses itself

by Stanley N. Cohen

Mythology is full of hybrid creatures such as the Sphinx, the Minotaur and the Chimera, but the real world is not; it is populated by organisms that have been shaped not by the union of characteristics derived from very dissimilar organisms but by evolution within species that retain their basic identity generation after generation. This is because there are natural barriers that normally prevent the exchange of genetic information between unrelated organisms. The barriers are still poorly understood, but they are of fundamental biological importance.

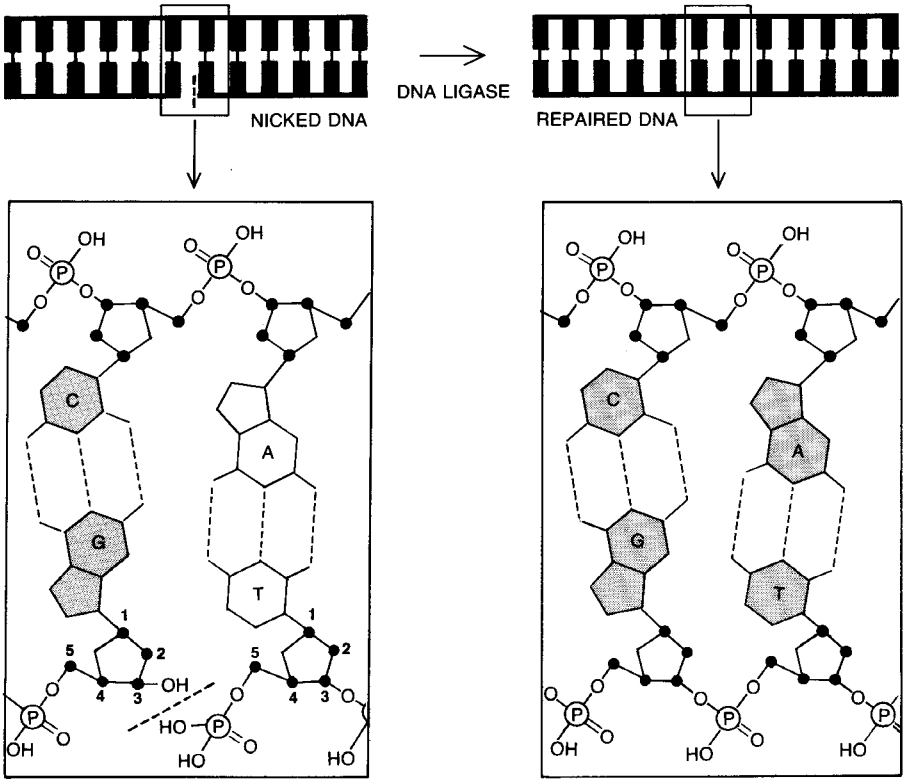
The basic unit of biological relatedness is the species, and in organisms that reproduce sexually species are defined by the ability of their members to breed with one another. Species are determined and defined by the genes they carry, so that in organisms that reproduce asexually the concept of species depends on nature's ability to prevent the biologically significant exchange of genetic material—the nucleic acid DNA—between unrelated groups.

The persistence of genetic uniqueness is perhaps most remarkable in simple organisms such as bacteria. Even when they occupy the same habitat most bacterial species do not exchange genetic information. Even rather similar species of bacteria do not ordinarily exchange the genes on their chromosomes, the structures that carry most of their genetic information. There are exceptions, however. There are bits of DNA, called plasmids, that exist apart from the chromosomes in some bacteria. Sometimes a plasmid can pick up a short segment of DNA from the chromosome of its own cell and transfer it to the cell of a related bacterial species, and sometimes the plasmid and the segment of chromosomal DNA can become integrated into the chromosome of the recipient cell. This transfer of genes between species by extrachromosomal elements has surely played some role in bacterial evolution, but apparently it has not been widespread in nature. Otherwise the characteristics of the common bacterial species would not have remained so largely intact over the huge number of bacterial generations that have existed during the era of modern bacteriology.

In 1973 Annie C. Y. Chang and I at the Stanford University School of Medicine and Herbert W. Boyer and Robert B. Helling at the University of California School of Medicine at San Francisco reported the construction in a test tube of biologically functional DNA molecules that combined genetic information from two different sources. We made the molecules by splicing together segments of two different plasmids found in the colon bacillus *Escherichia coli* and then inserting the composite DNA into *E. coli* cells, where it replicated itself and expressed the genetic information of both parent plasmids. Soon afterward we introduced plasmid genes from an unrelated bacterial species, *Staphylococcus aureus,* into *E. coli,* where they too expressed the biological properties they had displayed in their original host; then, applying the same procedures with John F. Morrow of Stanford and Howard M. Goodman in San Francisco, we were able to insert into *E. coli* some genes from an animal: the toad *Xenopus laevis.*

We called our composite molecules DNA chimeras because they were conceptually similar to the mythological Chimera (a creature with the head of a lion, the body of a goat and the tail of a serpent) and were the molecular counterparts of hybrid plant chimeras produced by agricultural grafting. The procedure we described has since been used and extended by workers in several laboratories. It has been called plasmid engineering, because it utilizes plasmids to introduce the foreign genes, and molecular cloning, because it provides a way to propagate a clone, or line of genetically alike organisms, all containing identical composite DNA molecules. Because of the method's potential for creating a wide variety of novel genetic combinations in microorganisms it is also known as genetic engineering and genetic manipulation. The procedure actually consists of several distinct biochemical and biological manipulations that were made possible by a series of independent discoveries made in rapid succession in the late 1960's and early 1970's. There are four essential elements: a method of breaking and joining DNA molecules derived from different sources; a suitable gene carrier that can replicate both itself and a foreign DNA segment linked to it; a means of introducing the composite DNA molecule, or chimera, into a functional bacterial cell, and a method of selecting from a large population of cells a clone of recipient cells that has acquired the molecular chimera.

In 1967 DNA ligases—enzymes that can repair breaks in DNA and under certain conditions can join together the loose ends of DNA strands—were discovered almost simultaneously in five laboratories. A DNA strand is a chain of nucleotides, each consisting of a deoxyribose sugar ring, a phosphate group and one of four organic bases: adenine, thymine, guanine and cytosine. The sugars and phosphates form the backbone of the strand, from which the bases project. The individual nucleotide building blocks are connected by phosphodiester bonds between the carbon atom at position No. 3 on one sugar and the carbon atom at position No. 5 on the adjacent sugar. Double-strand DNA, the form found in most organisms, consists of two



DNA LIGASE is an enzyme that repairs "nicks," or breaks in one strand of a double-strand molecule of DNA (*top*). A strand of DNA is a chain of nucleotides (*bottom*), each consisting of a deoxyribose sugar and a phosphate group and one of four organic bases: adenine (*A*), thymine (*T*), guanine (*G*) and cytosine (*C*). The sugars and phosphates constitute the backbone of the strand, and paired bases, linked by hydrogen bonds (*broken black lines*), connect two strands. The ligase catalyzes synthesis of a bond at the site of the break (*broken colored line*) between the phosphate of one nucleotide and the sugar of the next nucleotide.

chains of nucleotides linked by hydrogen bonds between their projecting bases. The bases are complementary: adenine (*A*) is always opposite thymine (*T*), and guanine (*G*) is always opposite cytosine (*C*). The function of the ligase is to repair "nicks," or breaks in single DNA strands, by synthesizing a phosphodiester bond between adjoining nucleotides [*see illustration above*].

In 1970 a group working in the laboratory of H. Gobind Khorana, who was then at the University of Wisconsin, found that the ligase produced by the bacterial virus T4 could sometimes catalyze the end-to-end linkage of completely separated double-strand DNA segments. The reaction required that the ends of two segments be able to find each other; such positioning of two DNA molecules was a matter of chance, and so the reaction was inefficient. It was clear that efficient joining of DNA molecules required a mechanism for holding the two DNA ends together so that the ligase could act.

An ingenious way of accomplishing this was developed and tested independently in two laboratories at Stanford: by Peter Lobban and A. Dale Kaiser and by David Jackson, Robert Symons and Paul Berg. Earlier work by others had shown that the ends of the DNA molecules of certain bacterial viruses can be joined by base-pairing between complementary sequences of nucleotides that are naturally present on single-strand segments projecting from the ends of those molecules: *A*'s pair with *T*'s, *G*'s pair with *C*'s and the molecules are held together by hydrogen bonds that form between the pairs. The principle of linking DNA molecules by means of the single-strand projections had been exploited in Khorana's laboratory for joining short synthetic sequences of nucleotides into longer segments of DNA.

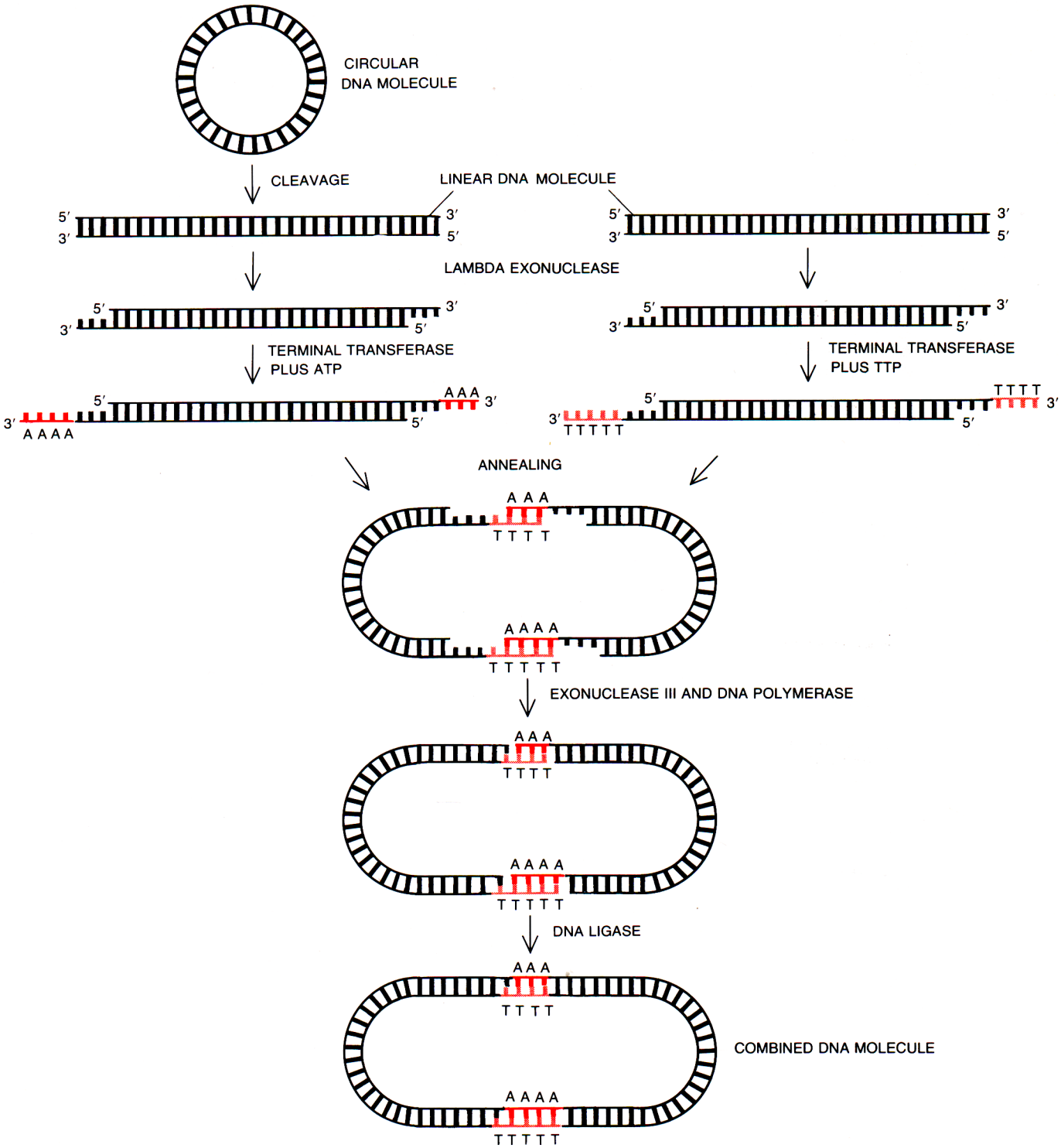
The Stanford groups knew too that an enzyme, terminal transferase, would catalyze the stepwise addition, specifically at what are called the 3′ ends of single strands of DNA, of a series of identical nucleotides. If the enzyme worked also with double-strand DNA, then a block of identical nucleotides could be added to one population of DNA molecules and a block of the complementary nucleotides could be added to another population from another source. Molecules of the two populations could then be annealed by hydrogen bonding and sealed together by DNA ligase. The method was potentially capable of joining any two species of DNA. While Lobban and Kaiser tested the terminal-transferase procedure with the DNA of the bacterial virus P22, Jackson, Symons and Berg applied the procedure to link the DNA of the animal virus SV40 to bacterial-virus DNA.

The SV40 and bacterial-virus DNA molecules Berg's group worked with are closed loops, and the loops had first to be cleaved to provide linear molecules with free ends for further processing and linkage [*see illustration on opposite page*]. (As it happened, the particular enzyme chosen to cleave the loops was the *Eco RI* endonuclease, which was later to be used in a different procedure for making the first biologically functional gene combinations. At the time, however, the enzyme's special property of producing complementary single-strand ends all by itself had not yet been discovered.)

The cleaved linear molecules were treated with an enzyme, produced by the bacterial virus lambda, called an exonuclease because it operates by cutting off nucleotides at the end of a DNA molecule. The lambda exonuclease chewed back the 5′ ends of DNA molecules and thus left projecting single-strand ends that had 3′ termini to which the blocks of complementary nucleotides could be added. The next step was to add, with the help of terminal transferase, a block of *A*'s at the 3′ end of one of the two DNA species to be linked and a block of *T*'s at the 3′ ends of the other species. The species were mixed together. Fragments having complementary blocks at their ends could find each other, line up and become annealed by hydrogen bonding, thus forming combined molecules. To fill the gaps at the 5′ ends of the original segments the investigators supplied nucleotides and two more enzymes: exonuclease III and DNA polymerase. Finally the nicks in the molecules were sealed with DNA ligase.

The method of making cohesive termini for joining DNA molecules in the first successful genetic-manipulation experiments was conceptually and operationally different from the terminal-transferase procedure. It was also much simpler. It depended on the ability of one of a group of enzymes called restriction endonucleases to make complementary-ended fragments during the cleavage of DNA at a site within the molecule, instead of requiring the addition of new blocks of complementary nucleotides to DNA termini.

