Acta Cryst. (1971). B**27**, 1099

A Probability Consideration for Evaluating the Reliability of the Relationships among the Symbols in the Symbolic Addition Procedure

BY TOORU TAGA AND KENJI OSAKI

Faculty of Pharmaceutical Sciences, Kyoto University, Yoshida, Sakyo-ku, Kyoto, Japan

(*Received* 24 *July* 1970)

Reliability of the relationships among the symbols which are found in the symbolic addition procedure has been evaluated on the basis of probability considerations. Formulae for such reliabilities have been presented for both centrosymmetric and non-centrosymmetric crystals, and some possible applications were discussed, which could be helpful in an automatic phase determining procedure.

Introduction

It is well known that for a centrosymmetric crystal the sign of $E_{\mathbf{h}}$ is approximately related to others by the sign relation

$$S\,E_{\mathbf{h}} \simeq S \sum_{\mathbf{k}_r} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}} \tag{1}$$

(Hauptman & Karle, 1953), and for a noncentrosymmetric crystal the phase $\varphi_{\mathbf{h}}$ associated with $E_{\mathbf{h}}$ is approximately related to other phases by the relation

$$\varphi_{\mathbf{h}} \simeq \langle \varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}} \rangle_{\mathbf{k}_r} \tag{2}$$

(Karle & Karle, 1966). In the symbolic addition procedure, the formula (1) or (2) is employed to define as many phases as possible of the largest $|E_{\mathbf{h}}|$ in terms of the initially assigned phases and symbols. In the course of the phase determination using (1) or (2), it may often happen that some of the unknown symbols are expressed in terms of others. Such relationships among the symbols may be used to eliminate one of the symbols included in the relationships and will help us in reducing the number of combinations of undefined signs or phases. As Karle & Karle (1966) pointed out, the relationship among the symbols should be accepted only with great care and a criterion which has been applied most commonly is based upon the number of inconsistencies among the contributors to (1) or (2). In this paper we introduce probabilities to evaluate the reliability of the relationships among the symbols. These probabilities could be used as the basis of a more quantitative criterion for accepting the relationships in an automatic phase determining procedure. The probability consideration for centrosymmetric crystals is treated first and that for noncentrosymmetric crystals is described afterwards.

Centrosymmetric crystal

The probability that the sign of $E_{\mathbf{h}}$ defined by (1) is positive was given by Woolfson (1954) and Cochran & Woolfson (1955) in the following form

$$P_{+}(\mathbf{h}) = \tfrac{1}{2} + \tfrac{1}{2}\tanh(\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}| \sum_{\mathbf{k}_r} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}})\,. \tag{3}$$

When the sign of a certain $E_{\mathbf{h}}$ is not given as either plus or minus but is defined in terms of a symbol a, the probability that the sign of $E_{\mathbf{h}}$ is equal to the assigned symbol is given also by the same formula (3). In this case, summation over $\mathbf{k}_r$ includes only $E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}$'s which define the sign of $E_{\mathbf{h}}$ in terms of the symbol a, and $E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}$'s are to be treated as $+|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|$ for those which assign $+a$ to the sign of $E_{\mathbf{h}}$ and as $-|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|$ for those which assign $-a$.

If there is a reflexion whose sign is given by two symbols, there will be a certain possibility of finding the relationship between these symbols. Reliability of the relationship in question could be evaluated as a probability which will be given in the following.

If both symbols a and b are assigned to the sign of $E_{\mathbf{h}}$ and the corresponding probabilities are $P_a(\mathbf{h})$ and $P_b(\mathbf{h})$, the probability of a being equal to b, denoted by $P_{a=b}(\mathbf{h})$, may be expressed as follows,

$$P_{a=b}(\mathbf{h}) = P_a(\mathbf{h})P_b(\mathbf{h}) + [1-P_a(\mathbf{h})][1-P_b(\mathbf{h})]\,, \tag{4}$$

for a is equal to b if both a and b are equal to the sign of $E_{\mathbf{h}}$ or if both are opposite to the sign of $E_{\mathbf{h}}$. By use of (3), (4) is expressed in the following form

