Molecular Biology: A Better Artificial Gene

The first wholly artificial gene with the potential for functioning inside a living cell has been synthesized by Nobel laureate Har Gobind Khorana and his associates at the Massachusetts Institute of Technology, Cambridge. The gene, only the second ever to be synthesized, is a 126-unit DNA fragment that codes for the production of tyrosine transfer RNA in *Escherichia coli*. Announcement of the synthesis was made last month in Chicago by one of Khorana's colleagues, Kanhiya Lal Agarwal, at the 166th National Meeting of the American Chemical Society.

Khorana's group also produced the first artificial gene: In 1970, while still at the University of Wisconsin in Madison, he announced the synthesis of a 77-deoxyribonucleotide gene that codes for the production of alanine transfer RNA in yeast. He had chosen that particular gene because yeast alanine transfer RNA was the only transfer RNA whose sequence was known when the work commenced in 1965. (Its sequence was announced that year by Robert W. Holley and his associates at Cornell University.) Since the sequences of the DNA gene and the transfer RNA are complementary, knowledge of one provides knowledge of the other.

That choice, however, proved unfortunate for at least two reasons. Most important, the artificial alanine transfer RNA gene, while structurally correct, was nonfunctional both in cells and in test tube experiments because it did not contain the initiator and terminator signals that start and regulate the synthesis of the transfer RNA. These signals are polydeoxyribonucleotides of undetermined length and are attached to both ends of the gene, enabling the cellular machinery to recognize and control the gene. The nucleotide sequence of these controls is not known for any gene.

Even if the control nucleotides were available, moreover, there was no way to determine whether the artificial gene would function in a yeast cell since such cells already produce alanine transfer RNA. Khorana is confident that both of these problems can be overcome with the *E. coli* gene.

The biologically active 85-nucleotide tyrosine transfer RNA was first sequenced in 1968 by John Smith's group at the Medical Research Council, Cambridge, England. In 1970, however, Smith and Sidney Altman found that the first product produced by the gene is a "precursor" transfer RNA containing 126 nucleotides. For some as yet unknown reason, the extra 41-nucleotide fragment is cleaved from the larger segment after synthesis. It is the DNA segment which codes for these 126 nucleotides that Khorana's group has synthesized.

The M.I.T. group actually synthesized two complementary 126-unit polynucleotides to form a two-stranded, helical gene. The overall synthesis was accomplished by first synthesizing short—10 to 14 nucleotides—segments of each strand. When complementary segments of unequal length were then placed in solution, they formed a two-stranded complex with the longer strand extending from one end. This short segment of single-stranded DNA then acts as a template for binding the adjacent segment of the complementary strand, which can then be linked to the first segment with a DNA ligase.

Each Segment Is a Template

Part of the newly added segment then extends as a single strand to serve as a template for the addition of another segment. In this fashion, the Khorana group produced four major subunits, which were then combined to produce the gene. This is the same procedure that was used to synthesize the yeast alanine transfer RNA, and at this stage the group had merely repeated its earlier work, although the product was about 70 percent larger.

The second synthesis is potentially far more significant, however, because of two other discoveries made by Smith's group. They found that the tyrosine transfer RNA gene could be readily incorporated into a bacteriophage—a virus that infects bacteria—which could be used to introduce the artificial gene into bacterial cells. They also isolated a mutant *E. coli* strain in which short, nonfunctional proteins are produced as a result of an apparent defect in the terminator signal for the tyrosine transfer RNA gene. This defect can be overcome by introduction of a functional gene through use of the bacteriophage. The mutant strain thus provides a means of determining whether the artificial gene is, in fact, functional.

That determination cannot be made, however, until the initiator and terminator segments have been identified and synthesized. So far, Agarwal says, the group has sequenced the first 24 nucleotides of the terminator segment, which they believe to be a significant fraction of the control unit. They have also begun sequencing the initiator signal, but say that they still have no idea how long it is.

Their sequencing method is extremely tedious, but very effective. One strand of the two-stranded natural gene is allowed to form a complex with a short segment of the artificial gene that complements a segment of the natural gene bordering on the unknown natural segment. In the presence of a DNA polymerase, they then add one nucleotide at a time to the complex until they find which type is attached to the end of the artificial segment. By continuing this process repeatedly, they are able to determine the complete sequences.

The genes that code for transfer RNA's are by far the simplest genes to work with, since there is a one-to-one correspondence between the number of deoxyribonucleotides in the gene and the number of ribonucleotides in the product. A gene that codes for a protein, in contrast, would have to have at least three nucleotides for every amino acid in the protein, and would thus be far larger and much more complicated. Since most potentially curable genetic abnormalities in humans involve protein deficiencies, the synthesis of a clinically useful gene will be very difficult, even after the major problems of identifying and sequencing the appropriate genes have been overcome. Nonetheless, the Khorana group's results represent a major step toward the day when biochemical cure of genetic defects can be a reality.

—THOMAS H. MAUGH II