Molecular Architecture and Biological Reactions

LINUS PAULING

Chairman, Division of Chemistry and Chemical Engineering,
California Institute of Technology, Pasadena, Calif.

Answers to many basic problems of biology—nature of growth, mechanism of duplication of viruses and genes, action of enzymes, mechanism of physiological activity of drugs, hormones, and vitamins, structure and action of nerve and brain tissue—may lie in knowledge of molecular structure and intermolecular reactions

★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★ ★

THERE are two subjects that I am deeply interested in—structure, the detailed nature of molecules, crystals, and cells, described in terms of their constituent atoms, with interatomic distances determined to within 0.01 Å., an interest that began in my youth and has received most of my attention until recent years; and the basis of the physiological activity of substances, an interest that is more recent but just as keen. It is with a deep feeling of satisfaction that I have reached the firm conclusion in recent years that these two fields are most intimately related.

Why have we still so little understanding of the structural basis of the physiological activity of chemical substances, despite the interest and effort of many able physiologists and chemists during recent decades? I believe that it is because the problem has been examined, in the main, from one point of view only—not the wrong point of view, but one which, unaided, gives a vista insufficient to reveal its true complex nature. This point of view is that which surveys the chemical reactivity of molecules—their tendency to break their chemical bonds, the very strong bonds between atoms, and to form new chemical bonds. The other point of view which is needed is that which directs the mind's eye to the detailed size and shape of the molecules and the nature of the weak interactions of molecules with other molecules, in particular with the macromolecules and macromolecular stromatic structures which characterize the living organism. Until very recently physiologists and pharmacologists have barely thought of this aspect of their great problem—and I am convinced that once they begin to use this new idea seriously a period of the greatest development will have started. I believe that the next twenty years will be as great years for biology and medicine as the past twenty have been for physics and chemistry.

Eddington has said that the study of the physical world is a search for structure rather than a search for substance. If we ignore the philosophical implications of the words, we may say that the chemist and biologist in their study of living organisms must carry on both a search for structure and a search for substance, and that the second of these must precede the first. Investigators have had great success in isolating chemical substances from living organisms, and in determining the chemical composition of the simpler of these substances. The chemical composition is also known of many substances of external origin which exert physiological activity on living organisms. We may consider this work of isolation and identification of active chemical substances as the search for substance in biology.

The Search for Structure

The search for structure has also made great progress. From the one side biologists have, by visual observation with the microscope, made thorough studies of the apparent structure of aggregates of cells, of cells themselves, and of certain constituents of cells, such as chromosomes. This visual observation has provided information about structures in size extending down to 10^{-4} cm., 10,000 Å. Forty years ago the dark forest of the dimensional unknown stretched from this limit of the visible microscope back indefinitely into the region of smaller dimensions. In recent years the region from 10^{-7} down to 10^{-12} cm., containing atoms and simple molecules, has been thoroughly explored by an expedition outfitted with x-rays and similar tools, and the physicists are strongly pushing back into the region of the structure of atomic nuclei, below 10^{-12} cm. Another detailed exploration is being carried out with the electron microscope. This has pushed the nearer boundary of the unknown back from 10^{-4} to 10^{-6} cm.,

although the major portion of this region has been only sketchily explored during the few years since the development of the electron microscope, and a very great amount of work still remains to be done.

The answers to many of the basic problems of biology—the nature of the process of growth, the mechanism of duplication of viruses, genes, and cells, the basis for the highly specific interactions of these structural constituents, the mode of action of enzymes, the mechanism of physiological activity of drugs, hormones, vitamins, and other chemical substances, the structure and action of nerve and brain tissue—the answers to all these problems are hiding in the remaining unknown region of the dimensional forest, mostly in the strip between 10 and 100 Å., 10^{-7} and 10^{-6} cm.; and it is only by penetrating into this region that we can track them down.

There are many ways of investigating this region—by x-rays, ultracentrifuges, light-scattering techniques, the study of chemical equilibria, the techniques of degradation, isolation, identification, and synthesis used by the organic chemist, serological methods, chemical genetics, the use of both radioactive and nonradioactive tracers, the use of electron microscopes of improved resolving power—but no one method is good enough to solve the problem, and all these methods must be applied as effectively as possible if the problem is to be solved.

At the present time we know in complete detail the atomic structure of many simple molecules, including a few amino acids; but we do not know in detail how the amino acids are combined to form proteins. We do not know, except very roughly, even the shapes of such important molecules as serum proteins, enzymes, genes, the substances which make up protoplasm—and if we are to obtain a thorough understanding of the structure of living organisms detailed information

about the atomic arrangement of these substances must be obtained.

