The Structure of the Insulin Molecule

BY D. M. WRINCH AND IRVING LANGMUIR

The cyclol hypothesis has yielded a cage structure C_2 for the insulin molecule.[1,2,3] This structure has a skeleton framework lying on the surface of a truncated tetrahedron (hereafter called the C_2 polyhedron) whose faces are parallel to those of a regular octahedron having 6 corners S, as shown in Fig. 1a. The arrangement of the carbon and nitrogen atoms in the skeleton (cyclol fabric) which forms two of the faces of the C_2 polyhedron is shown in Fig. 1b.

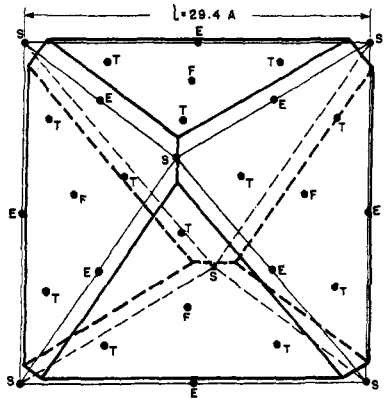


Fig. 1a.—The C_2 truncated tetrahedron (heavy lines) and the C_2 octahedron (light lines), showing on the front half the centers of lacunae at the midpoints of edges E, the centers of faces F and at points T midway between F and S.

The dimensions are determined by a, the only metrical parameter of the cyclol structure, which, being a mean between the C–C and the C–N bond lengths, was taken as $a = 1.50$ Å. The shortest distance between the centers of lacunae in the cyclol fabric is $4\sqrt{2a} = 8.48$ Å.; the length of the sides of the triangular faces of the C_2 polyhedron is $7\sqrt{6a} = 25.7$ Å.; the length of the slits which form the 3 short sides of the hexagonal faces is $\sqrt{6a} = 3.68$ Å., and the distance between the parallel opposite faces of the polyhedron is $16a = 24.0$ Å.

For many calculations we shall find it convenient to replace the C_2 polyhedron by an equivalent regular octahedron which we shall call the C_2 octahedron whose side length is $l = 8\sqrt{6a} = 29.4$ Å. This size is so chosen that the distance between the opposite faces is the same ($16a = 24.0$ Å.) as the corresponding distance in the C_2 polyhedron.

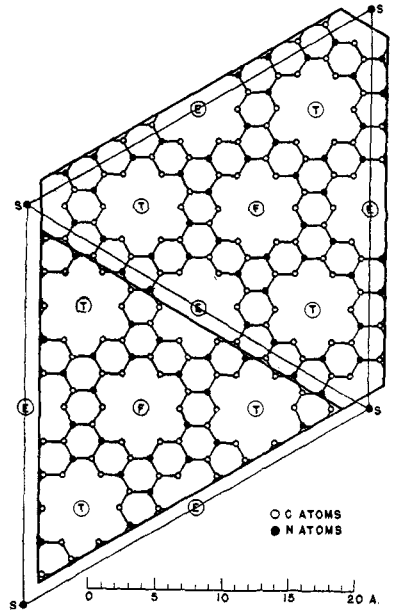


Fig. 1b.—The C, N skeleton of the cyclol fabric on two faces of the C_2 polyhedron, whose boundaries are given by heavy lines. The light lines are the projections of the boundaries of the faces of the C_2 octahedron which, after folding, lie 1 Å. above or below the faces of the C_2 polyhedron.

The faces and corners of the C_2 octahedron are shown in Figs. 1a and 1b. The distance SE between a corner S of a triangular face and E the center of an opposite side of the same triangle is $12\sqrt{2a} = 25.5$ Å., which is three times the shortest distance between lacunae. Thus there are lacunae at E, the midpoints of edges; at F, the centers of faces; and at T, the midpoints between F and S.

The faces of the C_2 polyhedron and the corresponding faces of the C_2 octahedron are parallel and are separated by the distance $(2/3)a = 1.0$ Å.; four of the faces of the polyhedron lie within and the other four lie outside the octahedron. The surface area and volume of the C_2 polyhedron are,

(1) D. M. Wrinch, *Nature*, **137**, 411 (1936); *Proc. Roy. Soc.* (London). **160A**, 59 (1937).

(2) D. M. Wrinch, *Nature*, **139**, 972 (1937); *Proc. Roy. Soc.* (London), **161A**, 505 (1937).

(3) D. M. Wrinch, *Science*, **85**, 566 (1937); *Trans. Faraday Soc.*, **33**, 1368 (1937).

respectively, $762\sqrt{3}a^2 = 2970$ Å.^{2} and $2003\sqrt{3}a^3 = 11{,}710$ Å.^{3}, which differ very little from the corresponding values for the C_2 octahedron $768\sqrt{3}a^2 = 2993$ Å.^{2} and $2048\sqrt{3}a^3 = 11{,}970$ Å.^{3}, respectively.

The final test for the correctness of a structure of molecules which form a crystal lattice is that it should be in agreement with deductions properly made from X-ray data. It was found[3] that the C_2 molecules, each having a trigonal axis parallel to the trigonal axis of the insulin lattice, fit the rhombohedral cell of the insulin lattice given by an X-ray analysis.[4] They can be arranged with any orientation δ in the corresponding hexagonal cell, and this single parameter was necessarily left undetermined until further information should be obtained from X-ray studies.

