ADVENTURES WITH GENES

Sydney Brenner

Genetics is a uniquely fascinating biological science because it is the study of the plan of the organism. Although for many centuries, men had known that like begets like, it was Mendel who first began the science with his experiments on garden peas. He crossed tall with dwarf peas and found that all of their progeny were tall and not half-way in between. Then, when he took these F1 hybrids and crossed them with each other, dwarf plants reappeared in the F2 generation in the ratio of one dwarf to three tall. His explanation of these observations was remarkable. He postulated that the external *character*, the height of the plant, is determined by internal *factors*. Organisms have two such factors for each character, one inherited from each parent. Variant factors exist and when the factor *T*, specifying tallness, is paired with the variant, *t*, specifying dwarfness - or rather lack of tallness - in the hybrid F1, the plants are tall because, as Mendel postulated, *t* was recessive to *T*. In the next generation the factors segregate and are received in all combinations to produce the ratio 1TT:2Tt:1tt. The first two classes are tall and this gives the 3:1 ratio of tall to dwarf plants. Later Mendel showed that two different pairs of factors segregated independently to give ratios of 9:3:3:1 and he thought that all factors would do so, but this was later shown not to be the case by Morgan. Mendel's work lay unheeded for more than 35 years, while all kinds of theories of inheritance were proposed and debated in Europe. It was the rediscovery of Mendel's work and its independent confirmation by Correns, Tschermak, and de Vries in 1900 that ushered in the modern era of genetics. New experimental organisms were introduced, most particularly, the fruitfly *Drosophila*. Mendelian factors become genes, and Morgan showed that there were regular exceptions to the independent segregation of different genes and that some pairs were linked more than others. Their degree of separation was due to recombination, the measured frequency of such recombinants was proportional to their distance apart and such distances were additive for different pairs. Thus, he was able to construct genetic maps of organisms and the linkage groups detected in this way were shown to be equivalent to the visible chromosomes. Later,

when the large-banded salivary chromosomes of *Drosophila* were discovered, the order of the mutants on the genetic map was shown to correspond to the order of their locations on the physical map. The recessive nature of most mutations allowed genetic complementation tests to be performed to decide whether or not the same gene was defective in two different mutants. A hybrid was formed; if the organism was wild type (normal), two different genes were involved, but if it was mutant, the defects were in the same gene. Until Müller discovered the induction of mutation by ionising radiation, most mutants were of spontaneous origin, and induced mutants considerably enlarged the repertoire of genetic variation available to the experimentalist.

Thus, by the 1930s the classical concept of the gene had been formulated. It was at once a singular unit of mutation, function, and recombination. The “beads on a string” model captured this in a simple pictorial form; each bead was a gene with a function, mutation altered the shape of the bead, and recombination occurred between the beads. But what could the function of genes be? Because the genes had to be replicated in each cell division, Ilaldane had as early as 1929 formulated the autosynthetic catalyst model, in which he thought that the gene not only had to exert some catalytic function in the cell, that is, was an enzyme, but it was a special enzyme, because it could also catalyse its own synthesis. The single advance was made by Beadle and Tatum who turned to the study of biochemical mutants of *Neurospora*. They were able to obtain nutritional mutants, and by 1945 Beadle formulated the hypothesis that each gene exerted its function by specifying the structure of an enzyme. It became known as the one gene—one enzyme hypothesis and one of the foundations of modern molecular genetics. There was much scepticism about this generalization, because while it was easy to see how this could be true for biosynthetic pathways, most geneticists worked on much more complex morphological and developmental mutants and could not readily accept the idea that elaborate biological structures and functions could be produced by enzyme action.

The advent of biochemical genetics and the introduction of bacteria, particularly *Escherichia coli*, as experimental organisms, gave rise to an explosive development of knowledge of biosynthetic pathways and their genetic specification. Lederberg's discovery of sex in *E. coli* allowed genetic complementation and genetic mapping. For example, mutants requiring a single amino acid could be easily isolated.

