has provided information about the nature of the bonds formed by an atom of a transition element with covalence 3 and with one *d* orbital available. When one *d* orbital is available and two bonds are formed, the bonds are linearly directed, as was shown by quantum mechanical calculation of the two best *spd* hybrid bond orbitals that can be constructed with only one *d* orbital available (Pauling, 1931). When four bonds are formed with use of one *d* orbital, they are directed toward the corners of a square. The configurations observed for gold and silver in the three minerals under discussion indicate that maximum stability is achieved for a tercovalent element with one *d* orbital available for bond formation through the formation of two oppositely directed single bonds, the remaining valences being used for the formation of bonds in the plane at right angles to the line formed by the first two (Pauling, 1931). These considerations suggest a refinement in the method of calculating bond numbers. On the assumption that the bond-forming power of an orbital is proportional to its concentration in the bond direction, the two equivalent orbitals corresponding to the two best bonds that can be formed with the use of a single *d* orbital are of the form

$$\frac{1}{2\sqrt{3}}s \pm \frac{1}{\sqrt{2}}p_z + \frac{\sqrt{5}}{2\sqrt{3}}d_z \qquad \text{(Pauling, 1931).}$$

They have $\frac{5}{12}d$ character (the square of the coefficient of d_z in the above equation), which corresponds to the single-bond radius for gold 1·309 Å.* We accordingly predict, with use of the single-bond radius for tellurium, 1·348 Å, that the gold–tellurium distance for two oppositely directed single bonds would be 2·66 Å, which agrees exactly with the average of the values for the corresponding bonds in Table 4, which range from 2·63 to 2·68 Å. With this refinement, the bond numbers of the longer gold–tellurium bonds would be calculated with use of the radius 1·43 Å† for gold, and their values would be correspondingly a little larger.

The authors wish to express their thanks to Dr Donald V. Higgs for making the drawings of Figs. 1 and 2.

References

ELLIOTT, N. & PAULING, L. (1938). *J. Amer. Chem. Soc.* **60**, 1846.
HARVEY, R. D. (1928). *Econ. Geol.* **23**, 778.
PAULING, L. (1931). *J. Amer. Chem. Soc.* **53**, 1367.
PAULING, L. (1947). *J. Amer. Chem. Soc.* **69**, 542.
PAULING, L. (1948*a*). *Nature, Lond.* **161**, 1019.
PAULING, L. (1948*b*). *Proc. Roy. Soc.* A, **196**, 343.
PAULING, L. (1948*c*). *Victor Henri Memorial Volume.* Liège: Desoer.
PAULING, L. (1948*d*). *J. Chem. Soc.* p. 1461.
PEACOCK, M. A. & THOMPSON, R. M. (1946). *Amer. Min.* **31**, 515.
TUNELL, G. (1941). *Amer. Min.* **26**, 457.
TUNELL, G. & KSANDA, C. J. (1935). *J. Wash. Acad. Sci.* **25**, 32.
TUNELL, G. & KSANDA, C. J. (1936*a*). *J. Wash. Acad. Sci.* **26**, 507.
TUNELL, G. & KSANDA, C. J. (1936*b*). *J. Wash. Acad. Sci.* **26**, 509.
TUNELL, G. & MURATA, K. J. (1950). *Amer. Min.* **35**, 959.

* Calculated by means of equation (12 *c*) (Pauling, 1948 *b*) with $\delta = \frac{5}{12}$.

† Calculated by means of equation (12 *c*) (Pauling, 1948 *b*) with $\delta = \frac{2}{12}$.

Acta Cryst. (1952). **5**, 381

The Height of the Vector Rods in the Three-Dimensional Patterson of Haemoglobin

BY F. H. C. CRICK

The Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge, England

(*Received* 15 *November* 1951)

A method is described for calculating the average height of the vector rods in the three-dimensional Patterson of haemoglobin. The calculations have been carried out for the α-helix. They show that the simple models previously considered for haemoglobin give peaks which are too high, and that to obtain even rough agreement either less than about half the protein must be put into chains parallel to the *X* axis or considerable irregularity must be introduced. The evidence suggests that the folded polypeptide chain does not run in one direction for more than 20 Å at a stretch.