$$P_{a=b}(\mathbf{h}) = \tfrac{1}{2} + \tfrac{1}{2}\tanh(\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}| \sum_{\mathbf{k}_a} E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}) \times \tanh(\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}| \sum_{\mathbf{k}_b} E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}})\,, \tag{5}$$

where $\mathbf{k}_a$ and $\mathbf{k}_b$ represent the restricted values of $\mathbf{k}$ for which the corresponding $E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}$'s give the symbols a and b to the sign of $E_{\mathbf{h}}$, respectively. If there is only one reflexion with the largest $E_{\mathbf{h}}$ whose sign is definable in terms of symbols a and b, the reliability of the relationship that the symbols a and b are equal may be evaluated by formula (5).

If there are several reflexions whose signs are definable in terms of both a and b, the probability that a and b are equal, indicated from each reflexion, may be given by $P_{a=b}(\mathbf{h}_1)$, $P_{a=b}(\mathbf{h}_2)$, $\cdots$ and the corresponding indication that a and b are opposite sign may be given by $1-P_{a=b}(\mathbf{h}_1)$, $1-P_{a=b}(\mathbf{h}_2)$, $\cdots$ respectively. An application of Bayes's theorem (Uspensky, 1937) shows that the overall probability of the equality of a and b, divided by the corresponding probability of the opposite sign is

$$\frac{P_{a=b}}{P_{a=-b}}=\frac{\prod\limits_{\mathbf{h}} P_{a=b}(\mathbf{h})}{\prod\limits_{\mathbf{h}}[1-P_{a=b}(\mathbf{h})]} \tag{6}$$

and since $P_{a=b}+P_{a=-b}=1$,

$$P_{a=b}=\frac{\prod\limits_{\mathbf{h}} P_{a=b}(\mathbf{h})}{\prod\limits_{\mathbf{h}} P_{a=b}(\mathbf{h})+\prod\limits_{\mathbf{h}}[1-P_{a=b}(\mathbf{h})]}\,. \tag{7}$$

Substituting (5) to (7), we obtain

$$P_{a=b}=\frac{1}{1+\prod\limits_{\mathbf{h}}\left\{\dfrac{1-q_{a=b}(\mathbf{h})}{1+q_{a=b}(\mathbf{h})}\right\}}\,, \tag{8}$$

where

$$q_{a=b}(\mathbf{h})=\tanh(\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}|\sum_{\mathbf{k}_a}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}})\tanh(\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}|\sum_{\mathbf{k}_b}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}})\,. \tag{9}$$

From (7) or (8), it may be seen that

$$P_{a=b}>\tfrac{1}{2} \text{ for } a=b \text{ and } P_{a=b}<\tfrac{1}{2} \text{ for } a=-b.$$

Formula (7) or (8) may be interpreted to include even the reflexion whose sign is not given by both a and b at the same time, if $P_{a=b}(\mathbf{h})$ is set equal to $\frac{1}{2}$ or if $q_{a=b}(\mathbf{h})$ is set equal to 0. The probability $P_{a=b}$ calculated by (7) or (8) will give the numerical reliability of the relationship between the symbols a and b considering all the reflexions.

So far the case with only two symbols has been considered. When three or more symbols have been used for the sign determination, the probabilities of the possible relationships among these symbols may be also evaluated by the application of the formula (5), (7) or (8). One example for the case using three symbols is shown in Table 1, where the probabilities that the sign of a certain reflexion has been defined in terms of these symbols are given in the left-hand part. Here the probability values less than 50 suggest that the sign of symbols assigned at the top line are to be reversed. In the right-hand part of the Table, the possible relationships among the symbols are given at the top line and the corresponding probabilities calculated from the values in the left-hand part, using the formula (5), are given at the appropriate positions. In this Table it may be noted that the identical relationships, such as $ab=+$, $a=b$, $ac=bc$, are collected on the same columns. Overall probabilities of each relationships considering all the reflexions can be evaluated by use of formula (7) from the probability values of the same columns. They are given in the last row corresponding to each relationship. The probabilities obtained in such a way may be useful in evaluating the reliabilities of the relationships among the symbols numerically.

