Acta Cryst. (1982). A**38**, 289–294

On Integrating the Techniques of Direct Methods and Isomorphous Replacement I. The Theoretical Basis*

BY HERBERT HAUPTMAN

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA

(*Received* 18 *June* 1981; *accepted* 27 *October* 1981)

Abstract

Recent advances in direct methods are applied to structurally isomorphous pairs; in particular the probabilistic theory of the three-phase structure invariant in $P1$ is worked out. The neighborhood principle plays the central role. Specifically, the six-magnitude first neighborhood of each of the four kinds of three-phase structure invariant is defined and the joint probability distribution of the corresponding six structure factors is derived. This distribution leads to the conditional probability distribution of each kind of three-phase structure invariant, assuming as known the six magnitudes in its first neighborhood. In the favorable case that the variance of the distribution happens to be small, one obtains a reliable estimate (0 or π) of the structure invariant [the neighborhood principle: Hauptman (1975). *Acta Cryst.* A**31**, 680–687].

1. Introduction

Crystal structures having as many as 80–100 independent nonhydrogen atoms are more or less routinely solvable nowadays by direct methods. For macromolecules, on the other hand, the method of isomorphous replacement is almost universally used. It is natural to suppose that the ability to combine the two techniques would lead to methods more powerful than either. This fusion of the two techniques has now been accomplished and is described here; the initial applications strongly suggest that the anticipated gain in power has in fact been realized (Hauptman, Potter & Weeks, 1982). It is noteworthy that the neighborhood principle, which has played such an important part in the recent development of the more traditional direct methods, here plays an indispensable role.

If f_j and g_j denote atomic structure factors for a corresponding pair of isomorphous structures in $P1$, then respective normalized structure factors $E_{\mathbf{H}}$ and $G_{\mathbf{H}}$ are defined by

$$E_{\mathbf{H}} = |E_{\mathbf{H}}| \exp(i\varphi_{\mathbf{H}}) = \frac{1}{\alpha_{20}^{1/2}} \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{H}.\mathbf{r}_j), \quad (1.1)$$

$$G_{\mathbf{H}} = |G_{\mathbf{H}}| \exp(i\psi_{\mathbf{H}}) = \frac{1}{\alpha_{02}^{1/2}} \sum_{j=1}^{N} g_j \exp(2\pi i \mathbf{H}.\mathbf{r}_j) \quad (1.2)$$

where

$$\alpha_{mn} = \sum_{j=1}^{N} f_j^m g_j^n, \quad (1.3)$$

some of the f_j's (or g_j's) may be zero (or negative in the neutron diffraction case), and $\mathbf{r}_j$ is the position vector of the atom labeled j. This formulation includes not only the case of two strictly isomorphous structures but also, for example, the special case that the f structure is a native protein and the g structure a heavy-atom isomorphous derivative, as well as the case that one or both sets of data are obtained by neutron diffraction. The f_j and g_j are zero-angle atomic scattering factors and are therefore equal to the atomic numbers Z_j in the X-ray diffraction case but may be negative in the neutron diffraction case.

In the case of no isomorphism the problem of crystal structure determination, and therefore the phase problem, has long been known to be solvable in principle because the number of observable structure factor magnitudes $|E_{\mathbf{H}}|$ in general exceeds by far the number $(3N)$ of atomic coordinates needed to define the structure. This redundancy has in fact been exploited to yield simple relationships among the structure factors having probabilistic validity, and these in turn have played a key role in the solution of the phase problem. If now one assumes as known, in addition to the structure-factor magnitudes $|E_{\mathbf{H}}|$, the structure-factor magnitudes $|G_{\mathbf{H}}|$ for an isomorphous structure, then, in effect, the number of available data has been doubled (or at least greatly increased) with little or no corresponding increase in the number of parameters (still approximately $3N$) needed to define the unknown structure(s). One therefore anticipates that the ability to use both sets of observed magnitudes $|E_{\mathbf{H}}|$ and $|G_{\mathbf{H}}|$, by increasing the number of available data relative to the number of unknowns, will facilitate the solution of the phase problem. This expectation is in fact realized in several ways as the results obtained in this paper show. In the first place, although there are now more

* Presented at the Am. Crystallogr. Assoc. winter meeting, Texas A&M Univ., 22–26 March 1981, Abstr. C4.

