PACIOREK, W. A. & CHAPUIS, G. (1992). *J. Appl. Cryst.* **25**, 317–322.

PACIOREK, W. A. & KUCHARCZYK, D. (1985). *Acta Cryst.* **A41**, 462–466.

PETRÍČEK, V. & COPPENS, P. (1988). *Acta Cryst.* **A44**, 235–239.

PETRÍČEK, V., COPPENS, P. & BECKER, P. (1985). *Acta Cryst.* **A41**, 478–483.

PETRÍČEK, V., MALÝ, K. & CÍSAŘOVÁ, I. (1991). *Methods of Structural Analysis of Modulated Structures and Quasicrystals*, edited by J. M. PÉREZ-MATO, F. J. ZÚÑIGA & G. MADARIAGA, pp. 262–267. Singapore: World Scientific.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A., VETTERLING, W. T. (1992). *Numerical Recipes: the Art of Scientific Computing*, 2nd ed. Cambridge Univ. Press.

REISS, H. R. (1980). *Phys. Rev. A*, **22**, 1786–1813.

WATSON, G. N. (1958). *A Treatise on the Theory of Bessel Functions*, 2nd ed. Cambridge Univ. Press.

WOLFF, P. M. DE (1974). *Acta Cryst.* **A30**, 777–785.

WOLFF, P. M. DE & VAN AALST, W. (1972). *Acta Cryst.* **A28**, S111.

WOLFF, P. M. DE, JANSSEN, T. & JANNER, A. (1981). *Acta Cryst.* **A37**, 625–636.

Structure Solution by Minimal-Function Phase Refinement and Fourier Filtering. I. Theoretical Basis

BY GEORGE T. DeTITTA AND CHARLES M. WEEKS

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

PAMELA THUMAN* AND RUSS MILLER

Department of Computer Science, State University of New York at Buffalo, Buffalo, NY 14260, USA

AND HERBERT A. HAUPTMAN

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

Abstract

Eliminating the N atomic position vectors r_j, $j = 1$, 2, ..., N, from the system of equations defining the normalized structure factors E_H yields a system of identities that the E_H's must satisfy, provided that the set of E_H's is sufficiently large. Clearly, for fixed N and specified space group, this system of identities depends only on the set $\{H\}$, consisting of n reciprocal-lattice vectors H, and is independent of the crystal structure, which is assumed for simplicity to consist of N identical atoms per unit cell. However, for a fixed crystal structure, the magnitudes $|E_H|$ are uniquely determined so that a system of identities is obtained among the corresponding phases φ_H alone, which depends on the presumed known magnitudes $|E_H|$ and which must of necessity be satisfied. The known conditional probability distributions of triplets and quartets, given the values of certain magnitudes $|E|$, lead to a function $R(\varphi)$ of phases, uniquely determined by magnitudes $|E|$ and having the property that $R_T < \frac{1}{2} < R_R$, where R_T is the value of $R(\varphi)$ when the phases are equal to their true values, no matter what the choice of origin and enantiomorph, and R_R is the value of $R(\varphi)$ when the phases are chosen at random. The following conjecture is therefore plausible: the global minimum of $R(\varphi)$, where the phases are constrained to satisfy all identities among them that are known to exist, is attained when the phases are equal to their true values and is thus equal to R_T. This 'minimal principle' replaces the problem of phase determination by that of finding the global minimum of the function $R(\varphi)$ constrained by the identities that the phases must satisfy and suggests strategies for determining the values of the phases in terms of N and the known magnitudes $|E|$. Equivalently, the minimal principle leads to the solution of the (in general redundant) system of equations satisfied by the phases φ_H.

Introduction

The structure invariants, in this paper only triplets and quartets, link the observed magnitudes $|E|$ with the desired phases φ of the normalized structure factors E. The traditional techniques of direct methods use the conditional probability distributions of the structure invariants to obtain estimates of their values and thus relationships among the individual phases having probabilistic validity. These rela-

---

* Present address: Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.

tionships, together with the ubiquitous tangent formula, in one version or another, constitute the essential core of most direct-methods programs. If data are available to atomic resolution, these methods work well for structures having up to some 150 non-H atoms per asymmetric unit but are not routinely applicable to structures of much greater complexity, because the required estimates of the structure invariants are not sufficiently reliable for such large structures. Our goal is to describe strategies, based on the minimal principle, that show promise in strengthening the traditional direct-methods techniques, in particular for larger structures.