Complementation could show which mutations were in the same or in different genes and the genes could be mapped. Each gene corresponded to a specific enzyme catalysing one step in the amino acid biosynthetic pathway. This is the powerful method of genetic dissection: to study biosynthetic pathways, all a geneticist had to do was to isolate mutants with requirements for added nutrients, classify the genes for each phenotypic class by genetic complementation, and put them on a genetic map. This could then be coupled with biochemical assays, for which the mutant bacteria themselves provided a valuable enhancement because they not only accumulated precursors before the missing step, but also allowed the purification of enzymes by *in vitro* complementation. With simple methods to find mutants and to move genes from one bacterium to another, the geneticist, simply by observing whether or not bacterial colonies could grow under specified conditions, was able to reach deep and remarkable conclusions about the general physiological structure of an organism and, at the same time, open the way to novel experimental material, not previously available.

At roughly the same time, bacteriophage research developed as an experimental system for solving the problem of self-replication. A virus particle enters a bacterium and 20 to 30 minutes later, the bacterium lyses and releases several hundred new virus particles. The virus particles contained DNA but not many people believed at the time that all the genetic information was contained in DNA. Indeed the earlier work of Avery and McCarty on pneumococcal transformation, in which bacteria were genetically altered by DNA coming from a different striae was also viewed with some scepticism, partly because it was just beyond the conceptual horizon of contemporary scientists to view this as a genetic cross between a naked gene and cell, and partly because everybody was still gripped by the notion that genes had to contain protein, an inheritance of the old idea of the gene as an autocatalytic enzyme. In 1952, Hershey and Chase did an experiment with bacteriophage which showed that after infection most of the protein of the virus remained on the outside of the cell, while most of the DNA entered. Today, this paper would not survive the stringent rules of entry into the pages of many journals, because it by no means excluded *all* of the protein, but it had great psychological impact in reinforcing the pneumococcal research and gave stronger credence to the idea that genetic information was lodged in DNA.

The discovery of the structure of DNA by Watson and Crick in 1953 ushered

in the modern era of molecular biology. It provided at one blow a basis for understanding how genes were replicated and how their functions might be expressed by specifying the structure of their corresponding proteins. It may be hard to believe now, but it was only when Sanger completed his determination of the amino acid sequence of insulin in 1952 that chemists accepted that proteins were not random polymers and had a *defined* chemical structure in the form of this sequence. The connection between DNA and its final protein product was stated in what Francis Crick called the sequence hypothesis, namely, that the one-dimensional nucleotide sequence corresponded in some way to the one-dimensional sequence of amino acids in the polypeptide chain and that this in turn determined the three-dimensional folding of the protein. The correspondence, of course, was the genetic code.

One final important contribution to the new molecular genetics was Benzer's fine-structure genetic map of the rII locus of bacteriophage T4. This work demolished the classical singularity of the gene as a unit of mutation, function, and recombination. Other geneticists, such as Green and Pontecorvo, had already found what they called pseudoalleles. These were mutations which were clearly in the same gene as defined by functional complementation tests, but which could still be separated by recombination. Benzer took this to the ruthless limit, by showing that each of the two functionally different genes of the rII locus in bacteriophage T4 had literally hundreds of different mutations, all separable by recombination. He showed that the scale of the fine structure map was consistent with the molecular scale of the DNA structure and that each of the mutations could correspond to a single base pair.

When I met Benzer in the summer of 1954 in Cold Spring Harbor and heard about the beginnings of this work, I realized that this might provide an empirical way to study the genetic code. At that time, a biochemical approach seemed out of the question and very far in the future because it would need the reconstitution, *in vitro*, of what was bound to be a complicated system. We thought it would take a long time to open the black box of biochemistry and that we would need to find a way forward without it. At the time, there were some who thought they could find the answer without doing any work at all, and that the code could be deduced theoretically. Many of the special codes proposed could be proved to be inconsistent, and in 1955 I was able to show that even

the small number of amino acid sequences available at the time, when used correctly, could eliminate all overlapping, degenerate triplet codes, given that the code was universal. Theoretical work on the code had the historically useful function of clarifying thoughts on the issue and I learned from my work here that it is not only important for theories to be correct, they should also be capable of being true. That is, you should always be able to provide a biologically plausible example for the physical implementation of any abstract model.

