Acta Cryst. (1956). **9**, 156

The Three-Dimensional Patterson Function of Ribonuclease II*

BY BEATRICE S. MAGDOFF†, F. H. C. CRICK‡ AND V. LUZZATI§

The Protein Structure Project, Polytechnic Institute, Brooklyn 1, *N.Y., U.S.A.*

(*Received* 3 *December* 1954 *and in revised form* 6 *May* 1955)

All the intensities of the monoclinic form ($P2_1$) of the protein ribonuclease have been measured out to $1/d = 0{\cdot}445$ Å^{-1}, and a three-dimensional Patterson function has been computed from them. The Patterson shows that it is highly unlikely that the structure is based on polypeptide 'rods' all parallel to the *c* axis. However, the data show some asymmetry in the general direction indicated by the infrared dichroism. A concentration of vector fluctuations on and near the mirror planes of the Patterson has been explained on statistical grounds as a space-group effect.

Introduction

As a preliminary attack on the problem of the structure of the protein ribonuclease it was considered desirable to examine a three-dimensional Patterson function in order to obtain some idea of the simplicity, or otherwise, of the structure. For example, it has been postulated by Carlisle & Scouloudi (1951) that ribonuclease consists of parallel α-ribbons, running in the direction of the *c* axis in ribonuclease II. Such a simple structure should show unmistakable signs in the three-dimensional Patterson.

For such a Patterson function the smaller the number of molecules in the unit cell the better; in the ideal case the space group would be $P1$ with one molecule in the asymmetric unit. Thus the only reasonable choice among the various crystal modifications available (King, Magdoff, Adelman & Harker, 1956) is ribonuclease II (space group $P2_1$; cell dimensions $a = 30{\cdot}28$, $b = 38{\cdot}39$, $c = 53{\cdot}16$ Å, $\beta = 105{\cdot}83°$; one molecule per asymmetric unit) and this also has the advantage of being the form most easily obtainable. It was decided, therefore, to collect the X-ray intensities diffracted by ribonuclease II out to $2\theta = 40°$ using Cu $K\alpha$ radiation and to compute a three-dimensional Patterson function from them. Carlisle and his co-workers have also calculated a similar Patterson function; at the time of writing it has not been published, though sections from it have been shown at various meetings. The authors inform us that it is now in the course of preparation for publication.

Experimental technique

The methods for growing ribonuclease crystals have been described elsewhere (King *et al.*, 1956). Those chosen were grown from 50% tertiary butyl alcohol. The intensities were collected from only two crystals. The first was used for $2\theta = 0°$ to 30°, and the second for $2\theta = 30°$ to 40°, with an occasional measurement between $2\theta = 0°$ and $2\theta = 30°$ to enable both sets to be put on the same scale.

The experimental data were collected by two of us (B. S. M. and V. L.) with the help of all our colleagues. The intensities were measured on a General Electric XRD-3 Spectrogoniometer modified for single crystal work by the addition of an Eulerian cradle for mounting the specimen (Furnas & Harker, 1955). This instrument has the advantage that the intensities can be measured accurately and without giving the crystal prolonged X-ray exposures, which are known to damage the crystal (Magdoff, 1953). The total exposures in this experiment were 20 and 25 hr. respectively.

The cell dimensions of the crystal were first measured under fine conditions, i.e. with narrow slits, etc. From these dimensions the angular coordinates of all the *hkl* reflections, with their values of 2θ, were calculated on IBM machines (Magdoff, unpublished). They were then in such a form that they could be used for setting the instrument to bring any required reciprocal-lattice point into the reflecting position. This was done under coarse conditions (i.e. with wide slits) and for a crystal less than about $\frac{1}{2}$ mm. in size. The geometry of the apparatus is such that with the crystal stationary the counting rate is proportional to the integrated intensity. Counts were made for a fixed time of 10 sec., and the background count, from an adjacent non-lattice point in reciprocal space, was subtracted.

An approximate correction, described by Magdoff & Crick (1955*a*), was applied for the different absorption of the capillary in different directions (Furnas, unpublished). This was obtained by measuring the intensity of one chosen reflection (actually 0,10,0) with various aspects of the capillary to the X-ray beam. For the first crystal the relative correction

* Contribution No. 8 from The Protein Structure Project.

