Acta Cryst. (1961). **14**, 459

Insulin. The Three-Dimensional Patterson Function for Insulin Sulfate Type *A* Crystals

BY CLARA BRINK SHOEMAKER*

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts, U.S.A.

AND J. RALPH EINSTEIN AND BARBARA W. LOW

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts, U.S.A.
and Department of Biochemistry, College of Physicians and Surgeons, Columbia University, New York, U.S.A.

(*Received* 9 *May* 1960 *and in revised form* 15 *July* 1960)

The three-dimensional Patterson function for the insulin sulfate type *A* crystal is presented. The Patterson function is discussed in terms of possible symmetry within the asymmetric unit and of the gross molecular structure.

Introduction

The X-ray studies of orthorhombic insulin sulfate crystals previously reported by Low and her associates include two- and three-dimensional Patterson functions of an air-dried crystal form (Low, 1952) and two-dimensional Patterson projections of each of two wet forms (Low & Shoemaker, 1959). The space group of all these forms is $P2_12_12_1$; there are two molecules (of molecular weight 5733) in the asymmetric unit. A tentative packing model for the insulin structure was early proposed on the basis of the Patterson functions of the air-dried form (Low, 1952, 1953*a*, *b*). The model structure may simply be described as made up of rod-like regions of high electron density (perhaps either coiled or folded peptide chain) in close-packed array parallel to the *a* axis. From their study of the Patterson projections of the wet crystal forms Low & Shoemaker concluded that these projections provide no clearcut evidence for the tentative model as originally described. Calculation of the three-dimensional Patterson function for the type *A* crystal form was undertaken in the hope that it would provide further information about the gross molecular structure.

Experimental

The type *A* bovine insulin sulfate crystals used in this experimental work have been described by Low & Shoemaker (1959). The wet crystals—with maximum dimension about 1 mm.—were mounted in thin-walled glass capillaries. Optical examination, crystal mounting and X-ray photography were carried out in the cold room in which the crystals were grown, at a temperature of 0 ± 2 °C. Equi-inclination Weissenberg photographs were taken with a modified Weissenberg camera (Low, unpublished) using Cu $K\alpha$ radiation and a film-holder of radius 57·3 mm. The multiple film method was employed, and, for some layers, two exposures of different duration.

Three crystals were used in all: one for the collection of intensity data from layers $l=0$–6, a second for layers $l=7$–12, and a third for putting the intensities from all the layers on the same scale. The unit-cell dimensions of these crystals (numbered I, II, and III respectively) are given in Table 1. The lack of exact identity between the unit cell dimensions of these crystals is discussed below. The intensities were estimated by visual comparison with intensity strips. About 2000 non-zero intensities with minimum spacing 2·5 Å were measured for one octant of the reciprocal lattice. The data are virtually complete to about 3 Å. For spacings between 3 Å and 2·5 Å the intensities of about 250 reflections could be measured—that is, about 15% of the total number of possible reflections within this range.

Table 1. *Unit-cell dimensions*

Crystal	*a*	*b*	*c*
I	57·8 Å	50·9 Å	40·2 Å
II	57·8	51·9	38·9
III	58·2	51·4	38·7

The intensities were corrected with Lorentz-polarization factors by the method described by Cochran (1948). They were put on the same scale by comparison with intensities of the 1*kl* and 2*kl* layers, obtained from crystal III. Corrections were not made for absorption.

Intensity distribution and Patterson function

Intensity distribution—Wilson plot

The corrected intensities were arranged in order of increasing Bragg angle and averaged over small ranges of Bragg angle. A plot of $\log \langle I_o \rangle / \langle f_N^2 \rangle$ as a function of $\sin^2 \theta/\lambda^2$ is given in Fig. 1. The atomic

* Present address: Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts.

scattering factor of nitrogen f_N was used as a reasonable approximation to the average scattering factor; the values used were those given by McWeeny (1951). As was expected the experimental data do not follow the smooth theoretical curve (Luzzati, 1955). On the assumption that an equation of the form $\langle I\rangle = Kf_N^2 \exp(-2B\sin^2\theta/\lambda^2)$ is applicable to these data we may attempt to estimate the 'temperature parameter' B from the plot in Fig. 1. (Use of the term 'temperature parameter' is not meant to imply that the parameter B is that of the Debye–Waller theory.) A value of $B=30$ Å^2 seemed reasonable for use in removal of the origin peak, although a unique value for B is by no means indicated by this plot.

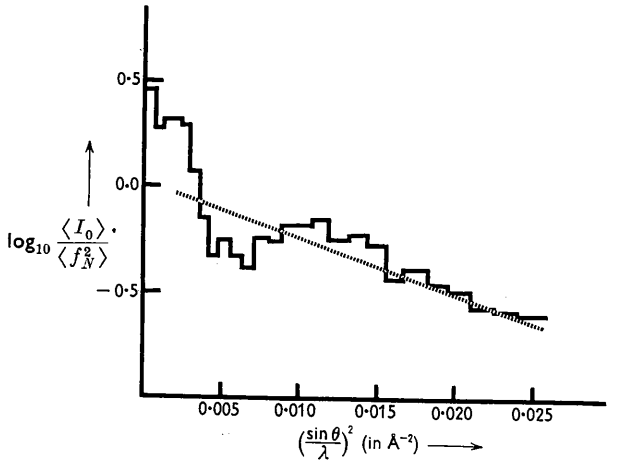

Fig. 1. Semi-logarithmic plot of $\langle I_o\rangle/\langle f_N^2\rangle$ against $\sin^2\theta/\lambda^2$. The broken line represents the theoretical Wilson plot corresponding to a 'temperature parameter' $B=30$ Å^2.

