The Cyclol Theory and the 'Globular' Proteins*

By Dr. D. M. Wrinch

A NUMBER of facts relating to proteins[1] suggest that the polypeptides in native proteins are in a folded state[2,3]. The type of folding must be such as to imply the possibility of the regular and orderly arrangement of hundreds of amino acid residues, which to some extent at least is independent of the particular residues in question.

At present two types of folding have been suggested, the cyclol type[3,4] and the hydrogen bond type[5]. The search for other types of folding is being continued. So far, it has not proved possible to discard either theory on the grounds that the type of link postulated is out of the question. It is, therefore, very desirable to test these theories by checking their implications against known facts. Accordingly it is now considered whether the cyclol theory can stand the test of the body of facts relating to the 'globular' proteins, established by Svedberg and his collaborators[6].

The Cyclol Fabric

In previous communications the cyclol postulate has been applied directly to polypeptides, and a number of molecules, cyclol 6, cyclol 18 . . . and the general cyclol fabric (Fig. 1) have been suggested for consideration. These are all of one polyhexagonal type, in which the individual hexagons are alternately '2-way' diazine hexagons sharing opposite sides with triazine hexagons and '3-way' triazine hexagons sharing alternate sides with diazine hexagons. In default of information as to the sides and angles of these hexagons, a mean value is at present adopted for the C—C and C—N distances indifferently, and the tetrahedral angle δ as the valency angle for C. The valency angle for N is also taken as δ, since it has this value in hexamethylene tetramine[7], in which, as in the cyclol fabric, each N atom is joined to three C atoms.

In accordance with these assumptions, the individual hexagons in the cyclol fabric are at present taken to be 'crumpled' as in cyclohexane. The midpoints of the sides of a 'crumpled' hexagon are the vertices of a plane hexagon, and the geometrical problems can be simplified by taking these plane hexagons as the fundamental units in a 'median' network which represents the cyclol in a new way, each 'crumpled' hexagon being replaced by a 'median' hexagon.

In previous communications the cyclols have been considered only in the case when all the median hexagons lie in one common plane. With this limitation there has, of course, been no question of building a closed (that is, a space-enclosing) cyclol. To do so, it is necessary to investigate the conditions under which a cyclol fabric can bend about a line. Evidently it is permissible for two abutting median hexagons to lie on different planes, if the angle between the planes is the tetrahedral angle δ. Thus a cyclol fabric need not have a single median plane, but may turn about certain lines provided that the angle between abutting median planes is δ. The problem of the possible existence of space-enclosing cyclols can then be stated precisely as follows. Is it possible for a cyclol network to bend across one line after another so that it joins up and thus surrounds a portion of space? In other words, can a cyclol network be drawn, not on a plane but on the surface

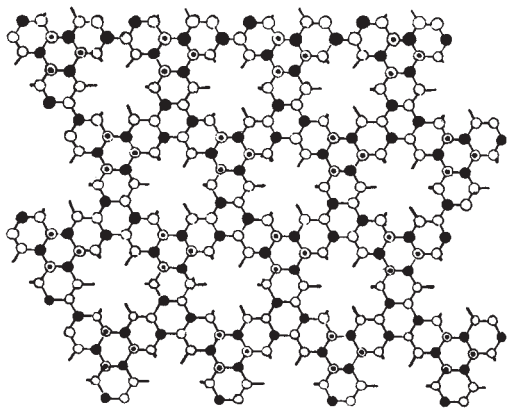

Fig. 1.

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), PEPTIDE HYDROXYL UPWARDS.
⊙ = C(OH), PEPTIDE HYDROXYL DOWNWARDS.
○— = CHR, DIRECTION OF SIDE CHAIN INITIALLY OUTWARDS.
○– = CHR, DIRECTION OF SIDE CHAIN INITIALLY UPWARDS.

of some polyhedron such that the faces at any edge crossed by the network abut at the tetrahedral angle?