In the present work, the structure invariants, specifically the triplets and negative quartets, again play the central role; however, instead of placing major emphasis on estimates of their values, the weak link in traditional techniques, we employ the conditional probability distributions themselves, which are known even for very large structures. Strictly speaking, of course, it must be assumed that the structure consists of N identical atoms, uniformly and independently distributed in the unit cell, for the distributions to be valid. To the extent that this condition is violated, the phases obtained by the method described here will be subject to errors in addition to those caused by errors in the observed intensities.

The algebraic background

Identities among the phases

The normalized structure factors are defined by

$$E_H = |E_H| \exp(i\varphi_H) = N^{-1/2} \sum_{j=1}^{N} \exp(2\pi i H \cdot r_j), \quad (1)$$

where H is an arbitrary reciprocal-lattice vector, N is the number of atoms in the unit cell, which for simplicity are here assumed to be identical, and r_j is the position vector of the atom labeled j. It should perhaps be emphasized that, by basing our analysis on the system of equations (1), we capture both the atomicity and non-negativity properties of the electron-density function in the crystal. Clearly, if the system of equations (1) is sufficiently large and the set $\{H\}$, consisting of n reciprocal-lattice vectors, is specified, elimination of the atomic position vectors r_j yields a system of identities among the E's, i.e. the magnitudes $|E|$ and phases φ, which depends only on N and on the space group – assumed to be fixed – and is independent of the crystal structure:

$$F(|E_H|; \varphi_H) \equiv 0. \quad (2)$$

If the magnitudes $|E|$ are known, these identities in turn lead to identities among only the phases φ,

dependent on the presumed known magnitudes $|E|$, which must of necessity be satisfied:

$$G(\varphi_H||E_H|) \equiv 0. \quad (3)$$

It is to be emphasized that the system of equations (2), each a function of the $2n$ variables $|E_H|$ and φ_H, is, for fixed N and fixed space group, independent of the crystal structure. In sharp contrast, the system of equations (3), each a function of the n variables φ_H, depends also on the magnitudes $|E_H|$, presumed to be known, and depends therefore on the crystal structure. If n is sufficiently large, then the system (3) is clearly redundant. For example, in the space group $P1$ let $n > 3(N - 1)$ and recall that one atomic position vector may be arbitrarily specified, thus fixing the origin. Then, by eliminating the remaining $N - 1$ atomic position vectors and equating real and imaginary parts of each equation in (2), one obtains the $2n - 3(N - 1)$ equations of the system (2) and, likewise, the $2n - 3(N - 1)$ equations of the system (3) in the n unknown phases φ_H. Since $n > 3(N - 1)$, $2n - 3(N - 1) > n$ so that the system (3) consists of more equations than unknowns φ_H. In short, the system (3), presumed to be independent, is redundant. Our major goal is to devise numerical algorithms that will solve the redundant system of equations (3) for the unknown phases φ_H and, to this end, the minimal principle plays the central role. It is to be emphasized that all we have done thus far is to show the existence of the system (3); the explicit functional forms of the left-hand sides, presumably polynomials in the cosine invariants with coefficients that are themselves polynomials in the magnitudes $|E_H|$, have not been found [with the minor exceptions of $N = 1$, 2, 3 in the space group $P1$ (unpublished work)] and are in fact not needed.

The structure invariants

The most important classes of structure invariants are the three-phase structure invariants (triplets),

$$\varphi_{HK} = \varphi_H + \varphi_K + \varphi_{-H-K}, \quad (4)$$

and the four-phase structure invariants (quartets),

$$\varphi_{LMN} = \varphi_L + \varphi_M + \varphi_N + \varphi_{-L-M-N}. \quad (5)$$

The probabilistic background

It is assumed that the position vectors r_j are random variables that are uniformly and independently distributed. Then, the structure invariants, as functions of random variables via (1), (4) and (5), are themselves random variables and their conditional probability distributions, with certain magnitudes $|E|$ assumed known, may then be found.

The conditional probability distribution of the triplet

For fixed reciprocal-lattice vectors **H** and **K**, the conditional probability distribution of the triplet φ_{HK} [(4)], with the three magnitudes

$$|E_H|, \quad |E_K|, \quad |E_{H+K}| \qquad (6)$$

assumed known, is known to be (Cochran, 1955)

$$P(\Phi|A_{HK}) = [2\pi I_0(A_{HK})]^{-1} \exp(A_{HK} \cos \Phi), \quad (7)$$

where

$$A_{HK} = 2N^{-1/2}|E_H E_K E_{H+K}| \qquad (8)$$

and I_0 is the modified Bessel function. From (7), it follows that the mode of φ_{HK} is zero, the conditional expectation value of $\cos \varphi_{HK}$, given A_{HK}, is