Non-centrosymmetric crystal

For the case of the phase relation (2), the probability distribution for $\varphi_{\mathbf{h}}$ was given by Cochran (1955) and Karle & Karle (1966) in the following form

$$P(\varphi_{\mathbf{h}})=[2\pi I_0(\alpha)]^{-1}\exp\{\alpha\cos(\varphi_{\mathbf{h}}-\langle\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}}\rangle)\}\,, \tag{10}$$

where I_0 is a modified Bessel function and

$$\alpha=2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}|\{[\sum_{\mathbf{k}_r}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}\cos(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})]^2+[\sum_{\mathbf{k}_r}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}\sin(\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}})]^2\}^{1/2}\,. \tag{11}$$

In the course of the phase determination using (2), the phase of a certain reflexion may be defined in terms of a symbol a from a number of contributions in the right-hand side of (2) and at the same time the phase of the same reflexion may be defined in terms of another symbol b from other contributions. The probability distribution for $\varphi_{\mathbf{h}}$ based on the contributions which assign the symbol a to $\varphi_{\mathbf{h}}$ has a centre at a and

Table 1. *Probability of the relationship among the symbols for the case with three symbols*

	Possible assignments of symbols and the associated probabilities (%)								Possible relationships among the symbols and the associated probabilities (%)						
Reflexion	+	a	b	c	ab	bc	ca	abc	$a=+$	$b=+$	$c=+$	$ab=+$	$bc=+$	$ca=+$	$abc=+$
$\mathbf{h}_1$		95	90									86			
$\mathbf{h}_2$	85				80							71			
$\mathbf{h}_3$						75	80					65			
$\mathbf{h}_4$		10	15									78			
$\mathbf{h}_5$		50	60									50			
$\mathbf{h}_6$		90	10									18			
$\mathbf{h}_7$	70	85							64						
$\mathbf{h}_8$			90		75				70						
$\mathbf{h}_9$						60		90	58						
$\mathbf{h}_{10}$		80	80	15								68	29	29	
$\mathbf{h}_{11}$			5	70		90				66	14		32		
$\mathbf{h}_{12}$		30			10	70	65			66	44	56	38	34	42
$\mathbf{h}_{13}$	80		75	20		40				(65 56	(32 45		(44 35		
$\mathbf{h}_{14}$	10	20						80	74					32	26
Overall probabilities									94·2	89·9	4·7	98·3	4·8	9·0	20·3

may be denoted by $P_a(\varphi_{\mathbf{h}})$. $P_a(\varphi_{\mathbf{h}})$ is expressed by a formula corresponding to (10) in which $\langle\varphi_{\mathbf{k}}+\varphi_{\mathbf{h}-\mathbf{k}}\rangle$ is a and α is replaced by α_a, for the summations in (11) are restricted to the terms assigning a to $\varphi_{\mathbf{h}}$. For the case of symbol b, the corresponding $P_b(\varphi_{\mathbf{h}})$ and α_b may be also defined in the same way. In the non-centrosymmetric crystals, the probability of the relationship between a and b can not be given in such a simple form as was shown by (5) for the centrosymmetric crystals. Then we have to consider the probability distribution for the difference x between the phases represented by the symbols a and b. If this probability distribution is denoted by $P_{a-b,\mathbf{h}}(x)$, it may be expressed by the convolution of the distribution functions $P_a(\varphi_{\mathbf{h}})$ and $P_b(\varphi_{\mathbf{h}})$,

$$P_{a-b,\mathbf{h}}(x)=P_a(\varphi_{\mathbf{h}})*P_b(\varphi_{\mathbf{h}})=[(2\pi)^2I_0(\alpha_a)I_0(\alpha_b)]^{-1} \times\int_{-\pi}^{+\pi}\exp\{\alpha_a\cos(\varphi_{\mathbf{h}}-a-\tfrac{1}{2}x) +\alpha_b\cos(\varphi_{\mathbf{h}}-b+\tfrac{1}{2}x)\}\mathrm{d}\varphi_{\mathbf{h}}\,. \quad (12)$$

When the integral in (12) is expressed by a Bessel function,

$$P_{a-b,\mathbf{h}}(x)=[2\pi I_0(\alpha_a)I_0(\alpha_b)]^{-1} \times\, I_0\left(\sqrt{\alpha_a^2+2\alpha_a\alpha_b\cos x+\alpha_b^2}\right). \quad (13)$$

The probability distribution for x, given by (13), has a maximum at the origin and its shape depends upon the values of both α_a and α_b. The peak of the distribution at the origin becomes sharp only when the values of both α_a and α_b are big enough.

