Acta Cryst. (1952). 5, 701

An X-ray Investigation of Air-Dried Lysozyme Chloride Crystals: The Three-Dimensional Patterson Function

By Robert B. Corey, Jerry Donohue and Kenneth N. Trueblood*

*California Institute of Technology**, Pasadena, California, U.S.A.*

and Kenneth J. Palmer

Western Regional Research Laboratory†, Albany, California, U.S.A.

(*Received* 14 *April* 1952)

The intensities of reflections from 198 separate forms of (*hkl*) estimated from oscillation photographs taken about the *a* and *c* axes of air-dried lysozyme chloride crystals with Cu $K\alpha$ radiation were used to calculate the complete three-dimensional Patterson function. The minimum spacing observed was 5·7 Å.

Attempts were made to use the Patterson diagram as a source of information concerning the probable arrangement and configuration of the protein molecules, and to derive distributions of scattering matter in the unit cell of the crystal which would explain the principal features of the Patterson diagram. The results of these attempts are discussed.

A table of relative values of $|F|^2$ for all 198 observed reflections (*hkl*) is included.

Introduction

The protein lysozyme can be crystallized readily, both from basic solution near its isoelectric point and from solutions below pH 7·0 in the form of its salts (Alderton & Fevold, 1946). Some of these salts, especially the halides, possess properties which recommend them for study by X-ray diffraction methods. X-ray measurements of air-dried lysozyme chloride crystals have been used to compute a value 13,900±600) for the molecular weight of the anhydrous, chloride-free protein (Palmer, Ballantyne & Galvin, 1948). In the present investigation this work has been extended: the intensities of Cu $K\alpha$ reflections obtained from these crystals have been estimated and used for the computation of a complete three-dimensional Patterson diagram.

Microscopic and optical examination (Jones, 1946) showed that crystals of lysozyme chloride grown at pH 4·5 (Alderton, Ward & Fevold, 1945) and dried slowly in air are tetragonal bipyramids giving positive uniaxial interference figures; their refractive indices were found to vary with moisture content, average values being $n = 1{\cdot}554$, whereas the birefringence (0·004) remained constant. The unit cell of these crystals, containing eight molecules, has the dimensions

$$a_0 = 71{\cdot}2,\ c_0 = 31{\cdot}4 \text{ Å } (\text{Cu } K\alpha\ \lambda = 1{\cdot}542 \text{ Å})$$

(Palmer *et al.*, 1948). A discussion of the possible space groups was given in the earlier study (Palmer *et al.*, 1948). The only systematic absences found were (*h*00) when $h \neq 2n$, (0*k*0) when $k \neq 2n$, and (00*l*) when $l \neq 4n$. These observed absences indicate that dry lysozyme chloride crystals have the space group D_4^4–$P4_12_1$ or its enantiomorph D_4^8–$P4_32_1$. Since only five orders of (00*l*) have spacings greater than the minimum spacing observed from the air-dried crystals, it is of interest that when photographed in the presence of their mother liquor lysozyme chloride crystals show the same systematic absences, and in this case, throughout the range (001) to (0,0,16). In view of the facts that the wet crystals contain a 4_1 (or 4_3) screw axis and that a relatively high degree of crystal perfection is retained by the crystals after the removal of water, it appears very unlikely that the space group of the dry crystals is other than D_4^4–$P4_12_1$, or its enantiomorph D_4^8–$P4_32_1$. In the present work no indications contrary to this space group assignment were found.

Experimental

The crystals of lysozyme chloride used in this investigation were grown at pH 4·5 (Alderton *et al.*, 1945). When they were about 1 mm. on a side they were carefully removed from their mother liquor and allowed to dry slowly in the air. In this way air-dried crystals which were free from flaws were readily obtained.

The X-ray data were obtained from a series of 5° oscillation photographs taken around both the *a* and *c* axes with nickel-filtered copper radiation at 3° intervals throughout 90°. Each photograph was re-

* Present address: Department of Chemistry, University of California at Los Angeles, Los Angeles, California, U.S.A.

** Contribution No. 1675 from the Gates and Crellin Laboratories.

† Bureau of Agricultural and Industrial Chemistry, Agricultural Research Administration, U.S. Department of Agriculture. Article not copyrighted.

corded on three superposed flat films. Because of the large specimen-to-film distance (15·1 cm.) and the short oscillation range, all spots were adequately resolved and no difficulty was experienced in the assignment of indices. About 650 spots corresponding to 198 separate forms of (*hkl*) were indexed. The minimum spacing observed was 5·7 Å. The 2° overlap on the oscillation photographs made it possible to correlate intensities on two adjacent films by measurements of spectra common to them. The intensities were corrected for Lorentz and polarization factors in the usual way to obtain relative values of $|F|_o^2$ on an arbitrary scale. These data are listed in Table 1.

The observed values of $|F|^2$ were treated by the method of Wilson (1942) in accord with the following equation:

$$\sum_{hkl} k|F|_o^2 = \sum_{hkl} \sum_i f_i^2 \exp\{-2B(\sin^2\theta/\lambda^2)\}, \qquad (1)$$

where k is the scale factor and other symbols have their usual significance. If the summations indicated are carried out over small intervals of $\sin\theta$, each of which contains a sufficiently large number of observed $|F|^2$ values, and the quantity $-\log_{10}(\sum_{hkl}|F|_o^2/\sum_{hkl}\sum_i f_i^2)$ is then plotted against $\sin^2\theta$, the intercept of the resulting straight line gives the value of $\log_{10} k$, and the slope the value of $2B/2{\cdot}3\lambda^2$. The method as outlined above is virtually the same as that used by Perutz (1949) in his work on methemoglobin. In the present investigation the values of $\sum_{hkl}\sum_i f_i^2$ were calculated in the following way: the best available analytical data for lysozyme chloride (Lewis, Snell, Hirschmann & Fraenkel-Conrat, 1950; W. A. Schroeder, California Institute of Technology, private communication) together with the value 15,680 for the molecular weight of the asymmetric unit in air-dried lysozyme chloride crystals (Palmer *et al.*, 1948), lead to the empirical formula $C_{584}H_{1104}O_{270}N_{180}S_{10}Cl_{11}$ for one-eighth of the content of the unit cell. If we assume that the f-curve for atomic nitrogen represents, with a change in scale of Z_i/Z_N, the f-curve for atom i, then the value of $\sum_{hkl}\sum_i f_i^2$ may be obtained very simply by substituting for it the quantity $\sum_{hkl}(\frac{1}{7}f_N)^2K$, where $K = \sum_i f_i^2$ at $\sin\theta = 0$, or 432,000.

