A REVIEW OF SOME RECENT X-RAY WORK ON PROTEIN CRYSTALS

DOROTHY CROWFOOT

Department of Mineralogy, Oxford University, Oxford, England

Received September 14, 1940

This is a difficult moment at which to try to give an account of protein structure from the x-ray side. A great deal of work is in progress, and it seems likely that many points about which we now conjecture will be settled soon if only the work can go on.

Three crystalline proteins have been studied in some detail by crystallographic methods,—insulin, horse methemoglobin, and lactoglobulin,—and this report will be mainly concerned with their investigation. The measurements on hemoglobin[1] have been carried out by Dr. M. F. Perutz, working at Cambridge University under Professor W. L. Bragg with a grant from the Rockefeller foundation. As Dr. Perutz is now interned[2] and may be unable to send an account of his work himself, Professor Bragg has kindly given me permission to use his results. The work on insulin and lactoglobulin has been carried out at Oxford University.

These three proteins have the advantage that they have all been x-ray photographed in both the wet and the dry states, and lactoglobulin in particular in several crystalline modifications (table 1). Further, a number of measurements have been made of the intensities of x-ray reflections from these crystal structures, and several Patterson Fourier projections have been calculated for each protein (figures 1 to 7). Unfortunately only in the case of air-dried insulin are these calculations anything like complete.

The first question that it seems relevant to ask about these crystal structures is whether or not there actually exist in the crystal separate individual molecules of the magnitude deduced by Svedberg and his coworkers from ultracentrifuge measurements. In x-ray analysis no absolute proof of the presence of a molecule is possible without a complete analysis which will demonstrate the degree of chemical binding in different directions in the crystal structure from the lengths of the interatomic

[1] The Patterson projections for wet hemoglobin were calculated by Mr. D. P. Riley from data supplied by Dr. Perutz.

[2] Since released (note added February, 1941).

TABLE 1

X-ray measurements on the proteins insulin, hemoglobin, and lactoglobulin

PROTEIN		*a*	*b*	*c*	*β*	CELL VOLUME	DENSITY	SPACE GROUP	MOLECULES PER CELL	BIREFRINGENCE
Insulin (hexagonal axes)	Wet	83.0		34.0		203,000	1.28	$R3$	3	+
	Dry	74.8		30.9		150,000	1.31	$R3$	3	+
Insulin (rhombohedral axes)	Wet	49.4			$\alpha = 114°16'$	68,000			1	
	Dry	44.4			$\alpha = 114°28'$	50,000			1	
Horse methemoglobin	Wet	110	63.8	54.2	112°	352,000	1.242	$C2$	2	− (?)
	Dry	102	51	47	130°	188,000	1.270	$C2$	2	
Lactoglobulin, tabular, orthorhombic	Wet	67.5	67.5	154	90°	702,000	1.257	$P2_12_12_1$	8	−
	Partly wet	67.5	67.5	148.5	90°	677,000		$P2_12_12_1$	8	−
	Dry	60	63	110	90°	416,000	1.27	$P2_12_12_1$	8	
Lactoglobulin, needle, tetragonal	Wet	67.5	67.5	133.5	90°	608,000		$P4_22_1$	8	+
	Dry	56	56	(130)	90°	408,000	(1.3)		8	

distances found. But the comparison of Patterson projections for wet and air-dried insulin (3) do provide strong evidence in favor of the Svedberg molecule in the crystal. In each of these projections there is a group of eighteen peaks around the origin, having the same positions relative to one another in both projections. The difference between the two projections is mainly due to a difference in the orientation of the adjacent groups of peaks. Figure 1 shows the effect. The whole group at the origin A is turned relative to the group at B through an angle of about 6° in the wet as compared with the dry structure. The group of eighteen peaks is there-

