

The functional dependence of $\text{Pr}(\alpha_0, \pi - \alpha_0)$ on α_0 is shown in Fig. 1 (*a* to *d*) for $(\sin \theta)/\lambda = 0.1, 0.2, \dots, 1.2$ when $\langle |\Delta r| \rangle$ has values 0.05, 0.10, 0.15 and 0.2 Å respectively.

The following points may be noted for making use of Fig. 1 in actual cases: (i) The value of $\text{Pr}(\alpha_0, \pi - \alpha_0)$ read from a *given* curve corresponding to a *given* value of $(\sin \theta)/\lambda$ represents the fraction with reference to reflexions in that range of $(\sin \theta)/\lambda$. Thus for example, if n denotes the total number of reciprocal-lattice points in the range $0.05 \leq (\sin \theta)/\lambda \leq 0.15$ for a given crystal, then for the case $\langle |\Delta r| \rangle = 0.1$ Å and $\alpha_0 = 30^\circ$, we have $\text{Pr}(\alpha_0, \pi - \alpha_0) = 0.04$. Thus under the above conditions the actual number of reflexions lying in the $(\sin \theta)/\lambda$ range 0.05 to 0.15 and having phases whose values deviate from the centrosymmetric values of 0 or π by more than 30° will be $0.04 \times n$; (ii) From the consideration given above it is clear that the value of $\langle |\Delta r| \rangle$ must be known *a priori* in order to make use of Fig. 1. The value of $\langle |\Delta r| \rangle$ for a given crystal exhibiting a degree of centrosymmetry could be estimated by a procedure discussed in the previous paper (PP, 1974).

It is seen from Fig. 1 that in a given crystal (*i.e.* $\langle |\Delta r| \rangle$ is fixed) the percentage of reflexions for

which the magnitude of the phase angles would lie in any given range, say $30^\circ \leq \alpha \leq 150^\circ$, is more for the high-angle than for the low-angle reflexions. For example when $\langle |\Delta r| \rangle \approx 0.1$ Å, only 4% of the reflexions have phase angle distributed in the range 30 to 150° when $(\sin \theta)/\lambda = 0.1$ while it is as high as 17% when $(\sin \theta)/\lambda = 0.4$. The deviation of the phase angles from the centrosymmetric values of 0 or π is thus more effectively exhibited for the reflexions with $(\sin \theta)/\lambda > 0.1$. This property might possibly be exploited for refinement of such structures by a modified Fourier synthesis and this problem is under investigation.

One of the authors (V.P.) would like to thank the Council of Scientific and Industrial Research, New Delhi, India, for financial assistance.

References

- LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
 PARTHASARATHY, S. (1965). *Acta Cryst.* **18**, 1028–1035.
 PARTHASARATHY, S. & PARTHASARATHI, V. (1974). *Acta Cryst.* **A30**, 43–46.
 RAMACHANDRAN, G. N. & SRINIVASAN, R. (1970). *Fourier Methods in Crystallography*. New York: John Wiley.

Acta Cryst. (1974). **A30**, 576

A Contribution to the Determination of a Correct System of Signs of Structure Factors for Centrosymmetric Crystals

BY J. HAŠEK

Institute of Macromolecular Chemistry, Czechoslovak Academy of Sciences, 162 06 Prague 6, Czechoslovakia

(Received 8 January 1974; accepted 26 February 1974)

To find a solution of the phase problem of centrosymmetric crystals, it is as a rule necessary to find among many possible solutions a correct system of signs of the structure factors. This paper describes the procedure used for finding a correct solution based on a comparison of the theoretical and empirical distribution functions of the positive signs of structure products.

1. Introduction

When solving the phase problem of centrosymmetric crystals by using direct methods it is usually necessary to decide at the end of the calculation which of the suggested sets of signs of the structure factors is the correct one (Karle & Karle, 1966; Ahmed, 1969; Woolfson, 1971, *etc.*). There exist several different procedures for determining which of the suggested solutions is the likeliest to be the correct one.

Schenk (1973*a*) has described an effective procedure for determining the correct set of signs using Harker-Kasper type relations.

Riche (1973) suggested for these purposes the so-called phase function based on the Sayre-Hughes formula.

It has been the objective of this paper to show how to create criteria suited for the determination of a correct system of signs of the structure factors based in the so-called statistical relationship (*e.g.* Hauptman & Karle, 1953; Cochran & Woolfson, 1955; Karle & Gilardi, 1973).

