uranium is to join together two orthouranate layers or two orthouranate chains by means of shared O_I atoms. If this suggestion proves to be correct, the di-orthouranates should be designated as $R_2U_2O_3O_4$, corresponding to the presence of $1\frac{1}{2}$ O_I and 2 O_{II} atoms per uranium. According to this guess as to the constitution of the di-orthouranates, each uranium forms two primary U–O_I bonds, but there are no uranyl groups (UO_2). Instead, two uranyl groups are joined together to form configurations O_I–U–O_I–U–O_I with the infinite layers or chains normal to this axis.

Anhydrous UO_3 is known to exist in several crystalline forms, but structural information has been reported for only one of them. The hexagonal form of UO_3 is said (Zachariasen, 1948*b*) to have a very simple structure with one molecule per unit cell with $a_1 = 3{\cdot}971$ Å, $a_3 = 4{\cdot}168$ Å. Positions were assigned to the oxygen atoms purely on the basis of steric considerations as follows:

$$1\ \mathrm{U\ in}\ (0,0,0),\quad 1\ O_I\ \mathrm{in}\ (0,0,\tfrac{1}{2}),\quad 2\ O_{II}\ \mathrm{in}\ (\tfrac{1}{3},\tfrac{2}{3},z)$$

with $z \approx 0{\cdot}17$. It is readily seen that the proposed structure can be described as an infinite stack of the hexagonal layers first observed in the $CaUO_2O_2$ structure. These layers are stacked directly on top of one another in such a way that the O_I atoms are shared. The formula, accordingly, should be written UOO_2. There are no uranyl groups; instead uranyl groups are joined together to form an endless linear chain –O_I–U–O_I–U–O_I–.

References

Asprey, L. B., Ellinger, F. & Zachariasen, W. H. (1954). *J. Amer. Chem. Soc.* In press.
Asprey, L. B., Penneman, R. & Staritzski, E. (1954). *J. phys. Chem.* **58**, 403.
Ellinger, F. (1953). Private communication.
Ellinger, F. & Zachariasen, W. H. (1954). *J. phys. Chem.* **58**, 405.
Fankuchen, I. (1935). *Z. Kristallogr.* **91**, 473.
Hoard, J. L. & Stroupe, J. D. (1949). National Nuclear Energy Series, Division III, vol. 2, chap. 2, p. 15.
Samson, S. & Sillén, L. G. (1947). *Ark. Kemi Min. Geol.* **25**, No. 21.
Zachariasen, W. H. (1945). Manhattan Project Report CP-2611, p. 14.
Zachariasen, W. H. (1948*a*). *Acta Cryst.* **1**, 281.
Zachariasen, W. H. (1948*b*). *Acta Cryst.* **1**, 265.
Zachariasen, W. H. (1949). *Acta Cryst.* **2**, 388.
Zachariasen, W. H. (1954*a*). *Acta Cryst.* **7**, 783.
Zachariasen, W. H. (1954*b*). *Acta Cryst.* **7**, 788.

Acta Cryst. (1954). **7**, 799

A Test of the Usefulness of Direct Mathematical Methods in the Structure Analysis of a Protein

By M. F. Perutz and V. Scatturin

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

(*Received* 20 *August* 1954)

A test was made to see how large a fraction of the $h0l$ reflexions of horse methaemoglobin obeyed sign relations of the type: $S(h0l) = S(h'0l')S(h+h',0,l+l')$. Thirty sign relations were found between the 16 strongest reflexions chosen for the test; 22 of these held and 8 failed. A further 56 sign relations were found between any two strong and any one medium reflexion; 30 of these held and 26 failed. The results suggest that direct mathematical methods are not likely to give correct results in the structure analysis of proteins, except perhaps in very favourable cases.

There has been some debate whether direct mathematical methods, such as inequalities (Harker & Kasper, 1948) or sign relations (Sayre, 1952; Cochran, 1952; Zachariasen, 1952), could be usefully applied to the structure analysis of proteins. The recent determination of the signs of the inner 100 $h0l$ reflexions of horse methaemoglobin by the isomorphous replacement method (Green, Ingram & Perutz, 1954) provides an opportunity for testing the validity of such procedures. The unitary structure factors of haemoglobin are too small to apply inequalities. However, Cochran (1952) has shown that this is not necessary, because whenever inequalities give a relationship between three large structure factors it is of the form

$$S(hkl) = S(h'k'l')S(h+h',k+k',l+l')\ . \qquad (1)$$

The first application of this type of sign relation to proteins is due to Kendrew (1952). He treated a particularly favourable case: a projection of myoglobin in which the polypeptide chains are probably viewed end-on, simulating in appearance a structure with only a few well-resolved atoms. Such simple projections, however, are rare. It was thought that a general measure of the usefulness of direct mathematical methods, for projections showing no specially favourable features, might be obtained by choosing the strongest among the reflexions of known sign in horse haemoglobin and examining the extent to which equation (1) is obeyed between them. If the signs of the strongest reflexions from a protein crystal could be derived by the use of sign relations, then a Fourier projection could be calculated with these reflexions as terms. Such a projection might conceivably show some of the major features of the correct structure, and these features might lead to the determination of further signs. On the other hand, if the number of sign relations obeyed proved to be insufficient, then no attempt to solve the structure by direct mathematical methods is likely to be successful.