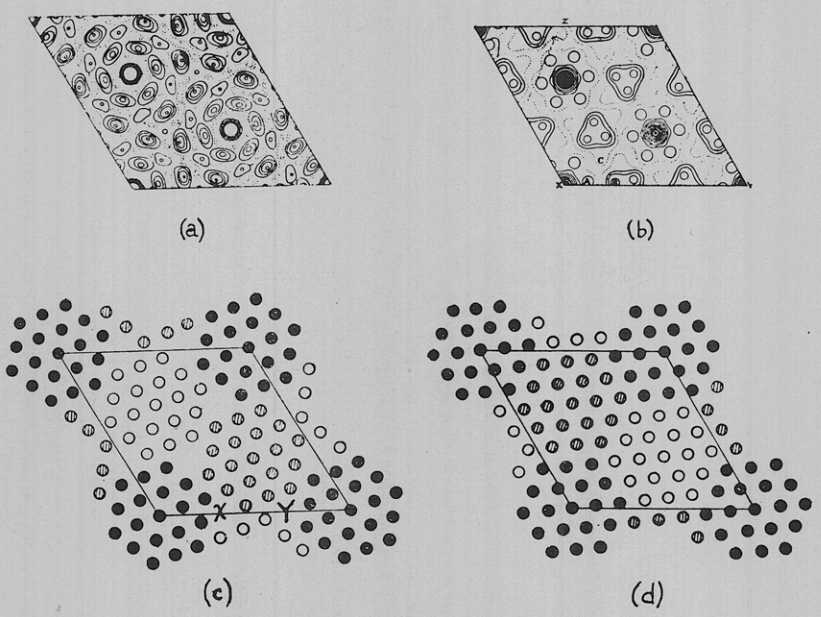


FIG. 1. Patterson projections on the basal plane of the structure of (a) wet and (b) dry insulin. Arrangement of hexagonal arrays of points: (c) expanded (*cf.* wet insulin) and (d) close-packed (*cf.* dry insulin).

fore interdependent, and the peaks move as if they corresponded to interatomic distances organized together in a single unit. The simplest interpretation of this organized unit makes it of the same size as the crystal unit cell, which in turn is equal in weight to one Svedberg molecule. One can derive somewhat similar evidence, though not so coherent, from a comparison of two of the lactoglobulin crystal structures (4). Here the crystallography is much more complicated. In the wet tabular crystals the molecular weight of the unit cell indicates the presence of eight Svedberg units plus a large proportion of liquid of crystallization. The crystals

shrink primarily in the direction of the *c*-axis, first by only 3.6 per cent to a partly wet form, and then by 41 per cent to the air-dried disordered structure. Comparison Patterson projections of the wet and partly wet structures show patterns which differ considerably more in the peak formations corresponding to large interatomic distances than in the arrangement of peaks about the origin. Such a distinction is in accordance

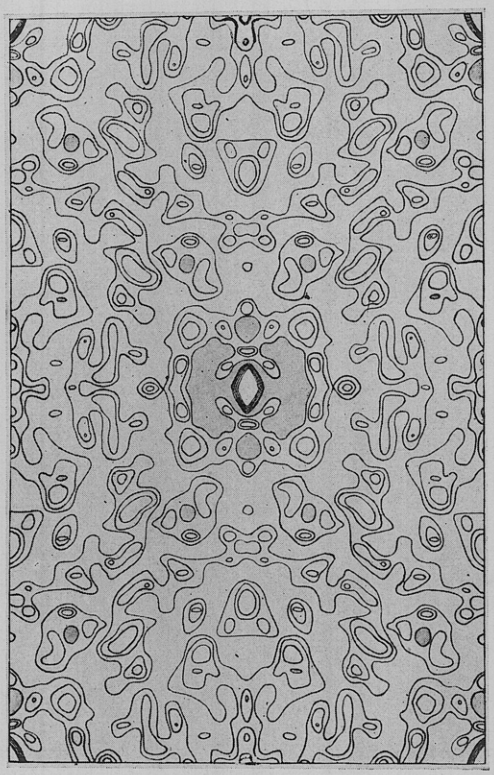


FIG. 2. *Pxy* derived from wet crystals of horse methemoglobin (after M. Perutz and D. P. Riley).

with the presence of molecules which move closer together during the shrinking process, maintaining their internal structure relatively unchanged.

