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

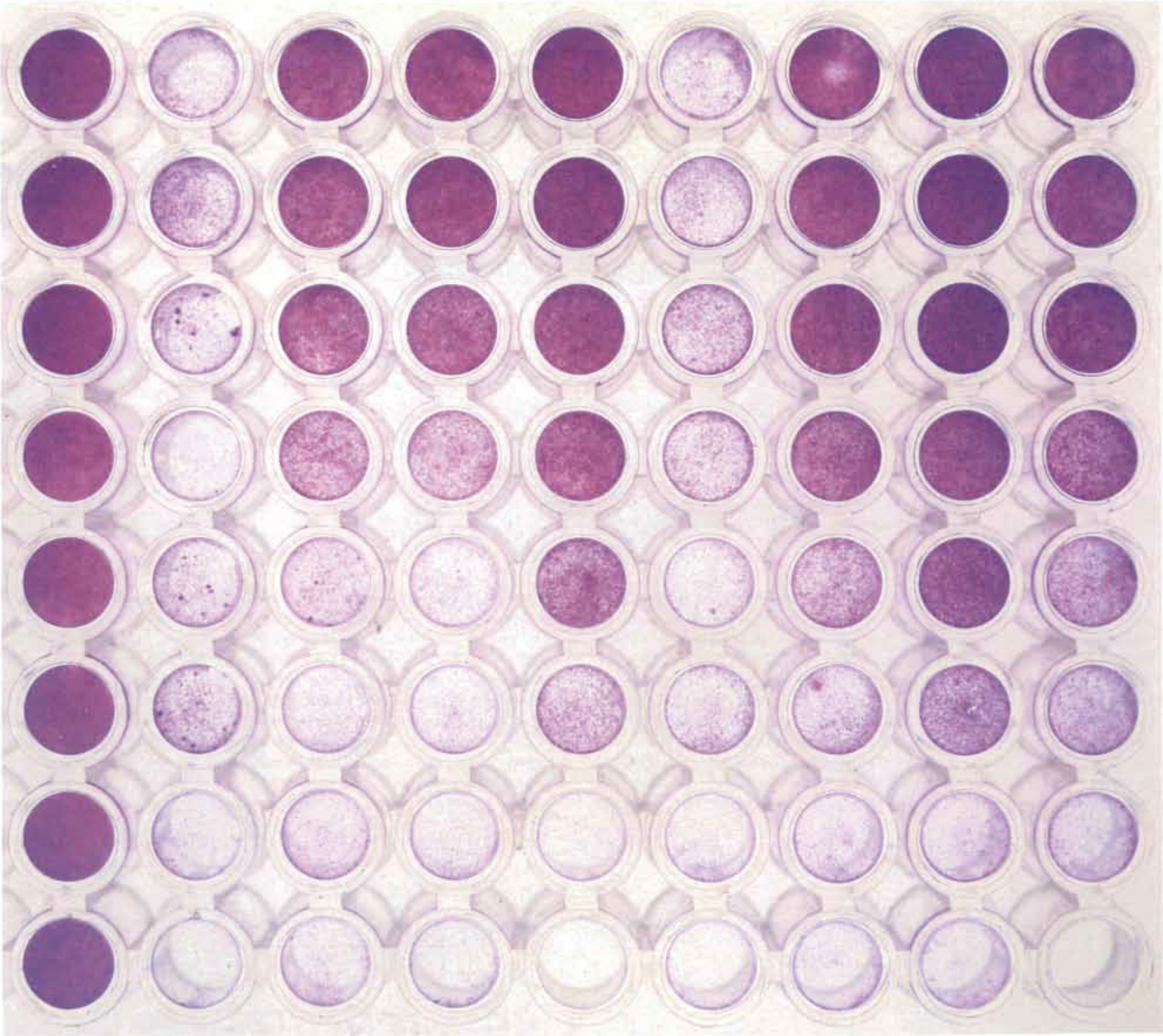
Cells make proteins by translating a set of commands arrayed along a strand of DNA. This hereditary information is held in the order of four chemical groups along the DNA: the bases adenine, thymine, guanine and cytosine. In sets of threes along DNA these bases specify which amino acids, the fundamental 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.