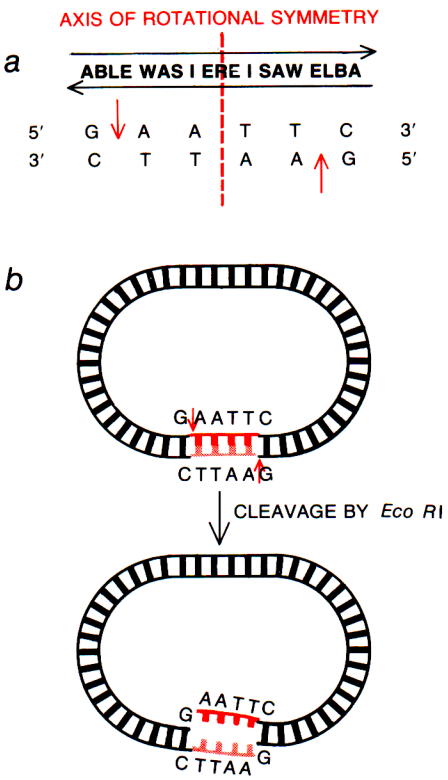
Viruses grown on certain strains of *E. coli* were known to be restricted in their ability to grow subsequently on other strains. Investigations had shown that this restriction was due to bacterial enzymes that recognize specific sites on a "foreign" viral DNA and cleave that DNA. (To protect its own DNA the bacterial cell makes a modification enzyme that adds methyl groups to nucleotides constituting the recognition sites for the restriction endonuclease, making them resistant to cleavage.) Restriction endonucleases (and modification methylases) are widespread in microorganisms; genes for making them were found on viral chromosomes and extrachromosomal plasmid DNA as well as on many bacterial chromosomes. During the early 1970's the nucleotide sequences at the cleavage sites recognized by several re-



TERMINAL-TRANSFERASE procedure for joining DNA molecules involves a number of steps, each dependent on a different enzyme. If one of the molecules to be joined is a closed loop, it must first be cleaved. The linear molecules are treated with lambda exonuclease, an enzyme that cuts nucleotides off the 5′ end of DNA strands (the end with a phosphate group on the No. 5 carbon). Then specific nucleotides are added to the 3′ end (the end with an OH group on the No. 3 carbon) by the action of the enzyme terminal transferase. One DNA species is supplied with adenosine triphosphate (ATP), the other with thymidine triphosphate (TTP), so that *A* nucleotides are added to one species and complementary *T* nucleotides to the other. When the two species are mixed, the complementary bases pair up, annealing the molecules. Nucleotides and the enzymes DNA polymerase and exonuclease III are added to fill gaps and DNA ligase is added to seal the DNA backbones. The result is a double molecule composed of two separate DNA segments.

striction endonucleases were identified. In every instance, it developed, the cleavage was at or near an axis of rotational symmetry: a palindrome where the nucleotide base sequences read the same on both strands in the 5′-to-3′ direction [*see illustration below*].

In some instances the breaks in the DNA strands made by restriction enzymes were opposite each other. One particular endonuclease, however, the *Eco RI* enzyme isolated by Robert N. Yoshimori in Boyer's laboratory in San Francisco, had a property that was of special interest. Unlike the other nucleases known at the time, this enzyme introduced breaks in the two DNA strands that were separated by several nucleotides. Because of the symmetrical, palindromic arrangement of the nucleotides in the region of cleavage this separation of the cleavage points on the two strands yielded DNA termini with projecting complementary nucleotide sequences: "sticky" mortise-and-tenon ter-



RESTRICTION ENDONUCLEASES cleave DNA at sites where complementary nucleotides are arranged in rotational symmetry: a palindrome, comparable to a word palindrome (*a*). The endonuclease *Eco RI* has the additional property of cleaving complementary strands of DNA at sites (*colored arrows*) four nucleotides apart. Such cleavage (*b*) yields DNA fragments with complementary, overlapping single-strand ends. As a result the end of any DNA fragment produced by *Eco RI* cleavage can anneal with any other fragment produced by the enzyme.

mini. The *Eco RI* enzyme thus produced in one step DNA molecules that were functionally equivalent to the cohesive-end molecules produced by the complicated terminal-transferase procedure.

The experiments that led to the discovery of the capabilities of *Eco RI* were reported independently and simultaneously in November, 1972, by Janet Mertz and Ronald W. Davis of Stanford and by another Stanford investigator, Vittorio Sgaramella. Sgaramella found that molecules of the bacterial virus P22 could be cleaved with *Eco RI* and would then link up end to end to form DNA segments equal in length to two or more viral-DNA molecules. Mertz and Davis observed that closed-loop SV40-DNA molecules cleaved by *Eco RI* would reform themselves into circular molecules by hydrogen bonding and could be sealed with DNA ligase; the reconstituted molecules were infectious in animal cells growing in tissue culture. Boyer and his colleagues analyzed the nucleotide sequences at the DNA termini produced by *Eco RI*, and their evidence confirmed the complementary nature of the termini, which accounted for their cohesive activity.

In late 1972, then, several methods were available by which one could join double-strand molecules of DNA. That was a major step in the development of a system for manipulating genes. More was necessary, however. Most segments of DNA do not have an inherent capacity for self-replication; in order to reproduce themselves in a biological system they need to be integrated into DNA molecules that can replicate in the particular system. Even a DNA segment that can replicate in its original host was not likely to have the specific genetic signals required for replication in a different environment. If foreign DNA was to be propagated in bacteria, as had long been proposed in speculative scenarios of genetic engineering, a suitable vehicle, or carrier, was required. A composite DNA molecule consisting of the vehicle and the desired foreign DNA would have to be introduced into a population of functional host bacteria. Finally, it would be necessary to select, or identify, those cells in the bacterial population that took up the DNA chimeras. In 1972 it still seemed possible that the genetic information on totally foreign DNA molecules might produce an aberrant situation that would prevent the propagation of hybrid molecules in a new host.

Molecular biologists had focused for many years on viruses and their relations with bacteria, and so it was natural that bacterial viruses were thought of as the most likely vehicles for genetic manipulation. For some time there had been speculation and discussion about using viruses, such as lambda, that occasionally acquire bits of the *E. coli* chromosome by natural recombination mechanisms for cloning DNA from foreign sources. It was not a virus, however, but a plasmid that first served as a vehicle for introducing foreign genes into a bacterium and that provided a mechanism for the replication and selection of the foreign DNA.

A ubiquitous group of plasmids that confer on their host bacteria the ability to resist a number of antibiotics had been studied intensively for more than a decade. Antibiotic-resistant *E. coli* isolated in many parts of the world, for example, were found to contain plasmids, designated *R* factors (for "resistance"), carrying the genetic information for products that in one way or another could interfere with the action of specific antibiotics [see "Infectious Drug Resistance," by Tsutomu Watanabe; SCIENTIFIC AMERICAN, December, 1967]. Double-strand circular molecules of *R*-factor DNA had been separated from bacterial chromosomal DNA by centrifugation in density gradients and had been characterized by biochemical and physical techniques [see "The Molecule of Infectious Drug Resistance," by Royston C. Clowes; SCIENTIFIC AMERICAN, April, 1973].

In 1970 Morton Mandel and A. Higa of the University of Hawaii School of Medicine had discovered that treatment of *E. coli* with calcium salts enabled the bacteria to take up viral DNA. At Stanford, Chang and I, with Leslie Hsu, found that if we made the cell membranes of *E. coli* permeable by treating them with calcium chloride, purified *R*-factor DNA could be introduced into them [*see illustration on opposite page*]. The *R*-factor DNA is taken up in this transformation process by only about one bacterial cell in a million, but those few cells can be selected because they live and multiply in the presence of the antibiotics to which the *R* factor confers resistance, whereas other cells die. Each transformed cell gives rise to a clone that contains exact replicas of the parent plasmid DNA molecules, and so we reasoned that plasmids might serve as vehicles for propagating new genetic information in a line of *E. coli* cells.

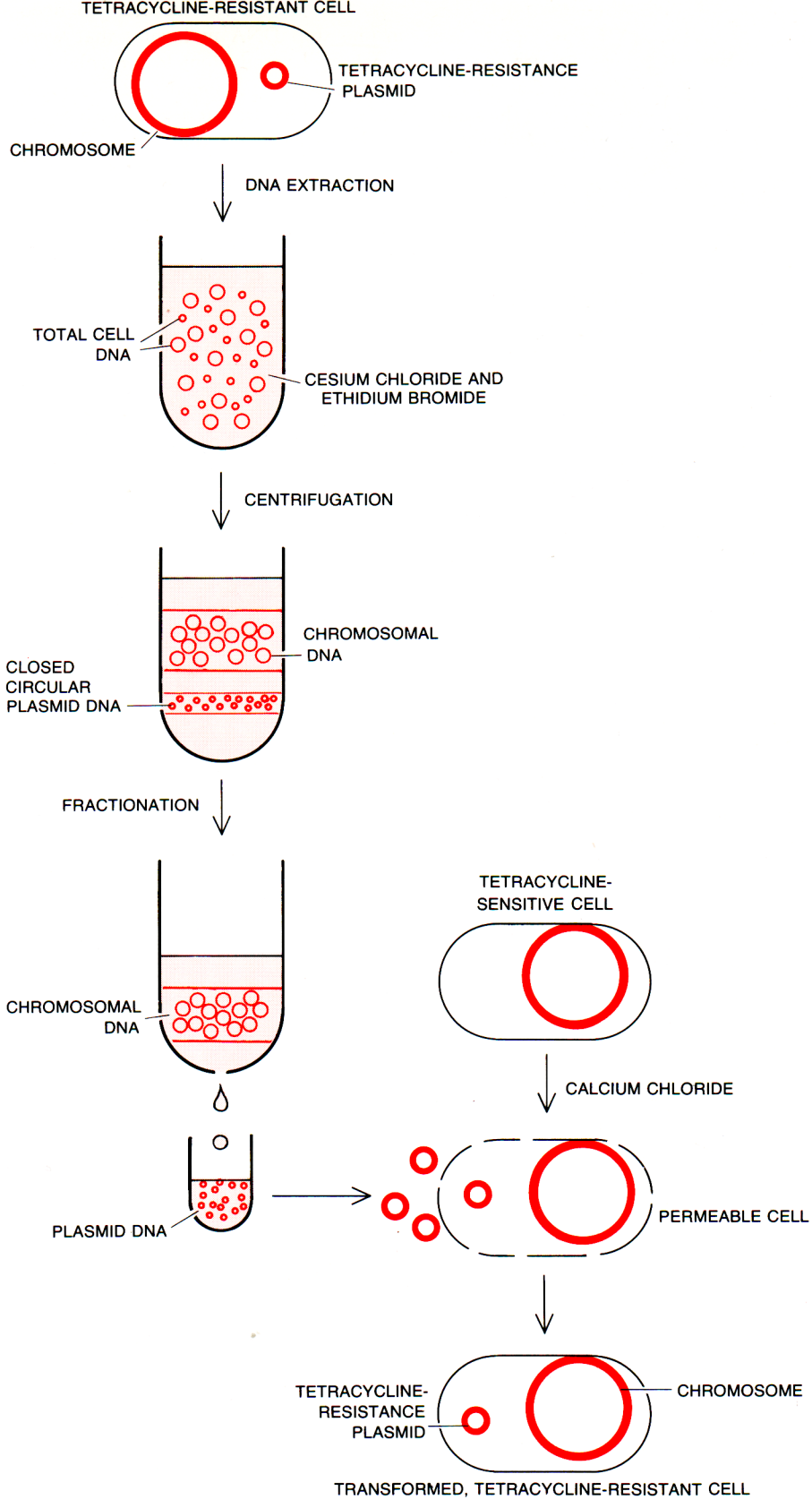
In an effort to explore the genetic and molecular properties of various regions of the *R*-factor DNA we had begun to take plasmids apart by shearing their DNA mechanically and then transforming *E. coli* with the resulting

fragments. Soon afterward we began to cleave the plasmids with the *Eco RI* enzyme, which had been shown to produce multiple site-specific breaks in several viruses. It might therefore be counted on to cleave all molecules of a bacterial plasmid in the same way, so that any particular species of DNA would yield a specific set of cleavage fragments, and do so reproducibly. The fragments could then be separated and identified according to the different rates at which they would migrate through a gel under the influence of an electric current.

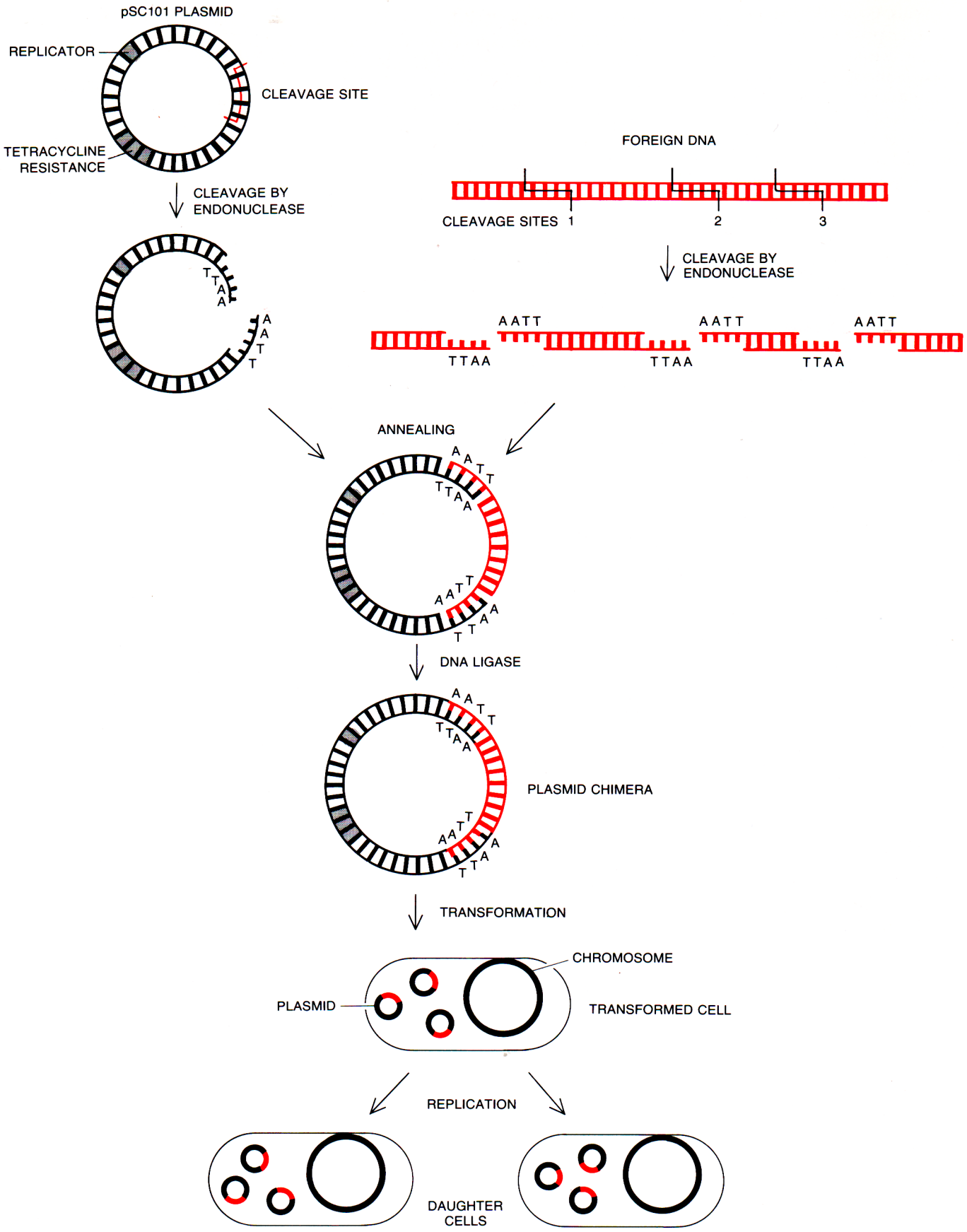
When the DNA termini produced by *Eco RI* endonuclease were found to be cohesive, Chang and I, in collaboration with Boyer and Helling in San Francisco, proceeded to search for a plasmid that the enzyme would cleave without affecting the plasmid's ability to replicate or to confer antibiotic resistance. We hoped that if such a plasmid could be found, we could insert a segment of foreign DNA at the *Eco RI* cleavage site, and that it might be possible to propagate the foreign DNA in *E. coli*.