$$\varepsilon(\cos \varphi_{HK}|A_{HK}) = [I_1(A_{HK})/I_0(A_{HK})] > 0 \qquad (9)$$

and the conditional variance of $\cos \varphi_{HK}$, given A_{HK}, is

$$\mathrm{var}\,(\cos \varphi_{HK}|A_{HK})$$

$$= \varepsilon(\cos^2 \varphi_{HK}|A_{HK}) - [\varepsilon(\cos \varphi_{HK}|A_{HK})]^2$$

$$= \tfrac{1}{2} + \tfrac{1}{2}[I_2(A_{HK})/I_0(A_{HK})] - [I_1^2(A_{HK})/I_0^2(A_{HK})], \quad (10)$$

i.e.

$$\mathrm{var}\,(\cos \varphi_{HK}|A_{HK}) = \tfrac{1}{2} + \tfrac{1}{2}t'_{HK} - t^2_{HK}, \qquad (11)$$

where

$$t_{HK} = I_1(A_{HK})/I_0(A_{HK}) = \varepsilon(\cos \varphi_{HK}|A_{HK}) \qquad (12)$$

and

$$t'_{HK} = I_2(A_{HK})/I_0(A_{HK}) = \varepsilon(\cos 2\varphi_{HK}|A_{HK}). \quad (13)$$

It is to be stressed that the conditional expected value of the cosine, (9), is always positive, since $A_{HK} > 0$.

The quartet

For fixed reciprocal-lattice vectors **L**, **M** and **N**, with the seven magnitudes

$$|E_L|, \quad |E_M|, \quad |E_N|, \quad |E_{L+M+N}|,$$
$$|E_{L+M}|, \quad |E_{M+N}|, \quad |E_{N+L}| \qquad (14)$$

assumed known, the conditional probability distribution of the quartet φ_{LMN} [(5)] is now known (Hauptman, 1975). For our purpose, since N is presumed to be large, it will be sufficient to use the approximation (Giacovazzo, 1976)

$$P(\Phi|B_{LMN}) = [2\pi I_0(B_{LMN})]^{-1} \exp(B_{LMN} \cos \Phi), \qquad (15)$$

where

$$B_{LMN} = (2/N)|E_L E_M E_N E_{L+M+N}|[(|E_{L+M}|^2$$
$$+ |E_{M+N}|^2 + |E_{N+L}|^2) - 2]. \qquad (16)$$

As in (9) and (10), we now find

$$\varepsilon(\cos \Phi_{LMN}|B_{LMN}) = I_1(B_{LMN})/I_0(B_{LMN}) \qquad (17)$$

and

$$\mathrm{var}\,(\cos \Phi_{LMN}|B_{LMN}) = \tfrac{1}{2} + \tfrac{1}{2}[I_2(B_{LMN})/I_0(B_{LMN})]$$
$$- [I_1^2(B_{LMN})/I_0^2(B_{LMN})]. \quad (18)$$

In sharp contrast to (9), the conditional expected value of the cosine [(17)] is now positive or negative according to

$$B_{LMN} \gtrless 0, \qquad (19)$$

i.e., in view of (16), corresponding to the three 'cross terms' $|E_{L+M}|$, $|E_{M+N}|$ and $|E_{N+L}|$ being mostly large or mostly small, respectively. Those quartets for which $B_{LMN} < 0$ are known as negative quartets because their cosines are probably negative. The special importance of the negative quartets will be emphasized subsequently. It is to be stressed that it is only the expected values of the cosines of the negative quartets that are negative; not all cosines of negative quartets are necessarily negative.

The minimal principle

The heuristic background

The mode of the distribution [(7)] is zero and the variance of the cosine [(11)] is small if A_{HK} [(8)] is large. In this way, one obtains the estimate for the triplet φ_{HK} [(4)]:

$$\varphi_{HK} = \varphi_H + \varphi_K + \varphi_{-H-K} \simeq 0, \qquad (20)$$

which is particularly good in the favorable case that A_{HK} [(8)] is large, *i.e.* that $|E_H|$, $|E_K|$ and $|E_{H+K}|$ are all large. The estimate given by (20) is one of the cornerstones of current techniques of direct methods. It is surprising how useful (20) has proven to be in applications, especially since it yields only the zero estimate of the triplet and only those estimates for which $|E_H|$, $|E_K|$ and $|E_{H+K}|$ are all large are reliable. Clearly, the coefficient $2N^{-1/2}$ in (8), and therefore A_{HK} as well, decreases with increasing N, *i.e.* with increasing structural complexity. Hence, the relationship given by (20) becomes increasingly unreliable for larger structures and the traditional direct-methods procedures dependent on (20) eventually fail. However, more recent developments in direct methods, for example those which make explicit use of the Sayre (1952) equation, have largely overcome this limitation of the older techniques and have proved to be more effective in applications (Debaerdemaeker, Tate & Woolfson, 1985, 1988; Woolfson & Yao, 1988, 1990).

