Acta Cryst. (1952). 5, 136

Arrangement of Polypeptide Chains in Horse Methaemoglobin

By W. L. Bragg and E. R. Howells

Cavendish Laboratory, Cambridge, England

and M. F. Perutz

Medical Research Council Unit, Cavendish Laboratory, Cambridge, England

(*Received* 7 *September* 1951)

The analysis in this paper is based on the assumption that the protein molecule consists for the greater part of parallel polypeptide chains in a hexagonal packing with an inter-chain distance of 10·5 Å. If this assumption is correct, the analysis assigns to the haemoglobin molecule an arrangement in which the chains, seen in end-on projection, are arranged in layers. There are three inner layers containing four, five and four chains respectively of nearly equal weight. The outer layers contain only two or three chains, and each of these chains is of less scattering power than those in the inner layer.

A study of the absolute intensities shows that the values of $F(0kl)$ due to the hexagonal array are only one-third of what would be expected if the polypeptide chains were straight and parallel throughout the molecule. Several factors are likely to reduce $|F(0kl)|$. The haemoglobin molecule consists of two halves joined, probably, along a plane which cuts across the chain direction. This implies that chains cannot continue in a straight line for more than about 30 Å before being forced to turn a corner. Corner turning, lack of alinement of the two halves of the molecule, non-uniformity in the distribution of the side-chains, and, perhaps, meandering of the chains from strict parallelism would all reduce $F(0kl)$. It is too early to judge whether this reduction would be sufficient to account for the low observed F's or whether part of the molecule is occupied by structures other than straight chains running parallel to the a axis.

1. Introduction

A Patterson projection along the a axis of crystals of horse methaemoglobin shows a hexagonal pattern of peaks spaced 10·5 Å apart (Boyes-Watson, Davidson & Perutz, 1947). The three-dimensional Patterson synthesis of the same crystal shows corresponding rods of vector density parallel to the a axis, which have been interpreted (Perutz, 1949) as arising from an array of polypeptide chains in the crystal parallel to the a axis. The present paper describes an attempt to find the positions in the molecule of the chains in end-on projection. Absolute intensity data were used; these were obtained by comparison of the intensities of reflexions from the $(0kl)$ zone with those from a standard anthracene crystal, using both precession photographs and a Geiger-counter spectrometer.

Whatever its precise configuration, the central part of a polypeptide chain may be expected to be a region of relatively high density and scattering power in an end-on projection, because the atoms –CO–$\underset{|}{\text{CH}}$–NH– are closely linked by homopolar bonds. For example, in a structure consisting of close-packed 3·7 residue helices (Pauling, Corey & Branson, 1951) spaced 10·5 Å apart, the mean electron density in the main –CO–$\underset{|}{\text{CH}}$–NH– chain is about 2·5 times greater than in the volume occupied by the side-chains. It will be assumed that the majority of the chains are of approximately equal length and scattering power. The overall molecular dimensions and the total number of residues limit the number of chains in a molecule, and standard methods of X-ray analysis can be used to find positions for the projected chains which explain the values of $|F(0kl)|$; they play the same role as the atoms in a simpler structure.

As in all X-ray analysis, no solution is possible unless definite assumptions are made about the physical nature of the scattering units. The solution proposed here is based on the chain interpretation of the Patterson diagrams, and is of significance only if it is correct.

2. Possible models

The Patterson diagrams indicate a hexagonal packing of the chains of the nature shown in Fig. 1 (*a*), with a distance of 10·5 Å between the chains. There are some considerations which limit the possible variations of this pattern. The packing of the molecules in this and other crystalline forms of haemoglobin (Bragg & Perutz, 1952) shows that when the molecule is viewed in this projection it may be enclosed in a circle of diameter between 50 and 55 Å. There cannot therefore be more than five chains in the b direction along the central row, or more than five layers of chains in the c direction. There are in all about 580 residues in the molecule, and since there is strong evidence that residues are spaced at intervals of 1·5 Å along a chain

(Pauling & Corey, 1951; Perutz, 1951), the combined length of all chains in the molecule cannot exceed 870 Å. Packing also indicates that the overall length of the hydrated molecule in the *a* direction is from 70 to 75 Å, so, at the most, individual chains cannot be more than 55–65 Å in length; about fifteen such chains would account for the 580 residues. These round figures serve as a guide in a first attempt to find a model.