In our collection at Stanford there was a small plasmid, *pSC101*, that had been isolated following the mechanical shearing of a large plasmid bearing genes for multiple antibiotic resistance. It was less than a twelfth as long as the parent plasmid, but it did retain the genetic information for its replication in *E. coli* and for conferring resistance to one antibiotic, tetracycline. When we subjected *pSC101* DNA to cleavage by *Eco RI* and analyzed the products by gel electrophoresis, we found that the enzyme had cut the plasmid molecule in only one place, producing a single linear fragment. We were able to join the ends of that fragment again by hydrogen bonding and reseal them with DNA ligase, and when we introduced the reconstituted circular DNA molecules into *E. coli* by transformation, they were biologically functional plasmids: they replicated and conferred tetracycline resistance.

The next step was to see if a fragment of foreign DNA could be inserted at the cleavage site without interfering with replication or expression of tetracycline resistance and thus destroying the plasmid's ability to serve as a cloning vehicle. We mixed the DNA of another *E. coli* plasmid, which carried resistance to the antibiotic kanamycin, with the *pSC101* DNA. We subjected the mixed DNA to cleavage by *Eco RI* and then to ligation, transformed *E. coli* with the resulting DNA and found that some of the transformed bacteria were indeed resist-



PLASMID DNA can be introduced into a bacterial cell by the procedure called transformation. Plasmids carrying genes for resistance to the antibiotic tetracycline (*top left*) are separated from bacterial chromosomal DNA. Because differential binding of ethidium bromide by the two DNA species makes the circular plasmid DNA denser than the chromosomal DNA, the plasmids form a distinct band on centrifugation in a cesium chloride gradient and can be separated (*bottom left*). The plasmid DNA is mixed with bacterial cells that are not resistant to tetracycline and that have been made permeable by treatment with a calcium salt. The DNA enters the cells, replicates there and makes the cells resistant to tetracycline.



FOREIGN DNA is spliced into the *pSC*101 plasmid and introduced with the plasmid into the bacterium *Escherichia coli*. The plasmid is cleaved by the endonuclease *Eco* RI at a single site that does not interfere with the plasmid's genes for replication or for resistance to tetracycline (*top left*). The nucleotide sequence recognized by *Eco* RI is present also in other DNA, so that a foreign DNA exposed to the endonuclease is cleaved about once in every 4,000 to 16,000 nucleotide pairs on a random basis (*top right*). Fragments of cleaved foreign DNA are annealed to the plasmid DNA by hydrogen bonding of the complementary base pairs, and the new composite molecules are sealed by DNA ligase. The DNA chimeras, each consisting of the entire plasmid and a foreign DNA fragment, are introduced into *E. coli* by transformation, and the foreign DNA is replicated by virtue of the replication functions of the plasmid.

ant to both tetracycline and kanamycin. The plasmids isolated from such transformants contained the entire *pSC*101 DNA segment and also a second DNA fragment that carried the information for kanamycin resistance, although it lacked replication functions of its own. The results meant that the *pSC*101 could serve as a cloning vehicle for introducing at least a nonreplicating segment of a related DNA into *E. coli*. And the procedure was extraordinarily simple.

Could genes from other species be introduced into *E. coli* plasmids, however? There might be genetic signals on foreign DNA that would prevent its propagation or expression in *E. coli*. We decided to try to combine DNA from a plasmid of another bacterium, the *pI*258 plasmid of *Staphylococcus aureus*, with our original *E. coli* plasmid. The staphylococcal plasmid had already been studied in several laboratories; we had found that it was cleaved into four DNA fragments by *Eco RI*. Since *pI*258 was not native to *E. coli* or to related bacteria, it could not on its own propagate in an *E. coli* host. And it was known to carry a gene for resistance to still another antibiotic, penicillin, that would serve as a marker for selecting any transformed clones. (Penicillin resistance, like combined resistance to tetracycline and kanamycin, was already widespread among *E. coli* strains in nature. That was important; if genes from a bacterial species that cannot normally exchange genetic information with the colon bacillus were to be introduced into it, it was essential that they carry only antibiotic-resistance traits that were already prevalent in *E. coli*. Otherwise we would be extending the species' antibiotic-resistance capabilities.)

Chang and I repeated the experiment that had been successful with two kinds of *E. coli* plasmids, but this time we did it with a mixture of the *E. coli*'s *pSC*-101 and the staphylococcal *pI*258: we cleaved the mixed plasmids with *Eco RI* endonuclease, treated them with ligase and then transformed *E. coli*. Next we isolated transformed bacteria that expressed the penicillin resistance coded for by the *S. aureus* plasmid as well as the tetracycline resistance of the *E. coli* plasmid. These doubly resistant cells were found to contain a new DNA species that had the molecular characteristics of the staphylococcal plasmid DNA as well as the characteristics of *pSC*101.

The replication and expression in *E. coli* of genes derived from an organism ordinarily quite unable to exchange genes with *E. coli* represented a breach in the barriers that normally separate biological species. The bulk of the genetic information expressed in the transformed bacteria defined it as *E. coli*, but the transformed cells also carried replicating DNA molecules that had molecular and biological characteristics derived from an unrelated species, *S. aureus*. The fact that the foreign genes were on a plasmid meant that they would be easy to isolate and purify in large quantities for further study. Moreover, there was a possibility that one might introduce genes into the easy-to-grow *E. coli* that specify a wide variety of metabolic or synthesizing functions (such as photosynthesis or antibiotic production) and that are indigenous to other biological classes. Potentially the *pSC*101 plasmid and the molecular-cloning procedure could serve to introduce DNA molecules from complex higher organisms into bacterial hosts, making it possible to apply relatively simple bacterial genetic and biochemical techniques to the study of animal-cell genes.

Could animal-cell genes in fact be introduced into bacteria, and would they replicate there? Boyer, Chang, Helling and I, together with Morrow and Goodman, immediately undertook to find out. We picked certain genes that had been well studied and characterized and were available, purified, in quantity: the genes that code for a precursor of the ribosomes (the structure on which proteins are synthesized) in the toad *Xenopus laevis*. The genes had properties that would enable us to identify them if we succeeded in getting them to propagate in bacteria. The toad DNA was suitable for another reason: although we would be constructing a novel biological combination containing genes from both animal cells and bacteria, we and others expected that no hazard would result from transplanting the highly purified ribosomal genes of a toad.

Unlike the foreign DNA's of our earlier experiments, the toad genes did not express traits (such as antibiotic resistance) that could help us to select bacteria carrying plasmid chimeras. The tetracycline resistance conferred by *pSC*101 would make it possible to select transformed clones, however, and we could then proceed to examine the DNA isolated from such clones to see if any clones contained a foreign DNA having the molecular properties of toad ribosomal DNA. The endonuclease-generated fragments of toad ribosomal DNA have characteristic sizes and base compositions; DNA from the transformed cells could be tested for those characteristics. The genes propagated in bacteria could also be tested for nucleotide-sequence homology with DNA isolated directly from the toad.

When we did the experiment and analyzed the resulting transformed cells, we found that the animal-cell genes were indeed reproducing themselves in generation after generation of bacteria by means of the plasmid's replication functions. In addition, the nucleotide sequences of the toad DNA were being transcribed into an RNA product in the bacterial cells.

Within a very few months after the first DNA-cloning experiments the procedure was being used in a number of laboratories to clone bacterial and animal-cell DNA from a variety of sources. Soon two plasmids other than *pSC*101 were discovered that have a single *Eco RI* cleavage site at a location that does not interfere with essential genes. One of these plasmids is present in many copies in the bacterial cell, making it possible to "amplify," or multiply many times, any DNA fragments linked to it. Investigators at the University of Edinburgh and at Stanford went on to develop mutants of the virus lambda (which ordinarily infects *E. coli*) that made the virus too an effective cloning vehicle. Other restriction endonucleases were discovered that also make cohesive termini but that cleave DNA at different sites from the *Eco RI* enzymes, so that chromosomes can now be taken apart and put together in various ways.

The investigative possibilities of DNA cloning are already being explored intensively. Some workers have isolated from complex chromosomes certain regions that are implicated in particular functions such as replication. Others are making plasmids to order with specific properties that should clarify aspects of extrachromosomal-DNA biology that have been hard to study. The organization of complex chromosomes, such as those of the fruit fly *Drosophila*, is being studied by cloning the animal genes in bacteria. Within the past few months methods have been developed for selectively cloning specific genes of higher organisms through the use of radioactively labeled RNA probes: instead of purifying the genes to be studied before introducing them into bacteria, one can transform bacteria with a heterogeneous population of animal-cell DNA and then isolate those genes that produce a particular species of RNA. It is also possible to isolate groups of genes that are expressed concurrently at a particular stage in the animal's development.

The potential seems to be even broader. Gene manipulation opens the pros-

pect of constructing bacterial cells, which can be grown easily and inexpensively, that will synthesize a variety of biologically produced substances such as antibiotics and hormones, or enzymes that can convert sunlight directly into food substances or usable energy. Perhaps it even provides an experimental basis for introducing new genetic information into plant or animal cells.

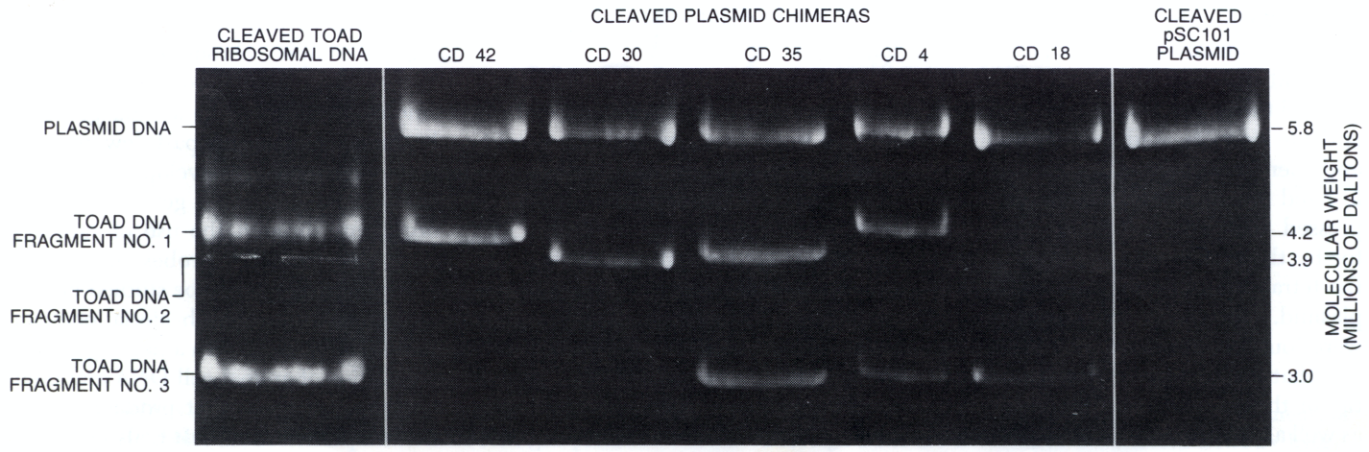
It has been clear from the beginning of experimentation in molecular cloning that the construction of some kinds of novel gene combinations may have a potential for biological hazard, and the scientific community has moved quickly to make certain that research in genetic manipulation would not endanger the public. For a time after our initial experiments the *pSC*101 plasmid was the only vehicle known to be suitable for cloning foreign DNA in *E. coli*, and our colleagues asked for supplies with which to pursue studies we knew were of major scientific and medical importance. Investigators normally facilitate the free exchange of bacteria and other experimental strains they have isolated or developed, but Chang and I were concerned that manipulation of certain genes could give rise to novel organisms whose infectious properties and ecological effects could not be predicted. In agreeing to provide the plasmid we therefore asked for assurance that our colleagues would neither introduce tumor viruses into bacteria nor create antibiotic-resistance combinations that were not already present in nature; we also asked the recipients not to send the plasmid on to other laboratories, so that we could keep track of its distribution.

When still other cloning vehicles were discovered, it became apparent that a more general mechanism for ensuring experimental safety in gene-manipulation research was advisable. The groundwork for such control had been established earlier: the National Academy of Sciences had been urged to consider the "possibility that potentially biohazardous consequences might result from widespread or injudicious use" of these techniques and had asked Paul Berg to form an advisory committee that would consider the issue. Berg too had been concerned about the potential hazards of certain kinds of experimentation for some years, and had himself decided to abandon plans to try to introduce genes from the tumor virus SV40 into bacteria because of the possible danger if the experiment were successful.

Berg brought together a number of investigators, including some who were then directly involved in molecular cloning, in the spring of 1974. In a report released in July and in a letter to leading professional journals the members of the committee expressed their "concern about the possible unfortunate consequences of indiscriminate application" of the techniques and formally asked all investigators to join them in voluntarily deferring two types of experiments (which had, as a matter of fact, been avoided by informal consensus up until that time). Experiments of Type I involved the construction of novel organisms containing combinations of toxin-producing capabilities or of antibiotic-resistance genes not found in nature. Type 2 experiments involved the introduction of DNA from tumor viruses or other animal viruses into bacteria; the committee noted that "such recombinant molecules might be more easily disseminated to bacterial populations in humans and other species, and might thus increase the incidence of cancer or other diseases."

The Academy committee was concerned largely because of our inability to assess the hazards of certain experiments accurately before the experiments were undertaken. Guidelines for safety had long been available in other areas of potentially hazardous research, such as studies involving known disease-causing bacteria and viruses, radioactive isotopes or toxic chemicals. Because of the newness of the microbial gene-manipulation methods, no such guidelines had yet been developed for work in this area, however; there was the possibility that potentially hazardous experiments might proceed before appropriate guidelines could be considered and implemented. We recognized that most work with the new methods did not and would not involve experiments of a hazardous nature but we recommended the deferral of Type I and Type II experiments until the hazards were more carefully assessed, until it was determined whether or not the work could be undertaken safely and until adequate safety precautions were available. The committee also proposed that an international meeting be held early in 1975 to consider the matter more fully.