Let us imagine ourselves increased in size by the linear factor 250,000,000—the commonly used factor in molecular models, which makes 1 Å., 10^{-8} cm., become approximately 1 inch, atoms on this scale being 2 or 3 inches in diameter. With this magnification we would become about equal in height to the distance from the earth to the moon. Let us consider ourselves examining the earth, which would appear to us to be about the size of a billiard ball; and let us concentrate our attention on a small organism on the surface of the earth—New York City—which would appear as a spot about 0.01 inch in diameter, barely visible to the naked eye, and showing itself to be living by slow changes in shape and size.

To obtain a better view of this organism we could use a microscope, the resolving power of which would be about 1,000 feet; we could distinguish Central Park, the rivers, and such aggregates of skyscrapers as Rockefeller Center, but the individual skyscrapers would not be clearly defined. By "chemical" methods we would know that, running through the veins and arteries of this organism, there were substances such as street cars, busses, automobiles, ships, and people; and we might, by the use of membranes of known pore size or by some similar method, obtain the molecular weight of these. In addition, we would have obtained, through the application of a strange method of experimental investigation, the diffraction of x-rays and electron waves, complete information about the structure of objects smaller than about 1 foot in diameter, such as a storage battery, a small electric motor, a piece of cable, a small gear wheel, a bolt or rivet.

The use of the electron microscope, with resolving power about 10 feet, would give us very much additional information. We would know exactly—that is, to within 10 feet—the shape of the Empire State Building, though we might not be sure about the separate smaller rooms into which it is divided, and we could not obtain by the electron microscope information about the elevators and the machinery for operating them, the steel girders of which the building is constructed, and other structural features of similar size. We would be able to see, with the electron microscope, an automobile only as a particle, barely discernible, and roughly spherical in shape, and the human beings in the city would not be visible. We could get complete information about a storage battery, a ring gear, a brake pedal—but not about the automobile built up of these and many other parts; and it is clear that to obtain an understanding of the structure of this city we would still need to find a method of exploring objects in the range 1 to 10 feet.

Our hope for achieving precise knowledge about biological structures and reactions is based largely on the electron microscope and on diffraction methods. The diffraction studies of simple molecules have been carried out in sufficient number to permit the formulation of generalizations about atomic radii, bond angles, and other features of molecular configuration; it is still very important that the exact structure be determined of vitamins, bacteriostatic agents, and other physiologically active substances—the complete crystal structure determination of the rubidium salt of penicillin so ably made by Dorothy Crowfoot and Barbara Rogers-Low (6) has provided not only decisive information about the chemical formula of the substance but also the structural basis for later consideration of the detailed mechanism of its bacteriostatic activity.

Structure of Protein

The most important of all structural problems is the problem of the structure of proteins: until this problem is solved all discussions of the exact molecular basis of biological reactions remain in some degree speculative. The polypeptide-chain structure of proteins proposed by Fischer is now generally accepted, and there is little doubt that the picture of folded chains held by hydrogen bonds, van der Waals forces, and related weak interactions in more or less well-defined configurations, as discussed eleven years ago by Mirsky and me (9) is essentially correct. But this whole picture remains very vague—for only a few proteins (such as β-lactoglobulin (3)) do we have nearly complete knowledge of the numbers of residues of the different amino acids in the molecule, and for no protein does there exist more than fragmentary information either about the sequence of the different residues in the polypeptide chain or about the way in which the chain is folded. Only for fibrous proteins in the completely extended state do we have knowledge (still very rough) of the configuration and relative orientation of the polypeptide chains (as originally determined by Astbury), and this knowledge applies only to the backbone of the chains and not to the side groups. There is urgent need for complete and accurate structure determinations of proteins and related substances. So far these determinations have been reported for only four such substances (5)—two amino acids and two simple polypeptides—all made in our Pasadena laboratories; and it is my hope that, now that the war is over, precise information will rapidly accrue, including ultimately detailed structures of fibrous proteins, respiratory pigments, antibodies, enzymes, reticular proteins of protoplasm, and others.

Importance of Shape

Despite the lack of detailed knowledge of the structure of proteins, there is now very strong evidence that the specificity of the physiological activity of substances is determined by the size and shape of molecules, rather than primarily by their chemical properties, and that the size and shape find expression by determining the extent to which certain surface regions of two molecules (at least one of which is usually a protein) can be brought into juxtaposition—that is, the extent to which these regions of the two molecules are complementary in structure. This explanation of specificity in terms of "lock-and-key" complementariness is due to Paul Ehrlich, who expressed it often, in words such as "only such substances can be anchored at a particular part of the organism which fit into the molecule of the recipient combination as a piece of mosaic fits into a certain pattern".