2. Consistency test

As a rule, sign relationships of the statistical type may be rearranged into a form in which the probability $P_{+\text{theor}}$ of the positive sign of the structure product* is

* The structure product is such a product of structure factors $\prod_{i=1}^n F_i^{\pm 1}$ for which the corresponding linear combination of their phases $\sum_{i=1}^n a_i$. φ_{HI} is a structure seminvariant.

expressed in terms of a non-decreasing function of the expression W consisting of the absolute values of the structure factors. The expression W will be referred to below as a weight of the respective statistical relationship.

It may be assumed that the most probable solution will be the one giving the best agreement between the relative frequency values of the positive signs of the structure products $P_{+exp}^{(i)}$ for given sufficiently large intervals of the weights and, on the other hand, the theoretical probability values of positive sign of the structure product $P_{+theor}^{(i)}$ corresponding to the average weights in these intervals. An appropriate measure for the fit of these values is a weighted sum K of their squared deviations:

$$K = \frac{\sum_{i=1}^p r_i \cdot (P_{+exp}^{(i)} - P_{+theor}^{(i)})^2}{\sum_{i=1}^p r_i}, \quad (1)$$

where the summation proceeds over all groups of the structure products, r_i is the statistical weight characterizing the degree of accuracy of the determination of the difference $P_{+exp}^{(i)} - P_{+theor}^{(i)}$ for an i th contribution, and p is the number of groups of the structure products.

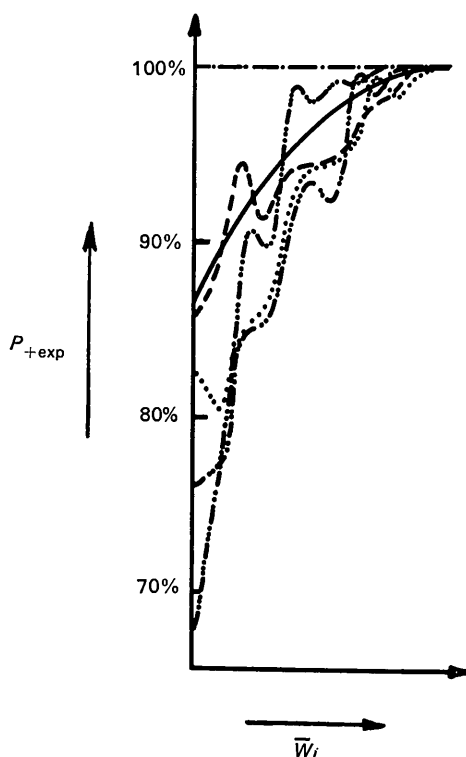


Fig. 1. Comparison of the curves P_{+exp} (for solution 1 ---, 2 ·····, 3 - · - · - ·, 4 - · - · - · and 7 - · - · - ·) with the corresponding P_{+theor} (full line) for $C_{16}H_{21}N_3O_3Cl$. The best fit between the curves P_{+exp} and P_{+theor} is obtained by using solution 1.

It is expected that the smaller the coefficient K the larger the probability that the associated set of signs will be the correct one.

The reliability of the consistency test is strongly dependent on the extent to which the premises of validity of the statistical relationship used for calculating $P_{+theor}^{(i)}$ are fulfilled for the structure under investigation. Since the existing methods of calculation of $P_{+theor}^{(i)}$ do not allow the special structure features to be taken into account (overlapping of Patterson peaks, pseudosymmetry, etc.), it may happen that for some structures the correct system of signs of the structure factors will not correspond to the lowest coefficient K , but to some of the subsequent ones.

3. Consistency test based on the Σ_2 relationship

(a) Method

Let us further restrict ourselves only to the case when the tested structure products are triplets of normalized structure factors $E_H E_K E_{H-K}$.

To calculate the probability P_+ that the product $E_H E_K E_{H-K}$ has the positive sign, the relationship (Cochran & Woolfson, 1955)

$$P_+ = \frac{1}{2} + \frac{1}{2} \tanh \{ \sigma_3 \cdot \sigma_2^{-3/2} |E_H E_K E_{H-K}| \} \quad (2)$$

is usually employed, where $\sigma_i = \sum_{j=1}^N Z_j^i$, N is the number of atoms in the unit cell, and Z_j is the number of electrons of the j th atom.

