DNA Sequencing: A New Era in Molecular Biology

Although the genetic code has been known for more than a decade, molecular biologists have so far been unable to explain how information is encoded in parts of DNA that are not structural genes (that is, genes that code for proteins rather than, say, ribosomal RNA's). Most of the DNA in cells of higher organisms, for example, does not consist of structural genes, and some of these nongene sequences are thought to be used for the control of gene expression. One step in understanding the relations between the structures and functions of segments of DNA that are not genes is to determine the nucleotide sequences of these regions. Until recently, DNA sequencing was so arduous that sequences of only very few short segments had been determined.

Within the past 2 years, fast and simple ways to determine DNA sequences have been developed. These techniques dramatically affect the time and effort required to obtain DNA sequences. For example, Walter Gilbert of Harvard University notes that, 3 years ago, he and his associates took 2 years to determine the sequence of a piece of DNA that is 20 bases long. Now, they can determine such a sequence in a day. Many investigators believe that these techniques herald a new era in molecular biology—an era in which so many DNA sequences will be known that solutions to long-standing problems of DNA structure, sequence organization, and function will become clear.

The enthusiasm among molecular biologists for the new DNA sequencing techniques is generated in part by the fact that only a few years ago they despaired of ever finding an easy way to sequence DNA. Robert W. Holley of the Salk Institute and later Frederick Sanger and his associates at the Medical Research Council in Cambridge, England, had developed methods to sequence RNA, but these methods could not be directly extended to DNA. Thus it appeared as if DNA sequences would have to be determined indirectly, by sequencing their RNA copies. This procedure greatly limits the number of sequences that can be obtained since it is not always feasible to isolate RNA copies of a particular DNA segment in quantities large enough to sequence. Furthermore, many DNA segments, such as control regions, may not normally be copied into RNA. And, although any DNA segment that can be isolated can be transcribed into RNA in vitro, until recently very few specific DNA segments could be isolated from the long, naturally occurring DNA's.

To sequence an RNA molecule, investigators first added radioactive RNA precursors when the RNA was being synthesized. This enabled them to extract labeled RNA's to use for sequencing. Then the RNA's were degraded with one of several enzymes that cleave the molecules at specific bases. One such enzyme, for example, cleaves at all guanosine bases. The resulting small fragments were separated, and the sequence of each fragment was determined. The same enzyme was then used under different conditions so that only a random subset of all guanosines would be cleaved. This allowed the partial ordering of the small fragments. The process was then repeated with an enzyme with a slightly different specificity, and a second group of small fragments, which overlap with the first group, was obtained. Finally, the sequence of the original RNA molecule was pieced together from this information.

Three crucial features of this RNA sequencing method could not be applied to DNA molecules. First, it is difficult to label most DNA's in vivo at a specific activity high enough for subsequent analysis, and no means were known to label only segments of interest for sequencing. Second, no DNA-degrading enzymes were known that were analogous to the RNA-degrading enzymes that cleave at specific bases. Third, even if such DNA-degrading enzymes were found, naturally occurring DNA molecules are so large that such enzymes would chop them into fragments too numerous to be separated.

Within the past few years, several discoveries were made that changed this gloomy outlook for DNA sequencing. Investigators were able to get around the requirement for base-specific enzymes with a new approach that turned out to be a vast improvement over the RNA sequencing methods. In addition, they found ways to cut small DNA segments of interest from large DNA molecules and then they could radioactively label these segments.

The new approach that avoids base-specific enzymes was introduced by Sanger and his associates. These investigators developed several DNA sequencing methods that differ in their details, but have the same underlying idea: namely, a small piece of DNA to be sequenced is identified. Then all sequences beginning at the one end of that DNA segment and extending to all possible lengths (initial segments) are obtained. This means that these initial segments consist successively of only the first nucleotide, of only the first two nucleotides, of only the first three nucleotides, and so on, until the final segments, which consist of the entire DNA piece to be sequenced. These initial segments are then separated by length. By various means, the investigators determine the identity of the terminal nucleotide of each initial segment. Thus they can read off the DNA sequence by citing the order of the terminal nucleotides of the initial segments in order of increasing length.

