Molecular Biology: Moving toward an Understanding of Genetic Control

The advances of molecular biology have again come into the public eye during the past few months. Immediately after a press conference called by Harvard biologists to announce that they had isolated a gene, London's *Evening Standard* carried the headlines, "Genetic 'bomb' fears grow." On the same day London's *Daily Mail* voiced the alarm of "The frightening facts of life. Scientists find secret of human heredity and it scares them." The official publication of the Harvard work which resulted in the foregoing headline was entitled "Isolation of pure *lac* operon DNA" and was published in *Nature* [J. Shapiro, L. Machattie, L. Eron, G. Ihler, K. Ippin, J. Beckwith, *Nature* **224**, 768 (1969)].

Two weeks later when results obtained at the University of Bristol and at Cambridge on the sequences of bases in the RNA of bacteriophages were published in *Nature*, London's *Evening News* hailed the British work as "more important a step on the way to genetic engineering than the much-publicized American achievement." (The Cambridge research was performed by an American on an NSF postdoctoral fellowship.)

Gene Isolation

The Harvard research team isolated an integrated group of genes from the intestinal bacterium *Escherichia coli*. More particularly, they isolated one section of *Escherichia coli* DNA which is called the *lac* operon and which is responsible for metabolizing lactose. The impetus for their work is the need to test models of the regulation of DNA-directed protein synthesis. It is generally recognized that protein synthesis occurs in several steps, but there is considerable doubt about which step is controlled. Most models are made in the assumption that the rate at which messenger RNA molecules are transcribed from DNA is controlled or that the rate at which messenger RNA is translated into proteins is controlled. Isolated genes will now allow geneticists to study the processes of transcription and translation independently, without having to account for complicating cellular effects.

The biologists infected the bacteria with bacteriophages which incorporate their host's DNA into their own. In particular, λ plac 5 phage integrates its genetic material next to the *lac* operon in the chromosomes of *E. coli*. When the phage leaves the bacterium, the *lac* operon goes along and constitutes about 5 to 10 percent of the genomes of these phage particles. As the *lac* operon constitutes less than 0.1 percent of the DNA in *E. coli*, this is in effect a 100-fold enrichment of *lac* operon DNA. The next problem for the biologists was to separate the *lac* operon from the phage.

DNA is, structurally, a double-stranded helix. However, only one of these strands is transcribed into messenger RNA, and hence is called the "sense" strand. The two DNA strands of the phage, known as the heavy (H) and the light (L) strand, are characterized chemically by different base arrangements and can be physically separated. The phage λ plac 5 incorporated the *lac* operon "sense" strand into its L strand. In another type of phage used by Shapiro and his colleagues, ø 80 plac 1, the *lac* operon was inserted into its DNA in the other orientation so that the "sense" strand of the *lac* operon was in the H strand. Thus when the H strands from the two different phages were mixed and annealed, the only part of the DNA that had a complementary or matching component was the *lac* sequence. The biologists had made a small double strand or duplex of *lac* operon in the middle with single unmatched strands of phage DNA dangling at the ends.

The *lac* duplex was freed of its four single-stranded tails by a nuclease that was specific for single-stranded nucleic acids. This enzyme attacks single strands of DNA but cannot digest double strands. The end product was not, however, absolutely pure *lac* operon since whole viral-like DNA duplexes were seen in the gene preparation. The possibility also exists that some single strand DNA and duplexes formed from a second region of homology between H strands of λ plac 5 and ø 80 plac 1 are present. It is important to determine the nature of any contaminating DNA since the presence of non-*lac* DNA might greatly complicate studies on the in vitro transcription of the *lac* operon.

The electron micrographs showed that the average length of the purified molecules was 1.4 μm. Most of this length (1.29 μm) is taken up by the structural gene (*z* gene) which specifies the first enzyme in the operon. The rest (about 0.14 μm) contains the control elements. One is the promoter region and the other is the operator region (p-o region). The enzyme RNA polymerase binds to the promoter region for the start of transcription of the operon into RNA, and the repressor protein binds to the operator region to shut off RNA transcription. From its length, Shapiro and his co-workers deduced that the operator region contains about 410 base pairs.

