19 September 1980, Volume 209, Number 4463 SCIENCE

A Revolution in Biology

John Abelson

While molecular biology has been a productive and growing discipline for 30 years, we are now in the middle of a period characterized by an unprecedented level of activity and excitement (*1*). Two related technologies, gene cloning and rapid DNA sequencing, have opened up vast new areas for research. The first development, recombinant DNA, made it as simple to isolate and study the structure of eukaryotic genes as it previously was to study bacterial genes (probably easier since much is still being learned about bacterial genes by the recombinant DNA approach).

The development of rapid DNA sequencing techniques by Sanger *et al.* (*2*) and by Maxam and Gilbert (*3*) has made it possible to determine the molecular structure of eukaryotic genes, an achievement that we could not have dreamed of 10 years ago.

The current technology of this revolution in biology has its origin in bacterial genetics and enzymology. Research in molecular biology is now focused on higher organisms, but *Escherichia coli* remains an essential resource material. The recombinant DNA technology derives from and depends on the elegant manipulations of bacterial genetics. The technology also depends on the growing battery of enzymes as reagents. Thanks to the development of dependable and reproducible methods for enzyme purification, a large collection of enzymes can now be prepared and are offered commercially. Host restriction and modification were first studied by Arber as an interesting phenomenon in microbial genetics. The type II restriction endonucleases were purified by H. Smith from *Haemophilus influenzae* and were first encountered as enzyme activities.

The articles in this issue present some of the current trends in this field.

The Structure of Eukaryotic Genes

The most dramatic revelation in the first encounters with eukaryotic genes is that the gene is often not colinear with its product. Instead genes contain intervening sequences or introns which interrupt the continuity of the genetic information. As a result, a eukaryotic gene coding for a protein may be much larger than is required for the simple coding of the amino acid sequence. The β-hemoglobin gene in mouse, for example, contains two intervening sequences of 116 and 646 base pairs (*4*). The coding sequence is 432 base pairs. This is hardly the most extreme example. The dihydrofolate reductase gene of the mouse contains five intervening sequences. The entire gene containing 568 base pairs of coding sequence is 32,000 base pairs in length (*5*).

A newly discovered enzymatic reaction, splicing, is required for the expression of the genes. The entire gene is transcribed as a long RNA precursor, the intervening sequences are then clipped out, and the ends are rejoined to yield the uninterrupted coding sequence in the mature messenger RNA.

Several sets of genes have been investigated in detail. Proudfoot *et al.* (page 1329) reviews the progress in the characterization of the human globin genes. From the perspective of 1970, the progress is incredible. All the globin genes have now been cloned, and the β and α clusters have been mapped. The complete nucleotide sequences of the five β-like globin genes are known, and it is not unlikely that in the end the entire 60,000 to 70,000 base pairs of the β cluster will be known. This is a region that is rich in biological significance. The expression of the different globin genes is differentially regulated so that the system is an excellent model for the study of development. A number of human mutations, including the thalassemias, are known which affect the expression of these genes, and already much information is known relating the mutations to the map of the gene cluster.

For some time it has been known that genomes are not static but that there is considerable scope for the movement of blocks of genetic information. The genetics of this phenomenon have been studied for at least 40 years. Soon after the pioneering work of Barbara McClintock with maize in the late 1950's, similar transposable elements were found to exist in bacteria. The recombinant DNA technology has now provided appropriate tools for demonstrating these genetic phenomena at the molecular level.

A unity in the mechanism by which transposable elements function is now beginning to emerge. Transposable elements are blocks of genetic information that carry with them the capability to catalyze their own movement. This is done by facilitating genetic recombination of the ends of the elements with other, nonhomologous sequences. This can lead to a number of genetic phenomena—gene instability, transversions, inversions, and deletions. Thus the mechanism of transposable antibiotic resistance genes in bacteria may be analogous to the movement of transposable elements in yeast and, in *Drosophila*, to the integration of RNA tumor viruses and to the mechanism of expression of immunoglobulin genes (*6*).