0567-7394/82/030289-06$01.00

phases to be determined (phases φ and ψ rather than phases φ alone) than in the case of no isomorphism, the far greater number of available structure invariants more than compensates for the increased number of unknown phases. For example, there now exist two-phase structure invariants

$$\varphi_{\mathbf{H}} - \psi_{\mathbf{H}}, \tag{1.4}$$

having no analogue in the non-isomorphous case, and four different kinds of three-phase structure invariant,

$$\omega_0 = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \varphi_{\mathbf{L}}, \quad \omega_1 = \varphi_{\mathbf{H}} + \varphi_{\mathbf{K}} + \psi_{\mathbf{L}},$$
$$\omega_2 = \varphi_{\mathbf{H}} + \psi_{\mathbf{K}} + \psi_{\mathbf{L}}, \quad \omega_3 = \psi_{\mathbf{H}} + \psi_{\mathbf{K}} + \psi_{\mathbf{L}}, \tag{1.5}$$

where

$$\mathbf{H} + \mathbf{K} + \mathbf{L} = 0, \tag{1.6}$$

whereas only the first (or fourth) kind of structure invariant (1.5) can exist in the non-isomorphous case. Next, depending on the values of the six magnitudes

$$|E_{\mathbf{H}}|, |E_{\mathbf{K}}|, |F_{\mathbf{L}}|, |G_{\mathbf{H}}|, |G_{\mathbf{K}}|, |G_{\mathbf{L}}| \tag{1.7}$$

in the first neighborhood of any of the structure invariants (1.5), it is possible, in favorable cases, to obtain far more reliable estimates for the structure invariants (1.5) than is possible in the non-isomorphous case. Finally, again depending on the values of the six magnitudes (1.7), certain of the three-phase structure invariants (1.5) having the value π may be reliably identified, whereas in the non-isomorphous case only those three-phase invariants having the value zero can ever be identified by means of the three magnitudes in the first neighborhood alone. In summary, the existence of six magnitudes $|E|$ and $|G|$ which are now available, rather than only the three magnitudes $|E|$ (or $|G|$) in the non-isomorphous case, implies that it is now possible to estimate reliably larger numbers of three-phase structure invariants than is otherwise the case.

Although the major goal in this paper is to initiate the probabilistic theory of the three-phase structure invariant for isomorphous pairs, a brief account of the two-phase structure invariant is also given, firstly because of its intrinsic interest and importance, secondly because it provides the needed introduction to the more advanced theory of this invariant, and finally because it makes possible a greatly abbreviated account of the more complex analysis needed for the three-phase structure invariant. The mathematical formalism adopted here follows in a general sort of way the one introduced recently (Hauptman, 1975*a*,*b*) but there are some differences in detail.

2. The probabilistic theory of the two-phase structure invariant $\varphi_{\mathbf{H}} - \psi_{\mathbf{H}}$

The first neighborhood of the two-phase structure invariant (1.4) is defined to consist of the two magnitudes

$$|E_{\mathbf{H}}|, |G_{\mathbf{H}}|. \tag{2.1}$$

2.1. *The joint probability distribution of the two structure factors* $E_{\mathbf{H}}$, $G_{\mathbf{H}}$.

It is assumed that an isomorphous pair of structures in $P1$ with atomic position vectors $\mathbf{r}_j$, $j = 1, 2, \ldots, N$ is fixed and that normalized structure factors E and G are defined by (1.1)–(1.3). The reciprocal-lattice vector $\mathbf{H}$ is assumed to be the primitive random variable which is uniformly distributed in reciprocal space. Then the structure factors $E_{\mathbf{H}}$ and $G_{\mathbf{H}}$, as functions of the primitive random variable $\mathbf{H}$, are themselves random variables. Denote by $P(R,S;\Phi,\Psi)$ the joint probability distribution of the magnitudes $|E_{\mathbf{H}}|$, $|G_{\mathbf{H}}|$ and the phases $\varphi_{\mathbf{H}}$, $\psi_{\mathbf{H}}$ of the complex normalized structure factors $E_{\mathbf{H}}$, $G_{\mathbf{H}}$. Then, following methods previously described (Karle & Hauptman, 1958) $P(R,S;\Phi,\Psi)$ is given by the fourfold integral

$$P(R,S;\Phi,\Psi) = \frac{RS}{(2\pi)^4} \int_{\rho,\,\sigma=0}^{\infty} \int_{\theta,\,\chi=0}^{2\pi} \rho\sigma \times \exp\{-i[R\rho\cos(\theta-\Phi) + S\sigma\cos(\chi-\Psi)]\} \times \prod_{j=1}^{N} q_j \,\mathrm{d}\rho\,\mathrm{d}\sigma\,\mathrm{d}\theta\,\mathrm{d}\chi, \tag{2.2}$$