A rough comparison of calculated and observed (0*kl*) intensities was carried out by the method of the 'fly's eye' optical analogy, the projection being represented by variations of a hexagonal array of scattering

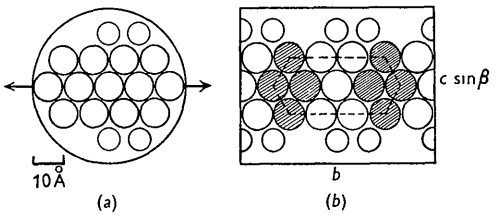

Fig. 1. (*a*) An idealized haemoglobin molecule, *a* projection; the fit in a circle of 52·2 Å diameter is shown. (*b*) Corresponding unit-cell projection.

points. It was found that an approximate imitation could be obtained only if the points were arranged in an odd number of layers with the middle layer lying on the twofold *b* axis. Accepting an odd number of layers, the intensities of the 00*l* reflexions were not compatible with a number of layers other than five. Better agreement is obtained if a smaller weight is given to the points in the top and bottom layers, and packing considerations also indicate that the chains must be shorter in this part of the molecule. The optical method merely served to suggest the type of model which it would be profitable to investigate by more precise methods.

3. The symmetry of the molecule

The space group of the crystal is *C*2. *F*(0*kl*) is therefore complex, and refinement by Fourier methods would at first sight seem unpromising. However, the molecular models which gave a measure of agreement in the 'fly's eye' test were all nearly centrosymmetric, and this suggested that the phases of *F*(0*kl*) might be nearly 0 or π, at any rate for the lower orders. This was confirmed by Wilson's statistical method (Wilson, 1949), illustrated in Fig. 2. In this method the reflexions are classed in order of the ratio *Z* of their intensity to the mean intensity, and *N*(*Z*) is plotted against *Z*. The upper curve is the theoretical curve for crystals with a symmetry centre or its equivalent in a projection, the lower curve for crystals without such a centre. The crosses, representing the statistical distribution for the lower orders, fall more closely on the upper curve, while the statistical distribution for the higher orders is nearer the lower curve. If the molecule is nearly centro-symmetric in its general structure, as this would appear to indicate, it is possible to make a model of it by assigning plus or minus signs to the *F* values. The model will of course have a false symmetry centre. It will, however, correspond to the superposition of the actual molecule and its image in a symmetry centre, and since it has been concluded that the assymmetry is a matter of its finer detail the general picture will be correct.

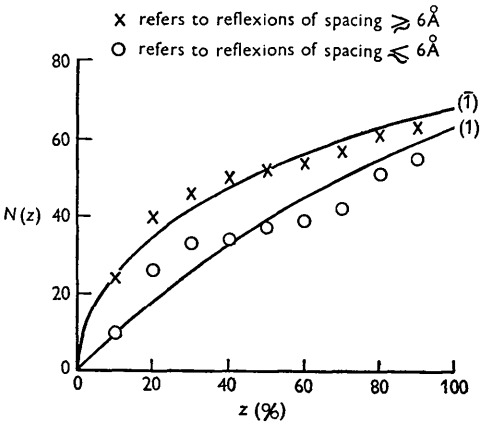

Fig. 2. Intensity distributions for normal horse methaemoglobin, *a* projection.

4. The assignment of signs to the values of *F*(0*kl*)

The haemoglobin lattice can be made to swell or shrink by changing its environment. The alteration takes place by changes in the relative positions of the molecules, but without changes in their internal structure (Boyes-Watson *et al.* 1947). A plot of the observed $|F|$ values in reciprocal space for three stages gives the moduli of the molecular transform at a large number of points, as shown in Fig. 3 (*a*). The high values occur in groups. There is a strong presumption that values of *F*(0*kl*) in any one of these groups have the same sign. A Fourier transform of the arrangement which gave promising agreement by the 'fly's eye' method is shown in Fig. 3 (*b*). Weights of 1 have been assigned to chains in the three central layers and of $\frac{1}{4}$, $\frac{1}{2}$, $\frac{1}{4}$ to the three chains in the upper and lower layers (see § 2). It is seen that the peaks in this transform coincide with the groups of values of high $|F|$'s, and give further indication that the $|F|$'s within any such group have the same sign. The zero contours of this transform, also, have been drawn in Fig. 3 (*a*). Their positions depend critically on the exact geometry of the idealized model; in the case of the 08*l* reflexions the signs indicated by the Fourier transform did not agree with those deduced from condition (*b*) below, and the choice indicated by the latter was adopted.