Introduction

The three-dimensional Patterson synthesis of horse methaemoglobin has, as one of its features, rod-shaped regions of high vector density, parallel to the *X* axis, which Perutz (1949) interpreted in terms of straight rods of high electron density in the molecule; he suggested that these rods were due to folded polypeptide chains arranged in hexagonal close packing

and spaced 10·5 Å apart, with their length parallel to X. Bragg, Howells & Perutz (1952) have shown that the $|F(0kl)|$'s are consistent with such an arrangement.

Pauling, Corey & Branson (1951) have proposed a helical fold for the polypeptide chain known as the 3·7 residue α-helix. When so folded, the amino acid residues repeat at regular intervals of 1·5 Å along the chain direction. Perutz (1951*a*, *b*) has reported a reflexion of 1·5 Å spacing from planes perpendicular to the fibre axis in artificial polypeptides and fibrous proteins of the α-keratin type and also, though less clearly, in haemoglobin. The reflexion in haemoglobin is possibly weaker than in hair, and certainly weaker than in the artificial polypeptides (Perutz, unpublished).

The above facts might suggest that the structure of haemoglobin consisted mainly of polypeptide chains folded into 3·7 α-helices, and packed side by side parallel to the X axis of the crystal. Rough calculations show that the vector rods to be expected from such a model would be more dense than those observed. The present paper describes an attempt to calculate rather more accurately the absolute vector density of the rods on the basis of certain simple models, and to compare the results with the observed density.

A parallel attack on this problem has been made by Bragg, Howells & Perutz (1952), who, working in two dimensions on the X projection, have shown that the absolute value of $|F(063)|$ is only one-third of the value calculated for a model consisting entirely of straight and parallel chains extending throughout the length of the molecule.

We shall use the terms 'coiled chains' or 'rods' to denote the coiled polypeptide chains, and 'vector rods' to denote the regions of high vector density they produce in a Patterson. In what follows 'haemoglobin' will refer to horse methaemoglobin in the usual monoclinic form.

Method of calculation

It is necessary to have a method which is sufficiently accurate without being too laborious. Simple methods involving merely counting vectors tend to be inaccurate; complete structure-factor and Patterson calculations would be prohibitively lengthy. The method described here is a compromise.

The basic idea is to reduce the three-dimensional calculation to a two-dimensional one. This is done by calculating the *average* vector height over a short length in the rod direction. The steps in the calculation are as follows:

(*a*) *The idealized model.*—A crystal is taken to consist entirely of infinite lengths of α-helices packed in an infinite regular hexagonal array. The Patterson of an end-on projection of the chains is calculated.

(*b*) *The real model.*—The haemoglobin molecule is considered to be made up of a certain number of lengths of α-helix distributed in a certain way. The electron density of the α-helix is assumed to be uniform in the rod direction; by counting vectors the *ratio* is found between the vector densities, in three dimensions, of the idealized and the real model at chosen points in their Pattersons.

(*c*) *The experimental data.*—Consider the chosen point in the observed Patterson. The *average* value over a length of 3 Å on either side of it in the rod direction is evaluated. This is then compared with the calculated value. It is assumed that this averaging will compensate for the theoretical assumption that the electron density of the α-helix was uniform in the rod direction.

The advantages of this method are that in step (*a*) the effects of side chains, heat motion, diffraction etc., can be easily and accurately allowed for. This is difficult to do if a vector-counting method is used throughout. Vector counting is then employed in step (*b*) to get the *ratio* of vector density between a known and an unknown case. Finally, since the observed three-dimensional Patterson has already been computed by Perutz (1949), the comparison with observation can easily be made.

In what follows, attention will be mainly concentrated on the part of the three-dimensional Patterson of the haemoglobin crystal in the neighbourhood of the X axis, and averaged between $X = +9$ and $+15$ Å. This stretch is chosen because it avoids the region very near the origin, and because the vectors in it are mainly due to each short length of α-helix with itself, and thus the relative arrangement of the various lengths of α-helix in the molecule is not important here.

All calculations will be made in terms not of the absolute height of the Patterson but of the height above the average (above zero as usually plotted with $F^2(000)$ omitted). It can be shown that this can be done, and the liquid in the crystal ignored, if that liquid has an electron density near the average electron density of the protein, as happens to be the case here (Wrinch, 1950).

Finally, it should be stated that the above procedure can be looked at in a more rigorous manner by considering how a Patterson can be built up from the 'self-Pattersons' and 'cross-Pattersons' of *parts* of the unit cell, and by working throughout with a fictitious electron density given by

(actual electron density)$-$(average electron density).