where

$$q_j = \left\langle \exp\left\{ \frac{if_j}{\alpha_{20}^{1/2}} \rho\cos(2\pi\mathbf{H}.\mathbf{r}_j - \theta) + \frac{ig_j}{\alpha_{02}^{1/2}} \sigma\cos(2\pi\mathbf{H}.\mathbf{r}_j - \chi) \right\} \right\rangle_{\mathbf{H}}. \tag{2.3}$$

Appendix I* contains some preliminary formulas, Appendix II the derivation of q_j, Appendix III the derivation of $\prod_{j=1}^{N} q_j$, and Appendix IV a brief description of the techniques devised to evaluate the fourfold integral (2.2). The final formula, taken from Appendix IV, is simply

$$P(R,S;\Phi,\Psi) = \frac{RS}{\pi^2(1-\alpha^2)} \exp\left\{ -\frac{1}{1-\alpha^2}[R^2 - 2\alpha RS\cos(\Phi-\Psi) + S^2] \right\}, \tag{2.4}$$

* Appendices I–IX have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 36520 (23 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

where α is defined by

$$\alpha = \frac{\alpha_{11}}{\alpha_{20}^{1/2}\,\alpha_{02}^{1/2}}. \tag{2.5}$$

Related distributions employing a different probabilistic background and mathematical formalism have been derived in a different context by Srinivasan and co-workers; see, for example, Srinivasan & Parthasarathy (1976, Ch. 5).

2.2. *The joint probability distribution of the two structure-factor magnitudes* $|E_{\mathbf{H}}|$, $|G_{\mathbf{H}}|$

Under the same assumptions as in § 2.1, $|E_{\mathbf{H}}|$ and $|G_{\mathbf{H}}|$ are random variables whose joint probability distribution $P(R,S)$ is obtained by integrating (2.4) with respect to Φ and Ψ between the limits 0 to 2π. Employing (I.8) (from Appendix I) with $m = 0$ one finds

$$P(R,S) = \frac{4RS}{1-\alpha^2}\exp\left\{-\frac{R^2+S^2}{1-\alpha^2}\right\} I_0\left(\frac{2\alpha RS}{1-\alpha^2}\right), \tag{2.6}$$

where I_0 is the modified Bessel function (Watson, 1958).

2.3. *The correlation coefficient r of the pair* $|E_{\mathbf{H}}|^2$, $|G_{\mathbf{H}}|^2$

The distribution (2.6) leads directly to a formula for the correlation coefficient r of the pair $|E_{\mathbf{H}}|^2$, $|G_{\mathbf{H}}|^2$ [Appendix V, equation (V.6)]:

$$r = \frac{\langle(|E_{\mathbf{H}}|^2 - \overline{|E_{\mathbf{H}}|^2})(|G_{\mathbf{H}}|^2 - \overline{|G_{\mathbf{H}}|^2})\rangle_{\mathbf{H}}}{\langle(|E_{\mathbf{H}}|^2 - \overline{|E_{\mathbf{H}}|^2})^2\rangle_{\mathbf{H}}^{1/2}\,\langle(|G_{\mathbf{H}}|^2 - \overline{|G_{\mathbf{H}}|^2})^2\rangle_{\mathbf{H}}^{1/2}}$$
$$= \alpha^2 = \alpha_{11}^2/\alpha_{20}\alpha_{02}. \tag{2.7}$$

Clearly,

$$0 \le r \le 1. \tag{2.8}$$

If the averages in (2.7) are taken over all reciprocal-lattice vectors **H** having a fixed value of $\sin\theta/\lambda$ then r, as a function of $\sin\theta/\lambda$ is constant in the case of perfect isomorphism. If, however, the isomorphism is imperfect, then r will be a monotonically decreasing function of $\sin\theta/\lambda$; in this case r, as a function of $\sin\theta/\lambda$, may be regarded as a measure of the degree of isomorphism between the two structures.

