LIVING MOLECULES*

DR. M. F. PERUTZ

Cavendish Laboratory, University of Cambridge, England

SCHOOL text-books of biology taught me that living cells are filled with *protoplasm*. At the University I was led to believe that protoplasm was a suspension of colloids, large molecular aggregates of indefinite structure which somehow endowed it with life. Today we realise that these names were mere repositories of ignorance and that protoplasm is really a microcosm of vast complexity and definite molecular structure. We are still extremely far from knowing this structure in detail, but we are beginning to understand some of the basic facts of reproduction, growth, breathing and movement in molecular terms.

Growth is essentially a process of chemical synthesis, and to understand it we must know how it is directed and controlled. The cell cannot work under the extreme conditions of temperature and pressure used in a chemical factory. Instead, it synthesises its constituents in a series of small chemical steps, each brought about by a specific chemical called an enzyme. Several thousand such enzymes are probably required to bring about all the diverse processes on which life depends. Well over 1,000 different enzymes are already known, and all those which have been isolated were found to be complicated substances of the type known as proteins. This makes the determination of the way in which proteins are built one of the fundamental problems in biology and medicine.

CONTROLLING LIFE CELL

If proteins are the enzymes which control the life of the cell, what controls the synthesis of the proteins? It cannot be other enzymes made of protein, since they would have to be made by yet more enzymes, and so on *ad infinitum*. We now know that the synthesis of enzymes is controlled by genes—the material which passes from generation to generation and determines inherited characteristics.

A gene must therefore possess a dual function: *it must be able to copy itself exactly*, so as to pass the message to the next generation and it must be able to determine the structure of a protein or a protein molecule.

The genes of nearly all organisms consist of a material which has been called deoxyribonucleic acid or DNA for short. This DNA is so made as to form a chemical code. Its molecules consist of long chains of atoms in which an identical chemical pattern of sugar and *phosphate repeats at regular intervals* like the links of a chain.

Attached to each link is a rather complicated group of atoms called a base. There are four different kinds of base. We do not yet know in what order they are arranged in any one nucleic acid, nor have we any direct information that they are arranged in any definite order at all; we know only that their proportions are constant and characteristic in the DNA from any particular species.

It is the fact that the bases are the only variable constituents which makes us believe that they are arranged in a definite order and that this sequence is the "code" which carries the inherited information. If this is true then the genetic language is written in a four-letter alphabet on an immensely long scroll. The actual number of "letters" in the DNA of a bacterial virus is 500,000 and in the chromosomes of a mammal about 3,000,000,000.

CONTINUITY OF INHERITANCE

To ensure continuity of inheritance an exact copy of this information has to be made each time a cell divides. In 1953, J. D. Watson and F. H. C. Crick at the Medical Research Council's Molecular Biology Research Unit at Cambridge proposed a structure for DNA which suggests a possible copying mechanism (Fig. 1).

It consists of two chains of DNA coiled round each other, like two snakes, to form a double helix. The actual model looks like a spiral staircase in which the links of the nucleic acid chain form the banisters, and the bases attached to them form the steps. Each step consists of two bases, one from each of the chains, which are linked together by chemical bonds.

Suppose now that the four bases which form the symbols of the genetic code are called A, T, G and C, then only specific pairs of bases can be linked to form a step, such that A is always linked to T and G to C. This means

---

* This article is specially written by a leading scientist in Britain to mark the 300th Anniversary of the Royal Society in July 1960. The Royal Society—or to use its full title, "The Royal Society of London for the Promotion of Natural Knowledge", received its first Charter from *King Charles II, who also described himself as its* founder and patron.

that a particular sequence of bases in one chain must be paired to a complementary sequence in its partner chain. Wherever A appears in one chain T must be in the other, not A, G or C. This is the vital idea of Watson and Crick's model, for from it we can deduce the way in which the genetic code might copy itself.