By their heavy dependence on the triplet relationship [(20)] (as well as the analogous negative quartet relationship), traditional direct-methods

techniques do not fully exploit our detailed knowledge of the triplet and quartet distributions [(7) and (15)]. We propose now to determine the values of the phases φ in such a way that they generate triplets and quartets which, for each fixed value of A_{HK} or B_{LMN}, have distributions that agree with their theoretical distributions, (7) or (15), respectively. More specifically, one determines the value of a set of phases as those that generate triplets φ_{HK} and quartets φ_{LMN} whose cosines have, for each fixed value of A_{HK} and B_{LMN}, conditional expectation values and variances in agreement with their theoretical values as given by (9), (10), (17) and (18). In connection with this, it should be noted that, for a sufficiently large basis set of phases, say more than 300 phases in the base, the number of structure invariants generated by them exceeds by far (two or three orders of magnitude at least) the number of unknown phases φ. Owing to this great redundancy, a large number of identities among the structure invariants, equal to the difference between the number of structure invariants and the number of phases, must be satisfied. An important aspect of our present formulation is that all identities among the structure invariants, supplemented by the identities (3) among the phases, which must of necessity also hold, will in fact be satisfied.

Triplets

In view of (9) and the previous discussion, one now replaces the zero estimate (20) of the triplet φ_{HK} [(4)] by the estimate

$$\cos \varphi_{HK} \simeq I_1(A_{HK})/I_0(A_{HK}) = t_{HK} \qquad (21)$$

and expects that the smaller the variance [(10)] is, the more reliable the estimate [(21)] will be. If one defines the weight W_{HK} by

$$W_{HK} = A_{HK}, \qquad (22)$$

then W_{HK} is positively correlated with the reciprocal of the variance (10). One is thus led to construct the function

$$R(\varphi) = \sum_{H,K} W_{HK}^{-1} \sum_{H,K} W_{HK}(\cos \varphi_{HK} - t_{HK})^2, \qquad (23)$$

which is seen to be a function of all those triplets φ_{HK} that are generated by a prescribed set of phases $\{\varphi\}$. Recall that, if the basis set of phases $\{\varphi\}$ is sufficiently large, then there are many more structure invariants φ_{HK} than individual phases φ. Thus, in addition to the identities that the individual phases must satisfy, a myriad of identities among the structure invariants must, of necessity, also be satisfied. It is therefore natural to suppose that the best set of values for the structure invariants φ_{HK} is that which minimizes the residual R, (23), subject to the constraint that all identities among the structure invariants are in fact satisfied.

Since the triplets φ_{HK} are defined by (4) as functions of the individual phases φ, (23) defines $R(\varphi)$ implicitly as a function of the individual phases. One therefore naturally anticipates that the best set of values for the individual phases is that which minimizes the residual $R(\varphi)$, (23), now regarded as a function of the individual phases φ provided, as always, that the phases themselves satisfy the required identities [(3)]. The advantage of this formulation is that all identities among the structure invariants will then automatically be satisfied and it is unnecessary to define in further detail what the nature of these identities must be. It remains only to derive the conditions under which this formulation of the so-called minimal principle is valid.

The minimal principle for triplets (Hauptman, 1988, 1989, 1991; Hauptman, Velmurugen & Han, 1990; Hauptman & Han, 1993; DeTitta, Weeks, Thuman, Miller & Hauptman, 1991; Miller et al., 1993; Weeks, DeTitta, Thuman, Miller & Hauptman, 1992; Weeks, DeTitta, Hauptman, Thuman & Miller, 1994; Weeks, DeTitta, Miller & Hauptman, 1993)

From (23), one finds

$$R(\varphi) = \left(\sum_{H,K} W_{HK} \right)^{-1} \sum_{H,K} W_{HK}(\tfrac{1}{2} + \tfrac{1}{2}\cos 2\varphi_{HK}$$
$$- 2t_{HK} \cos \varphi_{HK} + t_{HK}^2). \qquad (24)$$