2.4. *The conditional probability distribution of the structure invariant* $\omega = \varphi_H - \psi_H$, *given the two magnitudes* $|E_{\mathbf{H}}|$, $|G_{\mathbf{H}}|$

The joint conditional probability distribution $P(\Phi,\Psi|R,S)$ of the pair $\varphi_{\mathbf{H}}$, $\psi_{\mathbf{H}}$, given that $|E_{\mathbf{H}}|$ and $|G_{\mathbf{H}}|$ have the pre-assigned non-negative values R and S, respectively, is obtained directly from (2.4) by fixing R and S and multiplying by a suitable normalizing factor:

$$P(\Phi,\Psi|R,S) = \frac{1}{L}\exp\left\{\frac{2\alpha}{1-\alpha^2}RS\cos(\Phi-\Psi)\right\}, \tag{2.9}$$

where the normalizing parameter L is not relevant for the present purpose. Since the distribution (2.9) depends only on the difference $\Phi - \Psi$, (2.9) leads directly to the conditional probability distribution of the structure invariant $\omega = \varphi_{\mathbf{H}} - \psi_{\mathbf{H}}$, given $|E_{\mathbf{H}}|$ and $|G_{\mathbf{H}}|$:

$$P(\Omega|RS) = \frac{1}{K}\exp\left\{\frac{2\alpha}{1-\alpha^2}RS\cos\Omega\right\}, \tag{2.10}$$

where, from (I.8),

$$K = 2\pi I_0\left(\frac{2\alpha}{1-\alpha^2}RS\right). \tag{2.11}$$

Equation (2.10) implies that

$$\varphi_{\mathbf{H}} - \psi_{\mathbf{H}} \simeq 0 \tag{2.12}$$

provided that

$$\frac{2\alpha}{1-\alpha^2}|E_{\mathbf{H}}G_{\mathbf{H}}| \text{ is large.} \tag{2.13}$$

3. The probabilistic theory of the three-phase structure invariants

It has already been seen that there are four kinds of three-phase structure invariants, (1.5), the first neighborhood of each of which consists of the six magnitudes (1.7). Our major task therefore is to derive the joint probability distribution of the six structure factors whose magnitudes constitute the first neighborhood of each of the structure invariants (1.5).

3.1. *The joint probability distribution of the six structure factors* $E_{\mathbf{H}}$, $E_{\mathbf{K}}$, $E_{\mathbf{L}}$, $G_{\mathbf{H}}$, $G_{\mathbf{K}}$, $G_{\mathbf{L}}$, *where* $\mathbf{H}+\mathbf{K}+\mathbf{L}=0$.

Make the same assumptions as in § 2.1. Denote reciprocal space by S, and by $S \times S \times S$ the threefold Cartesian product, *i.e.* the collection of all ordered triples (**h**,**k**,**l**) of reciprocal-lattice vectors **h**,**k**,**l**. The primitive random variable is the ordered triple (**H**,**K**,**L**) of reciprocal-lattice vectors which is assumed to be uniformly distributed over the subset of $S \times S \times S$ defined by (1.6). Then the structure factors $E_{\mathbf{H}}$, $E_{\mathbf{K}}$, $E_{\mathbf{L}}$, $G_{\mathbf{H}}$, $G_{\mathbf{K}}$, $G_{\mathbf{L}}$, as functions of the primitive random variables **H**, **K**, **L**, are themselves random variables. Denote by

$$P = P(R_1,R_2,R_3,S_1,S_2,S_3;\Phi_1,\Phi_2,\Phi_3,\Psi_1,\Psi_2,\Psi_3) \tag{3.1}$$

the joint probability distribution of the magnitudes $|E_H|$, $|E_K|$, $|E_L|$, $|G_H|$, $|G_K|$, $|G_L|$ and the phases φ_H, φ_K, φ_L, ψ_H, ψ_K, ψ_L of the complex normalized structure factors E_H, E_K, E_L, G_H, G_K, G_L. Then, as in § 2.1, P is given by the twelvefold integral

$$P=\frac{1}{(2\pi)^{12}}R_1R_2R_3S_1S_2S_3$$
$$\times\int_{\rho_1,\rho_2,\rho_3,\sigma_1,\sigma_2,\sigma_3=0}^{\infty}\int_{\theta_1,\theta_2,\theta_3,\chi_1,\chi_2,\chi_3=0}^{2\pi}\rho_1\rho_2\rho_3\sigma_1\sigma_2\sigma_3$$
$$\times\exp\{-i[\rho_1R_1\cos(\theta_1-\Phi_1)+\rho_2R_2\cos(\theta_2-\Phi_2)$$
$$+\rho_3R_3\cos(\theta_3-\Phi_3)$$
$$+\sigma_2S_1\cos(\chi_1-\Psi_1)$$
$$+\sigma_2S_2\cos(\chi_2-\Psi_2)$$
$$+\sigma_sS_3\cos(\chi_3-\psi_3)]\}\prod_{j=1}^{N}q_j\,d\rho\,d\sigma\,d\theta\,d\chi, \quad (3.2)$$