† Present address: Boyce Thompson Institute, Yonkers, N.Y., U.S.A.

‡ Present address: Medical Research Council Unit, Cavendish Laboratory, Cambridge, England.

§ Present address: Centre de Recherches sur les Macromolécules, Strasbourg, France.

varied from 1·0 to 1·3, for the second from 1·0 to 2·1. It was therefore important not to omit it.

A record of the intensities, corrected for polarization and Lorentz factor, is available.*

Errors

It is now known that this set of data is not as accurate as it might be, but it is believed that the inaccuracies are not great, and that the percentage error is at least as small as that of any other three-dimensional set of intensities of a protein crystal, ribonuclease or otherwise.

The errors are due to three causes. First the crystals were mounted rather dry and thus their intensities were probably fluctuating somewhat, as has been described elsewhere (Magdoff & Crick, 1955*a*). Thus the recorded intensities probably do not all refer to exactly the same state. Secondly no elaborate checks were made to see that the instrument was always *exactly* set on the reflections, and this may have produced some errors at higher values of $1/d$. Thirdly, the capillary correction was in one case rather large, and as this correction can only be assessed approximately it would have been worth while to work with a thinner-walled capillary and/or a thicker crystal.

To get some idea of the inaccuracies, the $h0l$'s (out to $2\theta = 30°$ using Cu $K\alpha$ radiation) were compared with what is believed to be an accurate set (see Fig. 7 in Magdoff & Crick, 1955*a*). The value of $\Delta = \Sigma|\Delta I|/\Sigma I$ rises roughly linearly with $1/d$ to about 20% at $1/d = 0{\cdot}28$ Å^{-1}, so that the errors, though not as small as they might be, are certainly acceptable. Their effect is likely to smear over the Patterson a little, but we do not believe that these errors will materially affect our conclusions.

Computation

The Patterson function was calculated, using IBM machines, by a new method (Magdoff, unpublished). There were about 5700 terms, not allowing for symmetry. It was found necessary to include elaborate checks to prevent accidental errors, which were more common than was expected. The final Patterson function is believed to contain only minor errors, say less than 10 e.^{2}Å^{-3} at any point. The values were calculated at intervals of $a/40$, $b/40$ and $c/80$. The calculation took one person about 4 weeks.

Experimental results

(*a*) *The Wilson curve*

The plot of the average intensity, $\langle I \rangle$, against $1/d$ is given in Fig. 1. It shows a sharp maximum in the 5 Å region, and the suggestion of a small maximum in the 10 Å region. The powder diagram of ribonuclease published by Arndt & Riley (1955) confirms that this latter peak really exists, although it is small by comparison with the corresponding peak in many other proteins. This in itself suggests that ribonuclease is unlikely to consist mainly of parallel 'rods' spaced 10 Å apart.

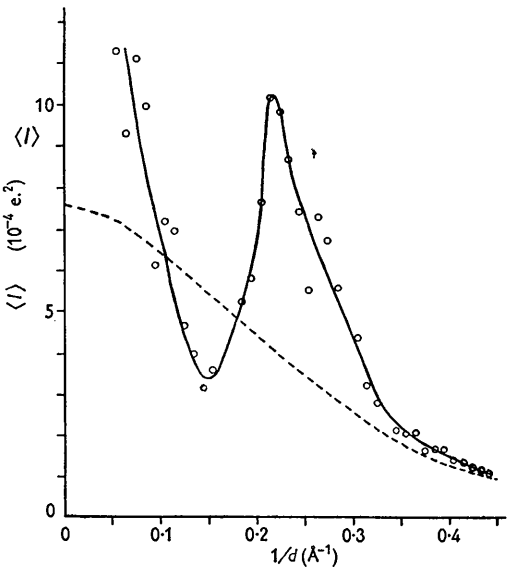

Fig. 1. Plot of $\langle I \rangle$ against $1/d$. The circles are computed values. The broken curve is $1560 f_N^2(s) \exp(-4{\cdot}56/d^2)$, which values were used to remove the origin peak.

The high values of Fig. 1 near the origin are caused by the low electron density of the solvent contrasted with that of the hydrated protein.