Such a meeting was held in February at the Asilomar Conference Center near Pacific Grove, Calif. It brought together 86 American biologists and 53 investigators from 16 other countries, who spent three and a half days reviewing progress in the field of molecular cloning and formulating guidelines that would allow most types of new hereditary characteristics to be introduced into bacteria and

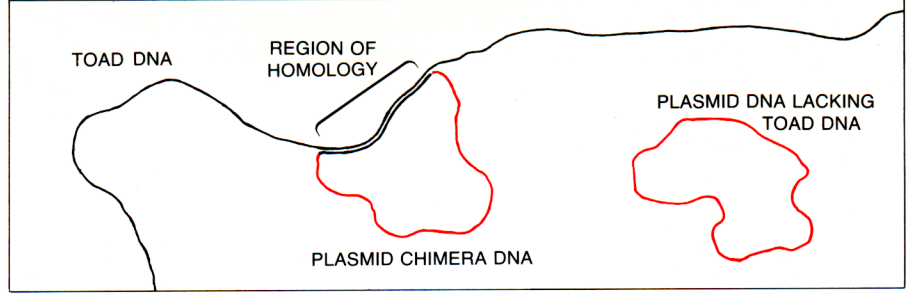


GEL ELECTROPHORESIS demonstrates the presence of toad DNA in chimeric plasmids. Fragments of DNA migrate through a gel at different rates under the influence of an electric current, depending on their size. Linear molecules of plasmid DNA (*right*) and the cleavage products of toad ribosomal DNA (*left*) therefore have characteristic sizes and migrate characteristic distances in a given time. The bands of DNA, visualized by a fluorescent dye, are photographed in ultraviolet. All five chimeric plasmids (*center*) contain a plasmid DNA molecule; in addition each chimera includes one or more fragments characteristic of original toad DNA.

viruses safely. Invited nonscientists from the fields of law and ethics participated in the discussions and decisions at Asilomar, along with representatives of agencies that provide Federal funds for scientific research; the meetings were open to the press and were fully reported. The issues were complex and there were wide differences of opinion on many of them, but there was consensus on three major points. First, the newly developed cloning methods offer the prospect of dealing with a wide variety of important scientific and medical problems as well as other problems that trouble society, such as environmental pollution and food and energy shortages. Second, the accidental dissemination of certain novel biological combinations may present varying degrees of potential risk. The construction of such combinations should proceed only under a graded series of precautions, principally biological and physical barriers, adequate to prevent the escape of any hazardous organisms; the extent of the actual risk should be explored by experiments conducted under strict containment conditions. Third, some experiments are potentially too hazardous to be carried out for the present, even with the most careful containment. Future research and experience may show that many of the potential hazards considered at the meeting are less serious and less probable than we now suspect. Nevertheless, it was agreed that standards of protection should be high at the beginning and that they can be modified later if the assessment of risk changes.

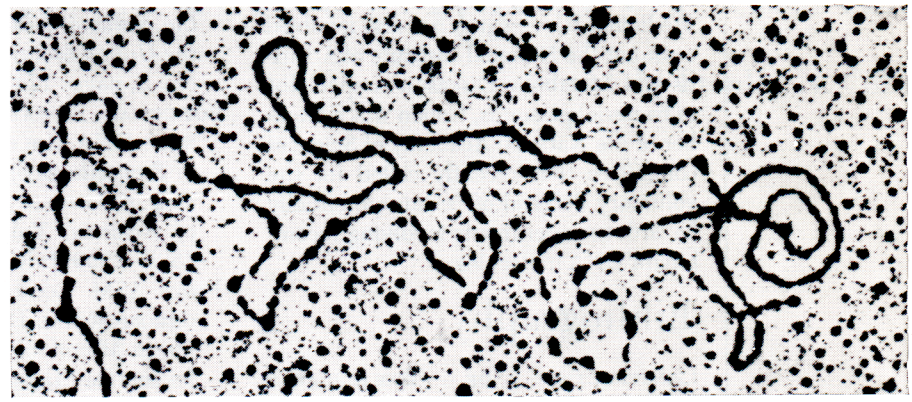
Physical containment barriers have long been used in the U.S. space-exploration program to minimize the possibility of contamination of the earth by extraterrestrial microbes. Containment procedures are also employed routinely to protect laboratory workers and the public from hazards associated with radioactive isotopes and toxic chemicals and in work with disease-causing bacteria and viruses. The Asilomar meeting formulated the additional concept of biological barriers, which involve fastidious cloning vehicles that are able to propagate only in specialized hosts and equally fastidious bacterial strains that are unable to live except under stringent laboratory conditions.

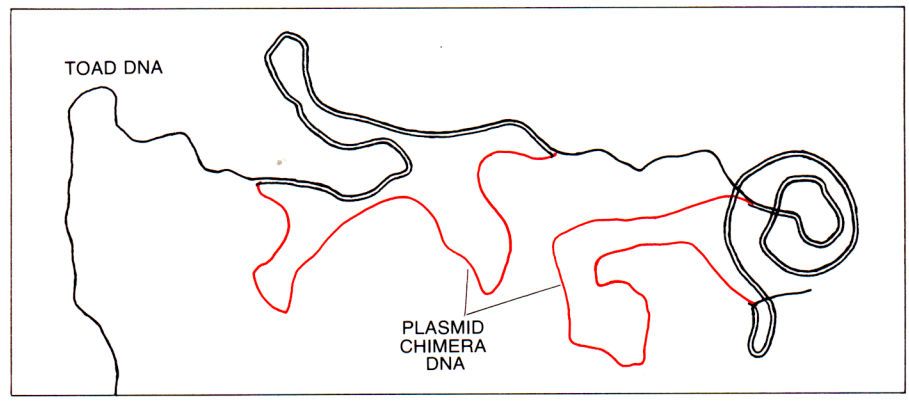
In the past the scientific community has commonly policed its own actions informally, responding to ethical concerns with self-imposed restraint. Usually, but not always, society at large has also considered the public well-being in determining how knowledge obtained by basic scientific research should be applied. Extensive public scrutiny and



HETERODUPLEX ANALYSIS identifies regions of a toad DNA (*black*) that have been incorporated in a chimeric plasmid DNA molecule. DNA isolated from toad eggs and the DNA of the chimera are denatured, that is, each natural double-strand molecule is split into two single strands of DNA, by alkali treatment. The toad and the chimeric DNA's are mixed together, and any complementary sequences are allowed to find each other. The toad DNA incorporated in the chimeras has nucleotide sequences that are complementary to sequences in the DNA taken directly from the animal source. Those homologous sequences anneal to form heteroduplex double-strand DNA that can be identified in electron micrographs.

open discussion by scientists and nonscientists of the possible risks and benefits of a particular line of basic research has been rare, however, when (as in this case) the hazards in question are only potential and, for some experiments, even hypothetical. As this article is being written it is still too early to know what the long-range outcome of the public discussions initiated by scientists working in genetic manipulation will be. One can hope that the forthright approach and the rigorous standards that have been adopted for research in the cloning of recombinant DNA molecules will promote a sharper focus on other issues relevant to public and environmental safety.





PRESENCE OF TOAD DNA in two separate chimeric plasmid molecules is demonstrated by an electron micrograph made by John F. Morrow at the Stanford University School of Medicine. As is indicated in the drawing (*bottom*), there are DNA strands from two plasmids and a strand of toad DNA. The micrograph shows thickened regions of DNA where nucleotide sequences are homologous and two single strands have been annealed. The toad DNA in the chimeras codes for ribosomes, and the space between the two heteroduplex regions is compatible with the spacing of multiple ribosomal genes in toad DNA.

The Author

STANLEY N. COHEN is associate professor of medicine at the Stanford University School of Medicine. A graduate of Rutgers University and the University of Pennsylvania School of Medicine, he joined the Stanford faculty in 1968 after spending several years teaching and doing research in molecular biology at the Albert Einstein College of Medicine. His research has also involved a stint at the National Institute of Arthritis and Metabolic Diseases. A specialist in bacterial plasmids, Cohen was a member of the National Academy of Sciences committee that recently called for the voluntary deferral of certain potentially hazardous experiments involving recombinant DNA molecules.

Bibliography

CONSTRUCTION OF BIOLOGICALLY FUNCTIONAL BACTERIAL PLASMIDS *IN VITRO*. Stanley N. Cohen, Annie C. Y. Chang, Herbert W. Boyer and Robert B. Helling in *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 70, No. 11, pages 3240–3244; November, 1973.

REPLICATION AND TRANSCRIPTION OF EUKARYOTIC DNA IN *ESCHERICHIA COLI*. John F. Morrow, Stanley N. Cohen, Annie C. Y. Chang, Herbert W. Boyer, Howard M. Goodman and Robert B. Helling in *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 71, No. 5, pages 1743–1747; May, 1974.

POTENTIAL BIOHAZARDS OF RECOMBINANT DNA MOLECULES. P. Berg et al. in *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 71, No. 7, pages 2595–2599; July, 1974.

FIRST ASM CONFERENCE ON EXTRACHROMOSOMAL ELEMENTS IN BACTERIA in *Microbiology–1974*. American Society for Microbiology, 1975.

REPORT OF THE WORKING PARTY ON THE EXPERIMENTAL MANIPULATION OF MICROORGANISMS. Her Majesty's Stationery Office, London, 1975.

1466
SCIENTIFIC AMERICAN OFFPRINTS

Useful Proteins from Recombinant Bacteria

by Walter Gilbert and Lydia Villa-Komaroff

SCIENTIFIC AMERICAN
APRIL 1980
VOL. 242, NO. 4 PP. 74-94

PUBLISHED BY **W. H. FREEMAN AND COMPANY** 41 MADISON AVENUE, NEW YORK, NEW YORK 10010

Useful Proteins from Recombinant Bacteria

Bacteria into which nonbacterial genes have been introduced are able to manufacture nonbacterial proteins. Among the proteins made by recombinant-DNA methods are insulin and interferon

by Walter Gilbert and Lydia Villa-Komaroff

A living cell is a protein factory. It synthesizes the enzymes and other proteins that maintain its own integrity and physiological processes, and (in multicelled organisms) it often synthesizes and secretes other proteins that perform some specialized function contributing to the life of the organism as a whole. Different kinds of cells make different proteins, following instructions encoded in the DNA of their genes. Recent advances in molecular biology make it possible to alter those instructions in bacterial cells, thereby designing bacteria that can synthesize nonbacterial proteins. The bacteria are "recombinants." They contain, along with their own genes, part or all of a gene from a human cell or other animal cell. If the inserted gene is one for a protein with an important biomedical application, a culture of the recombinant bacteria, which can be grown easily and at low cost, will serve as an efficient factory for producing that protein.

Many laboratories in universities and in an emerging "applied genetics" industry are working to design bacteria able to synthesize such nonbacterial proteins. A growing tool kit of "genetic engineering" techniques makes it possible to isolate one of the million-odd genes of an animal cell, to fuse that gene with part of a bacterial gene and to insert the combination into bacteria. As those bacteria multiply they make millions of copies of their own genes and of the animal gene inserted among them. If the animal gene is fused to a bacterial gene in such a way that a bacterium can treat the gene as one of its own, the bacteria will produce the protein specified by the animal gene. New ways of rapidly and easily determining the exact sequence of the chemical groups that constitute a molecule of DNA make it possible to learn the detailed structure of such "cloned" genes. After the structure is known it can be manipulated to produce DNA structures that function more efficiently in the bacterial cell.

In this article we shall first describe some of these techniques in a general way and then tell how we and our colleagues Argiris Efstratiadis, Stephanie Broome, Peter Lomedico and Richard Tizard applied them in our laboratory at Harvard University to copy a rat gene that specifies the hormone insulin, to insert the gene into bacteria and to get the bacteria to manufacture a precursor of insulin. In an exciting application of this technology Charles Weissmann and his colleagues at the University of Zurich recently constructed bacteria that produce human interferon, a potentially useful antiviral protein.

DNA, RNA and Proteins

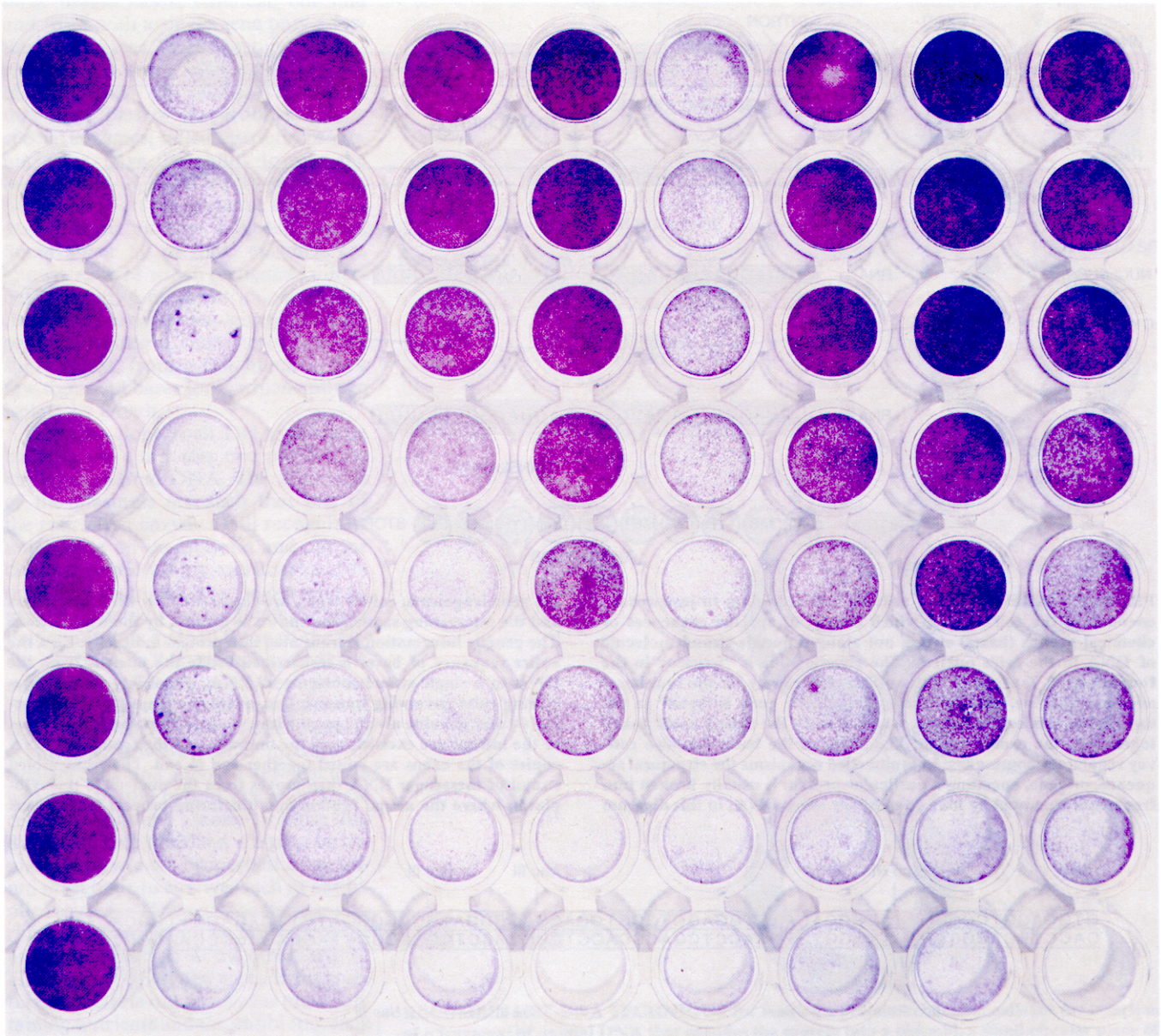
Cells make proteins by translating a set of commands arrayed along a strand of DNA. This hereditary information is held in the order of four chemical groups along the DNA: the bases adenine, thymine, guanine and cytosine. In sets of threes along DNA these bases specify which amino acids, the fundamental building blocks of proteins, are to be used in putting the protein together; the correspondence between specific base triplets and particular amino acids is called the genetic code. The part of a DNA molecule that incorporates the information to specify the structure of a protein is called a structural gene.