Several articles deal with transposable elements. Simon and his colleagues (page 1370) are studying phase variation in *Salmonella* in which the bacterium is able to switch between the production of two antigenically distinct flagella types in order to escape the immunological defenses of the host. Phase variation, it was discovered, is mediated by a recom-

John Abelson is professor of chemistry at the University of California, San Diego 92093, and member of the Agouron Institute, La Jolla, California.

0036-8075/80/0919-1319$00.75/0

binational switch, a 900-base pair fragment which undergoes inversion in the *Salmonella* genome. This fragment codes for a protein that mediates the recombination. Furthermore, this protein and the ends of the switch on which it must operate are homologous to those in other bacterial transposable elements—a finding suggesting the unity of the mechanisms involved.

As early as 1968 Britten and Kohne (*7*) showed that eukaryotes contain nucleotide sequences that are repeated many times throughout the genome. In *Drosophila* and in yeast at least some of the middle repetitive DNA segments are transposable elements. Roeder *et al.* (page 1375) present an analysis of mutations in yeast that are caused by the transposable elements *Ty*1 and *Ty*2. By cloning and sequence analysis, Roeder *et al.* show that these mutations bear similarities to the bacterial transposable elements in that they create a 5-base pair duplication of the target DNA in the process of transposition. By genetic analysis they demonstrate an exact analogy of the yeast transposed with the two-component *spm* system in maize. In both systems there are suppressors of unstable mutations and these produce the wild-type phenotype. In yeast these act by catalyzing the clean excision of the inserted transposable element. Thus again, the unity of these mechanisms is demonstrated.

The recombinant DNA technology has been used to elucidate the mechanisms by which immunoglobulin genes are expressed and the diversity of immunoglobulins is generated. To molecular biologists viewing the eukaryotic world from afar this was probably the most intriguing of problems. How is an organism able to generate a repertoire of 10^5 to 10^6 different antibodies? Amino acid sequence analysis of immunoglobulins led to some definite information about how diversity must be generated, but the cloning and sequencing of antibody genes has now revealed how it is done, virtually at the molecular level. In fact antibody genes are generated by DNA recombination (in this case deletion) between variable genes (which code for the portion of the antibody which recognizes the antigen) and constant gene (which provide the framework). It is now believed that there are enough variable genes, used in combination, to generate the diversity which is required. The articles on immunology in this issue illustrate the advanced state of this field despite the fact that the first such analysis is not yet 5 years old (*8*).

Directed Mutagenesis

Once a gene has been isolated it can be altered in vitro. This can be done with the use of restriction enzymes, exonucleases, ligases, and chemical mutagens to produce deletions, insertions, and point mutations (see Peden *et al.*, page 1392). The altered genes are, of course, recloned in order to purify and propagate them.

This strategy has been used, as described in a number of articles in this issue, in order to investigate the role of sequences in the gene or its control elements. In bacteria, promoter and operator mutants had been isolated, and the eventual sequence determination of the nucleotide changes helped to designate the functionally important regions of the control element. In eukaryotes, few such mutants have been isolated, and hence in vitro (or surrogate) genetic techniques should have wide application. Some of the early results with this technology have been surprising. Brown's group has demonstrated that the promoter recognition sequence for the transcription of the 5*S* gene in *Xenopus* is actually in the middle of the gene (*9*). Using these techniques, Corden *et al.* (see page 1406) and Proudfoot *et al.* have been able to designate promoter regions which specify the initiation start sites for the RNA polymerase II transcription of several different eukaryotic genes. Surprisingly, these promoters are not strikingly different in size and disposition to the start site from prokaryotic promoters.