One first restricts the summations in (24) to those reciprocal-lattice vectors H, K for which A_{HK}, from (8), has a fixed specified value. W_{HK}, from (22), is also fixed under this restricted summation. The expected values t_{HK} of $\cos \varphi_{HK}$, from (12), and t'_{HK} of $\cos 2\varphi_{HK}$, from (13), as functions of A_{HK}, are then also fixed. Two cases are distinguished. In the first case, the phases are assumed to be equal to their true values for some choice of origin and enantiomorph. Then, in view of (12) and (13), R, from (24), becomes (since $\tfrac{1}{2}t'_{HK} < t_{HK}^2$)

$$R_T = \tfrac{1}{2} + \tfrac{1}{2}t'_{HK} - t_{HK}^2 < \tfrac{1}{2}, \qquad (25)$$

where t_{HK} and t'_{HK} are given by (12) and (13), respectively. In the second case, the values of the phases are chosen at random, so that

$$\langle \cos \varphi_{HK} \rangle_{H,K} = \langle \cos 2\varphi_{HK} \rangle_{H,K} = 0. \qquad (26)$$

Now, (24) reduces to

$$R_R = \tfrac{1}{2} + t_{HK}^2 > \tfrac{1}{2}. \qquad (27)$$

In view of (25) and (27),

$$R_T < \tfrac{1}{2} < R_R. \qquad (28)$$

Hence, the residual $R(\varphi)$, equal to R_T when the phases take on their true values, is in fact less than the residual $R(\varphi)$, equal to R_R when the phases are chosen at random, as anticipated by the heuristic argument presented earlier.

It remains only to drop the restriction on the double sums in (24) that requires A_{HK} to be constant over the course of the summation. Again, denoting by R_T the residual $R(\varphi)$ when the phases are equal to their true values and by R_R the residual $R(\varphi)$ when the phases are chosen at random, one now obtains, with (25) and (27),

$$R_T = \tfrac{1}{2} + \left(\sum_{H,K} W_{HK} \right)^{-1} \sum_{H,K} W_{HK}(\tfrac{1}{2} t'_{HK} - t^2_{HK}) < \tfrac{1}{2} \quad (29)$$

and

$$R_R = \tfrac{1}{2} + \left(\sum_{H,K} W_{HK} \right)^{-1} \sum_{H,K} W_{HK} t^2_{HK} > \tfrac{1}{2} \quad (30)$$

so that once again the inequalities

$$R_T < \tfrac{1}{2} < R_R \qquad (28a)$$

hold, where the summations in (24) are now unrestricted. It should be emphasized that (29) and (30) give explicit formulas for calculating R_T and R_R in terms of the observed magnitudes of the normalized structure factors $|E|$ and do not require prior knowledge of the phases, in contrast to (24).

It has to be observed finally that, if the phases are not constrained, then the two cases considered so far are not exhaustive since (26) may fail to hold, for example, not only when the phases take on their true values but also in the singular case that every phase is equal to $+ \varphi$ or $- \varphi$ for some fixed value of φ. In these singular cases, every triplet φ_{HK} reduces simply to

$$\varphi_{HK} = \pm \varphi \quad \text{or} \quad \varphi_{HK} = \pm 3\varphi \qquad (31)$$

so that, in general,

$$\langle \cos \varphi_{HK} \rangle_{H,K} \neq 0 \quad \text{and} \quad \langle \cos 2\varphi_{HK} \rangle_{H,K} \neq 0 \quad (32)$$

since φ is fixed over the averaging process and (26), on which the derivation of R_R is based, fails to hold. There remains open the possibility then that there exist one or more values of φ, that is singular points, leading to residuals $R(\varphi)$ that are less than R_T. Detailed calculations show that this possibility is in fact realized, so that despite the fact that $R_T < R_R$, R_T is not the (unconstrained) global minimum of (23). This observation makes clear the need to force the phases to satisfy the required identities (3).

If one is to determine phases directly by this approach, it is desirable to modify the formalism described here so as to include the quartets. It is anticipated that the positive quartets will not be as useful in this respect since triplets and positive quartets are strongly correlated. It is expected that the negative quartets, on the other hand, will provide the kind of supplemental information that will prove to be useful, since they are strongly dependent on the weak reflections, the very ones that the triplets avoid. This expectation is confirmed subsequently. Never-

theless, the positive quartets may still have a role to play and they are readily incorporated into the formalism if desired.