When we want to copy a document we prepare a negative from which we make a positive print. The complementary sequence of

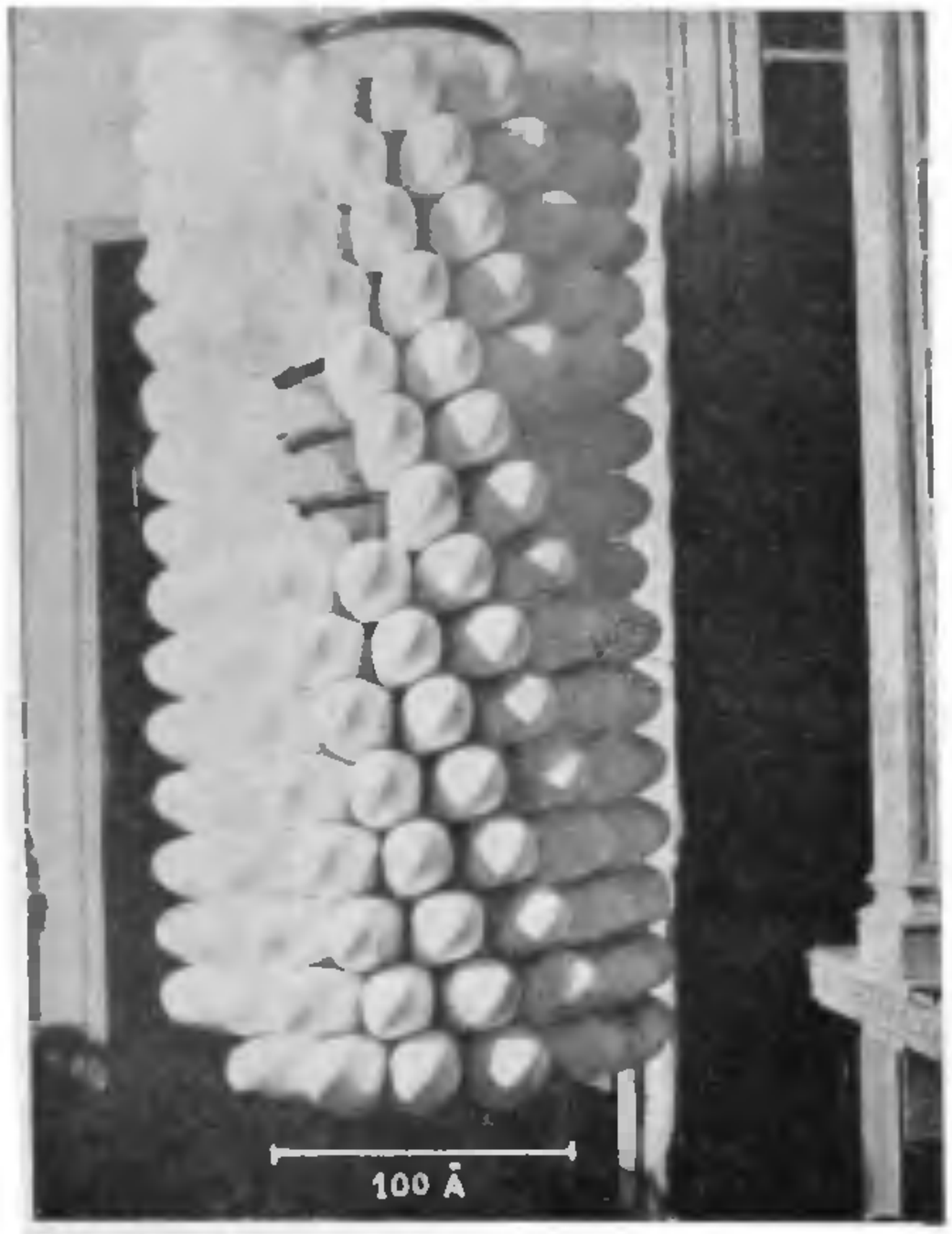

FIG. 1. A model of a double helix structure of DNA. The phosphate-sugar links are represented by wires and the bases by disks. The scale indicates Ångstrom units. 1 Å=1/10 000,000 of a millimeter or 1|250 000,000 of an inch. (One and a half Ångstroms is the size of a single carbon atom. The Ångstrom is the unit of measure on the atomic scale.)

bases in the two chains makes each DNA "staircase" a negative and positive combined. To reproduce themselves the two chains of the parent double helix would have to separate in a solution full of loose links similar to those which form the chains, and each chain would have to become a template for the building of a new complementary chain which grows around it.