It is not easy to choose suitable constants to fit a curve of the form $\Sigma f^2 \exp(-A/d^2)$ to Fig. 1, especially as the number of *ordered* atoms in the crystal is not known. Preliminary measurements to put the data on an absolute scale have been made by Dr Thomas C. Furnas, Jr., and his provisional value has been used throughout this paper. The dotted line of Fig. 1 shows the curve $1560 f_N^2(s) \exp(-4{\cdot}56/d^2)$, and these values were used to remove the origin peak.

(*b*) *The Patterson function*

The Patterson function was plotted by a novel method. The actual IBM sheets contained the figures set out in a rectangular array. Contours were first drawn on these sheets. They were naturally distorted from the correct shape because the basic array did not correspond to the monoclinic cell. Correct contours were then drawn on to the final sheets, having an underlying grid of the correct size and shape, by comparing the points where the contours intercepted their grids in the two cases. This method proved to be

* The record has been deposited as Document No. 4665 with the ADI Auxiliary Publications project, Photoduplication Service, Library of Congress, Washington 25, D.C., U.S.A. A copy may be secured by citing the Document number and by remitting \$21·25 for photoprints, or \$6·25 for 35 mm. microfilm. Advance payment is required. Make checks or money orders payable to: Chief, Photoduplication Service, Library of Congress.

quicker than first copying all the computed figures on to the scaled sheets and then drawing the contours.

Each contour represents a level of about 100 e.^{2}Å^{-3}. Fig. 2 shows the 21 sections of the three-dimensional Patterson function, from $y = 0$ to $y = 20b/40$. The IBM sheets from which they were plotted have been deposited with the American Documentation Institute,* whence copies can be obtained.

As might be expected from Fig. 1, the Patterson has a very marked 5 Å shell. This shell is not com-

* See footnote on p. 157.