Removal of the origin peak

The origin region of the Patterson function for a protein crystal is considerably more complicated than that for crystals composed of small molecules. Besides the usual self-vectors of ordered atoms there are contributions—even at the origin itself—from vectors between nearest-neighbor atoms because of the large 'temperature parameter'. There are also contributions from disordered regions: the liquid between the molecules and perhaps parts of the protein molecules as well. It is impossible to remove the contribution to the origin peak from the self-vectors of ordered atoms because the number of ordered atoms is not known. However, for convenience in computation a peak can usefully be subtracted from the origin region. The following rather arbitrary procedure was used for this purpose. A function was subtracted from the Patterson series proportional to the Patterson function of a single nitrogen atom with a 'temperature parameter' estimated from the Wilson plot. The proportionality constant was adjusted so that $P_c(0)=0$, where $P_c=P-F(000)^2/V$. The method used to subtract this function was to subtract values of its transform at each reciprocal lattice point from the corresponding $|F|_o^2$. The transform was taken to be $k\varphi(S)$, where $\varphi(S)=f_N^2\exp(-15S^2)$ smoothed to become zero at $S=0{\cdot}36$ Å^{-1}; $\varphi(S)$ is plotted in Fig. 2. The factor k was derived from the equation

$$\sum_{h,k,l}{}' |F(hkl)|_o^2 = kV4\pi\int_0^{0\cdot36}\varphi(S)S^2dS\ .$$

The prime indicates that $F(000)^2$ has been omitted from the summation. The function thus subtracted is of course spherically symmetrical. Because of series termination it consists of a large positive central peak surrounded by spherical shells of small absolute value and alternating sign. The central peak decreases to

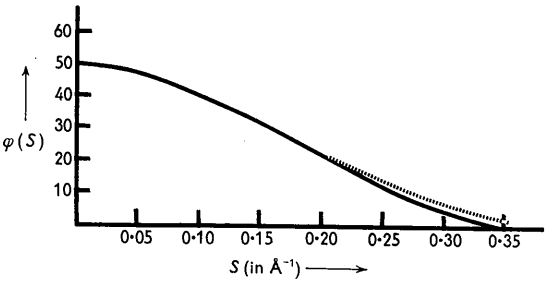

Fig. 2. Plot of $\varphi(S)$ against $S(=1/d)$. The function $\varphi(S)$ (solid line) was obtained from the function $f_N^2\exp(-15S^2)$ (broken line where not superposed) by altering it at high values of S.

zero at about 3 Å from the origin. There is a minimum of about $1\frac{1}{2}$ contour intervals (see below) at approximately $3\frac{1}{2}$ Å, and a maximum of about 2/3 contour interval at approximately 5 Å from the origin. At distances greater than 5 Å from the origin the absolute value of the subtracted function never exceeds a small fraction of one contour interval.

Calculation of the Patterson function

The Patterson function was calculated on I.B.M. machines using the M-card system (V. F. H. Schomaker, unpublished) at the Statistical Services Laboratory of Massachusetts Institute of Technology by Mr J. R. Steinberg. The intervals used were $a/60$, $b/60$ and $c/40$. The calculated sections are given in Figs. 3 and 4.

Insulin sulfate crystals, as other protein crystals, deteriorate after long exposure to X-rays, resulting in a general decrease in intensity of the diffraction pattern. This intensity change increases, in general, with increasing Bragg angle. Thus, intensity data taken from a crystal after different total times of exposure do not refer to the same state of the crystal. Each of the two crystals used for collection of intensity data was exposed for about 400 hr. in all, after which some deterioration had taken place. Therefore the practice of simply putting the data from all layers on a common scale, as followed here, involved the introduction of errors. These errors are probably not large

enough to affect the broad features of the Patterson function to any great extent.

A further source of error arises from the lack of exact identity between the unit cell dimensions of crystals I and II. By an unfortunate oversight the unit-cell dimensions of these crystals were not measured accurately until all the intensities had been measured. These two crystals came from the same batch, several crystals of which had previously been examined and had been shown to have nearly identical unit-cell dimensions. The dimensions for crystal II are close to average for the batch, but those of crystal I diverge from the average more than usual. The broad features of the Patterson function are probably not greatly affected by this circumstance. One basis for this assumption is that the large majority of the reflections of high intensity are among those measured using crystal I. To verify this assumption $h0l$ and $0kl$ Patterson projections were calculated using (*a*) coefficients with $l=0$–6 (crystal I), and (*b*) coefficients with all values of $l(l=0$–$12)$ (crystals I and II). For each projection in the Patterson functions (*a*) and (*b*) were closely similar to each other and to the corresponding projections calculated using complete data from one crystal alone (Low & Shoemaker, 1959). If the change in the unit-cell dimension along *c* between crystals I and II were accompanied by rotations of the molecules with respect to the axes, such close agreement would not be expected. Thus it seems likely that the lack of complete identity between crystals I and II brings about only a 'blurring' of the Patterson function.

Absolute scale

The absolute scale of the $|F|^2$ values and therefore of the Patterson function cannot be determined by the Wilson method. Of the criteria which must be satisfied if this method is to be used, two not satisfied here are (1) that the number of ordered atoms be known; (2) that the data go out to spacings of the order of 1·5 Å. Although a 'Wilson line' might be drawn by employing the Luzzati (1955) criteria, this would not lead to a determination of the absolute scale since there would still remain the problem of the number of ordered atoms. The absolute scale could of course be determined by comparison with standard crystals; this had not been done for type *A* insulin sulfate crystals. Following the collection of absolute intensity data (Traub & Hirshfeld, unpublished studies) for the isomorphous insulin citrate crystals (Low & Berger, 1960) we have been able to make a very rough estimate of the absolute scale of the insulin sulfate data by a comparison of Wilson plots for the two sets of data. The scale thus estimated for the Patterson function is, very approximately, 1 contour interval = 40 e.^{2}A^{-3}. In the discussion which follows this estimate of the absolute scale has been employed.

Discussion

The calculated Patterson function for type *A* insulin sulfate is shown in sections in Figs. 3 and 4. If $F^2(000)$ had been included in the calculation all heights would have been increased by 17,200 e.^{2}Å^{-3} or by approximately 430 contour levels. The function thus consists of a very high plateau modulated by small peaks and hollows. The origin peak, if not removed, would have been 95 contour levels above the average level $F^2(000)/V$. The highest peak outside the origin region is at $x=1/2$, $y=21/60$, $z=16/40$ and is 7·3 contour levels above the average level. The difference in level between this peak and the lowest hollow is only 3% of the average level of the plateau. As in Patterson functions of other protein crystals, the great relative height of the plateau is due to the overlapping of vector peaks, to a high 'temperature parameter', and to disorder in the unbound liquid of crystallization; perhaps parts of the protein molecules are also disordered.