The most precise of the mutagenic strategies involves the use of synthetic oligonucleotide to direct the mutation. The techniques required to synthesize blocks of DNA sequence developed by Khorana and his colleagues (*10*) have provided crucial and unique contributions such as in the elucidation of the genetic code. Itakura and Riggs (page 1401) describe the more recent use of oligonucleotides in the construction of synthetic genes and as specific mutagenic agents. Wallace *et al.* (page 1396) have used this technique to engineer the precise deletion of an intervening sequence in a yeast transfer RNA gene. Oligonucleotide-directed mutagenesis makes it possible for the experimenter to alter the gene in a precise manner. As an example, the technology can be used to probe the structural basis for function in enzymes. Suppose that the proposed mechanism for the catalytic action of an enzyme invokes a crucial role for a particular serine residue. It is possible, using oligonucleotide-directed mutation to change the serine codon to, say, an alanine, or any other amino acid codon that one may wish to insert at that site. The altered gene can then be inserted into *E. coli* in order to produce the mutant enzyme. Protein chemists are already thinking of how this technology might be applied, and thus projecting an insatiable demand for oligonucleotides.

Transformation

Once genes have been altered, one must be able to test for changes in function. There are a number of in vitro techniques for doing this, including injection of DNA into the nucleus of oocytes, incubation with oocyte extracts, and use of partially defined RNA polymerase II systems (see, for example, Proudfoot *et al.* and Corden *et al.*). Ultimately, however, the altered gene must be returned to the organism. Transformation techniques have now been worked out for yeast and for cultured mammalian cells.

Yeast transformation is described by Scherer and Davis (page 1380), who have used the technique to investigate the genetic phenomenon of gene conversion. A number of yeast cloning vectors are now available—some in which the DNA is maintained as a plasmid and some in which it is integrated at sites of sequence homology. The frequency of transformation can be high enough to allow cloning experiments to be carried out entirely within yeast. This has allowed the cloning and analysis of genes that would have no recognizable phenotype in bacteria. Two groups have succeeded in isolating the genetic elements of the yeast mating type system (*11*). In this system switches in the mating type of yeast are mediated by a recombination mechanism bearing some resemblance to the transposable elements we have already discussed.

The fact that transformed DNA sequences can be integrated at sites of homology has allowed the development of a procedure in which entering sequences (containing alterations) can perfectly replace the original resident sequences (*12*). The organism is unchanged except for the alteration that has been introduced and the altered gene is in the proper milieu. This allows for a perfectly controlled test for the phenotype of any mutation.

Using mammalian cells, Axel and his colleagues have demonstrated that exogenous DNA can be taken up and expressed. This has allowed (Pellicer *et al.*, p. 1414) the direct selection of cellular genes coding for thymidylate kinase and

the adenosine phosphoribosyltransferase gene. Again these techniques will be used to accomplish the isolation of genes that are not expressed in *E. coli* but which have distinct phenotypes in the mammalian cell. Transformed DNA is integrated into chromosomal DNA but apparently not at unique sites.

Mulligan and Berg (page 1422) have taken a different approach. In their experiments, the transforming DNA is introduced into cultured cells via an SV40 vector. It had already been established (*13*) that in order to express stable messenger RNA a transcription unit must contain at least one intervening sequence, that is, each transcript must be spliced. This can be done by constructing hybrid transcription units—a complementary DNA coupled with an SV40 late leader, for example. Mulligan and Berg have introduced the *E. coli* gene (*gpt*) for the enzyme xanthane-guanine phosphoribosyltransferase (HGPRT) into SV40.

When SV40 provides the promoter and the RNA processing signals, the bacterial gene can be expressed. The expression was demonstrated in a dramatic way by transformation of human Lesch-Nyhan cells. The Lesch-Nyhan syndrome results from a lack of HGPRT. As a result of the deficiency, cells cultured from Lesch-Nyhan patients cannot grow in a medium containing aminopterin, hypoxanthine, and thymidine. Cells transformed by the SV40-*gpt* hybrid are rescued and can grow in the media. Obviously, these results open up a wealth of possible experiments on the control gene expression in mammalian cells.