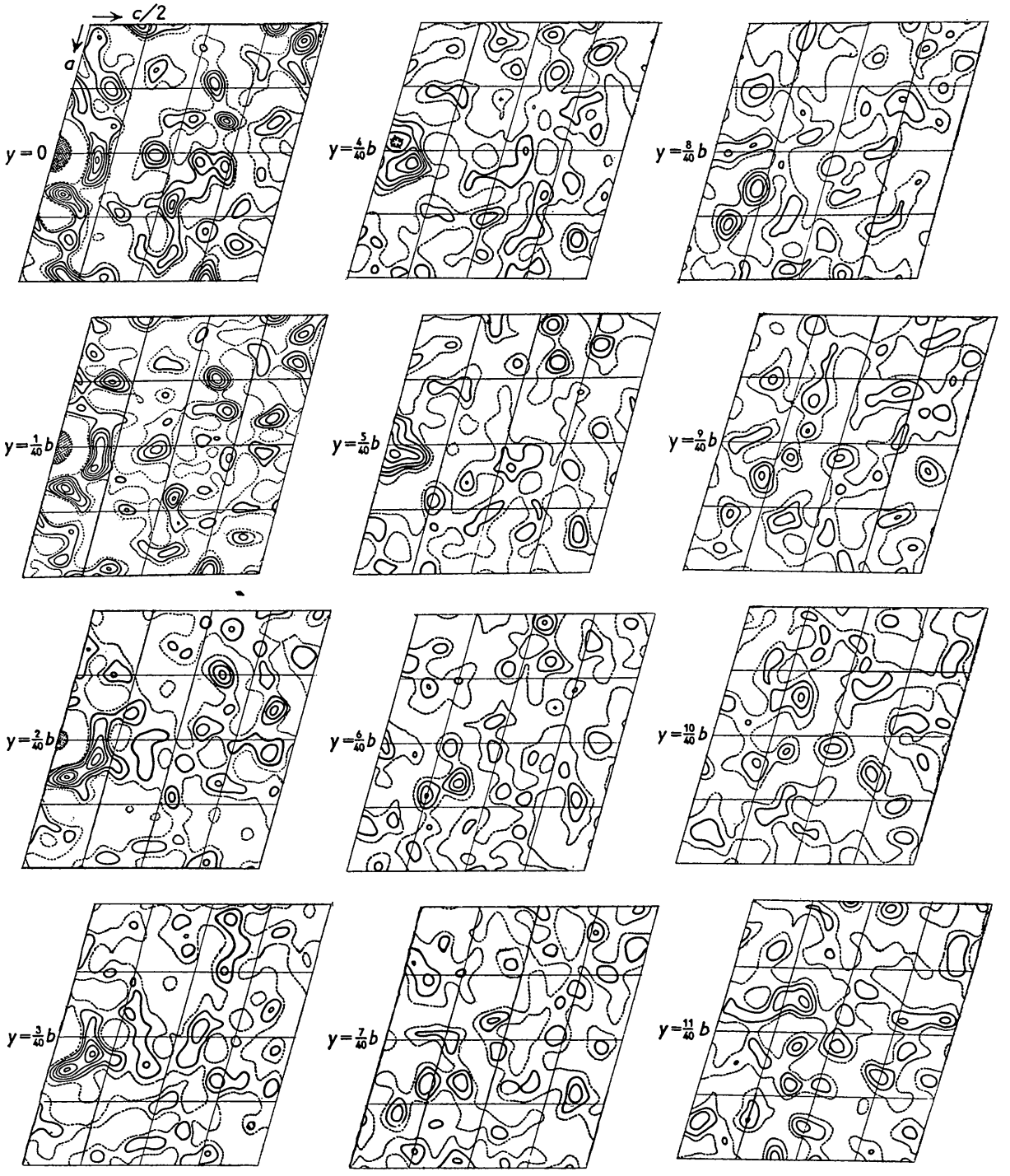

Fig. 2.