The necessary data are now available through Crowfoot's recent measurements[5] of insulin crystals (59 terms). Her data were summarized in five Patterson–Harker diagrams, which are reproduced herewith[6] in Figs. 1c to 5.

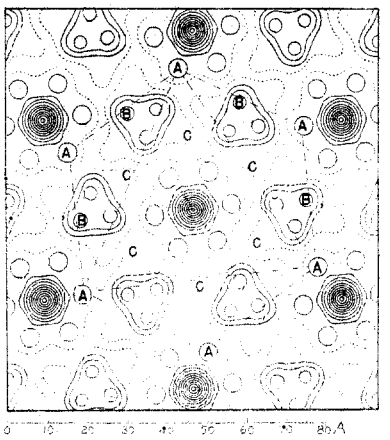


Fig. 1c.—P(xyo) projection of Patterson–Harker map for insulin. Origin in center.

Diagrams of this kind must be sharply distinguished from electron density maps which have played so large a part in establishing the structures of molecules containing relatively few atoms. The Patterson–Harker maps represent not distributions of matter in space but of vectors erected at a common point or origin, each vector giving in magnitude and direction the distance between two atoms in the original stucture. The Patterson–Harker diagrams obtained from the X-ray data are essentially sections or projections of a three-dimensional map showing the distribution of these vectors in space. They constitute an intermediate step between the X-ray photograph and the determination of the molecular structure. Whereas the electron density maps make it possible to confirm a structure directly, the Patterson–Harker diagrams need to be interpreted before they become of use.

In the case of the insulin lattice, Crowfoot has performed the transition from the X-ray pictures to the Patterson–Harker diagrams but has not accomplished their interpretation.

In a recent preliminary investigation,[7] however, it has been found that all the 18 peaks per molecule on Crowfoot's xy-plane projection (A, B, and C in Fig. 1c) lie in the positions to be expected if the insulin molecule is characterized by regions of high electron density at the 6 corners (S) of the C_2 octahedron shown in Fig. 1a.

Thus, taking in turn each of the 6 corners of the C_2 octahedron of side l, and drawing lines to the other 5 corners, we obtain 30 vectors. Erecting these from a common origin, we find that 6 give points at the corners of a similarly oriented octahedron of side $2l = 58.8$ Å. and the other 24 give two points at each of the midpoints of the 12 edges of the same octahedron. These points thus constitute the Patterson–Harker diagram of the S points of the C_2 octahedron.

If we project these 30 points upon a plane parallel to any face of the octahedron (the xy-plane), we find that 6 lie at the corners of a regular hexagon of side length $16\sqrt{2}\,a = 33.9$ Å., 2 at each midpoint of a side (lines connecting *adjacent* corners) and 2 at each midpoint of lines connecting *alternate* corners. A hexagon of this size has been drawn in Fig. 1c having its center at the origin with its axes turned through 6° corresponding to the previously undetermined[3] angle δ. The points, designated A, B, and C by Crowfoot, in Fig. 1c lie in positions which correspond to the projections of the Patterson–Harker diagram of the slits S of the C_2 molecule.

The Patterson–Harker diagram given by points

(4) D. Crowfoot, *Nature*, **135**, 591 (1935).

(5) D. Crowfoot, *Proc. Roy. Soc.* (London), **164A**, 580 (1938).

(6) The figures given by Crowfoot map the contours within a single rhombohedral cell. Since we wish to consider the molecular rather than the crystal structure, we have fitted together four photographic enlargements of each of Crowfoot's figures and have so obtained maps of the whole neighborhood around the central molecule. The lettering, the scales, and the polygons corresponding to the projections of the Patterson–Harker octahedron, of side $2l$, were drawn in on these composite enlargements and they were rephotographed for the figures of this paper.

(7) D. M. Wrinch, THIS JOURNAL, **60**, 2005 (1938); *Science*, **88**, 148 (1938).

which lie on or within the C_2 octahedron of side l thus consists of points on or within a similarly oriented octahedron (the PH octahedron) of side $2\,l$. In Fig. 2 the cross section of the PH octahedron, through its center and parallel to the xy-plane, has been drawn about the origin. This hexagon is obtained by joining the midpoints of adjacent sides of the hexagon in Fig. 1c. In Figs. 4 and 5 the projections of the edges of the PH octahedron on to the plane of the section are shown, neglecting the effect of the slight rotation of the PH octahedron through $\delta = 6°$.

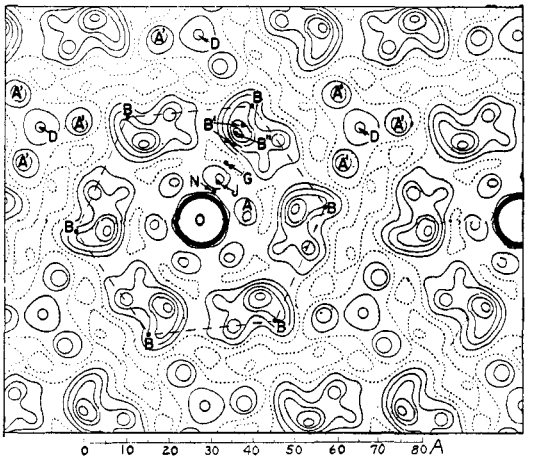


Fig. 2.—P(xyo) section of Patterson-Harker map for insulin.

