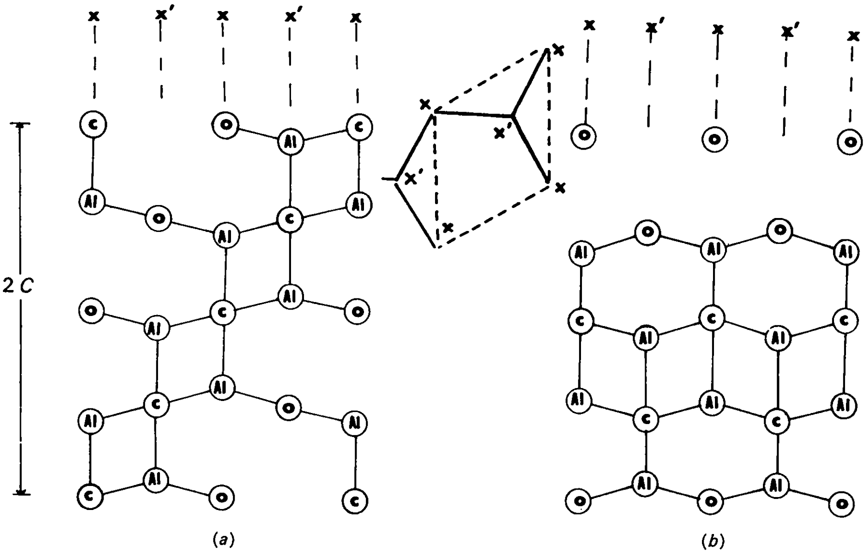

Fig. 3. Two possible short-range order structures for Al$_2$CO.

be apparent in the crystal owing to the long-range disorder.

This research was supported by the Office of Ordnance Research, U.S. Army, through Grant DA-ORD-31-124-61-G63.

References

Amma, E. L. & Jeffrey, G. A. (1961). *J. Chem. Phys.* **34**, 252.

Austerman, S. B. (1962). *Bull. Amer. Phys. Soc.* **7**, II, 607.

Carpenter, G. B. (1963). Technical Report No. C2, Chemistry Dept., Metcalf Research Laboratory, Providence, Rhode Island.

Cruickshank, D. W. J. & Rollett, J. S. (1953). *Acta Cryst.* **6**, 705.

International Tables for X-ray Crystallography (1959). Vol. II, p. 344. Birmingham: Kynoch Press.

Jeffrey, G. A., Parry, G. & Mozzi, R. L. (1956). *J. Chem. Phys.* **25**, 1024.

Jeffrey, G. A. & Slaughter, M. (1963). *Acta Cryst.* **16**, 177.

Jeffrey, G. A. & Wu, V. Y. (1963). *Acta Cryst.* **16**, 559.

Kohn, J. A. & Eckart, D. W. (1964). *Z. Kristallogr.* **119**, 454.

Parthé, E. (1963). *Z. Kristallogr.* **119**, 204.

Parthé, E. (1964). Private communication.

Pauling, L. (1960). *The Nature of the Chemical Bond.* 3rd Edn. Ithaca: Cornell Univ. Press.

Rundle, R. E. (1948). *Acta Cryst.* **1**, 180.

Shiono, R. (1963). Technical Report No. 43, Univ. of Pittsburgh Crystallography Laboratory.

Stackelberg, M. von & Schnorrenberg, E. (1934). *Z. Phys. Chem.* **B27**, 37.

Stackelberg, M. von, Schnorrenberg, E., Paulus, D. & Spiess, K. F. (1935). *Z. Phys. Chem.* **A173**, 127, 140.

Wadsley, A. D. (1963). *Advanc. Chem.*, No. 39, *Non-Stoichiometric Compounds.*

Wadsley, A. D. (1964). *Non-Stoichiometric Compounds*, Chapter 3. New York: Academic Press.

Acta Cryst. (1966). **20**, 547

A Powerful Phase-determining Equation for the Invariant Structure Factors in the Centrosymmetric Projection of the Non-centrosymmetric Space Group $P2_1$

By R. C. G. Killean

Department of Physics, The University, St. Andrews, Scotland

(*Received* 27 *April* 1965 *and in revised form* 8 *September* 1965)

A phase-determining equation is developed based on the Harker–Kasper inequality for a diad-screw axis. Probabilities for a given phase as high as 0·99 are obtained.

Introduction

A solution of the crystal structure problem involves solving the structure factor equation

$$F(\mathbf{h}) = \sum_{i=1}^{N} f_i \exp(2\pi j \mathbf{h} \cdot \mathbf{r}_i)$$

for the N triads (r_i). Hauptman & Karle (1953) have shown that for the special case of a centrosymmetric space group the pioneer work of Rayleigh (1919) on two-dimensional random walks may be used to provide a probability distribution for the phase of a structure factor and have extensively applied this technique to solving structures in centrosymmetric space groups.