To act on this information the cell copies the sequence of bases from its genetic storehouse in DNA into another molecule: messenger RNA. A strand of DNA serves as a template for the assembly of a complementary strand of RNA according to base-pairing rules: adenine always pairs with uracil (which in RNA replaces DNA's thymine) and guanine pairs with cytosine. In animal cells transcription takes place in the nucleus of the cell. The messenger-RNA molecules carry the information out of the nucleus into the cytoplasm, where a complex molecular machine translates it into protein by linking together the appropriate amino acids. In bacteria, which have no nucleus, transcription and translation take place concurrently. The messenger RNA serves as a temporary set of instructions. Which proteins the cell makes depends on which messengers it contains at any given time; to make a different protein the cell makes a new messenger from the appropriate structural gene. The DNA in each cell contains all the information required at any time by any cell of the organism, but each cell "expresses," or translates into protein, only a specific small portion of that information. How does the cell know which structural genes to express?

Along with the structural information, a DNA molecule carries a series of regulatory commands, also written out as a sequence of bases. The simplest of these commands say in effect "Start here" or "Stop here" both for the transcription and for the translation steps. More complicated commands say when and in which type of cell a specific gene should be used. The genetic code is the same in all cell nuclei, a given structural sequence specifying the same protein in every organism, but the special commands are not the same in bacteria and in animal cells. One of the most surprising differences was discovered only in the past two years. The information for a bacterial protein is carried on a contiguous stretch of DNA, but in more complicated organisms, such as pigs and people, the structural information is broken up into segments, which are separated along the gene by long stretches of other DNA called intervening DNA or "introns." In such a cell a long region (often 10 times more than might be needed) is transcribed into RNA. The cell then processes this long RNA molecule, removing the sequence of bases that does not code for the protein and splicing together the rest to make a messenger-RNA molecule that carries essentially just the "start," the structural sequence and the "stop" needed for translation.

To persuade a bacterium to make a nonbacterial protein one must put into bacteria a DNA molecule that has a sequence of bases specifying the protein's amino acids as well as the bacterial commands for transcription and translation. Moreover, the inserted DNA must be treated by the bacterium as its own so that it will be duplicated as the bacterium divides. The problem thus breaks down into three parts: to find the right structural sequence (insulin's, for example), to place it in bacteria in such a way that it will be maintained as the bacteria grow and then to manipulate the surrounding information, modifying the regulatory commands so that the structural sequence is expressed as protein. Once the protein is made, still further changes in its gene or modifications of the bacterium may be needed to obtain the protein in large enough amounts to be useful.

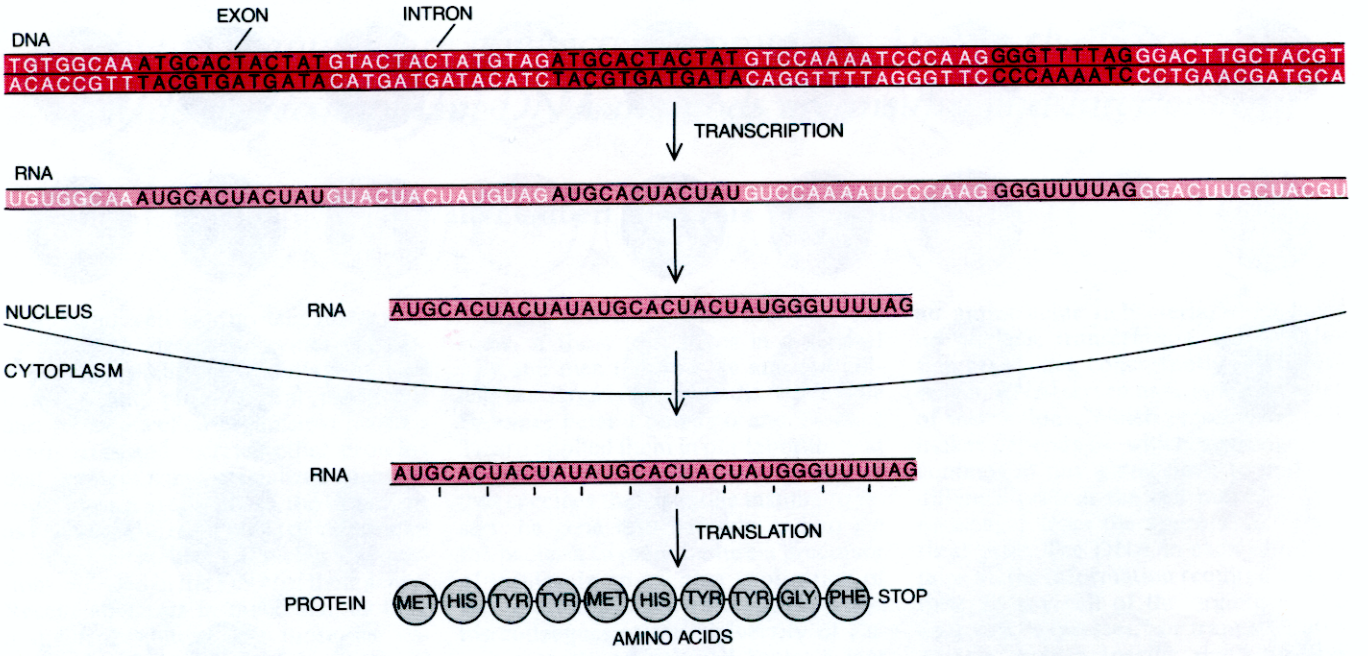
The constellation of recombinant-



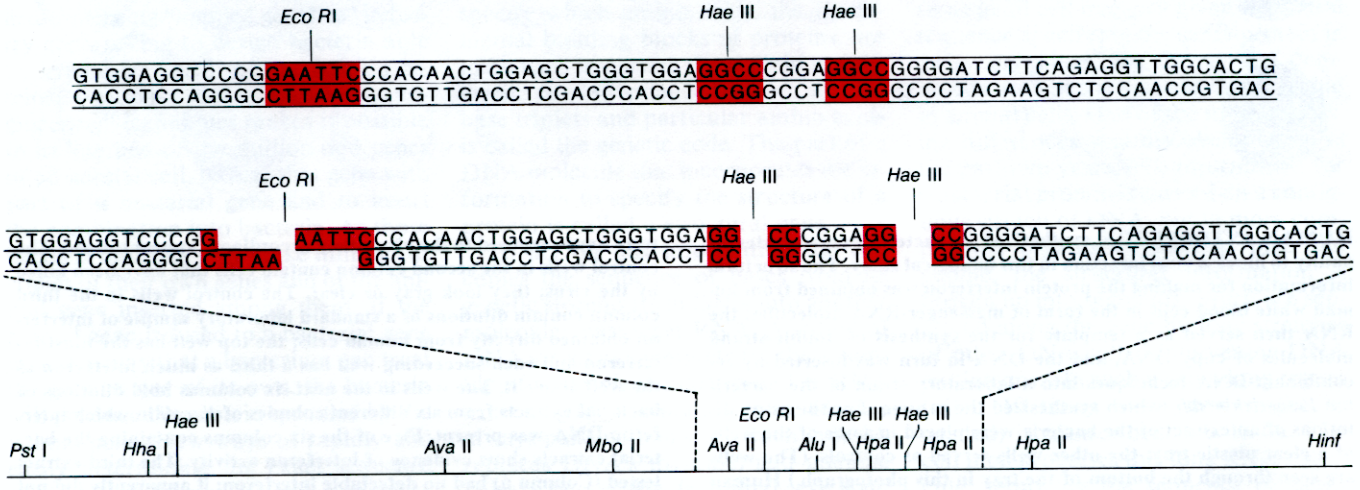
HUMAN INTERFERON synthesized in bacteria demonstrates its ability to block a viral infection in this biological assay. The structural information for making the protein interferon was obtained from human white blood cells in the form of messenger-RNA molecules; the RNA then served as a template for the synthesis of double-strand molecules of copy DNA, and the DNA in turn was inserted by recombinant-DNA techniques into a laboratory strain of the bacterium *Escherichia coli,* which synthesized the protein. For the assay dilutions of an extract of the bacteria were placed in some of the wells of a clear plastic tray; the other wells served as controls. (The wells are seen through the bottom of the tray in this photograph.) Human cells were added to the wells and were grown to form a layer of cells covering the bottom of each well. A virus preparation was then added to the cells. Twenty-four hours later the cell layer was stained. Where interferon in the extracts protected the cells against the virus the cells survived and were stained. Where there was no interferon the virus killed the cells and the dead cells did not pick up the stain. The control wells in the first column at the left contain a layer of cells that were never exposed to the virus; they accordingly appear stained. The control wells in the second column contain cells that have been killed by the virus; they look gray or clear. The control wells in the third column contain dilutions of a standard laboratory sample of interferon obtained directly from human cells; the top well has the most interferon and each succeeding well has a third as much interferon as the well above it. The wells in the next six columns hold dilutions of bacterial extracts from six different colonies of *E. coli* in which interferon DNA was present. Five of the six columns containing the bacterial extracts show evidence of interferon activity. The third extract tested (Column 6) had no detectable interferon; it apparently did not have a complete interferon gene. The synthesis of human interferon by the recombinant-DNA method was achieved by Charles Weissmann and his colleagues at the University of Zurich in collaboration with Kari Cantell of the Finnish Red Cross. The work was supported by Biogen, SA. Interferon is synthesized by many animal cells, but it is species-specific: only human interferon works for human beings, and it has been too scarce even for satisfactory experimentation.

DNA techniques for placing and maintaining a new gene in bacteria is called cloning, which in this sense means the isolation of a specific new DNA sequence in a single organism that proliferates to form a population of identical descendants: a clone. There are two convenient ways of doing this. In one method a small circular piece of DNA called a plasmid is the vehicle for introducing the new DNA into the bacterium. Plasmids carry only a few genes of their own and are maintained in several copies inside the bacterium by the bacterium's own gene functions; they remain separate from the main set of bacterial genes carried on a circle of DNA about 1,000 times larger. Alternatively the vehicle could be a virus that grows in bacteria. Such viruses normally have some 10 to 50 genes of their own (a bacterium has several thousand genes) and can often carry other new DNA segments in place of some of their own. All the techniques we shall describe apply to both plasmids and viruses.

A molecule of DNA resembles a very



PROTEINS ARE MADE in a living cell according to instructions encoded in the cell's genes, which consist of specific sequences of chemical groups (bases) strung out along a double-strand molecule of DNA in the cell's nucleus. The genetic code is "written" in the four letters *A, T, G* and *C,* which stand respectively for the four bases adenine, thymine, guanine and cytosine. The code is "read" in the three-letter sets called codons, which specify the amino acids linked together in the protein chain. The order of the bases can also convey regulatory commands. In multicelled organisms the structural sequence, or gene, encoding a particular protein is usually broken into fragments separated by long stretches of other DNA; in this diagram the gene fragments, called exons, are represented by the black letters and the intervening sequences, known as introns, by the white letters. The genetic information is translated into protein indirectly. First the entire sequence of bases is transcribed inside the nucleus from the DNA to a single-strand molecule of RNA. According to the base-pairing rules governing transcription, adenine always pairs with uracil (*U*) and guanine always pairs with cytosine. Next the RNA copies of the introns are excised from the message and the remaining RNA copies of the exons are joined together end to end. The reassembled strand of messenger RNA then moves from the nucleus to the cytoplasm, where the actual protein-manufacturing process takes place.



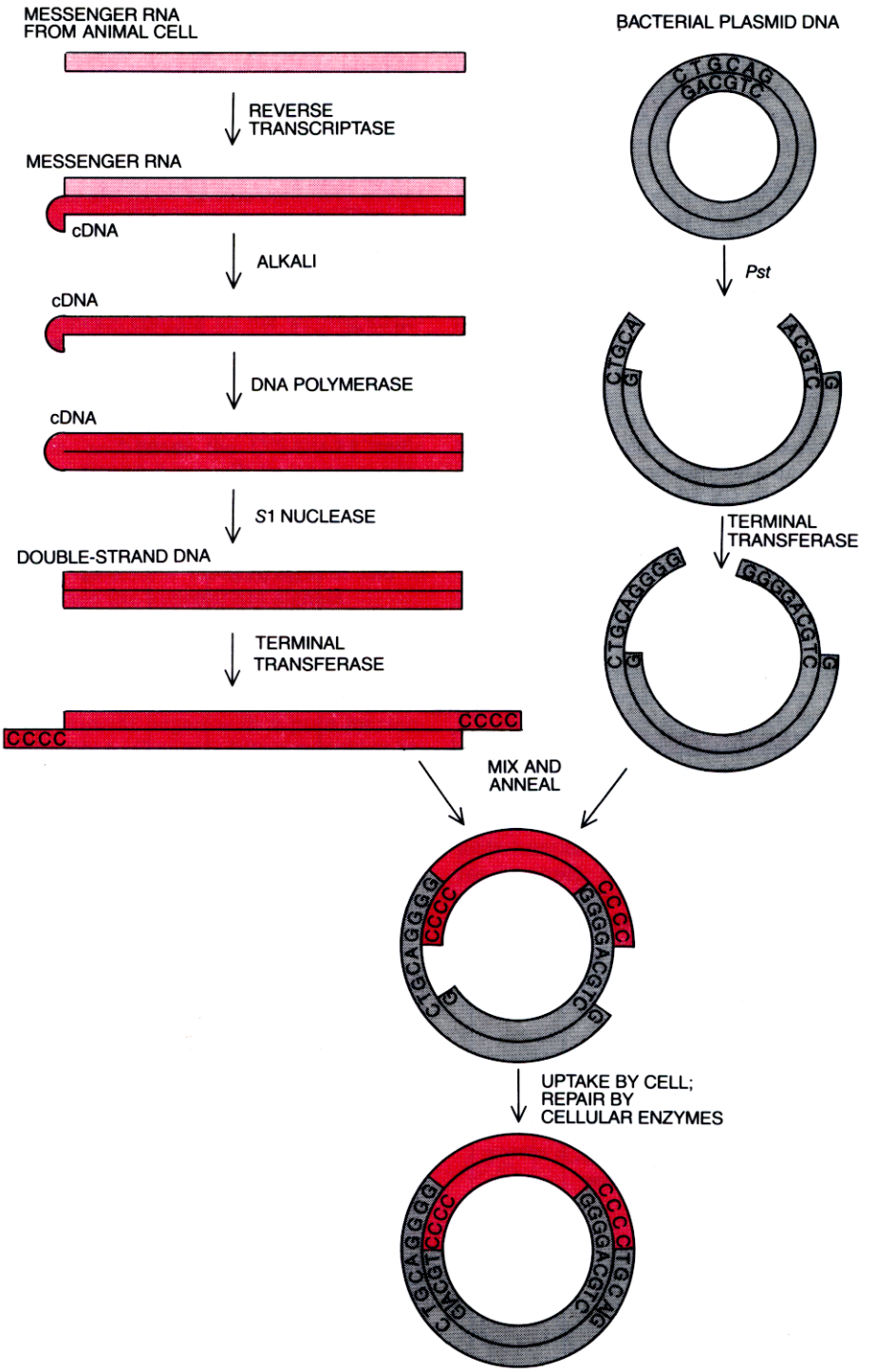
DNA CAN BE CUT into comparatively short lengths with the aid of restriction endonucleases, special enzymes that recognize specific base sequences at which they cause the molecule to come apart. For example, *Eco RI*, the first such enzyme discovered, recognizes a certain six-base sequence and cuts the molecule wherever this sequence appears, whereas *Hae* III, another restriction enzyme, operates at a certain four-base sequence. Since the probability of finding a particular four-base sequence is greater than that of finding a particular six-base sequence, one would expect *Hae* III to cut DNA more often than *Eco RI*. Accordingly one *Eco RI* site and two *Hae* III sites are represented in the DNA segment at the top, which corresponds to part of the gene coding for insulin in rat cells. The same DNA contains recognition sites for a number of other restriction enzymes, as is shown in the line diagram of a larger gene fragment at the bottom.

long, twisted thread. A bacterium has one millimeter of DNA in a continuous string of some three million bases folded back and forth several thousand times into a space less than a micron (a thousandth of a millimeter) across. In human cells the DNA is packed into 46 chromosomes, each one containing about four centimeters in a single piece, the total amount corresponding to about three billion bases. How can one find and work with a single gene only a few thousand bases long? Fortunately nature has devised certain enzymes (proteins that carry out chemical reactions) that solve part of the problem. These special enzymes, called restriction endonucleases, have the ability to scan the long thread of DNA and to recognize particular short sequences as landmarks at which to cut the molecule apart. Some 40 or 50 of these enzymes are known, each of which recognizes different landmarks; each restriction enzyme therefore breaks up any given DNA reproducibly into a characteristic set of short pieces, from a few hundred to a few thousand bases long, which one can isolate by length.