The results of this evaluation of scale and temperature factors are presented in Fig. 1, which is a plot of the experimental points and the straight line obtained from them by the method of least squares. In the present case the scale factor, k, and the temperature factor, B, cannot be determined with the degree of certainty which might be inferred from the method. The smallest spacing of any observed reflection is about 6 Å and the corresponding low resolution in Patterson space causes many interatomic peaks to overlap with the peak at the origin. No account of such overlaps is taken in computations

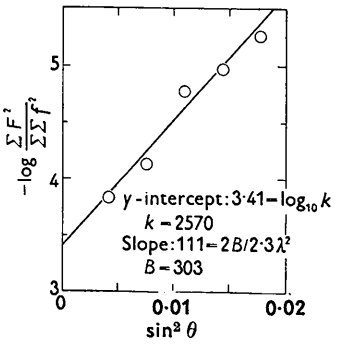

Fig. 1. The Wilson plot of the linear relationship used for determining an absolute scale factor k and a 'temperature' factor B. Correction of the data for overlapping of Patterson peaks with the peak at the origin yields a corrected value for k of 22,400.

leading to the values of the scale and temperature factors. A quantitative treatment of the effect of this overlapping on the value of the scale factor k is given at a later point in this paper.

The large value of 303 obtained for the temperature factor, B, does not indicate high-amplitude thermal vibrations in the crystal but arises rather from the disorder which develops during the drying of the crystals in air after removal from their mother liquor. Since this drying is accompanied by a loss in weight of about 40% (Jones, 1946), it is perhaps remarkable that the disorder is not even greater.

Treatment of the data

Attempted sign determination by method of inequalities

A direct method of structure determination suggested by Harker & Kasper (1948) makes use of the unitary structure factors, U_{hkl}, defined by the equation $|U_{hkl}| = |F_{hkl}|/\sum_i f_i'$, where f_i' is the form factor of the ith atom, including the temperature factor, $\exp(-B\sin^2\theta/\lambda^2)$. For the application of this method to its best advantage both the absolute scale factor and the temperature factor must be known. Wilson's method for the determination of the scale and temperature factors is based on the following relation:

$$|F_{hkl}|^2 = \sum_i \sum_j f_i' f_j' \exp[2\pi i(\theta_i - \theta_j)]. \qquad (2)$$

Separating terms for which $i = j$, we have

$$|F_{hkl}|^2 = \sum_i f_i'^2 + \sum_{i\neq j}\sum f_i' f_j' \exp[2\pi i(\theta_i - \theta_j)]. \qquad (3)$$

Wilson states that if the $|F_{hkl}|^2$'s are averaged in a small range of $\sin^2\theta$, the double sum on the right will add up to zero 'if $\lambda/\sin\theta$ is small compared to interatomic distances', since on the average the positive and negative terms will tend to cancel. Under these circumstances equation (3) reduces to equation (1), which was used in the determination of the scale and temperature factors described above. An alternative

Table 1. *Relative values of $|F|^2$ estimated from photographs of air-dried tetragonal lysozyme chloride crystals*

hkl	$\|F\|^2_{rel.}$ *	*hkl*	$\|F\|^2_{rel.}$	*hkl*	$\|F\|^2_{rel.}$	*hkl*	$\|F\|^2_{rel.}$
200	4	10,7,0	6	851	8	652	4
400	106			10,5,1	28	752	3
600	72	980	4			852	9
800	3			661	3		
10,0,0	30	990	3	761	13	762	11
12,0,0	5			861	4	962	4
		101	6	10,6,1	2	10,6,2	3
110	33	201	2	11,6,1	3		
210	727	301	4			103	5
310	27	401	95	771	14	303	3
410	76	501	10	971	9	403	75
510	71	601	2			503	3
610	180	801	19	881	3	603	14
710	17	901	8			803	30
810	190	10,0,1	15	202	205		
910	17	12,0,1	5	302	54	113	6
10,1,0	17			402	53	213	2
11,1,0	42	111	115	502	40	313	3
12,1,0	5	211	58	702	9	413	3
		311	11	802	2	513	6
220	173	411	43	902	3	613	5
320	9	511	119			713	10
420	28	611	13	112	69	813	4
520	37	711	70	212	14	913	5
620	1	811	65	312	4		
720	21	911	7	412	31	223	127
820	20	10,1,1	23	512	124	323	2
920	1			612	19	423	6
10,2,0	2	221	355	712	9	523	5
11,2,0	2	321	16	812	11	623	20
		421	136	912	8	723	5
330	46	521	264	10,1,2	3	823	3
430	171	621	60	11,1,2	4		
630	2	721	18			333	19
730	15	821	4	322	65	433	2
830	37	921	18	422	87	533	12
930	3	10,2,1	1	522	28		
10,3,0	1	12,2,1	7	622	87	443	71
11,3,0	1			722	17	543	15
12,3,0	2	331	2	10,2,2	32	743	3
		431	184	11,2,2	2		
440	299	531	42			753	8
540	3	631	5	332	186		
640	9	731	55	432	10	204	1
10,4,0	10	831	15	532	59	204	3
11,4,0	3	931	10	632	74	504	3
		10,3,1	8	732	3	704	3
550	25			832	12		
650	35	441	3	10,3,2	4	314	3
750	5	541	149	11,3,2	5	414	2
950	4	641	51			514	3
10,5,0	2	741	74	442	18	614	2
		841	4	542	47	714	3
660	115	941	3	642	12		
760	57			742	3	224	1
960	4	551	12	11,4,2	7	424	2
		651	13				
770	15	751	15	552	6	544	2
970	6						

* Application of the method of Wilson leads to the factor 2570 for converting these values of $|F|^2$ to the absolute scale.

interpretation of the summation of equation (3) over *hkl* is to consider that each side represents a calculation of the height of the origin peak of the three-dimensional Patterson function by a different path. In this regard Hughes (1949) pointed out that 'the double sum over *hkl* gives the contribution *at the origin* from all peaks having maxima elsewhere', and that it is expected to average to zero if 'the data are sufficiently extensive to resolve the peak at the origin ... from neighboring peaks'.