The way in which signs were chosen will here be indicated only in broad outline, because its justification mainly depends on the appearance of the final Fourier projections and only in part on the method of choice. The molecules are arranged in sheets parallel to (001), and the protein has a higher scattering power than the

liquid between the molecules. A choice of positive sign for $|F(001)|$ is therefore equivalent to the choice of an origin in these sheets. The hexagonal pattern of peaks in the Patterson projection is mainly due to the strong 062 and 063 reflexions, which are almost certainly of the same sign because they occur in so well marked a group of high values in Fig. 3 (*a*). A choice of positive sign for these reflexions is equivalent to placing the origin at a peak representing a chain in the layer along the *b* axis (see Fig. 1). If the negative sign the places where the units are shaded should be twice the height of the other peaks.

(*b*) An approximately uniform level in the region at height $\frac{1}{2}c$ which is believed to lie in the region occupied by the liquid, which is uniform in density at this resolution and should not show any marked peaks. In order that this may be so, $\Sigma F(0kl)_{l\,\mathrm{even}} - \Sigma F(0kl)_{l\,\mathrm{odd}}$ must be small for each value of k, and this imposes restrictions on the choice of signs, as was shown independently by Crick (1952).

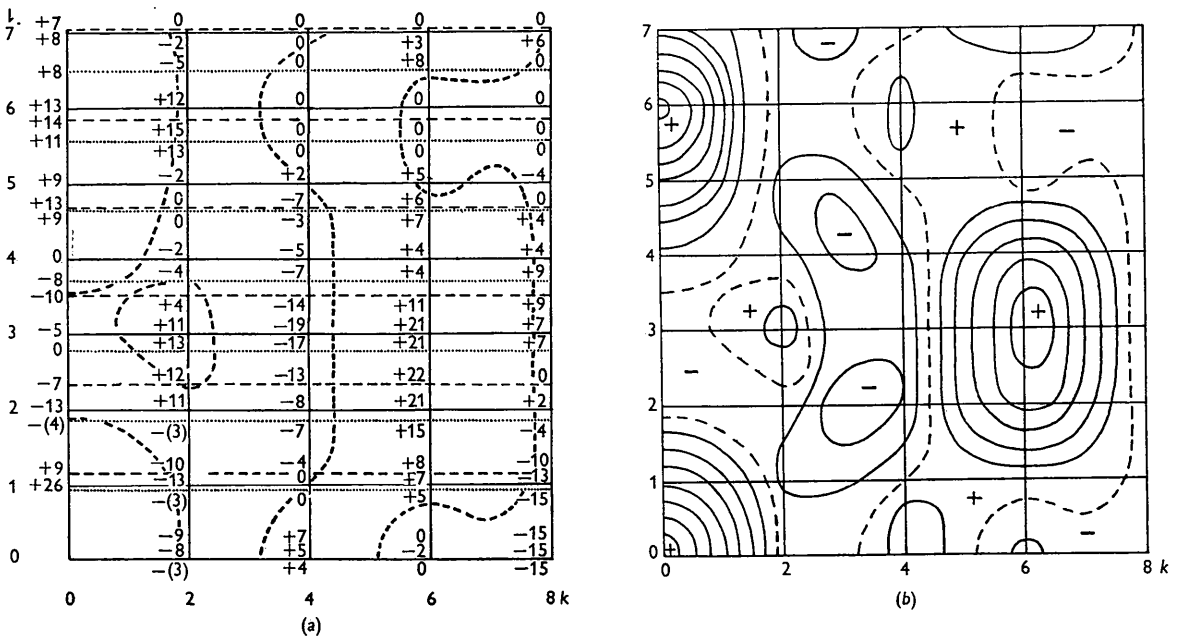

Fig. 3. (*a*) Plot of absolute F values (in units of 100 electrons), in reciprocal space, for three stages of shrinkage and swelling. The signs indicated are those used in Fourier summations. Zero level of molecular transform is shown by broken line. (The F values in brackets are small because the expanded stage contained 3·5M of $(NH_4)_2SO_4$.) (*b*) Transform of a structure composed of molecules like that in Fig. 1 (*a*). Chains taken as point scatterers of weights 1, $\frac{1}{2}$ and $\frac{1}{4}$.

