An exact expression for the joint p.d.f. which, at least for small values of n, can often be accurately evaluated, can be obtained in the important case in which all the components of the random vector $\mathbf{x}$ may be non-zero only in a bounded range, say $-x_M < x_i < +x_M$, $i = 1, \ldots, n$ (Weiss, Shmueli, Kiefer & Wilson, 1985; Shmueli & Weiss, 1985). One can then expand the joint p.d.f. $p(\mathbf{x})$ in a Fourier series

$$p(\mathbf{x}) = (2x_M)^{-n} \sum_{\mathbf{u}} C_{\mathbf{u}} \exp(-\pi i \mathbf{u}^T \mathbf{x}/x_M) \qquad (A3)$$

and write the Fourier coefficients, in the conventional manner, as

$$C_{\mathbf{u}} = \int_{-x_M}^{x_M} \ldots \int_{-x_M}^{x_M} p(\mathbf{x}) \exp(\pi i \mathbf{u}^T \mathbf{x}/x_M)\, \mathrm{d}^n\mathbf{x}. \qquad (A4)$$

Since, however, the random variables, x_i, are confined to the $[-x_M, +x_M]$ range, the probability of finding any of them outside this range is necessarily equal to zero. We can thus replace, with no loss of generality, the limits of integration in ($A4$) by $\pm\infty$ and obtain for the Fourier coefficients an expression analogous to ($A1$). In fact, the Fourier coefficients, $C_{\mathbf{u}}$, are then just the values of the characteristic function at the points: $(\omega_1, \ldots, \omega_n) = (\pi u_1/x_M, \ldots, \pi u_n/x_M)$. The practical significance of ($A3$) is now conditioned by our ability to evaluate the characteristic function, and by the convergence properties of the resulting Fourier summation.

In some applications one has to deal with random vectors in which not all the components are independent. If, for example, $\mathbf{x} = (x_1, x_2, x_3)$ and x_3 depends on x_1 and x_2, we shall still have a triple Fourier series to compute, but the integration leading to the characteristic function (or the Fourier coefficients) involves only the independent variables. For examples of such situations see Shmueli & Weiss (1985) and the derivation in the text.

References

Abramowitz, M. & Stegun, I. (1972). *Handbook of Mathematical Functions*. New York: Dover.
Bertaut, E. F. (1955). *Acta Cryst.* **8**, 823-832.
Cochran, W. & Woolfson, M. M. (1955). *Acta Cryst.* **8**, 1-12.
Cramér, H. (1951). *Mathematical Methods of Statistics*. Princeton Univ. Press.
Giacovazzo, C. (1977). *Acta Cryst.* A**33**, 50-54.
Giacovazzo, C. (1980). *Direct Methods in Crystallography*. New York: Academic Press.
Gradshteyn, I. S. & Ryzhik, I. M. (1980). *Tables of Integrals, Series and Products*. New York: Academic Press.
Harker, D. & Kasper, J. S. (1948). *Acta Cryst.* **1**, 70-75.
Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* ACA Monograph No. 3. Pittsburgh: Polycrystal Book Service.
Hauptman, H. & Karle, J. (1959). *Acta Cryst.* **12**, 846-850.
International Tables for X-ray Crystallography (1952). Vol. I, *Symmetry Groups*, edited by N. F. M. Henry & K. Lonsdale. Birmingham: Kynoch Press.
Karle, J. (1972). *Acta Cryst.* B**28**, 3362-3369.
Karle, J. & Gilardi, R. D. (1973). *Acta Cryst.* A**29**, 401-407.
Kendall, M. G. & Stuart, A. (1969). *The Advanced Theory of Statistics*, Vol. 1, 3rd ed. London: Charles Griffin.
Klug, A. (1958). *Acta Cryst.* **11**, 515-543.
Shmueli, U. (1982). *Acta Cryst.* A**38**, 362-371.
Shmueli, U. & Weiss, G. H. (1985). *Acta Cryst.* A**41**, 401-408.
Shmueli, U., Weiss, G. H., Kiefer, J. E. & Wilson, A. J. C. (1984). *Acta Cryst.* A**40**, 651-660.
Weiss, G. H., Shmueli, U., Kiefer, J. E. & Wilson, A. J. C. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. Wilson, pp. 23-42. Guilderland, NY: Adenine Press.
Woolfson, M. M. (1954). *Acta Cryst.* **7**, 61-64.

Acta Cryst. (1986). A**42**, 246-253

Triplet Phase Invariants from Single Isomorphous Replacement or One-Wavelength Anomalous Dispersion Data, Given Heavy-Atom Information

By Jerome Karle

Laboratory for the Structure of Matter, Naval Research Laboratory, Washington, DC 20375-5000, *USA*

(*Received* 17 *September* 1985; *accepted* 14 *January* 1986)

Abstract

Certain general algebraic formulas for computing triplet phase invariants become accessible when structural information is available concerning the replacement atoms in isomorphous replacement or the predominant type of anomalously scattering atoms in one-wavelength anomalous dispersion experiments. The formulas of interest are presented and subjected to a number of test calculations to obtain insight into their accuracy and to determine the effects of errors in the data. The formulas are simple to calculate and some possible strategies for their use are discussed.

Introduction

On the basis of certain mathematical and physical considerations that pertain to isomorphous replacement or anomalous dispersion experiments, rules

0108-7673/86/040246-08$01.50

(Karle, 1983, 1984*a*, *b*) and algebraic formulas (Karle, 1984*c*, 1985*a*) were derived for the evaluation of triplet phase invariants. Among these was a general formula, applicable to both types of experimental data and not yet evaluated, that requires information concerning the structure factor magnitudes for the structure of the heavy atoms (or anomalously scattering atoms). This general formula is investigated here. Further accuracy would accrue if the phases for the heavy-atom structure were also known. This latter information is not mandatory, however, since the heavy-atom phases enter the formula as triplet phase invariants which can often be set equal to zero to good approximation. In the case of anomalous dispersion the formula is applicable when there is one type or one predominant type of anomalous scatterer. In the case of isomorphous replacement, the formula permits the calculation of the cosines of phase invariants in the range from −1 to 1 and for anomalous dispersion it permits the calculation of the sines of phase invariants in the range −1 to 1.