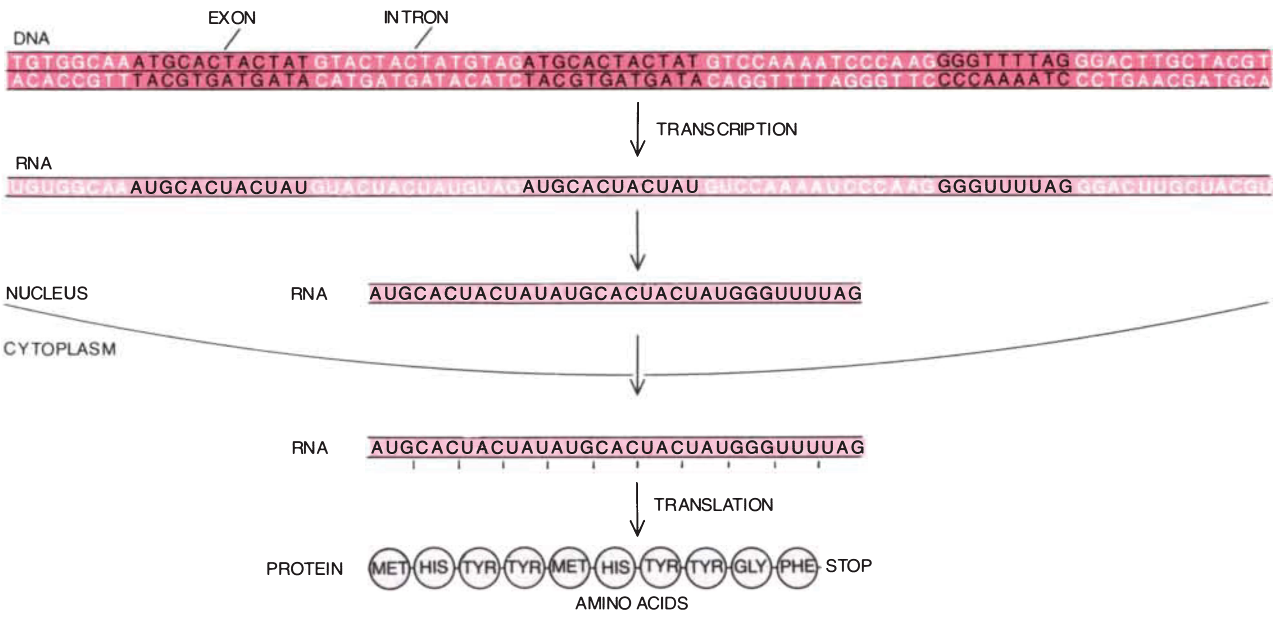
The constellation of recombinant-



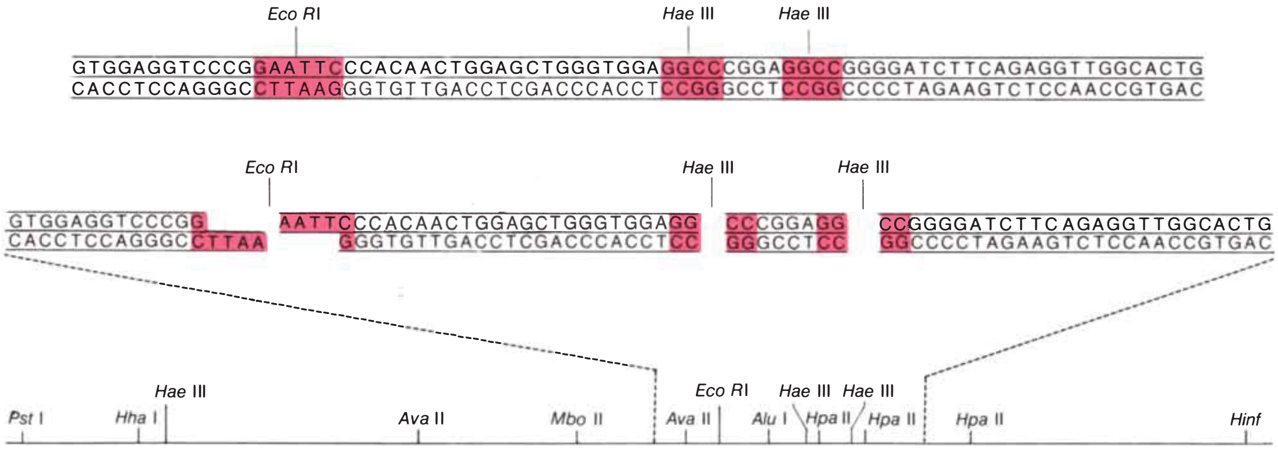
HUMAN INTERFERON synthesized in bacteria demonstrates its ability to block a viral infection in this biological assay. The structural information for making the protein interferon was obtained from human white blood cells in the form of messenger-RNA molecules; the RNA then served as a template for the 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
mann and his colleagues at the University of Zurich in collaboration with Kari Cantell of the Finnish Red Cross.
by Biogen, SA. Interferon is synthesized by many animal cells, but it is species-specific: only human interferon works for
ings, 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 R*I, 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 R*I. Accordingly one *Eco R*I 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.

What's the right wine for candlelight and Saturday night?

Almadén.

Vintage-dating makes it right. As you come to appreciate fine wines, you learn that there are subtle differences from one vintage year to the next. So at Almadén, we vintage-date our fine varietal wines to enhance your appreciation and enjoyment of these

BMW. THE CAR COMPANY THAT HASN'T BEEN LEGISLATED INTO MEDIOCRITY.

The challenge that faces the automaker of the 1980's lies not in meeting tougher pollution standards, nor in coping with ever more restrictive safety regulations—not even in accommodating the petroleum shortages that are fast becoming a regular

No indeed. The challenge lies in meeting the myriad of societal demands and still building automobiles that are worth driving.

Pessimists darkly predict that it cannot be done and point to the depressing number of uninspired automobiles on the

Yet, fortunately, these gloomy predictions do not take into account the inventive genius and the obsessive determination of the engineers at BMW in Munich, Germany to build extraordinary automobiles.

Against all odds the BMW

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.

WHAT WILL YOU BUILD? | **sales@docketalarm.com** | **1-866-77-FASTCASE**