If we take the existence of protein molecules in the crystal as to some extent established, the next problem we have to consider is that of their shape. This question has recently received considerable attention from a number of workers. The most interesting calculations are probably those of Polson and Neurath, who have calculated the dissymmetry of

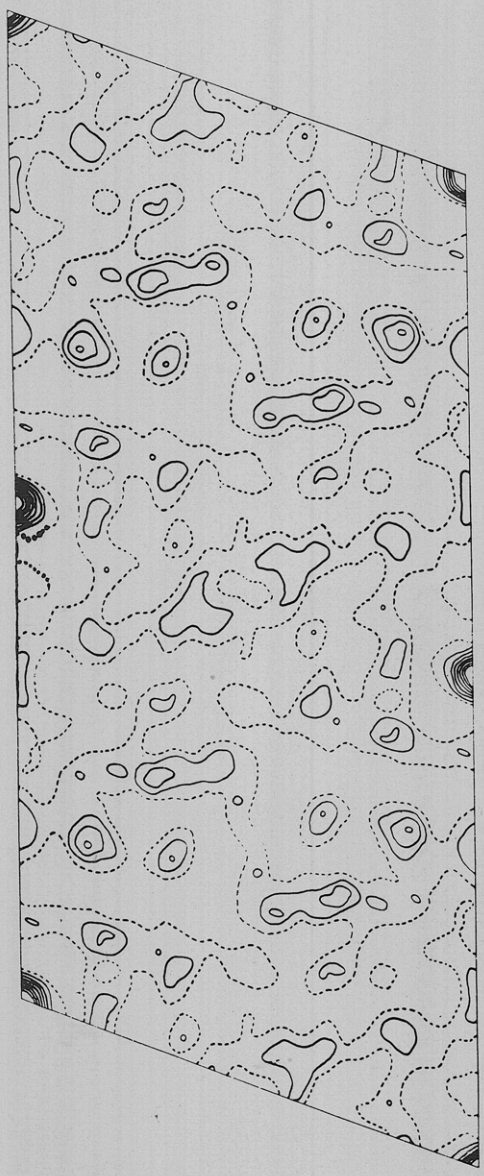


FIG. 3

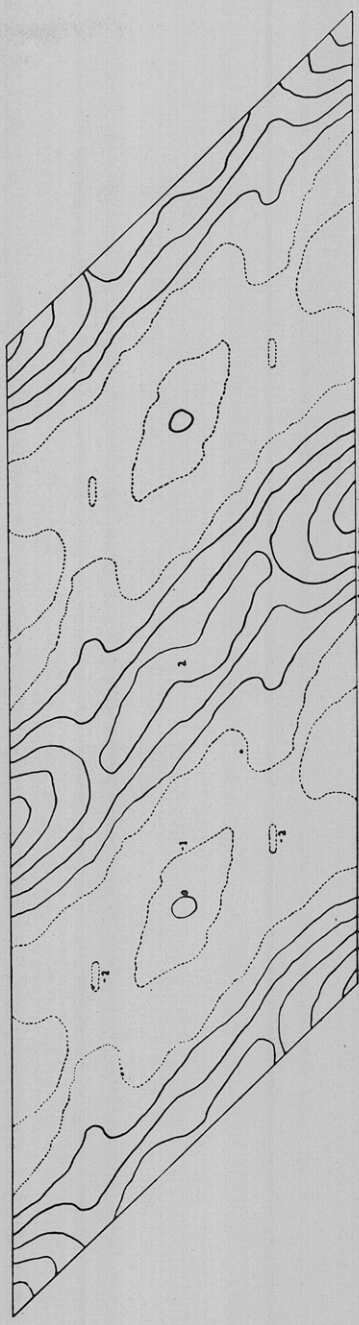
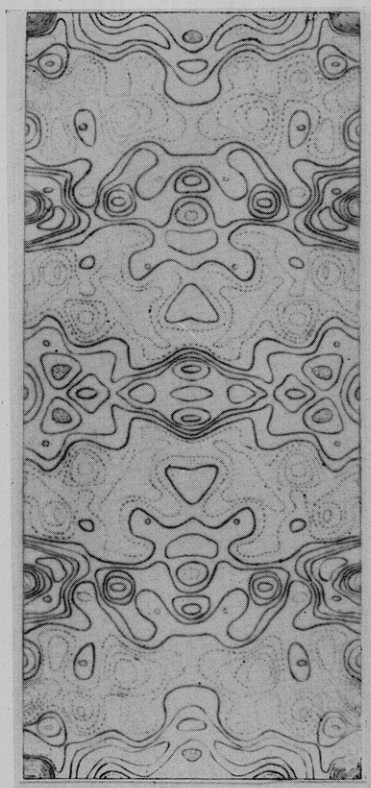
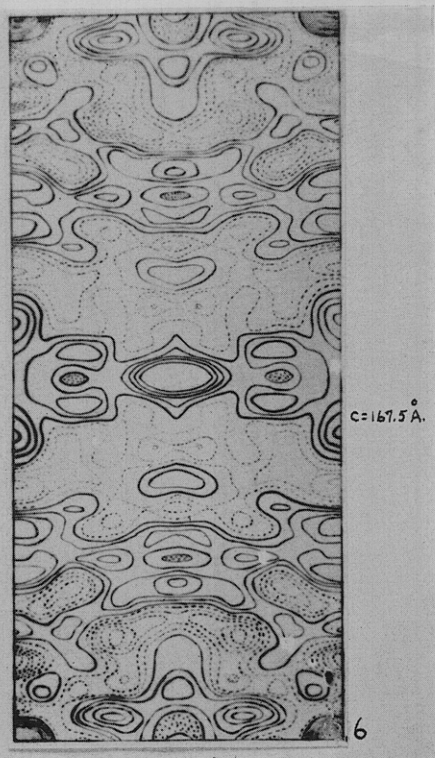