The probability distribution (13) can be used as a basis for evaluating the reliability of the equality of the symbols a and b. If $P_{a-b,\mathbf{h}}(x)$'s are given for several reflexions, we may obtain the overall probability distribution by multiplying individual distributions.

$$P_{a-b}(x)=A\prod_{\mathbf{h}}P_{a-b,\mathbf{h}}(x)\,, \quad (14)$$

where A is a normalizing constant given by

$$A=1\Big/\int_{-\pi}^{+\pi}\prod_{\mathbf{h}}P_{a-b,\mathbf{h}}(x)\,\mathrm{d}x\,. \quad (15)$$

Formula (14) corresponds to formula (7) for the centrosymmetric case. The probability distribution (14) can be calculated numerically at the end of the procedure employing (2). If the distribution of $P_{a-b}(x)$ has a sharp and high peak at the origin, the symbols a and b may be assumed to be equal. The reliability of the equality of a and b could be evaluated by the corresponding variance, which is given in this case by the expected value of x^2, calculated by the following equation,

$$\langle x^2\rangle=\int_{-\pi}^{+\pi}x^2P_{a-b}(x)\,\mathrm{d}x\,. \quad (16)$$

We may accept the equality to the symbols a and b, if the value of $\langle x^2\rangle$ is less than a suitable constant which is to be chosen taking into account of the permissible error.

So far a special case of the equality of two symbols has been treated for the sake of convenience, but in actual cases possible relationships among the symbols will be given in the form of linear equations including various unknown symbols. As an example, we may assume that the phase of a certain reflexion is assigned at the same time by two sets of linear combinations of symbols and phase terms, such as $a+3b-\pi$ and $b+c-\frac{3}{4}\pi$, where each expression is assumed to have been averaged over those which differ only in numerical terms. In this case, the possible relationship will be given by a linear equation, $a+2b-c=\frac{1}{4}\pi$, where all the symbol terms are assembled on the left-hand side, and the probability distribution for the symbol combination, $a+2b-c$, with its center at $\frac{1}{4}\pi$, will be given by a slight generalization of (13). If many probable relationships having the same symbol combination, $a+2b-c$, are obtained from a number of reflexions, and if from each of the relationships the probability distribution for this symbol combination is given with its centre at the phase value which appears in the corresponding relationship, the overall probability distribution for the symbol combination, $a+2b-c$, may be obtained by multiplying these individual distributions. The most probable value for the symbol combination, $a+2b-c$, will be given by the position of the maximum peak of this probability distribution and the reliability of this most probable value will be evaluated by the corresponding variance calculated by use of (16). Relationships among various unknown symbols could be found in the way as was shown in the above example.

Conclusion

For centrosymmetric crystals, the reliability of the relationships among the symbols, which have been found in the phase determination procedure, may be evaluated on the basis of the formulas (5) and (7). For non-centrosymmetric crystals, the corresponding reliability can be evaluated on the basis of the probability distribution functions (13) and (15), which represent the distributions for the linear combination of symbols. The treatments introduced here are expected to be useful in an automatic calculation of the symbolic addition procedure.

The authors would like to express their gratitude to Professor S. Naya and Professor K. Nakatsu of Kwansei Gakuin University for helpful discussions.

References

COCHRAN, W. (1955) *Acta Cryst.* **8**, 473.
COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1.
HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem.* I. *The Centrosymmetric Crystal.* A.C.A. Monograph No. 3. Pittsburg: Polycrystal Book Service.
KARLE, I. L. & KARLE, J. (1966). *Acta Cryst.* **21**, 849.
USPENSKY, J. V. (1937). *Introduction to Mathematical Probability,* New York: McGraw Hill.
WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 61.