The Harvard group is not the first to isolate nuclear genes of known biological function. Over a year ago, Max Birnsteil and his colleagues at the University of Edinburgh reported the isolation of gene material from cells of the toad *Xenopus laevis* [*Nature* **219**, 454 (1968)]. In that work DNA molecules, called ribosomal DNA satellites, were separated from the bulk of the nuclear DNA. About one half of this isolated fraction consists of the ribosomal RNA genes of *Xenopus*. The other half is composed of DNA sequences with a very high guanine plus cytosine content. These sequences are linked to the ribosomal RNA genes and serve no known function.

Also D. E. Kohne of the Carnegie Institution of Washington isolated the sequences in *E. coli* DNA that are homologous to ribosomal RNA [*Biophys. J.* **8**, No. 10, 1104 (1968)]. The key to these experiments was reducing the piece size of the DNA to a point where the amount of nonribosomal DNA isolated with the specific ribosomal genes is less than 10 percent.

One of the most perplexing problems about the mechanism of gene control is how the cell knows where and when to initiate protein synthesis. Even a simple messenger RNA molecule contains several thousand bases with information for synthesis of three or four proteins. In addition, there are regions of the RNA where the bases cannot be translated into protein struc-

ture. It is believed that instructions for the initiation and termination of protein and RNA synthesis are contained in these untranslated regions. Although they did not discover the function of the untranslated regions, investigators at Cambridge and Bristol have shown that they are distributed between the genes and at the ends of the RNA molecule. The research at Cambridge and Bristol was directed toward learning the exact base sequence that initiates protein synthesis in RNA phages.

Protein Synthesis

When the four RNA bases, adenine (A), guanine (G), cytosine (C), and uracil (U), are combined in groups of three, they form a "codon" which specifies amino acids in protein synthesis. For example, GCU codes for the amino acid alanine. The signal for the initiation of protein synthesis is either of the codons AUG or GUG. As messenger RNA must bind to the ribosomes for protein synthesis, it was at first believed that the ribosome starts to translate messenger RNA by binding to one end and then sliding along to the first AUG or GUG codon. However, it has now been shown that ribosomes do not necessarily start at the end of an RNA messenger. This finding complicates our understanding of the genetic code. The codons AUG and GUG appear at many positions along the RNA sequence, and their position apparently determines which parts of the RNA are translated. Also, the triplet combinations can change by a simple shifting of their phase. That is, the bunches of three bases are different depending on where you begin. How then does the ribosome recognize the triplet that initiates the process?

Joan Argetsinger Steitz, working at Britain's Medical Research Council Laboratory of Molecular Biology, Cambridge, succeeded in isolating and deciphering sequences of RNA from the bacteriophage R17 [*Nature* **224**, 957 (1969)]. The phage RNA has approximately 3300 nucleotides or bases, and about 3000 of these accommodate the structural genes that direct the synthesis of three proteins. One protein makes the coat of the virus, one is the enzyme or replicase protein which helps in replications of RNA molecules, and the third is the maturation protein which is involved in the assembling of the protein coat and RNA into a complete virus. Steitz isolated three different sections of the RNA, which included the beginning of each of the three genes. The sections also included the initiator regions immediately preceding the gene on the RNA molecule. Using methods developed by F. Sanger and his colleagues at Cambridge, Steitz deciphered the order of the bases in each section, enabling her to search for any clues about protein-chain initiation.

In the same issue of *Nature*, J. Hindley and D. H. Staples of the University of Bristol, using the same technique as Steitz, determined the sequence of 26 nucleotides in the initiation region of the coat protein gene of the phage Qβ. Both R17 and Qβ phages infect *E. coli* and take over the bacterium's synthesizing pathways for their own benefit. This involves the binding of phage RNA to the *E. coli* ribosomes. The three biologists were able to mix RNA from their phage with ribosomes from *E. coli* in a chemical environment appropriate for initiation of protein synthesis. With the initiation region and the beginning of a gene bound to the ribosomes, pancreatic ribonuclease was added to digest the excess viral RNA lying outside the ribosomes. Then the fragments of RNA protected by the ribosomes were isolated.

