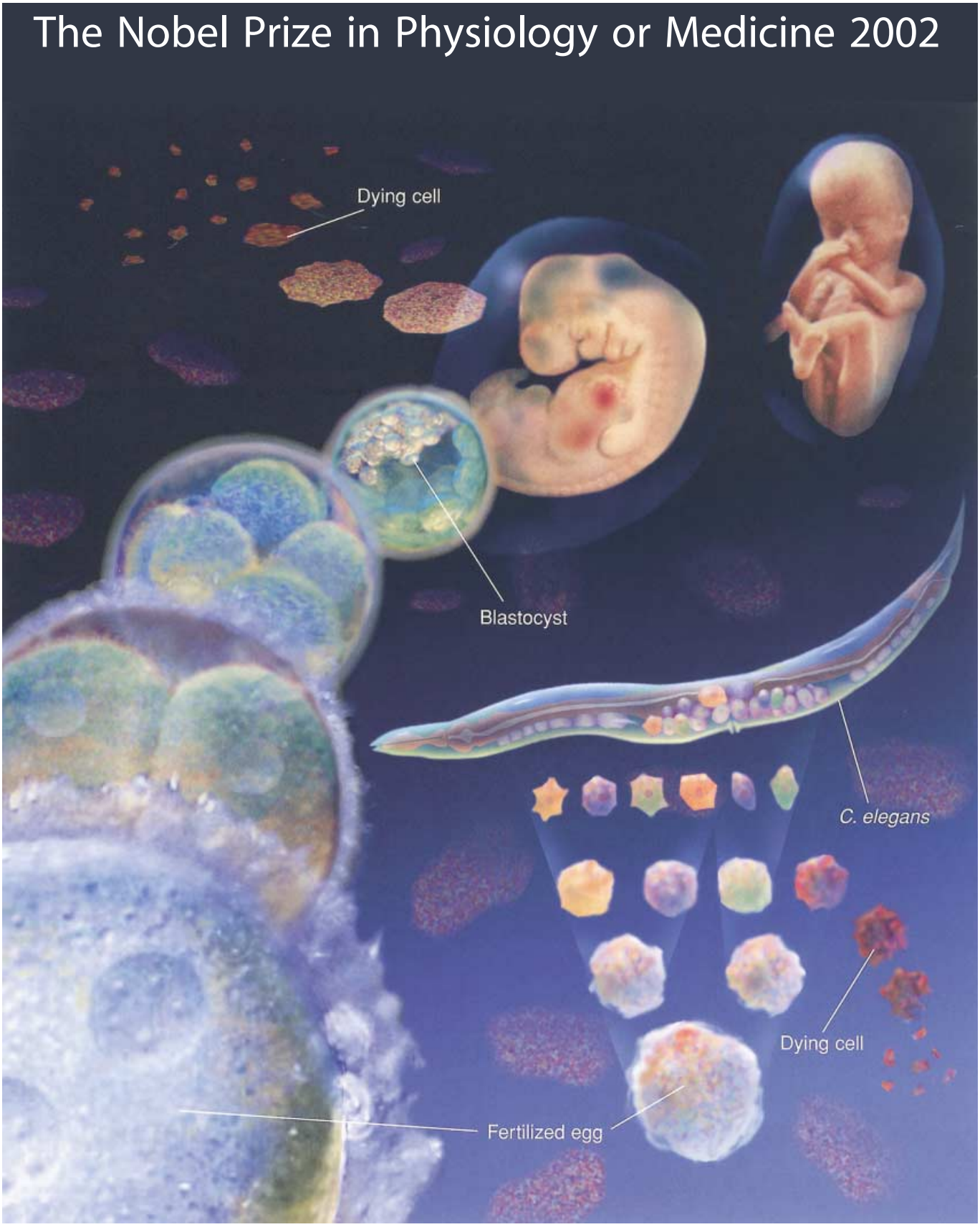
The Nobel Prize in Physiology or Medicine 2002



Nature's Gift to Science (Nobel Lecture)**

Sydney Brenner*[a, b]

KEYWORDS:

gene sequencing · gene technology · model organism · Nobel lecture

The title of my lecture is "Nature's gift to Science". It is not a lecture about one scientific journal paying respects to another, but about how the great diversity of the living world can both inspire and serve innovation in biological research. Current ideas of the uses of model organisms spring from the exemplars of the past and choosing the right organism for one's research is as important as finding the right problems to work on. In all of my research these two decisions have been closely intertwined. Without doubt, the fourth winner of the Nobel prize this year is *Caenohabditis elegans*; it deserves all of the honour but, of course, it will not be able to share the monetary award.

