18 November 1983, Volume 222, Number 4625 SCIENCE

Biological Frontiers

Frederick R. Blattner

This is the fourth issue of *Science* in as many years to be devoted to the "revolution" in biology. The emphasis of this collection of papers is fundamental research rather than techniques or practical applications. Philosophers and historians of science will probably regard the current period as a time of explosive advance and refinement of information rather than a true revolution. It has not been necessary, as was the case with quantum theory or relativity, to overturn major paradigms of accepted thought; rather, an enormous increase in the power of experimental techniques is now producing answers to complex long-standing biological questions.

A cornerstone of this advance is the technique of gene cloning, which allows a segment of genetic material to be removed from its normal context in a genome and replicated to high yields for studies in isolation, or to be reintroduced into a variety of cells where it can be studied in new genetic contexts. The basic principle of physically segmenting the genome of interest is an old one. Long before the discovery of restriction enzymes, what amounted to gene cloning (although not called that), the use of transducing phages and sex factors, was a method of choice in the study of bacterial genetics. But the in vitro methods of gene splicing have enormously extended the range of application and the precision of execution of the principle. This has led for the first time to practical methods for study of eukaryotic genomes, including those of human and important plants and animals, although the techniques have been extraordinarily valuable in the study of prokaryotes as well.

The author is a professor in the Department of Genetics at the University of Wisconsin, Madison 53706.

An initial aim of many cloning experiments has been to determine gene structure through DNA sequencing. This purely structural approach has also been remarkably rewarding in providing insight into function. No more striking example can be found than the immune system, in which the first genes to be sequenced were those that code for antibody molecules. Many long-standing issues concerning the mechanisms involved in the generation of diversity, the control of expression, and the path of evolution of these genes have been brought from the level of vague conjecture to one of refined and testable hypotheses built on the discoveries of DNA rearrangement, gene conversion, alternate messenger RNA splicing, and somatic mutation. In this issue, Teillaud *et al.* describe one such detailed study focused on the role of somatic mutation in controlling the affinity of antigen for antibody.

Recently, the structural approach, in which DNA sequencing is used, has advanced to the histocompatibility genes that regulate the immune response. The rapid progress in this area and the precise analysis achieved so far (reviewed in this issue by Steinmetz and Hood) exemplify the effectiveness of gene sequencing to unravel the complexities of the most difficult biological problems.

A major line of research toward interpretation of DNA sequence involves the detailed study of "sites" in DNA or RNA that have specific regulatory functions. A site is roughly defined as a small (about 100 base pairs) region that, through its interaction with other cellular components, brings about an effect on the expression of adjacent genetic material. This concept is a very general one, specific examples of which include promoters, operators, terminators, attenuators, cap sites, retroregulators, translational control sites, RNA splice sites, polyadenylation sites, enhancers, origins of replication, ribosome binding sites, site-specific recombination sites, viral encapsidation sites, switch sites, and more. In this issue, Rosenberg *et al.* provide a review of the methods developed to study the behavior of transcription initiation and termination sites through the use of specially designed vectors. The key to this type of analysis is the provision of a standard gene whose expression under the control of the particular "site" under investigation can be quantitatively monitored. Recombinant DNA molecules containing regulatory sites are also useful as substrates for in vitro investigation of the cellular factors that are involved in control. Several factors that are involved in the expression by eukaryotes of low molecular weight RNA's (class III genes) are described in the article by Lassar *et al.* Recently a new type of site, termed "enhancer," has been described which may turn out to be involved in tissue-specific and developmental regulation of eukaryotic messenger RNA expression—the Holy Grail of molecular biology. The paper by Rosenthal *et al.* represents a recent contribution to this burgeoning area of research.

The analysis of nucleic acid structure by hybridization of labeled probes to various types of "blots" has provided a global view that has revealed a remarkable degree of plasticity in both prokaryotic and eukaryotic genomes. The ability of DNA to rearrange is apparently mediated in many cases by specific transposable elements that have evolved the capacity to jump from one point to another in DNA, and at the same time carry neighboring sequences with them. The mechanism by which this takes place in the specific example of the bacterial transposon Tn*903* is the topic of the contribution by Weinert *et al.* DNA rearrangements also take place in eukaryotes. In some instances these are a part of a normal developmental process (for example, gene assembly prior to the expression of antibody genes) but rearrangements may also be involved in neoplastic transformation. The papers by Leder *et al.* and by Land *et al.* focus on

these phenomena whose understanding will be a necessary prerequisite to the conquest of cancer.