The minimal principle for negative quartets

In complete analogy with the previous discussion of triplets and with reference to (5) for the definition of the quartet φ_{LMN}, to (16) for the definition of B_{LMN}, and to (17) and (18), one now defines t_{LMN} and t'_{LMN} by

$$t_{LMN} = I_1(B_{LMN})/I_0(B_{LMN})$$
$$= \varepsilon(\cos \varphi_{LMN} | B_{LMN}) \qquad (33)$$

$$t'_{LMN} = I_2(B_{LMN})/I_0(B_{LMN})$$
$$= \varepsilon(\cos 2\varphi_{LMN} | B_{LMN}). \qquad (34)$$

Then, from (18),

$$\text{var}(\cos \varphi_{LMN} | B_{LMN}) = \tfrac{1}{2} + \tfrac{1}{2} t'_{LMN} - t^2_{LMN}. \quad (35)$$

Now, the weight W_{LMN} is defined by

$$W_{LMN} = |B_{LMN}|, \qquad (36)$$

which is seen to be positively correlated with the reciprocal of the variance (35). As before, in the derivation of (23), one is now led to construct the residual

$$R(\varphi) = \left(\sum_{L,M,N} W_{LMN} \right)^{-1} \sum_{L,M,N} W_{LMN}$$
$$\times (\cos \varphi_{LMN} - t_{LMN})^2, \qquad (37)$$

a function of all negative quartets φ_{LMN}, i.e. those for which B_{LMN}, from (16), satisfies the inequality

$$B_{LMN} < 0 \qquad (38)$$

and which are generated by a prescribed set of phases $\{\varphi\}$. It should be stressed that phases φ_L, φ_M, φ_N and φ_{-L-M-N}, belonging to the basis set of phases, generally correspond to the most intense reflections, but the phases φ_{L+M}, φ_{M+N} and φ_{N+L}, associated with the relatively weak magnitudes $|E_{L+M}|$, $|E_{M+N}|$ and $|E_{N+L}|$, the 'cross terms' of the quartet, in general do not belong to the basis set of phases. As before, $R(\varphi)$ may be regarded as a function of phases alone. In analogy with the derivation of (29) and (30), one now finds

$$R_T = \tfrac{1}{2} + \left(1 \Big/ \sum_{L,M,N} W_{LMN} \right)$$
$$\times \sum_{L,M,N} W_{LMN}(\tfrac{1}{2} t'_{LMN} - t^2_{LMN}) < \tfrac{1}{2} \quad (39)$$

and

$$R_R = \tfrac{1}{2} + \left(1 \Big/ \sum_{L,M,N} W_{LMN} \right)$$
$$\times \sum_{L,M,N} W_{LMN} t^2_{LMN} > \tfrac{1}{2}, \qquad (40)$$

where R_T is the value of $R(\varphi)$ in (37) (now regarded as a function of phases) when the phases take on their true values for some choice of origin and enantiomorph and R_R is the value of $R(\varphi)$ when the phases are chosen at random. Hence, as before, in (28a),

$$R_T < \tfrac{1}{2} < R_R. \qquad (41)$$

As before, R_T, the value of $R(\varphi)$ in (37) when every phase is equal to its true value, is the constrained global minimum of (37).

The minimal principle combining triplets and negative quartets

By combining (23) and (37), one is led to the definition of the so-called minimal function:

$$R(\varphi) = \left[\sum_{\mathbf{H,K}} W_{\mathbf{HK}}(\cos\varphi_{\mathbf{HK}} - t_{\mathbf{HK}})^2 \right.$$
$$\left. + \sum_{\mathbf{L,M,N}} W_{\mathbf{LMN}}(\cos\varphi_{\mathbf{LMN}} - t_{\mathbf{LMN}})^2 \right]$$
$$\times \left(\sum_{\mathbf{H,K}} W_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} W_{\mathbf{LMN}} \right)^{-1}, \qquad (42)$$

where the double sum is taken over all reciprocal-lattice vectors $\mathbf{H}$, $\mathbf{K}$ associated with triplets $\varphi_{\mathbf{HK}}$ that are generated by a specified basis set of phases $\{\varphi\}$ corresponding to the largest value of $|E|$ and the triple sum is taken over all reciprocal-lattice vectors $\mathbf{L}$, $\mathbf{M}$, $\mathbf{N}$ corresponding to the negative quartets $\varphi_{\mathbf{LMN}}$ generated by the basis set of phases $\{\varphi\}$. It has already been pointed out that $A_{\mathbf{HK}}$ and $|B_{\mathbf{LMN}}|$ are positively correlated with the reciprocals of the variances of the corresponding conditional distributions (7) and (15). It is for this reason that one defines the weights $W_{\mathbf{HK}}$ and $W_{\mathbf{LMN}}$ by (22) and (36), respectively, and in this way employs the known variances of (7) and (15). Then, provided that the basis set of phases $\{\varphi\}$ is chosen to be sufficiently large, the minimal function $R(\varphi)$ has a constrained global minimum at the point that all the phases are equal to their true values for some choice of origin and enantiomorph (the minimal principle). In this way, the problem of phase determination is replaced by the problem of finding the constrained global minimum of the minimal function $R(\varphi)$, from (42), a known function of the phases. Since the system of equations (3) is redundant, constraining the phases to satisfy these identities is tantamount to fixing their values to be equal to their true values for some choice of origin and enantiomorph. Thus, the minimal principle must be understood to mean that any values for the phases that lead to a value of the minimal function less than R_T, the constrained global

minimum, cannot satisfy the system (3) and hence must differ from the true values of the phases.