In the case of lysozyme it therefore seemed that the Wilson method would not give a reliable value for the

scale factor. However, a more reliable value for the scale factor can be obtained if account is taken of the overlaps of neighboring peaks with the peak at the origin. The height of the peak at the origin of the Patterson function is given by the relation

$$P(0,0,0) = \frac{1}{V}\sum_{hkl}|F|^2_{hkl} = \frac{1}{V}\sum'_{hkl}|F|^2_{hkl} + \frac{1}{V}|F|^2_{000}\,.$$

The observed values of $|F|^2_{hkl}$ (Table 1) must of course be multiplied by the scale factor k before the above relation applies. The quantities V and $|F|^2_{000}$ are known, so we find that

$$P(0,0,0) = 0{\cdot}462k + 28{,}200\,.$$

The height of the peak $P(000)$ may be evaluated by use of the known contents of the unit cell and the expected shape of a Patterson peak. For a Patterson function which has been sharpened by dividing the terms by an average unitary scattering factor and then multiplying by a convergence function, $M(H)$, the shape of a peak is given by the Fourier transform (Shoemaker, 1947)

$$P_{ij}(r) = 4\pi Z_i Z_j \int_0^\infty M(H)\frac{\sin 2\pi Hr}{2\pi Hr}H^2 dH\,,$$

where $H = 2\sin\theta/\lambda$ and r is the distance from the center of the peak. In the case of lysozyme, the effect on the data of the large temperature factor is tantamount to the sharpening and multiplication by a convergence function. We then obtain, since $M(H) = \exp[-\frac{1}{2}BH^2]$,

$$P_{ij}(r) = Z_iZ_j(2\pi/B)^{\frac{3}{2}}\exp[-2\pi^2r^2/B] = 0{\cdot}00303\,Z_iZ_j\exp[-0{\cdot}0658r^2]\,,$$

where B has been given the value 303 as obtained from the Wilson plot. It is then found that

$$\sum_j P_{ii}(0) = \sum_j 0{\cdot}00303\,Z_i^2 = 1{\cdot}28\times10^3\,,$$

where the summation is carried over the j atoms in the unit cell. The quantity $\sum_j P_{ii}(0)$ is equal to the value of $P(0,0,0)$ if there is no overlap from peaks near the origin. In order to estimate the contribution of the neighboring peaks, we assume that the number, $N(r)$, of atoms within a distance r of each atom is given by $N(r) = ar^3$, where a is a constant. The value of a is found to be 0·25 Å^{-3} by making use of the average volume per atom (excluding hydrogen), 17 Å^3 in the crystalline amino acids for which data are available. The integral

$$P_{ij}(0) = 1{\cdot}28\times10^3\int_0^\infty 0{\cdot}25r^3\exp[-0{\cdot}0658r^2]dr = 3{\cdot}7\times10^4$$

will then give the contribution to $P(000)$ of the overlapping peaks. Thus

$$P(0,0,0) = \sum_j P_{ii}(0) + \sum_j P_{ij}(0) = 1{\cdot}28\times10^3 + 3{\cdot}7\times10^4 = 3{\cdot}8\times10^4\,.$$

Moreover, since $P(0,0,0) = 0{\cdot}462k + 28{,}200$, we find that $k = 2{\cdot}2\times10^4$. The scale factor given by the Wilson plot (Fig. 1) is therefore too small by a factor of 8·6. If a is increased to 0·30, the scale factor is increased to $3{\cdot}8\times10^4$, i.e. by an additional factor of 1·7. The value $a = 0{\cdot}30$ corresponds to a volume of 14 Å^3 per atom, a value which appears unreasonably small. If the temperature factor B is increased to 325, it is found that k is $4{\cdot}9\times10^4$, an increase over our corrected value by a factor of only 2·2. However, it is unlikely that B has been undervalued by as much as 25 units, because of the satisfactory fit with the straight line in the Wilson plot. It therefore seems very likely that the value $k = 2{\cdot}2\times10^4$ is not in error by more than a factor of about 2.

If the corrected scale factor of $2{\cdot}2\times10^4$ is used to place the values of $|F|^2$ on an absolute scale, and these values are then used to calculate the corresponding values of $|U_{hkl}|$, it is found that the largest $|U_{hkl}|$ is 0·10. Of the 198 observed forms, only about 20 have values of $|U_{hkl}|$ as large as 0·05. An error of a factor of about 9 in the corrected scale factor (which as defined refers to $|F|^2$; $\sqrt{k}$ refers to $|F|$ and hence to $|U|$) would be required in order that the method of inequalities could be used to fix the signs of any reflections when atom form factors are used for the calculation of $|U_{hkl}|$. On the other hand, if the arrangement of atoms in the crystal is such that a molecule, or a part of a molecule, effectively scatters as a single unit, the form factors of these units could be used in place of atom form factors in the calculation of the alternative values of $|U_{hkl}|$, some of which might then be sufficiently large to permit evaluation of their signs. Since nothing whatever is known about the shapes of the hypothetical units, evidence for their presence must be sought in the experimental data.