The Patterson function contains the usual 5 Å shell, here somewhat extended along the *y* axis. There is a maximum within this region, lying at $x=0$, $y=3/60$, $z=2/40$, at a distance of $3\frac{1}{2}$ Å from the origin. Surrounding the 5 Å shell is a mainly negative region of irregular shape, (i.e., a region where the true value of the function is less than the average value), beyond which there are a number of positive peaks about 10–15 Å from the origin. Similarly located peaks are found in the Patterson functions for ribonuclease and hemoglobin (Magdoff, Crick & Luzzati, 1956; Perutz, 1949). The maxima near 5 Å and 10 Å correspond, respectively, to the broad maximum near 5 Å and to the shoulder near 10 Å in the Wilson plot.

Sections of the Patterson function on mirror planes (basal and Harker planes) exhibit greater peakiness than do other sections. Magdoff *et al* (1956) have explained this effect.

A remarkable feature of the Patterson function is the presence of several very extensive 'negative' regions. There is a negative channel running the full length of the *c* axis (except for a few positive bridges) and dividing the cell into two parts. The channel may be seen in projection (with several positive islands emerging from it) in the *c*-plane Patterson projection (Low & Shoemaker, 1959). The channel encloses a region about the *c* axis of variable cross-sectional shape, but typically extends from approximately $x=0$, $y=1/2$ to $x=1/3$, $y=0$ (sections $z=0$–$7/40$, $12/40$–$20/40$). On sections $z=8/40$–$11/40$ inclusive the region enclosed about the *c* axis narrows, the channel extending from approximately $x=0$, $y=1/3$ to $x=1/4$, $y=0$ (except for bridging on section $z=10/40$). The width of the channel is quite variable. It is most extensive near $x=0$, $y=1/2$, sections $z=0$–$7/40$; near $x=1/2$, $y=1/6$, sections $z=0$–$3/40$; and near $x=1/3$, $y=0$, sections $z=12/40$–$18/40$. The latter large negative regions are the main reasons for the appearance

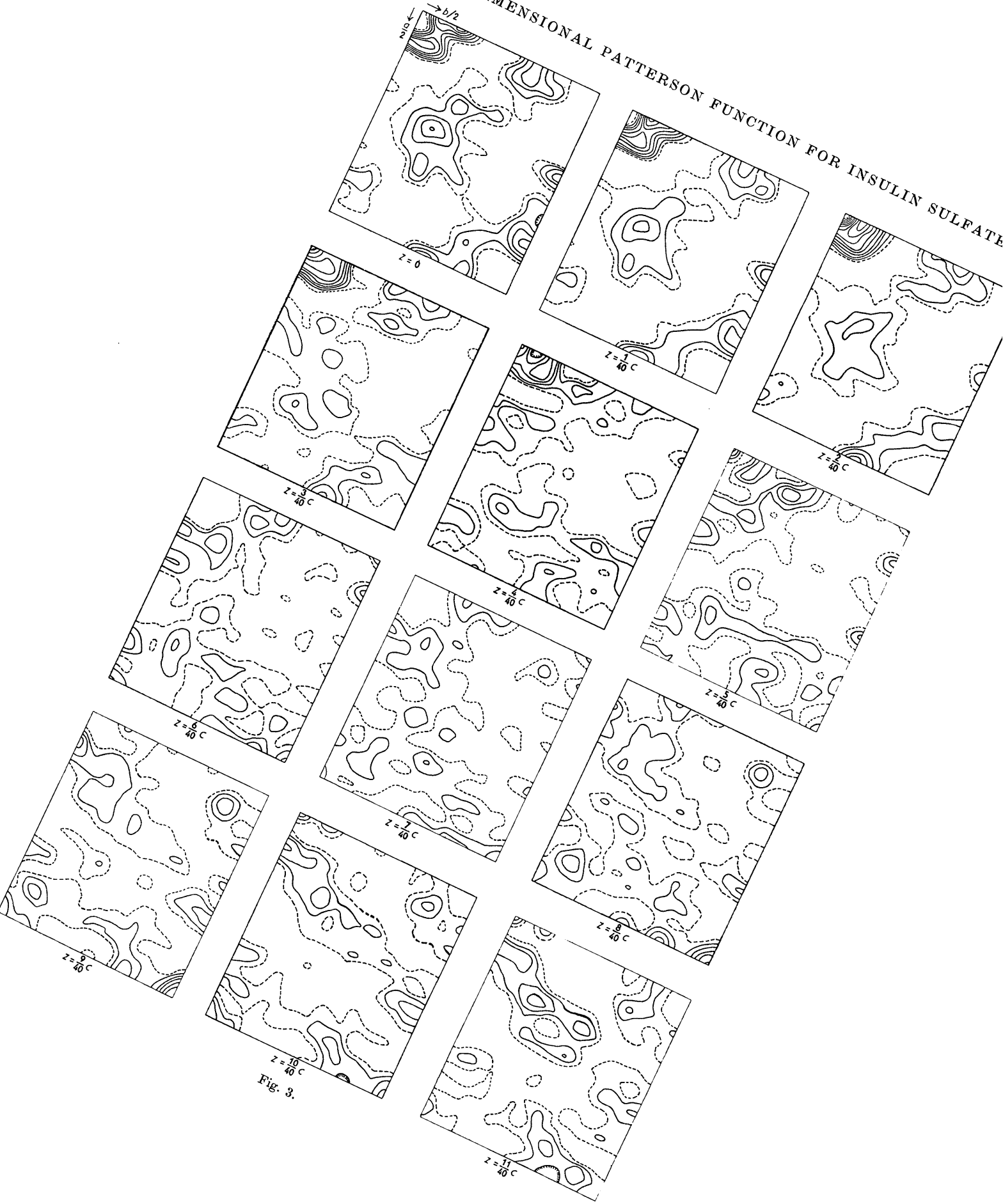
b/2
a/2
z = 0
z = 1/40 c
z = 2/40 c
z = 3/40 c
z = 4/40 c
z = 5/40 c
z = 6/40 c
z = 7/40 c
z = 8/40 c
z = 9/40 c
z = 10/40 c
z = 11/40 c

Fig. 3.