FIG. 4

FIG. 3. Horse methemoglobin: Patterson projection *Pxz* from wet crystals (after M. Perutz and D. P. Riley).

FIG. 4. Horse methemoglobin: *Pxz* derived from air-dried crystals; contour lines at 25 units apart (after M. Perutz).





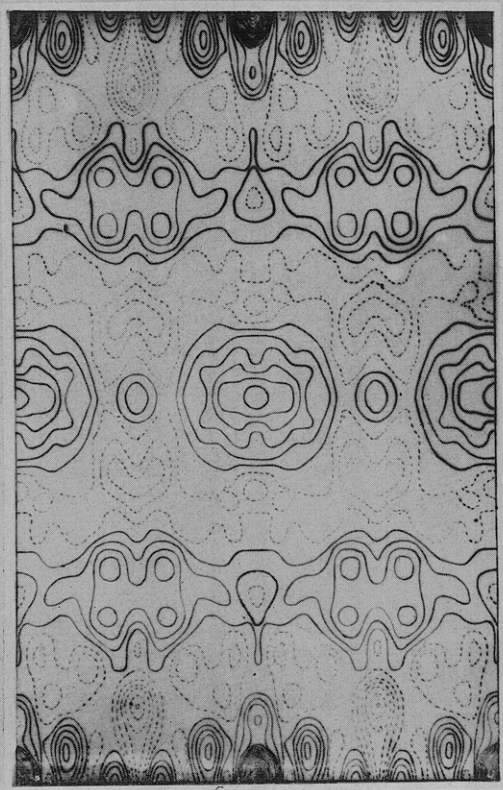


FIG. 5. Lactoglobulin: tabular, wet; Patterson projection on (100)
FIG. 6. Lactoglobulin: tabular, partly wet; Patterson projection on (100)
FIG. 7. Lactoglobulin: tabular, wet; Patterson projection on (110)

protein molecules from the Svedberg dissymmetry constant, f/f_0, obtained from diffusion constant measurements. Neurath (5), in particular, uses the dissymetry constant to deduce possible dimensions of the protein molecules concerned, assuming these to be prolate ellipsoids of revolution. His values for insulin, lactoglobulin, and hemoglobin are as shown in table 2.