were chosen, it would merely be equivalent to shifting the origin by $\frac{1}{4}b$ along the layer so that it fell between two chains. We have chosen the former alternative. The group of high values in the region (005), (006), (007) must also be positive if there is an odd number of layers and the origin lies in the central layer. In effect, the three clusters of high positive values in reciprocal space around the origin, (062) and (006) appear in any transform of a hexagonal net in real space with 10·5 Å spacing and one of its points coinciding with the origin. Two criteria were used in selecting the remaining signs:

(*a*) A clear pattern of peaks representing the chains in the three central layers which indicates an approximate equality in weight of these chains. It was mainly this condition which decided the negative sign of $|F(043)|$, the strongest reflexion after 062 and 063, since a positive sign for this reflexion results in a peak of dominant height at the origin. The overlap shown in Fig. 1 (*b*) must be taken into account; the peaks at

5. The Fourier projections

The signs finally chosen are indicated in Fig. 3 (*a*), and Fig. 4 shows the Fourier series formed of terms with the observed amplitudes and these signs. For the purpose of these Fourier summations an artificial temperature factor was applied to the observed F's, such that the amplitudes of all high-order reflexions ($d < 7$ Å) were reduced to less than half the mean value of the low-order ones ($d > 7$ Å). Terms with F values of less than half the mean were omitted from the Fourier summations, as were all terms with $d \leqslant 7$ Å. Trial calculations with all the low-order terms included showed that they do not affect the positions of the peaks, though their shapes and heights alter slightly depending on the signs of the weak terms. The three upper figures represent three stages of shrinkage or expansion, the origin being at the left-hand bottom corner. The lower figure shows the complete unit cell for the normal state with the origin at its centre. The *b* axis is halved in this projection because the *c* face

of the crystal is centred. The following points may be noted:

(*a*) The three representations of the molecule are almost identical, although such very different values of

(*c*) The area where it is believed that there is only water or salt solution is at an approximately uniform level in all three projections, and the density corresponds to the density of the water or the salt solution.

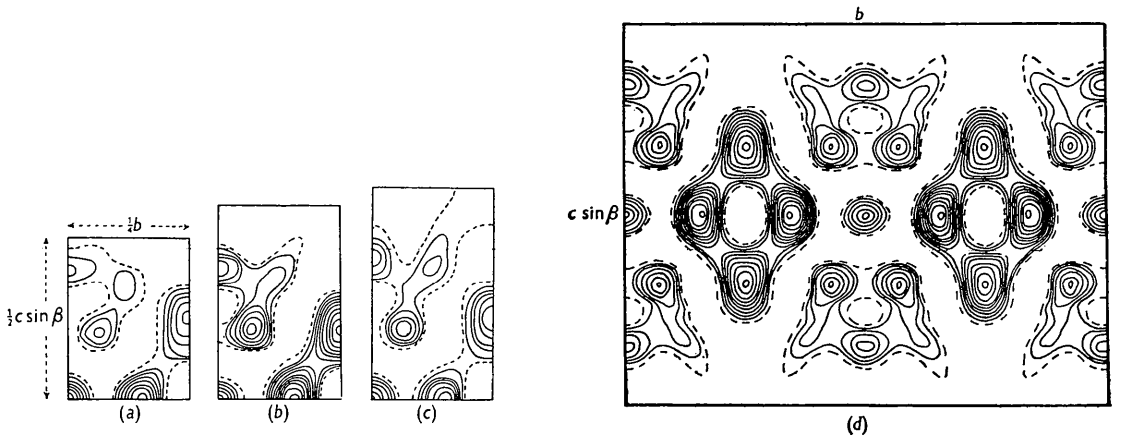

Fig. 4. (*a*), (*b*), (*c*) Fourier projections for three stages. (*d*) Fourier projection of whole unit cell, normal crystal. Positive contours drawn at intervals of 1 e.Å^{-2}. Zero contour shown by broken line. F (000) = 40 e.Å^{-2} approximately.