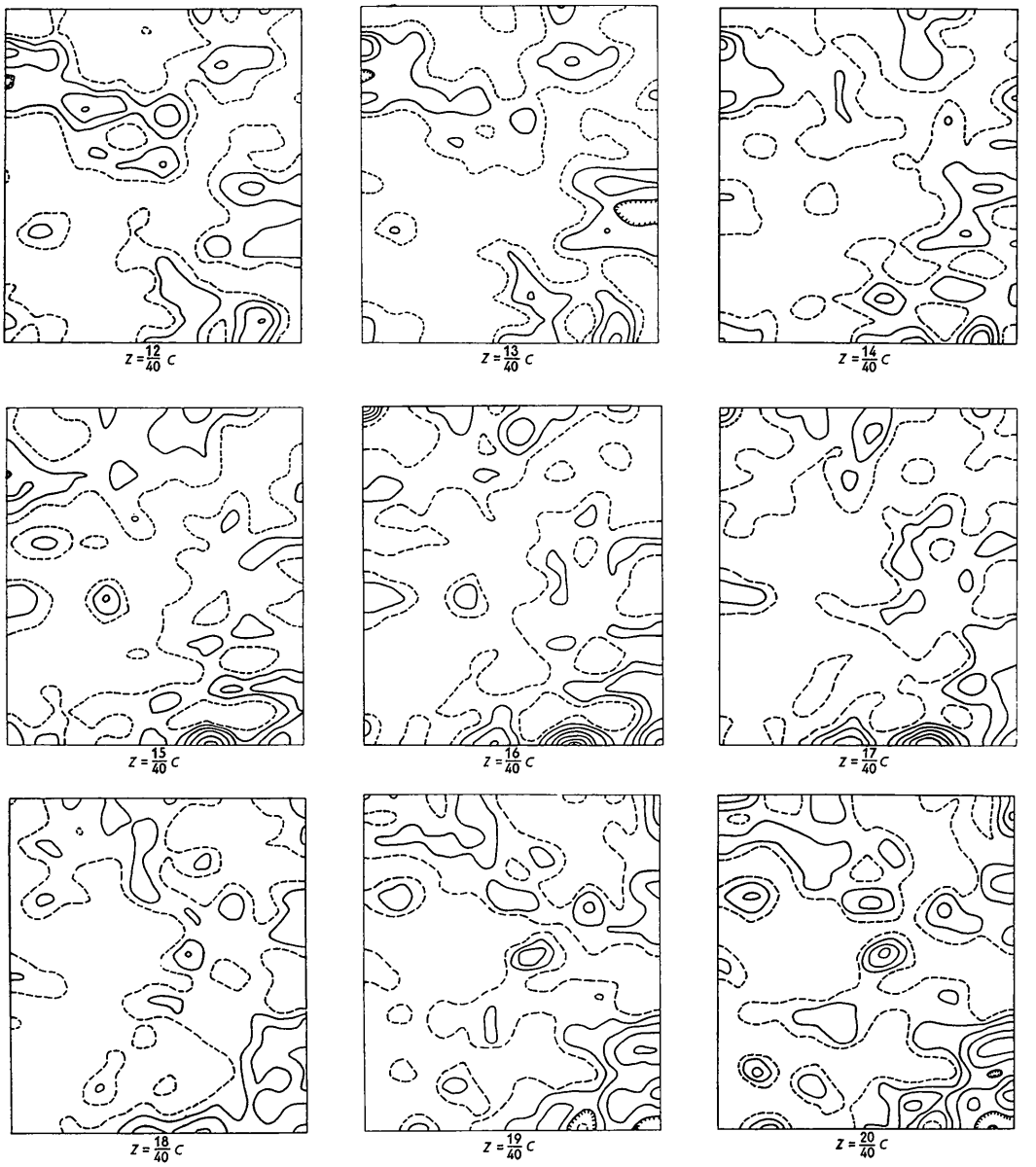

Fig. 3 (*cont.*).

Sections of the three-dimensional Patterson function. Contours are at approximately 40 e.^{2}Å^{-3}. The broken 'zero' contour represents the average level $F^2(000)/V$. 'Negative' contours are omitted.

of a negative channel in the a-plane Patterson projection.

Pseudo-origin peak

Low & Shoemaker (1959) noted that in the b-plane Patterson projection for type A insulin sulfate crystals, the highest non-origin peak, at $x=1/2$, $z=16/40$, has about it a peak distribution similar to that about the origin. The highest non-origin peak in the c-plane projection, at $x=1/2$, $y=21/60$, is also a pseudo-origin peak. There is a peak in the a-plane projection at $y=21/60$, $z=16/40$, but it is not the highest non-origin peak, nor does its position appear to be a pseudo-origin for the projection. (As has been noted in a previous paper, the highest peak in the a-plane projection is at $y=1/2$, $z=1/2$, and there are certain similarities between the peak distributions about this point and about 0, 0 (Low & Shoemaker, 1959). This apparent pseudo-centering is not related to nor explained by the discussion which follows.)

