THE STRUCTURE OF PROTEINS

Caltech researchers discover the essential atomic structure of several proteins—including those found in bone, muscle and red blood cells

Linus Pauling and Robert B. Corey with molecular model showing configuration of atoms in a protein molecule

ONE OF THE MOST important problems in the field of biochemistry is to find out how proteins, the principal building blocks of the body, are put together.

At the Diamond Jubilee meeting of the American Chemical Society in New York last month, Dr. Linus Pauling, Chairman of the Institute's Division of Chemistry and Chemical Engineering, and Dr. Robert B. Corey, Professor of Chemistry at the Institute, reported "significant progress" after some 15 years of work on this fundamental problem. They have, in fact, discovered the essential atomic structure of several proteins—including those found in bone, muscle, and red blood cells. And work with other proteins has progressed to the point where knowledge of their structures appears to be imminent.

In learning how proteins are put together, we are gaining a deeper insight into the nature of living matter. We are on the way to a better understanding of the nature of physiological reactions—and on the way to gaining greater control over disease.

The major components of all living cells are proteins, fats, carbohydrates in various forms, salt and water. Though all of them are essential, the proteins among these are responsible for many of the activities which we associate primarily with living things—and, in this respect, they may be considered more important than any of the other components of living cells.

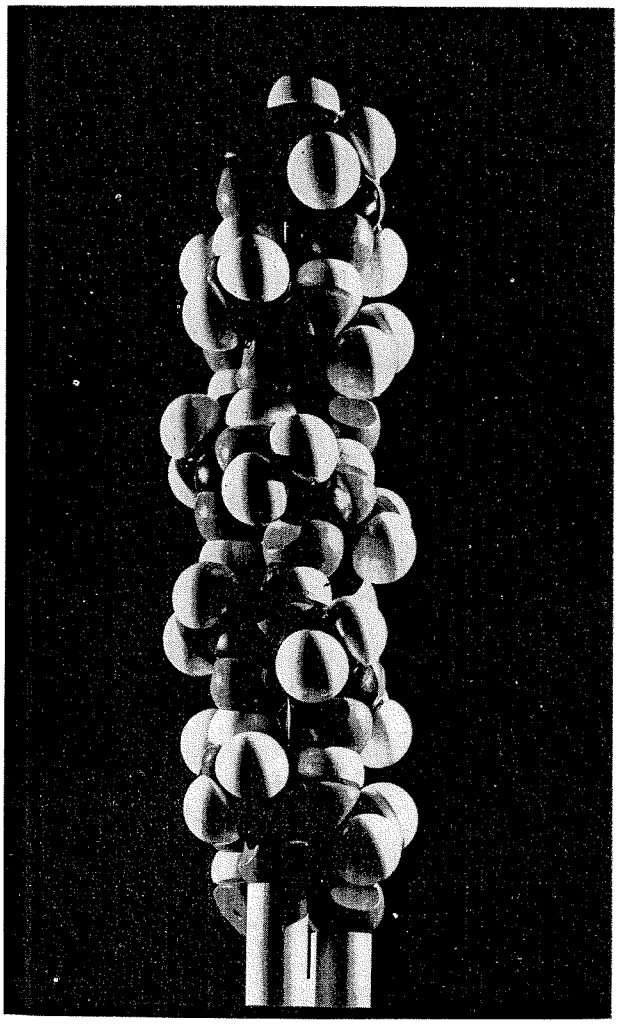
Some of the most important proteins include hair, wool, horns, fingernails and feathers. The major components of skin and muscle are proteins. Egg white is chiefly protein.

Hemoglobin, the red coloring matter of the blood, is another kind of protein. Some of the hormones, which regulate bodily activity, are proteins. All antibodies, all viruses, and the genes, the carriers of the mechanism of heredity, are proteins.

Altogether, there are thousands of different kinds of proteins in the human body. Unlike most other chemicals, which consist of a score or two individual atoms, protein molecules are made up of many thousands—and sometimes even millions—of individual atoms, each occupying a specific place in the architecture of the molecule.

Protein molecules, as one of their characteristic properties, are very large. For instance, the molecular weight of sodium chloride is about fifty-five; the molecular weight of the smallest protein that we know, insulin, is 12,000, and that is something of an exception. Of the common proteins, the smallest molecular weight is about 35,000.

Proteins are important constituents of foods. They are digested by the digestive juices in the stomach and intestines, being split in the process of digestion into small molecules. The small molecules are able to pass through the walls of the stomach and intestines into the blood



Wooden molecular model (250,000,000 linear magnification) of the 3.7-residue helical configuration of the polypeptide chain of some proteins

stream, by which they are carried around into the tissues, where they may then serve as building stones for the manufacture of the body proteins.