This investigation may be regarded as a further exploration of the mathematical tools available for application to the analysis of macromolecular structure. The particular role that triplet phase invariants may play remains to be determined. When the heavy-atom structure is known, it is possible in single isomorphous replacement to obtain numerous initial phase values that could possibly be extended and refined by use of the triplet phase invariants.

Diffraction experiments concerning isomorphous substitution of macromolecules are normally accompanied by anomalous dispersion effects. Because both types of data are readily measurable, it can be expected that data of sufficient accuracy to permit the application of both techniques will often be attainable, as the history of macromolecular structure determination has already shown (*e.g.* Adman, Sieker & Jensen, 1973). The combining of both techniques would be the optimal way to handle the data. A way to do this with the use of exact algebraic equations has already been discussed (Karle, 1984*c*). Here single isomorphous replacement and one-wavelength anomalous dispersion are treated individually as if the other did not exist.

A way to obtain essentially unique values for phase differences from one-wavelength anomalous dispersion data, *i.e.* the two intensities measured at $\mathbf{h}$ and $-\mathbf{h}$, has been described (Karle, 1985*b*). As has been previously noted (Karle, 1985*b*), the possible existence of such a calculation was indicated by the uniqueness of values for the triplet phase invariants derived by Hauptman (1982*b*), Giacovazzo (1983) and subsequently by Karle (1984*a*). Knowledge of the structure of the predominant type of anomalous scatterers is not required to obtain the phase differences but, if it is known, phase values for the structure of the macromolecule can be obtained from the values of the phase differences. Evaluations of the triplet phase invariants of the type examined here may be useful for refining the phase values for macromolecules.

Table 1. *Quantities involved in the applications of* (1) *to single isomorphous replacement and one-wavelength anomalous dispersion data*

The quantities ${}_m\mathscr{F}_{1,\mathbf{h}}$, ${}_m\mathscr{F}_{2,\mathbf{h}}$ and ${}_m\mathscr{F}_{3,\mathbf{h}}$ are defined by the corresponding entries in columns 2, 3, 4, respectively.

Case m	${}_m\mathscr{F}_{1,\mathbf{h}}$	${}_m\mathscr{F}_{2,\mathbf{h}}$	${}_m\mathscr{F}_{3,\mathbf{h}}$
i	$F_{\mathbf{h}PH}$	$F_{\mathbf{h}P}$	$F_{\mathbf{h}H}$
1	$F_{\lambda\mathbf{h}}$	$F^*_{\lambda\bar{\mathbf{h}}}$†	$F^a_{\lambda\mathbf{h}} - F^{a*}_{\lambda\bar{\mathbf{h}}}$†

† The asterisk denotes complex conjugate.

Theory

A general formula has been derived (Karle, 1984*a*, equation 31) of the form

$$\cos({}_m\varphi_{2,\mathbf{h}} + {}_m\varphi_{2,\mathbf{k}} + {}_m\varphi_{2,(\bar{\mathbf{h}}+\bar{\mathbf{k}})} - {}_m\varphi_{3,\mathbf{h}} - {}_m\varphi_{3,\mathbf{k}} - {}_m\varphi_{3,(\bar{\mathbf{h}}+\bar{\mathbf{k}})})$$
$$\simeq T/|{}_m\mathscr{F}_{3,\mathbf{h}}\,{}_m\mathscr{F}_{3,\mathbf{k}}\,{}_m\mathscr{F}_{3,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}|, \quad (1)$$

where

$$T = (|{}_m\mathscr{F}_{1,\mathbf{h}}| - |{}_m\mathscr{F}_{2,\mathbf{h}}|)(|{}_m\mathscr{F}_{1,\mathbf{k}}| - |{}_m\mathscr{F}_{2,\mathbf{k}}|)$$
$$\times (|{}_m\mathscr{F}_{1,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}| - |{}_m\mathscr{F}_{2,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}|). \quad (2)$$

There are numerous definitions of the ${}_m\mathscr{F}_{i,\mathbf{h}}$ ($i = 1, 2, 3$), listed in a recent publication (Karle, 1985*a*). When $m = i$, the $\mathscr{F}$ concern quantities associated with isomorphous replacement and when $m = 1$, the $\mathscr{F}$ concern quantities associated with anomalous dispersion, as shown in Table 1.

For isomorphous replacement,

$$F_{\mathbf{h}PH} - F_{\mathbf{h}P} = F_{\mathbf{h}H}, \quad (3)$$

where $F_{\mathbf{h}PH}$ is the structure factor for a macromolecule substituted with heavy atoms, $F_{\mathbf{h}P}$ is the structure factor for the original macromolecule and $F_{\mathbf{h}H}$ is the structure factor for the structure formed by the heavy atoms alone.

For anomalous dispersion,

$$F_{\lambda\mathbf{h}} - F^*_{\lambda\bar{\mathbf{h}}} = F^a_{\lambda\mathbf{h}} - F^{a*}_{\lambda\bar{\mathbf{h}}}, \quad (4)$$

where $F_{\lambda\mathbf{h}}$ is the structure factor for a macromolecule containing anomalous scatterers at some incident wavelength, λ, and $F^a_{\lambda\mathbf{h}}$ is the structure factor for the structure formed by the anomalous scatterers. For the case of one type of anomalous scatterer present (Karle, 1984*a*, equation 10 when $j = 2$)

$$F^a_{\lambda\mathbf{h}} - F^{a*}_{\lambda\bar{\mathbf{h}}} = 2i(f''/f^n_{2,\mathbf{h}})F^n_{2,\mathbf{h}}, \quad (5)$$

where f'' is the imaginary correction to the normal atomic scattering factor, $f^n_{2,\mathbf{h}}$, of the anomalous scatterers and $F^n_{2,\mathbf{h}}$ is the structure factor for the anomalous scatterers, scattering as if there were no anomalous effects present.