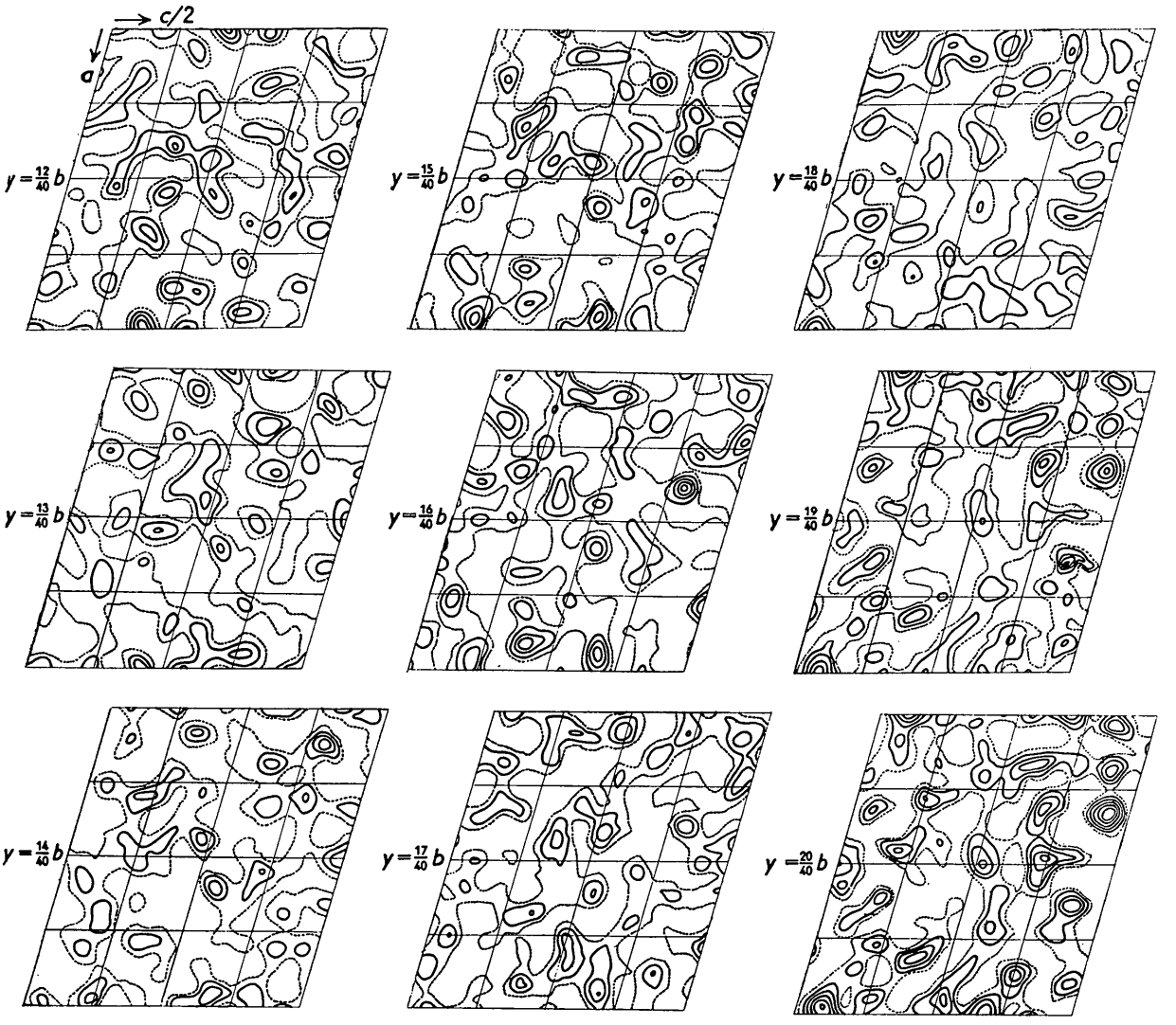

Fig. 2 (*cont.*)

Fig. 2. Twenty-one sections of the three-dimensional Patterson function. Contours at every 100 e.^{2}Å^{-3}; zero contours broken; negative contours omitted. Scale: 1 cm. = 7 Å.

pletely isotropic, more vectors falling in one general direction (about 30° from c in the direction of a) than in the plane perpendicular to this direction. This can be seen from Fig. 3(a), which shows a polar stereogram of the peak height of the 5 Å shell in all directions. The basal section of the Patterson function (Fig. 4) also shows this.

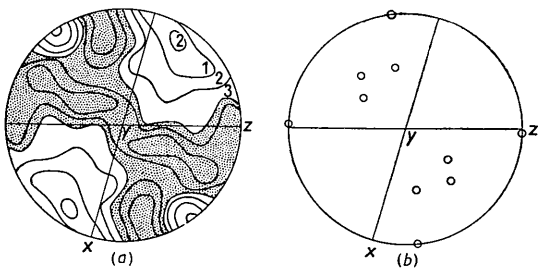

Fig. 3. (a) Polar stereogram of the 5 Å shell. Dotted regions are above 300 e.^{2}Å^{-3}. Other contours are marked in 10^{-2} e.^{2}Å^{-3}. (b) Polar stereogram of possible 'rod' directions.

The Mirror-plane effect.—The other striking feature of the Patterson function is the markedly greater variations in height of the contours in the basal and Harker sections ($y = 0$ and $y = b/2$) than in the general sections (Fig. 2). This was at first sight very puzzling, as the effect only extends a small distance (~ 1 Å) on either side of these sections in the y direction. This is shown in Table 1, which lists the number of peaks with heights greater than the third and fourth contour respectively. Thus any explanation depending merely on the structure of the molecule would have to be of a very unexpected character. On the assumption that the Patterson vectors are distributed rather

Table 1. *Number of peaks above* 300 e.^{2}Å^{-3} *in the three-dimensional Patterson*

Section (40y/b)	No. in the 5 Å shell	No. in the 10 Å shell	No. elsewhere	Total
0	4 (3)	3	9 (6)	16
1	3 (3)	2 (1)	8 (2)	13
2	2 (2)	—	2	4
3	2 (2)	—	—	2
4	2 (2)	—	—	2
5	2 (2)	—	—	2
6	—	2	1	3
7	—	—	—	0
8	—	2	—	2
9	—	1	—	1
10	—	—	1	1
11	—	—	1	1
12	—	—	2	2
13	—	—	—	0
14	—	—	1	1
15	—	—	3	3
16	—	—	3 (1)	3
17	—	—	1	1
18	—	—	1	1
19	—	—	4 (3)	4
20	—	—	15 (3)	15

The values in parentheses are the number of peaks which are above 400 e.^{2}Å^{-3}.