Proteins act by interaction with each other and with other substances, and what they interact with and how they interact are governed by their structure. Each protein has a special structure which enables it to do a specific job.

The first great advance toward an understanding of protein structure was made in 1900 when Emil Fisher, a German, found that a protein molecule is composed of simpler substances known as amino acids (of which there are 24 in all). The amino acids, in turn, were found to be linked together into larger groups known as peptides, and the peptides are linked together into even larger groups, known as polypeptides.

The problem of determining the structure of proteins then became one of finding the sequence of various amino acids in the chain and the way in which the chain is coiled.

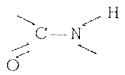
Thirty years ago scientists in Germany and England began making x-ray diffraction photographs of proteins in the hope that photographic analysis would reveal their structure. Work along these lines — the x-ray analysis of simple substances — was begun at the California Institute of Technology in 1916, by Burdick and Ellis, and was continued with great energy by Professor Roscoe G. Dickinson. Thus Caltech was a pioneer in the use of the technique in the United States. In making such photographs, x-rays, played on the substance, are scattered by the atoms. The resultant photographic pattern, as given by simple substances, allows scientists to determine the arrangement of their atoms—which, of course, can't be seen themselves. But protein structures proved to be too complicated for x-ray analysis to reveal their structure.

Indirect attack

Some fifteen years ago Dr. Pauling and Dr. Corey decided to attack the problem by a more roundabout method. Instead of trying to study the complicated proteins directly, the researchers went to work on the component parts of the proteins. They began an investigation of the crystal structure of the amino acids, of simple peptides (those made of two or three amino acids, rather than several hundred), and of other simple substances related to proteins.

In this way Drs. Pauling and Corey were ultimately able to obtain enough structural information to permit the precise prediction of reasonable configurations of the polypeptide chains (the backbones of several proteins). Intricate scale models of these molecules were then constructed.

The whole process was necessarily slow, of course. It took from six months to two years each to find the atomic structure of the four simplest amino acids, for example.

The structural units with which Drs. Pauling and Corey worked are rigid groups of atoms, in the form of planar amide groups  which are held together by carbon atoms. Each of these linking carbon atoms has a hydrogen atom and a side chain, characteristic of the amino acid, attached to it. The two bonds formed by each of these carbon atoms with the two adjacent amide groups are at an angle of 110° with one another. At first the assumption was made that any orientation could be assumed around each of the bonds.

An engineering problem

The investigators were then faced with a geometrical or engineering problem—the problem of finding orientations around these bonds that would not lead to steric hindrance between one part of the chain and another part of the chain, and also that would permit the oxygen atom and the NH group to form hydrogen bonds, N—H . . . O, with length 2.8 A.

Only four acceptable structures were found in three years of search. It was also found that all of the structures that had previously been proposed for proteins

had to be rejected, as not conforming to the requirements of interatomic distances, bond angles, and other structural features indicated by the investigations of the simpler substances.

One of the structures which has been found is shown in the adjacent figure. It is a helical structure (a spiral structure), in which each amide group forms hydrogen bonds with the amide groups that are removed by three in either direction along the chain. This helix has approximately 3.7 residues per turn, and it satisfies all of the structural requirements.

In a later stage in the investigation Drs. Pauling and Corey have assumed that certain orientations around the N—C bond and the C—C′ bond (the two bonds between the linking carbon atom and the adjacent amide groups) are favored relative to other orientations, 60° apart.