Month-by-Month Information

This means that a loose chain link carrying A joins on to T in the parent chain, a loose chain link carrying G joins on to C in the parent chain and so on. When all the loose chain links are joined, each parent chain will be intertwined with a new daughter-chain made up of a complementary sequence of bases. In this way one parent double helix gives rise to two daughter double helices each carrying on an exact copy of the genetic information.

We do not yet know how nucleic acids control the synthesis of enzymes. Certain parts of the mechanism have been discovered and others are still obscure, but almost every month some new observation is published in the scientific literature which adds a new piece to this great jigsaw puzzle.

Great progress has recently been made in Britain in elucidating the structure of the proteins themselves. Like the nucleic acids, proteins are giant molecules made up of molecular chains, but whereas DNA has only four different groups attached to the links of the chain, proteins possess 20 and these are called amino-acids. Like the nucleic acids, protein chains tend to take up helical configurations.

One protein may consist of one or several such chains, which in turn may contain anything from 20 to several thousand amino-acids, so arranged that the different kinds of side-chains occur in a definite order. This sequence is the translation of the genetic code into protein structure. Ten years ago the structure of protein was still largely unknown. In November 1958 the Nobel Prize for Chemistry was awarded to F. Sanger, a member of the Medical Research Council's Staff at Cambridge University, for having been the first to work out the chemical constitution of a protein, by determining the order of the 51 amino-acids in the two chains composing the molecule of insulin.

Chemistry Milestone

This discovery was one of the milestones in protein chemistry. First, it removed the last shadow of doubt from the protein chain hypothesis which Hofmeister, a German Chemist, had enunciated more than 50 years earlier. It established the fact that the different amino-acids really are arranged in a definite, genetically determined sequence, but disproved the widely held belief that this sequence was regular. It revealed the part played by sulphur bridges in the architecture of protein molecules and the chemical nature of the differences between animal species. Most important of all, Sanger demonstrated that the complete formula of a protein can be determined by chemical methods and thereby stimulated a great new volume of research all over the world.

The problem of protein structure really required a two-fold approach : the chemical one, used by Sanger to find the number of chains and the sequence of amino-acids, and a physical one to discover the way the chains are coiled and folded. The physical approach is based on X-ray analysis. It involves studies of the X-ray diffraction patterns, generally from single crystals, and is a technique that has been widely used to determine the atomic arrangement in simpler compounds. Most of these compounds, however, were at least 100 times smaller than protein molecules, and it was a matter of great difficulty to extend the methods of X-ray analysis to molecules of such enormous size and complexity. None of the approaches yielded much information, until I discovered, in 1953, that the problem could be solved by studying the X-ray diffraction patterns from a pair of crystals, one containing the protein alone and the other a derivative of the protein incorporating a heavy atom such as mercury. This method has now become the basis for the structure analysis of crystalline protein and viruses in many laboratories.

Its first great success was achieved in 1957 when my colleague J. C. Kendrew was able to build a three-dimensional model of myoglobin, a protein containing pigment group called haem to which the oxygen becomes attached. The first X-ray analysis was calculated with a limited power of resolution, sufficient to show the general configuration of the chain and the position of the hæm group. but incapable of resolving atoms.