Crowfoot used rectangular coördinates x, y, z to describe the position of peaks, and we shall describe the sections cut through the Patterson-Harker maps in this way; but, because of the trigonal symmetry of the lattice and molecule, we prefer for other purposes to use polar coördinates z, θ, ρ where θ is $\tan^{-1} y/x$ and ρ is the distance from the z-axis.

We shall use M to designate molecular centers: M_0 for that at the origin, M′ for the next one along the z-axis at $z = c = 30.9$ Å., and M_1, M_2, . . . M_6 for the 6 nearest neighbors near the x, y plane in the positions indicated in Fig. 3. Thus M_1 and M_6 are located, respectively, at $z = \pm 10.3$ Å., $\theta = \pm 30°$, $\rho = 43.2$ Å., while the positions of the others are obtained by successive rotations through 120°. The distance M_1 to M_3 is 74.8 Å.

The agreement we have noted between the projection of Fig. 1c and the projection of the PH octahedron suggests that the A, B, and C peaks represent the S points of the C_2 octahedron. However, the points on the PH octahedron are not all in one plane. The A and C points lie at $z = \pm 24$ and the B points at $z = 0$. The calculated coördinates are given in Table II.

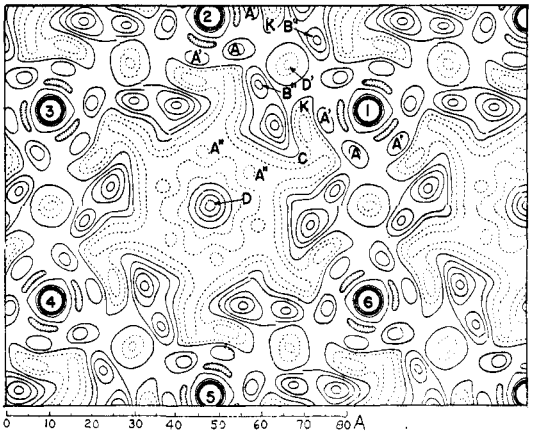


Fig. 3.—P($xy^1/_2c$) section of Patterson-Harker map.

It is thus important to examine the sections of Figs. 2–5 to see if the maxima whose projections give the A, B, and C points in Fig. 1c really lie at the proper values of z.

We are handicapped in this study by the fact that the sections of Figs. 4 and 5 are taken at $\theta = 0$ and 90°, while the maxima are to be expected at 6 and 96°. Figure 2 is a section which, for different molecules, corresponds to values of $z = 0$, and ± 10.3. Similarly, Fig. 3 gives $z = \pm 5.2$, and 15.5. None of these sections comes at $z = 24$, where the A or C peaks lie. However, the section at $z = 25.8$ is close enough to locate the peaks reasonably well.

Referring to Fig. 5, we see in fact that the A peak for M_0 lies close to the upper M_2 molecule at $z = +24$ and at a value of ρ only a few per cent. less than $\rho = 33.9$ which corresponds to the calculated PH octahedron (see Table II).

In Fig. 4 at $z = +3$ Å., $\theta = 0°$, $\rho = 10$ Å., there is a well-defined peak which we identify as the A peak belonging to M_1 (which lies at $z = -20.6$). With respect to M_1 this peak lies at $z = +23.6$. From Figs. 1c and 2 we see that its position with respect to M_1 corresponds to $\theta = 216°$, for which z should be $+24$.

The A peaks appear in their proper positions in Figs. 1c, 2, 3, 4 and 5, and these sections all show that the A peaks are comparatively isolated. Figure 4 shows, however, that in one direction, parallel to the z-axis, the A maximum is connected through a point J to another maximum A′, which, although really a separate peak, contributes to the

A peak seen in the projection of Fig. 1c. We shall discuss the origin of J and A′ later.

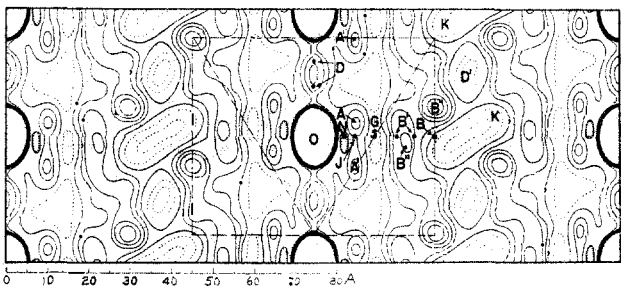


Fig. 4.—P(*xoz*) section of Patterson–Harker map. $\theta = 0°$.