Steitz was able to identify the nucleotide sequence which corresponds to the beginning of the three genes in R17. Each had an AUG at the start, and the succeeding codons conformed with previous work on the structure of the coat protein, replicase protein, and maturation protein. However, in none of the cases was the AUG immediately preceded by any one of codons UAG, UAA, or UGA, which are known to terminate protein synthesis.

Hindley and Staples identified their RNA fragment as the beginning of the Qβ coat protein gene. They also did not observe a termination codon preceding the AUG. In fact, none of the linear sequences in the initiation region of all four RNA fragments could be decoded. This led the researchers to speculate how the higher ordering of the RNA molecule affects gene control.

Through hydrogen bonding, the RNA molecules can form two-dimensional "hairpin" loops called secondary structure, and they can even fold further into a three-dimensional tangle called the tertiary structure. It is possible that the secondary or tertiary structure shields internal codons from the ribosomes and insures that only the correct AUG is on the exterior available for binding. Steitz observed a potential secondary structure in R17 coat protein which gave maximum exposure to the correct AUG. Hindley and Staples postulated a similar structure in the Qβ coat protein sequence.

Steitz concluded that the absence of termination triplets immediately to the left of the initiator codon AUG in all the fragments demonstrated that untranslated regions are present between the genes of R17 RNA. The codons specifying termination of one protein and initiation of the next may be quite widely spaced. Moreover, unpublished work by J. M. Adams at Cambridge showed that the first 80 to 90 nucleotides from the 5′ end of the RNA molecule do not overlap with any of the three known genes. Steitz and three colleagues (R. Gesteland, P. Spahr, and J. Jeppesen) have made experiments indicating that the gene order in the R17 RNA is: 5′ end, maturation protein, coat protein, replicase protein, 3′ end.

Similar untranslated regions exist in the phage Qβ. Experiments conducted by Hindley in collaboration with C. Weissman and M. A. Billeter of the University of Zurich and J. E. Dahlberg and H. M. Goodman of the University of Geneva showed that the coat protein sequence does not appear in the first 300 nucleotides from the 5′ end of the RNA molecule and that no AUG or GUG codon appears before position 62 [*Nature* **224**, 1083 (1969)].

The British-Swiss team used a new method to determine the order of bases in Qβ. They used purified Qβ RNA replicase enzyme obtained from infected *E. coli* to synthesize RNA in a test tube. By synchronizing the start of the reaction, they could isolate homogeneous strands of RNA ranging in length from 5 to 300 nucleotides. From these they determined the sequence of the first 175 bases from the 5′ terminus. Two AUG codons appeared but not until positions 62 and 102. The sequence following both the AUG's does not correspond to the Qβ coat protein sequence. However, unpublished work by the same group showed that the coat protein gene is located between 800 and 1600 nucleotides from the 5′ end.

They also found evidence for secondary structure in the first few nucleotides. The function of the nucleotides at the beginning of the RNA molecule remains an enigma. It is

possible that they are involved with the replication of phage RNA and carry no genetic message. To solve this problem, it would be of great interest to decipher all 3300 nucleotides in one phage RNA molecule. As all the necessary techniques are available, it is probably only a matter of time before this step in the control of gene expression is understood.

—GERALD L. WICK

Dr. Gerald L. Wick is assistant science editor of New Scientist.

Power Generation: The Next 30 Years

For the past 3 decades the consumption of power in the United States has doubled every 10 years, and a number of people believe that this doubling rate is likely to continue to the end of the century. If this power is generated with existing technology, completely unacceptable levels of pollution will result.

At a symposium on "Power Generation and Environmental Change" held as part of the AAAS meeting in Boston, 16 speakers addressed themselves to the pollution problems associated with power production. Since the increased population accounts for only 20 percent of the increased power consumption, several members of the audience asked about the possibility of cutting down per capita power consumption as a solution to pollution problems. Although some of the speakers thought that this would be desirable, it was generally agreed that our society would demand the power, that it would be produced, and that changes in the technology of power production and waste disposal would be made in attempts to control pollution.