Thus, studying the correlation between the genetic fine structure and the chemical sequence of its protein became the only practical way we could see of analyzing the genetic code. The gene went in one end of the black box, the protein came out of the other; we were going to study the rules of this transformation by comparing the two.

I launched an experimental program to find a gene-protein pair that could be used to study how information was transferred from gene to protein. It had to be in a bacteriophage because fine structure genetics could be easily carried out and it should be for a protein of the virus coat, because at least that was accessible and might be more easily purified. George Streisinger began to work on host range genetics in T2 and, on my return to South Africa in 1955, I began work on tryptophan cofactor genetics of the related bacteriophage T4. Our excitement in this field increased when Ingram invented the protein fingerprinting technique and was able to show that the sickle cell mutation caused the change of a single amino acid in haemoglobin. This was exactly the method we needed, because to determine the complete amino sequence of even a small protein was a formidable task in those days, and to have to do it for a wild type and several mutants was out of the question. With fingerprinting, we would only have to look at the short peptides that were different. During this early period, one particular problem came to be formulated that could be solved by this approach—this was the question of co-linearity. Although no one believed otherwise, it had not been proved that the two sequences were co-linear, that is, that the mutations in DNA could be put in a one-to-one linear correspondence with the amino acid changes in the polypeptide chain. We could imagine all kinds of mappings and so, when I came to Cambridge at the beginning of 1957 to begin a molecular genetics group with Francis Crick, co-linearity became the first target of our research. Later that year, we were reinforced by George

Streisinger and Seymour Benzer and his student, Sewell Champe. We set ourselves the initial task of purifying tail fibers from bacteriophage T2. Our approach was to dissociate phage particles into various structural components and then separate these by simple means; most protein chemists thought this was nonsense, and told us that we should first solubilize all of the proteins and then separate the different ones by standard methods. Fortunately, we did not listen to them. In this work, we opened up the whole area of the structure of complex viruses and as a by-product, together with Robert Horne, I introduced the negative staining method as a new and powerful method for the electron microscopy of biological structures. Soon the tail fiber project got into difficulties. Tail fibers composed a very small fraction of the total phage, and they appeared to have an alarmingly large molecular weight. We turned to work on the major components of the particles such as the tail sheaths and the head protein, but for these we had no defined genes. In the meantime, other groups started to work on the same problem with different pairs of genes and proteins, mostly in bacteria. Our competitors seemed to be doing well, whereas we had excellent genes with no proteins, and excellent proteins with no genes. The head protein is the major component of the bacteriophage particle, and a fingerprint of total phage shows essentially only the fingerprint pattern of this protein. Thus, the work of isolating the protein could be trivialised to the purification of phage by a simple centrifugation step, and I therefore set out to find a genetic system for the head protein of bacteriophage T4. I used the property of osmotic shock, and found ways of selecting for and isolating different osmotic shock-resistant mutants and, what was more important, ways of assaying for the shock-sensitive recombinants. A fine structure map was constructed, and I tried to prove that the osmotic shock mutants were in the same gene as that specifying the head protein by looking at the genetic linkage of fingerprint differences between T2 and T4 heads to the gene containing the shock mutants. At one time, I thought I had mapped the osmotic shock mutants right into the head protein gene, but searches for fingerprint differences in the mutants proved fruitless. Later it was shown that osmotic shock was localized in gene 24, which is just next to 23, the gene specifying the head protein, but being almost right is not much better than being totally wrong. We later used the head protein to prove co-linearity by an entirely novel method. Anand Sarabhai discovered that the *amber* class of suppressible mutants were chain-terminating; such mutants did

not make viable phage under nonpermissive conditions, but since the head protein represents the major component of all proteins synthesized late in phage infection, we could easily see the fingerprint of the head protein in a digest of the complete cell without any purification at all. Each of the chain-terminated proteins was shown to have a different pattern of peptide components, such that longer chains contained all of the peptides of shorter ones, together with new peptides defining the additional segments. This simple topological argument establishes the order of the segments, and all we had to do was to show that this was the same as the linear order of the mutations on the fine structure genetic map. We never had to determine the total sequence of the protein at all and, indeed, such ladders form the basis of many ordering techniques in modern-day sequencing.