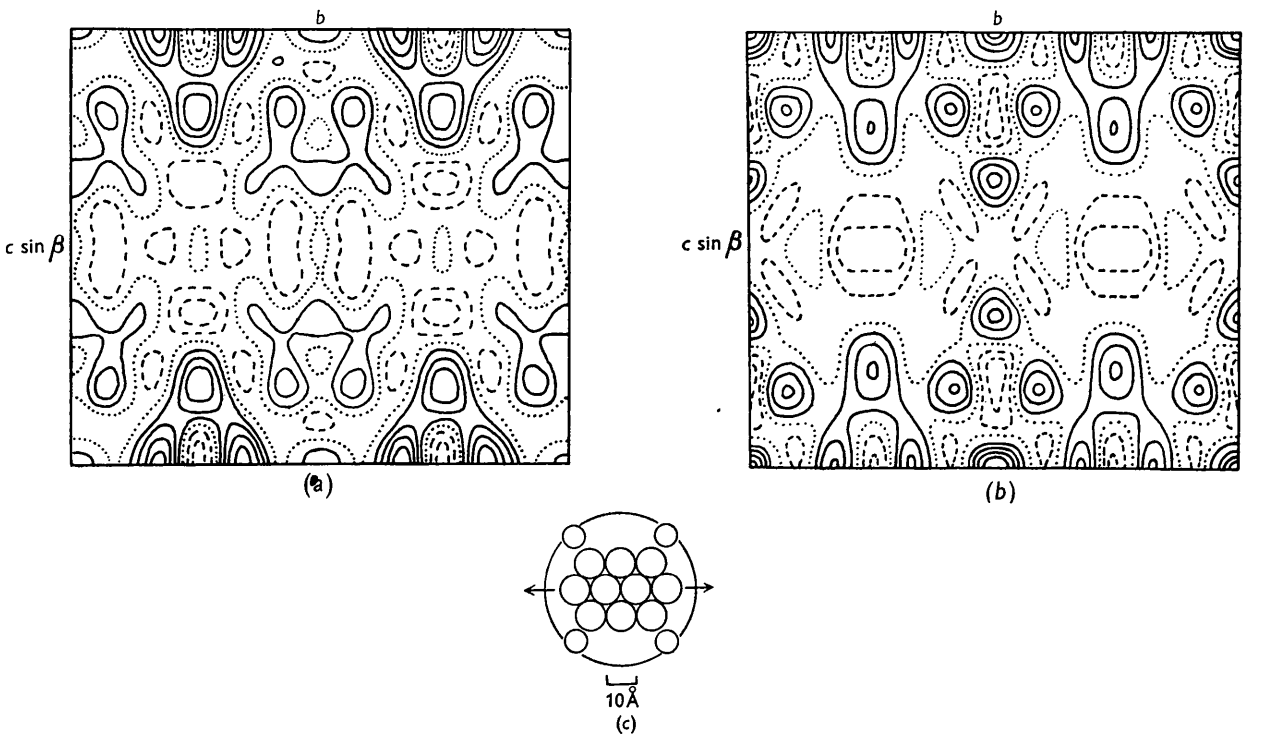

Fig. 5. (*a*) Fourier projection corresponding to Fig. 4(*d*) calculated on the Pepinsky machine. (*b*) A Fourier projection with an alternative choice of signs. (*c*) Representation of the molecule implied by Fig. 5(*b*).

$|F(0kl)|$ have been used in their formation. The peaks representing the chains are not so high in the expanded stage (*c*), but in this case a denser solution of $(NH_4)_2SO_4$ replaces the water and the peaks emerge less from the average level.

(*b*) The peaks where chains from neighbouring molecules overlap are higher than those representing single chains, as is to be expected.

We are indebted by Dr Ray Pepinsky at State College, Pennyslvania, for the use of his machine to test a large number of alternative choices of the signs. Fig. 5 shows two sample photographs of the contour lines drawn by the machine. Fig. 5(*a*) represents a series formed with the signs we have chosen, and Fig. 5(*b*) one of the most plausible alternatives. Fig. 5(*c*) shows a representation of the molecule implied by

the Fourier projection of Fig. 5 (*b*). The central layer now contains four chains, the next layer three and the outer layer two shorter chains placed in improbable positions at the edge of the molecule. The molecule contains only twelve chains of an average length of 73 Å, while its diameter along *b* would be only 42 Å; these dimensions are not in accord with those deduced by Bragg & Perutz (1952), and it seems improbable, therefore, that this picture of the structure is right. Other alternatives were rejected because they did not give clearly marked peaks, or gave peaks in the area occupied by the liquid. The photographs confirmed our tests, and showed that a picture of the molecule in accord with the assumptions made here could only be formed with signs which in the main agreed with those we have chosen.