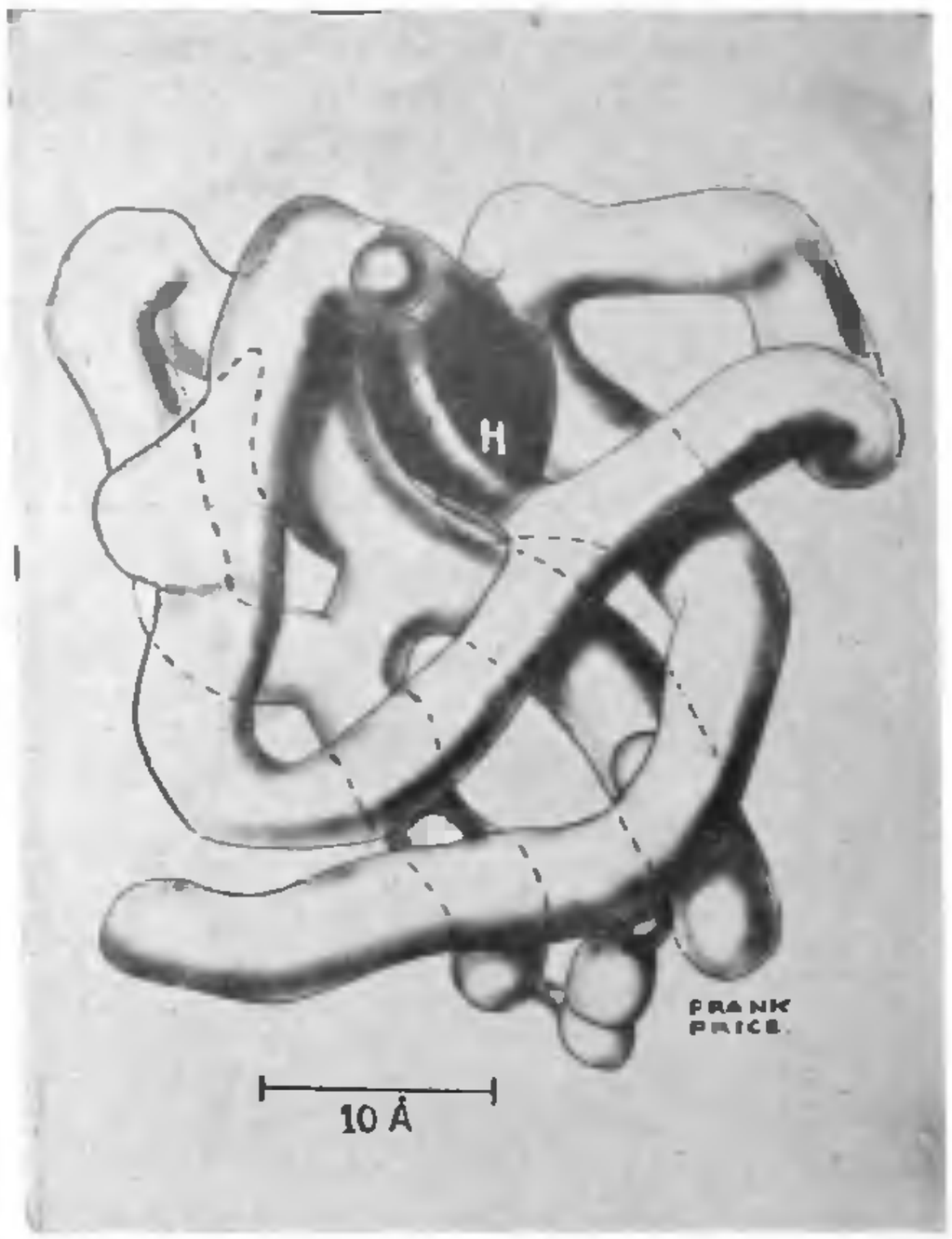

FIG. 2. J. C. Kendrew's model of the myoglobin molecule. The light winding rod represents the protein chain, the dark disk the hæm group.

New X-ray Analysis

Late last year Kendrew and his collaborators completed a new X-ray analysis at three times the resolution of the first, which shows the structure in almost atomic detail. The straight stretches of chain in Fig. 2 are now resolved into helices like right-handed screws. These make up about two-thirds of the structure. Departures from the helical configuration occur mainly where the chain bends or turns a corner.

The manner of linkage of the hæm group to the chain, which has long been a matter of great interest to biochemists, is well resolved. As we expected, the link is made through the nitrogen atom of an amino-acid side-chain called histidine which forms a chemical bond with the iron atom in the hæm group.