The Closed Cyclols

To solve this problem, all the polyhedra in which some at least of the dihedral angles are equal to the tetrahedral angle will be considered in turn. As a first step it is remarked that among the regular and semi-regular polyhedra[8], only four satisfy the conditions. These are the truncated tetrahedron, the octahedron, the truncated octahedron and the skew triangular prism. On this occasion, as an example of this method of building megamolecules, attention is directed to the truncated tetrahedron, on which it has proved possible to draw closed cyclol networks. These networks form a linear series C_1, C_2, . . . C_n, . . . which comprise 72, 288, . . . $72n^2$, . . . amino acid residues. Figs. 2 and 3 show models of C_1 and C_2 in which the cyclol fabric is represented by the median hexagons. These models have 4 hexagonal faces, 4 triangular faces and 6 slits. The actual distribution of the (C—C—N) groups of atoms in the amino acid residues in the molecule can be inferred from the median hexagons, which are to be regarded simply as a shorthand notation.

The possibility of building closed cyclols on the other polyhedra mentioned above and on polyhedra which are not regular or semi-regular is being investigated. If they can be constructed they will also

* This article is based upon lectures given at the Harvard, Yale and George Washington University Medical Schools and at the Medical Centre, Columbia University, in October and November 1936. Communicated to the Editor of NATURE on March 10, 1937.

exist in linear series, each comprising numbers of residues represented by a quadratic function of the natural numbers 1, 2, . . . n, . . .

The cyclol hypothesis therefore predicts the existence of one or more series of 'space-enclosing' protein molecules, each series comprising numbers of residues given by a quadratic function of the natural numbers 1, 2, 3, . . . n, . . . and in particular of a series C_1, C_2, . . . C_n, . . . having the shape of truncated tetrahedra and comprising 72, 288, . . . $72n^2$, . . . residues.

These predictions are, it is claimed, confirmed by the results obtained by Svedberg and his collaborators[6].

(1) It is found that certain proteins are 'globular' molecules which, in appropriate circumstances, are monodisperse. The cyclol theory implies the existence of space-enclosing molecules containing certain specific numbers of amino acid residues: their polyhedral character is in accord with and offers an interpretation of the nature of this 'globularity'.

(2) It is found that certain molecules exist in different degrees of association in solutions of different *p*H, having a maximum molecular weight in a certain *p*H range and dissociating reversibly into molecules with submultiple molecular weights on one or both sides of this range. This is interpreted to mean that, in certain types of protein, the molecule will form multiple molecules at appropriate *p*H values by linkages between peptide hydroxyls or by salt linkages between side chains, the process of dissociation on changing the *p*H being reversible.

(3) It is found that the molecular weights of proteins are not distributed at random, but fall into a sequence of widely separated classes, the molecular weights in one class varying by as much as 15 per cent from a mean value. This is interpreted to mean that the proteins falling into one of these classes have a common structure as regards the arrangement of the constituent amino acids, and it is further suggested that each class connotes one closed cyclol network or an association of a certain number of such units. The variation in molecular weight within a class is then accounted for by the different selections of residues in the various proteins which can yield

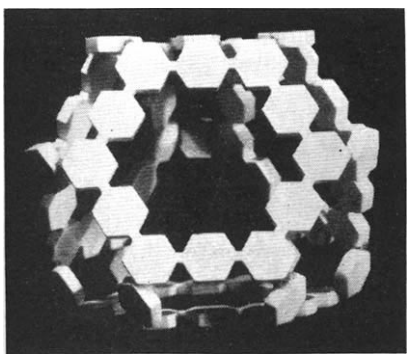

Fig. 2.

an average residue weight varying (say) from 100 to 135, which may also entail the presence of different numbers of water molecules in the molecule[9]. A further modification is also introduced if imino acid residues are present.

These tests are only qualitative. It would be of greater interest to apply more stringent quantitative tests. This will only be possible when data are available which give, for some of the 'globular' proteins, the shape of the molecule, the average value of the weights of the contained amino acid residues, the number of imino acid residues and the numbers of water molecules which form an integral part of the molecular structure.