6. The form of the molecules

In interpreting Fig. 4 (*d*) it must be borne in mind that the molecules are overlapping in the *b* direction and that a false symmetry has been introduced by the assumption that $F(0kl)$ is real. The most simple interpretation is a molecule much as in Fig. 1 (*a*). The higher peaks in Fig. 4 (*d*) would be expected to represent the region of overlapping of the molecules, and indicate a molecular centre at the centre of the figure. In other words, it is more plausible to suppose that a chain coincides with the origin at the centre of the molecule, as it would if there were five chains in the *b* direction; the centre of the molecule does not fall between two chains, as it would if there were four chains in this direction. There is only one overlap in the next layer, as would be the case if it consisted of four chains (see Fig. 1). The chains in the top and bottom layers have less weight. If they had half the weight, and were arranged as in Fig. 1 (*a*), the false symmetry would give a chain of weight $\frac{1}{2}$ at the centre of the layer, flanked by two chains of weight $\frac{1}{4}$. This is in broad agreement with the appearance of the Fourier projections. This last suggestion is of course purely speculative, but it illustrates the extent to which the false symmetry of the projection may conceal a departure of the actual molecule from centro-symmetry.

7. The heights of the Fourier peaks

The absolute values of $F(0kl)$, and consequently the heights of the peaks in the Fourier series, are much less than would be expected from straight polypeptide chains parallel to the *a* axis of the crystal. 'Straight' applies here to the central axis around which the elements of the chain are coiled or folded in the various models of the chain which have been put forward; in all of them the –CO–NH–$\underset{|}{\text{C}}$H– atoms are within a circle of about 3 Å from this axis. There are 580 residues in the molecule, with 29 electrons per residue in the main chains, so that the total number of electrons per cell in the main chains, as distinct from the side groups, is $2 \times 29 \times 580 = 33{,}600$. The total number of electrons in the unit cell, $F(000)$, is 140,000. The contribution of a strong reflexion such as 063 from these chain electrons alone will therefore be,

$$33{,}600 \times 0{\cdot}66 \times 0{\cdot}5 \times 0{\cdot}9 \times 0{\cdot}6 = 6000,$$

where 0·66 is the factor by which F is reduced owing to the arrangement of the chains in the unit cell, 0·5 represents the fall in diffracting power of the chain owing to its finite cross-section for the case of the 3·7 residue helix which is taken as a representative model, 0·9 is the temperature factor deduced from the smallest observable spacing, and 0·6 the factor by which the chain height is reduced by the background of the atoms in the side-chains if these are assumed to be uniformly distributed. The actual value of $|F(063)|$ is 2050, only one-third of the calculated estimate. The contribution to F by the main chain may of course be partly offset by the non-uniform distribution of the side-chains and liquid, but it is unlikely that their distribution should depart so far from uniformity as to balance the concentrated scattering matter in the chains to the extent of reducing 6000 to 2050.

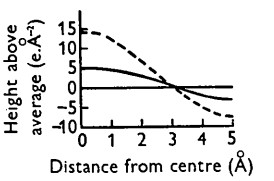

Fig. 6. Comparison of the distribution of electron density in a typical Fourier peak (full line), with that calculated for the end-on view of 54 Å length of Pauling α helix (broken line), seen at the same resolution as that of the Fourier projections of Fig. 4.

Another estimate of the discrepancy between the actual structure and an idealized model consisting of straight parallel chains was obtained by comparing the electron-density distribution of the 3·7 residue helix seen in end-on projection (at the resolution corresponding to our Fourier projections) with the actual electron-density distribution of one of the peaks in Fig. 4. This comparison (Fig. 6) shows the height of the observed peak to be only about one-third of the calculated one. Similar results were obtained by Crick (1952) by comparing the observed vector densities in the 'rods' of the three-dimensional Patterson synthesis with those calculated for simple models consisting of straight parallel chains.

To explain the discrepancy we must assume either (*a*) that only part of the molecule consists of straight parallel chains; or (*b*) that the peaks in Fig. 4 do not represent single lengths of chain, but two (or possibly more) separate lengths of chain which are not exactly alined; or (*c*) that the chains are far from straight, so that their scattering power falls off very rapidly with increasing $\sin\theta/\lambda$. It seems probable that all three causes contribute to the effect; (*a*) and (*b*) because the molecule consists of two equal halves joined, probably,

along a plane which cuts across the chain direction, so that part of the total length of chain is taken up with corner turning, and (*c*) because the bulkier side-chains may cause the main chains to meander.