I intend to tell you a little about the early work on the nematode to put it into an intellectual perspective. It bridges, both in time and concept, the biology we practice today and the biology that was initiated some fifty years ago with the revolutionary discovery of the double-helical structure of DNA by Watson and Crick. My colleagues who follow will tell you more about the worm and also recount their incisive research on the cell lineage and on the genetic control of all death.

To begin with, I can do no better than to quote from the paper I published in 1974.[1] The paper was unhesitatingly entitled, "The Genetics of *Caenorhabditis elegans*", and the opening sentence reads: "How genes might specify the complex structures found in higher organisms is a major unsolved problem of biology." This is still true today. The paper outlined how a genetic approach coupled with detailed studies at the cellular level might be a way of studying this important question. It introduced *C. elegans* as the organism of choice for this work.

This choice had a long history. Twenty years earlier, we had posed a different question. Then, the central problem in biology was how the one-dimensional sequence of nucleotides in DNA specified the one-dimensional sequence of amino acids in proteins. Today, any student would give this question a very simple answer. 'All you have to do is to find a gene and have it sequenced and then make some protein using the gene and get someone to determine its amino acid sequence.' In those early days, the techniques for determining amino acid sequences of proteins were primitive and needed large amounts of proteins, which had to be purified first. There were no methods to isolate genes and no techniques for the chemical determination of their sequences. Our analysis of genes was limited to genetics. Indeed, the only way we could assert that there was a gene in an organism was by finding a mutant allele for it. Like Mendel, we could not say that there was a gene for the character of tallness until dwarf mutants were discovered suffering from a heritable lack of tallness. Genetic analysis of linkage used recombination to analyze the structure of chromosomes, to determine the locations of genes and their linear order along a chromosome. But in order to probe the structure of the gene something special was needed. That something was provided by bacteriophages, viruses with tiny genomes which grow on bacteria and which show recombination. In 1954, Seymour Benzer developed a system using the *rII* gene of bacteriophage T4, in which powerful selection could be exerted for the $r+$ phenotype. Forward mutants could be easily picked and large numbers of phage crosses could be readily performed. The selection method allowed recombinants to be measured at very high resolution, limited only by the rates of reverse mutations. These experiments revealed that the gene contains hundreds of sites at which mutation could occur and that the scale of separation is of the order of the distances between adjacent base pairs on the DNA structure. The experiments not only exploded the classical idea of the gene as an indivisible unit of function, mutation, and recombination, but allowed the fine structure map to now be viewed as a picture of the nucleotide sequence of the DNA. This map therefore provided an approach to studying how the sequence of bases in DNA might correspond to the sequence of amino acids in proteins by using genetics for the first and chemistry for the latter sequence. This programme of molecular genetics was never quite completed, but the exploitation of the properties of T4 bacteriophages played important roles in studies of mutagenesis, the general nature of the genetic code and genetic suppression, and in the demonstration of the existence of messenger RNA.

For our purposes today we need to reflect on the properties that made phages the ideal 'model organisms' for this phase of research in molecular biology. They are easy to grow and maintain in a laboratory, large numbers could be readily generated, and many experiments could be conducted in parallel. The final readout was the presence or absence of lysis

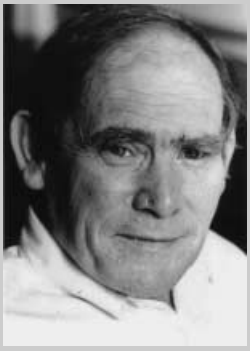
[a] *S. Brenner*
Salk Institute
100010 North Torrey Pines Road
La Jolla, California (USA)
Fax: (+1) 858-5528285
E-mail: sbrenner@salk.edu
and
[b] *King's College, Cambridge (England)*

of bacteria and this assay could be applied to single particles each of which left a plaque—an area of lysis in a lawn of bacteria. For mapping purposes, scoring was digital—yes or no—and single particles could be easily counted to obtain accurate frequencies.