When Sanger first introduced this concept for obtaining initial segments, DNA sequencing was limited by a lack of methods available for isolating specific short DNA fragments from long naturally occurring DNA molecules. Sanger and his associates got around this problem in one instance when they used their knowledge of part of the sequence of a viral DNA to devise a means to sequence the DNA from that virus. This approach, however, cannot readily be extended to most naturally occurring DNA molecules.

The key to obtaining short DNA fragments for sequencing lies in the use of restriction enzymes. Molecular biologists have begun to make extensive use of these enzymes, which can be isolated from bacteria. Each restriction enzyme recognizes a specific DNA sequence and cuts the DNA at that sequence. Because these enzymes are so important for DNA sequencing, Richard Roberts of Cold Spring Harbor Laboratories has spent the past 3 years engaged in an extensive search for such enzymes. The enzymes, he says, are so common that about one-third of the bacterial strains he screens produce them. So far, he has screened nearly 300 strains and he and others have isolated a collection of about 45 enzymes with different specificities.

When short DNA segments have been isolated, they are relatively easy to label radioactively. For example, enzymes can be used to add a radioactive label to one end of a DNA segment. This is crucial for DNA sequencing since all initial

segments must start at the same position. If that position is radioactively labeled, only segments with that label are detected as initial segments.

Methods for separating initial segments by length have been greatly improved in the few years since DNA sequencing began. About 2 years ago, investigators used a two-step procedure to separate molecules no longer than about 20 nucleotides. Now, with recent improvements of a one-step method (gel electrophoresis), Gilbert reports that he can separate molecules up to 100 nucleotides in length. Gel electrophoresis is not a new technique, but it could not previously be used to separate DNA molecules that differ in length by only one nucleotide. Sanger, and independently, Gilbert, however, modified conditions for gel electrophoresis until they achieved this resolution. When initial segments are separated by gel electrophoresis, the smallest segment forms a band at the bottom of the gel, the next largest forms a band just above the smallest, and so on.

As a consequence, two extremely rapid methods to sequence DNA have been developed to obtain small DNA fragments of interest from large DNA molecules, to label the initial segments, and to separate initial segments of up to 100 nucleotides by length. The major difference between the two DNA sequencing techniques is the means by which the initial segments are obtained and their terminal nucleotides identified. A method, published last year by Sanger and A. R. Coulson, relies on enzymatic reactions. An alternate approach, based on chemical reactions, was developed by Allan Maxam and Gilbert, and is not yet published.

Sanger and Coulson obtain their initial segments from an enzyme-catalyzed synthesis of them. They take a single-stranded piece of DNA and begin to copy it, starting from a specific position. These copies are made to terminate at specific bases under different reaction conditions. For example, to obtain all initial segments ending just before a cytosine, they first let the synthesis of the copy of the DNA fragment of interest begin. Then, they stop the reaction and start it again in the presence of all the nucleotides except cytosine. Thus each initial segment in this reaction mixture will be elongated until a cytosine must be added to it, whereupon the copying of the DNA template will stop for lack of cytosine.

Maxam and Gilbert developed a completely different method to obtain initial segments. They use chemical reagents to break DNA fragments of interest at specific bases. One reagent breaks DNA specifically at purines (adenine and guanine) and the other breaks DNA specifically at pyrimidines (cytosine and thymine). They can adjust their reaction conditions, though, so that the purine-specific reagent cleaves preferentially at an adenine or so that it cleaves preferentially at a guanine. Similarly, they can cause the pyrimidine-specific reagent to cleave preferentially at a cytosine.

To obtain all initial fragments that end just before a guanine, Maxam and Gilbert proceed as follows: First, they label their DNA fragment of interest at one end. Then, they allow the purine-specific reagent to react with the DNA under conditions in which only one base, on the average, is hit per DNA fragment and in which guanines are far more likely to react than adenines. When they break a large collection of copies of that DNA fragment, they should obtain all initial fragments that end just before a guanine and some of those that end just before an adenine. The initial segments can be distinguished from segments that begin elsewhere in the DNA fragment since only the initial segments will carry the radioactive label. The initial segments that end just before a guanine can be distinguished from those that end just before an adenine since there will be many more copies of each of the guanine segments than of the adenine segments. Thus the guanine segments will form dark bands on the gel (indicating that there are many copies of each of them) and the adenine segments will form light bands. The other initial segments are isolated and identified in an analogous way (Fig. 1).