With the definitions in Table 1 and use of (5), it is possible to calculate the right side of (1) and thereby evaluate the cosine function. The $|{}_m\mathscr{F}_{1,\mathbf{h}}|$ and $|{}_m\mathscr{F}_{2,\mathbf{h}}|$ are obtained from experimental measurement and the $|{}_m\mathscr{F}_{3,\mathbf{h}}|$ are obtained from derived information, for example, the heavy-atom structure. The angles on the left side of (1) are associated with the corresponding $|{}_m\mathscr{F}_{2,\mathbf{h}}|$ and $|{}_m\mathscr{F}_{3,\mathbf{h}}|$. It had not been brought out in the original derivation of (1) that, with the approximations used, the same function on the right side of (1) would be associated with any of the eight cosine functions obtained by substituting any or all of the ${}_m\varphi_{2,\mathbf{h}}$, ${}_m\varphi_{2,\mathbf{k}}$ and ${}_m\varphi_{2,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}$ with the corresponding ${}_m\varphi_{1,\mathbf{h}}$, ${}_m\varphi_{1,\mathbf{k}}$ and ${}_m\varphi_{1,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}$. In other words, the formula is insensitive to the distinction between the values of ${}_m\varphi_{1,\mathbf{h}}$ and ${}_m\varphi_{2,\mathbf{h}}$. In recognition of this and for the purpose of facilitating quantitative comparison between the cosine function and the right side of (1), the equation is rewritten

$$\cos\left({}_m\bar{\varphi}_{\mathbf{h}}+{}_m\bar{\varphi}_{\mathbf{k}}+{}_m\bar{\varphi}_{(\bar{\mathbf{h}}+\bar{\mathbf{k}})}-{}_{\mathrm{m}}\varphi_{3,\mathbf{h}}-{}_m\varphi_{3,\mathbf{k}}-{}_m\varphi_{3,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}\right)$$
$$\simeq T/|{}_m\mathscr{F}_{3,\mathbf{h}}{}_m\mathscr{F}_{3,\mathbf{k}}{}_m\mathscr{F}_{3,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}|, \qquad (6)$$

where

$${}_m\bar{\varphi}_{\mathbf{h}}=0\cdot5({}_m\varphi_{1,\mathbf{h}}+{}_m\varphi_{2,\mathbf{h}}). \qquad (7)$$

This becomes for isomorphous replacement ($m=i$)

$${}_i\bar{\varphi}_{\mathbf{h}}=0\cdot5(\varphi_{\mathbf{h}PH}+\varphi_{\mathbf{h}P}) \qquad (8)$$

and for anomalous dispersion ($m=1$)

$${}_1\bar{\varphi}_{\mathbf{h}}=0\cdot5(\varphi_{\lambda\mathbf{h}}-\varphi_{\lambda\bar{\mathbf{h}}}). \qquad (9)$$

The argument of the cosine function on the left side of (6) contains a triplet phase invariant that is composed of average phases, as defined in (7), and a second triplet phase invariant composed of phases associated with the heavy-atom structure or anomalous scatterers. If the heavy-atom structure is known, the value of this second triplet phase invariant can be calculated and from knowledge of the value of the right side of (6), the value of the triplet phase invariant of interest ${}_m\bar{\varphi}_{\mathbf{h}}+{}_m\bar{\varphi}_{\mathbf{k}}+{}_m\bar{\varphi}_{\bar{\mathbf{h}}+\bar{\mathbf{k}}}$ can be evaluated, usually with a twofold ambiguity.

Under certain circumstances, the exact formulas for isomorphous replacement

$$\cos(\varphi_{\mathbf{h}P}-\varphi_{\mathbf{h}H})=(|F_{\mathbf{h}PH}|^2-|F_{\mathbf{h}P}|^2-|F_{\mathbf{h}H}|^2)/2|F_{\mathbf{h}P}||F_{\mathbf{h}H}| \qquad (10)$$

and

$$\cos(\varphi_{\mathbf{h}PH}-\varphi_{\mathbf{h}H})=(|F_{\mathbf{h}PH}|^2-|F_{\mathbf{h}P}|^2+|F_{\mathbf{h}H}|^2)/2|F_{\mathbf{h}PH}||F_{\mathbf{h}H}| \qquad (11)$$

may be useful, namely, when the right sides of (10) and (11) have values in the vicinity of ±1. In such cases, the $\varphi_{\mathbf{h}P}$ or $\varphi_{\mathbf{H}PH}$ are approximately equal to or π away from $\varphi_{\mathbf{h}H}$.

Test calculations

Test calculations were performed on exact data and also on data into which errors were introduced. The data were computed at 2·5 Å resolution for Cu $K\alpha$ radiation from the coordinates for cytochrome c550.PtCl_4^{2-} from *Paracoccus denitrificans* (Timkovich & Dickerson, 1976). For the isomorphous replacement tests, the Pt atoms were regarded as comprising the heavy-atom structure. For the anomalous dispersion tests, the structure factors were computed in two ways. One calculation introduced anomalous effects from the Pt atoms alone and the second included anomalous effects from the Pt, Fe, S and Cl atoms. The first calculation models the case when there would be only one type of anomalous scatterer. This calculation not only represents an important experimental case, but also provides a basis of comparison for determining the effect on the errors of including all four types of anomalous scatterers in the data while treating the data as if the Pt atoms were the one predominant type of anomalous scatterer.

Isomorphous replacement

Calculations of values for triplet phase invariants in an isomorphous replacement experiment by use of (6) are presented in Table 2. Appropriate definitions for the use of (6) are given in Table 1. The calculations were based on 400 independent reflections except for row 6 in which 800 independent reflections were selected from among the 3252 acentric ones available from 2·5 Å data. They were chosen on the basis of the largest values for $||F_{PH}|-|F_P||$. The first column shows the number of invariants that were formed. When a cut-off value different from zero is given in the second column, it means that reflections were omitted from the calculations when the values for their corresponding $|F_P|$ and $|F_{PH}|$ were less than the cut-off value. For those data into which random errors were introduced (rows 3-7), an acceptance criterion was used for handling those instances in which the right side of (6) exceeded 1·0. If the right side exceeded 1·5, the calculation was rejected, otherwise the value was set back to 1·0.