The form factors for uniform spheres and other simple geometrical figures have nodes when plotted against $\sin\theta$; consequently a plot of the observed

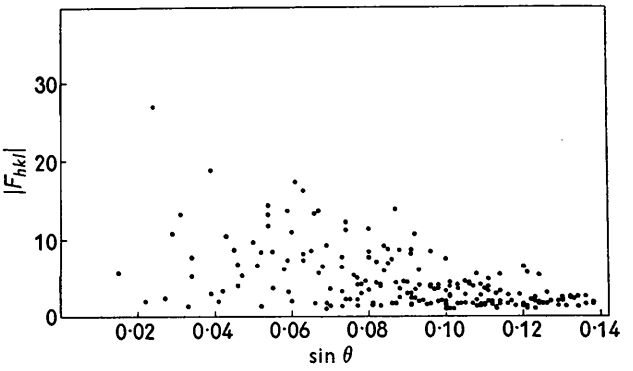

Fig. 2. A plot of $|F_{hkl}|$ *v.* $\sin\theta$ for the 198 reflections observed from air-dried lysozyme chloride crystals.

amplitude of the structure factors against sin θ might reveal such a variation in form factor. Fig. 2 is a plot of the *hkl* data for dry lysozyme chloride. Since this plot fails to indicate the presence of nodes, we may conclude that the scattering matter in these crystals is not distributed in the form of uniform spheres or spherical shells. The absence of other spherical distributions of scattering matter whose form factors do not decrease monotonically with sin θ is also indicated. The presence of non-spherical scattering masses would of course not be detected in this fashion.

In the absence of information from other sources which might indicate more complicated distributions of scattering matter, it is fruitless to attempt to deduce possible form factors from the intensity data alone; consequently this approach to the structure determination was not carried further.

The radial distribution function

Evidence for the presence of the 3·7-residue α-helix in hemoglobin (Pauling & Corey, 1951) and in bovine serum albumin (Riley & Arndt, 1952) has been found by comparing the radial distribution functions of these proteins with two calculated for a single α-helix. The radial distribution of dry lysozyme chloride is shown in Fig. 3. It shows one broad maximum at about 13·2 Å, which certainly consists in large part of inter- rather than intra-chain vectors. There is no distinctive detail in the region between 5 and 10 Å, such as that characteristic of the α- and γ-helixes (Pauling & Corey, 1951); however, we do not feel that important intra-chain vectors would characterize a radial distribution curve based on data which extend only to spacings of 5·7 Å.

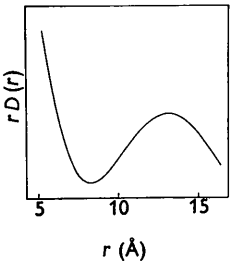

Fig. 3. The radial distribution function for air-dried lysozyme chloride.

Calculation of the Patterson function

The values of $|F|^2$ may be used for the calculation of a complete three-dimensional Patterson plot of all interatomic vectors in the unit cell of the crystal. This plot would be characterized by low resolution, not only because of the large number of atoms in the unit cell and the consequent overlapping of vector maxima, but also because of the absence of data from short interplanar spacings. Therefore it was not anticipated that the Patterson function would yield much information about the details of the structure. Atoms, or even atomic aggregates, such as phenyl groups and other protein side-chains, would not even be resolved, much less identified. Under the most favorable circumstances, the approximate profile of the protein molecule and perhaps some gross features of the distribution of matter within it might be indicated.

The calculation of the Patterson function,

$$P(u, v, w) = \sum_h \sum_k \sum_l |F_{hkl}|^2 \cos 2\pi(hu+kv+lw) ,$$

was carried out with a set of cards punched to correspond to Beevers–Lipson strips (V. Schomaker, unpublished work). The function was evaluated at intervals of $\frac{1}{60}$ in u, $\frac{1}{60}$ in v, and $\frac{1}{20}$ in w, corresponding to 1·19, 1·19, and 1·56 Å, respectively. The general appearance of the three-dimensional plot is shown in

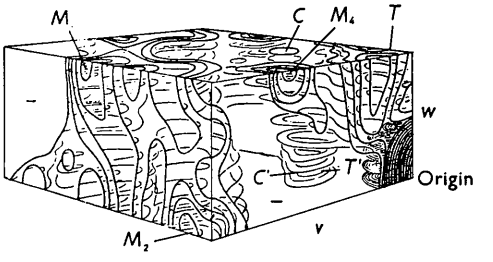

Fig. 4. Perspective view of the function $P(u, v, w)$. Minus signs do not indicate negative regions, but regions of comparatively low vector density. Throughs and craters in the vector density are designated by minus signs to distinguish them from maxima. Contours are drawn at equally spaced arbitrary intervals.

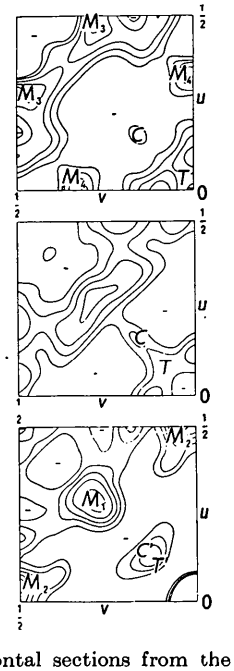

Fig. 5. Three horizontal sections from the three-dimensional Patterson plot; the upper, middle, and lower sections are taken at $w = \frac{1}{2}$, $\frac{1}{4}$, and 0 respectively. Contours and minus signs drawn as described for Fig. 4, except that the higher contours of the origin peak are omitted.

Fig. 4. Horizontal sections at $w = 0, \frac{1}{4}$, and $\frac{1}{2}$, respectively, are reproduced in Fig. 5 and the diagonal vertical section, $u = v$, in Fig. 6. Contours are drawn at the same regularly spaced arbitrary intervals; minus signs do not indicate negative regions, but merely regions of relatively low density.

The principal features of the Patterson plot can be seen from Figs. 4 and 5. The function is symmetrical about diagonal planes at $u = \pm v$. The region close to and above the origin, representing relatively short vectors, is joined to the main body of vector density by a narrow diagonal 'bridge' near $w = \frac{3}{8}$; this bridge is plainly visible in Figs. 4 and 6. Also in the $u = \pm v$ planes there are somewhat elongated torus-shaped regions TT' of relatively high densities. Several other

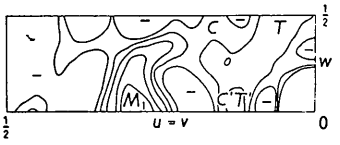

Fig. 6. The diagonal vertical section of the three-dimensional Patterson function taken at $u = v$. Contours and minus signs drawn as described for Fig. 5.

prominent maxima are to be seen at various levels throughout the plot; M_1, M_2 and M_3, etc. designate maxima which are discussed in the text. Especially striking are two regions of comparatively low density: that surrounding $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$, shown clearly in Fig. 4, and the large region which almost divides the plot into separate portions corresponding to short and long vectors respectively (Figs. 4 and 5).