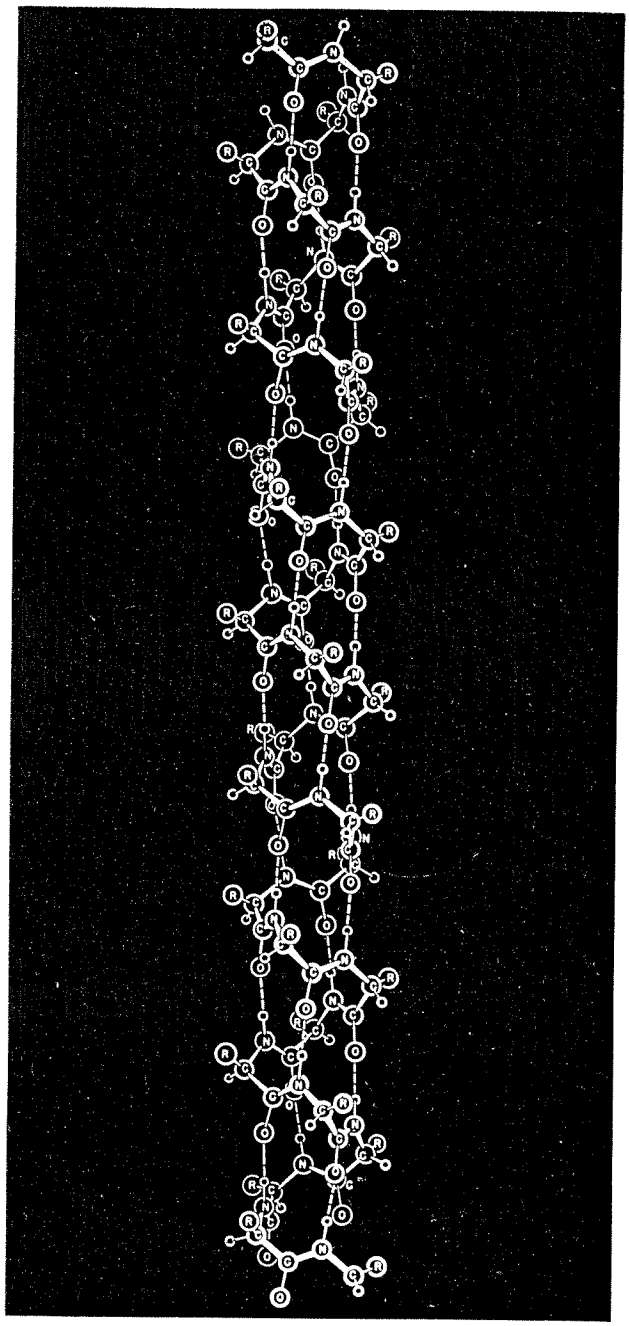
If this assumption is made, it becomes possible to make an exhaustive analysis of conceivable structures of different stages of complexity. The simplest stage of complexity is that in which the amino acid residues are assumed to be all equivalent except for differences in the side chains. There are then six possible orientations around the N—C bond, and six around the C—C′ bond, and hence a total of 36 structures.

Molecular models

Using a special set of molecular models built by Dr. Corey, on a scale of 250,000,000 linear magnification (that is, with 1 inch = 1 Angstrom), and fitted with set screws to give rigidity to the model, Dr. Pauling investigated the 36 structures. In a short time he found that 32 are impossible: either the chain is required to bend back on itself in such a way as to try to force two atoms into the same position in space, and thus to give rise to steric hindrance that would make the molecule unstable, or the NH groups and CO groups are oriented in such a way as not to permit them to form hydrogen bonds.

Two of the remaining four structures were found to be the 3.7-residue helix that had already been discovered. These two structures differ from one another in having all the side chains pointing upward or all pointing downward, relative to the axis of the helix. The other two structures were entirely new ones, in which the polypeptide chains have a zig-zag configuration that permits them to form hydrogen bonds with adjacent chains, producing a layer or sheet. There is evidence that these two structures occur in silk, stretched hair, stretched muscle, and some other fibrous proteins.

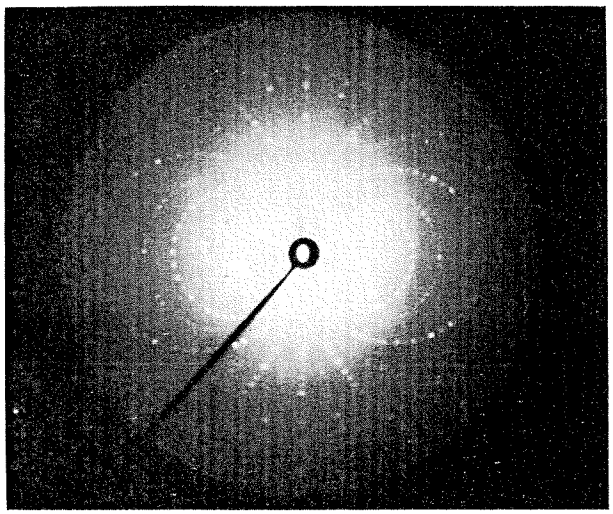
The 3.7-residue helix shown in the adjacent photograph and drawing is indicated by x-ray data to be present in ordinary hair, contracted muscle, horn, finger nail, and other proteins, and also in synthetic polypeptides. X-ray data for these proteins and the synthetic polypeptides provide values for the diameter of the spring-shaped molecules, and for the distance between the successive turns of the spring. These experimental values are in excellent agreement with the values predicted from the data for simple substances.