The procedure for determining a correct system of signs of the structure factors based on the validity of relationship (2) is as follows:

(1) The triplet list is divided into p groups so that each group contains approximately 100 triplets and any arbitrary triplet in the $(i+1)$ th group has a higher weight $|E_H E_K E_{H-K}|$ than any arbitrary triplet in the i th group.

(2) For the tested system of signs of the structure factors the signs of all triplets in the triplet list are determined.

(3) In each group of triplets the number of triplets with the positive sign N_i^+ , the number of triplets with the negative sign N_i^- , the total number of signed triplets $N_i = N_i^+ + N_i^-$ and the average weight of the triplets

$$\bar{W}_i = (1/N_i) \sum \sigma_3 \cdot \sigma_2^{-3/2} |E_H E_K E_{H-K}| \quad (3)$$

are determined. The summation is over all signed triplets in the i th group.

(4) For each group of triplets the theoretical probability of the positive sign of the triplet with the average weight \bar{W}_i is calculated using relation (3),

$$P_{+theor}^{(i)} = \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) \tanh (\bar{W}_i), \quad (4)$$

and also the relative frequency of the positive signs of the triplets in the i th group of triplets,

$$P_{+exp}^{(i)} = N_i^+ / (N_i^+ + N_i^-). \quad (5)$$

$P_{+theor}^{(i)}$ [and $P_{+exp}^{(i)}$] values may be plotted depending on the corresponding average weight of the triplets, \bar{W}_i . A curve P_{+theor} (and P_{+exp}) may be plotted through these points.

(5) Coefficients which we shall call \mathcal{K} (for statistical weights $r_i = N_i$) and \mathcal{K}' (for statistical weights $r_i = N_i \bar{W}_i$) are calculated from relationship (1).

(6) The whole procedure is carried out for each of the tested solutions.

(7) The most probable system of signs of the structure factors of all the tested ones is that which provides the best fit between the P_{+exp} and P_{+theor} curves. Since the coefficients \mathcal{K} and \mathcal{K}' express a certain measure of the area between these two curves, the most probable solution is indicated by the lowest coefficients \mathcal{K} and \mathcal{K}' .

When comparing the theoretical and empirical distribution functions of the positive signs of triplets, it should be borne in mind that the course of the relative frequency of the positive-sign triplets P_{+exp} depending on \bar{W}_i may differ considerably, even for a correct system of signs of the structure factors, from the course of the curve P_{+theor} obtained by using relationship (4). The deviations of the curve P_{+exp} from the curve P_{+theor} calculated from (4) may be due to some characteristic features of the structure solved (the presence of a heavy atom, overlap of Patterson peaks, non-crystallographic symmetry, etc.), which disturb the validity of the assumptions used for deriving the curve P_{+theor} , and/or to unreliably determined experimental values of normalized structure factors.

Schenk (1973b) shows in some examples how in the case of real structures the curves P_{+exp} for a correct solution differ from P_{+theor} calculated according to (4). Further studies on the course of the curves P_{+exp} can be found in a paper by Krieger & Schenk (1973).

(b) Comparison with the figures of merit

Coefficients M called the 'figures of merit' (Woolfson, 1971) are very frequently used for the identification of a correct system of signs of the structure factors:

$$M = \frac{\sum w_i}{\sum |w_i|} \quad (6)$$

In this expression the summation proceeds over all triplets \mathbf{H} , \mathbf{K} , $\mathbf{H}-\mathbf{K}$ with determined signs. The weights w_i of the triplets are taken either as $w_i = \text{sign}(E_{\mathbf{H}}) \cdot \text{sign}(E_{\mathbf{K}}) \cdot \text{sign}(E_{\mathbf{H}-\mathbf{K}})$ or $w_i' = E_{\mathbf{H}} \cdot E_{\mathbf{K}} \cdot E_{\mathbf{H}-\mathbf{K}}$ or $w_i'' = w_i [\frac{1}{2} + \frac{1}{2} \tanh(\sigma_3 \cdot \sigma_2^{-3/2} |E_{\mathbf{H}} E_{\mathbf{K}} E_{\mathbf{H}-\mathbf{K}}|)]$; the associated coefficients are M , M' , and M'' respectively. It is expected that the correct set of signs of the structure factors will be found among those solutions which have the largest values of M .

