Fred Sanger was awarded the rare distinction of two Nobel Prizes for Chemistry, in 1958 and 1980. The first was for his work on the structure of proteins, particularly that of insulin, the latter for his contribution concerning the determination of nucleic acid sequences, the foundation work that ultimately led to The Human Genome Project..

The early days of DNA sequences

FRED SANGER

To an inveterate sequencer like myself, the recent achievements in the Human Genome Project are particularly exciting. Ever since the 1940s, when I discovered the first tetra-peptide sequences at the ends of the insulin chains, I have been hooked and have worked with sequences on and off until my retirement in 1982. During the course of that work, it became increasingly clear that sequences had an essential role in living systems. The activity of proteins depended on amino acid sequences and, as more was discovered about DNA, we realized that all the necessary information in living matter was contained within the nucleotide sequence. If we could determine this we would learn much more about the workings of living matter and hopefully contribute a more rational approach to medicine.


Fred Sanger

During the 1960s we gradually turned our attention from proteins to nucleic acids. In those early days, no substantial work had been done on DNA sequencing. Wu and Kaiser[1] published the first successful DNA sequence detailing the 'sticky ends' of lambda in 1968. Although the method was tedious and only applicable to the few special residues on this bacteriophage, this was an important step. Curiously enough, their copying method employed DNA polymerase, the enzyme used in determining the human genome, though under very different conditions.

When we started working on DNA I don't believe we were thinking about sequencing the entire human genome—perhaps in our wildest dreams but certainly not within the next 30 years. After all, the human genome was probably the ultimate goal of all our work. We were finding interesting things as we progressed and—at least for me—there was a lot of fun in overcoming some of the difficulties that arose.

It was surprising to me that there were so few other people working in the field as whole-heartedly as me, and this was probably due to the way science is supported. Then, as now, grants are typically given to projects that are expected to give tangible results within a few years and it was difficult to get support for long-term projects. We were lucky that the British Medical Research Council was willing to adopt a long-term perspective.

When we started working on nucleic acids we used the general approach we had used on proteins: the 'jigsaw puzzle' method. We broke down large molecules to smaller bits that could then be analyzed and fitted together. The work was slow by today's standards, but didn't seem so at the time.

Whereas the 20 amino acids have different properties, nucleic acids comprise four rather similar components. It gradually became clear that we could take advantage of this simpler composition to develop techniques for identifying sequences based on the behavior of fragments upon fractionation and use DNA polymerase in a copying procedure. Progress was slow, but eventually Alan Coulson and I developed the 'dideoxy' method, which is still the basic method used.

Throughout all our sequencing work, fractionation of large breakdown products has been a major technical obstacle. At first, nucleotides presented severe difficulties and would not migrate as clearly separated products on the systems used. We achieved a major breakthrough when we tried using ionophoresis on denaturing acrylamide gels. Rather unexpectedly, the fragments migrated as sharp bands and separated almost exactly according to size. We could, for instance, separate a fragment of 50 residues from one of 51 residues, something we had never done before. Nowadays an improved version of this system is an integral component of the dideoxy method and fragments as large as 1000 residues can be differentiated by a single additional base. I have spent much of my time working with fractionation systems, often with disappointing results—smears or streaks instead of well-defined bands or spots. It is still amazing to me that such a faithful system is possible.

Another problem with the dideoxy method as we first developed it was that it only worked on single-stranded DNA, whereas most DNA is, of course, double-stranded. This difficulty was overcome by the ingenious work of Gronenborn & Messing[2] who prepared a ssDNA bacteriophage vector in which fragments of DNA could be cloned and sequenced. Not only did this make it possible to work with dsDNA, it also solved the problem of fractionating the fragments. This was done simply by isolating pure clones, which is probably the ideal fractionation procedure.

Initial studies on DNA were limited by the absence of any small DNA templates on which to develop new methods. The smallest ones were those of the ssDNA bacteriophages and most of our early studies were carried out on one of these, Φx174, the sequence of which is 5386 nucleotides and was completed in 1978. A pay-off from this work was that we could deduce the amino acid sequence of the 10 proteins for which it was coding. So although we were no longer working on amino-acid sequencing, we had developed a more rapid method for doing it.

The next sequence that we tackled was that of the 15,569 residues of the human mitochondrion, which perhaps could be considered the start of the Human Genome Project. Most of the mitochondrion's components are encoded by the cellular DNA, but there is a small DNA inside the mitochondrion that was found to code for 13 proteins, 22 transfer RNAs and 2 ribosomal RNAs. Thus mitochondria can be considered to be individual organisms and are generally believed to have evolved from some primitive, free-living organism that lost much of its synthetic capacity when it became incorporated into the cells of higher organisms.

One of our most interesting and unexpected findings was that mitochondria have a slightly different genetic code from the norm, which had previously been thought to be universal. Whereas UGA is normally a termination codon, in mitochondria it codes for tryptophan.

I had envisaged that increasingly longer DNA templates would be studied, but I didn't anticipate that determination of the human genome would be initiated so soon. This has largely been made possible by progress in other fields, particularly robotics and

in data processing, and also, of course, by financial support.