In recent years the concept of complementariness of surface structure of antigen and antibody was emphasized by Breinl and Haurowitz (4), Mudd (10), and Alexander (1), and then was strongly supported by me (11) in the course of an effort to understand and interpret serological phenomena in terms of molecular structure and molecular interactions. Since 1940 my collaborators (Dan H. Campbell, David Pressman, Carol Ikede, L. H. Pence, G. G. Wright, S. M. Swingle, D. H. Brown, J. H. Bryden, A. L. Grossberg, L. A. R. Hall, Miyoshi Ikawa, Frank Lanni, J. T. Maynard, and A. B. Pardee) and I have gathered a great amount of experimental evidence about antigen-antibody interaction (12), which not only supports the general thesis that serological specificity is the consequence of structural complementariness, but provides information about extent of complementariness.

It has been verified that the closeness of fit of an antibody molecule to its homologous haptenic group is to within better than 1 Å.—that a methyl group (van der Waals radius 2.0 Å.) can replace a chlorine atom (radius 1.8 Å.) in a haptenic group with little interference with its combination with antibody (as was first shown by Landsteiner), but that interference is caused by replacing a hydrogen atom (radius 1.2 Å.) by a methyl group. The complementariness in structure with respect to proton-donating and proton-accepting hydrogen bond-forming groups has been found to be very important in determining the strength of attraction of antibody and haptenic group; and the complementary electrical charge in antibody homologous to the p-azophenyltrimethylammonium group has been shown to be within about 2 Å. of the minimum possible distance from the charge of opposite sign in the haptenic group. The great amount of quantitative data which has been gathered for scores of different haptens and antigens and successfully interpreted in terms of molecular structure and the concept of complementariness leaves no doubt that this structural explanation of serological specificity is correct.

The phenomenon of specificity, so common in biology, is rare in chemistry (with

the sole general exception mentioned below). Only very occasionally does there occur a unique representative of a class of compounds, such as the ion $W_2Cl_9^{---}$, which owes its special stability to the ratio of radii of the atoms of chlorine and tripositive tungsten, which permits a covalent bond to be formed between the two tungsten atoms in the complex. The one general chemical phenomenon with high specificity is closely analogous in both its nature and its structural basis to biological specificity: this phenomenon is crystallization. There can be grown from a solution containing molecules of hundreds of different species, crystals of one substance which are essentially pure. The reason for the great specificity of the phenomenon of crystallization is that a crystal from which one molecule has been removed is very closely complementary in structure to that molecule, and molecules of other kinds cannot in general fit into the cavity in the crystal or are attracted to the cavity less strongly than a molecule of the substance itself. Only if the foreign molecule is closely similar in size and shape and the location and nature of active (hydrogen bond-forming) groups to the molecule it is replacing will it fit into the crystal; and it is indeed found that the tendency to solid-solution formation depends upon the same structural features (such as replacement of a chlorine atom by a methyl group) as the tendency to serological cross reaction.

Examples of Biological Specificity

Many isolated examples of biological specificity and biological similarity determined by molecular size and shape and the detailed nature of intermolecular forces might be mentioned, such as the similarity in physiological (antipyretic-antineuralgic) activity of 4-isopropylantipyrine and 4-dimethylaminoantipyrene (pyramidon), which is clearly the result of the similarity in size and shape of the isopropyl group, and the dimethylamino group. I shall, however, discuss in detail only the specificity of enzymatic reactions.

From the standpoint of molecular structure and the quantum mechanical theory of chemical reaction, the only reasonable picture of the catalytic activity of enzymes is that which involves an active region of the surface of the enzyme which is closely complementary in structure not to the substrate molecule itself, in its normal configuration, but rather to the substrate molecule in a strained configuration, corresponding to the "activated complex" for the reaction catalyzed by the enzyme: the substrate molecule is attracted to the enzyme, and caused by the forces of attraction to assume the strained state which favors the chemical reaction—that is, the activation energy of the reaction is decreased by the enzyme to such an extent as to cause the reaction to

proceed at an appreciably greater rate than it would in the absence of the enzyme. This is, I believe, the picture of enzyme activity which is usually accepted.

Experimental data have not been gathered which permit the induction of so precise a representation of the structure and configuration of the active region of any enzyme as for the antibodies discussed above, but there do exist some data which support the general concept. If the enzyme were completely complementary in structure to the substrate, then no other molecule would be expected to compete successfully with the substrate in combining with the enzyme, which in this respect would be similar in behavior to antibodies; but an enzyme complementary to a strained substrate molecule would attract more strongly to itself a molecule resembling the strained substrate molecule than it would the substrate molecule. Examples of this behavior have been found: the hydrolysis of benzoyl-l-tyrosylglycine amide by either chymotrypsin or papain was found by Bergmann and Fruton (2) to be practically completely inhibited by an equal amount of benzoyl-d-tyrosylglycine amide. This suggests that the strained configuration of the l-isomer during the enzymatic hydrolysis is somewhat similar to the normal configuration of the d-isomer.