Killean (1962) has used an identical technique to examine the distribution of phases in the centrosymmetric projection of some known structures in the space group $P2_1$ and Killean, Ferrier & Young (1962) have solved the structure of α-D-glucose monohydrate by this method. In these investigations only zonal data were used and the efficiency of the phase determining method was low, yielding at the best only positive phases with any degree of accuracy because of the skew distribution of the normalized structure factors. A more efficient method has been developed which utilizes the three-dimensional structure factor data and gives both positive and negative phases for the centrosymmetric projection. The probability equation developed is based on the Harker–Kasper inequality for a diad-screw axis, but the power of this probability technique when applied to diad-screw axes has not been generally realized, although its possible use has been commented on.

A phase determining equation

The assumptions that are usually made in deriving the probability relationships for the phases of structure factors are (i) that hydrogen atoms may be disregarded, (ii) that the scattering factors of the heavier atoms are equal, and (iii) that their coordinates are random variables. The structure factor expression then represents a two-dimensional random walk, and adopting these assumptions the structure factor equation for space groups $P2_1$ can be written as

$$F(hkl)=f\sum_{i=1}^{N/2}\{\exp[2\pi j(hx_i+ky_i+lz_i)] + (-1)^k\exp[2\pi j(-hx_i+ky_i-lz_i)]\}$$

and hence by simple algebra,

$$\{F(hkl)F^*(hkl)\}/Nf^2=E^2(hkl) = 1+(-1)^kU(2h,0,2l)+\Phi(x_iy_iz_ix_jy_jz_j) \quad (1)$$

where $\Phi(x_iy_iz_ix_jy_jz_j)$ is a double summation over all independent atoms and has the properties

$$\langle\Phi(x_iy_iz_ix_jy_jz_j)\rangle=0$$
$$\langle\Phi^2(x_iy_iz_ix_jy_jz_j)\rangle=1\,.$$

Rearranging (1) and summing over all possible values of k with h and l constant gives

$$\sum_{k=0}^{t}(-1)^k\{E^2(hkl)-1\}-(t+1)U(2h,0,2l) = \sum_{k=0}^{t}(-1)^k\Phi(x_iy_iz_ix_jy_jz_j)\,, \quad (2)$$

where t will be determined by the amount of data available and in general will be equal to the value of $k_{\max}$ for the given constant values h and l. The right-hand side of equation (2) has mean value zero, and a variance given by

$$\sigma^2=\sum_{r=0}^{t}\sigma_r^2+\sum_{\substack{r=1\\ r>s}}^{t}\sum_{s=0}^{t-1}R_{rs}\sigma_r\sigma_s$$

where all σ_r are unity and the R_{rs} are the correlation coefficients which to a first approximation will be taken as zero. Hence $\sigma^2=(t+1)$ and the left-hand side of (2) is, by the central limit theorem, distributed according to the function

$$P\propto\exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{k=0}^{t}(-1)^k\{E^2(hkl)-1\} -(t+1)U(2h,0,2l)\right\}^2\right].$$

It follows in the usual way that if $P_+(2h,0,2l)$ is the estimate of the probability of $U(2h,0,2l)$ having a positive phase then

$$P_+(2h,0,2l) =\tfrac{1}{2}\left[1+\tanh\left\{|U(2h,0,2l)|\left\{\sum_{k=0}^{t}(-1)^k(E^2(hkl)-1\right\}\right\}\right]. \quad (3)$$

It should be emphasized that although equation (3) predicts phases for a projection in which there may be overlap of atoms, it is not rendered invalid by this overlap as its value is derived using three-dimensional data, and not just the data of the centrosymmetric projection. It is clear that the greater the value of t the closer $P_+(2h,0,2l)$ will be to the true probability for the phase and consequently equation (3) will produce more accurate results for comparatively long screw-axes.

Effect of scaling error on probability equation

It is well known that the valid application of the Harker–Kasper inequalities and probability equations containing terms of the form $\{E^2(hkl)-1\}$ is critically dependent upon accurate scaling of the experimental data. Scaling errors do not effect equation (3) to the same extent and although the absolute probability value will be effected by such an error it is extremely unlikely that the phase angle will be altered unless there is a gross error in the postulated B value, or the absolute probability is nearly 0·5. It is easy to see that the scale factor has no effect upon the phase angle which is determined by the sign of the expression

$$\sum_{k=0}^{t}(-1)^k\{\alpha E^2(hkl)-1\}$$

where α is the term expressing the error in the postulated normalized structure factors. Suppose t is odd, then

$$\sum_{k=0}^{t_{\text{odd}}}(-1)^k\{\alpha E^2(hkl)-1\}=\alpha\sum_{k=0}^{t_{\text{odd}}}(-1)^kE^2(hkl)\,,$$

and the sign of this expression is independent of α.

Probably the most effective way of dealing with errors in the overall B value is to postulate upper and lower limits for its value and, using the appropriate scale factors, compute two probabilities for each structure factor. The closest of these probability values to 0·5 can then be used as the worst possible estimate of the absolute probability.