Subsidiary calculations from this standpoint have shown, for example, that diffraction effects in this particular problem are probably negligible, and that the averaging process will not produce serious errors.

Simple theory

Consider a single cylinder of constant electron density ϱ e.Å^{-3}, of cross-sectional area A, and of total length L placed between $X = -\frac{1}{2}L$ and $X = +\frac{1}{2}L$, where X

is the axial direction. The average vector density of its Patterson between the planes $X = +l_1$ and $X = +l_2$ is

$$[L-\tfrac{1}{2}(l_1+l_2)]\varrho^2 A/k \text{ e.}^2\text{Å}^{-3},$$

all dimensions being in Ångström units. Here kA is the cross-sectional area over which the Patterson vectors spread, and over which the average is taken.

Consider such rods, but of infinite length in an infinite hexagonal lattice and so spaced that the vectors in Patterson space from one rod to itself do not overlap those between neighbouring rods; then the vector density, averaged over the volume covered by the vector rods, is

$$L'\varrho^2 A/k \text{ e.}^2\text{Å}^{-3},$$

where L' is the length of the nominal unit cell in the X direction. Since this vector density will be independent of X we can calculate it from the two-dimensional end-on Patterson projection.

In practice the Patterson vectors between adjacent rods will overlap, but by comparing cases in which the overlapping effects are very similar we can avoid appreciable errors due to this.

For the more complicated case of a rod whose electron density is a function of Y and Z, but independent of X, the above expressions will still hold, as can be seen by projecting the electron density on to the X axis. Note that the contours of the vector density in cross-section (perpendicular to X) will have the same *shape*, though different absolute values, in the two cases for which formulae are given above. Thus the ratio of the values at corresponding points in these two cases will be the same as the ratio of the *average* values over the corresponding cross-sections. Thus the *ratio* of the above expressions—one referring to a model with rods infinitely long, the other to a model consisting of rods of limited length—may be used to obtain the ratio of the vector densities at the appropriate points.

Results

(*a*) *Height of vector peaks*

A section through the Fourier projection of the α-helix, placed in a hexagonal lattice 10·4 Å apart, and viewed end-on, is given in Fig. 1. Note that it is volcano shaped. Each unit of ordinate represents 2·25 e.Å^{-2} for 27 Å length of chain. The corresponding Patterson projection is given in Fig. 2. Each contour

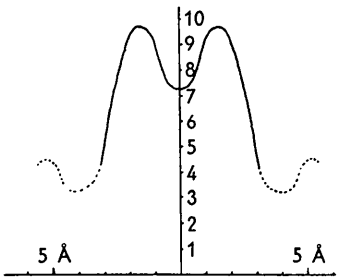

Fig. 1. A line section through the end-on projection of the α-helix, under limited resolution.

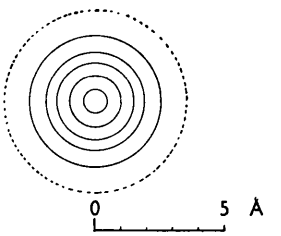

Fig. 2. The central part of the end-on Patterson projection of the α-helix, under limited resolution. Average contour broken.

represents an average vector density of 15 e.^{2}Å^{-3} for a unit cell 27 Å long. The chain and the Cβ atom have been given the co-ordinates listed by Pauling & Corey (1951) for the 18-residue 5-turn α-helix, and the remainder of the side chains have been put in uniformly in the area outside a circle of radius 4 Å, with sufficient electron density to make the overall average density 0·43 e.Å^{-3}. The 'heat-motion' assumed was that calculated for haemoglobin. It is rather stronger than the value estimated by Perutz, α being taken as 25 instead of 20, where α is defined by $\langle I\rangle = \langle I_0\rangle \exp[-\alpha\eta^2]$ and η is the reciprocal spacing in Å^{-1}. The limiting circle was taken as 2·8 Å, to correspond with the haemoglobin data. The calculated height of the peak of the central vector rod, averaged over the chosen strip between $X = 9$ and $X = 15$ Å, is given in Table 1 for several models. The last column includes a very rough allowance for vectors between chains with different X co-ordinates, assuming a plausible model for the molecule, while no such allowance is made in the preceding column. The observed height, obtained from the three-dimensional data of Perutz, and put on an absolute scale by measurements of Perutz, is 230 e.^{2}Å^{-3}.

Table 1. *Calculated Patterson peaks*

(Values in e.^{2}Å^{-3}.)