The data prove that the A peaks lie alternately above and below the x, y plane at distances ± 24 Å., which is the distance between opposite parallel faces of the C_2 octahedron. We note from the section $\theta = 90°$ (Fig. 5) that the *molecule* at the origin must have one of its slits S at $\theta = 96°$, $z = +12$ Å., but there is no slit at -12 Å. for this value of θ. Since there is only one molecule in each rhombohedral cell, we conclude that all the molecules in the crystal lattice given by these diagrams are so oriented that an apex of the triangle forming the upper face of the C_2 octahedron is located at $\theta = 96$ but none at 276°.

The points in the PH octahedron which correspond to the B peaks lie at $z = 0$ on the corners of the hexagon shown in Fig. 2. There is, in fact, a prominent extension of the contours toward this point B, but the presence of other maxima B′ and B″ prevents us from estimating the exact contribution of B. We shall discuss the factors that cause these complications in this region.

The calculated coördinates for the C points in Table II give values of θ which are the same as for A, but the values of ρ are half as great, and z is of opposite sign although it has the same magnitude (24 Å.). Examination of Fig. 5 shows a maximum at $z = -27$, $\theta = 90°$, $\rho = 20$, not far from the calculated position of C at $z = -24$, $\theta = 96°$, $\rho = 17$. The agreement might be much better if the section had been taken at $\theta = 96°$.

Although the A, B, and C peaks in all the sections agree satisfactorily with the positions calculated from the slits of the C_2 octahedron, it will be recognized that the 6 slits provide a very incomplete description of the C_2 structure. It is reasonable to hope that the 62 additional maxima and minima per molecule listed in Table II, which are seen only in the sections of Figs. 2–5, correspond to other characteristic features of the C_2 molecule, such as the polyhedral nature of the skeleton. We may also expect to detect effects due to the 44 hexagonal lacunae in the positions shown in Figs. 1a and 1b by F, T, and E.

Density Deviations in Insulin.—Each molecule in the insulin crystal contains several thousand atoms of an average atomic weight of about 7. The heaviest atoms are 3 zinc atoms per molecule, and there are about 36 sulfur atoms. With the exception of these heavy atoms and a few groups of atoms having slightly higher or lower density than the average, we may look upon the volume distribution of electrons as being substantially uniform or at least as varying continuously in such a way as to show no fine structure. A slowly varying electron distribution would contribute to the general scattered X-ray radiation, but its effects would be eliminated automatically when the Patterson–Harker analysis is made by taking only photographic lines whose intensity is estimated visually. The finer structure given by the lines, which is the basis of the Patterson–Harker diagrams, can thus be regarded as representing positive and negative deviations from this assumed smooth volume distribution of electrons.

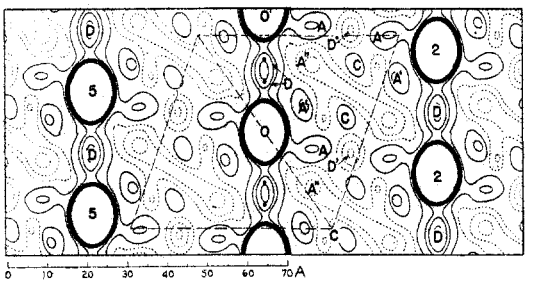


Fig. 5.—P(*oyz*) section of Patterson–Harker map. $\theta = 90°$.

In the protein molecule we may roughly assume 0.50 electron per unit of molecular weight, whereas in a water molecule we have 0.56 electron per unit. Considering that the density of the insulin crystal as a whole is about 1.32 and that of water is 1.0, it appears that water has an electron density about 16% lower than that of the average of the protein molecule. Thus low density may be expected in regions where the water content is high or where there are actual voids. Similarly, side chains comprising aliphatic or aromatic hydrocarbons would also correspond to minima, while the groups containing many oxygen or

TABLE I

DENSITY DEVIATIONS IN THE INSULIN LATTICE

Symbol	Description (L = lacunae)	No. per molecule	Coördinates Z, Å.	θ deg.	ρ Å.	Symmetry	Density	Relative magnitude (approx).
O	Center	1	0		0	...	$-o$	−4
S	Slit (corner)	6	±12	6 ∓ 30	17.0	3	$+s$	+5
E	L in edge	12	±12	6 ± 30	8.5	3	$-e$	−1
			0	6	14.7	6		
F	L in face center	8	±12	...	0	...	$-f$	−3
			± 4	6 ± 30	11.3	3		
T	Other L in faces	24	±12	6 ∓ 30	8.5	3	$-t$	−1
			± 8	(6 ∓ 6) and (6 ± 66)	12.3	3		
			± 4	6 ∓ 30	14.1	3		
Z	Zinc complex	6/2	± 5	± 30	21.6	3	$+z$	+4
U	Group near Z	3 or 6	± 5	40 ∓ 30	17	3	$+u$	+3
V	Group near Z	3 or 6	± 3	38 ± 30	22	3	$-v$	−5
Q	Midpoints between L	144	..	...	...	...	$+q$	+0.1

especially sulfur atoms correspond to maxima.