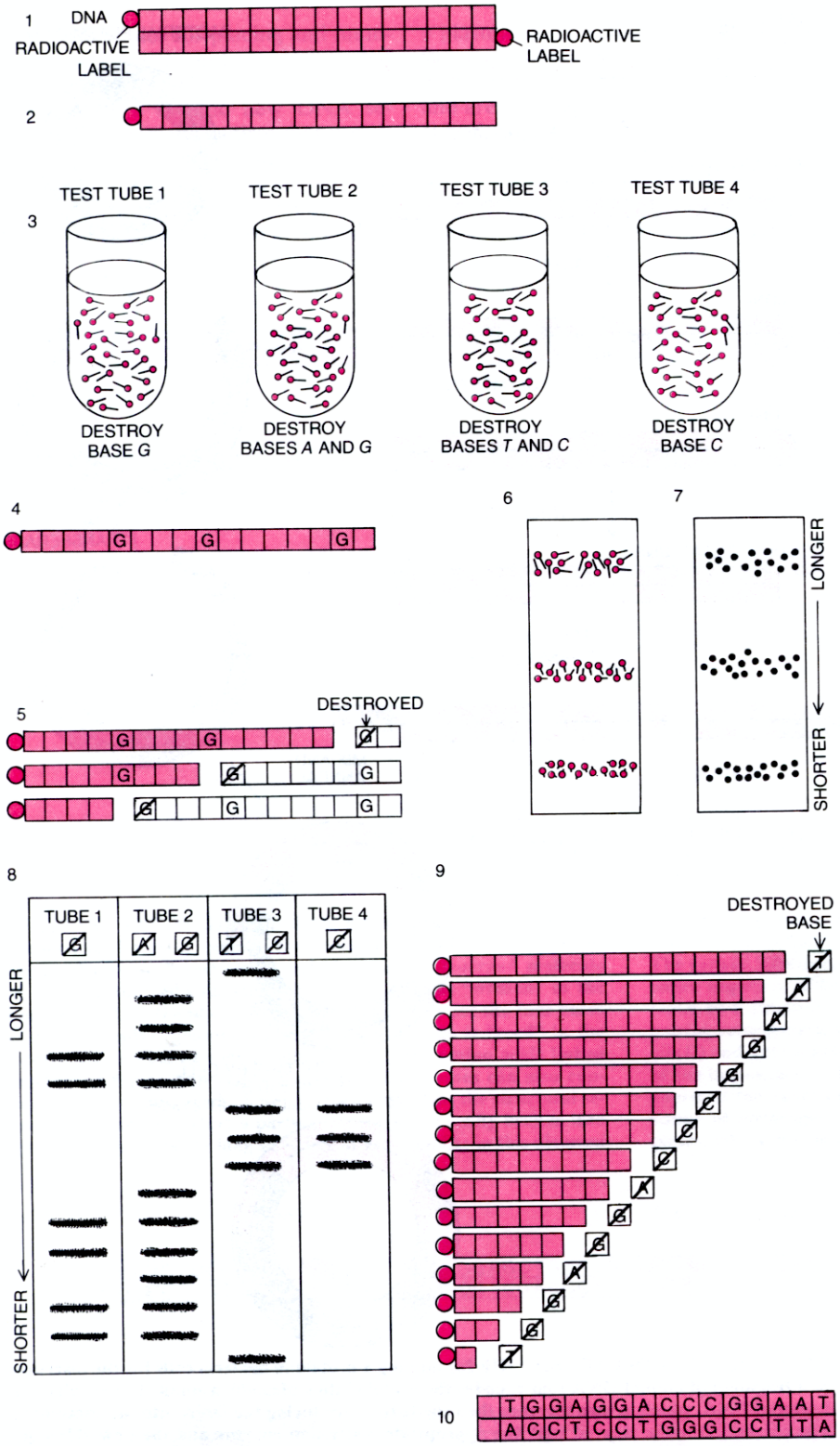
One can clone such DNA pieces in bacteria. As a first step one purifies the circle of plasmid DNA. The sequences of the plasmids are such that one of the restriction enzymes will recognize a unique site on the plasmid and cut the circle open there. One can insert a chosen DNA fragment into the opening by using a variety of enzymatic techniques that connect its ends to those of the circle. Ordinarily this recombinant-DNA molecule could not pass through the bacterial cell wall. A dilute solution of calcium chloride renders the bacteria permeable, however; in a mixture of treated cells and DNA a few bacteria will take up the hybrid plasmid. These cells can be found among all those that did not take up the DNA if a gene on the plasmid provides a property the bacterium must have to survive, such as antibiotic resistance. Then any bacterium carrying the plasmid will be resistant to the antibiotic, whereas all the others will be killed by it. When one spreads the mixture of bacteria out on an agar plate containing nutrients and the antibiotic, each single bacterium with a plasmid will grow into a separate colony of about 100 million cells. A single colony can be chosen and grown further to yield billions of cells, each of which contains identical copies of the new DNA sequence in a recombinant plasmid.

The Sequencing of DNA

The procedures we have outlined so far are followed in "shotgun" cloning experiments. One breaks up the DNA of an animal cell into millions of pieces and inserts each piece into a different bacterium. In this way a number of collections of all the fragments of human,



RECOMBINANT-DNA TECHNIQUE for making a protein in bacteria calls for the insertion of a fragment of animal DNA that encodes the protein into a plasmid, a small circular piece of bacterial DNA, which in turn serves as the vehicle for introducing the DNA into the bacterium. The plasmid DNA is cleaved with the appropriate restriction enzyme and the new DNA sequence is inserted into the opening by means of a variety of enzymatic manipulations that connect the new DNA's ends to those of the broken plasmid circle. In the procedure illustrated here, for example, a special enzyme, reverse transcriptase, is first used to copy the genetic information from a single-strand molecule of messenger RNA into a single strand of copy DNA. The RNA template is then destroyed, and a second strand of DNA is made with another enzyme, DNA polymerase. Still another enzyme, *S1* nuclease, serves to break the covalent linkage between the two DNA strands. In the next step the double-strand DNA is joined to the plasmid by first using the enzyme terminal transferase to extend the ends of the DNA with a short sequence of identical bases (in this case four cytosines) and then annealing the DNA to the plasmid DNA, to which a complementary sequence of bases (four guanines) has been added. Bacterial enzymes eventually fill the gaps in the regenerated circular DNA molecule and seal the connection between the inserted DNA and the plasmid DNA. The particular plasmid used by the authors to make rat proinsulin in bacteria, designated *pBR*322, incorporates two genes that confer resistance to two antibiotics: penicillin and tetracycline. The plasmid is cleaved by the restriction enzyme *Pst* at a recognition site that lies in the midst of the gene encoding penicillinase (the enzyme that breaks down penicillin). The added DNA destroys this enzymatic activity, but the tetracycline resistance remains and is used to identify bacteria containing the plasmid.



SEQUENCING OF DNA, in the method devised by one of the authors (Gilbert) and Allan M. Maxam, begins with the attachment of a radioactive label to one end of each strand of double-strand DNA (*1*). The strands of trillions of molecules are separated (*2*) and a preparation of one of the two kinds of strands is divided among four test tubes (*3*). Each tube contains a chemical agent that selectively destroys one or two of the four bases *A*, *T*, *G* and *C*, thereby cleaving the strand at the site of those bases; the reaction is controlled so that only some of the strands are cleaved at each of the sites where a given base appears, generating a set of fragments of different sizes. A strand containing three *G*'s (*4*), for example, would produce a mixture of three radioactively labeled molecules (*5*). The reactions break DNA at the *G*'s alone, at the *G*'s and the *A*'s, at the *T*'s and the *C*'s, and at the *C*'s alone. The molecules are separated according to size by electrophoresis on a gel; the shorter the molecule, the farther it migrates down the gel (*6*). The radioactive label produces an image of each group of molecules on an X-ray film (*7*). When four films are placed side by side (*8*), the ladderlike array of bands represents all the successively shorter fragments of the original strand of DNA (*9*). Knowing what base or pair of bases was destroyed to produce each of the fragments, one can start at the bottom and read off a left-to-right sequence of bases (*10*), which in turn yields the sequence of the second strand.

mouse, rat and fly DNA have been made. One can determine the structure of any one of these cloned DNA's by breaking up the hybrid plasmid with a restriction enzyme, separating the resulting DNA fragments, determining the base sequence of each of the fragments and then putting the sequences together to deduce the entire structure of the cloned DNA.

There are two methods for sequencing DNA. Both exploit reference points created by restriction-enzyme cleavage of the DNA at a specific short sequence and then work out the rest of the sequence by measuring the distance of each base from that cut. They do this by creating a set of radioactively labeled molecules, each of which extends from the common point to one of the occurrences of a specific base. When these molecules are separated by size and detected by their radioactivity, the length of the smallest one shows the position of the first occurrence of that base; longer molecules correspond to later occurrences. The pattern created by the analysis of these molecules looks like a ladder. From the positions of the rungs one reads off the lengths. By comparing four such patterns one reads off a sequence.

One technique, devised by Allan M. Maxam and one of us (Gilbert), makes use of chemical reagents that detect the different chemical properties of the bases and break the DNA there. To generate the set of fragments the reactions are done for a short time, so that the molecule is broken only occasionally instead of everywhere the base occurs; different molecules will be broken at different places. Four different sets of reagents are used to generate the four patterns. The radioactive label is attached directly to the end of the particular restriction fragment one wants to sequence, so that only the molecules stretching from the labeled end to the break are detected by their radioactivity.

The other sequencing method, devised by Frederick Sanger of the British Medical Research Council Laboratory of Molecular Biology in Cambridge, makes a DNA copy with an enzyme and stops the sequential synthesis, and hence the elongation of the copy, by blocking the movement of the enzyme at a specific base. Here the radioactive label is incorporated into the newly synthesized molecule in four different reactions. Both methods can provide the sequence of from 200 to 300 bases in a single experiment. One of the small plasmids involved in our cloning experiments was sequenced in a year by Gregory Sutcliffe, who worked out the order of the 4,357 bases on one strand and checked them by working out the complementary strand.

Any DNA region carried on a plasmid can be isolated and sequenced. The difficulty is not in determining the sequence but in obtaining the specific

DNA fragments needed. The recombinant-DNA technique serves almost as a microscope to isolate and to magnify, by making many copies, a DNA region, but one does not want to look through a million bacteria to find a specific gene. The fundamental problem, which has no general solution, is to place only the desired DNA sequence—the desired structural gene—in a bacterium.

Getting the Right Gene

One straightforward approach is suitable for very small proteins. The amino acid sequence and the genetic code will predict a sequence of bases that can specify those amino acids. One can then chemically synthesize a corresponding DNA molecule. Exactly this was done by Keiichi Itakura and his co-workers at the City of Hope National Medical Center in Duarte, Calif., who constructed a DNA sequence 42 bases long that dictates the structure of somatostatin, a small hormone consisting of 14 amino acids. The longer the stretch of DNA, however, the harder it is to make; the synthesis of a stretch of DNA 100 bases long is extremely difficult. Many small hormones consist of from 50 to 100 amino acids, and enzymes and other proteins range from 200 to several thousand amino acids in length. Furthermore, one does not know the amino acid sequence of many interesting proteins. (Indeed, the amino acid sequence of some of these proteins has become available only through the sequencing of cloned DNA.)

The desired structural gene is present, of course, somewhere on the DNA of the animal cell. The problem is to find it, but even if that were possible, the structural information would be broken up (as we mentioned above) by long stretches of other DNA. The information does exist in a continuous form, however, on the messenger RNA. Moreover, different cells specialize in the synthesis of different proteins, so that the appropriate tissue will contain the desired messenger RNA along with other messengers for the common proteins made by all cells. Insulin, for example, is made by the beta cells of the pancreas; those cells contain insulin messenger RNA and other cells do not, even though the insulin gene is present in the DNA of every cell.

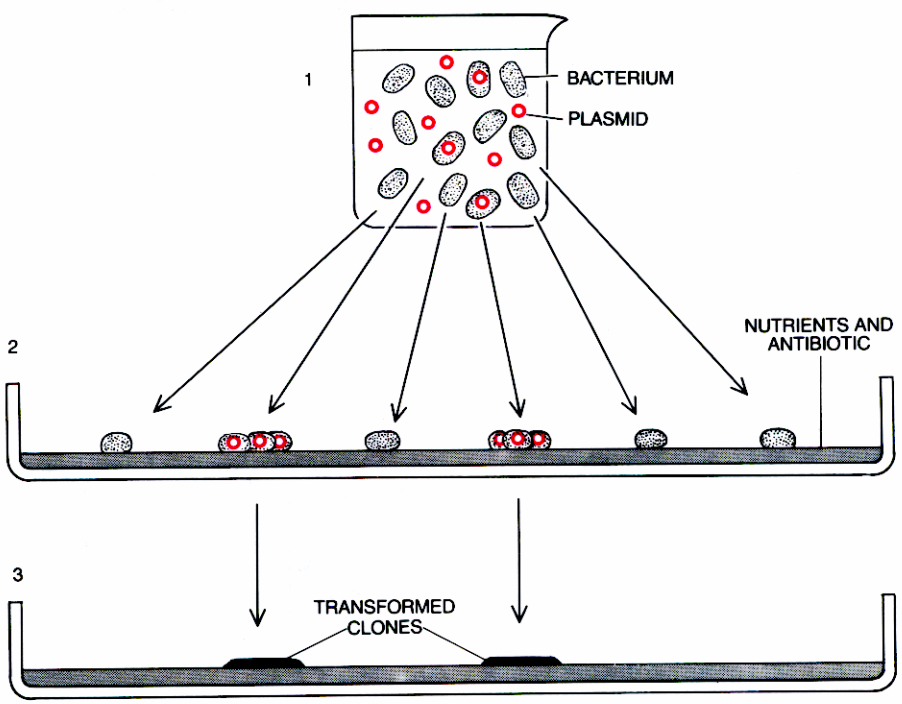
The task is then to convert the desired structural information from the cell's messenger RNA into DNA, which can be cloned. For this one takes advantage of a special enzyme, reverse transcriptase, that can copy a single strand of RNA to make a complementary strand of DNA. (The enzyme is found in certain RNA viruses that reverse the normal DNA-to-RNA transcription. Such viruses depend on RNA rather than DNA to carry their information from one cell to another and convert the RNA back into DNA with the help of reverse transcriptase after they infect a new cell.) One takes this strand of complementary DNA, called copy DNA, and makes a second strand of DNA with the more usual DNA-copying enzyme. The resulting double-strand cDNA fragments are more or less complete copies not only of the desired messenger RNA but also of all the other messenger RNA's that were present in the tissue. At best, however, only a few of the DNA fragments contain all the wanted structural information. Even in those fragments the regulatory signals that surround the structural sequences refer to translation in the animal cell, not in bacteria, and (since the DNA was made from RNA) there will be no transcriptional commands. Although the cDNA can be cloned, two problems remain: to detect any clones containing the sought-after structural DNA fragment and to provide the appropriate signals.