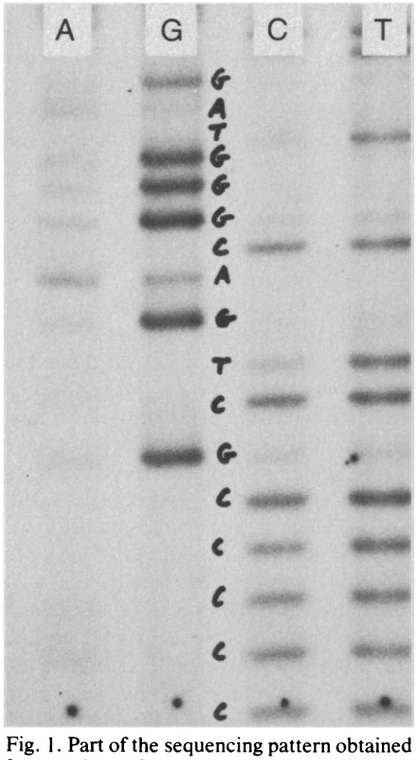

Fig. 1. Part of the sequencing pattern obtained from a piece of DNA about 130 base pairs in length. The letters at the top of the columns, A, G, C, and T (adenine, guanine, cytosine, and thymine) indicate which base was preferentially cleaved by chemicals. The darkest band in each column represents the base missing from the end of the initial segments, with the exception of cytosine (C). All dark bands in the C column represent cytosines even if bands also appear in the T column at that position. Bands that appear in the T column but not in the C column represent thymines (T). To read the sequence of the DNA, read off the base represented by each band, starting from the bottom of the columns. [Source: Walter Gilbert and Allan Maxam, Harvard University]

Although Gilbert and Maxam have so far sequenced DNA segments of at most 100 nucleotides, they believe that much longer segments can be sequenced with their method. This is because the chemical reagents cut the DNA in a uniform way independent of the length of the DNA. Thus the length of DNA that can be sequenced is restricted only by the ability of gel electrophoresis to resolve the initial segments, and this resolution can be increased by increasing the lengths of the gels and the strength of the current used for electrophoresis. Maxam and Gilbert estimate that they can increase the number of initial segments separable by gel electrophoresis so that DNA fragments 800 nucleotides long can be sequenced. (This will require a 5 m gel with a potential of 20,000 volts across it, Maxam estimates.)

The new DNA sequencing methods are now being used to study a wide variety of DNA segments. Maxam and Gilbert, for example, are sequencing regions of a set of genes and control regions of *Escherichia coli*. They are using their technique to determine directly what sections of the control regions of this DNA are bound by proteins. When a protein binds to DNA, the DNA is protected from the chemicals that react with DNA bases and hence such a DNA segment can be identified.

The method of chemical cleavage, unlike the enzymatic methods, works as well with single-stranded as with double-stranded DNA molecules. Maxam and Gilbert are exploiting this feature to sequence single-stranded viral DNA's and to sequence RNA molecules. A single-stranded DNA molecule can be copied from an RNA molecule with a reverse transcriptase. Since DNA is now so

much easier to sequence than RNA, this provides an easy way to obtain RNA sequences.

Tom Maniatis of Harvard University and, independently, Nina Fedoroff of the Carnegie Institute of Washington in Baltimore, are developing a modification of the enzymatic method of Sanger and Coulson for the direct sequencing of restriction enzyme fragments. Maniatis has obtained relatively large quantities of the hemoglobin gene by molecular cloning methods (*Science*, 19 March 1976, page 1160), and is now sequencing regions of this gene that are not transcribed into protein. He and his associates are also sequencing another class of DNA sequences —the satellite (small DNA) sequences— that do not consist of genes and that are present in huge amounts in cells of higher organisms. Winston Salser and his associates at the University of California at Los Angeles are using the chemical method to sequence the hemoglobin gene and satellite sequences. Salser, Randolph Wall, and their colleagues are, in addition, cloning and sequencing immunoglobulin genes. Fedoroff is sequencing the control regions of a set of genes, the 5*S* ribosomal RNA genes, that are repeated many times in cells of toads. Other investigators are now beginning to sequence nearly any DNA that they can obtain in large amounts.