It is a great source of joy to me that the dideoxy method is still the basic technique used. It was perhaps the climax of my career and makes me feel that all our previous studies on sequences with their successes and failures were not only enjoyable but also a worthwhile contribution to the future of medicine. Of course, I have not been alone in this work and I feel I owe a great deal to those with whom I have worked. I have been very fortunate to work in two first-class laboratories, both fairly young and occupied by enthusiastic scientists, who were interested not only in what they were doing but also in other people's work and keen to exchange ideas. It is easy to thrive in such an atmosphere. I may have had some good ideas but from where did I get them? Perhaps from some chance conversation that I have now completely forgotten.

I have never had a very large group working with me at any one time. I preferred to work at the bench myself rather than to plan experiments for others to do. I have worked largely with PhD students and post-doctoral fellows, starting with Rodney Porter, my first student. I have been lucky in having many excellent post-doctoral fellows. The progress of the work owed much to them. They were already experienced workers and often brought new ideas and expertise to the group. Without them we could not have achieved what we did. Bart Barrell and Alan Coulson are the two colleagues who have been with the group the longest. Both started out as technical assistants but are now in charge of their own groups.

It should be pointed out that the entire human genome is not yet finished, and it seems to me that it ought to be completed with maximum possible accuracy while the facilities and motivation are in place. The function of much of the sequence is not yet understood, much of it now termed 'junk' DNA, but it may well be that further research reveals some function and that will require the exact sequence.

It seems unfortunate that a certain amount of competition has developed in the field and I hope that this will not affect the standards of accuracy during the later stages of the work. We have always found that the final stages of any sequencing project are relatively slow and there may be a temptation to leave them to someone else.

One of the largest sequencing groups has decided not to make their results wholly available without certain restrictions, in hope of commercial gain. This is unfortunate when so many other groups have dedicated their results freely for the public good and the future of medicine. The genome is a vital part of us all, probably the most vital, and information about it should be freely available and must not become the private property of anyone, however rich. No other part of the human body has ever been committed to such selfish conditions and it seems completely immoral to try to place restrictions on such a vital part of humanity. What would happen if an expert on brain anatomy decided to place restrictions on the brain so that only he could work on it?

The human genome sequence is of no commercial value in itself. It must be kept freely available so that many scientists of differing expertise can work on it to advance medicine. This would then be the appropriate stage at which to make commercial gains through established patenting practices.

1. Wu, R. & Kaiser, A.D. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *J. Mol. Biol.* **35**, 523–527 (1968).
2. Gronenborn, B. & Messing, J. Methylation of single-stranded DNA *in vitro* introduces new restriction endonuclease cleavage sites. *Nature* **272,** 375–377 (1978).

Head of Protein & Nucleic Acid Chemistry
Medical Research Council Laboratory of Molecular Biology, Hills Rd. Cambridge

Society has entered uncharted territory regarding how, when and where genetic information can be used. This article discusses the major issues raised by increased access to genomic information, which will ultimately be resolved by legislation or the courts.

Legal Issues in Genomic Medicine

PHILIP R. REILLY

Advances or apparent advances in our understanding of human genetics have elicited societal responses for nearly a century. Many of these responses have been expressed as new legislation or new case law in the United States. During the first half of the twentieth century, the enactment in more than 30 states of involuntary sterilization laws targeting persons residing in institutions for the mentally retarded because of the erroneous belief that most mental retardation was monogenic, is an example of a failed social policy that was based on a presumed genetic advance. In other circumstances, such as the enactment of mandatory state newborn screening laws in the 1960s to implement an inexpensive test to detect infants with phenylketonuria, or the expansion of newborn screening to identify children with sickle cell anemia in the 1980s, the social policies (though imperfect) have been clearly beneficial. Sometimes new genetic information has stimulated a public discourse far beyond what scientists might think makes sense. For example, reports in the late 1960s suggesting that men with XYY syndrome were found at high rates in mental penal institutions elicited much speculation about the impact of mutations that drive aberrant behavior on criminal jurisprudence.

In the late 1970s in the US, the field of prenatal screening was dramatically altered by judicial decisions. By 1980 a significant number of state supreme courts had held that a woman of 35 years or older who gave birth to a child with Down Syndrome (trisomy 21) could sue her physician if he had not informed her about the age-associated risk of bearing children with this disorder and that there was diagnostic test available (fetal karyotyping after amniocentesis). In 1987, the American College of Obstetricians and Gynecologists warned their members that a physician could be sued by a woman who gave birth to a child with a neural tube defect, if he failed to inform her in a timely fashion during her pregnancy about the available screening test[1].

The flow of genetic information into the practice of medicine, enabled by the completion of the Human Genome Project, has already stimulated the enactment of new laws aimed at preventing genetic discrimination. The impact of genetics, however, is much wider. It has begun to challenge thinking in regard to well-established medico-legal principles such as confidentiality and privacy, and raised profound questions about intellectual property. The burgeoning field of pharmacogenomics has created a new set of property interests, placing high values on tissue samples and DNA data. The advent of new kinds of genetic research has caused the US federal government to reconsider its policies on the oversight of research involving human subjects. Advances in stem-cell biol-