We have compared the peak distribution in the three-dimensional Patterson about the point 1/2, 21/60,

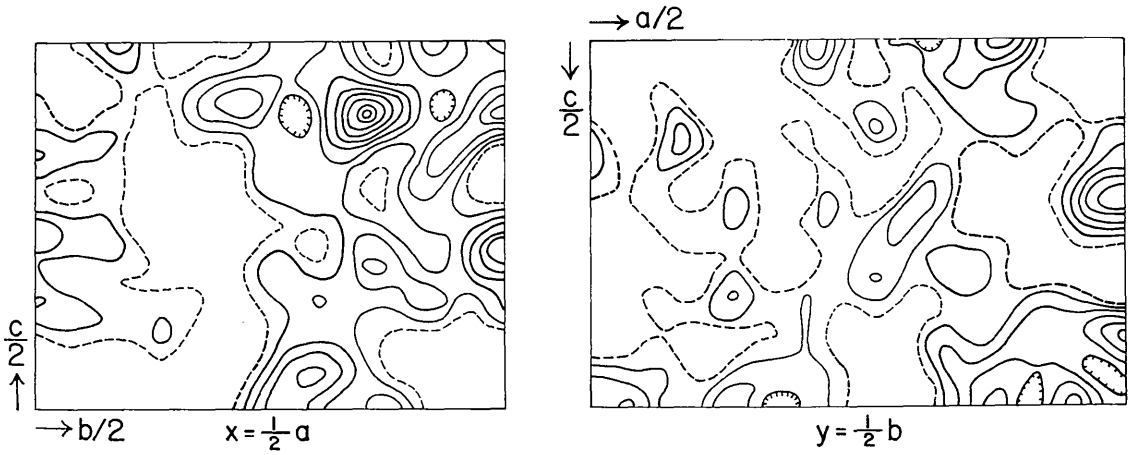

Fig. 4. Sections of the three-dimensional Patterson function at $x=a/2$ and $y=b/2$. Contours as in Fig. 3.

16/40 with that about 0, 0, 0. Although there is some correspondence, the two peak distributions are rather dissimilar.

These observations may be explained on the basis of a non-crystallographic two-fold or pseudo-two-fold axis which relates the two molecules in the asymmetric unit. A detailed exposition of all the arguments involved has been given in the Appendix by two of us (B. W. L. and J. R. E.). There it is shown that if the asymmetric unit of a crystal, space group $P2_12_12_1$, should contain a two-fold axis parallel to a at $y=\beta$, $z=1/4+\gamma$, then the Patterson function would have peaks at the equivalent positions $1/2, 1/2\pm2\beta, 1/2\pm2\gamma$, each peak having 1/4 the weight of the origin peak. Furthermore, the position of each peak would be a pseudo-origin, with pseudo-origin character masked heavily in the three-dimensional function and in the a-plane projection, but relatively slightly in the b-plane and c-plane projections. The effects observed are for values of $\beta=\pm4\cdot5/60$ and $\gamma=\pm2/40$. Since it is impossible to estimate the relative peak heights accurately we cannot distinguish between a two-fold or pseudo-two-fold axis, nor between an axis parallel to a and one somewhat inclined to the a axis. We can define approximate y and z coordinates of the two-fold or pseudo-two-fold axis as either (1) $y=4\cdot5/60, z=8/40$; (2) $y=4\cdot5/60$, $z=12/40$; (3) $y=-4\cdot5/60$, $z=8/40$; (4) $y=-4\cdot5/60$, $z=12/40$. It should be noted that this four-fold choice of y and z is simply dependent upon the absolute choice of origin for the co-ordinate system of the unit cell. There are two insulin molecules (of molecular weight 5733) in the asymmetric unit. Therefore the two-fold or pseudo-two-fold axis relates one molecule to another.

Without further evidence we might offer the above conclusion rather tentatively. However, Einstein & Low (1960) have found from completely independent evidence that the two molecules of the asymmetric unit of the type *A* crystal are very probably related by a two-fold or pseudo-two-fold axis. Given two independent sources of evidence this conclusion appears to be established.

Gross molecular structure

The Patterson function does not appear to provide any further information about the gross molecular structure and intermolecular packing. We have looked specifically for certain features which might be expected if, in the real structure, there were extensive lengths of α-helical coil packed parallel to each other. There are not any continuous regions of high vector density extending from the origin in any direction. Nor is there any vector from the origin along which there is a series of peaks at regular intervals of 5, 10, and 15 Å, etc., which would occur (Crick, 1956) if extensive lengths of α helix were packed parallel to each other. Further, the 5 Å shell shows very little anisotropy.

If the insulin molecule should contain relatively extensive lengths of α-helical coil, and these were not oriented even approximately parallel to each other, then it would be difficult to find evidence for them in the vector structure. Moreover, if there were relatively short lengths of α-helical coil, these too would be difficult to detect even with the most favorable orientation.

APPENDIX

Pseudo-origins in the Patterson function for space group $P2_12_12_1$ where the asymmetric unit contains a non-crystallographic two-fold rotation axis parallel to a crystal axis

Suppose that in a crystal structure having space group $P2_12_12_1$ there is a two-fold rotation axis, parallel to a and with coordinates $y=\beta$, $z=1/4+\gamma$, which relates the two halves of the asymmetric unit. (We note that if $\beta=\gamma=0$, the space group is $I2_12_12_1$ with a real origin at 1/2, 1/2, 1/2). Then there are eight general

symmetry-related positions, the coordinates of which are given below.

(1)	$(0, 0, 0) + (x, y, z)$	(5)	$(0, 2\beta, \frac{1}{2}+2\gamma) + (x, \bar{y}, \bar{z})$
(2)	$(\frac{1}{2}, 0, \frac{1}{2}) + (\bar{x}, \bar{y}, z)$	(6)	$(\frac{1}{2}, \overline{2\beta}, 2\gamma) + (\bar{x}, y, \bar{z})$
(3)	$(\frac{1}{2}, \frac{1}{2}, 0) + (x, \bar{y}, \bar{z})$	(7)	$(\frac{1}{2}, \frac{1}{2}-2\beta, \frac{1}{2}-2\gamma) + (x, y, z)$
(4)	$(0, \frac{1}{2}, \frac{1}{2}) + (\bar{x}, y, \bar{z})$	(8)	$(0, \frac{1}{2}+2\beta, \overline{2\gamma}) + (\bar{x}, \bar{y}, z)$

The set in each column are related by the space-group symmetry; pairs on the same line are related by the two-fold axis of the respective asymmetric unit. The eight equivalent atoms of type r will be labeled $r_1, r_2, \ldots, r_8$, where the coordinates of r_1 are $(0, 0, 0)+(r_x, r_y, r_z)$, those of r_2 are $(1/2, 0, 1/2)+(\bar{r}_x, \bar{r}_y, r_z)$, etc. The number of atom types $r, s, \ldots$ is $n/2$, where n is the number of atoms in the asymmetric unit of space group $P2_12_12_1$. The vector from atom r_j to atom s_k will be denoted $\mathbf{r}_j\mathbf{s}_k$, and the corresponding interatomic Patterson peak will be called an rs peak.

