

# Researchers Complete Analysis of Make-Up of Gene

By JANE E. BRODY

A Belgian research team has for the first time completely analyzed the chemical make-up of a gene that dictates the production of a protein, an achievement that appears to bring closer the day when genes can be made to order and turned on and off at will.

The work is regarded by experts in genetics as a significant technological feat that, among other things, provides further proof that scientists have accurately translated the genetic code and shows which parts of the code are most often used in nature.

Previously, simple artificial genes with known structure have been synthesized, and two years ago the Nobel laureate Jiar Gobind Khorana manufactured a major part of a real gene using ingenious linking techniques to piece the molecule together. Dr. Khorana's gene itself did not direct the production of a protein.

The Belgian work, hailed this month in the British journal *Nature* as one of the year's most outstanding achievements in molecular biology, was made possible by sophisticated analytic techniques worked out by the Nobel laureate Frederick Sanger and colleagues in Cambridge, England, about five years ago.

These techniques, which showed how to determine in sequence the constituents of a molecule of ribonucleic acid (RNA), have not previously been applied to so complex and laborious a task as the complete analysis of a gene.

Genetics research is widely considered to be one of the most important avenues to understanding how living things are built at the most basic level.

This understanding may enable future doctors to correct inherited genetic defects and to treat more effectively a variety of degenerative diseases thought to be related to the breakdown of genes that control the function of cells.

Many scientists hope a better ability to manipulate genes in plants and animals will also lead to safer pesticides, new microbes to break down pollutants and more nutritious foods.

### Like Mass Production

The achievement by the Belgian team, headed by Dr. Walter Fiers of the State University of Ghent, might be likened to the manufacture of the first mass-produced automobile after others had worked out the techniques of the assembly line.

"The work demonstrates how far science has come," said Dr. David Baltimore, a leading molecular biologist from the Massachusetts Institute of Technology. "It is a landmark of summation of many years of work, going back to many other groups. The Belgian team took on the hardest task. They brought Sanger's idea to fruition."

Dr. Sanger's team at the Medical Research Council Laboratory of Molecular Biology is currently trying to develop methods to analyze sequences of deoxyribonucleic acid (DNA)—the material that the genes of man and nearly all other living things are made of.

The RNA gene deciphered by the Belgian group came from a virus called MS2 that infects bacteria. The gene dictates the production of the protein that

forms most of the outer coat of the virus. Although a number of isolated pieces of this gene had previously been determined, the Fiers group was the first to work out the entire gene structure and link it to the critical punctuation marks that tell where the gene starts and stops.

"If you ever want to manipulate a gene—to turn it on or off—you have to know where the signals are that say 'start here' and 'stop here,'" Dr. Baltimore explained.

The ability to turn off (that is, make nonfunctional) an RNA gene might one day be important in cancer control, since most of the viruses known to cause cancer in animals consist of RNA genes.

Once the sequence of the MS2 gene was known, the Fiers group was able to suggest a probable model of how the gene exists when it is not working. The group concluded that the molecule must fold up on itself into hairpin-like loops, forming a structure that roughly resembles a flower. Hence, they dubbed it the "flower" model.

Dr. Norton Zinder of Rockefeller University, a leader in the field of RNA virus research who called the Belgian work "a real technological feat," said that knowledge of the folded structure "explains certain important biological properties of the molecule."

For example, the flower structure must first be opened up and the information of that gene "read" before the cell substances that translate the genetic code can recognize the start of the next attached gene.

### Words of 3 Letters

The "letters" in the words of the genetic code are molecules called nucleotides. RNA consists of the nucleotides adenine, cytosine, guanine and uracil, commonly abbreviated A, C, G and U. Each word in the genetic code consists of three nucleotide letters, yielding 64 different possible words known as triplets.

Most of the triplets serve as code words for one or another of the 20 amino acids that, when strung together in orderly fashion, make up protein molecules. A few of the triplets act as punctuation marks.

Thus, the triplets GCU, GCC and GCA are translated as specifying the amino acid alanine, the triplets UCU, AGC, UCA, UCG and UCC specify the amino acid serine, and so forth.

The MS2 coat protein, a rather small molecule as proteins go, contains 129 amino acid units, the exact sequence of which had been previously determined. Thus, the gene that codes for that protein would contain three times that number of nucleotides—or 387—plus a number of untranslated nucleotides that perform punctuation and other as yet unknown functions.

### Enzymes Used

To determine precisely which code words are used and in what order, the Fiers team used enzymes to "digest" the RNA molecule in orderly fashion into ever smaller fragments. Then, by recognizing overlapping fragments, making educated guesses based on knowledge of the protein's structure and combining this new information with previously analyzed pieces of the gene, the team was able to deduce the precise nucleotide sequence of the complete gene.

In all, they found that 49 different code words were used in the gene.

It was "a real tedious job that took a lot of people two to three years to complete," Dr. Zinder said.

The authors of the original report of the achievement, published last May in *Nature*, were W. Min Jou, G. Haegeman, M. Ysebaert and Dr. Fiers.

The team's achievement may also be useful when science is ready to manufacture genes to perform certain desired functions, such as to supply a missing protein or neutralize the effects of an unwanted one.

Dr. Mark Ptashne, a geneticist at Harvard University, said that before a gene is made to order, it would be helpful to know which words in the genetic code are most frequently used in actual biological organisms.

Dr. Baltimore said there may also be important biological influences exerted by the "nearest neighbor" triplets.

"As we collect more and

more knowledge of how genes are constructed, we can better unravel how they work," he added.

