is equal to

$$\lambda_A^{\min} = \sum_r w_r W_r^2 , \quad (11)$$

where the W axis coincides with the normal of the 'best' plane. Since $U_r^2 + V_r^2 + W_r^2 = R_r^2$ is constant for every orientation of the Cartesian system, the relation

$$\mathbf{I} + \mathbf{A} = \mathbf{E} \sum_r w_r R_r^2 \quad (12)$$

is valid, where **I** denotes the inertial tensor and **E** the unit matrix. From (12) it can easily be shown that **I** and **A** are diagonalized by the same transformation, thus we have the analogous relation for the eigenvalues

$$\lambda_I^i + \lambda_A^i = \sum_r w_r R_r^2 , \quad i = 1, 2, 3 . \quad (13)$$

With respect to the procedure described in this paper we first note that equation (4) is equivalent to searching for the maximum possible moment of inertia about the normal of the plane. Thus the normal of the 'best' plane is a principal inertial axis with moment (eigenvalue of the inertial-tensor matrix **I**)

$$\lambda_I^{\max} = \sum w_r (U_r^2 + V_r^2) . \quad (14)$$

However, in the first instance we do not look for this eigenvalue but rather for the corresponding eigenvector, the principal axis of **I**. Then the eigenvalue $\lambda_I^{\max}$ (and hence $\lambda_A^{\min}$) can easily be calculated from equations (3) and (14). (The other two eigenvectors – and eigenvalues – are not determined since they are not relevant.) Thus both approaches give the same result. In principle, the solution described here is of a simpler type; in practice, the differences in computation do not matter when using electronic computers.

Note added in proof: – The program has also been found useful for constructing crystal drawings when used in the following way: The plane of projection has to be defined by three points. The weights of these points must be non-zero, whereas the weights of all other points (atoms) in the unit cell are set to zero. Then the positions of the atoms are given in the final output with in-plane coordinates U and V, and W coordinates normal to the plane. Since the plane of projection is defined by three points in the unit cell, it may not be a lattice plane. In this regard the program is different from that described by Minor & Dyson (1970); see also Buerger (1965).

I am indebted to the Deutsche Forschungsgemeinschaft for financial support.

References

BLOW, D. M. (1960). *Acta Cryst.* **13**, 168.
BUERGER, M. J. (1965). *Tschermaks Min. Petr. Mitteilungen*, **10**, 595.
HAMILTON, W. C. (1961). *Acta Cryst.* **14**, 185.
MINOR, J. C. & DYSON, D. J. (1970). *J. Appl. Cryst.* **3**, 185.
SCHERINGER, C. (1963). *Acta Cryst.* **16**, 546.
SCHOMAKER, V. WASER, J., MARSH, R. E. & BERGMAN, G. (1959). *Acta Cryst.* **12**, 600.

Acta Cryst. (1971). B**27**, 1472

A procedure for representing arbitrary phase probability distributions in a simplified form.

By WAYNE A. HENDRICKSON,* *Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, D.C.* 20390, *U.S.A.*

(*Received* 2 *November* 1970)

This note presents a least-squares method for fitting the phase probability distributions obtained in protein crystallography by the function $P(\alpha) = \exp(K + A\cos\alpha + B\sin\alpha + C\cos 2\alpha + D\sin 2\alpha)$. The method has been tested with data from crystals of lamprey hemoglobin.

An alternative, and algebraically simplified, representation for the phase probability distributions used in protein crystallography has recently been described (Hendrickson & Lattman, 1970). It adds generality to the treatment of various types of phase information, affords computational advantages over the conventional functional forms and simplifies the combination of phase information from independent sources. It proved to be a convenient and useful aid in the structure analysis of lamprey hemoglobin (Hendrickson & Love, 1971). Unfortunately, the new representation required a reformulation of the error model for the isomorphous replacement method. This revision has been validated by experiment, but it nonetheless renders the new representation incommensurate with the formulations from other error models. Thus, unless computations are begun *de novo*, the advantages of the simplified form are lost to the structure analyses of the many proteins for which phase probability distributions have been computed by other error models. The analysis of such structures might benefit if one could cast the distributions at hand in the alternative representation. In particular, this would facilitate the inclusion of additional phase information, such as from a partial structure or direct methods, in the refinement of atomic models. The close similarity of phase probability curves computed by the usual isomorphous replacement error models with those calculated by the new procedure (Hendrickson & Lattman, 1970) suggests that a good fit by the simplified representation should be possible.

The problem, then, is to find the values of the parameters in the simplified representation,

$$P_c(\alpha) = \exp(K + A\cos\alpha + B\sin\alpha + C\cos 2\alpha + D\sin 2\alpha) , \quad (1)$$

which provide a best fit to an arbitrary 'observed' phase probability distribution, $P_o(\alpha)$. A least-squares minimization of the direct discrepancy between $P_c(\alpha)$ and $P_o(\alpha)$ leads to a set of non-linear normal equations which must be solved by iteration. However, logarithms of the probabi-

* A Resident Postdoctoral Research Associate of the National Research Council.