Interpretation of the three-dimensional Patterson function

The absence of rod-like features

In the Patterson function of horse methemoglobin (Perutz, 1949) rod-like features parallel to a were observed; in the Patterson function of horse metmyoglobin (Kendrew, 1950) similar rods were found parallel to $[20\bar{1}]$. In both proteins these rods were identified as the vector equivalents of polypeptide chains running parallel to a and $[20\bar{1}]$ respectively. It was further suggested that the molecules consist of parallel bundles of chains perpendicular to the monoclinic axes; the space-group symmetry of these crystals then requires that all chains be parallel. The chains in hemoglobin have recently been interpreted in terms of the 3·7-residue α-helix (Pauling & Corey, 1951).

No similar rod-like features are present in the Patterson function of lysozyme; however, rods would presumably not be an outstanding feature of the Patterson unless all chains were parallel. If lysozyme molecules consist of bundles of parallel chains the high symmetry of these crystals would result in parallelism of all chains only for one particular molecular orientation. Thus, the absence of rod-like features in the three-dimensional Patterson function of air-dried lysozyme chloride is not surprising.

The Patterson projection $P(u, v)$

In order to facilitate a simple approach to the solution of the Patterson diagram, a two-dimensional Patterson projection $P(u, v)$ was prepared. It was hoped that this projection could be interpreted by direct inspection of conspicuous maxima and minima, and that it could be used to assign positions to molecules or masses of scattering matter in the unit cell. The function $P(u, v)$, shown in Fig. 7, has three main

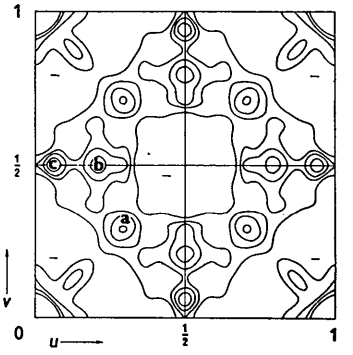

Fig. 7. The Patterson projection, $P(u, v)$. Contours and minus signs drawn as described for Fig. 5.

peaks, designated **a**, **b**, and **c**, and three smaller peaks. It is possible to account for peak **b** and a part of peak **a** by placing scattering masses in the eight equivalent points of the space group $P4_12_1$: (1) x, y, z; (2) $\bar{x}, \bar{y}, \frac{1}{2}+z$; (3) $\frac{1}{2}-y, \frac{1}{2}+x, \frac{1}{4}+z$; (4) $\frac{1}{2}+y, \frac{1}{2}-x, \frac{3}{4}+z$; (5) $y, x, \bar{z}$; (6) $\bar{y}, \bar{x}, \frac{1}{2}-z$; (7) $\frac{1}{2}-x, \frac{1}{2}+y, \frac{1}{4}-z$; (8) $\frac{1}{2}+x, \frac{1}{2}-y, \frac{3}{4}-z$, with $x = y = 0{\cdot}146$. A peak is then expected to occur at $u = v = 2x$ and another, having twice the height of the first, at $u = \frac{1}{2}-2x$, $v = \frac{1}{2}$. These positions, and their relationship to peaks **a** and **b**, are shown in Fig. 8.

If for this arrangement the scattering masses are identified as molecules, these would then be grouped in pairs having the same x and y coordinates. In the

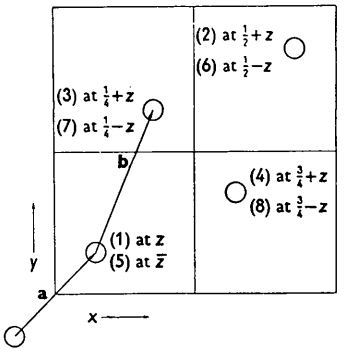

Fig. 8. An arrangement of equivalent points and their relationship to peaks **a** and **b** of the Patterson projection.

z direction each molecule would have two neighbors, one at a distance $2z$ and the other at $1-2z$. Considerations of simple packing would suggest a value of $\frac{1}{4}$ for the z-parameter, which would result in an equal spacing of all molecules $\frac{1}{2}c$ apart along the z axis. A maximum in the three-dimensional Patterson diagram would then be expected at $0, 0, \frac{1}{2}$. Actually, $0, 0, \frac{1}{2}$ is a center of comparatively low vector density, as can be seen in Fig. 5. Regardless of the value of z, the futility of assuming this arrangement can be shown by considering points (1) and (2) which are related by the vectors $2x, 2y, \frac{1}{2}$. Reference to Fig. 5 shows that the point $u = v = 0{\cdot}292$, $w = \frac{1}{2}$ lies in a region of exceptionally low vector density. It was therefore obvious that the apparent relationship between the two peaks **a** and **b** was merely a coincidence, and further exploration of this structure was abandoned.

Packing of spherical molecules

As a second starting point for an attempt at a direct interpretation of the Patterson diagram spherical molecules were assumed, arranged in accord with the eight general positions of the space group $P4_12_1$ and so as to give minimum voids. Simple calculation shows that these conditions are satisfied by spheres 27·1 Å in diameter with $x = 0{\cdot}256$, $y = 0{\cdot}476$, $z = 0{\cdot}250$. In this arrangement, shown in Fig. 9, each sphere (1) is in contact with five other spheres, (3), (4′), (6), (5) and (5′) in the cell below. Each sphere is also nearly (within 4·3 Å) in contact with its equivalent spheres in unit cells above and below it. The spheres occupy 52·4% of the volume of the unit cell.

If the vectors which join the centers of these spheres are compared with the Patterson diagram many vectors are found to occupy regions of relatively high density. Serious conflicts, however, occur. In particular, the vector relating points (1) and (8), for which $u = \frac{1}{2}$, $v = 0{\cdot}452$, $w = \frac{1}{4}$, and that relating points (1) and (5), for which $u = v = 0{\cdot}220$, $w = \frac{1}{2}$, fall in regions of conspicuously low vector density in the Patterson plot.