Even if nuclear reactor development proceeds rapidly, coal combustion will still account for almost half of the power produced at the end of the century. Thus, all the problems now associated with coal—including the difficult problem of controlling sulfur emission —will remain.

Most of the panelists agreed that it is easier to control pollution from nuclear reactors than from coal furnaces. Of the radioactive materials produced in nuclear reactors it now appears that most of them can be handled as solid wastes by economically acceptable techniques. The most difficult problems remaining are the development of methods for capturing krypton-85, which is now released to the atmosphere and contributes to the general background radiation, and tritium, which is part of the effluent and may be incorporated into biological systems.

Typical of the estimates for power consumption presented at the AAAS symposium are a set giving 1.5×10^{12} kilowatt-hours for 1970, 2.7×10^{12} for 1980, and 8.0×10^{12} for the end of the century. The trillion tons of coal in known, economically recoverable, reserves in the United States will last well past the end of the century, but recently passed legislation and other, expected legislation setting limits on the sulfur content of coal cannot be met with the existing supply of low-sulfur coal. Wallace Behnke, of Commonwealth Edison Company, said that short-term plans are to use more natural gas (which is the least plentiful of all fossil fuels) and foreign oil and liquid gas. When these low-sulfur fuels are exhausted, sulfur control will depend on the development of new methods of burning domestic coals.

The rate at which uranium reserves are used will depend on progress in the breeder reactor program and on our willingness to use expensive (hard to recover) deposits.

The fuel rods in the light water reactors now in use contain uranium-238 and small amounts of uranium-235 and plutonium-239, but only the last two isotopes are utilized as fuel. Uranium-238 can be used as fuel if it is converted to plutonium-239 by the capture of a fast neutron, but the fast neutrons in light water reactors are slowed down by the water coolant, and significant amounts of plutonium-239 are not produced. In the most promising types of breeder reactors, liquid sodium or fused fluoride salt coolants would be used and the concentration of fast neutrons should be high enough so that more fuel is produced than consumed.

The Atomic Energy Commission estimates that full-scale commercial use of breeder reactors will be achieved by the mid-1980's. Several observers in the coal industry and in academic and consulting positions believe that this is a very optimistic estimate, but few doubt that most of the country's nuclear power will eventually be generated by breeder reactors.

The time required for development of breeder reactors is important because they require fuels with some uranium-235, and the supply of inexpensive ores containing this isotope is limited. If reactor development proceeds as the AEC expects, the inexpensive reserves of uranium (less than \$10 per pound of U_3O_8) will be used up in the mid-1980's—the same time that the AEC estimates breeder reactors will come into use. Medium-priced fuel (up to \$30 per pound of U_3O_8) will be used up between 1990 and 2000, the rapid development of breeder reactors delaying the time by a few years. With breeder reactors the use of new reserves would level off around the turn of the century, after more than 1 million tons of U_3O_8 had been consumed. Without breeder reactors we would begin using up the 10 million tons of known reserves a few decades after the year 2000. The often-heard criticism that the AEC is pushing the deployment of the current generation of light water reactors too fast is based on the fear that these inefficient devices will deplete the supplies of inexpensive and medium-priced uranium.

Research for Pollution Abatement

All large coal-burning power plants in the United States use the pulverized fuel technique in which powdered coal is blown into furnaces with very large volumes. This method produces much particulate matter and leaves the waste gases in an oxidized state in a large volume of air.

The particulate matter can be collected with electrostatic precipitators. These utilize well-established technology, but they are usually larger than the furnace and are expensive to install. For example, the precipitator for the 1000-megawatt plant at Ravenswood, New York, cost \$10 million.

The search for an economical method of collecting the sulfur from pulverized-fuel combustion has been going on for 30 years, but there are still no efficient devices for collecting sulfur on any large power plants now in operation. Some observers believe that with either economic incentives or laws setting lower limits on sulfur emission the power companies would have effective sulfur controlling devices developed within a few years. Others believe that there are difficult technological prob-