After the basic principles of information transfer from genes to proteins had been established with the identification of messenger RNA, the discovery of the mechanism of protein synthesis, and the structure of the genetic code, it was natural for some of us to ask whether the lessons learnt in molecular biology could be applied to the genetics of more complex phenotypes. All questions in genetics involve asking how the phenotype is represented in the genotype or, reflexively, how the genes map onto the phenotype. It is clear that one could not have understood how metabolism or biosynthesis was represented on the genome until one had conceived of the 'one gene – one enzyme relationship,' and that a gene specifies the amino acid sequence of a polypeptide chain, which after folding correctly, is able to perform a specific catalytic function in a metabolic pathway. In the same way, the elaborate structure of the protein coat of a bacteriophage could not be understood without knowing that it is built of many different kinds of subunits and that these are able to self-assemble into the final particle. We have to know how the gene specifies the construction of the entity that carries out the function. The same is true at a higher level of organization. In studying the genetics of behavior, it is difficult, if not impossible, to go directly from the gene to behavior, because there is no simple mapping that connects the two. In my paper, I put it in this way: "Behaviour is the result of a complex ill-understood set of computations performed by nervous systems and it seems essential to decompose the question into two: one concerned with the question of the genetic specification of nervous systems and the other with the way nervous systems work to produce behaviour." Thus, just as the structure and function of protein molecules is the necessary connection between the genes and metabolism, the link between genes and behavior resides in understanding the structure of nervous systems and how they are constructed. These are questions of anatomy and embryological development. This set the requirements for the experimental organism as one which was not only suitable for genetical study in the laboratory but also allowed the structure of the nervous system to be accurately defined. Since a nervous system is essentially a cellular network, we had to be able to observe junctions between cells and their processes and this could only be achieved with the electron microscope, which has the necessary resolution. Since the electron microscope provides only a small window because of its high magnification, we needed a small animal which would also need to have a nervous system with a small number of cells. After some searching, my choice finally settled on the small nematode, *Caenorhabditis elegans*. This nematode is a self-fertilizing hermaphrodite with rare spontaneous males. The adults are about 1 mm in length and the life cycle is completed in $3^1/_2$ days. The animals live in a two-dimensional world feeding on *Escherichia coli* on the surface of agar plates. They are easy to grow in bulk, each animal producing about 300 progeny during a cycle. My paper reported the isolation of several hundred mutants together with their complementation and mapping to define about one hundred loci. In parallel with the genetic work, I launched a program with Nichol Thompson to determine the complete structure of the worm by serial section electron microscopy.[2] This project was completed by John White and Eileen Southgate who joined the group soon after it was formed and whose work resulted in the determination of the complete structure of the nervous system (and much more besides) in *C. elegans*.[3]

The paper I referred to summarized the work on genetics but in the early seventies, trying to understand the functions of genes specifying the development of the nervous system in molecular terms seemed impossibly remote. However, going to the molecular level was inevitable, and we very early initiated studies of mutants which affect the movement of the worm by disrupting muscle structure function.[4–6] The defects were easily characterized by electron microscopy and we began to work on two of the genes, *unc*-15 and *unc*-54, which had abnormal thick filaments in the body muscles. We chose muscle because the structural proteins constitute a considerable fraction of the total protein of the worm and could be easily isolated and characterized by biochemical analysis. Nonetheless, this seemed a rather limited and oblique way to tackle a very large problem.