where q_j is now defined by

$$q_j=\langle\exp\{(if_j/\alpha_{20}^{1/2})[\rho_1\cos(2\pi\mathbf{H}.\mathbf{r}_j-\theta_1)$$
$$+\rho_2\cos(2\pi\mathbf{K}.\mathbf{r}_j-\theta_2)$$
$$+\rho_3\cos(2\pi\mathbf{L}.\mathbf{r}_j-\theta_3)]$$
$$+(ig_j/\alpha_{02}^{1/2})[\sigma_1\cos(2\pi\mathbf{H}.\mathbf{r}_j-\chi_1)$$
$$+\sigma_2\cos(2\pi\mathbf{K}.\mathbf{r}_j-\chi_2)$$
$$+\sigma_3\cos(2\pi\mathbf{L}.\mathbf{r}_j-\chi_3)]\}\rangle_{\mathbf{H}+\mathbf{K}+\mathbf{L}=0}. \quad (3.3)$$

In view of Appendices I–IV, Appendices VI–VIII, which contain the derivation of q_j, $\prod_{j=1}^{N}q_j$, and the evaluation of the twelvefold integral (3.2), are greatly abbreviated. The final formula [from equation (VIII.13)], the first major result of this paper, is simply

$$P=[1/\pi^6(1-\alpha^2)^3]R_1R_2R_3S_1S_2S_3$$
$$\times\exp\{-[1/(1-\alpha^2)]$$
$$\times(R_1^2+R_2^2+R_3^2+S_1^2+S_2^2+S_3^2)$$
$$+2\beta[R_1S_1\cos(\Phi_1-\Psi_1)$$
$$+R_2S_2\cos(\Phi_2-\Psi_2)+R_3S_3\cos(\Phi_3-\Psi_3)]$$
$$+2\beta_0R_1R_2R_3\cos(\Phi_1+\Phi_2+\Phi_3)$$
$$+2\beta_1[R_1R_2S_3\cos(\Phi_1+\Phi_2+\Psi_3)$$
$$+R_1S_2R_3\cos(\Phi_1+\Psi_2+\Phi_3)$$
$$+S_1R_2R_3\cos(\Psi_1+\Phi_2+\Phi_3)]$$
$$+2\beta_2[R_1S_2S_3\cos(\Phi_1+\Psi_2+\Psi_3)$$
$$+S_1R_2S_3\cos(\Psi_1+\Phi_2+\Psi_3)$$
$$+S_1S_2R_3\cos(\Psi_1+\Psi_2+\Phi_3)]$$
$$+2\beta_3S_1S_2S_3\cos(\Psi_1+\Psi_2+\Psi_3)\}, \quad (3.4)$$

where α is defined in (2.5) and

$$\beta=\alpha/(1-\alpha^2), \quad (3.5)$$

$$\beta_0=\frac{1}{\alpha_{20}^{3/2}\alpha_{02}^{3}(1-\alpha^2)^3}[\alpha_{30}\alpha_{02}^3-3\alpha_{21}\alpha_{02}^2\alpha_{11}$$
$$+3\alpha_{12}\alpha_{02}\alpha_{11}^2-\alpha_{03}\alpha_{11}^3], \quad (3.6)$$

$$\beta_1=\frac{1}{\alpha_{20}^{2}\alpha_{02}^{5/2}(1-\alpha^2)^3}[(\alpha_{21}\alpha_{20}-\alpha_{30}\alpha_{11})\alpha_{02}^2$$
$$-2(\alpha_{12}\alpha_{20}-\alpha_{21}\alpha_{11})\alpha_{02}\alpha_{11}$$
$$+(\alpha_{03}\alpha_{20}-\alpha_{12}\alpha_{11})\alpha_{11}^2], \quad (3.7)$$

$$\beta_2=\frac{1}{\alpha_{20}^{5/2}\alpha_{02}^{2}(1-\alpha^2)^3}[(\alpha_{12}\alpha_{02}-\alpha_{03}\alpha_{11})\alpha_{20}^2$$
$$-2(\alpha_{21}\alpha_{02}-\alpha_{12}\alpha_{11})\alpha_{20}\alpha_{11}$$
$$+(\alpha_{30}\alpha_{02}-\alpha_{21}\alpha_{11})\alpha_{11}^2], \quad (3.8)$$