My own work is concerned with hæmoglobin, the protein in the red-blood cells which carries oxygen from the lungs to the tissues and carbon dioxide back to the lungs. It contains about 10,000 atoms and consists of four chains each with about 140 amino-acids, and four hæm groups, each capable of carrying one oxygen molecule. The result of an X-ray analysis carried out by several colleagues and myself show that each of the four chains is bent into the same shape as the chain in myoglobin. The four chains are then assembled at the corners of a tetrahedron and together make up a molecule which is very nearly a sphere. The vitally important hæm groups lie in separate pockets at the surface of the molecule, each pocket being formed by the folds in one of the chains.

Complex Structure

How is the enormously complex structure of this molecule related to its function? It serves to carry four molecules of oxygen, minute in size by comparison, from the lungs or gills to the tissues. Each oxygen molecule is attached to one hæm group. It is one of the vital physiological properties of hæmoglobin that these hæm groups interact, so that the combination of the first one with oxygen makes it easier for the next one to combine, and so on. The present model is sufficiently detailed to rule

out some wrong theories which were current to explain these interactions, but not yet detailed enough to tell us the right explanation. This may come at the next stage of the analysis when we hope to work out most of the structure of hæmoglobin in atomic detail.

X-ray analysis can be applied to any crystalline substance, no matter how big its molecules, including some of the smaller plant and animal viruses which can be crystallized like any chemical and are yet in a certain sense alive. They contain millions, rather than thousands, of atoms.

X-ray work on the mosaic virus, which causes mottling in the leaves of the tobacco plant, was started at Cambridge University by J. D. Bernal and I. Fankuchen in the late 30's and has been continued in the past decade by D. Casper in the United States and by J. D. Watson, Rosalind Franklin and A. Klug in this country. As a result of their work we now know that the virus is a rod in which nucleic acid and protein are interwoven in a beautiful pattern (Fig. 3). A helical chain of nucleic