In order to obtain an estimate of the accuracy of the calculations obtainable from (6), two types of error were computed, type I and type II. Type I is an average magnitude of error

$$\langle|\Phi-\cos^{-1}[\text{right side of (6)}]|\rangle, \qquad (12)$$

where Φ represents the sum of angles (of known value in the test problems). The arc cosine function is generally twofold ambiguous and may also require a shift of $\pm2\pi$ to find the value closest to that of the

Table 2. *The evaluation of average triplet phase invariants in isomorphous replacement by use of* (6) *from* 2·5 Å *data for cytochrome* $c550.\mathrm{PtCl}_4^{2-}$

The calculations were based on 400 independent reflections except for row 6 in which 800 independent reflections were chosen from among the 3252 acentric ones available on the basis of the largest values for $||F_{PH}|-|F_P||$. The last row was added to show the effect of including centric reflections in the data set and comparing the result from (6) to the correct value for a single triplet phase invariant, the one having the largest triple product of associated structure factor magnitudes, instead of the average of eight of them. For those data into which random errors were introduced (rows 3–7), an acceptance criterion was used for handling those instances in which the right side of (6) exceeded 1·0. If the right-side value exceeded 1·5, the calculation was rejected, otherwise the value was set back to 1·0.

Number of invariants	Cut-off for $\lvert F_P\rvert, \lvert F_{PH}\rvert$	Av. magnitude of error for $\lvert F_P\rvert$ $\lvert F_{PH}\rvert, \lvert F_H\rvert$ (%)	Av. magnitude of error type I (rad)	Av. magnitude of error type II
18 933	0	0	0·42	0·30
17 110	150	0	0·38	0·27
15 940	0	5	0·59	0·37
13 125	150	5	0·57	0·34
6453	0	10	0·81	0·49
55 880	0	10	0·82	0·54
5524	150	10	0·76	0·46
18 007	0	0	0·19	0·08

known Φ. In computing (12), the latter value of the arc cosine was used. The purpose of the computation was solely to present an estimate of the accuracy in radians. Type II is an average magnitude of error

$$\langle|\cos\Phi - \text{right side of (6)}|\rangle. \qquad (13)$$

Type I and type II errors appropriate to the calculations in Table 2 are shown in columns 4 and 5, respectively.

It is seen from Table 2 that large numbers of invariants can be computed with no apparently serious increase in the average error. This is shown by the calculations in rows 5 and 6 which differ in the use of 400 and 800 independent data, respectively, for the computation of the values of the invariants. It would appear that the number of independent data used could be safely increased and significant increases in the number of invariants computed would also accrue from the use of one- and two-dimensional data. Having a cut-off value for $|F_P|$ and $|F_{PH}|$ does not appear to have a great effect on the average error of the calculations, but it may prevent the occurrence of large errors from structure factors of small magnitude. Significant benefits would be obtained from keeping the average error in the structure factor magnitudes in the vicinity of 5% rather than 10%. The calculation of the 55 880 invariants and associated errors in row 6 of Table 2 required about 1 min and 40 s on the Cray X-MP/12 with the use of a rather simple program.

Table 3. *Average magnitude of discrepancy in isomorphous replacement between sets of eight triplet phase invariants and their averages calculated from* 2·5 Å *data for cytochrome* $c550.\mathrm{PtCl}_4^{2-}$

In any set, the eight invariants are formed by adding $\varphi_{\mathbf{h}P}$ or $\varphi_{\mathbf{h}PH}$, $\varphi_{\mathbf{k}P}$ or $\varphi_{\mathbf{k}PH}$ and $\varphi_{(\bar{\mathbf{h}}+\bar{\mathbf{k}})P}$ or $\varphi_{(\bar{\mathbf{h}}+\bar{\mathbf{k}})PH}$. The invariants in the first row were formed from 400 and those in the second row were formed from 800 independent reflections selected from the 3252 acentric reflections on the basis of the largest values for the $||F_{PH}|-|F_P||$. Since (6) is interpreted as providing the values of the average triplet invariants, this computation gives some insight into how well an average represents any member of the set of eight.

Number of invariants	Av. magnitude of error (rad)
18 933	0·34
99 367	0·37

As implied by the left side of (6), only the values of average triplet invariants, ${}_m\bar{\varphi}_{\mathbf{h}} + {}_m\bar{\varphi}_{\mathbf{k}} + {}_m\bar{\varphi}_{\mathbf{h}+\mathbf{k}}$, are expected to be the useful and meaningful context in which the invariants will be obtained from theory with the use of experimental data. The average invariants are the averages of eight different invariants formed by adding $\varphi_{\mathbf{h}P}$ or $\varphi_{\mathbf{h}PH}$, $\varphi_{\mathbf{k}P}$ or $\varphi_{\mathbf{k}PH}$ and $\varphi_{(\bar{\mathbf{h}}+\bar{\mathbf{k}})P}$ or $\varphi_{(\bar{\mathbf{h}}+\bar{\mathbf{k}})PH}$. It is of interest to gain some insight into how well the value of an average invariant represents the value of any one member. Calculations of the average magnitude of error between the value of the average invariants and those of their eight members are shown in Table 3. The invariants in the first row were formed from 400 and those in the second row from 800 independent reflections selected from the 3252 acentric reflections available on the basis of the largest values for the $||F_{PH}|-|F_P||$. The first column of Table 3 gives the number of average invariants and the second gives the average magnitude of error in radians for all of them.