Finding the Right Clone

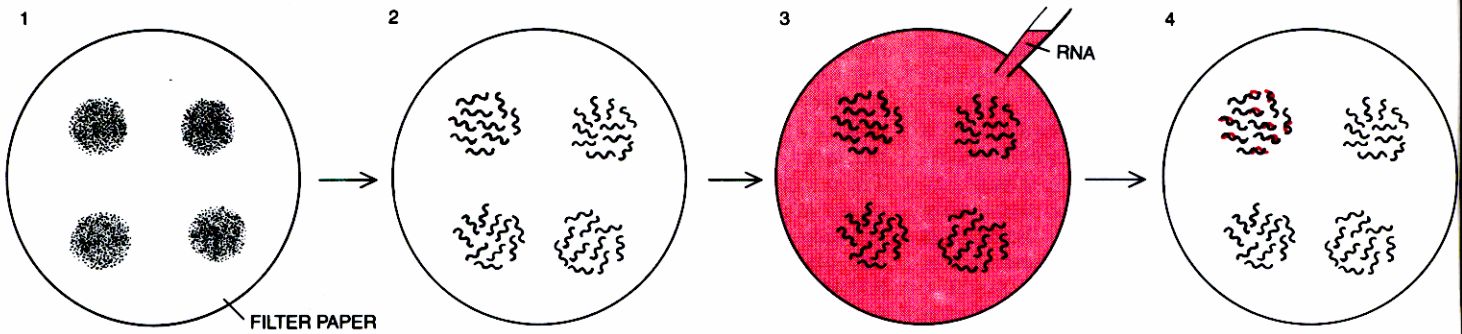
It is simple to find the right clone if the experiment began with a pure messenger RNA. One can detect matching sequences by the process called hybridization. The two strands of a DNA molecule can be separated by heating, which breaks the weak bonds that hold the two strands together without breaking the strong chemical bonds between bases along the chain. When a mixture of such strands is cooled, those sequences that match will find each other. The first step of this process is called denaturation, the second step reannealing. The same process serves to identify sequence matches between RNA and DNA.

One grows bacterial colonies on a disk of cellulose nitrate paper, breaks open the bacterial cells where they lie and fixes the released DNA to the paper. When the DNA is denatured and reannealed to radioactive RNA, only the remains of those colonies that contained a plasmid whose sequence matches the messenger become radioactive. Since one keeps a replica (a living duplicate set of the colonies), one can obtain bacteria containing the desired DNA. One grows these bacteria to provide material to identify, in further hybridization tests, other clones that contain the same sequence in different surroundings and may turn out to be more effective in producing the wanted protein.

If one cannot purify the messenger RNA because the specific messenger is a small fraction of all the messengers in a cell, there are other ways to search for the DNA sequence. One useful property is the detailed shape of the corresponding protein molecule. Those shapes that are most different and distinctive can be recognized by the protein molecules called antibodies. Animals make antibodies as part of their protective response to foreign substances. If one injects human insulin into a guinea pig, for example, the guinea pig will make antibodies that bind to human insulin. These antibodies will not bind to guinea pig insulin because they "see" only the shapes that make the human protein different. A purified antibody, then, can



RECOMBINANT PLASMIDS (*color*) bearing the inserted animal-protein genes and genes for resistance to tetracycline are mixed with bacteria (*1*). Some cells take up the plasmid. The mixture of cells is spread on a culture medium containing the antibiotic (*2*), which kills all the cells that do not have the plasmid. The cells that have taken up the plasmid are antibiotic-resistant; they live, and each of them gives rise to a clone, a colony of genetically identical cells (*3*).



CLONE CONTAINING DESIRED DNA can be found among all the successfully transformed clones (*1*) by means of RNA-DNA hybridization if one has a pure messenger-RNA probe for the desired sequence. The cells are broken open and their DNA is denatured and fixed to filter paper (*2*). The RNA probe (RNA molecules labeled with a radioactive isotope) is added (*3*). The RNA (*color*) will anneal to any DNA whose sequence it matches, forming RNA-DNA hybrids (*4*); the remainder of the RNA is washed away. The presence of the hy-

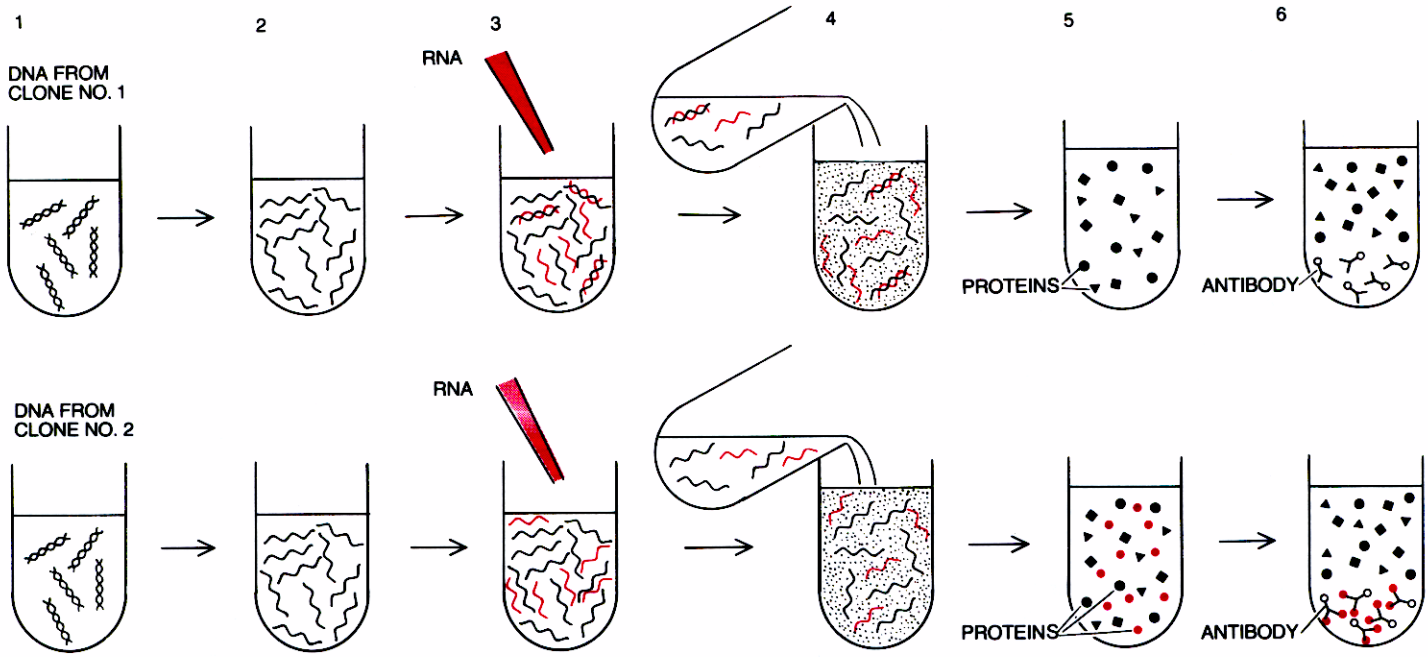
serve as a reagent to detect a particular protein. (This is the way vaccines work. If an animal is injected with an inactivated virus, it is stimulated to make antibodies against the viral proteins. Thereafter the antibodies will protect the animal against infection by that virus by binding to the virus particle and signaling other cells to remove the invader. Without the earlier stimulation the antibody response to the invading virus is too slow to block the infection.)

Even without purifying a specific messenger RNA one can make the RNA molecules function in the test tube by adding the machinery needed to translate the messengers (obtained from the cytoplasm of broken cells) along with radioactive amino acids. Among the small amounts of radioactive proteins that are synthesized one can recognize the protein of interest with antibodies. This provides a means of detecting the presence of a specific messenger. If one takes a recombinant plasmid and hybridizes it to the mixture of RNA's, only the RNA that matches a sequence in the plasmid will anneal to it and therefore no longer function in translation; the plasmid of interest is detected by its ability to block the synthesis of the desired protein. This identification can be verified because the RNA bound to the DNA can be separated from all the other RNA's and then released from the DNA, whereupon it will function to direct the synthesis of the protein.

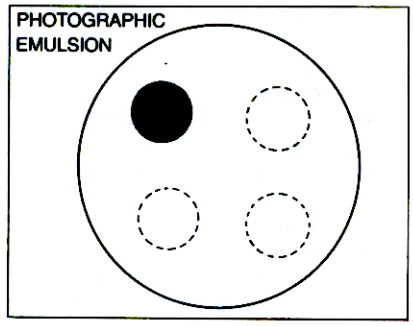
Regulatory Signals

With these techniques one can clone and identify DNA fragments carrying the information that dictates the structure of a protein. Will the information work in bacteria?

One must provide regulatory signals the bacterium can use. One of them is the signal to start the synthesis of a messenger RNA; in bacteria it is a region of DNA immediately in front of the segment of DNA that will be transcribed into RNA. The second important signal functions as part of the messenger RNA, telling the bacterial translation machine to "Start here." All bacterial genes have these two kinds of start signals (some of which work better than others). They also have two stop signals, one for translation and one for transcription. A simple way to make the new protein sequence is to cut a bacterial gene open in its middle with a restriction enzyme and to insert the new DNA there. This results in a hybrid protein that starts out as some bacterial protein and then continues as the string of amino acids one wants. That is how the chemically synthesized gene for somatostatin was made to work in bacteria. The DNA



HYBRID-ARRESTED TRANSLATION, a technique developed by Bryan Roberts of the Harvard Medical School, identifies a clone (*top*) containing the desired DNA even in the absence of a purified RNA probe. DNA from clones being tested (*1*) is denatured (*2*). Unpurified RNA (the same RNA used to make the inserted DNA) is added (*3*); it anneals to any matching DNA. Placed in a "translation system" containing radioactively labeled amino acids (*4*), the unhybridized RNA directs the synthesis of radioactive proteins, but the hybridized RNA cannot be translated; the specific protein (*color*) encoded by the desired DNA is not synthesized in the presence of the clone containing that DNA (*5*). The presence or absence of that protein is determined by an antibody test. Antibody to the protein, fixed to plastic beads,

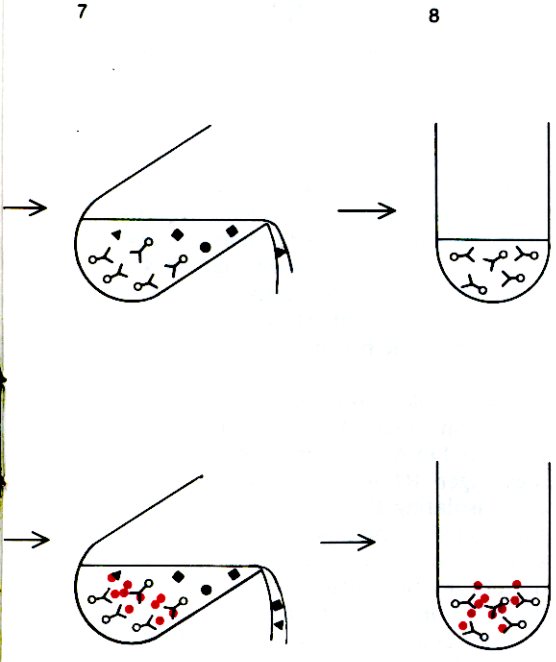


brids is revealed by autoradiography: a photographic emulsion is placed on the filter paper and after exposure the clone containing the desired DNA is identified as a dark spot (*5*).

for those 14 amino acids, followed by a stop signal, was inserted near the end of a 1,000-amino-acid protein. After the bacterium made the hybrid protein the somatostatin part was cleaved off chemically and purified.

Not only can the bacterial gene serve to provide the regulatory signals but also it may endow the hybrid protein with further useful properties. For example, a few bacterial proteins are secreted through the membrane that surrounds the cell. If one inserts the animal DNA into the gene for such a protein, the bacterial part of the hybrid protein will serve as a carrier to move the new protein through the membrane so that it is more easily observed and purified.

We exploited all the techniques described above to obtain a copy of the insulin gene and to insert it into bacteria to make proinsulin. Insulin is a small hormone made up of two short chains, one chain 20 amino acids long and the other 30 amino acids long. These two chains are initially part of a longer chain of 109 amino acids, called preproinsulin. As preproinsulin is synthesized in the beta cells of the pancreas, the first 23 amino acids of the chain serve as a signal to direct the passage of the molecule through a cell membrane. As this happens those amino acids are cleaved off, leaving a chain of 86 amino acids: proinsulin. The proinsulin chain folds up to bring the first and last segments of the chain together, and the central portion is cut out by enzymes to leave insulin. The role of the central portion is to align the two chains comprising insulin correctly. If the two chains are taken apart later, they do not reassemble easily or efficiently. (In spite of these difficulties Itakura and his co-workers synthesized two DNA fragments corresponding to the two chains of human insulin and attached them separately, like somatostatin, to the same large bacterial gene in order to synthesize two separate hybrid proteins in two different bacteria. Then they cut off the two short pieces, purified them and put them together to form insulin.)



is added and binds the protein, precipitating the protein out of the solution (*6*), which is poured off (*7*). Measurement of the precipitates' radioactivity (*8*) shows that one clone (*top*) contains the desired DNA, because it blocked the synthesis of the specific protein.

The Proinsulin Experiment

In our experiments we started with a tumor of the insulin-producing beta cells of the rat. (We worked with rat insulin because at the time we began our experiments the guidelines established by the National Institutes of Health for recombinant-DNA investigations would not allow us to insert the human insulin gene into bacteria; that prohibition has since been removed.)

We made DNA copies of the beta-cell messenger RNA and put them into a plasmid, in the middle of a gene for a bacterial protein, penicillinase, that would be secreted through the membrane of the bacterial cell. We looked among the bacterial colonies by hybridization, we proved that we had the right hybrid plasmid by blocking the synthesis of insulin in a test tube as we described above and we sequenced the DNA to see exactly what part of the insulin gene we had. Once we had found one hybrid plasmid, we used it to find 48 more by repeating the hybridization test. These 48 clones represented 2 percent of all the clones we had made.

Would any of those clones actually synthesize insulin? We looked among the clones containing insulin DNA for any that were synthesizing a hybrid protein part of which was proinsulin. For this we relied on a sensitive radioactive-antibody test. We coated plastic disks with antibody directed against either insulin or penicillinase and exposed them to the contents of cells from each clone. Any insulin (or penicillinase) present in the cells binds to the antibody and is thereby fixed to the plastic disks. Then we applied radioactively labeled anti-insulin antibody to detect the presence of proteins with insulin shapes. One clone gave positive responses, both on disks coated with anti-insulin and on those coated with antipenicillinase, to radioactive antibody to insulin, thereby demonstrating the presence of a penicillinase-insulin hybrid protein.