Finally, as in the derivations of (29), (39), (30) and (40), one now finds for the minimal function (42) the constrained global minimum

$$R_T = \tfrac{1}{2} + \left(\sum_{\mathbf{H,K}} W_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} W_{\mathbf{LMN}} \right)^{-1}$$
$$\times \left[\sum_{\mathbf{H,K}} W_{\mathbf{HK}}(\tfrac{1}{2} t'_{\mathbf{HK}} - t^2_{\mathbf{HK}}) \right.$$
$$\left. + \sum_{\mathbf{L,M,N}} W_{\mathbf{LMN}}(\tfrac{1}{2} t'_{\mathbf{LMN}} - t^2_{\mathbf{LMN}}) \right] < \tfrac{1}{2} \quad (43)$$

and for R_R

$$R_R = \tfrac{1}{2} + \left(\sum_{\mathbf{H,K}} W_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} W_{\mathbf{LMN}} \right)^{-1}$$
$$\times \left[\sum_{\mathbf{H,K}} W_{\mathbf{HK}} t^2_{\mathbf{HK}} + \sum_{\mathbf{L,M,N}} W_{\mathbf{LMN}} t^2_{\mathbf{LMN}} \right] > \tfrac{1}{2}. \quad (44)$$

It is to be stressed again that R_T and R_R, as given by (43) and (44), depend only on known magnitudes $|E|$ and do not require prior knowledge of the phases φ.

Concluding comments and implementation

Qualifications

A number of qualifications of the preceding arguments need to be made. The first is that, strictly speaking, 'elimination' of the atomic position vectors from the set of equations (1) is only possible when the equations themselves are relevant, *i.e.* a set of ideal normalized structure-factor magnitudes $|E_{\mathbf{h}}|$ is available to infinite precision for an equal-atom non-vibrating point-scattering structure. Therefore, (3) must be clearly understood, in the context of real structures and real data, to be equations of loose constraint, so the equal signs of those equations should be replaced by 'approximately' equal signs. The second is that, even for ideal structures and ideal data, the equations of constraint (3) apply only at that point in many-dimensional space where the unique unknown phases $\varphi_{\mathbf{H}}$ assume their true values. An argument based on the continuity of the exponential function in (1) makes plausible the conjecture that (3) are pretty well satisfied when the n unknown phases $\varphi_{\mathbf{H}}$ are close to their true values, but at present we attempt no rigorous proof of the assertion nor, more importantly, can we speculate on the quantitative breakdown of (3) as the phases move away from their true values. The third is that nothing in the calculation of $R_{\min}$ acts to constrain the phases according to the equations of constraint. In addition to the singular points where all phases take on a special value, there are values of the

phases, notably values close to their true values, but unconstrained by the system (3), that minimize R_{min} but, on the basis of comparison with the true values, have larger average phase errors. The fourth is that the minimization of the weighted mean-square differences between the I_1/I_0 estimates of the *means* of the invariant distributions and the calculated values based on a model force the calculated values towards the means – and not necessarily towards the distribution themselves. Finally, we need to appreciate clearly that the mere existence of the equations of constraint (3) does not suggest ways to impose the constraints. Nothing in the minimization of R_{min} does so. Nonetheless, the conceptual framework suggested by (3) permits us to propose an approximate scheme for the imposition of these constraints and thereby leads us to a practical direct method of structure determination.

The nature of the constraints and the underlying dimensionality of the phase problem

A traditional formulation of direct methods might be something like this: for an N-atom structure, we try to secure the values for $\sim 10N$ phases (invariably those for the $\sim 10N$ largest $|E_h|$ magnitudes); we do so, *via* the tangent formula, by employing $\sim 300N$ three-phase structure invariants assuming they all have $\cos(\varphi_H + \varphi_K + \varphi_L)$ values close to $+1$. By the preceding arguments, we have greatly overdetermined the structure, there being $3N$ unknown positional parameters in three-space. We cast the minimal function originally in terms of invariants $R\{\Phi\}$ involving no constraints at all. If we recast it in terms of phases $R\{\Phi\}$, we automatically impose all the constraints among the structure invariants themselves (Fortier, DeTitta, Fronckowiak, Smith & Hauptman, 1979). These identity relationships among the structure invariants must hold exactly, no matter what the values of the phases, and hence have little phasing power; however, by recasting the process we have reduced the space of the minimization from $\sim 300N$ to $\sim 10N$ dimensions.