It would appear that a large number of initial phase values may be obtained from (10) and (11) with acceptable accuracy. This is indicated by the results presented in Table 4. The reflections were ordered according to the largest values for the $||F_{PH}|-|F_P||$ and the number of acentric reflections available was 3252. The samples consisting of 400 and 800 reflections were ordered subsets of the latter. The second column indicates limits for the magnitude of the right sides of (10), (11). The limit of 0·9397 implies that as much as a calculated 20° difference between the two angles would be acceptable and 0·7660 implies that as much as a 40° difference would be acceptable. When there are errors in the data, as shown in the third column, the actual differences that are accepted can exceed 20 or 40°. Columns 4–7 show the numbers of phases, $\varphi_{\mathbf{h}P}$ and $\varphi_{\mathbf{h}PH}$, that were set equal to the corresponding $\varphi_{\mathbf{h}H}$ and the average magnitude of error. It is seen that little is gained in terms of accuracy by restricting the number of reflections and, evidently, fewer phases are evaluated under such circumstances.

Table 4. *The evaluation of individual phases,* $\varphi_{\mathbf{h}P}$ *and* $\varphi_{\mathbf{h}PH}$, *in isomorphous replacement by use of* (10), (11) *and* 2·5 Å *data for cytochrome* c550.PtCl$_4^{2-}$

Those values of the phases, $\varphi_{\mathbf{h}P}$ and $\varphi_{\mathbf{h}PH}$, that were determined to be sufficiently close to the values of the corresponding $\varphi_{\mathbf{h}H}$ for the heavy-atom structure by use of (10), (11) were set equal to the $\varphi_{\mathbf{h}H}$. This was done when the values of the right sides of (10), (11) were equal to or larger than the values in column 2, otherwise the phases were not evaluated. The subsets of 400 and 800 reflections were selected on the basis of the largest values for the differences, $||F_{PH}|-|F_P||$. Average percentage random errors introduced into the data are indicated in column 3.

Number of reflections	Limit for |right sides| of (10), (11)	Av. magnitude of error for $\|F_P\|, \|F_{PH}\|, \|F_H\|$ (%)	Number of φ_P accepted	Av. magnitude of error for φ_P (rad)	Number of φ_{PH} accepted	Av. magnitude of error for φ_{PH} (rad)
3252	0·9387	0	721	0·17	843	0·17
3252	0·7660	0	1463	0·35	1629	0·34
3252	0·9397	5	856	0·40	932	0·36
3252	0·7660	5	1403	0·47	1591	0·43
3252	0·9397	10	1094	0·57	1152	0·51
3252	0·7660	10	1522	0·59	1698	0·54
800	0·9397	10	571	0·51	598	0·45
800	0·7660	10	665	0·51	734	0·46
400	0·9397	10	333	0·46	348	0·41
400	0·7660	10	364	0·47	388	0·41

The one- and two-dimensional data in the space group of cytochrome c550, $P2_12_12_1$, are centric. Application of (10), (11) to such data should give many errorless phase evaluations.

A possibly useful strategy in single isomorphous replacement would be to use (10) and (11) to obtain a large set of initial phase values and then apply the evaluations of the triplet phase invariants from (6) or from probabilistic methods (Fortier, Moore & Fraser, 1985) an extension of an analysis of Hauptman (1982*a*) to refine further and extend the phases by use of, for example, a least-squares procedure. The work of Fortier *et al.* is relevant to the analysis in this paper because it presents a formula that can evaluate the cosines of triplet phase invariants having any value in the range −1 to +1.

It is possible to compare only the results in the last row of Table 2 of this paper directly with those in Table 2 of Fortier *et al.* The other entries in Table 2 are based on average invariants that are averages of eight of them, as noted previously. In addition, several items in Table 2 were computed from data containing random errors. The entries in Table 2 of Fortier *et al.* concern individual invariants computed from exact data. Comparison shows that the results of the last row of Table 2 of this paper are of an accuracy similar to the best calculations given in Table 2 of Fortier *et al.* The intention of the presentation in the first seven rows of Table 2 of this paper is to give an insight into what might be expected in actual experimental circumstances. The results would be improved if centric data were included in the calculations.

Anomalous dispersion

Calculations of the values for triplet phase invariants in an anomalous dispersion experiment involving one predominant type of anomalous scatterer by use of (6) are presented in Table 5. The radiation used was Cu $K\alpha$. Appropriate definitions for the use of (6) are given in Table 1. For $m=1$, we find from (5) that

$$ {}_m\varphi_{3,\mathbf{h}} = \pi/2 + \varphi^n_{2,\mathbf{h}}, \tag{14}$$

where $\varphi^n_{2,\mathbf{h}}$ is the phase associated with $F^n_{2,\mathbf{h}}$, the structure factor in (5) for the single type of anomalous scatterers present, scattering as if there were no anomalous effects. The same mathematics holds approximately when one predominant type of anomalous scatterer is present among others. For the case of one predominant type of anomalous scatterer present, (6) becomes from use of (9) and (14)

$$\begin{aligned} &\sin\left({}_1\bar{\varphi}_{\mathbf{h}} + {}_1\bar{\varphi}_{\mathbf{k}} + {}_1\bar{\varphi}_{\bar{\mathbf{h}}+\bar{\mathbf{k}}} - \varphi^n_{2,\mathbf{h}} - \varphi^n_{2,\mathbf{k}} - \varphi_{2,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}\right) \\ &\quad \simeq -T/|{}_m\mathscr{F}_{3,\mathbf{h}}\,{}_m\mathscr{F}_{3,\mathbf{k}}\,{}_m\mathscr{F}_{3,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}|, \end{aligned} \tag{15}$$

where T is obtained from (2) and Table 1 and the denominator of the right side of (15) is obtained from taking appropriate magnitudes in (5), *e.g.*

$$ {}_1\mathscr{F}_{3,\mathbf{h}} = 2(f''/f^n_{2,\mathbf{h}})|F^n_{2,\mathbf{h}}|. \tag{16}$$

The calculations shown in Table 5 were based on 400 independent reflections except for the sixth row in which 800 independent reflections were chosen from among the 3252 centric ones available from 2·5 Å data on the basis of the largest values for $||F_{\lambda\mathbf{h}}|-|F_{\lambda\bar{\mathbf{h}}}||$. The first column shows the number of invariants that were formed. The structure factors used for the calculations in rows 1–4 were based on the Pt atoms alone as the anomalous scatterers. In all cases, the data were analyzed as if the Pt atoms were the sole predominant anomalous scatterers. For the second row, in which a cut-off value different from zero is given, reflections were omitted from the calculations when the values for their corresponding $|F_{\lambda\mathbf{h}}|$ and $|F_{\lambda\bar{\mathbf{h}}}|$ were less than the cut-off value. Errors