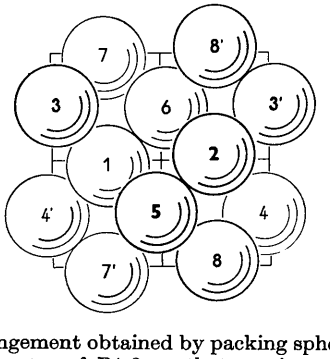

Fig. 9. The arrangement obtained by packing spheres in accord with the symmetry of $P4_12_1$ so that maximum space-filling is achieved.

Regions of relatively high scattering power

Subsequent inspection of the Patterson plot showed that good agreement could be obtained by use of the parameters $x = 0{\cdot}29$, $y = 0{\cdot}40$, $z = \frac{1}{8}$. Equivalent positions corresponding to these parameters are shown as large circles, A, in Fig. 10; fractions indicate the

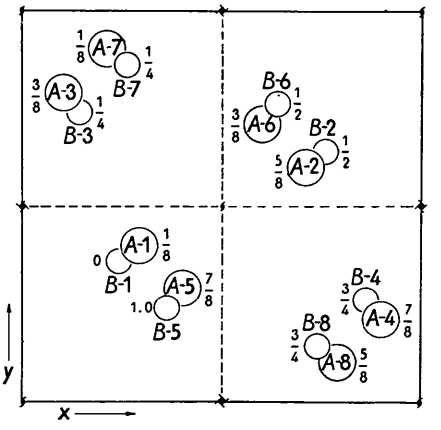

Fig. 10. Projection along z of one unit cell, showing the positions of the regions of relatively high scattering power, A and B; z-parameters are indicated by fractions.

z parameter. The positions in the three-dimensional Patterson plot of the interactions (AA) between these points are shown as large circles in Fig. 11. The remaining conspicuous maxima appearing in the Patterson section $w = \frac{1}{4}$, which correspond to long intermolecular vectors (M_2 and M_3), can be explained by assuming the presence of scattering matter in the neighborhood of $x = 0{\cdot}24$, $y = 0{\cdot}36$, $z = 0{\cdot}00$, and equivalent points in the crystal. These points (B) are shown as small circles in Fig. 10, and the corresponding maxima in the Patterson plot by similar circles in Fig. 11. Vectors representing interactions between the points of sets A and B produce maxima in the Patterson plot at levels $w = \frac{1}{8}$ and $w = \frac{3}{8}$ respectively. Plots of these levels are shown in Fig. 12. They were prepared by interpolation between the levels at $\frac{2}{20}$ and $\frac{3}{20}$, and $\frac{7}{20}$ and $\frac{8}{20}$. The positions of the maxima (AB) are represented by circles in Fig. 12, all are seen to lie in regions of relatively high vector density. A search was made for other significantly different sets of positions in the crystal which, through interactions within the set or with other sets, would give rise to features of the Patterson plot not explained by the positions A and B. This search was unsuccessful in the sense that no positions were found which would give rise to maxima such as M_1 or M_4 without at the same time giving rise also to other maxima, some of which occurred in regions of low vector density in the Patterson plot.

Discussion

It is possible to reach certain reasonable conclusions about the packing in the dry lysozyme chloride

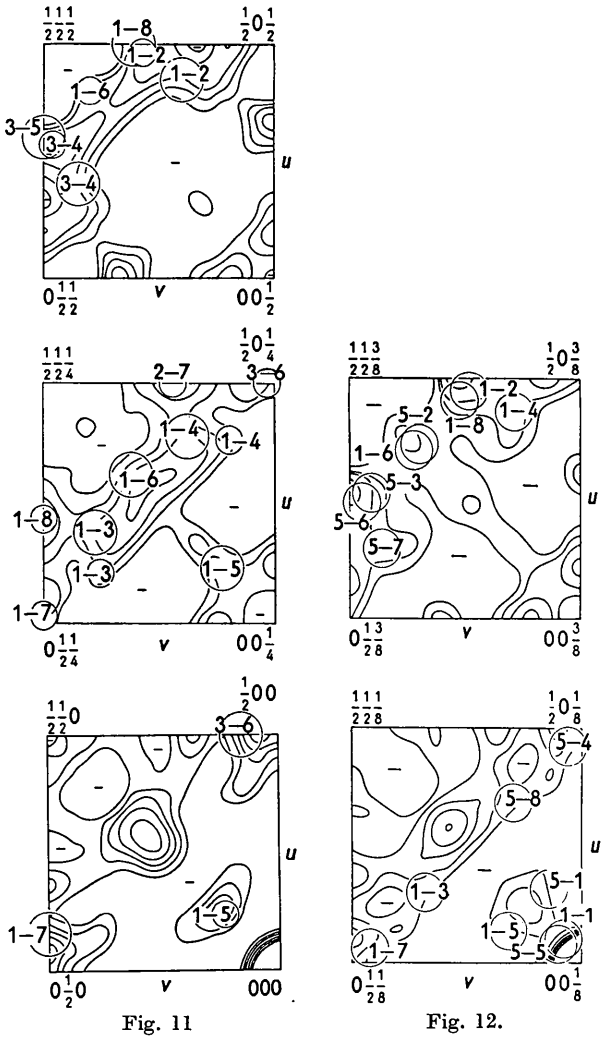

Fig. 11 Fig. 12.

Fig. 11. Patterson sections at $w = \frac{1}{2}, \frac{1}{4}$, and 0, contours and minus signs drawn as described for Fig. 5. Large circles show positions of interactions AA, small circles show positions of interactions BB.

Fig. 12. Patterson sections at $w = \frac{3}{8}$ and $\frac{1}{8}$, contours and minus signs drawn as described for Fig. 5. Circles show positions of interactions AB.

crystal from consideration of the density of these crystals and of similar substances. Palmer *et al.* (1948) found the density to be 1·305 g.cm.$^{-3}$, after correction for the small content of sodium chloride. We may further correct for the 9% of water in the crystals; if we assume that this water has approximately its usual density, we find that the density of the molecules of lysozyme chloride themselves plus whatever voids exist in the crystal is about 1·35 g.cm.$^{-3}$.