The resultant of a two-fold screw axis and a parallel two-fold rotation axis is a pure translation, which is the vector sum of the translational component of the screw axis, plus twice the vector distance between the two symmetry axes. Thus

$$\mathbf{r}_1\mathbf{r}_7 = \mathbf{r}_5\mathbf{r}_3 = \mathbf{s}_1\mathbf{s}_7 = \mathbf{s}_5\mathbf{s}_3 = \ldots$$
$$= \mathbf{T}_1 \text{ with components } (1/2,\ 1/2-2\beta,\ 1/2-2\gamma)$$
$$\mathbf{r}_2\mathbf{r}_8 = \mathbf{r}_6\mathbf{r}_4 = \mathbf{s}_2\mathbf{s}_8 = \mathbf{s}_6\mathbf{s}_4 = \ldots$$
$$= \mathbf{T}_2 \text{ with components } (1/2,\ 1/2+2\beta,\ 1/2-2\gamma)\ .$$

There are similar equations for the inverse vectors. The vectors $\pm\mathbf{T}_1$, $\pm\mathbf{T}_2$ will be called special vectors. The positions of their termini in the Patterson will be called special positions and denoted $\pm T_1$, $\pm T_2$.

There are two rr peaks, two ss peaks, etc., superposed at each of the four special positions, whereas there are eight rr peaks, eight ss peaks, etc., at the Patterson origin. Therefore each of the resultant peaks at a special position has 1/4 the weight of the origin peak. (We note that in the case $\beta=\gamma=0$, space group $I2_12_12_1$, these four peaks coincide at 1/2, 1/2, 1/2 to give a peak of the same weight as that at 0, 0, 0.)

Consider the distribution of all the interatomic Patterson peaks in relation to the origin and to the four special positions. It may be shown that each special position is a pseudo-origin for a large fraction of these peaks. If for example each (interatomic) peak is displaced by the vector $\mathbf{T}_1$, a large number are brought into exact coincidence with the original positions of other peaks, from which, with certain exceptions, they are indistinguishable. Displacing the rs peaks at $\mathbf{r}_3\mathbf{s}_j$, $\mathbf{r}_7\mathbf{s}_j$, $\mathbf{r}_j\mathbf{s}_1$, and $\mathbf{r}_j\mathbf{s}_5$ (where j designates any equivalent position),

$$\mathbf{T}_1 + \mathbf{r}_3\mathbf{s}_j = \mathbf{r}_5\mathbf{r}_3 + \mathbf{r}_3\mathbf{s}_j = \mathbf{r}_5\mathbf{s}_j$$
$$\mathbf{T}_1 + \mathbf{r}_7\mathbf{s}_j = \mathbf{r}_1\mathbf{r}_7 + \mathbf{r}_7\mathbf{s}_j = \mathbf{r}_1\mathbf{s}_j$$
$$\mathbf{r}_j\mathbf{s}_1 + \mathbf{T}_1 = \mathbf{r}_j\mathbf{s}_1 + \mathbf{s}_1\mathbf{s}_7 = \mathbf{r}_j\mathbf{s}_7$$
$$\mathbf{r}_j\mathbf{s}_5 + \mathbf{T}_1 = \mathbf{r}_j\mathbf{s}_5 + \mathbf{s}_5\mathbf{s}_3 = \mathbf{r}_j\mathbf{s}_3\ .$$

Each of these peaks is brought into coincidence with the original position of another peak of the same weight, except for cases in which either one or the other is a multiple peak. The symmetry gives rise to the doubling of peaks shown below. (The further multiplicities of rr peaks are unimportant in the cases to which these results are to be applied (many atoms per unit cell)).

$$\mathbf{r}_1\mathbf{s}_1 = \mathbf{r}_7\mathbf{s}_7,\ \mathbf{r}_1\mathbf{s}_5 = \mathbf{r}_7\mathbf{s}_3,\ \mathbf{r}_2\mathbf{s}_2 = \mathbf{r}_8\mathbf{s}_8,\ \mathbf{r}_2\mathbf{s}_6 = \mathbf{r}_8\mathbf{s}_4,$$
$$\mathbf{r}_3\mathbf{s}_3 = \mathbf{r}_5\mathbf{s}_5,\ \mathbf{r}_3\mathbf{s}_7 = \mathbf{r}_5\mathbf{s}_1,\ \mathbf{r}_4\mathbf{s}_4 = \mathbf{r}_6\mathbf{s}_6,\ \mathbf{r}_4\mathbf{s}_8 = \mathbf{r}_6\mathbf{s}_2\ .$$

Thus when the entire set of rs peaks, including in general 48 single and 8 double peaks, are displaced by vector $\mathbf{T}_1$, four double peaks are brought into coincidence with the original positions of four single peaks, four single peaks with those of four double peaks, and 16 single peaks with those of 16 single peaks. The special position $\mathbf{T}_1$ is a pseudo-origin for all interatomic peaks (any r, s) associated with vectors having origins at equivalent positions 3 or 7, and/or having termini at equivalent positions 1 or 5 (except for the differences in weight of some of the related pairs of peaks). This set of interatomic peaks comprises 7/16 of the total number of all the peaks, since 1/4 of all vectors have the required origins, 1/4 have the required termini, but 1/16 have both. The other special positions show similar pseudo-origin character.

In the three-dimensional Patterson the pseudo-origin character of the special positions is masked to a large extent, since each special position is a pseudo-origin for less than half the interatomic peaks. However, the pseudo-origin character of these positions is much more marked in two of the Patterson projections, those in which special positions coincide. In the b-plane projection T_1 coincides with T_2, and $-T_1$ with $-T_2$. In the c-plane projection T_1 coincides with $-T_2$, and $-T_1$ with T_2.