We were also interested in mutagenesis. Seymour Benzer and Ernst Freese had worked on chemically induced mutations in the rII locus, using analogs of the normal bases, which are incorporated into DNA and had been shown to be mutagenic. The idea was to try to find very specific mutagens which could be uniquely assigned to one particular base pair substitution, for example, G-C→A-T. Then, if we had a gene-protein pair, not only could we use it for establishing co-linearity but, by looking at the amino acid changes produced by specific mutagens, we might even be able to work out the genetic code. As will be seen later, this was only to be partially realized in our work on the assignments of the nonsense triplets, but the tobacco virus protein mutants later provided a useful source of supporting information in the determination of the genetic code. As a sideline, Seymour Benzer, Leslie Barnett and I decided to look at the spectrum of mutants induced by proflavine, an acrideine dye, and one in which I had a long interest, going back to work that I had done a decade earlier on supravital standing. The remarkable result was that not a single acridine mutation mapped to the same place as mutants in the previously determined spectra of base analog mutants and, more importantly, as we discovered later, no base analog mutant could be reverted by acridines. However, each set reverted spontaneously and showed enhanced reversion by the same mutagen that induced it. This was paradoxical. At first sight, one expectation of this work on mutagenesis would have been to have found two mutagens, M1 and M2, such that mutants induced with M1, were not reverted by M1, but by M2, and *vice versa*. That would give us two very specific agents, one producing G-C→A-T and, the other

A-T→G-C, for example. But this is an over-simplification because, as we realized, it would be wrong to expect such pure results; there would always be exceptions produced by mutations at sites different from the initial mutation, either in the same triplet or even at a remote site producing a compensating alteration in the protein. Thus, when Freese suggested that base analog produced transitions (pu-py→pu-py) and proflavine produced transversions, (pu-py→py-pu), we knew that this theory could not be right, since exceptions would be expected in both cases. The hard experimental fact was that the two classes of mutants were totally disjointed. Then, in a conversation with Francis Crick, I had the idea that acridines might cause a special class of mutations such as short, perhaps single, base additions and deletions. This would explain how proflavin could revert acridine-induced mutations and, at the same time, why it did not induce reversions of base analog mutations. Arguing that such additions and deletions would have drastic effects on gene function, we quickly supported this notion by showing that while both bromouracil and acridine would produce mutations in the rII gene, whose products were dispensable, only bromouracil could induce mutations in the *h* gene which had an *essential* protein product.

We had already shown that revertants of acridine mutants occurred at a site distant from the original mutant. These suppressors therefore compensated the effects of the first, so that if the first was assigned the sign (＋), all the suppressors would then be (−). This is the notion of *phase* in the reading of the genetic message, and we realized that we could use this purely *genetic* method to analyze the general nature of the genetic code. Thus, imagine that the (＋) mutation introduces an extra base into the message, this would throw it out of phase completely. For example, if the message was read in groups of three letters:

CATCATCATCATCAT……

putting in an extra base would advance the phase:

CAATCATCATCATCATCA……
　＋

This could then be compensated for by a (−) mutant to give

CAATCATCATATCATCAT……
　　－

and restore the correct phase of reading after the change. The distance apart of the two mutations would depend on how much of an altered sequence the protein could tolerate. This had two consequences which we could test. If the code was a triplet code, there should be only two classes of mutants, one of sign (−) and one of sign (＋). Furthermore, putting two mutants of the same sign together should still be mutant, but adding a third of the same sign should restore the phase. We were able to prove this. Thus, we showed that the suppressors of the suppressors of a mutant had the same sign as the original mutant and that adding three mutants of the same sign was in fact a wild type. It was fortunate but deliberate that we chose to work in the first part of the rII B cistron where the exact sequence of the protein is not critical, and that Francis Crick correctly realized the importance and significance of the so-called barriers, many of which turned out later to be chain-terminating mutants produced in the phase shifted sequence. This work proved that the genetic message was read in groups of three, that is, it was a non-overlapping triplet code, and since we had viable phase-shifts over a considerable length of the gene, it also showed that the code was highly degenerate with very little nonsense and with the majority of triplets corresponding to amino acids. There were, of course, some anomalies, which we could not understand, but, which at that time we resolved to set firmly aside. Several years of additional work were required to disentangle these, but what is interesting is that each anomaly had a special and different explanation. Thus, there were mutants which were clearly suppressed, but from which only the original mutant could be isolated; these turned out to be duplications in which the original mutant of sign (＋) was compensated by a (＋) at the duplication junction, the third (＋) being provided by the second copy of the first mutant. A unique exception to the expectation that no base analog mutation of sign (0) should compensate an acridine mutant allowed Anand Sarabhai to discover novel reinitiation sites in the gene that corrected the phase of reading of the message.

It is hard to convey the almost miraculous nature of this genetic research. The experiments were very simple and required only agar plates, a few tubes and paper-strips. Hundreds of experiments could be done at the same time and the results were available in half a day or so. The observations were simply to score whether or not growth of the phage occurred—only two numbers, 0 and 1. From such simple

observations made at the level of populations of bacteriophage, powerful deductions could be made about the detailed molecular structure and workings of a living system. The black box of biochemistry did not have to be opened. However, paradise does not come cheaply and the most important step in this work was to have understood the correct difference between the base analog and acridine mutations and also to have had a clear notion of their physical nature. Without this, all the experiments would have degenerated into a gigantic set of incomprehensible facts.

The black box had to be opened. Many things happened in close succession. Zamecnik and others had developed *in vitro* biochemical systems for protein synthesis; Hoagland had discovered the activation of amino acids; and Zamecnik had discovered transfer RNA. Crick's adaptor hypothesis of protein synthesis was totally vindicated and given physical realization. We knew that protein synthesis occurred on ribosomes but the nature of the information intermediate between the gene and the protein was unknown and surmised by all to be the ribosomal RNA itself. In their experiments in Paris, Jacob and Monod had come to the conclusion that either there was a relatively short-lived intermediate or that perhaps that some proteins were made directly on DNA. Another possibility was that a small number of special ribosomes could be made which were capable of prodigious protein synthesis. It was during a historic discussion in my rooms at King's College Cambridge in April 1960 that I suddenly realized that the small amount of new RNA made after bacteriophage infection, discovered by Volkin and Astrachan, had the properties expected of a genetic intermediate. In particular, its apparent base composition strongly resembled that of the infecting DNA, which had a higher A＋T/G＋C ratio than that of the DNA of its host, *Escherichia coli*, and was also very different from that of the ribosomal RNA of *E. coli*. I should remind you of one of the deep paradoxes of the time, which was that while the A＋T/G＋C ratio of DNA of different organisms varied over an enormous range, that of the ribosomal RNA was constant and this was certainly not to be expected of a copy of the genetic material. Volkin and Astrachan had also shown that this RNA was unstable and turned over rapidly which would be essential for explaining the results of enzyme induction experiments, but only important for us in that it could explain how the phage messenger could displace all of bacterial protein synthesis quickly after infection. The base composition was *the* telling point. Now Volkin and Astrachan's explanation of their

results was that this RNA was a *precursor* to phage DNA, while, later, Spiegelman concluded that it was evidence for a small number of new ribosomal particles making phage protein after infection. As is now well known, Francois Jacob and I did the critical experiments in Pasadena in the summer of 1960, and we found totally convincing evidence for messenger RNA. We also did a large number of control experiments to eliminate other explanations, many of which sound totally farfetched today.