In the present work we have taken the 16 strongest among the *h0l* reflexions of known sign (Table 1).

Table 1

hkl	Sign	*F*	$U_1[F(000)=16{,}450]$	$U_2[F(000)=72{,}000]$
001	+	2650	0·162	0·037
002	−	1190	0·072	0·017
006	+	1280	0·078	0·018
205	+	1060	0·065	0·015
$20\bar{2}$	−	1070	0·065	0·015
402	−	1220	0·074	0·017
400	+	1020	0·062	0·014
$60\bar{2}$	−	1930	0·0117	0·027
$60\bar{7}$	+	1030	0·062	0·014
10,0,1	−	1020	0·062	0·014
10,0,0	−	1190	0·072	0·015
$12,0,\bar{1}$	+	1400	0·085	0·019
$12,0,\bar{4}$	+	1130	0·069	0·016
$12,0,\bar{7}$	+	1120	0·069	0·016
$14,0,\bar{3}$	−	1440	0·088	0·020
$16,0,\bar{7}$	−	1090	0·066	0·015

Their unitary structure factors were calculated on the basis of two alternative assumptions. For U_1 the value of $F(000)$ was taken as the difference between the number of electrons contained in two haemoglobin molecules and in an equivalent volume of water (for the justification of this procedure see Bragg & Perutz, 1954). For U_2 the value of $F(000)$ was taken as the total number of electrons contained in two haemoglobin molecules, but omitting the electrons in the liquid of crystallization which may be regarded as uniform at the resolution considered here.

$\sigma' = (\overline{F^2})^{\frac{1}{2}} = 600$ for the 100 reflexions of known sign. For the 16 strongest reflexions $(\overline{F^2})^{\frac{1}{2}} = 1360 = 2{\cdot}3\sigma'$. Their values range from $1{\cdot}7$ to $4{\cdot}4\sigma'$. Altogether, 30 sign relations were found to apply between the 16 reflexions; 22 of these agreed with the experimentally determined signs, and 8 failed. Thus the fraction which held was 75%.

We now took any two reflexions *h0l* and *h′0l′*, belonging to the original set of 16, and chose any reflexion of medium intensity for the third term $h+h',0,l+l'$. In this way a further 56 sign relations were found of which 30 (or 54%) proved consistent with the experimentally determined signs.

It is surprising, in view of the complexity of the structure, to find as many as 75% of the sign relations holding, even among the strongest reflexions. This result can hardly be fortuitous, as the following argument shows. Suppose the chances of any sign relation holding are even. If the total number of sign relations is n, and the number found to hold good is r, then the probability of this happening is

$$p = \frac{n!}{r!\,(n-r)!}\left(\frac{1}{2}\right)^n = \frac{1}{157} \quad \text{for} \quad n = 30 \quad \text{and} \quad r = 22.$$

The probability of there being exactly equal numbers of consistent and inconsistent relations is $\frac{1}{7}$. Thus the consistency of 22 out of 30 in haemoglobin must be considered significant.

Cochran & Douglas (1954) have devised a method of solving crystal structures directly with the help of such sign relations, using an electronic calculator to work out the correct combination of signs. They find that more than 90% of the sign relations among the strong and medium reflexions have to hold before their method becomes practicable, since otherwise the calculator works out a very large number of equally probable sign combinations. If this were done in a protein there would be few safe criteria for choosing the right one among them. Thus even if the fraction of sign relations holding among the strongest reflexions in haemoglobin is large, considering the complexity, it is nevertheless too small to offer any prospects that direct mathematical methods might be useful in the solution of its structure. Moreover, the fraction of sign relations holding between two strong and one medium reflexion is hardly greater than one-half. This failure does not necessarily imply that sign relations might not be useful in special cases, such as Kendrew's myoglobin projection mentioned earlier, but before any reliance can be placed on the results of a direct approach it would be necessary to establish the required simplicity of the projection by independent methods, such as three-dimensional Patterson analysis.

One of us (V. S.) is indebted to the Consiglio Nazionale delle Ricerche of Italy for a grant enabling him to work at Cambridge.

References

Bragg, W. L. & Perutz, M. F. (1954). *Proc. Roy. Soc.* A, **225**, 315.

Cochran, W. (1952). *Acta Cryst.* **5**, 65.

Cochran, W. & Douglas, A. S. (1954). Unpublished.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. Roy. Soc.* A, **225**, 287.

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

Kendrew, J. C. (1952). Private communication.

Sayre, D. (1952). *Acta Cryst.* **5**, 60.

Zachariasen, W. H. (1952). *Acta Cryst.* **5**, 68.