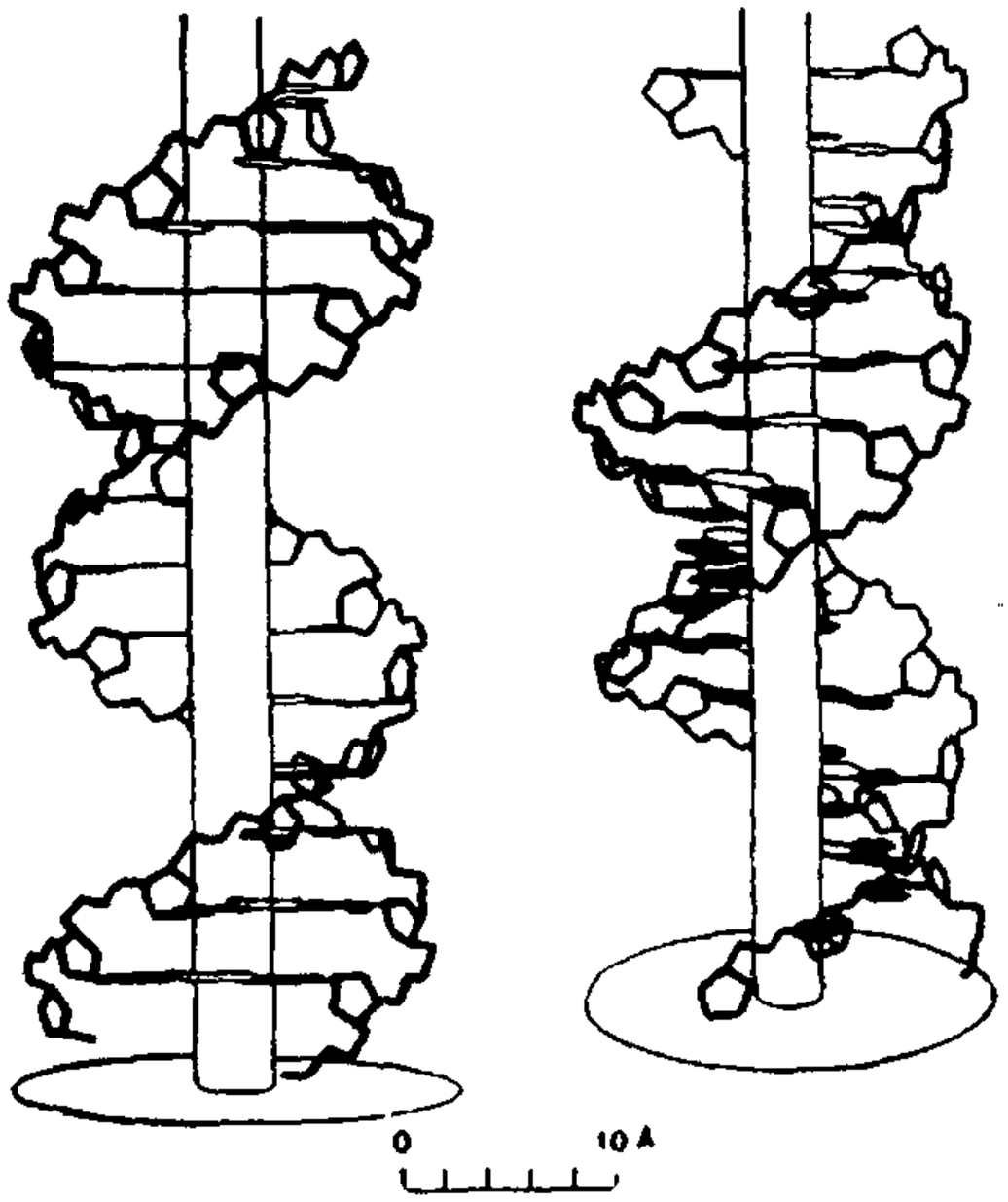

FIG. 3. A model of tobacco mosaic virus. Each "loaf" represents a protein sub-unit of molecular weight 17,000. Seen in detail it would look somewhat like the model in Fig. 2. The opening shows the coils of RNA inside the particle. The Pitch of the coil is 23 Å. (Rosalind Franklin and A. King).

acid is wound around a central cylindrical hole like the filament of an electric lamp, and is surrounded by a helical array of 2,000 protein molecules all exactly alike. Of the two components of the virus only the nucleic acid is infective. It acts as the gene which reorganises the infected tobacco leaf cell for the purpose of producing virus, while the protein seems to be merely a protective coat for the gene. On the atomic scale their size is truly gigantic and yet their structure is almost as definite as that of a molecule of sugar, where every atom occupies its rightful place.

Structure of Muscle

Movement is one of the most important manifestations of life. How is it accomplished on a molecular scale? The answer to this question has not yet been found, but it has been brought much nearer by the work of H. E. Huxley and Jean Hanson at London University, and of A. F. Huxley and F. Niedergerke at Cambridge University, who have studied the structure of muscle by a variety of techniques, especially electron microscopy.

Muscle contains two kinds of protein filaments which are known as myosin and actin. In most muscles, like those of the arms and legs, these filaments are arranged in a series of bands, visible under the microscope at striations one two-thousandth of an inch wide, which runs across the length of the muscle. The myosin filaments are arranged in parallel so that each one is surrounded by six others at the corners of a regular hexagon. At the centre of each triangle formed by three myosin filaments is a filament of actin. Changes in the length of the muscle are achieved by the actin filaments sliding into or out of the spaces between the myosin filaments. By an ingenious choice of material Jean Hanson and J. Lowy at the Medical Research Council's Biophysics Research Unit at London University, were recently able to show that the mechanism of contraction in the slower-acting, smooth muscle, like that of the uterus, is essentially the same as in striated muscle.

Sliding motion between two different kinds of protein filaments may well turn out to be the universal mechanism of movement throughout the animal kingdom. It gives a picture of movement on the sub-microscopic, but not yet on the molecular scale, and leaves the chemical forces responsible for sliding between the two kinds of filaments still to be discovered. This will be the next great step in muscle physiology.

The greatest advances in our understanding of biological function are likely to come through a knowledge of structure on the molecular scale. Recent work in Britain has contributed significantly towards this aim.