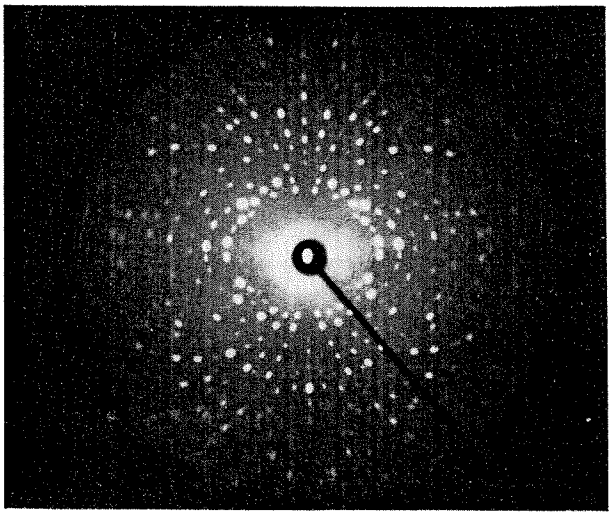


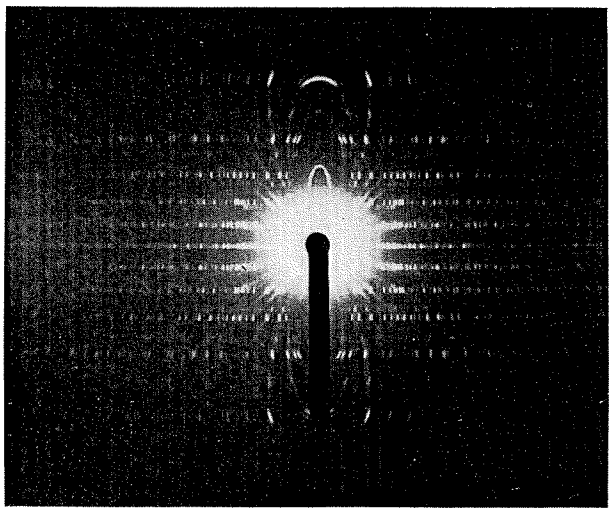
A drawing of the 3.7-residue helix shown on page 8

The fact that the number of amino acid residues in each turn of the spring is a fraction, 3.7, was surprising. Professors Pauling and Corey pointed out that Sir Lawrence Bragg and his collaborators in the Cavendish Laboratory at Cambridge, England, who had been working with great vigor on the same problem, had been led astray by making the unjustified assumption that the number of residues per turn in a helical structure had to be integral.

The same spiral structure has also been found to occur in the molecules of hemoglobin, the red protein inside the red cells of the blood. This protein contains iron







X-ray Diffration Photographs

X-ray diffraction photographs of amino acid crystals helped Caltech researchers determine how amino acid residues link up in protein molecules. The spots on the photographs shown here are reflections from the planes of atoms in various amino acid crystals. Positions of the atoms can be determined from the positions and intensities of the spots.

atoms (four to each molecule) and has the power of combining with oxygen in the lungs in order to transport the oxygen to the body tissues.

A surprising atomic structure has been found for another protein, collagen, which makes up tendons and is also present in skin and bone. The collagen molecule consists of three chains twisted around each other to form a three-strand cable. The chains are fastened to one another by means of hydrogen atoms. The molecules of gelatin also have this twisted three-strand structure.

Possible applications

The hope is, of course, that knowledge of the atomic structure of proteins will be found useful in medical research. The diseases treated with ACTH and cortisone, for example, are sometimes called collagen diseases because this protein—whose structure has now been determined—is thought to be involved in the diseases. Also, Dr. Pauling and other Caltech researchers have recently discovered that certain types of anemia, such as sickle-cell anemia, are associated with an abnormality in the hemoglobin molecules of the patients. In general, it is believed that drugs act by combining with proteins.

These are not the only possible fields of application of the research findings. They may also be found useful in the synthesis of proteins, since a knowledge of the atomic structure of a substance is fundamental in any attempt to synthesize it.

Dr. Pauling and Dr. Corey were assisted in their research at Caltech by E. W. Hughes, J. H. Sturdivant, Jerry Donohue, Verner Schomaker, D. P. Shoemaker, Gustav Albrecht, Henry A. Levy, W. J. Moore, Jr., G. B. Carpenter, H. R. Branson, and H. L. Yakel, Jr. The work has been supported by grants from the Rockefeller Foundation, the National Foundation for Infantile Paralysis, and the United States Public Health Service.