Once it was realized that messages were not part of ribosomes, but could be added to them, the genetic code became amenable to direct biochemical study. The work of Nirenberg, Ochoa, and Khorana allowed the direct assignment of triplets to amino acids, first by binding to random homopolymers and later by the use of specifically synthesized sequences. Eventually, the code was determined directly and only in the case of the nonsense chain-terminating mutants was there the last flicker of the genetic approach. This was based on the unique mutagenic specificity of hydroxylamine which produces the transitions, G-C→A-T, almost to the exclusion of other changes, by modifying C to C', say, which then behaves like T. Now only one strand of DNA carries the message; it is the strand that will be copied into RNA. If this has an altered sequence, the messenger will be mutant and will not function. If the other strand contains the alteration, the messenger made off the normal strand will be normal, and a non-functional messenger will appear only when the defect is copied into that strand by DNA replication. One characterized class of chain-terminating mutants, *amber* mutants, were known to be induced by hydroxylamine, and our experiment then was to compare the spectra of induced *amber* mutants of the rII gene generated under different conditions. In the first, the mutagenized phage was passed through a strain where the function was not required, so that all mutations were recovered; in the second, the phage was passed through a strain in which the expression of the rII gene was needed for growth, so that modifications in the strand copied into the messenger would be immediately expressed and *amber* mutants produced in that configuration would be lost. Thousands of rII mutants were analyzed to obtain many recurrences of ambers at the same site. The result was most satisfying. There were several *amber* sites which were missing in the set which required function and these, therefore, had to be due to G→A changes in the messenger. Thus, the *amber* triplet had one A. However, because

only some of the sites are absent under the excluding conditions, *amber* mutants could also be produced by C→U changes in the messenger and, therefore, the triplet also had to have a U, and its composition was therefore (UA). In our study of the chain-terminating mutants of the head protein we had determined the amino acids connected to ambers by mutation, as had Alan Garen in his work on the phosphatase gene. For example, ambers reverted in one step to tyrosine, but this was not induced by base analog mutagens and was, therefore, a transversion. We also found that another class of chain-terminating mutants, *ochre* mutants, could be converted to *amber* mutants and that this was enhanced by base analog mutations, but not by hydroxylamine. Thus, it was likely that ambers were (UAG) or (UAC), ochres, (UAA) or (UAU) and tyrosine (UApy) or (UApu). The first of these alternations proved to be the correct one with the bases in the order shown.

By then, the field had become very strongly based on biochemistry, and rightly so. And although one could lament the passing of a unique style of research, there was now no need to use elegant cunning when everything could be done directly by chemistry. We did use genetics in our studies of the suppressor transfer RNAs, but we had learned the important lesson that powerful means of analyzing mutant phenotypes considerably enhances genetic analysis. In that case, we used a genetic trick to amplify the suppressor gene and its product—today it would be called cloning—so we could apply the new methods of RNA sequencing developed by Sanger. I also did some work with Francois Jacob on DNA replication; we created the theory of the replicon and started to study conditional mutants of DNA replication.

But already my mind had turned to other things and in the early 1960s I had begun to look seriously at studying new areas of biology, using the same combination of genetics and phenotypic analysis, and in this way initiated the nematode project. Elsewhere, I have briefly described how my work on *C.elegans* began and what I had set out to do with it. *Caenorohabditis elegans* is a small free-living nematode, easy to grow and handle in the laboratory, with a short life cycle and a small number of cells. It has a nervous system of some 300 cells. The idea was to discover how a complex organism was produced from its genetic program, by isolating mutants that altered its development or behavior and then studying these as deeply as we could. I isolated a few

thousand mutants of the worm and classified and mapped them. I began a project to reconstruct the entire anatomy of the worm by serial section electronmicroscopy. Not many people took this work seriously. My more polite critics thought it was too long-term and that I was 20 years ahead of my time. But, of course, somebody has to be 20 years ahead of the time, so that others do not stay 20 years behind. The successful succeed, and nobody knows about those who do not.