Recombinant DNA strategies have been of extraordinary value in the identification and analysis of proteins. In the past few years, more protein sequences have been determined by predictions from DNA than were ever ascertained by direct protein sequence methods. Many proteins that are present in cells in such minute quantities that they would be difficult to identify (much less sequence) by conventional methods have become accessible to analysis. The ease with which genes coding for proteins can now be isolated is illustrated in the paper of Young and Davis, who developed the technique of antibody screening to isolate clones of a yeast RNA polymerase. Equally important is the capability to express such proteins in large quantities and to high levels of purity for biochemical analysis. The capability to modify precisely the genetic code through site-specific mutagenesis has also been of paramount importance. The study of mutations has always been a cornerstone of genetics but in the past it was necessary to rely on accidents of nature or the haphazard processs of mutagenesis to obtain them. Through strategies based on synthetic oligonucleotide chemistry it is now possible to generate mutations with absolute precision so that the effect of a single amino acid substitution can be evaluated. The paper by Villafranca *et al.* provides a beautiful example of the power of this method in the case of the dihydrofolate reductase gene. Much has been made in the general public press of the role of expression systems in providing medically important products such as insulin. (However, the predominant use of this technique is surely for the acquisition of knowledge of expression of proteins having little potential for commercial exploitation.)

The analysis of genes and proteins, their biochemistry and their regulation, is but a small part of the subject matter of biology. Issues of cell movement, cellular communication, organ formation, and organism behavior are, despite initial skepticism, rapidly becoming part of the frontier to which molecular biology is contributing. The use of monoclonal antibody probes to trace the growth and movement of individual identified leach neurons by McKay and co-workers points to the identification of surface proteins which mediate cell-to-cell interactions guiding the neurons to the cells to which they will make connections. The paper of Nirenberg *et al.* provides insight into the mechanism of synapse formation itself. McAllister *et al.* deal with a similar theme although in this case the probe method is in situ nucleic acid hybridization. This study demonstrates migration of cells in the marine snail *Aplysia*, which produces an egg-laying hormone. These cells start from an embryonic site in the ectoderm of the body wall and then move to specific locations in the nervous system where they presumably stimulate egg production by secretion of the hormone.

Finally, attention should be drawn to the papers by Palmiter *et al.* and by Caplan *et al.*, which describe the introduction of genetic material into animals and plants, respectively. Methodology of this sort is necessary for the acid test of how a particular gene functions in its normal state (that is, integrated at the gene's normal chromosomal locus). What has been accomplished will need to be refined since current methods do not achieve correct positioning or gene copy number with any reliability. Even so, dramatic results have been obtained in some cases, such as the production of larger than normal mice.

The cross-section of basic research presented in this issue will, I hope, put into perspective the tremendous number of basic research applications that have been opened in all fields of biology, not just molecular biology, by the new technical developments. There is a special need to apply them with vigor in a whole range of fields by scientists who are not trained in molecular biology. In general, molecular biologists have been very good citizens in making techniques and materials available, and the contributions of companies that make restriction enzymes and other tools of quality available commercially should make it possible for more and more scientists to apply these methods in their respective fields. The techniques are basically simple and very well documented. I see no need for the molecular biologists themselves to take over, as some have feared. (Although as a practitioner I have been continually amazed at how expensive it can be to operate a laboratory engaged in gene cloning.) This is to underscore the need to preserve and enhance the money allocated to investigator-initiated basic research from all sources despite the temptation to aim for rapid reduction to practical applications.

To end on a philosophical point, the relative ease of DNA sequencing has had a quantitative effect on the way information is acquired in molecular biology and this had led to a qualitative difference in the way in which it is and will be used. At present the worldwide accomplishment in DNA sequence amounts to 2.3×10^6 base pairs, representing 2500 individual sequences. It is now becoming more or less routine to sequence completely the DNA of whole genetic entities ranging from single genes through multigene families to simple life forms such as viruses and phages. Currently the largest single DNA molecule to have been sequenced is the phage lambda genome (48502 base pairs). We are beginning to recognize that determination of the total genetic specification of more advanced life forms may be a possibility in the relatively near future. Extension of this principle to bacteria (genome size 5×10^6 base pairs)—the simplest free living forms—would require an increase of the worldwide technical effort by only a factor of 2. Some three orders of magnitude more would be needed to progress to the total human genome.

Acquisition of "total" genetic information about a life form adds a new philosophical dimension. In some sense the rest of biology becomes merely a matter of interpretation. One might wonder, given the total DNA sequence of *Escherichia coli* and a big enough computer, could one reconstruct what the organism looks like, what it lives on, what it could be poisoned by and how it behaves? Or will a biological "uncertainty principle" be discovered that would preclude such a development?

Part of our advance depends on whether the information obtained by specific experiments on function can provide general principles. By study of a finite sample of promoters will we be able to derive a rule to tell from a sequence whether it is a promoter and what its characteristics would be. Progress in this area is quite likely and could be extended in principle to all "sites." Progress will also be needed in prediction of secondary and tertiary structures of biological molecules including the proteins and their complexes with substrates. These structural problems have so far seemed beyond the capacity of most computing facilities. However, computer chess programs have already reached the level below the grand master. Perhaps the solution to the protein-folding problem is nearer than we think. Regardless of the particular outcomes of these endeavors, there will surely come some major changes in our perception of ourselves and our place in the universe.