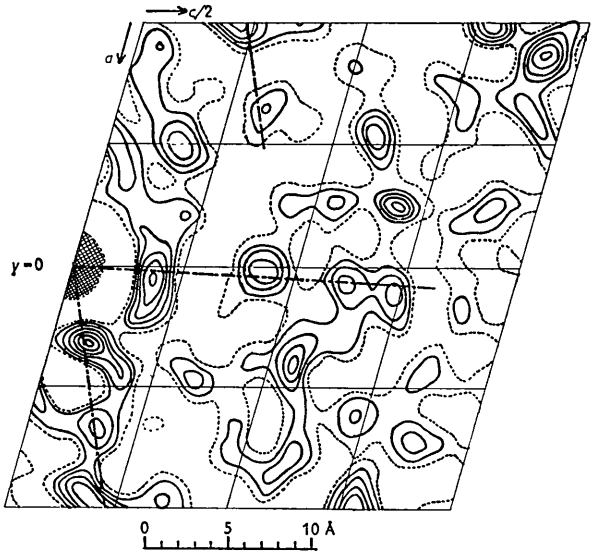

Fig. 4. Patterson section at $y = 0$. Zero contours broken; negative contours omitted. Heavy chain lines show 'rod' directions.

randomly we have been able to explain this as an effect due to the mirror planes in the Patterson.

Consider for the moment that the two protein molecules are made up of point atoms, and that the points are distributed more or less randomly throughout the whole unit cell. Next imagine the corresponding Patterson. This will also be made up of points, and they, too, will be randomly distributed throughout the volume of the whole unit cell. However, owing to the symmetry of the space group ($P2_1$) the Patterson will have mirror planes at the basal and Harker sections. Because of this any point representing a *general* vector has a companion point at its mirror image. (The Harker vectors, which necessarily fall on a mirror plane, are excluded here. They are in any case insignificant in structure determination when the number of atoms in the unit cell is very large.) Now if one of these general points happens by accident to lie exactly on one of the mirror planes it will coincide with its own image, and that point will have double weight.

The effect of having real atoms, rather than point atoms, will be to smear over all the points of the point Patterson. On this picture the variations of the Patterson from its average value of F_{000}^2/V are due to the random fluctuations produced by the small number of vector atoms which contribute to any point in the Patterson. These fluctuations will be proportional to the square root of the number of vector atoms 'contributing' to a point. But on the mirror planes we have *half* this number of vector atoms, each of *twice* the weight, so the fluctuations are $2/\sqrt{2} = \sqrt{2}$ times those found elsewhere. Naturally, owing to the finite size of the vector atoms, the effect spreads a little on either side of the mirror planes, but no more than a fraction of the width of the vector atom.

Notice particularly that it is not the average *number* of vectors which is greater near the mirror planes, but their *coherence*, and it is this which produces the enhanced peakiness.

The position of the molecules.—In actual practice the two protein molecules by no means fill the whole unit cell uniformly, and in fact they occupy only about half its volume, the rest being solvent, the effective electron density of which is probably very flat. This means that the point vectors are not uniformly distributed throughout the volume, and that there are slow fluctuations in the average density of the vector points. If there were no solvent (i.e. if the solvent had zero electron density) these effects would be very marked, but the solvent reduces the variations in the average level very considerably. It does not effect, however, the slow variations, due to the shape of the molecule, in the fine-scale peakiness, and this effect is superposed on the mirror-plane effect already discussed.

For example, the region of the Harker section (Fig. 2) near the y axis is noticeably more uniform than the rest of this section. This makes it very probable that the vector from the 'center' of one molecule to that of the other, which necessarily lies on the Harker section, does not fall into this central region. This is in line with other evidence about the relative location of the molecules (Magdoff & Crick, 1955*b*).

This argument should not be carried too far. Except for very symmetrical shapes with very sharp edges we would not expect to find a high peak at *exactly* the location of the intermolecular vector. Indeed, for most reasonable models of a protein molecule a 'center' is not a well-defined point but a general region, the

exact location of the 'center' depending upon how it is defined. Sharp peaks in the Patterson are due to the internal structure of the molecule, not to its overall shape.