The new interest in DNA sequencing does not mean that RNA sequences are no longer of interest. Since RNA molecules carry important biological information, work on their sequences is continuing unabated at a number of laboratories. However, because DNA sequencing is now so extremely simple the direction of research in molecular biology will surely change. Useful information can now be quickly gathered—information that a few years ago investigators had little hope of gathering.

—GINA BARI KOLATA

Marihuana: A Conversation with NIDA's Robert L. DuPont

Robert L. DuPont, director of the National Institute on Drug Abuse (NIDA), made headlines recently when he became one of the first Administration officials publicly to recommend a liberalization of marihuana laws. The occasion was a press conference accompanying the release of *Marihuana and Health*, NIDA's fifth annual report to the Congress on marihuana research. Recently, *Science* talked with DuPont and asked him to elaborate on those views.

DuPont, a 40-year-old Harvard M.D., served as Director of the District of Columbia's Narcotics Treatment Administration from 1970 to 1973, where he conducted a comprehensive program for treatment of heroin addiction. In June 1973, Richard Nixon appointed him director of the White House Special Action Office for Drug Abuse Prevention. He held that position until the office was terminated on 30 June 1975. He has been director of NIDA since September 1973.—THOMAS H. MAUGH II

How would you summarize the new report to the Congress?

I think that it is hard to summarize. People are looking for simple statements that marihuana is safe or that it is dangerous, and the report defies that kind of summary. But there is a broad range of biological concerns reflected in the report, from the decrease in testosterone levels and effects on cell-mediated immunity to bronchitis and the potential for cancer. There is a growing concern about the seriousness of these negative health effects—although they are something less than has been searched for by many, such as some tremendous evidence that marihuana users' ears fall off or that their noses turn green, or something like that. Also, we have overlooked the problems of marihuana intoxication in the past. They are related most urgently to driving performance, but they are also related to work performance, to studying, and to interpersonal relations—to the activities where, we are aware, most of the problems of alcohol are concentrated. We've just ignored that in the past discussion of marihuana's health hazards.

Is the pattern of marihuana use changing?

There is evidence not merely of more widespread use, but of much more frequent use. This is highlighted in the report. Those people who do use marihuana tend to use it more frequently than most of us thought. In a study of high school seniors who used marihuana at the time of the survey, 23 percent of the user group reported daily use. Now that's different from alcohol, where the equivalent figure was about 9 percent. It suggests that there is a tendency to more frequent use if there is any use. I've talked to many marihuana users who report that they have to make an effort to limit their use—and it is not all that easy for many people to do that. In this way marihuana use is more like cigarette smoking than it is like alcohol. And I think we've been, perhaps, misled, or have misled ourselves, into thinking that there is not a very large potential for a lot of people to be on the very heavy use end of the spectrum.

Is the potency of street marihuana changing?

There has been a tendency toward rising potency levels. The conventional wisdom is that marihuana is of low potency—often said to be less than 1 percent THC [tetrahydrocannabinol]—and that the THC content of hashish is around 10 percent. But more recent evidence suggests that much marihuana in the United States has 2 or 3 percent THC and that most hashish is down around 4 or 5 percent; so the distinction between marihuana and hashish is not as great as people had thought. And there is good reason to believe that we're going to have marihuana in the potency range of 5 or 6 percent in the near future.

Is the more potent form potentially more hazardous?

Of course, all the effects are exaggerated with the more potent form. But the most important effect is an increased likelihood of having an acute adverse reaction—a panic reaction. But that's largely compensated for by the experience of the user.

Are marihuana use and alcohol use mutually exclusive?

The old idea was that somehow marihuana was *replacing* alcohol for young people—that if people used one, they wouldn't use the other. The argument was that marihuana is a safer drug than alcohol so, therefore, let them smoke pot. All of the available evidence, unfortunately, cuts the other way. It suggests that people who use alcohol are more likely to use marihuana and that people who use marihuana are more likely to use alcohol. So what we have is a drug-using behavior, particularly as exemplified in the two most common intoxicating substances, marihuana and alcohol, advancing as a front rather than moving like two ends of a teeter-totter. The effects on driving appear to be partic-