Our work and indeed the whole field was revolutionized by the discovery of DNA cloning and by the invention by Sanger



Sydney Brenner was born in South Africa and educated at the University of Witwatersrand, Johannesburg (Medicine and Science). He went to Oxford University and received a D.Phil. degree in 1952 working in the Physical Chemistry Laboratory. After a brief return to South Africa, he joined the Medical Research Council (MRC) Unit in the Cavendish Laboratory in Cambridge in 1956 and went on to serve as director of the successor of that laboratory, the MRC Laboratory of Molecular Biology, in Cambridge from 1979 to 1987. In 1987 he became Director of the MRC Unit of Molecular Genetics and retired in 1992 from the MRC. In 1996 Dr. Brenner founded the Molecular Sciences Institute in Berkeley, California. He is now Distinguished Professor at The Salk Institute, La Jolla, California.

Dr. Brenner's early research was in molecular genetics, working with bacteriophages and bacteria; he discovered messenger RNA (with Francois Jacob and Matthew Meselson) and, with Francis Crick, showed that the code was composed of triplets. In the 1960s he changed direction and initiated his research on C. elegans*, establishing it as a powerful experimental system for the analysis of complex biological processes. As a geneticist, he saw that the techniques of cloning and sequencing would open up new ways of approaching genetics. He turned to studying vertebrate genomics and has established the pufferfish genome as a powerful tool in genome analysis.*

and Gilbert a little while later of methods for sequencing DNA. It seemed possible that we could obtain the whole genome as a collection of DNA fragments and then proceed to characterize the genes directly by chemical means. At the Asilomar Conference in 1975, I gave a talk on our work on *C. elegans* and on the promise of applying the new methods for obtaining insights into the molecular basis of gene action in complex organisms. We had already characterized the DNA complement of *C. elegans* and, in a paper[7] accompanying the genetics paper, John Sulston and I showed that the genome consisted largely of unique DNA sequences and contained 20 times the amount of DNA in *Escherchia coli*. This study estimated the haploid amount of DNA as 80 megabases, which was an underestimate, partly because the value taken for *E. coli* was too low. We began to work in earnest to try to clone the genes we had assigned to components of thick filaments, the heavy chain of myosin, and paramyosin. This was accomplished by Jonathan Karn and Alexander MacCleod,[8] who used a myosin mutation previously shown to be a small internal deletion of the protein[9] to validate that the correct gene had been cloned. The great power of the new genetic approach was revealed by subsequent work of Jonathan Karn. Very quickly three other loci were identified for myosin heavy chains, and all were sequenced. It became clear that the best way to obtain the amino acid sequence of a protein was to sequence the gene specifying it; indeed, for the heavy chain of myosin the work involved had been reduced about a thousand fold.

The cloning of genes involved in the development of *C. elegans* and the identification of their proteins led quickly to insights into the molecular basis of the defects in the mutants that had been isolated. The irony was that, while *C. elegans* had been proposed as a model organism to understand genetical specification in the more complex mammalian organisms, by the time we were able to analyze our mutant genes, a few thousand genes had been cloned from humans, rats, and mice, and these were frquently the sequences that provided instant identification of many *C. elegans* genes. These early studies led John Sulston and others to map the genome of *C. elegans* into ordered clones, and this work was followed by obtaining first a draft sequence and, more recently, the complete sequence of the genome of the worm. At the same time John Sulston, as you will hear later, was able to derive the complete lineage of the cells in the hermaphrodite and the male. Thus, there had become available a model organism in which we knew the positions of every cell in the body and how they were connected to each other, the complete cell lineage, which told us where every cell came from in development, and the total DNA sequence together with a host of methods to alter the DNA and its expression. A similar path has been followed by those working on *Drosophila*, but the fly is much more complex than the worm and the anatomy of its nervous system has not reached the level of completeness achieved for *C. elegans*. The complete genome sequence of the fly is now also available and both the fly and the worm will continue to be important models for the study of function in complex organisms.

The new techniques of cloning and sequencing, which greatly enlarged the scope of genetical analysis in the nematode, led to the question of whether they could be extended to other organisms in which standard genetical analysis was difficult or even impossible. The new technology had liberated genetics from the tyranny of the reproductive cycles of organisms and, in principle, could now be applied to any organism. Its power stems from the properties of DNA, which enable its isolation, amplification, and expression in simple microbes, and from the uniqueness of DNA sequencing, a technology that allows us to extract the essential information—the linear sequence of nucleotides—in DNA from any source: viruses, microbes, plants, and animals, as well as from molecules that can be copied into DNA.