Others joined me in the nematode work and we were able to start other lines of work, particularly on the molecular biology of muscle, to which field the worm has made a distinctive contribution. John Sulston took over work begun by a talented student and determined the complete cell lineage of the worm. How far could we have got by genetics alone? It seems that, at least for some systems, the broad scheme of the genetic logic can be worked out by these methods as has been done for lineage mutants and for sex determination in the nematode. But what has changed everything for everybody is the development of the new genetics. From the mid-1970s two technical developments, DNA cloning and sequencing, have pervaded the entire field of biological research. We very early saw that our work on the genetics of muscle in the nematode could be really carried down to the molecular level if we could clone the genes and determine their structure by sequencing. What was satisfying is that when John Karn and Sandy MacLeod cloned the nematode myosin gene, we actually used the mutants to help us find the gene and to prove that the right one had been found. Although, initially, there was some resistance to this approach, the new methods of molecular genetics have pervaded all fields of biological research, and the nematode project has grown into a large international effort by many people.

Working on the nematode was a very different experience from working on phage. In phage genetics, ideas could be very rapidly converted into experimental practice, and so ingenuity in design of both models and experiments became highly prized. It was positively dazzling and made all other research seem pedestrian. It was just the subject for a young scientist. On the other hand, the nematode was a weighty project, requiring many long-term strategic moves (I do not call them decisions), and with fewer immediate rewards. Although it was a very risky project, it nonetheless had a much more solid air about it and fit my middle age. And what became prized were the

qualities of persistence and endurance, although every now and then a clever genetic experiment could be carried out.

And so to the final chapter. The new genetics is creating a revolution in biology. It is a deep and profound revolution. In classical experimental genetics we could only define a wild type gene by finding a mutant allele, and we were totally dependent on breeding experiments to classify and map genes. The whole of experimental research in biology was then devoted to the study of the functions of gene products. Today, all is different; we identify genes by cloning them and we characterize them by sequencing them. We have become liberated from the life cycles of organisms, and we no longer have to use the organism as an instrument to read its genetic sequence; we have direct access to it. We can study any organism for which we have DNA, and genetic analysis is now possible for everything, including Man. The main reason for what has become known as the Human Genome Project is that it can now be done, and done more efficiently if we all work together on it.

But what is most important about this new work is that it will change the paradigm of biological explanation. Our explanations have depended very much on mechanisms; while we are very good at finding out how these work, we have difficulties in analyzing complex assemblages of them and understanding their organization. This is especially true for biological systems which have evolved by natural selection, and are loaded with opportunistic accretion, unlike a man-built system where careful study might ultimately reveal the logical mind of the designer. Studying organisms from their genomes, from within, so to speak, produces a new point of view. We can ask ourselves whether it is possible to compute organisms from their genomes. I use the term compute in its broadest possible sense, in the sense that I view an organism as a system of natural computation, with its genome going in at one end and an organism coming out at the other. Achieving this would be equivalent to having a complete understanding, but it will be done not by some grand equation, but by embedding everything in existing biological knowledge. And this must be the manifesto of genetics, if not its program.

* * * * * *

I thought the other day that genetics had become like astronomy and that we are now engaged in exciting voyages of discovery rather than in the practice of everyday experimental science. The genetic heavens are still almost totally uncharted. In astronomy, when one looks at objects very far away, the light we see has taken a long time to reach us and so we can look far back in time and the astronomical telescope becomes an astronomical chronoscope. I now spend a lot of time looking at sequences; streams of symbols, many with meaning for the present structure and function of organisms, others probably with none and perhaps relics of a long-ago past. If we study these, we might be looking back at the genes of organisms of a very long time ago, and so our genetic microscope could become a genetic chronoscope. It would be fitting that the genomes of organisms should contain not only the library of genes that they use now but also a voluminous archive documenting their past, and that both could be discovered and understood by the study of the gene.