Table 1. *Measures of the fit of the simplified representation to* 300 *Blow–Crick phase probability distributions from lamprey hemoglobin*

Weighting function and statistic	Relative residual $\frac{\int_0^{2\pi} \lvert P_c(\alpha) - P_o(\alpha)\rvert d\alpha}{\int_0^{2\pi} P_o(\alpha) d\alpha}$	Discrepancy in figures-of-merit $\lvert m^c - m^o\rvert$	Discrepancy in centroid phases $\lvert \alpha^c{}_B - \alpha^o{}_B\rvert$
$w(\alpha)=1$			
Average	0·123	0·0135	1·40°
Worst value	1·01	0·168	28·2
$w(\alpha)=P_o(\alpha)$			
Average	0·036	0·0012	0·09
Worst value	0·25	0·015	5·3
$w(\alpha)=P_o^2(\alpha)$			
Average	0·031	0·0096	0·66
Worst value	0·35	0·301	22·9

lity functions can be used instead to effect a simple solution. Thereby the function to be minimized is

$$E=\int_0^{2\pi} w(\alpha)\,[\ln P_c(\alpha)-\ln P_o(\alpha)]^2\,d\alpha \qquad (2)$$

where weighting is provided by $w(\alpha)$. While the use of logarithms fails to guarantee a 'best' fit to $P_o(\alpha)$, proper weighting can assure an entirely adequate fit as demonstrated experimentally below. Since both $P_c(\alpha)$ and $P_o(\alpha)$ are of exponential form, their logarithms are given directly by the exponents. The minimum of E is found by the usual condition that partial derivatives of E with respect to each of the parameters must simultaneously vanish. The resulting normal equations are given in matrix form as

$$\mathbf{M}\cdot\mathbf{p}=\mathbf{d}\,. \qquad (3)$$

The elements of the normal matrix, **M**, are

$$m_{ij}=\int_0^{2\pi} w(\alpha)c_ic_j\,d\alpha$$

and vector **d** has elements

$$d_i=\int_0^{2\pi} w(\alpha)c_i\,\ln P_o(\alpha)\,d\alpha$$

where the c_i are defined by $\mathbf{c}=[1, \cos\alpha, \sin\alpha, \cos 2\alpha, \sin 2\alpha]$. The desired vector of parameters, $\mathbf{p}=[K, A, B, C, D]$, can easily be found by inversion of (3) once the weighting function $w(\alpha)$ has been defined. If (2) is left unweighted, *i.e.* $w(\alpha)=1$, only diagonal elements of the normal matrix are non-zero and the solution is particularly simple:

$$p_1=K=\frac{1}{2\pi}\int_0^{2\pi}\ln P_o(\alpha)\,d\alpha\,,$$

$$p_i=\frac{1}{\pi}\int_0^{2\pi} c_i\,\ln P_o(\alpha)\,d\alpha\,,\quad i=2,3,4,5\,. \qquad (4)$$

It is somewhat difficult to ascertain the proper least-squares weights, $w(\alpha)$. In principle the appropriate weights are given by

$$w(\alpha)=\frac{1}{\sigma^2[\ln P_o(\alpha)]}=\frac{P_o^2(\alpha)}{\sigma^2[P_o(\alpha)]}\,. \qquad (5)$$

However, the evaluation of $\sigma(P_o)$ is complicated and no useful, general simplifications are readily apparent. Special cases where $\sigma(P_o)$ is approximately proportional to P_o and others where it is more nearly constant suggest weights of $w(\alpha)=1$ or $w(\alpha)=P_o^2(\alpha)$. Alternatively, on an intuitive basis, it seems that since it is the peaks of the probability distributions which are crucial for phase determination, $w(\alpha)=P_o(\alpha)$ might be an effective weighting function.

The above procedures for evaluating parameters for the alternative representation of phase probability distributions have been tested with data from 300 of the reflections used in the structure analysis of lamprey hemoglobin (Hendrickson & Love, 1971). 'Observed' phase probability distributions, $P_o(\alpha)$, were computed by the method of Blow & Crick (1959) using data from three isomorphous derivatives. These distributions were fitted with both weighted and unweighted equations. The integrals were evaluated numerically with a 5° interval. Some measures of the goodness of these fits are presented in Table 1. Although the fit by the unweighted equations (4) was satisfactory for nearly all of the distributions, if the ratio of peak to valley probability exceeded 10^{10} quite poor fits sometimes occurred. In the case of weighting by $P_o^2(\alpha)$, nearly all fits were satisfactory but some were not and minor maxima tended to be over-emphasized. However, when $P_o(\alpha)$ was used as a weighting function, all of the distributions were fit very well.

It should be noted that for centrosymmetric reflections exact parameters can be obtained from the best phase, α_B, and the figure-of-merit, m, by inversion of equations (22) of Hendrickson & Lattman (1970). Thus,

$$A=\cos\alpha_B\,\tanh^{-1}m$$

and

$$B=\sin\alpha_B\,\tanh^{-1}m \qquad (6)$$

for pure real or pure imaginary structure factors.

Discussions with Drs Jerome Karle and Eaton Lattman contributed to this work and are greatly appreciated. Discussion of the theoretical basis for weighting was suggested by a referee.

References

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**. 794.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst*, B**26**, 136.

HENDRICKSON, W. A. & LOVE, W. E. (1971). In preparation.