Model no.	No. of rods per molecule	Length of each rod (Å)	Fraction of polypeptide chain in the rods	Peak height; no overlapping	Peak height; perfect overlapping
1	13	50	0·75	3000	3400
2	13	40	0·60	2200	2200
3	26	25	0,75	2000	3300
4	26	16	0·48	625	790

(If allowance is made for the fact that some of the important intensities were rather underestimated, this can be corrected to 280 e.^{2}Å^{-3}.)

There is thus a discrepancy by a factor of 10 between the vector density to be expected for a simple model of straight parallel chains and the observed density. This density is of the same order as the discrepancy by a factor of 9 in $|F(063)|^2$ noted by Bragg, Howells & Perutz (1952), and means that the haemoglobin molecule cannot consist almost entirely of straight parallel α-helices.

Moreover, the first three models in Table 1 are unlikely because they give vector distributions of the wrong shape. The observed distribution along the X axis (see Fig. 3(*a*)) of Bragg, Kendrew & Perutz, 1950), when roughly smoothed, has a minimum between 10 and 15 Å and a maximum around 27 Å. This is compatible with models having short rods about 18 Å long and about 27 Å apart in the X direction. Models in which the molecule is split into two by a plane roughly perpendicular to the X direction are plausible on other grounds, as shown by Bragg, Howells & Perutz (1952) in their discussion of models with an odd number of layers. The discrepancy in vector density is also least for the model (no. 4 in Table 1) with short, separated, rods. Note, however, that this model has only about half the polypeptide chain in the rods.

(b) Shape of the vector rods

So far we have only compared the peak density of the vector rods and its variation along the X axis of the Patterson and have not yet considered the vector distribution at points near the X axis. Fig. 3

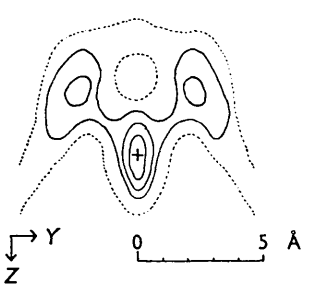

Fig. 3. Projection in the X direction of part of the central rods of the haemoglobin Patterson. X axis marked with a cross. Average contour broken, lower contours omitted.

is a projection in the X direction of the central vector rod in the three-dimensional Patterson calculated for the interval $X = 8{\cdot}8$–15 Å from the origin. Each contour represents 61 e.^{2}Å^{-3} of average vector density. It shows a central peak about 3 Å wide, flanked by two subsidiary ones which do not come from a rod-shaped structure but from a peak at $X = +8{\cdot}3$, $Y = +2{\cdot}7$, $Z = -3{\cdot}7$ Å together with its mirror image. By contrast, the corresponding Patterson peak of the α-helix (Fig. 2) has a width of 7 Å, or more than twice that observed in the three-dimensional Patterson. Since it could be argued that a fairer comparison would be between the total positive vectors near the axis, and not between the peak densities, this difference in shape increases rather than diminishes the remaining discrepancy for model no. 4. While too much weight should not be placed on the details of this shape, the fact that it is narrower than expected, rather than broader, has to be borne in mind when considering certain possible irregularities discussed below.

Possible irregularities

Under this heading will be discussed various factors which might reduce the density in the vector rods.

(a) Lack of alignment of the two halves of the molecule

I have found (Crick, unpublished) that a displacement of the two halves of about 2·0–2·5 Å in the c direction is consistent with the form of the rods in the $Y = 0$ plane of the three-dimensional Patterson. This displacement affects the X projection and has been considered independently by Bragg, Howells & Perutz (1952). It is one of the advantages of working in three dimensions, rather than two, that such a displacement would have little effect on the regions of vector space considered here. It does not affect the calculated values of vector density given in column 5 of Table 1 but it would reduce the figures given in column 6, as the vectors between chains with different X co-ordinates are now displaced from the X axis.

(b) Tilting of the rods

If the rods were straight and parallel, but were tilted about the Y axis, the effect would show clearly in the $Y = 0$ plane of the Patterson, which it does not; if about the c^* axis it would spread out the vectors in a plane perpendicular to c^*, which is not found to be so.