In these deductions it is assumed that the other factors which may influence the dissymmetry constant, the charge on the protein molecule and hydration, are negligible compared with the dissymmetry factor. Perutz has recently examined the other extreme hypothesis (which was first considered by Adair and Adair (1)),—namely, that the Svedberg dissymmetry constant is mainly a measure of the hydration of the protein mole-

TABLE 2

Dimensions of protein molecules

PROTEIN	f/f_0	b/a	a	b
Insulin	1.13	3.3	31	102
Lactoglobulin	1.2	4.3	28	122
Hemoglobin (horse)	1.24	4.8	32	155

TABLE 3

Radii of proteins

PROTEIN	r_1 (DRY "RADIUS")	r_2 (WET "RADIUS")	r_3 (WET "RADIUS", CALCULATED FROM VOLUME OF UNIT CELL)
Insulin	22.4	25.8	25.3
Lactoglobulin	23.2	29.2	28.8
Hemoglobin	27.2	29.9 33.7 34.6	34.8

cule in solution. To conform to the observations, the protein molecules must increase in effective size and have a lower density on hydration. Perutz (6) therefore calculates first the radii of the dry proteins from their measured molecular weights and partial molar volumes, assuming they are spherical, and then the radii of the wet proteins required to agree with the diffusion constant measurements (see table 3). He shows that a considerable degree of asymmetry,—as much as 1.4 or 0.71,—may be present without appreciably affecting the dissymmetry constant, and that therefore these radii are not to be taken as an exact measure of the shape of the molecules concerned. The third column of table 3 gives r_3, the radius of a hypothetical molecule, having the volume of the crystallographic unit cell found for the wet crystals, divided by the number of

molecules present. It is clear that the calculations show that the deviation of the dissymmetry constant from unity *could* be explained by hydration of the protein molecules in solution roughly equivalent to their hydration in the crystals.[3]

As both authors agree, in reality it is probable that both hydration and shape play a part in determining the observed dissymmetry constant. It should be possible to judge their relative importance by an examination of the x-ray evidence in more detail.

To take the case of insulin first. The unit cell of the air-dried crystals is a flat rhombohedron of very marked asymmetry. As this cell is of

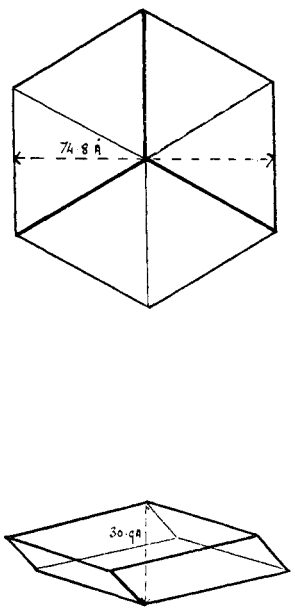


FIG. 8. The unit cell of air-dried insulin

weight equivalent to one Svedberg molecule, it might itself represent the shape of the molecule. But it is not possible actually to accommodate in it a molecule of the shape calculated by Neurath. If considered as approaching an ellipsoid of revolution, this ellipsoid would in the first case have to be oblate, not prolate, and its dimensions might be given roughly, to compare with Neurath's calculations, as follows: height, 30 Å.; diameter, 74.8 Å. (see figure 8). Actually, it is probable from the crys-

[3] In this connection it is interesting that the direct calculation of the molecular weight of pepsin from diffusion constant measurements gives a value twice as large as that given in the ultracentrifuge, while the crystals contain roughly 50 per cent of liquid of crystallization.

tallographic placing of adjacent molecules that the molecule would not have so extreme a form. The most compact alternative is represented by a prism of hexagonal base 43 Å. across and height 30 Å., and the truth probably lies between these limits. The study of the Patterson projection (see below) suggests that the close packing of the molecules in the crystal is determined not simply by their approximate shape, but to a large degree by their internal structure.