While possibility (*c*) is entirely speculative at the present stage, there are sufficient data to allow more detailed discussion of (*a*) and (*b*). The haemoglobin molecule consists of two chemically and structurally identical halves; this is shown by its dyad symmetry in the crystal, by its dissociation in dilute solution and by the fact that the molecular weight of globin is

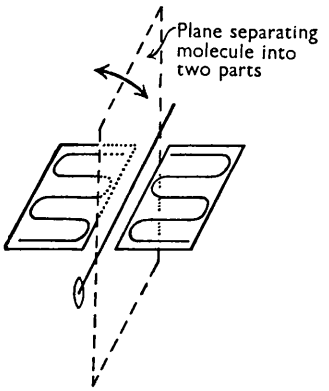

Fig. 7. The splitting of the central layer of polypeptide chains by the crystallographic dyad axis.

33,000, while that of haemoglobin is 67,000. If the two halves of the molecule are related by the dyad axis, they are likely to be joined together along a plane containing that axis. Since the dyad is level with the central layer of chains, the plane joining the two halves must cut across the chain direction (Fig. 7). The total length of the molecule along *a* is about 60–70 Å (Bragg & Perutz, 1952), which means that no chain can continue along a straight line for more than about 30 Å before it is forced to turn a corner. According to Porter & Sanger (1949, p. 121) globin contains three separate chains, comprising 290 residues, which would be about 440 Å long altogether. If these are to be split up into 30 Å lengths, there must be 15 such lengths and the three chains must turn 12 corners, making 24 corners for the whole haemoglobin molecule.

The fraction of chain taken up with turning corners would diminish $|F(063)|$ both by reducing the total length of straight chain and by superimposing a non-uniform background between them. No quantitative estimate of this effect is as yet possible. Further reduction may be due to possibility (*b*), i.e. lack of alinement of the chains in the two halves of the molecule. Some such lack of alinement is indicated by the weakness of $|F(006)|$ relative to $|F(063)|$. The ratio of these two amplitudes (0·62) would be compatible with a relative displacement of the two halves by about 2·5 Å in the direction normal to 001, which would reduce $|F(063)|$ by a factor of about 0·9.

For the time being it is impossible to judge whether corner turning, lack of alinement, non-uniform background and meandering from straight parallelism can account for the low absolute value of F, or whether part of the molecule is occupied by structures other than straight chains running parallel to the *a* axis.

The authors wish to thank Dr Ray Pepinsky of State College, Pennsylvania, for his kindness in calculating sixty-four alternative Fourier projections on his X RAC machine, and Mr H. E. Huxley for taking two of the precession photographs. They are grateful to the Medical Research Council for financing this research. One of them (E.R.H.) is indebted to the University of Wales for a Fellowship.

References

Boyes-Watson, J., Davidson, E. & Perutz, M. F. (1947). *Proc. Roy. Soc.* A, **191**, 83.
Bragg, W. L. & Perutz, M. F. (1952). *Acta Cryst.* In the press.
Crick, F. H. C. (1952). *Acta Cryst.* In the press.
Pauling, L., Corey, R. B. & Branson, H. R. (1951). *Proc. Nat. Acad. Sci., Wash.*, **37**, 205.
Perutz, M. F. (1949). *Proc. Roy. Soc.* A, **195**, 474.
Perutz, M. F. (1951). *Nature, Lond.*, **167**, 1053.
Porter, R. R. & Sanger, F. (1949). *Haemoglobin.* London: Butterworth.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318.

Short Communications

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 500 words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible; and proofs will not generally be submitted to authors. Publication will be quicker if the contributions are without illustrations.

Acta Cryst. (1952). **5**, 141

An illustration of the optical basis of Wilson's X-ray method for detecting centres of symmetry.

By C. A. Taylor, *College of Technology, Manchester* 1, *England*

(*Received* 27 *July* 1951)

During the development of optical methods for interpreting X-ray diffraction data (Lipson & Taylor, 1951; Taylor & Lipson, 1951), it has been necessary to prepare optical diffraction patterns of both centrosymmetrical and non-centrosymmetrical objects. The patterns given by these two types of object differ in a most striking way, as can be seen from the typical examples given in the figures. Diffraction patterns derived from centrosym-