Schenk (1972) has shown that the application of the figures of merit brings about a number of difficulties in the symmorphic space groups, *i.e.* space groups not containing glide planes or screw axes. The solution corresponding to the largest value of M in symmorphic space groups is always the highly improbable case of a

structure with all phases $\varphi_{\mathbf{H}} = 0$, but for the other solutions also the coefficients give no reliable indication of correctness.

It may be shown that between the coefficient M and a special case of the coefficient \mathcal{K} [for $p=1$, $P_{+theor}^{(1)} = 1$] there exists a simple relationship $M = 1 - 2\sqrt{\mathcal{K}}$. Thus, on the one hand, the form of the curves P_{+exp} is not taken into account in the coefficient M , and on the other, it is based on rather a rough assumption that the probability of the positive sign of an arbitrary triplet in the list of triplets is 100%.

Consequently, the main drawback of the figures of merit consists in that they do not take into account information on the negative signs of triplets provided by relationship (2). According to this relationship, a group of triplets with the same value $|E_{\mathbf{H}} E_{\mathbf{K}} E_{\mathbf{H}-\mathbf{K}}|$ contains $50 - 50 \cdot \tanh(\sigma_3 \cdot \sigma_2^{-3/2} |E_{\mathbf{H}} E_{\mathbf{K}} E_{\mathbf{H}-\mathbf{K}}|)$ per cent of triplets with the negative sign of the product $E_{\mathbf{H}} E_{\mathbf{K}} E_{\mathbf{H}-\mathbf{K}}$. Since the consistence test makes use of this information, it may be shown that the difficulties appearing in symmorphic space groups when the figures of merit are used are greatly reduced. Also the coefficients \mathcal{K} and \mathcal{K}' show the fit between the whole courses of the curves P_{+exp} and P_{+theor} , in contrast with the coefficient M . It is quite likely, therefore, that coefficients of type \mathcal{K} may lead to a certain improvement also in the case of non-symmorphic space groups where the figures of merit give good results.

(c) Experimental results

The use of the test of consistency has been demonstrated by using the chlorate of 4,4'-bis(dimethylamino)diphenylamine radical, $\text{C}_{16}\text{H}_{21}\text{N}_3\text{O}_3\text{Cl}$, $P\bar{1}$,

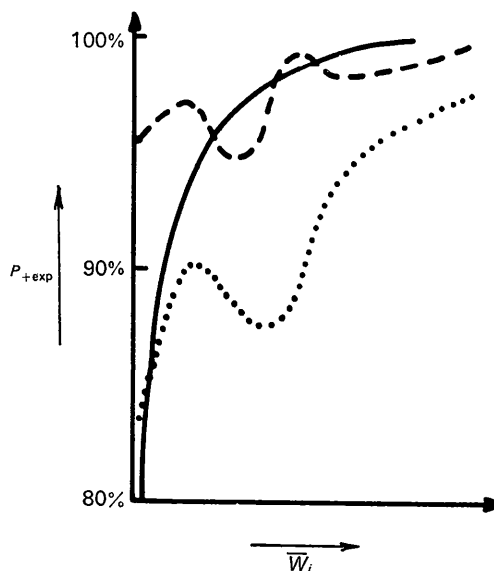


Fig. 2. Comparison of the curves P_{+exp} (for solutions 1 --- and 2 ···· having the lowest coefficients \mathcal{K}) with the curve P_{+theor} (full line) for *cis*-1,3,5-trichlorocyclohexane. The best fit of the curves P_{+exp} and P_{+theor} is given by solution 1.

$Z=2$ (Hlavatá, 1971) and *cis*-1,3,5-trichlorocyclohexane, $C_6H_9Cl_3$, $C2/c$, $Z=8$ (Huml & Hašek, 1973).

The coefficients \mathcal{K} , \mathcal{K}' and the corresponding figures of merit M , M' and M'' for eight possible solutions with the lowest \mathcal{K} for $C_{16}H_{21}N_3O_3Cl$ are summarized in Table 1. Solution No. 1 gave a correct picture of the structure. The coefficients \mathcal{K} and \mathcal{K}' in this case indicate the correct solution quite unequivocally, while figures of merit M , M' and M'' assume their highest values for solutions 4 and 7.