In 1985, when the first suggestions were made to sequence the human genome, I thought that the sequencing techniques, even with incremental improvements, would not be equal to the task, and would require a factory scale operation to do it. I had also come to the conclusion that most of the human genome was junk, a form of rubbish which, unlike garbage, is not thrown away. My view at the time was that we should treat the human genome like income tax and find every legitimate way of avoiding sequencing it. It could therefore be asked whether a genome existed in nature that perhaps had very much less junk but nevertheless had the full repertoire of vertebrate genes? It is easy to ask the question if one already has the answer. Towards the end of the 1960s I spent time in Woods Hole and took advantage of the library, where I first discovered the papers of Hinegardner.[10] At the time, I was puzzled by the enormous variations in the amounts of DNA in different organisms. Indeed, whereas most physicists thought that organisms did not have enough DNA to specify their complexity, it was clear to me that many organisms had too much. I discovered from Hinegardner that one group of fish, the Tetraodontidae, which included the Japanese pufferfish, *Fugu*, had very small genomes, with a haploid content of about 400 megabases as opposed to the 3000 megabases of mammalian genomes. Although teleost fish are distant from humans they are still vertebrates, with the same body plans, development, and physiological systems as ourselves. Because of these basic similarities it seemed unlikely that *Fugu*, with a haploid DNA content one-eighth that of mammals, would have eight times fewer genes, which made it much more probable that what was missing in *Fugu* was junk DNA. If *Fugu* had the same gene repertoire as humans, then its genome would be more compact, giving us the human gene inventory for eight times less work and expense. We proved that this was indeed the case and proposed the genome of *Fugu* as the ideal model vertebrate genome,[11] with a DNA content only four times that of *C. elegans*. I failed to persuade any of the official genome organizations of the virtues of such a model genome and it remained a personal project until quite recently when, with the collaboration of the US Department of Energy and the generous support of the Agency for Science, Technology and Research of Singapore, we were able to produce a first draft sequence which vindicated everything we had been saying for the past decade.[12] This project also reunited several members of the group who helped me start this work in Cambridge: S. Aparicio, G. Elgar, and B. Venkatesh. Sam Aparicio, originally,[13] and B. Venkatesh, and I[14–16] have used the *Fugu* genome in an

interesting application of the new genetics to discover the structure of control sequences in mammalian genomes by asking whether the mouse is able to read the *Fugu* genome in the same way as it reads its own by seeing whether there are sequences in *Fugu* that control expression with the same cell specificity as those in the mouse. So far apart are these genomes on the evolutionary scale that time has randomized all irrelevant sequences to reveal only those with conserved function unobscured by remnants of common origin. In fact, these experiments are tantamount to crossing a mouse with a fish, exploiting the power of modern DNA technology to penetrate natural reproductive barriers and extending the power of functional genetical analysis. *Fugu* is not a model organism but is rather an organism that possesses a model genome. The mouse is the model organism in this case, but, unlike forty years ago, when I began to work on *C. elegans*, we now do not require a single model on which everything—genetics, physiology, and biochemistry—can be done. Today we can take the genome from one source and cells from another, and we can create unique biological material by moving genes from one organism to another.

What of the future? I want to discuss briefly two ideas about future research which I think will become the challenge for the future. If we take 2020 as the year of good vision, we may note that this is seventeen years in the future and it represents about the same time as that between today and 1985, when the first discussions began about sequencing the human genome. It would not be excessively outrageous to suggest that the projects begun today could be completed by 2020, especially if we continued to work as a community of scientists.