It may be remarked, parenthetically, that the previous paragraph but one suggests a method for obtaining information about molecular location in cases where unfortunately the solvent has about the same electron density as the protein. In theory one should plot a smoothed value of p^2 (where p is the height of the Patterson above average). Notice that the Fourier components of this can be obtained directly from the observed intensities by a folding operation—it is not necessary to calculate the Patterson at all. Such Fourier components, especially the very low orders, might supply the same sort of information that one can obtain from the normal low-order reflections from a cell having a solvent of low electron density, i.e. in fortunate cases information about the positions and perhaps the shapes of the molecules.

Patterson rods.—A careful inspection of the Patterson has revealed nothing which could reasonably be described as a 'rod direction', that is, there is no continuous region of high vector density running straight out from the origin in any direction. We have therefore looked not for continuous rods, but lumpy ones—directions in which, say, there are large peaks at 5, 10 and 15 Å. It is to be expected on general grounds that the Patterson of an α-helix, or any other regular helix, when allowance is made for its side-chains, is by no means a continuous rod of high electron density, but has peaks corresponding to the distances between turns, and regions of lower electron density between these peaks. This has been confirmed for the α-helix by direct computation (Crick, unpublished).

We have found several directions along which, with a degree of faith, one might see such a sequence of peaks. These directions are marked on the polar stereogram shown in Fig. 3(*b*), and sections of the Patterson including five of these directions are given in Figs. 4 and 5.

Notice that a 'rod' in the basal plane has double weight, since the self-vectors of a rod in a general position occur in two positions in the Patterson (owing to the mirror planes) which coincide when it lies in the basal plane. Thus we do not feel that the *c*-axis direction singled out by Carlisle & Scouloudi (1951) is particularly significant. We have not calculated the expected height of vectors from an α-ribbon, but we should be very surprised if the structure they propose does not give vector heights near the *c* axis very much higher than are observed. We have this view on previous experience in computing the height of the vector rods in horse haemoglobin (Crick, 1952).

Whether the structure is likely to consist of rods lying not parallel, but in a variety of directions, is a more open question. It is perhaps significant that if this were so the mean rod direction would be not far from 30° from *c* in the direction of *a*, a direction given by Elliot (1952) as the approximate direction for the minimum infra-red dichroism at 4855 cm.$^{-1}$. Apart from the possible interference by the amide groups of the side chains, this might be expected to coincide with the mean chain direction.

Conclusion

We may reasonably conclude: (1) If the structure consists of rods their directions must be among those we have described; a structure with all the rods parallel to the *c* axis is unlikely. (2) The evidence for a rod structure is too indefinite to form the basis for a further attack on the structure at the present time. (3) Both the X-ray and the infra-red data show a small but significant degree of asymmetry, and both in the same direction.

It seems to us that these studies only serve to underline the need for the isomorphous-replacement method.

We should like to thank Dr Murray Vernon King for both growing and mounting all the crystals, and all our colleagues at the Protein Structure Project who

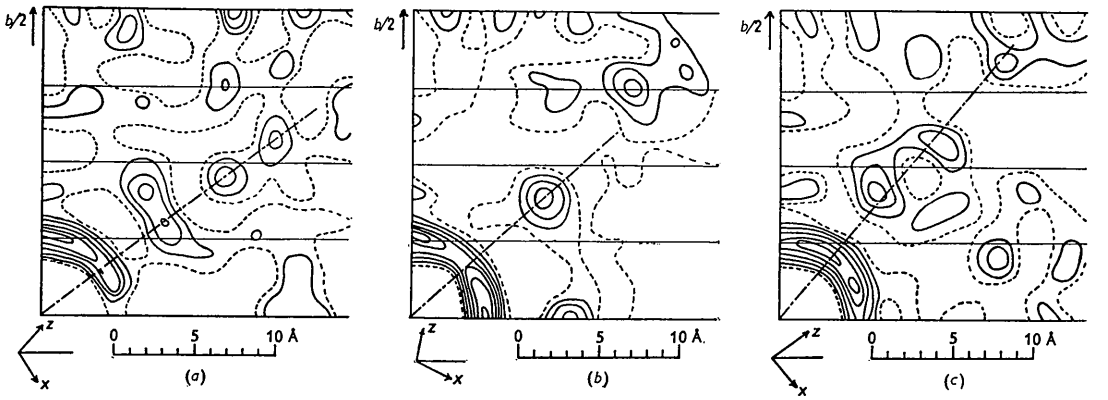

Fig. 5. Patterson sections showing 'rod' direction (chain line). Zero contours broken; negative contours omitted. The sections are perpendicular to the *ac* plane and parallel to a line (*a*) 50° from *c*, (*b*) 78° from *c*, (*c*) 36° from *c*.

assisted in collecting the data. The calculations were carried out on IBM machines at the Watson Laboratories, for which facilities we are very grateful.