The x-ray results therefore support Neurath's deduction,—namely, that the insulin molecule is markedly asymmetric,—while excluding such a high degree of asymmetry as that proposed. The additional deviation of the dissymmetry constant is probably due to hydration, as Perutz suggests. The comparison of the Patterson projections from wet and from dry insulin indicates that the molecules have moved from loose packing in the wet crystals to close packing in the dry crystals. The patterns suggest that in the wet crystals there are gaps at X, Y, etc. (figure 1) between the molecules which are filled with liquid of crystallization. It would be unlikely that this particular liquid would be carried with the molecules in solution. But the shrinking of the crystals on drying is greater than corresponds theoretically to the calculated shift in position of the molecules. The excess shrinking is due to hydration that might persist in solution and consequently further affect the diffusion constant.

The evidence from lactoglobulin and hemoglobin is less precise than that from insulin, owing to crystallographic complications, but it tends in the same direction. The Patterson projections from wet tabular lactoglobulin indicate that the molecules are packed in a pseudo-face-centered cubic array. Such an arrangement suggests that the molecules are not far from spherical, though a dissymmetry of the axes as great as 2:1 is possible,—again much smaller than that calculated by Neurath. It is also improbable that the hydration of the molecules in solution is quite as great as that of the wet tabular crystals, since these very easily shrink to the partly wet form, closely related in structure, and, further, the metastable needle crystals also have a lower hydration.

The crystal structure of hemoglobin presents some interesting problems. If the chemical molecular weight of 66,700 is accepted, the positions of the molecules are fixed in the lattice on twofold axes at the corners of the unit cell and C face centers. Perutz calculates that the most probable form of the molecule is a triaxial ellipsoid with $x = 22$, $y = 24$, and $z = 37.6$ Å. Setting $x = y = 23$ Å., the axial ratio is 1.63,—very much smaller than that found by Neurath. In fact, molecules of Neurath's calculated dimensions are, as in the case of insulin, excluded by the dimensions of the unit cell. As before, the hydration of the crystals is high, and probably a considerable proportion persists in solution.

Perutz proceeds to consider the methods of distributing the water of hydration. It is clear, as mentioned above, that this water must both increase the size and reduce the density of the molecules,—i.e., it does not simply enter into a central cavity of the molecule, as might be possible with molecules of the cyclol structure put forward by Dr. Wrinch. It may either be distributed over the surface of the molecule, or be disposed between structural units within the molecule in such a way that it causes an internal expansion. Or, of course, both these processes may occur. Perutz is rather in favor of the second hypothesis, to explain the fact that the introduction of heavy atoms into hemoglobin crystals through the liquid of crystallization produces no noticeable alteration in the intensities of x-ray reflections from the crystal. For insulin it should be possible to distinguish between the two alternatives through the calculated wet and dry Patterson projections on (0001). In the first case one would expect the characteristic peak system of dry insulin to be distorted in the wet crystal; in the second, the distances of the peaks from the origin should increase. Actually, for certain theoretical structures the two effects may lead to the same results and both changes appear to be present. Also, it is doubtful whether the accuracy of the present wet projection is sufficient for it to be employed in this argument, and better data are in preparation. One point is worth making,—the smallness of the change involved in atomic terms. The calculated change, if it is intramolecular in the projected distance of the inner insulin peak, is only from 10 to 10.5 Å. while, if the water of hydration is disposed over the surface of the protein molecule, it would correspond in quantity to a layer of water only one molecule thick.

The problem of hydration has an important bearing on that of the internal structure of the protein molecule, which is the third (and much the most speculative) point that I wish to discuss. The evidence to be considered here is mainly that provided by the Patterson projections of the different proteins.

These projections should first be distinguished according to resolving power. The limits of spacing used in their construction are roughly as follows:

	Å.
Wet hemoglobin	2.4
Wet insulin	4.5
Wet tabular lactoglobulin (100)	5.5
Partly wet tabular lactoglobulin (100)	6.5
Dry insulin	7.5
Wet tabular lactoglobulin (110)	10.5
Dry hemoglobin	13

As would be expected, the patterns for lactoglobulin (110) and for dry hemoglobin show broad, largely unresolved groups of maxima which prob-

ably indicate the main positions of the molecules, but little more. That from wet hemoglobin at the other end of the scale is the only projection which shows peaks as close as 4.5 Å. to the origin. The peaks are of very great interest, but it is impossible to be sure that the actual interatomic vectors concerned are 4.5 Å. until the third projection for this structure has been calculated. Apart from these features it is the similarities between the different projections that are most notable. Those of wet hemoglobin and insulin in particular show a general distribution of low maxima roughly in a series of rings about the origin with, very roughly, intervals of 10, 20, 30 Å., etc. In the lactoglobulin (100) projection the peak systems are only clearly resolved near the origin, where they show marked resemblances to the others.