The first of these projects I call **CellMap**. We are all conscious today that we are drowning in a sea of data and starving for knowledge. The biological sciences have exploded, largely through our unprecedented power to accumulate descriptive facts. How to understand genomes and how to use them is going to be a central task of our research for the future. We need to turn data into knowledge and we need a framework to do it. So genocentric has modern biology become that we have forgotten that the real units of function and structure in an organism are cells and not genes. The genome has gives us the inventory of gene loci and we must now get on to the discovery of the actions of the products of genes and how these are integrated in the physiology of cells. First, we will need to define all of the noncontingent states of gene expression in an organism, which is proposed as the correct way of defining a cell type. How many different cell types do we possess? There may be 200 in the body but there are likely to be many more in the brain. On top of this are all the contingent states; cells respond to outside stimuli and change their patterns of gene expression, but I do not consider a naïve neuron as a different cell type from its near identical neighbor which may have learnt something, although clearly the capacity to learn is a non-contingent property of that cell type. The next task will be to see how the gene loci map onto these cell types, to solve what may be called the instantiation problem. Many gene loci have multiple instantiations distinguished by different promoters, which specify the cell type for expression, or by additions or deletions of coding sequences by differential splicing that determine the location in the cell where that instantiation may act. There may be other differentially spliced sequence differences that control the lifetime of the messenger RNA or of the protein product. Not only is the cell the only physical locus for gene action but it is the correct level of abstraction to construct a framework for understanding functions. **CellMap** is seen as a map in many dimensions; it is at once a map of the cells in the organism onto which are projected the map of instantiations, as well as a map of the molecules in the cell. It is also a temporal map connecting cells with their predecessors and successors in development. By studying how such cells are connected with their homologues in different organisms we can see how these maps are layered in evolutionary space and what has been added to or removed from any particular subsystem as we move up and down on the evolutionary scale. The architecture of **CellMap** will be couched in a form that would facilitate computation, so that we can develop it into a predictive system, and, in the future, a system that we could use for the synthesis of new cell types and new organisms.

My second *gedanken* project is called **Humanity's Genes**. It arose in my mind during a discussion of a proposal to take the inbred lines of mice and extensively intercross them to generate 30 000 different mice representing different mosaics of the initial gene pools. Specially trained mouse phenotypers would then analyze the physiological properties in these mice and correlate them with their individual genomes. Unfortunately, the latter is the difficult task, as today there is no reasonable technology that can achieve this in any depth. However, suppose technology existed which made it easy to characterize 30 000 genomes, perhaps even to the point of resequencing them, would we bother to do this work with mice? We could go directly to humans, where we already have large numbers of diverse genomes, with skilled and expensively trained phenotypers, called doctors, studying them. Thus, since the technology does not exist, it now needs to be invented to provide the means of accurately analyzing large populations of genomes for detailed studies of natural human genetic variation and its correlation with phenotypes of health and disease. I believe that this will be the major challenge in human biology and medicine in the next decade. I am convinced that we will make our significant discoveries in humans and that the mouse will be used to validate the human findings by genetic synthesis, much in the same way as the chemist confirms a structure analysis by chemical synthesis. Chakravarti has shown the way forward by his synthesis of Hirschsprung's disease in mice.[17]

CellMap and **Humanity's Genes** are not really separate projects. We need the first to tell us where to look initially in genomes and to interpret what we may discover in the studies of human genetic variation. What will be the significant differences that we may find? Will they be amino acid changes in polypeptides, or changes in control sequences that affect the timing and amount of the products expressed, or in RNA molecules whose functions we are only now beginning to glimpse? We may also come to understand the enormous changes in structure and function that were brought about by evolution and gave our brains the immense capacities that they possess.

Nature has been generous to science and has provided us with many model systems. I have mentioned only the few that have been important in my own scientific work. *C. elegans* will continue to yield fruitful discoveries and insights in spite of my argument that we do not need model systems any longer for the study of human biology. However, there are many aspects of humanity that we still need to understand for which there are no useful models. Perhaps we should pretend that morality is known only to the gods and that if we treat humans as model organisms for the gods, then in studying ourselves we may come to understand the gods as well.