Two of us (F. H. C. C. and V. L.) would like to thank Dr David Harker for the hospitality shown to us during our stay at the Project.

References

ARNDT, U. W. & RILEY, D. P. (1955). *Philos. Trans.* A, **247**, 409.

CARLISLE, C. H. & SCOULOUDI, H. (1951). *Proc. Roy. Soc.* A, **207**, 496.

CRICK, F. H. C. (1952). *Acta Cryst.* **5**, 381.

ELLIOT, A. (1952). *Proc. Roy. Soc.* A, **211**, 490.

FURNAS, T. C. & HARKER, D. (1955). *Rev. Sci. Instrum.* **26**, 449.

KING, M. V., MAGDOFF, B. S., ADELMAN, M. B. & HARKER, D. (1956). In manuscript.

MAGDOFF, B. S. (1953). *Acta Cryst.* **6**, 801.

MAGDOFF, B. S. & CRICK, F. H. C. (1955*a*). *Acta Cryst.* **8**, 461.

MAGDOFF, B. S. & CRICK, F. H. C. (1955*b*). *Acta Cryst.* **8**, 468.

Acta Cryst. (1956). **9**, 162

Calculation of the Electron Distribution in the Hydrogen Atom for Different Values of the Temperature Factor

BY T. R. R. MCDONALD*

Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England

(*Received* 12 *July* 1955)

Tables are presented showing the effect of thermal vibration on the electron distribution in the hydrogen atom, both in three dimensions and in projection. The calculations, which were performed on the EDSAC, are based upon the assumption that the charge density in a stationary atom is adequately represented by the square of the normalised (1*s*) wave-function. In the three-dimensional case, $\varrho(r)$, the electron density in a vibrating atom, and $n(R)$, the fraction of the electron contained within a sphere of radius R measured from the centre of the atom, can be expressed explicitly as functions of the temperature parameter B. It is found that the atom is quite diffuse, a sphere of radius 1·0 Å, for example, containing only about 70% of the electron. In view of this, due care must be taken in interpreting the results of electron counting on electron-density maps computed from X-ray diffraction data. The magnitudes of the errors introduced by termination of the Fourier series are also discussed.

The results of these calculations agree reasonably well with recent measurements of the electron density. We conclude tentatively that the electron distribution in hydrogen atoms bonded to carbon, nitrogen or oxygen, or weakly hydrogen-bonded to nitrogen or chlorine, is adequately approximated by the electron distribution in an isolated atom.

Introduction

In a recent publication, Higgs (1953) has discussed the effect of thermal vibration upon the electron distribution in the carbon atom, and has demonstrated the very critical nature of the dependence of the peak density on the amplitude of thermal vibration, especially at low temperatures. In the present paper, it is proposed to investigate the effect of thermal motion on the electron distribution in the hydrogen atom. In view of the fact that the technique of X-ray diffraction has advanced to the stage where it is possible to measure the electron density in hydrogen atoms with fair accuracy, it is of some importance to know what values to expect under different experimental conditions.

We shall assume in what follows that the charge density in a stationary atom is adequately represented by the square of the normalised (1*s*) wave-function, that is, we shall ignore the perturbations of the electron cloud due to bonding. Justification for this apparently drastic simplification is afforded by recent experimental results. It is well known that the electron distribution in heavier atoms is relatively little affected by bonding, the departures from spherical symmetry being in general very small (apart from the effects of anisotropic thermal vibration). Recent accurate X-ray diffraction results suggest that this may also be true in the least favourable case of the hydrogen atom. It has been generally found (Table 9) that, if a spherically symmetrical distribution of electron density is subtracted from X in a covalent bond H–X by the technique of the F_o-F_c synthesis, the resultant distribution has, on the average, spherical symmetry, and

* Present address: 23 Balcarres Street, Edinburgh 10, Scotland.