A striking feature of the (0001) projection of air-dried insulin was pointed out by J. D. Bernal (2). In position all the observed peaks fall on a hexagonal network, the axes of which lie at an angle to the crystallographic axes. The angle observed is closely that required if the insulin molecule itself has a structure in which the eighteen points of the network around the origin are occupied by units which are arranged in a close-packed array, not only within one molecule but also with reference to the unit structure of neighboring molecules (figure 1). The change from dry to wet insulin then appears to involve an angular shift of the molecules from these close-packed positions. Further, the new peak positions in the wet (0001) projection are not far from a second hexagonal network, which might again bring the unit points into close contact.

It is interesting that a similar effect can be, very roughly, traced in the wet hemoglobin projection on (010). Nearly all the peaks present lie close to points on a network formed by intersection of the planes (006) and $(\bar{1}201)$ (figure 9). Bernal's procedure would lead one to deduce that here again the molecular pattern was roughly hexagonal in projection and formed by thirty-six points about the origin,—significantly double the number in insulin. Counting the origin peaks, the structures in insulin and hemoglobin would consist of nineteen and thirty-seven subgroups in projection, respectively.

Whatever is the interpretation of these effects, it is certain that the $(\bar{1}201)$ plane is important in the hemoglobin crystal structure. The two greater refractive indices of the crystal lie parallel to this plane, and also probably, from the examination of the absorption spectra (6), the heme groups of hemoglobin.

There are considerable difficulties in seeking an explanation of these Patterson projections, particularly in the more complicated structures such as lactoglobulin and hemoglobin. In these cases every maximum in the Patterson projection may correspond either to an intermolecular or

to an intramolecular vector and probably to both. The case of insulin is simpler, since the individual maxima must represent intramolecular vectors, the only doubt being with which molecule and with how many molecules in the lattice they are associated. It seems possible at first sight to make the distinction by comparing the wet and dry (0001) patterns from insulin. Here one would naturally associate the vectors that appear to move together with intramolecular vectors within one molecule, and this assumption has, in fact, underlain much of the previous discussion. It is clear that the peak movement is such that it tends to bring the peaks associated by Dr. Wrinch as due to vectors within one molecule out of coordination with one another. Other assumptions than simple movement would therefore have to be introduced to relate the Wrinch structure to

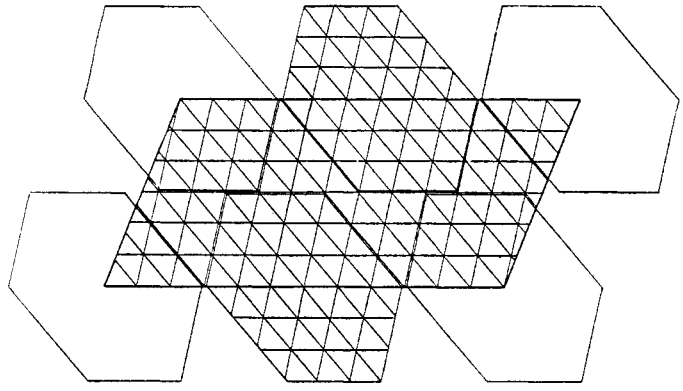


FIG. 9. Hemoglobin. Idealized picture of the xz projection, showing the network of points with the outlines of six intramolecular patterns. The distances apart of the intersections of the network are 8.6, 9.1, and 11 Å. The two central hexagons are at $z = \frac{1}{2}$; the others at $z = 1$.