Industrial Applications of Molecular Biology

No one can have missed the fact that there may be commercial applications of this technology. It is not yet clear how soon the technology will bring a product to market but there is no doubt that eventually vaccines and various naturally occurring polypeptides such as insulin and interferon (Streuli *et al.*, page 1343 and Masucci *et al.*, page 1431), which are in short supply, will be produced by this route. The technology already exists for engineering the overproduction of proteins in *E. coli* (Guarente *et al.*, page 1428) and for altering the genes to improve the usefulness of the protein product. On the whole, these developments should have a positive effect on progress in biology. The goals of the basic research and the engineering have many features in common. For example, what determines the optimum expression of a gene and the transport of the product through the membrane? What is the best way to engineer a particular mutation? How many interferon genes are there and where are they expressed?

Participation of molecular biologists in industry is a relatively new phenomenon. Heretofore, molecular biologists have found jobs in universities. Now, with university budgets and grants being curtailed, talented young and energetic investigators are finding places in industry to continue their research. We can anticipate some problems and some possible short-term losses. The age-old conflict between competition for the edge in the marketplace and the scientific ideal of open communication will not go away. Certainly the new recruits to industry will arrive wishing to continue communication with their colleagues. I hope that can be worked out. However, some of the most talented of our members may lose sight of the big picture. One cannot do everything, and the complicated problems of mammalian development probably will not be solved by the same people who are engineering the production of a hepatitis vaccine. But these are questions of value and who is to say which is more important?

Perspectives

We need to know much more about the structure and organization of eukaryotic genes and the factors that control their expression. The means are now available to collect these data, and the results are pouring in. Already sophisticated computer techniques are required to analyze the information, and central data banks must be organized in order to store and disseminate the information.

The techniques of gene modification and cloning are relatively simple to learn, and a large number of people are applying them to an ever broadening range of biological problems. For example, they will be increasingly used in the study of biological evolution. The DNA sequence contains the primary record. The comparison of sequences among gene families (see, for example, Proudfoot *et al.*) is already providing fascinating insights and raising new questions.

At the applied level, I believe that the most important and pressing applications will come in providing solutions to our energy problems. Photosynthesis is a solution to the need for a renewable energy source that has been perfecting itself for more than 3 billion years. Almost certainly an important component to the solution of the energy problem will have to come from biology. Engineering may be required to adapt to our needs the biological coversion of sunlight to carbon skeletons. It should be evident that many of the tools needed for this task are already at hand.

References and Notes

1. For an excellent history of molecular biology, see H. F. Judson, *The Eighth Day of Creation* (Simon and Schuster, New York, 1979).
2. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
3. A. M. Maxam and W. Gilbert, *ibid.*, p. 560.
4. D. A. Konkel, S. M. Tilghman, P. Leder, *Cell* **18**, 865 (1978).
5. J. Nunberg, R. Kaufman, A. C. Y. Chang, S. N. Cohen, R. T. Schimke, *ibid.* **19**, 355 (1980); G. Krause, C. Simonsen, R. Schimke, unpublished results.
6. See A. I. Bukhari, J. A. Shapiro, S. L. Adhya, Eds., *DNA Insertion Elements, Plasmids and Episomes* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1972); *Cold Spring Harbor Symp. Quant. Biol.*, in press.
7. R. J. Britten and D. E. Kohne, *Science* **161**, 529 (1968).
8. N. Hozumi and S. Tonegawa, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3628 (1976).
9. D. F. Bogenhagen, S. Sakonju, D. D. Brown, *Cell* **19**, 27 (1980); S. Sakonju, D. F. Bogenhagen, D. D. Brown, *ibid.*, p. 13.
10. H. G. Khorana, *Science* **203**, 614 (1979).
11. J. Hicks, J. Strathern, A. J. S. Klar, *Nature (London)* **282**, 478 (1979); K. A. Nasmyth and K. Tatchell, *Cell* **19**, 753 (1980).
12. S. Scherer and R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4951 (1979).
13. P. Gruss, C.-J. Lai, R. Dhar, G. Khoury, *ibid.*, p. 4317; R. C. Mulligan, B. H. Howard, P. Berg, *Nature (London)* **277**, 108 (1979); D. H. Hamer and P. Leder, *Cell* **18**, 1299 (1979).

13 August 1980