To see if the bacteria were secreting the hybrid protein we grew the clone in liquid culture and tried to extract the protein by a method that does not burst the bacterial cell membrane. The test showed the fused protein to be present outside the membrane: it was secreted, as we had hoped it would be.

Sequencing the DNA showed that the DNA fragment and the details of the fusion were such that the structural information in the clone was only for proinsulin and did not contain the "pre" region. In order to make insulin we removed most of the bacterial protein and the middle segment of the proinsulin with the digestive enzyme trypsin. Would the insulin made from the bacteria be an active hormone? Stephen P. Naber and William L. Chick of the Elliot P. Joslin Research Laboratory in Boston tested the molecule by showing that it affected the metabolism of sugar by fat cells, as it should.

Improving the Yield

The amount of proinsulin made by the original clone was very small; we are currently engaged in various manipulations to improve the yield. Regulatory signals must be not only efficient but also optimally placed. One need not be satisfied with the signals that happen to surround preexisting bacterial genes. With restriction enzymes one can clip out small DNA fragments that carry only the regulatory signals and tie them together with a DNA-linking enzyme to make new combinations. One can trim back the ends of these fragments by nibbling off bases with still other enzymes before reconnecting them. This will alter the spacings between the signals and the structural sequence. Although each of these manipulations generates only a small number of correct molecules, by cloning after each step one can make large amounts of the DNA and work out its sequence, and then continue the tinkering.

Moreover, one can synthesize short desired DNA sequences and tie them to other fragments. For example, David V. Goeddel and his co-workers at Genentech, Inc., took a piece of DNA containing the structural information for human growth hormone (168 amino acids), connected it to a synthetic piece of DNA containing part of the translational start signal and attached that combination in turn to a fragment containing the rest of the regulatory signals. When this DNA construction was cloned, the

bacteria made a protein of the shape (as recognized by antibodies) and size of growth hormone (although not yet with demonstrated hormone activity).

Although we do not yet know the optimal combinations of the DNA elements for making insulin in bacteria, finding them is only a matter of time. There are other problems to be considered. Often the new animal proteins are broken down in the bacterial cell because their structure is such that enzymes normally present in the bacteria can digest them. Ways have to be found to stabilize the proteins either by removing these enzymes, by embedding the new protein in a hybrid protein to protect it or by secreting it from the cell. Messenger-RNA molecules themselves are often unstable within the cell; modifications in their structure and in the cell itself can make them more effective and lead to increased protein synthesis. And if the number of copies of the plasmid carrying the gene in each cell can be increased, more of the product will be made.

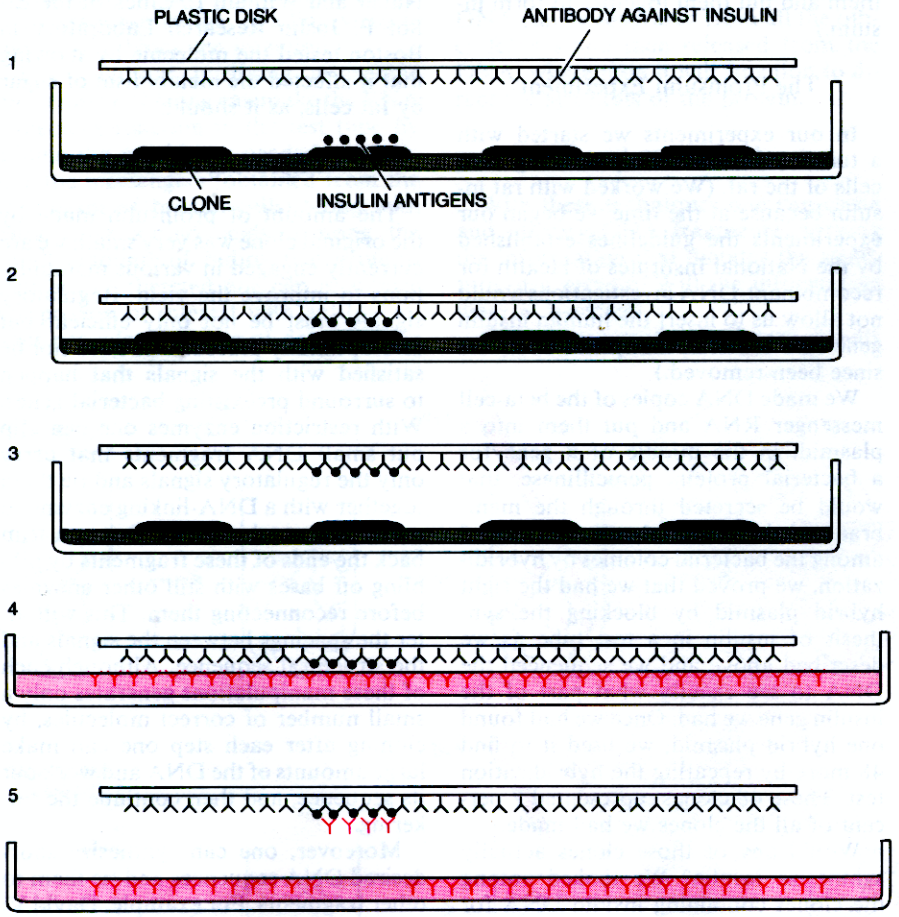
While we work to improve the yield of rat proinsulin and to purify it we expect to apply the same methods to the bacterial synthesis of human insulin. Investigators in other laboratories are also working on the problem, and one can hope that eventually the manufacture of human insulin by bacteria will be cheaper than the purification of insulin from pigs and cattle, the present sources of the hormone. Clearly other human hormones can also be prepared by these procedures. What other therapeutic proteins might be made in bacteria? In general any human protein that cannot be obtained in useful form from animals is an excellent prospect.

Other Proteins from Bacteria

Many genetic diseases are caused by the lack of a single protein. Replacement therapy may be possible if such proteins can be made in bacteria. Vaccines against viral or parasitic infections are a further wide class of possibilities. Today in order to make a vaccine one must be able to grow the disease organism in large amounts; often this is impossible or dangerous. Furthermore, the vaccine must be rendered harmless before it is administered, which can be difficult. The new technology offers the chance to make in bacteria only the protein against which the antibody response needs to be directed. This would eliminate any need to work with the intact disease organism. For example, the hepatitis *B* virus, which causes serum hepatitis, cannot be grown outside the body. The only source of this small DNA virus is the blood of infected human beings. The DNA of the virus has now been cloned in several laboratories and its complete sequence has been worked out, revealing the structure of the viral proteins; now the proteins are being made in bacteria. A flood of new information has resulted from this work.

A particularly promising candidate is interferon, a protein cells make to block viral infections quickly. (The antibody response is much slower.) Interferon appears to be the body's first line of defense against viruses. It may also have a therapeutic effect in some cancers. Interferon has never been available in sufficiently large amounts, however, to determine how effective it might really be in protecting against disease. The ability to test the activities of human interferon will soon be a reality because the protein has now been made in bacteria. Weissmann, with his colleagues Shigekazu Nagata, Hideharu Taira, Alan Hall, Lorraine Johnsrud, Michel Streuli, Josef Ecsödi and Werner Boll, along with Kari Cantell of the Finnish Red Cross, applied many of the techniques we have described to clone and to express this protein. The problem they faced was that the messenger RNA for interferon is far rarer than the one for insulin, even in white blood cells that have been stimulated by infection with a virus to make interferon. They took messenger RNA from these white blood cells (17 liters at a time), made double-strand cDNA and cloned it by the procedures we have described.

They looked through some 20,000 clones (in batches) by hybridizing the plasmid DNA from the clones to the messenger RNA of the white blood cells, isolating the RNA that annealed and checking the RNA to see if it was able to direct the synthesis of interferon (not in the test tube but by injecting the RNA into a particularly large cell, a frog's egg). Fortunately interferon is a remarkably potent substance, and so the amount synthesized in the frog's egg could be detected by its ability to protect cells against viruses.



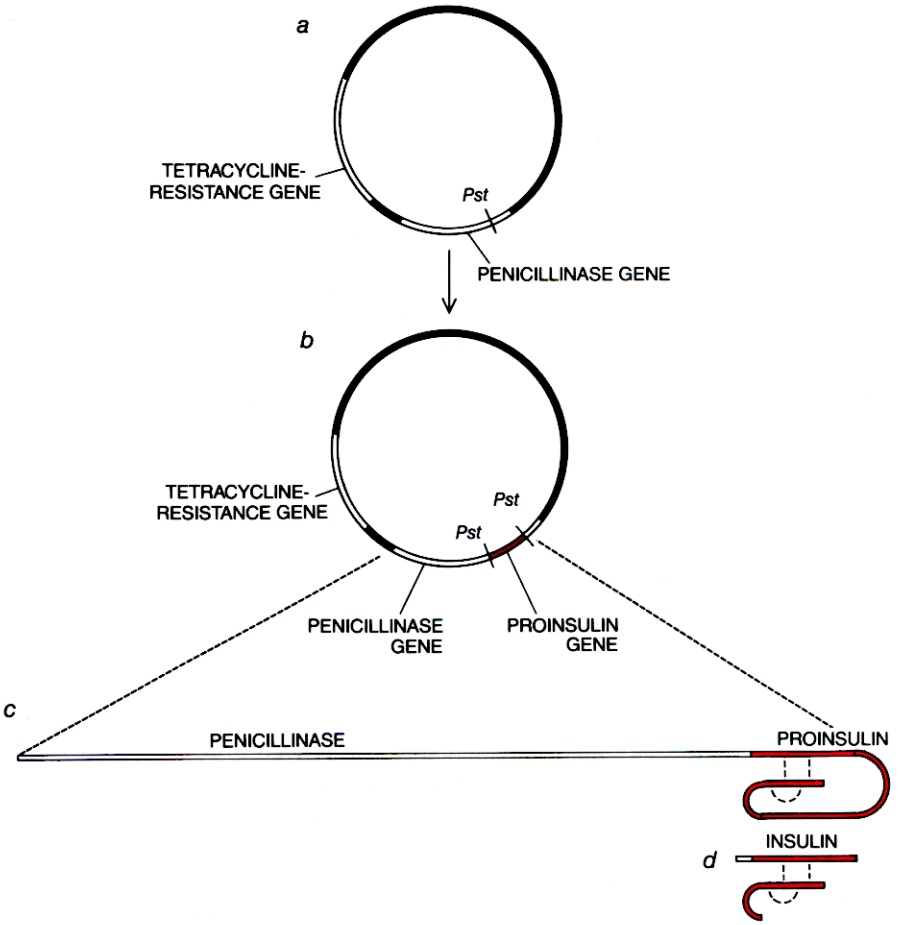
RADIOACTIVE-ANTIBODY TEST, developed by Stephanie Broome and one of the authors (Gilbert), is used to search among the bacterial clones containing insulin DNA for signs that insulin is indeed being synthesized. A plastic disk coated with an anti-insulin antibody is first exposed to the contents of cells from each clone (*1*). Any insulin present in the cells is bound to the antibody (*2*) and thereby fixed to the plastic disk (*3*). Radioactively labeled antibody (*color*) to insulin is then applied to the disk in order to detect the presence of the protein (*4, 5*). When the test is repeated with a plastic disk coated with an antipenicillinase antibody, only a hybrid protein, part penicillinase and part insulin, will bind the labeled antibody.

Once Weissmann and his colleagues had found a batch of clones that could hybridize to interferon messenger RNA they tested progressively smaller groups of those clones to find the correct one. Then, with that clone as a probe, they found other clones by means of hybridization testing. Finally they tested extracts of the bacteria carrying the interferon DNA (inserted into the penicillinase gene) directly to see if any of the bacterial clones made biologically active interferon. A number of clones did, confirming that the interferon structural DNA had been correctly identified. The sequencing of the DNA of those clones will determine the structure of interferon, which is still not known.

The amount of interferon made in the bacteria was extremely small: only one or two molecules per cell. (Bacterial proteins are usually made in from 1,000 to 100,000 copies per cell.) We are confident that the methods we have described will solve this problem and lead to the production of enough interferon for clinical tests.

The Recombinant-DNA Debate

The development of the genetic-engineering techniques described in this article was greeted, over the past decade, with both excitement and alarm. The possible benefits of the techniques were obvious, but some people felt there was reason for concern. Biologists called for an evaluation of the possible hazards of this research; the result was an unprecedented national and international effort in which the public, governments and the scientific community joined to monitor research activities. New knowledge about the properties of genes and the behavior of the bacteria used in this work (usually *Escherichia coli*) has led to a steady lessening of these concerns and to a relaxation of the guidelines that once restricted such experiments. In retrospect, with the advantage of hindsight, the concerns about hypothetical hazards seem to have been unwarranted.



RAT INSULIN WAS OBTAINED by the authors from a hybrid protein composed of part of the bacterial penicillinase molecule and a molecule of proinsulin, an insulin precursor. The map of the plasmid that served as a vehicle, *pBR322* (*a*), shows the location of the genes for the two enzymes conferring antibiotic resistance and the site of cleavage by the restriction enzyme *Pst*. The next map (*b*) shows the structure, as determined by DNA sequencing, of the recombinant plasmid in the bacterial clone that synthesized proinsulin. The proinsulin sequence (*color*) lies between two *Pst* sites that were regenerated in the insertion process. The hybrid protein synthesized by the clone (*c*) comprises most of the penicillinase and also the proinsulin molecule (*color*); broken lines represent disulfide bonds. The authors cut away most of the penicillinase and the middle segment of the proinsulin (*light color*) to make biologically active insulin (*d*).

We know of no adverse effects from this research. The great potential of the new techniques, both in promoting the growth of basic knowledge and in making possible the synthesis of products of direct benefit to society, is much closer to realization than seemed likely only a few years ago.

The Authors

WALTER GILBERT and LYDIA VILLA-KOMAROFF have collaborated on the development of techniques for the enzymatic manipulation of DNA molecules. Gilbert is American Cancer Society Professor of Molecular Biology at Harvard University. A Harvard graduate, he obtained his D.Phil. in mathematics from the University of Cambridge in 1957. He began his career as a theoretical physicist but switched to experimental work in molecular ge-netics about two decades ago. Gilbert is a founder of Biogen, SA, an applied-genetics company. Villa-Komaroff is assistant professor of microbiology at the University of Massachusetts Medical Center. She was graduated from Goucher College in 1970 and received her Ph.D. in cell biology from Harvard in 1975.

Bibliography

A BACTERIAL CLONE SYNTHESIZING PROINSULIN. Lydia Villa-Komaroff, Argiris Efstradiadis, Stephanie Broome, Peter Lomedico, Richard Tizard, Stephen P. Naber, William L. Chick and Walter Gilbert in *Proceedings of the Academy of Sciences of the United States of America,* Vol. 75, No. 8, pages 3727–3731; August, 1978.

SYNTHESIS IN *E. COLI* OF A POLYPEPTIDE WITH HUMAN LEUKOCYTE INTERFERON ACTIVITY. Shigekazu Nagata, Hideharu Taira, Alan Hall, Lorraine Johnsrud, Michel Streuli, Josef Escödi, Werner Boll, Kari Cantell and Charles Weissmann in *Nature,* in press.

THE SYNTHESIS OF EUKARYOTIC PROTEINS IN PROKARYOTIC CELLS. Lydia Villa-Komaroff in *Gene Structure and Expression,* edited by D. H. Dean, L. F. Johnson, P. C. Kimball and P. S. Perlman. Ohio State Press, in press.