$$\beta_3=\frac{1}{\alpha_{20}^{3}\alpha_{02}^{3/2}(1-\alpha^2)^3}[\alpha_{03}\alpha_{20}^3-3\alpha_{12}\alpha_{20}^2\alpha_{11}$$
$$+3\alpha_{21}\alpha_{20}\alpha_{11}^2-\alpha_{30}\alpha_{11}^3]. \quad (3.9)$$

Equation (3.4) should be compared with (2.4).

3.2. *The conditional probability distribution of the three-phase structure invariant* $\omega_0=\varphi_H+\varphi_K+\varphi_L$, *given the six magnitudes* $|E_H|$, $|E_K|$, $|E_L|$, $|G_H|$, $|G_K|$, $|G_L|$ *in its first neighborhood*

Assume again that an isomorphous pair of structures in $P1$ is fixed and that the six non-negative numbers R_1, R_2, R_3, S_1, S_2, S_3 are also specified. Suppose that the primitive random variable is the ordered triple (**H**,**K**,**L**) which is now assumed to be uniformly distributed over the subset of $S\times S\times S$ defined by (1.6) and

$$|E_H|=R_1, \quad |E_K|=R_2, \quad |E_L|=R_3, \quad (3.10)$$
$$|G_H|=S_1, \quad |G_K|=S_2, \quad |G_L|=S_3. \quad (3.11)$$

Then the structure invariant $\omega_0=\varphi_H+\varphi_K+\varphi_L$, as a function of the primitive random variable (**H**,**K**,**L**), is itself a random variable. Denote by $P_0(\Omega_0|R_1,R_2,R_3,S_1,S_2,S_3)$ the conditional probability distribution of $\varphi_H+\varphi_K+\varphi_L$, given (3.10) and (3.11). Then $P_0(\Omega_0|R_1,R_2,R_3,S_1,S_2,S_3)$ is derived from (3.4) [or equation (VIII.13) of Appendix VIII] by fixing R_1, R_2, R_3, S_1, S_2, S_3, integrating with respect to Ψ_1, Ψ_2, Ψ_3 from 0 to 2π, and multiplying by a suitable normalizing factor. This analysis is briefly described in Appendix IX. The final formula [from equation (IX.1)], the second major result of this paper, is

$$P_0(\Omega_0|R_1,R_2,R_3,S_1,S_2,S_3)\simeq\frac{1}{K_0}\exp(A_0\cos\Omega_0), \quad (3.12)$$

where

$$K_0 = 2\pi I_0(A_0), \tag{3.13}$$

$$\begin{aligned} A_0 = 2\{&\beta_0 R_1 R_2 R_3 + \beta_1[R_1 R_2 R_3 T(2\beta R_3 S_3) \\ &+ R_1 S_2 R_3 T(2\beta R_2 S_2) \\ &+ S_1 R_2 R_3 T(2\beta R_1 S_1)] \\ &+ \beta_2[R_1 S_2 S_3 T(2\beta R_2 S_2) T(2\beta R_3 S_3) \\ &+ S_1 R_2 S_3 T(2\beta R_1 S_1) T(2\beta R_3 S_3) \\ &+ S_1 S_2 R_3 T(2\beta R_1 S_1) T(2\beta R_2 S_2)] \\ &+ \beta_3 S_1 S_2 S_3 T(2\beta R_1 S_1) T(2\beta R_2 S_2) \\ &\times T(2\beta R_3 S_3)\}, \end{aligned} \tag{3.14}$$

and the function T is the ratio of two modified Bessel functions:

$$T(z) = \frac{I_1(z)}{I_0(z)}. \tag{3.15}$$

3.3. *The conditional probability distribution of the three-phase structure invariant* $\omega_1 = \varphi_H + \varphi_K + \psi_L$, *given the six magnitudes* $|E_H|$, $|E_K|$, $|E_L|$, $|G_H|$, $|G_K|$, $|G_L|$ *in its first neighborhood*