(c) Meandering of the rods

If the rods were broadly parallel, but meandered, vectors would be thrown away from the central vector rod and its peak density would be reduced. Fig. 4 shows the reduction in peak density due to a meander expressed as a root-mean-square deviation normal to the chain axis. This shows that to reduce the peak

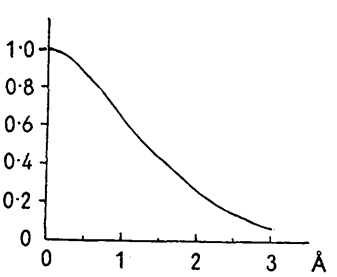

Fig. 4. Ordinate: reduction factor for peak vector density due to meandering of the chains. Abscissa: root mean square value of the meander.

height by a factor of 2, a r.m.s. deviation of $1\frac{1}{3}$ Å is required. This is quite a large variation over a length of 12 Å when it is considered that the r.m.s. radial deviation from the average direction evaluated for *all* the atoms in the polypeptide main chain must amount to $1\frac{1}{3}$ Å. It remains to be seen whether an α-helix can meander as much as this without breaking its hydrogen-bond sequence. In my opinion this is unlikely.

It should be noted, moreover, that a meander of the type considered would broaden the vector peak as well as flatten it. As noted above, there is little suggestion of this in Fig. 3.

(*d*) *Non-parallelism and kinking*

If the rods were straight, but only approximately parallel, vectors would again be thrown away from the central peak. If the angular deviations were small this should again lead to broadening of the peak; if large, the vectors might be thrown right outside, but one might expect packing difficulties unless such rods were on the surface of the molecule.

If the rods were kinked—that is, straight for short distances and then bent at an angle—similar remarks would apply.

(*e*) *Turning corners*

If the rods are short, there must be a large number of them making up the molecule. Since it is suspected from end-group assay that there are only six separate chains in horse haemoglobin (Porter & Sanger, 1948), plausible models are found to contain over 20 corners, where a 'corner' means a piece of polypeptide chain connecting the ends of two rods. This has been discussed independently by Bragg, Howells & Perutz (1952).

The number of corners is thus comparable with the number of rods. In Table 1 no allowance has been made for the reduction of peak vector density due to this complication. This is the biggest source of error in the calculations.

If the additional vectors were distributed largely at random, they would make no difference to the vector heights measured above average. It is more likely that the vectors between the rods and the corners will tend to fall into the space *between* the main vector rods. They might reduce the apparent vector density of the rods, but are unlikely to do so by a large factor. Without a definite model for the corners it is impossible to be more precise. If the peak height were reduced in this way, the same effect might explain the narrowness of the peak in Fig. 3, just as the apparent width of Mount Ararat was decreased by the Flood.

Other types of fold

All these calculations have been done for the α-helix, but the broad results (not the details about shape of peaks) are likely to apply to *any* regular fold of the polypeptide chain. They can be repeated for any particular case, though the pseudo-cylindrical symmetry of the α-helix makes this particular calculation peculiarly simple. Previous approximate calculations (unpublished) on the Astbury 2_{13} chain (see Bragg, Kendrew & Perutz, 1950) gave peak heights of the same order of magnitude as for the α-helix.

Asymmetry

There is one further general feature of the three-dimensional Patterson of horse haemoglobin which is puzzling, and that is the asymmetry both of Fig. 3 and of the 5 Å shell (see Perutz, 1949). The Patterson of an infinite α-helix has practically cylindrical symmetry, and one might have expected this to be more apparent in the haemoglobin Patterson, unless the lengths of α-helix were *very* short. While the asymmetry might be due to the individual arrangements of the different side-chains, their general arrangement is likely to be loosely based on hexagonal symmetry, and one would have surmised that deviations from this would tend to be random. The observed asymmetry suggests, rather, a more profound asymmetry in the arrangement of the folded polypeptide chains.

Conclusions and general discussion

We have found that:

(1) The absolute height of the vector rods cannot be reconciled with any model having more than about half the protein in straight rods parallel to the X axis.

(2) Better agreement in peak height can be obtained by postulating certain irregularities, but these cannot easily be reconciled with the shape of the vector peak.

(3) There is no evidence to support rods of 50 Å or so in length. The data suggest rods of, say, 18 Å in length, or less, and that the molecule is split into two halves, about 27 Å apart, probably by a plane roughly perpendicular to the X axis.

(4) None of the simple models considered here gives really good quantitative agreement with the height and shape of the vector peak.

What we have shown is that the haemoglobin structure cannot be as simple as, say, a synthetic polypeptide. This is in line with other evidence, such as the smallness of the infra-red dichroism (Elliott & Ambrose, 1950) and the diffuseness of the 1·5 Å reflection (Perutz, 1951*a*), both of which support the rod model qualitatively.