In Table I we have listed the points associated with the C_2 molecule where density deviations are to be expected. Thus, the center O of the molecule is probably a region of negative density deviation $-o$, since the side chains within the cage are not long enough to reach the center, a distance of 12 Å. from the surface of the C_2 octahedron. To the slits S we assign an intensity $+s$, since even apart from side chains the atoms forming the skeleton are crowded at these points. To the lacunae F, T, and E, presumably regions of negative deviations, intensities $-f$, $-t$, and $-e$ are assigned. Furthermore, we note from symmetry considerations, and for reasons already given,[3] that the zinc atoms are located near the midpoints along lines connecting the centers of adjacent molecules (except along the z-axis). These zinc atoms and the polar oxygen or sulfur atoms probably associated with them we designate by Z (intensity $+z$). We shall find later that there are two regions U and V having deviations $+u$ and $-v$, respectively, probably associated with side chains contributed by neighboring molecules which bind the zinc.

If the lacunae are regions of low density, the spaces between them must be of comparatively higher density. We therefore may consider that the midpoints Q between the centers of lacunae are regions having a positive density deviation $+q$.

The foregoing assumptions regarding density distribution are designed as a description of the characteristic features of the C_2 structure. The effects of individual side chains (except U and V) have been neglected. The C_2 structure is not actually an octahedron and has at least 2 kinds of trigonal faces (Fig. 1a). The trigonal symmetry characteristic of insulin could be preserved even if the C_2 molecule had faces of four kinds,[3] *i. e.*, with 4 different distributions of side chains. It may be hoped that the effects of these features may also ultimately be taken into account in analyzing the X-ray data.

Construction of Patterson–Harker Diagrams.—To consider the Patterson–Harker diagrams given by any distribution of atoms in space, we take first two points a and b of intensities α and β, respectively. The diagram associated with these two points comprises the vectors of a on a, of b on b, of a on b, and of b on a. Each has associated with it the product of the intensities of its two constituents. We therefore write

$$V([\alpha a + \beta b]^2) = \alpha^2 V(a^2) + \beta^2 V(b^2) + \alpha\beta V(ab) + \alpha\beta V(ba)$$

The terms $V(a^2)$ and $V(b^2)$ correspond to locations at the origin. It will be noticed that the cross terms $V(ab)$ and $V(ba)$ will be positive if α and β have the same sign, negative if they have different signs. Thus a positive cross term will result both when two points of positive intensity are concerned and when two points of negative intensity are concerned. On the other hand, the cross terms will be negative only when a point with positive intensity is associated with a point of negative intensity. These are located at positions ab and ba with respect to the origin.

The A, B, and C peaks that we have previously considered have the intensities s^2, $2s^2 - e^2$, and $2s^2 - e$ as given in Table II. It hardly seems possible at present to compare the relative intensities of the A, B, and C peaks to get the actual ratio in the coefficients.

Let us now consider the peak A′. This point

TABLE II

MAXIMA AND MINIMA IN THE PATTERSON-HARKER MAPS FOR INSULIN

The coördinates give the positions with reference to the center of the molecule to which they are associated.

Symbol	Character[a]	Vector	Symmetry	Coördinates z, Å.	Coördinates θ, deg.	ρ, Å.	Intensities
A	Max. I	OA	3	±24	6 ± 30	33.9	$+s^2$
A′	Max. I	1/3 OA	3	± 8	6 ± 30	11.3	$+4f^2$
A″	Min. I	1/2 OC	3	±12	6 ± 30	8.4	$-4es + 4e^2 + 4t^2 + 2eo$
B	Max. II	OB	6	0	6	29.4	$+2s^2 + e^2$
B′	Max. III	2/3 OB	6	0	6	19.6	$+2f^2$
	Max. III	5/6 OB	6	0	6	24.5	$+4t^2$
B″	Max. I	....	3 or 6	± 3	ca 38 ± 30	ca 22	$+2ov$
C	Max. II	OC	3	±24	6 ± 30	17.0	$+2s^2 + e^2$
D	Max. I	1/2 OM′	...	±12	...	0	$+fo$
D′	Min. I	1/2 OM	3	± 5	±30	21.6	$-2oz$
G	Min. I	1/2 OB	6	0	6	14.7	$-4es + 4e^2 + 4t^2 + 2eo$
J	Max. II	1/3 OB	6	0	6	9.8	$+2t^2$
K	Min. II	....	3	± 5	ca 40 ± 30	ca 17.	$-2ou$
N	Min. I	1/4 OB	6	0	6	7.4	$-32lq$

[a] In the second column of Table II, maxima or minima shown in the diagrams are classed as follows: I, well-defined isolated maxima or minima; II, maxima or minima whose effects are observable although partly masked by neighboring maxima or minima; and III maxima or minima which contribute to but are individually masked by neighboring terms.

(Fig. 5) lies close to the z-plane about one-third of the way from the origin to A (vector 1/3 OA). We note that in the C_2 octahedron of Fig. 1a the lines connecting F points in adjacent faces lie parallel to the lines S–S between opposite corners and are of one-third the length of S–S. There are four such vectors, corresponding to each of the diagonals between opposite points, so that the intensity of A′ corresponds to $4f^2$. This is necessarily positive and produces a maximum. The point A′ in Fig. 5 at $z = +8$, $\rho = 11.3$ Å. is, in fact, a maximum and happens to come almost directly above one of the A points, at $z = -4$, although it is unrelated since A′ belongs to M_0, while A belongs to M_2.

