Structure of Proteins and of Certain Physiologically Active Compounds

By Dr. D. M. Wrinch, Mathematical Institute, Oxford

RECENT studies have discovered the fact that a number of classes of physiologically active compounds (for example, the carcinogens, the sterols and bile acids, the sex hormones, the heart poisons) have all a similar framework[1]. Two questions then arise. (1) What is the significance of this similarity of structure? (2) Why do these compounds which have certain similarities of structure possess physiological properties so different as those, for example, of the sex hormones and the cardiac poisons? In answer to these questions we suggest that for certain types of interaction with the organism it is necessary (though not sufficient) that a molecule possess an affinity for, in the sense of superposability on, some 'substrate' in the organism. This suggestion, which simply gives precise expression to an idea long since well established in physiological chemistry[2], accounts for the similarity of framework. The particular nature of the interaction may then be attributed to the degree of unsaturation (compare specially the carcinogenic hydrocarbons and cholesterol), to the occurrence of a lactone ring and to the occurrence and particular distribution of carbonyl, hydroxyl and methyl groups in the compounds in question.

In the course of an orderly search for some suitable substrate sufficiently widely distributed through the organism and yet possessing a common molecular surface pattern or mosaic, attention is naturally directed first to the proteins. More than any other chemical substance proteins have an inevitable connexion with living matter; wherever there is living matter there proteins are found; wherever there are proteins there is or was living matter[3]. Undeniably, proteins are connected with the physicochemical changes which occur during the life of an organism. They form a more or less permanent part of it, since only under conditions of starvation are they oxidized to release the energy required to perform the work of the cell[4]. Furthermore, many facts belonging to protein chemistry (for example, the interchangeability of proteins as substrates for certain enzymes) suggest that proteins can possess a common surface pattern. For these reasons there is an inherent plausibility in postulating that the substrate whose mosaic is to have a geometrical affinity to the classes of compounds mentioned above is protein*.

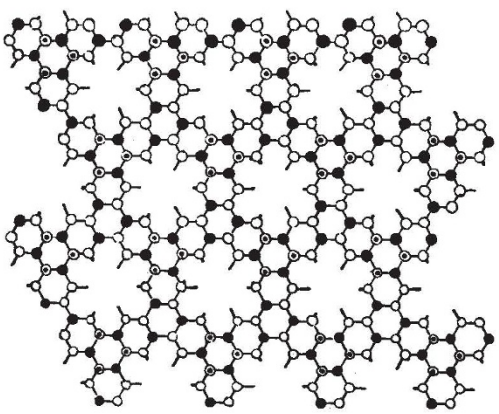

FIG. 1. The cyclol pattern. The median plane of the lamina is the plane of the paper. The lamina has its 'front' surface above and its 'back' surface below the paper.

● = N.
○ = C(OH), hydroxyl upwards.
⊙ = C(OH), hydroxyl downwards.
○— = CHR, direction of side chain initially outwards.
○- = CHR, direction of side chain initially upwards.

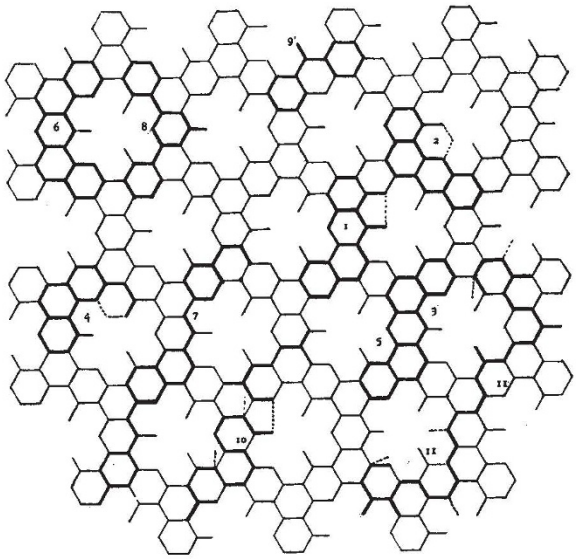

FIG. 2.