Table 1. Coefficients \mathcal{K} , \mathcal{K}' and figures of merit M , M' and M'' for the chlorate of 4,4'-bis(dimethylamino)diphenylamine radical (symmorphic space group $P\bar{1}$)

The correct solution is solution No. 1.

Solution No.	$\mathcal{K} \cdot 10^5$	$\mathcal{K}' \cdot 10^5$	$M \cdot 10^3$	$M' \cdot 10^3$	$M'' \cdot 10^3$
1	42	28	952	965	953
2	156	86	934	956	936
3	246	133	936	965	933
4	274	145	1000	1000	1000
5	323	158	934	957	937
6	395	192	924	951	927
7	428	190	951	972	954
8	445	221	922	951	926

The curves P_{+exp} for solutions 1, 2, 3, 4 and 7 (with the lowest coefficients \mathcal{K} , \mathcal{K}') are compared in Fig. 1 with the corresponding curve P_{+theor} . Solutions 4 and 7 give a higher number of the positive signs of triplets with a high weight $|E_H E_K E_{H-K}|$ than solutions 1; consequently, according to the figures of merit, they are wrongly suggested as the most probable ones. On the other hand the coefficients \mathcal{K} , \mathcal{K}' indicate quite unambiguously solution 1 as the most correct one, since the curve P_{+exp} for solution 1 gives the best fit with the curve P_{+theor} (full line).

Table 2 shows the coefficients \mathcal{K} , \mathcal{K}' and the corresponding figures of merit for eight possible solutions for *cis*-1,3,5-trichlorocyclohexane; Fig. 2 gives a comparison of the curves P_{+exp} for solutions 1 and 2 with the corresponding curve P_{+theor} . Table 2 and Fig. 2 indicate quite unequivocally that the correct solution is solution 1.

The calculations were carried out by means of the program 'Consistence test' (Fortran-IV, IBM-370) which may be used as a continuation of the sets of

Table 2. Coefficients \mathcal{K} , \mathcal{K}' and figures of merit M , M' and M'' for *cis*-1,3,5-trichlorocyclohexane (non-symmorphic space group $C2/c$)

The correct solution is solution No. 1.

Solution No.	$\mathcal{K} \cdot 10^5$	$\mathcal{K}' \cdot 10^5$	$M \cdot 10^3$	$M' \cdot 10^3$	$M'' \cdot 10^3$
1	93	29	988	991	988
2	213	156	950	961	951
3	1441	2782	906	865	906
4	2265	3575	858	828	858
5	2821	1964	837	873	837
6	3020	2070	831	870	832
7	5126	5547	768	759	768
8	5547	5986	752	755	763

programs *SAP* (Ahmed, 1969) and *MULTAN* (Woolfson, 1971).

Conclusion

The newly suggested coefficients of type \mathcal{K} indicate the correct solution more reliably than the figures of merit, mainly in symmorphic space groups. Consequently they seem to be a suitable aid for finding the correct solution to the phase problem of centrosymmetric crystal structures.

The author thanks Dr H. Schenk and Dr K. Huml for their helpful criticism of the manuscript.

References

- AHMED, F. R. (1969). *Crystallographic Computing*. Copenhagen: Munksgaard.
- COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1–12.
- HAUPTMAN, H. & KARLE, J. (1953). *Solution to the Phase Problem 1. The Centrosymmetric Crystal. ACA Monograph No. 3*. Pittsburgh: Polycrystal Book Service.
- HLAVATÁ, D. (1971). *Acta Cryst.* **B27**, 1483–1493.
- HUML, K. & HAŠEK, J. (1972). *Acta Cryst.* **B28**, 2852–2856.
- KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.
- KARLE, J. & GILARDI, R. D. (1973). *Acta Cryst.* **A29**, 401–407.
- KRIEGER, W. & SCHENK, H. (1973). *Acta Cryst.* **A29**, 720–723.
- RICHE, C. (1973). *Acta Cryst.* **A29**, 133–137.
- SCHENK, H. (1972). *Acta Cryst.* **A28**, 412–422.
- SCHENK, H. (1973a). *Acta Cryst.* **A29**, 31–34.
- SCHENK, H. (1973b). *Acta Cryst.* **A29**, 503–507.
- WOOLFSON, M. M. (1971). Summer School in York (Lecture notes E). Physics Department, University of York, Heslington, York YO1 5DD, England.