In the b-plane projection the 8 equivalent positions of the unit cell are translationally biperiodic with period $\mathbf{T}_1=\mathbf{T}_2$, henceforth denoted $\mathbf{T}$:

$$\mathbf{r}_1\mathbf{r}_7 = \mathbf{r}_2\mathbf{r}_8 = \mathbf{r}_5\mathbf{r}_3 = \mathbf{r}_6\mathbf{r}_4 = \mathbf{T} \text{ (any r)}\ .$$

Any atom in equivalent position 1, 2, 5, or 6 will be denoted type A, and any atom in equivalent position 3, 4, 7, or 8 will be denoted type B. Thus the type A and type B atoms each comprise two of the four asymmetric units (of space group $P2_12_12_1$). The two sets are related by the translational period $\mathbf{T}$. The Patterson function may now be considered to be made up of four parts: P_{AA}, P_{BB}, P_{AB}, and P_{BA}, where the subscripts refer to the origin and terminus, respectively, of the contributing vectors (i.e., P_{AB} represents the contribution from vectors with origins at type A atoms and termini at type B atoms). Because of the translational periodicity these functions are related as follows (where $\mathbf{u}$ is any position in vector space):

$$P_{AA}(\mathbf{u}) = P_{BB}(\mathbf{u})$$
$$P_{AB}(\mathbf{u}) = P_{AA}(\mathbf{u}-\mathbf{T}) = P_{BB}(\mathbf{u}-\mathbf{T})$$
$$P_{BA}(\mathbf{u}) = P_{AA}(\mathbf{u}+\mathbf{T}) = P_{BB}(\mathbf{u}+\mathbf{T}) .$$

That is, P_{AA} and P_{BB} are identical, P_{AB} is the same function displaced by $\mathbf{T}$ in the Patterson cell, and P_{BA} is the same function displaced by $-\mathbf{T}$.

We now consider the relationship between $P(\mathbf{u})$ and $P(\mathbf{u}+\mathbf{T})$, for a case in which the unit cell contains a great many atoms of nearly equal atomic numbers, arranged in fairly random fashion. The projected Patterson function and each of the four component functions defined above will then be quite smooth functions, especially when the temperature parameter is large.

$$\begin{aligned} P(\mathbf{u}) &= P_{AA}(\mathbf{u}) + P_{BB}(\mathbf{u}) + P_{AB}(\mathbf{u}) + P_{BA}(\mathbf{u}) \\ &= [2P_{AA}(\mathbf{u}) + P_{BA}(\mathbf{u})] + P_{AB}(\mathbf{u}) \\ P(\mathbf{u}+\mathbf{T}) &= [P_{AA}(\mathbf{u}) + 2P_{BA}(\mathbf{u})] + P_{BA}(\mathbf{u}+\mathbf{T}) , \end{aligned}$$

where the brackets have been inserted to draw attention to the close relationship between these functions. The functions enclosed in brackets represent, on the average, 3/4 of the value of $P(\mathbf{u})$ and $P(\mathbf{u}+\mathbf{T})$. Evidently they are the same except for the different weighting of P_{AA} and P_{BA} in the two cases. For ease in discussion they will be denoted

$$\begin{aligned} P' &= 2P_{AA} + P_{BA} \\ P'' &= P_{AA} + 2P_{BA} . \end{aligned} \qquad (1)$$

Then

$$\begin{aligned} P(\mathbf{u}) &= P'(\mathbf{u}) + P_{AB}(\mathbf{u}) \\ P(\mathbf{u}+\mathbf{T}) &= P''(\mathbf{u}) + P_{BA}(\mathbf{u}+\mathbf{T}) . \end{aligned} \qquad (2)$$

We wish to consider here the general locations of the broad maxima and minima of the Patterson function. (Such maxima and minima represent the superposition of a great many individual interatomic peaks. For example, in the case of type *A* insulin sulfate, there are probably on the order of 2000 ordered atoms in the unit cell, and therefore about 4×10^6 interatomic peaks. In any projection there are on the order of 2000 centers of interatomic peaks per Å^2.) We shall now demonstrate in a non-rigorous fashion that if $P(\mathbf{u})$ has a maximum at $\mathbf{u}_1$, then $P(\mathbf{u})$ very probably also has a maximum in the vicinity of $\mathbf{u}_1+\mathbf{T}$. The argument is based on the following considerations: (*a*) a maximum in P at $\mathbf{u}_1$ very probably is due mainly to a maximum in P' in that location; (*b*) given a maximum in P' near $\mathbf{u}_1$, there is very probably a maximum in P'' near $\mathbf{u}_1$; (*c*) given a maximum in P'' near $\mathbf{u}_1$, there is very probably a maximum in P near $\mathbf{u}_1+\mathbf{T}$.

The first of these considerations seems evident. If there is a maximum in P at $\mathbf{u}_1$ which is due to a maximum in P_{AB}, P' being fairly level, then there will be no corresponding maximum at $\mathbf{u}_1+\mathbf{T}$, except if, accidentally, P_{BA} has a maximum at $\mathbf{u}_1+\mathbf{T}$. However, the large majority of the maxima in P must be due mainly to maxima in P'.

Part (*b*) of the argument follows from the close relationship between P' and P'' (equation (1)). A maximum in P' may in general be due to a maximum in P_{AA} where P_{BA} is fairly level, to one in P_{BA} where P_{AA} is fairly level, or to a combination of maxima in both of the component functions. In the first two cases, there is obviously a maximum in P'' very close to that in P'. In the last case the positions of the resultant maxima in P' and P'' depend on the shapes and positions of the two component maxima. In general the highest points of both resultant maxima will lie somewhere between those of the component maxima. Therefore the maxima of P' and P'' should lie close together in many cases, i.e. those in which the maxima in P_{AA} and P_{BA} lie close together.

The considerations involved in part (*c*) are much the same as in part (*a*) and are based on the relative sizes of P and P''. Given a peak in P'' at $\mathbf{u}_1$, there will be a peak in P at $\mathbf{u}_1+\mathbf{T}$ except in those cases in which there is a sufficiently deep depression in P_{BA} at $\mathbf{u}_1+\mathbf{T}$ to counteract it.

The same argument may be applied to the positions of depressions in the Patterson projection.