The fundamental fact about the structure of proteins is that they consist of condensations of amino (and imino) acid molecules. Leaving aside the intermediate stages, polypeptides or substituted diketopiperazines, which may or may not occur in biosynthesis, and which in any event yield an identical pattern by means of cyclization and polymerization respectively, we find that it is possible to construct a simple polycondensation pattern of the amino acids. This 'cyclol' pattern depends upon

* In a recent research on the effect of methyl cholanthrene on mouse fibroblasts *in vitro* (NATURE, **138**, 291; 1936), Hearne finds that this carcinogen induces chromosome pairing and chiasma formation in somatic cells. In view of the protein in the nuclear membrane and in the chromosomes themselves (see Wrinch, NATURE, **134**, 978; 1934. ibid., **135**, 788; 1935), this result affords some support for the suggestion that the 'substrates' upon which carcinogens act are protein in nature.

© 1936 Nature Publishing Group

one and only one assumption, itself a simple extension of Fischer's classical hypothesis, namely, that amino acids can condense by means of single, double and triple peptide links. Given this assumption, the 'cyclol' theory of protein structure follows by means of purely logical and mathematical arguments[5]. The plane form of the cyclol pattern is shown in Fig. 1. Owing to the lævo character of all the amino and imino acids in proteins[6], the cyclol pattern is a dorsi-ventral lamina in the sense that from one surface all the side chains emerge, whereas the other surface is free from side chains. Here, then, is further evidence that it is possible for proteins with the most diverse chemical composition to share one common surface pattern, namely, that on the surface free from side chains. A number of the requirements of protein chemistry are thereby met; and the type of substance which is *a priori* most likely to be involved in reactions with the physiologically active compounds listed above, turns out actually to possess a common mosaic. Affinity with this mosaic then seems likely to be the explanation of the similarity of structure of these compounds. We have now to consider the details of this mosaic and its relation to the molecular framework of these compounds. On this occasion the question is discussed with reference to the carcinogens. The considerations adduced may also be applied to the sterols, the sex hormones, etc.

It will give a fair picture of the situation as it stands to-day with respect to carcinogens if we consider the following substances, which are listed in a roughly decreasing order of potency: (1) methyl cholanthrene[7]; (2) 1,2-benzpyrene[8]; (3) 1,2,5,6-dibenzanthracene[9]; (4) 3,4-benzphenanthrene[10]; (5) 1,2,5,6-dibenzacridine[10]; (6) 3,4,5,6-dibenzacridine[10]; (7) 2-(*p*-aminostyryl)-6-(*p*-acetylaminobenzoylamino)-quinoline methoacetate[11]; (8) triphenyl benzene[12]; (9) tetraphenyl methane[12]. All these structures have a geometrical affinity to the cyclol pattern which is shown in Fig. 2 with its 'back' surface upwards so that all the side chains emerge below the plane of the paper. The frameworks of (3), (5), (6) and (7) are completely superposable on this mosaic: that of (2) and of (4) are superposable except for one atom, as shown by dotted lines. The case of methyl cholanthrene is interesting, in that complete superposability can be obtained by including an imino acid residue in the cyclol pattern. Alternatively, the framework is superposable, as shown in Fig. 2, except for the distortion of the five-membered ring, shown by the dotted line. Triphenyl benzene is superposable to the extent of 21 of its 24 carbons: furthermore, it can attain this superposability in three different ways. The position with respect to tetraphenyl methane is rather different since this compound is not even approximately plane. In Fig. 2, (9) shows part of its skeleton: 14 of its 25 carbons are superposable in the cyclol pattern. Since this molecule is non-planar[13], the possibility of the superposition of these 14 atoms on the cyclol pattern requires that the cyclol pattern shall not be plane at that point, but take up an orientation in which two neighbouring diazine rings do not share the same median plane, but have median planes making the tetrahedral angle with one another. The affinity of (9) to proteins is therefore not to the plane cyclol pattern as shown in Figs. 1 and 2, but to the general three-dimensional cyclol pattern, which has already proved useful in devising structures for the globular proteins[14].