With the same probabilistic background as in § 3.2, the structure invariant $\omega_1 = \varphi_H + \varphi_K + \psi_L$, as a function of the primitive random variable (**H**,**K**,**L**), is itself a random variable whose conditional probability distribution $P_1(\Omega_1 | R_1,R_2,R_3,S_1,S_2,S_3)$, given (3.10) and (3.11), is derived in the same way as is P_0 (Appendix IX):

$$P_1(\Omega_1 | R_1,R_2,R_3,S_1,S_2,S_3) \simeq \frac{1}{K_1} \exp(A_1 \cos \Omega_1), \tag{3.16}$$

where

$$K_1 = 2\pi I_0(A_1), \tag{3.17}$$

and

$$\begin{aligned} A_1 = 2\{&\beta_1 R_1 R_2 S_3 + \beta_0 R_1 R_2 R_3 T(2\beta R_3 S_3) \\ &+ \beta_2[R_1 S_2 S_3 T(2\beta R_2 S_2) \\ &+ S_1 R_2 S_3 T(2\beta R_1 S_1)] \\ &+ \beta_1[R_1 S_2 R_3 T(2\beta R_2 S_2) T(2\beta R_3 S_3) \\ &+ S_1 R_2 R_3 T(2\beta R_1 S_1) T(2\beta R_3 S_3)] \\ &+ \beta_3 S_1 S_2 S_3 T(2\beta R_1 S_1) T(2\beta R_2 S_2) \\ &+ \beta_2 S_1 S_2 R_3 T(2\beta R_1 S_1) T(2\beta R_2 S_2) \\ &\times T(2\beta R_3 S_3)\}, \end{aligned} \tag{3.18}$$

the third major result.

3.4. *The conditional probability distribution of the three-phase structure invariant* $\omega_2 = \varphi_H + \psi_K + \psi_L$, *given six magnitudes* $|E_H|$, $|E_K|$, $|E_L|$, $|G_H|$, $|G_K|$, $|G_L|$ *in its first neighborhood*

With the usual probabilistic background the structure invariant $\omega_2 = \varphi_H + \psi_K + \psi_L$ is a random variable whose conditional probability distribution, given the six magnitudes (3.10) and (3.11), is found in the usual way (Appendix IX):

$$P_2(\Omega_2 | R_1,R_2,R_3,S_1,S_2,S_3) \simeq \frac{1}{K_2} \exp(A_2 \cos \Omega_2), \tag{3.19}$$

where

$$K_2 = 2\pi I_0(A_2) \tag{3.20}$$

and

$$\begin{aligned} A_2 = 2\{&\beta_2 R_1 S_2 S_3 + \beta_1[R_1 R_2 S_3 T(2\beta R_2 S_2) \\ &+ R_1 S_2 R_3 T(2\beta R_3 S_3)] \\ &+ \beta_3 S_1 S_2 S_3 T(2\beta R_1 S_1) \\ &+ \beta_0 R_1 R_2 R_3 T(2\beta R_2 S_2) T(2\beta R_3 S_3) \\ &+ \beta_2[S_1 R_2 S_3 T(2\beta R_1 S_1) T(2\beta R_2 S_2) \\ &+ S_1 S_2 R_3 T(2\beta R_1 S_1) T(2\beta R_3 S_3)] \\ &+ \beta_1 S_1 R_2 R_3 T(2\beta R_1 S_1) T(2\beta R_2 S_2) \\ &\times T(2\beta R_3 S_3)\}, \end{aligned} \tag{3.21}$$

the fourth major result.

3.5. *The conditional probability distribution of the three-phase structure invariant* $\omega_3 = \psi_H + \psi_K + \psi_L$, *given the six magnitudes in its first neighborhood*

As in the previous sections, the structure invariant $\omega_3 = \psi_H + \psi_K + \psi_L$ is a random variable whose conditional probability distribution, given the six magnitudes (3.10) and (3.11), is found in the usual way (Appendix IX):

$$P_3(\Omega_3 | R_1,R_2,R_3,S_1,S_2,S_3) \simeq \frac{1}{K_3} \exp(A_3 \cos \Omega_3), \tag{3.22}$$

where

$$K_3 = 2\pi I_0(A_3) \tag{3.23}$$

and

$$\begin{aligned} A_3 = 2\{&\beta_3 S_1 S_2 S_3 + \beta_2[R_1 S_2 S_3 T(2\beta R_1 S_1) \\ &+ S_1 R_2 S_3 T(2\beta R_2 S_2) + S_1 S_2 R_3 T(2\beta R_3 S_3)] \\ &+ \beta_1[R_1 R_2 S_3 T(2\beta R_1 S_1) T(2\beta R_2 S_2) \\ &+ R_1 S_2 R_3 T(2\beta R_1 S_1) T(2\beta R_3 S_3) \\ &+ S_1 R_2 R_3 T(2\beta R_2 S_2) T(2\beta R_3 S_3)] \\ &+ \beta_0 R_1 R_2 R_3 T(2\beta R_1 S_1) T(2\beta R_2 S_2) \\ &\times T(2\beta R_3 S_3)\}, \end{aligned} \tag{3.24}$$

which is the fifth major result of this paper.