the insulin patterns. It is reasonable to compare with this conclusion the effects to be expected with the model proposed by Bernal for insulin. Bernal has suggested that the molecules can be considered as consisting of a number of subgroups arranged in cubic close-packing. In projection the point-group associated with this structure is shown in figure 10, together with that on which the Wrinch interpretation is based, for comparison. The Patterson projection derived from the Bernal arrangement[4] is one in which all the peaks are complex and include three or four intramolecular vectors which coincide in the pattern from air-dried insulin,

[4] Examination of figure 10 will show that the two groups at 3 and 3′ in the Bernal structure are not at the same z level. The criticism of this theory by Wrinch and Langmuir (Proc. Phys. Soc. (London) **51**, 617 (1939)) is therefore not valid.

owing to the close-packed arrangement described above. The short vectors within one molecule contribute most to the peak intensity. On turning the molecule out of the close-packed positions one would expect the movement to be reflected in the peak positions, but these should also be considerably blurred if the movement is not exactly into the next close contact positions. Blurring certainly appears to be present, but not quite in the direction expected from simple theory. It is clear that in both cases it may be argued that deductions from the point-groups alone are oversimplified and that complete agreement will depend on the molecular model adopted.

The two theories are derived from quite different molecular models. According to the Wrinch theory, the insulin molecule is a single polyhedron; in the Bernal theory it consists of subgroups of twelve or twenty-

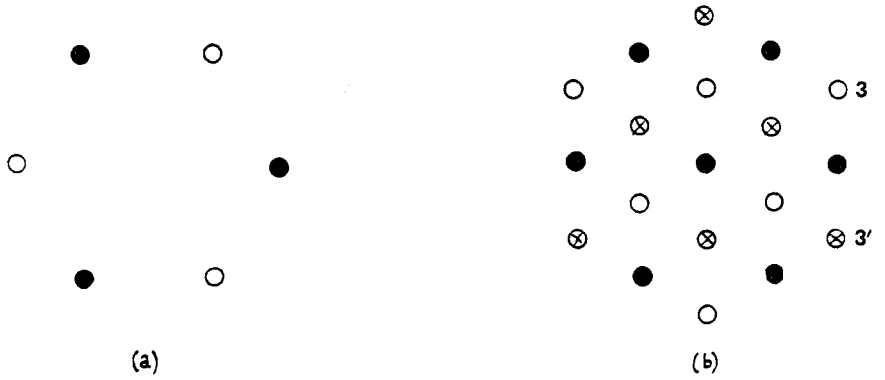


FIG. 10. Insulin: (a) Wrinch point group; (b) Bernal point group. The points at different heights are indicated by the variation in shading.

four amino-acid residues which are arranged in cubic close-packing, and probably linked through R groups such as cystine or glutamic acid. The amino-acid residues in each subgroup are linked in a chain and can be ordered in several ways, e.g., to cover the surface of a cube. Other polyhedra midway in size between these two have been devised by Talmud (7), and certainly there are quite different possibilities. It is clear from the existing controversy that no theory so far put forward is completely convincing; and in conclusion I should like to emphasize two points in the insulin data that seem important.

First, as an experimentalist, I should prefer to associate angular shift of the Patterson peaks for wet and dry insulin with that of the molecule.

Secondly, I am strongly of the opinion that the peak arrangement in air-dried insulin and its relation to close-packing structures is no accident,

but is an expression of some characteristic of the actual protein structure, whether this has the form discussed by Bernal or one completely different.

REFERENCES

(1) ADAIR, G. S., AND ADAIR, M. E.: Proc. Roy. Soc. (London) **B120,** 422 (1936).
(2) BERNAL, J. D.: Proc. Roy. Soc. (London) **A170,** 75 (1939).
(3) CROWFOOT, D., AND RILEY, D. P.: Nature **144,** 1011 (1939).
(4) CROWFOOT, D., AND RILEY, D. P.: Unpublished work.
(5) NEURATH, H.: J. Am. Chem. Soc. **61,** 1841 (1939).
(6) PERUTZ, M. F.: Ph.D. Thesis, University of Cambridge, 1939.
(7) TALMUD, D. L.: Compt. rend. U.R.S.S. **25,** 484 (1939).