This preliminary study of the affinities of the carcinogens for the cyclol mosaic presents two points of interest. Evidently a structure may be partially or even wholly superposable on the cyclol pattern without possessing the phenanthrene nucleus. Further, since the cyclol pattern need not be planar, there is no reason to expect that all carcinogens are planar. The substance (9), far from being an anomaly, may be the first of a new type of non-planar carcinogens which have a geometrical affinity to the non-planar cyclol pattern.

The suggestion that ordinal similarity of molecular structures is a necessary (though not sufficient) condition for certain types of physiological interaction, and that the special nature of these interactions is then determined by special atomic groupings, is in good accord with phenomena belonging to many fields of physiological chemistry. As an example we cite the case of the specific polysaccharides of types I, II, III, IV pneumococcus, which are antigenic when coupled to a protein, although the antigenic reaction is specific to the polysaccharides[15]. We may suppose that the importance of the protein constituent resides in its superposability on a protein substrate, the nature of the action being conditioned by the polysaccharide. It is perhaps appropriate to direct attention at this point to the far-reaching consequences which are logically entailed if, as is now suggested, proteins of different chemical composition can yet possess a common surface pattern.

The geometrical relationship of the cyclol pattern and the framework of the physiologically active compounds mentioned above further suggest that proteins may be the seat of biosynthesis. We suggest that cholesterol and carotene, for example, are formed in the organism by superposition on a cyclol template. The 'ingredients' may be pictured as long carbon chains which on superposition cyclize as shown in Fig. 2 (see (10) for cholesterol and (11) for carotene) or possibly as shorter carbon chains which then join up intermolecularly.

It appears that there is an important field for experimental work in the study of the processes of (intramolecular) cyclization and (intermolecular) polymerization with special reference to the use of large molecules of various types, including proteins, as templates. Such researches may throw light upon the mechanism at work in Nature which, from a vast array of possible structures, selects in fact only a few. Incidentally, they should confirm or refute the present suggestions, which amount to a simple interpretation of this mechanism in terms of the common mosaic of proteins predicted by the cyclol theory of protein structure.

[1] Fieser, "The Chemistry of Natural Products related Phenanthrene", New York (1936).

[2] Ing, *Science Progress*, **118**, 252 (1935). Marrack, "The Chemistry of Antigens and Antibodies", London, 1934.

[3] Jordan Lloyd, "Chemistry of the Proteins", London (1926).

[4] Jordan Lloyd, *Proc. Roy. Soc.*, B, **88**, 1 (1914).

[5] Wrinch, NATURE, **137**, 411 (1936); **138**, 241 (1936).

[6] Jordan Lloyd, *Biol. Rev.*, **7**, 256 (1932).

[7] Cook and Haslewood, *J. Chem. Soc.*, 428 (1934).

[8] Cook, Hewett and Hieger, *J. Chem. Soc.*, 395 (1933).

[9] Kennaway and Hieger, *Brit. Med. J.*, (i), 1044 (1930).

[10] Barry, Cook, Haslewood, Hewett, Hieger and Kennaway, *Proc. Roy. Soc.*, B, **117**, 318 (1935).

[11] Browning, Cohen, Cooper, Ellingworth and Gulbrauson, *Proc. Roy. Soc.*, B, **113**, 300 (1933).

[12] Morton, Clapp and Branch, *Science*, **82**, 134 (1935).

[13] George, *Proc. Roy. Soc.*, A, **113**, 587 (1927).

[14] Unpublished work.

[15] Morgan, *Chem. and Ind.*, **55**, 284 (1936).

© 1936 Nature Publishing Group