The densities of crystals of representative amino acids vary widely; the more polar and compact molecules (e.g. serine, threonine, hydroxyproline, glycine, aspartic acid) have crystal densities in the range 1·46–1·66 g.cm.$^{-3}$, while those with large non-polar side chains (e.g. leucine, α-amino-n-butyric acid, valine) have densities between 1·16 and 1·24 g.cm.$^{-3}$. Substances which may be considered intermediate between these classes have intermediate densities (e.g. alanine, 1.40 g.cm.$^{-3}$).

Since the best available amino acid analysis of lysozyme indicates that more than half of the residues consist of the more polar amino acids, which pack best in the solid state, it is possible that the protein chains in lysozyme may be significantly more closely packed than those in other proteins. However, it seems very unlikely that even the most favorable packing of such chains would lead to an intrinsic density of more than about 1·40–1·45 g.cm.$^{-3}$ for the protein molecule in this crystal. Allowance for the small content of chloride (2·4%) in the dry lysozyme chloride will not markedly affect this figure.

If then we accept the approximate figure of 1·45 g.cm.$^{-3}$ as an upper limit for the intrinsic density of the protein molecules, and compare this figure with the density 1·35 g.cm.$^{-3}$ calculated for the protein plus whatever voids exist unfilled by water, we see that these voids can amount at most to only about 7% of the volume of the crystal, and probably represent an even smaller fraction. Since the water occupies only about 12% of the total volume, the space occupied by the protein molecules is of the order of 80–85% of the volume of the crystal, or considerably more than the 52% which close-packed spheres would occupy when arranged in accord with the space-group symmetry $P4_12_1$. These considerations provide a strong argument in favor of a molecular shape which is markedly different from spherical, and render unlikely the interpretation of A and B as anything other than regions of relatively high scattering power.

In an early approach to the interpretation of the Patterson diagram use was made of the intensities of the ($hk0$) reflections and of the trigonometric portions of the corresponding structure factors, in the hope that it might be possible to obtain from these X-ray data an indication of the position and perhaps even of the shape and orientation of a hypothetical homogeneous molecule. Data for (110), (200) and (210) were inspected first with the introduction of 'molecules' having the form of spheres and of circular and elliptical cylinders, but no satisfactory intensity agreement could be obtained. However, an ellipsoid of uniform density having semi-axes 19, 12 and 12 Å respectively, when placed with its center at x=0·23, y=0·43, and its major axis inclined 22° to the x axis of the crystal, gave excellent agreement for the relative intensities of (110), (200), (210), (220) and (310). For reflections at larger angles agreement was not at all good, and clearly could not be significantly improved without the assumption of a much more complex distribution of scattering matter in the crystal. Nevertheless, the agreement of these planes of low indices may not be wholly without significance. Although it could not be construed as evidence that the lysozyme molecule has the shape of an ellipsoid

of revolution or, for that matter, of any simple geometrical solid, it suggested that the positions of the ellipsoids might correspond roughly to regions of relatively high density of scattering matter, and thus to the general domain of the protein molecules in the crystal.

To gain further insight into the possible justification of this use of intensity data, and of any significance which might be attributed to this distribution of ellipsoidal 'molecules', a Fourier projection was calculated with five strong (*hk*0) planes, (210), (220), (400), (430) and (440), these being the planes for which the signs of the structure factors were strongly indicated by the assumed distribution. A plot of this projection (Fig. 13) shows a conspicuous maximum centered

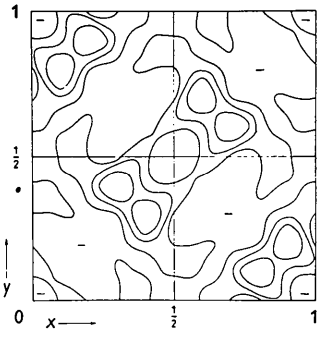

Fig. 13. Fourier projection on (001), made with relative amplitudes of the reflections (210), (220), (400), (430) and (440). Contours drawn at equally spaced arbitrary intervals. Minus signs indicate negative regions.

about the point $x = 0{\cdot}29$, $y = 0{\cdot}40$, the exact x- and y-coordinates of point A already derived on the basis of the three-dimensional Patterson diagram. The relationship between this Fourier plot and points A and B can be seen by comparison of Figs. 10 and 13. Since this plot failed to develop additional features which might be helpful in the location of scattering matter in the cell and thus aid in the establishment of signs of structure factors, it failed to extend our information beyond that already derived from the consideration of the three-dimensional Patterson diagram as a whole.

The interpretation of this Fourier plot as indicating the distribution of scattering matter within the lysozyme chloride crystal is probably unwarranted because of the oversimplifications made in the method of sign determination. On the other hand, the remarkable agreement of the Fourier projection with the results of the interpretation of the three-dimensional Patterson function may be construed as evidence that regions in the vicinity of A may really represent maxima in electron density. Lysozyme chloride contains eleven chloride ions and twelve sulfur atoms per molecule. Regions around A and B might then be interpreted as the loci of the basic amino acid residues (17 per molecule) or the cystine residues (5 or 6 per molecule) and the methionine residue (1 per molecule).

The authors are deeply indebted to Prof. Linus Pauling and Prof. Verner Schomaker for helpful discussions and criticisms. Part of the work of indexing the photographs and estimating the intensities was performed by Mrs Merle B. Platt.

This work was aided by a grant from the National Foundation for Infantile Paralysis.

APPENDIX

Variation of form factors

In order to determine the effect of the form factor of the scattering centers on the appearance of a Patterson projection, studies were made on a simple hypothetical structure. It should be emphasized that the interpretation of this structure in terms of lysozyme molecules was not intended. In the Patterson functions of simple compounds the peaks arise from the interaction between two atoms, and overlapping of peaks, although a serious deterrent to a rapid elucidation of the structure, is not expected to alter radically the shapes and relative heights of the maxima. For complex crystals, where data are observed only from planes of large spacing, the resolution is such that individual atoms are no longer important, but rather groups of many atoms. The form factor of a group of atoms will be quite different from that of a single atom, and it might be anticipated that this difference would have an effect on the Patterson function in a way not easily predicted.