Table 5. *The evaluation of average triplet phase invariants in anomalous dispersion by use of* (15) *from* 2·5 Å *data for cytochrome* c550.PtCl$_4^{2-}$

The calculations were based on 400 independent reflections except for row 6 in which 800 independent reflections were chosen from among the 3252 acentric ones available from 2·5 Å data on the basis of the largest values for $||F_{\lambda\mathbf{h}}|-|F_{\lambda\bar{\mathbf{h}}}||$. The structure factors used for the calculations in rows 1-4 were based on the Pt atoms alone as the anomalous scatterers, whereas those in rows 5 and 6 were based on the Cl, S, Fe and Pt atoms as the anomalous scatterers. In all cases, the data were analyzed as if the Pt atoms were the sole predominant anomalous scatterers. The radiation used was Cu $K\alpha$. Errors were introduced for the individual differences, $||F_{\lambda\mathbf{h}}|-|F_{\lambda\bar{\mathbf{h}}}||$, by multiplying randomly by values that ranged from 0·2 to 1·8, accounting for the 40% average error listed in the last four rows of column 3. The values of $|F_{\lambda\mathbf{h}}|$ and $|F_{\lambda\bar{\mathbf{h}}}|$ were suitably adjusted to account for the changed difference. For those data into which errors were introduced (rows 3-6), an acceptance criterion was used for handling those instances in which the right side of (15) exceeded 1·0. If the right side exceeded 1·5, the calculation was rejected, otherwise the value was set back to 1·0.

Number of invariants	Cut-off for $\lvert F_{\lambda\mathbf{h}}\rvert, \lvert F_{\lambda\bar{\mathbf{h}}}\rvert$	Av. magnitude of error for $\lvert F_{\lambda\mathbf{h}}\rvert-\lvert F_{\lambda\bar{\mathbf{h}}}\rvert$ (%)	Av. magnitude of error $\lvert F_{2,\mathbf{h}}^n\rvert$ (%)	Av. magnitude of error type I (rad)	Av. magnitude of error type II
16 662	0	0	0	0·41	0·28
16 099	100	0	0	0·42	0·28
5960	0	40	3	0·80	0·47
6046	0	40	5	0·79	0·47
3141	0	40	3	0·88	0·56
41 797	0	40	3	0·83	0·55

were introduced for the individual differences, $|F_{\lambda\mathbf{h}}|-|F_{\lambda\bar{\mathbf{h}}}|$, by multiplying randomly by values that ranged from 0·2 to 1·8, accounting for the 40% average magnitude of error listed in the last four rows of column 3. Random errors of average magnitude 3 and 5% for $|F_{2,\mathbf{h}}^n|$ are listed in the last four rows of column 4. For those data into which errors were introduced (rows 3-6), an acceptance criterion was used for handling those instances in which the right side of (5) exceeded 1·0. If the right side exceeded 1·5, the calculation was rejected, otherwise the value was set back to 1·0. Comparison of row 3 with row 1 indicates that the acceptance criterion caused a dramatic drop in the number of invariants accepted. There may well be a better criterion so that more invariant evaluations could be accepted without significant loss of accuracy. It does not seem from comparison of rows 1 and 2 in columns 5 and 6 that the introduction of the cut-off had a significant effect on the resulting errors in the calculations. The cut-off value for $|F_{\lambda\mathbf{h}}|$ and $|F_{\lambda\bar{\mathbf{h}}}|$ may, however, similarly to isomorphous replacement calculations, prevent the occurrence of large errors from structure factors of small magnitude. Errors of type I and type II in columns 5 and 6 are describable by use of (12) and (13), respectively, or, equivalently, by use of

$$\langle|\Phi'-\sin^{-1}[\text{right side of (15)}]|\rangle \qquad (17)$$

and

$$\langle|\sin\Phi'-\text{right side of (15)}|\rangle, \qquad (18)$$

where Φ' represents the sum of angles on the left side of (15). The calculation of (17) was carried out in a fashion comparable to that for (12). A comparison of rows 3 and 4 in which the calculations were the same except for the different average magnitudes of error for the $|F_{2,\mathbf{h}}^n|$ indicates that even greater average errors in $|F_{2,\mathbf{h}}^n|$ could be tolerated.

It is seen from Table 5 that large numbers of invariants can be computed without reaching unacceptable average errors. This is indicated by the calculations in rows 5 and 6 which differ in the use of 400 and 800 independent reflections, respectively, for the computation of the values of the invariants. The calculations for row 5 required 75 s on the Cray X-MP/12. It would appear from the indicated accuracies and the time involved that the number of independent reflections used could be readily increased beyond 800 with a consequent further increase in the number of invariants evaluated.

The triplet phase invariants composed of average phases, ${}_1\bar{\varphi}_{\mathbf{h}}+{}_1\bar{\varphi}_{\mathbf{k}}+{}_1\bar{\varphi}_{\bar{\mathbf{h}}+\bar{\mathbf{k}}}$, that are evaluated by use of (15) are the averages of eight different invariants formed by adding $\varphi_{\lambda\mathbf{h}}$ or $-\varphi_{\lambda\bar{\mathbf{h}}}$, $\varphi_{\lambda\mathbf{k}}$ or $-\varphi_{\lambda\bar{\mathbf{k}}}$ and $\varphi_{\lambda,(\bar{\mathbf{h}}+\bar{\mathbf{k}})}$ or $-\varphi_{\lambda,(\mathbf{h}+\mathbf{k})}$. The values for the average invariants represent the eight different invariants rather accurately for cytochrome c550.PtCl$_4^{2-}$. A calculation of the average magnitude of difference between $\varphi_{\lambda\mathbf{h}}$ and $-\varphi_{\lambda\bar{\mathbf{h}}}$ (Karle, 1985*a*, Table 3, column 5, $m=1$) for 2900 reflections and Cu $K\alpha$ radiation gave 0·07 rad. In these calculations data for which $|F_{\lambda\mathbf{h}}|$ and $|F_{\lambda\bar{\mathbf{h}}}|<100$ were not included, thus eliminating the smallest magnitudes.