Thus under the assumed conditions there would be a considerable degree of pseudo-centering of the *b*-plane projection of the Patterson function about $\mathbf{T}$ and about $-\mathbf{T}$, where $\mathbf{T} = \mathbf{T}_1 = \mathbf{T}_2$ (see above). The same would be true for the *c*-plane projection, where $\mathbf{T} = \mathbf{T}_1 = -\mathbf{T}_2$. There would be no such degree of pseudo-centering in the *a*-plane projection, since in that projection there is no translational biperiodicity within the unit cell.

It should be noted that these results will hold even if there are widely different temperature factors for different groups of atoms, or if there are large regions of complete disorder within the unit cell, as in wet protein crystals. There would still exist an exact translational biperiodicity for the ordered atoms in two projections. The part of the Patterson function due to vectors originating and terminating in the disordered regions must be almost level everywhere in the unit cell, and would therefore not interfere with the pseudo-centering of ordered-atom Patterson peaks.

Some of this work was carried out at Harvard University where it was supported by research grants G-1038 and G-1773 from the National Science Foundation. This investigation, at Columbia University, was supported by a research grant A 1320 (C1 and C2) from the National Institute of Arthritis and Metabolic Diseases, Public Health Service.

The authors are grateful to the Lilly Research Laboratory, Eli Lilly and Company, who supplied the insulin preparations. We wish to thank Dr J. E. Berger, who prepared the beef insulin sulfate crystals used. We wish also to thank Miss Nancy A. Wilson, Miss Celia Chen, and Mr J. A. Rosencrantz, who plotted the graphs and drew the contour maps.

References

CRICK, F. H. C. (1956). Unpublished study quoted in MAGDOFF *et al.* (1956).
COCHRAN, W. (1948). *J. Sci. Instrum.* **25**, 253.
EINSTEIN, J. R. & LOW, B. W. (1960). (In preparation.)
MAGDOFF, B. S., CRICK, F. H. C. & LUZZATI, V. (1956). *Acta Cryst.* **9**, 156.
MCWEENY, R. (1951). *Acta Cryst.* **4**, 513.
LOW, B. W. (1952). *Nature, Lond.* **169**, 955.
LOW, B. W. (1953*a*). *Nature, Lond.* **172**, 1146.
LOW, B. W. (1953*b*). In NEURATH, H. & BAILEY, K., Eds. *The Proteins*, Vol. I, Part A, Chapter 4. New York: Academic Press.
LOW, B. W. & SHOEMAKER, C. B. (1959). *Acta Cryst.* **12**, 893.
LOW, B. W. & BERGER, J. E. (1961). *Acta Cryst.* **14**, 82.
LUZZATI, V. (1955). *Acta Cryst.* **8**, 795.
PERUTZ, M. F. (1949). *Proc. Roy. Soc.* A, **195**, 474.

Acta Cryst. (1961). **14**, 467

A Bent Hydrogen Bond Model for the Structure of Ice-I

BY R. CHIDAMBARAM

Physics Department, Indian Institute of Science, Bangalore 12, *India*

(*Received* 2 *September* 1960)

A bent-hydrogen-bond model for the structure of ice-I is proposed which uses a value for the H–O–H angle close to the vapour value of 104½°. It is shown that while it is at least as consistent with the neutron-diffraction data of Peterson & Levy as the 'half-hydrogen' Pauling model, it is to be preferred from bond energy and spectroscopic considerations.

Recent studies on the structures of hydrated crystals by neutron diffraction and proton magnetic resonance indicate that the water molecule, when hydrogen-bonded into the structure, is not much distorted from its vapour configuration. Thus in oxalic acid dihydrate (Garrett, 1954) and sodium sesquicarbonate (Bacon & Curry, 1956), where the angles subtended at the donor water oxygen by the acceptor oxygens are respectively 84° and 114°, the corresponding H–O–H angles have been determined as 106° and 107°. This is to be expected, for it can be shown that the energy changes involved in the deformation of the H–O–H bond angle from the equilibrium value are at least several times higher than those involved in the formation of bent O–H · · · O hydrogen bonds. Hence, it would appear unlikely that in ice-I the H–O–H angle should increase by about 5° from the vapour value of 104½° to *exactly* the tetrahedral value, in order that linear O–H · · · O hydrogen bonds may be formed.

The main reason for the assumption in recent literature (for example, Frank, 1958) of a tetrahedral value for the H–O–H angle in ice is the neutron-diffraction investigation of hexagonal D_2O ice by Peterson & Levy (1957), who have shown that their diffraction data are in good agreement with the 'half-hydrogen' model of Pauling (1935). It is the purpose of this paper to suggest a modification of the Pauling model which uses a value for the H–O–H angle close to 104½°; to show that it is also consistent with the neutron-diffraction data; and then to advance some arguments in its favour.

If we suppose that the H–O–H angle is smaller than the tetrahedral value (*without being necessarily equal to the vapour value*), the so-called 'half-hydrogen' position will be split into three 'one-sixth hydrogen' positions distributed at the vertices of an equilateral triangle perpendicular to the O · · · O line, which will pass through the centroid of the triangle (Fig. 1). Here we have assumed that O · · · H–O–H · · · O lie in one plane (a reasonable assumption because for a given H–O–H angle this involves the minimum non-linearity in the hydrogen bonds) and that the non-linearity in the two hydrogen bonds from each oxygen is the same. It may be noted that this model retains the mean statistical space group D^4_{6h}–$P6_3/mmc$ of the 'half-hydrogen' Pauling model for hexagonal ice. If O–H = 1 Å and the angle H–O–H = 104½°, each 'one-sixth hydrogen' position is shifted from the O · · · O line by

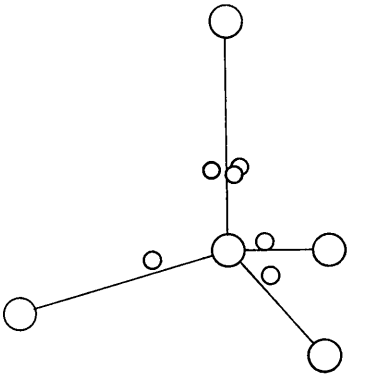

Fig. 1. Illustration of the splitting of a 'half-hydrogen' position into three 'one-sixth hydrogen' positions: ○–O, ○–⅙ H. (The splitting has been heavily exaggerated.)