At point D′ there is a well-defined and very symmetrical minimum, shown in Figs. 3, 4, and 5. Since it lies at the midpoint along the line connecting adjacent molecules, it must receive equal contributions from these. This appears to result from the interaction between O and Z, giving a minimum of intensity $-2oz$ at D′. Since the two molecular faces adjacent to the zinc are not of identical structure, it is possible that the zinc atom is not exactly at the midpoint between the two molecules. However, the contours around the point D′ (Fig. 5) are not elongated along the line connecting the atoms, so that the zinc atom is probably not far from this midpoint. The observed spheroidal nature of the D′ minimum may be an effect of neighboring A maxima.

The maximum D seen in Figs. 3, 4, and 5 lies approximately halfway between centers of adjacent molecules along the z axis from which it must receive equal contributions. The interaction between O and F lacunae in the upper and lower faces gives a maximum of intensity $+fo$ at $z = \pm 12$ Å. Thus we recognize that the maximum D represents the overlapping of two peaks at $z = \pm 12$ Å., the elongation in the vertical direction being due to the composite nature of this maximum. The size of the maximum gives a rough estimate of the maximum dimensions of the low density region O. The D′ minimum is of about the same size as the D minimum, both being dependent on the dimensions of O.

The well-defined minimum at A″ (Figs. 3 and 5) corresponds to the vector 1/2 OC, and appears to be due principally to the interaction of E and S points along the edges between adjacent lateral faces of the C_2 octahedron. This gives an intensity $-4es$. However, we find that there are other terms which contributed to the intensity at this point: four vectors EE, four TT, and two DE, so that the total intensity is $-4es + 4e^2 + 4t^2 + 2eo$. The fact that we have here a minimum indicates that the $4es$ term must predominate.

This, and similar observations, enable us to make the rough estimates of the relative intensities of the various points given in the last column of Table I which seem consistent with all the observed maxima and minima of Table II. Accurate values of these intensities must probably await knowledge of the distribution of the side chains on the surface of the C_2 molecule.

The minimum G in Figs. 2 and 4 corresponds

to the vector $^1/_2$ OB and contains exactly the same terms as the A″ minimum and should thus have the same relative intensity.

Figs. 1c, 2, 3, and 4 show high vector concentrations in 6 regions symmetrically arranged about the z-axis. This "B complex" lies between $\rho = 19$ and $\rho = 31$ Å., covers the range of angles θ from $-20°$ to $+20°$, and extends over all values of z. Figure 5, which corresponds to $\theta = 30°$, $90°$, etc., shows that in these planes the B complex is wholly absent. We recognize several definite maxima within the B complex. The principal one, B″, seen best in Fig. 4, appears to be due to the interaction of the low density regions O and V. Since O is at the origin, the OV term is located at V (see Tables I and II). The very strong maximum B″ at $z = +7$, $\theta = 0°$, $\rho = 29$ belongs to M_1, which lies at $z = 10.3$ Å. Another of the B″ points, the one belonging to M_0, thus lies at $z = -3$ Å., $\theta = 8°$, and $\rho = 22$ Å. The point marked B″ in approximately this position in Fig. 4 is really 3 Å. in front of the true B″ maximum, and this fact accounts for its lower intensity as compared to the B″ point located at $z = +7$, $\rho = 29$.

This maximum corresponds to a strong negative density deviation near the zinc atom Z. Perhaps it is associated with the phenyl group of one or two tyrosine side chains which bind the zinc atom.

In Table II we have listed two other maxima (B′) which correspond to the vectors $^2/_3$ OB and $^5/_6$ OB and result from FF and TT terms. They contribute to the B complex at $z = 0$ and ± 10.3.

The rest of the B complex, extending as it does over all values of z, is most readily interpreted as due to the interactions of T lacunae with other rather distant T, F, or E lacunae. With a total of 44 E, F, and T lacunae, there would be over 1900 vectors connecting them. We have attempted in Table II to consider particularly the single terms which are of greatest intensity because they contain the more important factors S, F, O, or U or because they involve a large number of contributing vectors as a result of symmetry. The terms of most significance have been those between extreme points in the molecules, such as the S points, for the Patterson–Harker map of these points has a simple structure with widely separated points.

We have naturally not attempted to analyze in detail all the terms involving T, E, and F. It does appear, however, from numerous trials that a particularly large fraction of the vectors which involve the more distant lacunae give terms that lie within the B complex, well distributed over various values of z. The terms are individually of small intensity but add up to substantial values because of their large number.[8]

We have now considered in Table II 80 well-defined points shown in the Crowfoot diagrams. These include nearly all points which might be expected to have high intensities. The total number of points per molecule for which we have assumed characteristic density deviations (neglecting the Q points), as shown in Table I, is 60 to 66 which corresponds to a total of about 3600 vectors. Many of these, however, coincide. We have not attempted to consider all the maxima and minima which would result from these vectors. We have considered a large number not given in Table II and in no case have we found that they conflict seriously with the Crowfoot diagrams. Some of the points come in positions where they are masked by neighboring maxima or minima, but many fall in positions where their effect is to modify the contour lines in a manner which improves the agreement with the Crowfoot diagrams.