It is possible that the measured data would be accurate enough to permit the evaluation of individual phases in a fashion similar to that for (10) and (11). The appropriate general equations are

$$\cos({}_m\varphi_{2,\mathbf{h}}-{}_m\varphi_{3,\mathbf{h}})=(|{}_m\mathscr{F}_{1,\mathbf{h}}|^2-|{}_m\mathscr{F}_{2,\mathbf{h}}|^2 -|{}_m\mathscr{F}_{3,\mathbf{h}}|^2)/2|{}_m\mathscr{F}_{2,\mathbf{h}}||{}_m\mathscr{F}_{3,\mathbf{h}}| \qquad (19)$$

$$\cos({}_m\varphi_{1,\mathbf{h}}-{}_m\varphi_{3,\mathbf{h}})=(|{}_m\mathscr{F}_{1,\mathbf{h}}|^2-|{}_m\mathscr{F}_{2,\mathbf{h}}|^2 +|{}_m\mathscr{F}_{3,\mathbf{h}}|^2)/2|{}_m\mathscr{F}_{1,\mathbf{h}}||{}_m\mathscr{F}_{3,\mathbf{h}}|, \qquad (20)$$

where, for anomalous dispersion, case 1, ${}_1\varphi_{3,\mathbf{h}}$ is given by (14),

$$ {}_1\varphi_{1,\mathbf{h}} = \varphi_{\lambda\mathbf{h}} \quad (21) $$

$$ {}_1\varphi_{2,\mathbf{h}} = -\varphi_{\lambda\bar{\mathbf{h}}} \quad (22) $$

and $|{}_1\mathscr{F}_{1,\mathbf{h}}|$, $|{}_1\mathscr{F}_{2,\mathbf{h}}|$ and $|{}_1\mathscr{F}_{3,\mathbf{h}}|$ are defined in Table 1 and by taking the magnitude of (5). This gives

$$ \sin(-\varphi_{\lambda\bar{\mathbf{h}}} - \varphi^n_{2,\mathbf{h}}) = (|F_{\lambda\mathbf{h}}|^2 - |F_{\lambda\bar{\mathbf{h}}}|^2 - 4(f''/f^n_{2,\mathbf{h}})^2|F^n_{2,\mathbf{h}}|^2) \times [4(f''/f^n_{2,\mathbf{h}})|F_{\lambda\bar{\mathbf{h}}}||F^n_{2,\mathbf{h}}|]^{-1} \quad (23) $$

$$ \sin(\varphi_{\lambda\mathbf{h}} - \varphi^n_{2,\mathbf{h}}) = (|F_{\lambda\mathbf{h}}|^2 - |F_{\lambda\bar{\mathbf{h}}}|^2 + 4(f''/f^n_{2,\mathbf{h}})^2|F^n_{2,\mathbf{h}}|^2) \times [4(f''/f^n_{2,\mathbf{h}})|F_{\lambda\mathbf{h}}||F^n_{2,\mathbf{h}}|]^{-1}. \quad (24) $$

Comparable equations for isomorphous replacement, (10) and (11), also follow from (19) and (20) and the use of Table 1. In a manner similar to that described for isomorphous replacement, (23) and (24) can be applied by specifying a magnitude somewhat less than unity that the right sides of (23) and (24) must attain. When this value or greater is attained, the sine functions may be set approximately equal to +1 or −1, as appropriate, and then the $-\varphi_{\lambda\bar{\mathbf{h}}}$ or $\varphi_{\lambda\mathbf{h}}$ may be set approximately equal to $\varphi^n_{2,\mathbf{h}} + \pi/2$ or $\varphi^n_{2,\mathbf{h}} - \pi/2$, corresponding to the values of +1 and −1, respectively. This could give an initial set of values for a number of phases which could perhaps be refined and extended by use of computed values of triplet phase invariants from application of (15).

There is an alternative way to obtain individual phase values by algebraic means. This has been described (Karle, 1985*b*). The phases evaluated by these means are the $\varphi^n_{1,\mathbf{h}}$, the phases corresponding to the structure factors for the structure of the non-anomalously scattering atoms. The formation and evaluation of triplet phase invariants consisting of the $\varphi^n_{1,\mathbf{h}}$ has also been discussed (Karle, 1984*c*).

Concluding remarks

Formulas (6) and (15) provide values for triplet phase invariants in single isomorphous replacement and one-wavelength anomalous dispersion experiments when the structure of the replacement or that for a predominant anomalous scatterer is known. Actually, only the magnitudes of the structure factors for the replacement structure or the structure of the predominant anomalous scatterer are required since the triplet phase invariants for the heavy-atom phases may be set equal to zero to good approximation when the associated products of structure factor magnitudes are large. Such information is obtainable, for example, from application of the exact linear theory (Karle, 1980). When the latter theory is applied, there are also alternative ways to evaluate the triplet phase invariants (Karle, 1984*c*).

The potential utility of (6) and (15) is indicated by the accuracy obtained in the test calculations when reasonable random errors are introducted into the data. The use of the formulas may be facilitated by the determination of initial values for a number of phases by application of (10) and (11) for isomorphous replacement or (23) and (24) for anomalous dispersion when the phases associated with the replacement or the predominant anomalous scatterer structure are known. When (10), (11), (23) and (24) are used in the manner described, they give essentially unambiguous phase values. In general, however, they give values with a twofold ambiguity that may also be useful if, in the course of the application of the triplet phase invariants, the ambiguity could be resolved by use of phase values developed by the triplet phase invariants combined with the initially determined essentially unambiguous phase values. Throughout the history of the application of isomorphous replacement and anomalous dispersion, many ways have been developed to overcome the twofold ambiguity (see, for example, Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing, 1984).