More extensive quantitative studies of inhibition of enzyme activity might well provide very interesting information about the configuration of the enzyme molecules. Carl Niemann and I have studies of this kind under way.

It is highly probable that many chemotherapeutic agents exercise their activity by acting as inhibitors to an enzymatic reaction through competition with an essential metabolite of similar structure. It was shown by Woods (16) in 1940 that the bacteriostatic action of sulfanilamide results from an inhibitory competition with p-aminobenzoic acid, and can be overcome by increasing the concentration of the latter substance. The metabolite and its inhibitor are closely related in molecular shape, differing in the replacement of a carboxyl group by a sulfonamide group. Other pairs in which a carboxyl group is replaced by a sulfonic acid or sulfonamide group are nicotinic acid and pyridine-3-sulfonic acid or its amide (7), pantothenic acid and pantoyltaurine (14) and the α-aminocarboxylic acids and the corresponding α-aminosulfonic acids (8).

An interesting case of inhibition is that of thiamine by pyrithiamine (13), the corresponding substance with the 6-membered pyridine ring in place of the 5-membered thiazole ring. The effective competition of pyrithiamine with thiamine for combination with the enzyme or other macromolecule involved might well have been predicted from the known cross reactivity of aromatic 5-membered rings containing sulfur and 6-membered rings not containing sulfur, as is strikingly shown by the formation of solid solutions

by thiophene and benzene. An analogous situation has been reported (15) by D. S. Tarbell of the University of Rochester. He has found that any substitution in the benzenoid ring of 2-methylnaphthoquinone destroys its vitamin K activity, but that the substance with a sulfur atom in place of —CH=CH— in the benzenoid ring retains this activity. These facts indicate that in the process of exerting vitamin K activity the benzenoid end of the molecule must fit into a pocket carefully tailored to it; that the other end is not so surrounded is shown by the retention of activity on changing the alkyl group in the 2-position. On the other hand, the failure of pyrithiamine to replace thiamine as a metabolite indicates that the sulfur atom of the thiazole ring in thiamine not only is effective in binding the molecule into its seat of action but also takes part in some way in the subsequent chemical reactions involved in the metabolic process.

Many Sciences Cooperate

The complete understanding of physiological activity will require consideration not only of molecular structure and weak intermolecular forces, but also of the chemical reactivity of the substances and of such other properties as solubility in different phases and degree of ionization, as well as of those properties of living organisms which may long defy simplification to chemical description; the importance of the problem for practical medicine as well as for fundamental biology is so great as to justify the attention and effort of many workers, in various fields of science, through whose cooperative effort the solution will some day be found.

Literature Cited

(1) Alexander, J., *Protoplasma*, **14**, 296 (1931).
(2) Bergmann, M., and Fruton, J. S., *J. Biol. Chem.*, 138, 124, 321 (1941).
(3) Brand, E., Saidell, L. J., Goldwater, W. H., Kassell, B., and Ryan, F. H., *J. Am. Chem. Soc.*, 67, 1524 (1945).
(4) Breinl, F., and Haurowitz, F., *Z. physiol. Chem.*, 192, 45 (1930).
(5) Corey, R. B., *Chem. Rev.*, 26, 227 (1940).
(6) Crowfoot, D., and Rogers-Low B., mentioned in *Science*, 102, 627 (1946).
(7) McIlwain, H., *Brit. J. Exptl. Path.*, 21, 136 (1940).
(8) McIlwain, H., *J. Chem. Soc.*, 1941, 75; *Brit. J. Exptl. Path.*, 22, 148 (1941).
(9) Mirsky, A. E., and Pauling, L., *Proc. Nat. Acad. Sci.*, 22, 439 (1936).
(10) Mudd, Stuart, *J. Immunol.*, 23, 423 (1932).
(11) Pauling, L., *J. Am. Chem. Soc.*, 62, 2643 (1940).
(12) Pauling, L., and collaborators, *J. Am. Chem. Soc.*, 68, 250 (1946), and earlier papers.
(13) Robbins, W. J., *Proc. Nat. Acad. Sci.*, 27, 419 (1941); Woolley, D. W., and White, A. G. C., *J. Exptl. Med.*, 78, 489 (1943).
(14) Snell, E. E., *J. Biol. Chem.*, 139, 975, 141, 121 (1941).
(15) Tarbell, D. S., private communication to author.
(16) Woods, D. D., *Brit. J. Exptl. Path.*, 21, 74 (1940).