It should be noted that the sections of the Patterson–Harker diagrams given by Crowfoot are chosen with reference to axes of the crystal lattice. The x, y projection and the sections in this plane show that the trigonal axis of crystal and molecule coincide, but that the hexagonal axes of the molecule do not coincide with those of the crystal but are turned through an angle about the z-axis of $\delta = +6°$.

The Patterson–Harker sections chosen with reference to the crystal lattice are, of course, useful in determining the symmetry of the crystal as a whole. For example, Crowfoot showed that the insulin crystal has symmetry corresponding to the R3 space group. She was not able, however, to determine the symmetry or orientation of the molecule itself.

These sections through the crystal lattice are also useful in determining the positions and the

(8) (Note added by I. Langmuir, August 14, 1938.) An alternative and perhaps preferable although nearly equivalent interpretation of the B complex is to regard it as being the result of the interaction of the slits S with a uniform positive density deviation over the faces of the C_2 octahedron. This gives a PH distribution very closely like that of the B complex. Upon this uniform positive surface distribution over the faces there is then to be superposed the negative deviations corresponding to the lacunae which have already been considered. This suggests that in Table I q should be taken larger than 0.1 so as to give a positive value for the average density deviation over the whole surface.

structure in the neighborhood of all atoms which lie midway between the adjacent molecules in the crystal. Thus Figs. 3, 4, and 5 give various sections through the maximum at D which is due to the overlapping of the EO vectors contributed by neighboring molecules along the z-axis. Furthermore, the maximum D′ is located approximately midway between other pairs of adjacent molecules and so furnishes direct information regarding the location of the zinc atoms and the groups which bind them to adjacent molecules. In other words, the crystal lattice sections give the location of the molecules and intermolecular structure.

To determine the intramolecular structure, however, sections should be chosen with reference to the axes of the C_2 octahedra. The projection of the x, y plane, in Fig. 1c, gives directly the value of δ. If sections parallel to the z-axis had then been taken at $\theta = 6°$, 21°, 36°, etc., simply related to the axes of the C_2 octahedron, the diagrams would have shown the intramolecular structure in far greater detail. For this purpose it would also be desirable to have x, y sections at the height $z = h = +24$ Å. and at halves, thirds, and sixths of this distance. To obtain these sections needed to bring out the molecular structure, it will not be necessary to obtain new X-ray data. These Patterson–Harker sections can be calculated directly from Crowfoot's data on the intensities of the X-ray reflections.

It has been shown[3] that the distance between faces of adjacent C_2 structures (not along the c-axis) varies from 7 to 13 Å. as δ varies. Since the zinc atoms and their attached groups lie between these faces, it appears that the angle δ is determined by the zinc atom. If we should substitute cadmium or if possible mercury for the zinc, we should expect to obtain different values of δ. This would change in marked degree the Patterson–Harker sections chosen with reference to the crystal lattice but would leave unchanged many of the intramolecular features characteristic of the sections which have been chosen with reference to the molecular orientation.

Attention should be called to the particular features of the Patterson–Harker diagrams near D′, the midpoint between neighboring molecules. The cross terms between the zinc atom and all the other points in the C_2 molecules reproduce about D′ as a center, an image of the whole of the C_2 structure of unaltered size and of intensity proportional to z. The minimum at D′ is merely the image of the hollow at the center of the C_2 molecule transposed to the point D′. If we could substitute a much heavier atom than zinc and so increase the factor z, this image should become more distinct and show more clearly all the features of the molecular structure.

We have examined the diagrams in the neighborhood of D′ to see if the 6 S points in this transposed image of the C_2 molecule are observable as maxima of intensity $+ sz$ in planes given by $z = \pm 5 \pm 12$ Å. We find that points in 3 locations are obtained: 2 of these fall at the edges of the B complex at $z = \pm 7$ and are masked by it. The third set of maxima lie with hexagonal symmetry at $\theta = 10°$ and $\rho = 5$ Å. and so appear to account exactly for the hexagonal form of the outermost of the closed curves that surround the molecular centers in Fig. 1c.

Just as we have seen that a high density point half way between molecular centers gives about that point an undistorted image of the whole molecule, so we must conclude that the vectors between a low density point O at the center of each molecule and all the other points in the molecule will produce another undistorted molecular image about O as a center but of intensity proportional to $-o$. The central part is, of course, lost because of the high central density, but the terms of intensity $-os$ resulting from the interaction of O and S should be present as minima in the diagrams. These 6 points occur at $^1/_2$ OA and in Fig. 5 all lie in the deep trough connecting the A″ points. The OE interactions we have already taken into account in points A″ and G.

The molecular image produced by the zinc around D′ could be greatly enhanced if a very heavy atom such as mercury could replace the zinc. It is conceivable that in this way simple structural analyses may be made of very complicated molecules.

The foregoing analysis shows that all the prominent features of the Crowfoot diagram are deducible from the C_2 structure whose octahedron has a side 29.4 Å. if $a = 1.50$ Å. During the development of this analysis we repeatedly found that as we introduced one by one the more delicate features of the C_2 molecule the more perfect became the concordance between the Patterson–Harker diagrams and the Crowfoot pictures. We feel therefore that these X-ray data, in giving so perfect a picture of the C_2 structure, provide the experimental basis for the cyclol theory.