I need to acknowledge the great debt I owe to the large number of scientists who joined me in converting Caenorhabditis elegans *from what was once regarded as a joke organism into the powerful experimental system it has become today. The recognition given to our work by the award of the Nobel Prize owes much to their labours. I want particularly to record the patient and generous support given to me by the Medical Research Council of Great Britain, who allowed me to initiate and develop the research on* C. elegans *in the MRC Laboratory of Molecular Biology in Cambridge. Such longterm research could not be done today, when everybody is intent only on assured short term results and nobody is willing to gamble. Innovation comes only from the assault on the unknown.*

My research on Fugu *was initiated with funding by the prize awarded to me by the Fondation Louis Jeantet de Medicine of Switzerland, and by generous grants from E. I. du Pont de Nemours & Company, Wilmington (USA). It has also enjoyed continued support from the Institute of Molecular & Cell Biology and the Agency for Science, Technology and Research in Singapore. Some of the research was also carried out in The Molecular Sciences Institute, Berkeley (USA).*

[1] S. Brenner, *Genetics* **1974**, *77*, 71 – 94.
[2] S. Ward, J. N. Thomson, J. G. White, S. Brenner, *J. Comp. Neurol.* **1975**, *160*, 313 – 338.
[3] J. G. White, E. Southgate, J. N. Thomson, S. Brenner, *Phil. Trans. R. Soc. London, Ser. B.* **1986**, *314*, 1 – 340.
[4] H. F. Epstein, R. H. Waterston, S. Brenner, *J. Mol. Biol.* **1974**, *90*, 291 – 300.
[5] R. H Waterston, H. F. Epstein, S. Brenner, *J. Mol. Biol.* **1974**, *90*, 285 – 290.
[6] A. R. MacLeod, R. H.Waterston, R. M. Fishpool, S. Brenner, *J. Mol. Biol.* **1977**, *114*, 133 – 140.
[7] J. E. Sulston, S. Brenner, *Genetics* **1974**, *77*, 95 – 104.
[8] A. R. MacLeod, J. Karn, S. Brenner, *Nature* **1981**, *291*, 386 – 390.
[9] A. R. MacLeod, R. H. Waterston, S. Brenner, *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5336 – 5340.
[10] R. Hinegardner, *Am. Nat.* **1968**, *102*, 517 – 523.
[11] S. Brenner, G. Elgar, R. Sandford, A. MacRae, B. Venkatesh, S. Aparicio, *Nature* **1993**, *366*, 265 – 267.
[12] S. Aparicio, J. Chapman, E. Stupka, N. Putnam, J. M. Chia, P. Dehal, A. Christoffels, S. Rash, S. Hoon, A. Smit, M. D. Gelpke, J. Roach, T. Oh, I. Y. Ho, M. Wong, C. Detter, F. Verhoef, P. Predki, A. Tay, S. Lucas, P. Richardson, S. F. Smith, M. S. Clark, Y. J. Edwards, N. Dogget, A. Zharkikh, S. V. Tavtigian, D. Pruss, M. Barnstead, C. Evans, H. Baden, J. Powell, G. Glusman, L. Rowan, L. Hood, Y. H. Tan, G. Elgar, T. Hawkins, B. Venkatesh, D. Rokhsar, S. Brenner, *Science* **2002**, *297*, 1301 – 1310.
[13] S. Aparicio, A. Morrison, A. Gould, J. Gilthorpe, C. Chaudhuri, P Rigby, R. Krumlauf, S. Brenner, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 1684 – 1688.
[14] B. Venkatesh, S. L. Si-Hoe, D. Murphy, S. Brenner, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12 462 – 12 466.
[15] D. H. Rowitch, Y. Echelard, P. S. Danielian, K. Gellner, S. Brenner, A. P. McMahon, *Development* **1998**, *125*, 2735 – 2746.
[16] S. Brenner, B. Venkatesh, W. H. Yap, C. F. Chou, A. Tay, S. Ponniah, Y. Wang, Y. H. Tan, *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *99*, 2936 – 2941.
[17] M. Angrist, S. Jing, S. Bolk, K. Bentley, S. Nallasamy, M. Halushka, G. M. Fox, A. Chakravarti, *Genomics* **1998**, *48*, 354 – 362.

Received: April 16, 2003 [A 625]