It is clear that somehow we must introduce into our simple initial picture sufficient irregularity to throw the vectors away from the central Patterson rod. It is possible that this can be done by making the chain irregular, but the irregularities must be large, probably involving the rearrangement of bonds, and not merely the straining of them. Alternatively we can make the chains kinked—that is, straight for short stretches and then bent at an angle. Other possibilities, such as the

rods being curved or kinked into a super helix, suggest themselves. In any case allowance must be made for the chains to turn corners.

The most important positive conclusion is that the length of the rods is probably rather small (15–20 Å). This suggests that globular proteins are really three-dimensional in their 'architecture', and not two-dimensional like the synthetic polypeptides. That is to say, whereas the broad structure of a synthetic polypeptide can be conveniently represented diagrammatically by a projection in the rod direction, a globular protein may be more like a three-dimensional framework, and may need a perspective drawing to show its main features. Whether this three-dimensional architecture conforms to a single general plan, or whether it is specific for each protein, or bit of a protein, remains to be seen.

Finally it must not be concluded that because of the poor quantitative agreement between observation and calculation we can immediately reject the rod model and α-helices. The method of calculation used is suitable for simple models, and becomes progressively more unreliable as we introduce complications. A more lengthy and accurate calculation would be needed to show whether the discrepancies in absolute height, projected shape, and asymmetry of the Patterson require a completely new model or whether they arise naturally out of simple modifications to the present model with short rods.

I should like to thank Dr M. F. Perutz both for allowing me to use unpublished material and for considerable advice on the presentation of the results. I should also like to thank Prof. Sir Lawrence Bragg, Dr W. Cochran and Dr J. C. Kendrew for helpful discussions.

References

Bragg, W. L., Howells, E. R. & Perutz, M. F. (1952). *Acta Cryst.* **5**, 136.
Bragg, W. L., Kendrew, J. C. & Perutz, M. F. (1950). *Proc. Roy. Soc.* A, **203**, 321.
Elliott, A. & Ambrose, E. J. (1950). *Disc. Faraday Soc.* **9**, 246.
Pauling, L. & Corey, R. B. (1951). *Proc. Nat. Acad. Sci. Wash.* **37**, 235.
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*a*). *Nature, Lond.* **167**, 1053.
Perutz, M. F. (1951*b*). *Nature, Lond.* **168**, 653.
Porter, R. R. & Sanger, F. (1948). *Biochem. J.* **42**, 287.
Wrinch, D. (1950). *Acta Cryst.* **3**, 475.

Acta Cryst. (1952). **5**, 386

Determination of Elastic Constants of Lithium Fluoride from Photographs of Diffuse Reflexions of X-rays

By J. Hoerni and W. A. Wooster

Crystallographic Laboratory, Cavendish Laboratory, Cambridge, England

(*Received* 27 *October* 1951)

A photographic method of determining elastic constants from the study of diffuse X-ray reflexions has been developed and applied to lithium fluoride. The results are as follows: $c_{11} = 9{\cdot}9 \times 10^{11}$, $c_{12} = 4{\cdot}3 \times 10^{11}$, $c_{44} = 5{\cdot}4 \times 10^{11}$ dyne cm^{-2}.

Thermal vibrations of the atoms in a crystal give rise in reciprocal space to scattering from regions other than the points of the reciprocal lattice. The theoretical distribution of this extra scattering density has been given by Jahn (1942) for cubic crystals, and it depends on the elastic constants of the crystal. Lonsdale & Smith (1941, 1942) have shown that there was qualitative agreement with this theory when the diffuse X-rays were photographically recorded, and Ramachandran & Wooster (1951*a*, *b*), using a Geiger counter, have obtained the elastic constants for several cubic crystals.

We have used the photographic method to study quantitatively the diffuse reflexions obtained with LiF. A camera of radius 10 cm. and monochromatized Cu $K\alpha$ radiation have been used, and the film blackening was measured with a Dobson microphotometer. Charts have been constructed to determine conveniently which point in reciprocal space corresponds to a given point of the diffuse spot (Hoerni & Wooster, in press). Although measurements at only three different points of the diffuse spot are in principle required to solve Jahn's equation for the elastic constants, many more observations have been made in order to check the results. The measurement of relative intensities within any one spot permits