Application to three known structures

Equation (3) has been investigated for three known structures: (*a*) cellobiose (Jacobson, Wunderlich & Lipscomb, 1961) with 13·05 Å unique axis and poor resolution on the (010) projection, (*b*) α-D-glucose monohydrate (Killean, Ferrier & Young, 1962) with 5·10 Å unique axis, and (*c*) 2′-methyl-1,2-benzanthraquinone (Ferrier & Iball, 1963) with 4·06 Å unique axis, but only data available from the first three layer lines. The results for all invariant structure factors with probabilities for a given phase greater than 0·6 are listed in Table 1. It is seen that there is an 85% success for cellobiose, an 86% success for α-D-glucose monohydrate and a 71% success for 2′-methyl-1,2-benzanthraquinone.

The most satisfactory method of analysis of these results is obtained by computing the most probable number of predictions correct within a given group of structure factors. Suppose that there exists a group of N structure factors, then the probability of M of these N structure factors having been assigned the correct phases is obtained by developing the product

$$(P_1\theta+Q_1)\,(P_2\theta+Q_2)\cdots(P_N\theta+Q_N)$$

where P_i is the probability for the phase of the ith structure factor (always $P_i > 0{\cdot}5$) and Q_i is the probability of the opposite event, and finding the coefficient of θ^M. The most probable number correct is then given by the value of M for which θ^M has the largest coefficient. It is worth noting that because of the skew distribution of these coefficients it is to be expected that the actual result will be better on the average than the most probable result. The results obtained for the three test structures are given in Table 2, where the computed most probable number correct is compared with the actual number correct on the basis of the known structures. It will be seen that in only one case is the actual number correct less than the computed most probable number correct, and in four cases it is better.

Table 1. *Predicted phases for three structures using equation* (3)

Predicted phases for cellobiose h k l	$\lvert F(hkl)\rvert$	Predicted phase	t	Probability of correctness of predicted phase	True phase
12 0 0	9·4	+	16	0·99	+
8 0 $\bar{2}$	34·2	+	12	0·99	+
0 0 4	26·9	+	16	0·99	+
6 0 $\bar{2}$	25·3	−	16	0·97	−
6 0 0	48·1	+	13	0·92	+
4 0 $\bar{4}$	18·1	+	15	0·89	+
8 0 2	24·8	+	14	0·88	+
4 0 2	35·5	+	14	0·87	+
0 0 2	24·7	−	16	0·86	−
6 0 2	12·9	−	16	0·77	−
10 0 0	20·9	+	16	0·75	−
4 0 4	5·0	−	14	0·64	+
2 0 2	12·1	−	15	0·60	−
Predicted phases for α-glucose monohydrate					
6 0 4	31·2	−	7	0·74	−
0 0 8	27·8	−	7	0·71	−
4 0 $\bar{8}$	20·4	−	7	0·69	+
6 0 2	16·8	+	7	0·67	+
10 0 $\bar{4}$	10·6	−	6	0·65	−
0 0 12	8·2	+	6	0·61	+
6 0 $\bar{8}$	11·2	−	7	0·60	−
Predicted phases for 2’-methyl-1,2-benzanthraquinone					
2 0 $\bar{6}$	31·9	+	4	0·75	+
6 0 0	34·8	−	4	0·68	−
22 0 2	2·6	+	4	0·65	−
16 0 4	18·8	−	4	0·63	−
$\overline{12}$ 0 6	8·1	+	4	0·62	+
10 0 2	9·5	+	4	0·62	+
14 0 4	17·4	−	4	0·62	+

Table 2. *Analysis of results obtained for the three test structures*

Analysis of cellobiose results													
Number of phases predicted	1	2	3	4	5	6	7	8	9	10	11	12	13
Most probable number correct	1	2	3	4	5	6	7	8	9	10	10	11	11
Actual number correct	1	2	3	4	5	6	7	8	9	10	10	10	11
Analysis of α-glucose monohydrate results													
Number of phases predicted	1	2	3	4	5	6	7						
Most probable number correct	1	2	2	3	4	4	5						
Actual number correct	1	2	2	3	4	5	6						
Analysis of 2′-methyl-1,2-benzanthraquinone results													
Number of phases predicted	1	2	3	4	5	6	7						
Most probable number correct	1	2	2	3	3	4	5						
Actual number correct	1	2	2	3	4	5	5						

I should like to thank Professor D. Rogers for some helpful discussion.

References

FERRIER, R. P. & IBALL, J. (1963). *Acta Cryst.* **16**, 513.

HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem I. The Centrosymmetric Crystal.* A.C.A. Monograph.

JACOBSON, R. A., WUNDERLICH, J. A. & LIPSCOMB, W. N. (1961). *Acta Cryst.* **14**, 598.

KILLEAN, R. C. G. (1962). Ph. D. Thesis. Univ. of St. Andrews.

KILLEAN, R. C. G., FERRIER, W. G. & YOUNG, D. W. (1962). *Acta Cryst.* **15**, 911.

RAYLEIGH, LORD (1919). *Phil. Mag.* 6th series, **37**, 341.