It is, however, suggested for consideration in the future, that the group of proteins with molecular

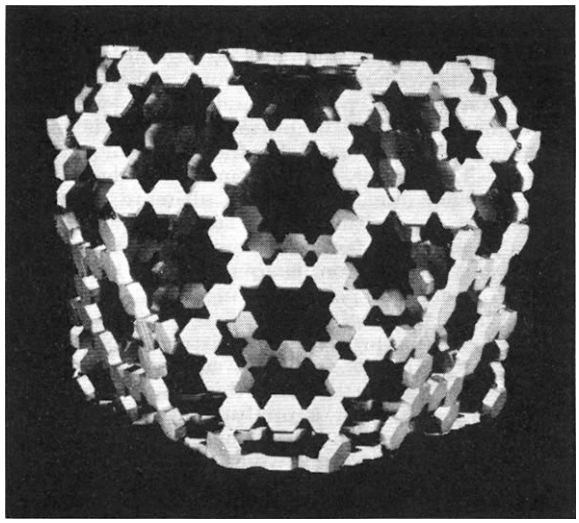

Fig. 3.

weights ranging from 33,600 to 40,500[10] are closed cyclol molecules of type C_2. This molecular weight class is of particular importance, since Svedberg has suggested that very many, possibly most, other proteins have molecular weights which are multiples of (say) 36,000.

Full details of this work will appear shortly.

Postscript added May 28, 1937: My attention has been directed to recent work by Bergmann and his collaborators, which strongly confirms the conclusion reached above. In particular, it may here be put on record that one of the predictions has now been verified: Bergmann and Niemann (*J. Biol. Chem.*, **118**, 301; 1937) deduce from the chemical analysis of egg albumin that this molecule (which belongs to the group of proteins with molecular weights in the neighbourhood of 36,000) consists of exactly 288 residues as predicted by the cyclol hypothesis. Full details of my work are in course of publication in the *Proceedings of the Royal Society*.

[1] Pryde, "Recent Advances in Biochemistry", chap. i (1931): Langmuir, Schaefer and Wrinch, *Science*, **85**, 76 (1937).

[2] Astbury, NATURE, **132**, 593 (1933). *Koll. Z.*, **69**, 340 (1934). Astbury and Street, NATURE, **126**, 913 (1930). *Phil. Trans. Roy. Soc.*, A, **230**, 75 (1931).

[3] Wrinch, NATURE, **137**, 411; **138**, 241: **138**, 651 (1936). *Proc. Roy. Soc.*, A, in the Press.

[4] Astbury, *J. Text. Inst.*, **17**, P. 281 (1936). NATURE, **137**, 803 (1936). *Chem. Weekblad*, **33**, 778 (1936).

[5] Jordan Lloyd, *Biol. Rev.*, **7**, 254 (1932). Mirsky and Pauling, *Proc. Nat. Acad. Sci.*, **22**, 439 (1936). Wrinch and Jordan Lloyd, NATURE, **138**, 758 (1936).

[6] Svedberg *et al.*, *Koll. Z.*, **51**, 10 (1930). *Trans. Far. Soc.*, **26**, 72 and 737 (1930). *Science*, **79**, 327 (1934). *Biol. Bull.*, **66**, 191 (1934). *Chem. Rev.*, **14**, 1 (1935), and a series of papers in *J. Amer. Chem. Soc.*, from 1929.

[7] Dickinson and Raymond, *J. Amer. Chem. Soc.*, **45**, 22 (1923). Wyckoff and Corey, *Z. Krist.*, **89**, 426 (1934).

[8] Andreini, *Soc. Ital. d. Scienze*, **14**, 75 (1907).

[9] Jordan Lloyd, *Biochem. J.*, **14**, 147 (1920); **21**, 1352 (1927); **25**, 1580 (1931). *Biol. Rev.*, **8**, 463 (1933).

[10] Svedberg and Eriksson-Quensel, "Tabulæ Biologicæ Periodicæ", **5**, 351 (1935–36).