The hypothetical structure chosen for study consisted of scattering points at $x = y = 0{\cdot}146$ in the coordinates of the general positions of a unit cell having the symmetry $P4_12_1$ and $a_0 = 71{\cdot}2$ Å. Three different types of scattering centers were placed at these points: (1) an atom with form factor $f = \exp(-300 \sin^2\theta/\lambda^2)$; (2) a sphere of uniform density, with $f = 1/s^3 (\sin rs - rs \cos rs)$, where r is the radius of the sphere and $s = 4\pi \sin\theta/\lambda$; and (3) a spherical shell, with $f = \sin rs/rs$ (James, 1948, p. 467). The value of the radius, 13·1 Å, was chosen on the basis of the following considerations: The form factor for a uniform sphere has nodes at values of $\sin\theta$ when $rs = \tan rs$. On a plot of $|F_{hk0}|$ for lysozyme $v.\sin\theta$, it was noted that the observed data were not in great

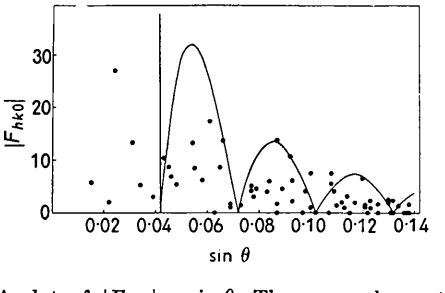

Fig. 14. A plot of $|F_{hk0}|$ $v.$ $\sin\theta$. The curve shows the form factor of a uniform sphere of radius 13·1 Å.

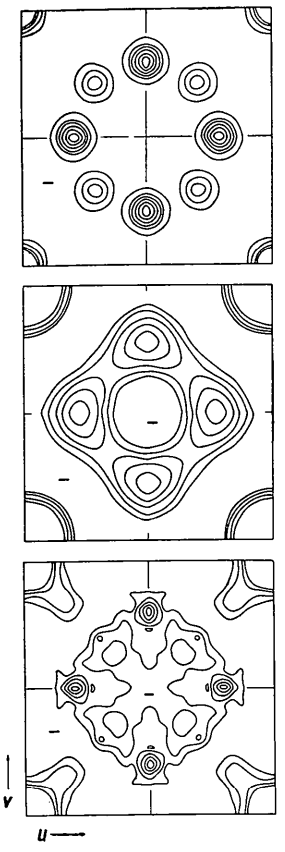

Fig. 15. Patterson projections for the same arrangement as shown in Fig. 8, but with different form factors assigned to the equivalent points. The upper, middle and lower projections were made with form factors for an atom, a sphere of uniform density, and a spherical shell, respectively. Contours and minus signs drawn as described for Fig. 5.

disagreement with a form-factor curve calculated with $r = 13{\cdot}1$ Å. This comparison is made in Fig. 14, where the f-curve, on an arbitrary scale, is superposed on the plot of the values of $|F_{hk0}|$. This figure should not be construed as indicating the presence of spherical molecules or scattering masses in lysozyme crystals, since there is a variety of shapes for which the f-curves have nodes, and not many experimental data were available. Although the nodes of the f-curve of a spherical shell of radius 13·1 Å occur at values of $\sin\theta$ differing from those shown in Fig. 14, this radius was used for the shell as well as for the sphere in order to make comparisons simpler.

The structure factors for all ($hk0$) reflections for which $\sin\theta < 0{\cdot}14$ were calculated for the three different structures, and the Patterson projections were prepared; these are shown in Fig. 15. The appearance of the three functions is of course quite different: the peaks in $P_{\text{atom}}(u, v)$ are quite sharp; the peaks in $P_{\text{sphere}}(u, v)$ are very much spread out, so much so that the interaction at $u = 2x$, $v = 2x$ is nearly lost; in the plot of $P_{\text{shell}}(u, v)$ several minor maxima appear for which there is no obvious structural interpretation. Significant changes take place with regard to the peak at $2x$, $2x$. In the plot of P_{sphere} it has nearly vanished, and in that of P_{shell} it is 8 Å, or more than one-tenth of the cell edge, from its expected position. These changes are doubtless in large part characteristic of the differences in the form factors, although rounding-off errors in the computation and lack of convergence also contribute.

A principle of fundamental significance for the interpretation of Patterson functions of complex crystals has been confirmed by the preceding analysis, namely, that the interpretation of peaks as vectors relating centers of scattering masses is not necessarily valid.

References

Alderton, G. & Fevold, J. (1946). *J. Biol. Chem.* **164**, 1.

Alderton, G., Ward, W. H. & Fevold, H. L. (1945). *J. Biol. Chem.* **157**, 43.

James, R. W. (1948). *The Optical Principles of the Diffraction of X-Rays*. London: Bell.

Jones, F. T. (1946). *J. Amer. Chem. Soc.* **68**, 854.

Harker, D. & Kasper, J. (1948). *Acta Cryst.* **1**, 70.

Hughes, E. W. (1949). *Acta Cryst.* **2**, 34.

Kendrew, J. C. (1950). *Proc. Roy. Soc.* A, **201**, 62.

Lewis, J. C., Snell, N. S., Hirschmann, D. J. & Fraenkel-Conrat, H. (1950). *J. Biol. Chem.* **186**, 23.

Palmer, K. J., Ballantyne, M. & Galvin, J. A. (1948). *J. Amer. Chem. Soc.* **70**, 906.

Pauling, L. & Corey, R. B. (1951). *Proc. Nat. Acad. Sci., Wash.* **37**, 282.

Perutz, M. F. (1949). *Proc. Roy. Soc.* A, **195**, 474.

Riley, D. P. & Arndt, U. W. (1952). *Nature, Lond.* **169**, 138.

Shoemaker, D. P. (1947). Dissertation, California Institute of Technology.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318.

Wilson, A. J. C. (1942). *Nature, Lond.* **150**, 151.