Summary

Crowfoot's data on X-ray reflections from insulin crystals, as summarized in a set of Patterson–Harker vector maps, are found to correspond in great detail to the C_2 polyhedral structure predicted for the insulin molecule on the basis of the cyclol theory. The features of the insulin molecule confirmed by the X-ray data include: (1) six high density points at the corners of an octahedron of side 29.4 Å., which is the size given by the cyclol theory; (2) low density regions near the centers of the 44 lacunae in the cyclol fabric that are located in definite positions in the edges and faces of the polyhedron; (3) a comparative hollow at the molecular center, *i. e*, a cage structure; (4) three zinc atoms per molecule on lines connecting molecular centers (except along the *z*-axis) and low and high density groups attached to these zinc atoms outside the C_2 structure.

OXFORD UNIVERSITY, ENGLAND
RESEARCH LABORATORY
GENERAL ELECTRIC CO.
SCHENECTADY, NEW YORK

RECEIVED JULY 30, 1938

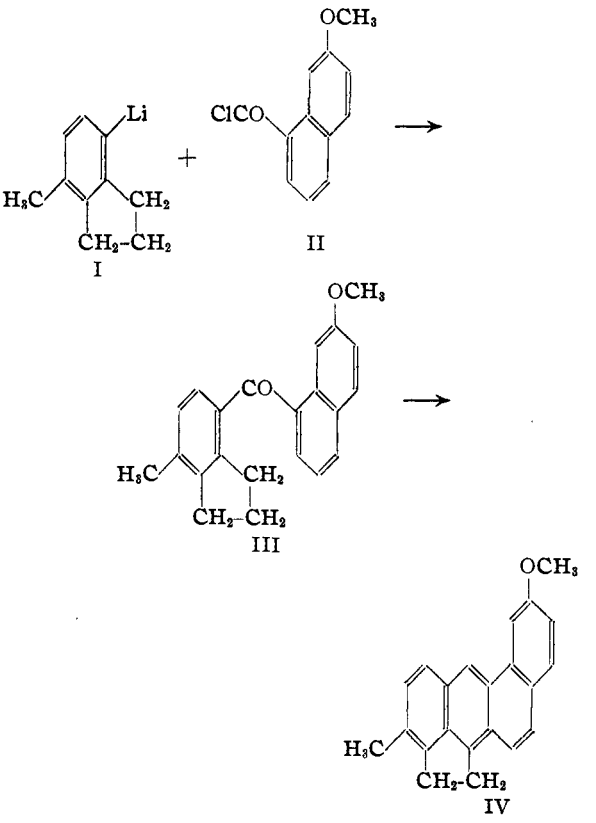
[CONTRIBUTION FROM THE CHEMICAL LABORATORY OF HARVARD UNIVERSITY]

The Synthesis of 2- and 6-Substituted Derivatives of 20-Methylcholanthrene[1]

BY LOUIS F. FIESER AND VICTOR DESREUX[2]

Derivatives of carcinogenic hydrocarbons having active functional groups are of interest to the general problem of attempting to correlate structure and biological activity, and if, among such compounds, derivatives can be found which possess marked carcinogenic activity they may be of value in providing a route to the preparation of interesting glycosides or conjugated proteins.

The only functional derivatives of methylcholanthrene previously known are 3-hydroxy-20-methylcholanthrene,[3,4] and its ether and acetate, 3-chloro-,[4] and 3-cyano-20-methylcholanthrene.[4] These were all made available by application of the Fieser–Seligman methylcholanthrene synthesis[5] to suitable methoxy- and chloro-substituted ketones, and the successful outcome of the syntheses shows that these substituent groups are capable of withstanding the pyrolytic conditions of the Elbs reaction when located in a position corresponding to the 3-position of methylcholanthrene. This is a β-position and is located in the terminal angular nucleus at a point rather remote from the site of ring closure, and it seemed likely that the other β-position 2 in this ring also would offer an environment favorable for the retention of an active functional group. Consequently, as a next step in the search for compounds of the type desired we synthesized the ketone III and investigated its behavior on pyrolysis. For the preparation of the ketone the lithium derivative



I from 4-methyl-7-chlorohydrindene[5,6] was condensed with the chloride of 7-methoxy-1-naph-

(1) This investigation was conducted as part of a program of research receiving support from the National Cancer Institute.

(2) Fellow of the Belgian American Foundation; Aspirant du Fonds National de la Recherche Scientifique Belgique.

(3) Cook and de Worms, *J. Chem. Soc.*, 1825 (1937).

(4) Fieser and Riegel, THIS JOURNAL, **59**, 2561 (1937).

(5) (a) Fieser and Seligman. *ibid.*, **57**, 942 (1935); (b) **58**, 2482 (1936).

(6) If the condensation of such a lithium derivative with a suitable nitrile constitutes a "modification of the original method of Fieser and Seligman," as stated by Cook and de Worms,[3] it may be noted that the modification was introduced not by these workers but by Fieser and Hershberg. *ibid.*, **59**, 394 (1937).