It is noteworthy that in the important special case that the f structure is a native protein and the g structure a heavy-atom isomorphous derivative then, on the basis of some preliminary calculations, it appears that

$$\beta_0 < 0, \quad \beta_1 > 0, \quad \beta_2 < 0, \quad \beta_3 > 0, \quad (3.25)$$

and that $|\beta_0|$, $|\beta_1|$, $|\beta_2|$, $|\beta_3|$ form an approximate arithmetic progression.

4. Concluding remarks

Recent advances in direct methods have been here integrated with the method of isomorphous replacement, and the probabilistic theory of the three-phase structure invariants for an isomorphous pair of structures has been worked out in some detail. The analysis includes the special case that one member of the pair is a native protein and the other member is a heavy-atom isomorphous derivative. A great deal of additional work remains to be done, *e.g.* the theory of the two-phase structure invariant and the higher-order structure invariants and seminvariants, the theory of structurally isomorphous triples, quartets, *etc.*, and the role of anomalous dispersion.

The initial applications of the work described here, using error-free data from a native protein and a single heavy-atom derivative, have been made, and these are presented in the following paper (Hauptman, Potter & Weeks, 1982). Although these initial results are very encouraging, attempts to apply these methods to the solution of unknown macromolecular structures must first overcome the obstacle presented by structures in which the number and occupancy factors of the heavy atoms may be unknown *a priori*. For such structures methods must be devised for estimating the parameters, β, β_0, β_1, β_2, and β_3 [equations (3.5)–(3.9)], which appear in the distributions (3.12), (3.16), (3.19), and (3.22), in terms of observed intensities only. Already (3.5) expresses β in terms of α, the square root of the correlation coefficient of the pair $|E_H|^2$, $|G_H|^2$. In a similar way it may be shown that β_0, β_1, β_2, and β_3, are likewise expressible in terms of the $|E_H|^2$ and the $|G_H|^2$ alone, but this work is outside the scope of the present paper and will be published separately.

Finally, the effect of errors inherent in experimental data as well as imperfect isomorphism has not been considered in this paper, the major purpose of which has been to formulate the basic theory. It is intended to present a study of the effect of errors in real data at a later date.

The research described here was supported by DHHS Grant No. GM26195 and Grant No. CHE-7911282 from the National Science Foundation, as well as a grant from the James H. Cummings Foundation, Inc. The author is indebted to Dr Suzanne Fortier who verified the lengthy calculations briefly summarized in this paper.

References

HAUPTMAN, H. (1975*a*). *Acta Cryst.* A**31**, 671–679.
HAUPTMAN, H. (1975*b*). *Acta Cryst.* A**31**, 680–687.
HAUPTMAN, H., POTTER, S. & WEEKS, C. (1982). *Acta Cryst.* A**38**, 294–300.
KARLE, J. & HAUPTMAN, H. (1958). *Acta Cryst.* **11**, 264–269.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. New York: Pergamon Press.
WATSON, G. N. (1958) *A Treatise on the Theory of Bessel Functions*. Cambridge Univ. Press.

Acta Cryst. (1982). A**38**, 294–300

On Integrating the Techniques of Direct Methods and Isomorphous Replacement. II. The First Applications

BY HERBERT HAUPTMAN, STEPHEN POTTER AND CHARLES M. WEEKS

Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA

(*Received* 18 *June* 1981; *accepted* 27 *October* 1981)

Abstract

With error-free diffraction data from the protein cytochrome c_{550} from *Paracoccus denitrificans*, having molecular weight $M_r \simeq 14\,500$, space group $P2_12_12_1$, and a single $PtCl_4^{2-}$ derivative, estimates (0 or π) of the three-phase structure invariants are obtained by recently secured direct methods employing the six-magnitude first neighborhood [Hauptman (1982). *Acta Cryst.* A**38**, 289–294] and compared with the known values.

0567-7394/82/030294-07$01.00