The use of (10), (11), (23) and (24) has the potential to establish the appropriate enantiomorph for a macromolecule. If the structure of the replacement atoms in single isomorphous replacement is centrosymmetric and the method is used independently of anomalous dispersion information, the enantiomorph will have to be established in some other way, *e.g.* by the use of the triplet phase invariants in some manner comparable to that used for small-molecule structure determination. One approach to this matter has been presented by Fan Hai-fu & Gu Yuan-xin (1985) and Yao Jia-xing & Fan Hai-fu (1985).

It has been the intention in this paper to illustrate the potential information available in the individual techniques of single isomorphous replacement and one-wavelength anomalous dispersion from the use of certain algebraic formulas that can be computed quite rapidly. In actual practice, the potential would be enhanced if the techniques were combined. In addition to the formulas presented here, there are numerous algebraic (Ramaseshan & Abrahams, 1975) and probabilistic (Hauptman, 1982*a*, *b*; Giacovazzo, 1983; Pontenagel, Krabbendam, Peerdeman & Kroon, 1983; Fortier, Moore & Fraser, 1985) formulas available for fashioning strategies for the analysis of experimental data. There are also relations from an exact algebraic analysis (Karle, 1980, 1984*c*, 1985*b*) which give an essentially unique result with one-wavelength anomalous dispersion data and, when used with exact data, give exact phase values. It is likely that optimal strategies will depend upon the quality and character of the data. The formulas and procedures considered here should, in any case, enhance the variety of options that can be considered.

I wish to thank Mr Stephen Brenner for writing the programs and making the computations reported here.

This work was supported in part by USPHS grant GM 30902.

References

Adman, E. T., Sieker, L. C. & Jensen, L. H. (1973). *J. Biol. Chem.* **248**, 3987-3996.
Fan Hai-fu & Gu Yuan-xin (1985). *Acta Cryst. A*41, 280-284.
Fan Hai-fu, Han Fu-son, Qian Jin-zi & Yao Jia-xing (1984). *Acta Cryst.* **A40**, 489-495.
Fortier, S., Moore, N. J. & Fraser, M. E. (1985). *Acta Cryst.* **A41**, 571-577.
Giacovazzo, C. (1983). *Acta Cryst* **A39**, 585-592.
Hauptman, H. (1982*a*). *Acta Cryst.* **A38**, 289-294.
Hauptman, H. (1982*b*). *Acta Cryst.* **A38**, 632-641.
Karle, J. (1980). *Int. J. Quantum Chem. Symp.* **7**, 357-367.
Karle, J. (1983). *Acta Cryst.* **A39**, 800-805.
Karle, J. (1984*a*). *Acta Cryst.* **A40**, 4-11.
Karle, J. (1984*b*). *Acta Cryst.* **A40**, 366-373, 374-379.
Karle, J. (1984*c*). *Acta Cryst.* **A40**, 526-531.
Karle, J. (1985*a*). *Acta Cryst.* **A41**, 182-189.
Karle, J. (1985*b*). *Acta Cryst.* **A41**, 387-394.
Pontenagel, W. M. G. F., Krabbendam, H., Peerdeman, A. F. & Kroon, J. (1983). Eighth Eur. Crystallogr. Meet., Liège, Belgium, 8-12 August 1983. Abstract 4.01-P.
Ramaseshan, S. & Abrahams, S. C. (1975). Editors. *Anomalous Scattering*. Copenhagen: Munksgaard.
Timkovich, R. & Dickerson, R. E. (1976). *J. Biol. Chem.* **251**, 4033-4046.
Yao Jia-xing & Fan Hai-fu (1985). *Acta Cryst.* **A41**, 284-285.

Acta Cryst. (1986). A**42**, 253-256

Propagation of Phase Errors During Phase Refinement and Extension

By Lee Makowski

Department of Biochemistry and Molecular Biophysics, Columbia University, 630 W 168th Street, New York, NY 10032, USA

(*Received* 17 *September* 1985; *accepted* 14 *January* 1986)

Abstract

A formalism is developed for estimating phase errors during refinement using non-crystallographic symmetry, solvent flattening or density modification. This formalism, based on a separation of all structure factors into known (best estimate) and unknown (random variable) parts, leads directly to simple expressions for the propagation of phase errors during a refinement process. Phase extension and treatment of unmeasured reflections are readily incorporated into this scheme. The formulation provides a direct method for evaluating the success of a refinement process. This may be useful in cases where examination of the resulting electron density map does not provide a quantitative evaluation of the calculations, such as at low to moderate resolution or when phase extension has been used. It may also provide a basis for designing optimal refinement strategies.

Introduction

It is becoming common in crystallographic studies of macromolecules to use density modification, non-crystallographic symmetry and solvent flattening to improve electron density maps or to extend and refine initial phase sets. These refinement strategies benefit from a weighting scheme in which accurately phased reflections are included in the calculation of the electron density map with higher weights than poorly phased reflections (*e.g.* Sim, 1959; Bricogne, 1976; Rayment, Baker & Caspar, 1983). Estimation of errors in the initial phase sets as obtained, for instance, by isomorphous replacement, can be calculated taking into account observational errors and lack of closure (*e.g.* Blow & Crick, 1956; Dickerson, Kendrew & Strandberg, 1961). Errors during refinement have generally been estimated by some variation of the method put forth initially by Sim (1959) who calculated the phase errors resulting from using a partial structure for the calculation of structure factors. In most applications the electron density map being refined cannot be divided into known and unknown regions. Consequently, the phase error is estimated by some measure of the mean discrepancy between calculated and observed intensities (*e.g.* Hendrickson & Lattmann, 1970; Bricogne, 1976). Although these measures provide good relative estimates of the decrease in phase error during refinement (as judged from the corresponding electron density maps), they are not absolute measures, and the success of a refinement procedure is usually based on the interpretability of the resulting electron density map. This leaves open the question of the reliability of the results, particularly at low resolution or after phase extension.

In this paper, the reciprocal-space formalism first put forward by Crowther (1967, 1969) is used as a starting point for deriving simple expressions for

0108-7673/86/040253-04$01.50