Of course, we could attempt to reduce the dimensionality of the problem to its minimum number and recast the minimal function yet again in terms of the $3N$ positional parameters $R\{\Gamma\}$, but there would be a penalty to pay. Unlike $R\{\varphi\}$, which involves the phases φ_H in equations of the type $[t - \cos(\varphi_H + \varphi_K + \varphi_L)]^2$, where the φ's have integer coefficients (± 1) (Han, DeTitta & Hauptman, 1991), $R\{\Gamma\}$ involves the unknowns in equations like the structure-factor equations: transcendental, with the x, y, z's multiplied by h, k, l's and raised to an exponential power. We may as well tackle the phase problem head-on and resort to trial and error, where the function minimized would be the standard crystallographic R

factor. Can we retain the calculability of $R\{\varphi\}$ yet retain the more robust constraints of $R\{\Gamma\}$?

We can proceed as follows. Allow the minimal function to retain its computationally convenient phase form $R\{\varphi\}$. Permit a set of phases $\{\varphi_h\}$ to adjust somewhat to a new set of phases $\{\varphi'\}$ in order that $R\{\varphi'\} < R\{\varphi\}$. Now 'close to' $\{\varphi'\}$ is yet a third set of phases $\{\varphi'_{con}\}$, a set that respects, as much as possible, the equations of constraint (3). We suggest that $\{\varphi'_{con}\}$ can be derived from $\{\varphi'\}$ by a density-modification procedure. Take the adjusted phases $\{\varphi'\}$ that reduce the value of the minimal function and the observed magnitudes $|E_h|$, calculate a Fourier map, select from the map the N maxima we hope will eventually become atomic positions and calculate phases $\{\varphi'_{con}\}$ *via* structure factors. What justifies the conjecture that the $\{\varphi'_{con}\}$ are in any sense 'constrained'? We argue that the $\{\varphi'_{con}\}$ are at least more likely to obey the equations of constraint simply because they result from the model of the crystal structure that underlies the whole probabilistic development – they are derived from a point-scatterer non-vibrating N-atom structural model.

We suggest that a plausible scheme for structure determination relies equally on the two pillars of minimization of $R\{\varphi\}$ in phase space and re-imposition of constraints in real space. If computing were no problem, we could postulate a starting model $\{\Gamma\}$, calculate phases $\{\varphi_{con}\}$, minimize $R\{\varphi\}$ by infinitesimally adjusting the $\{\varphi_{con}\}$ to $\{\varphi'\}$, carry out the Fourier inversion and map interpretation to secure $\{\Gamma'\}$, which is infinitesimally different from $\{\Gamma\}$, calculate, *via* structure factors, $\{\varphi'_{con}\}$ and iterate to convergence. We do not mean to imply that such a minimization itself is a trivial exercise; it is probably beset by the 'local-minimum' problem common to many-dimensional minimization problems that are far from linear in nature. Nonetheless, the prescription yields a set of phases $\{\varphi'_{con}\}$ that are 'atomic' and that yield approximately 'non-negative' electron-density maps at each step of the minimization.

As a practical matter, we in fact allow phases to adjust much more than infinitesimally before we re-impose the constraints *via* Fourier inversion, map interpretation and Fourier back-inversion. Still, we restrict the phase excursions enough for an approximate relationship between $\{\varphi'\}$ and $\{\varphi'_{con}\}$ to be maintained. To fail to do so would be tantamount to generating random phase sets and we present some evidence in an accompanying article (Weeks, DeTitta, Hauptman, Thuman & Miller, 1994) that argues against such an approach.

Applications

The procedures we have described have been implemented in a number of ways for a number of computer architectures, both serial and parallel. In

order to gather statistics that are in any way meaningful, we have largely, but not exclusively, restricted our attention to known, albeit real, crystal structures. [See the accompanying paper by Weeks *et al.* (1994).] A few concluding remarks are in order.

The implementations have been of the multisolution variety, each starting point being defined by a new random structure. Currently, we are searching over hundreds or thousands of starting points and each starting point has been subjected to tens or hundreds of cycles: {structure factor, R minimization, Fourier transformation, map interpretation, ...}. Clearly, to make the process more computationally bearable, we need to focus our attention on three areas. Can we identify promising starting structures *prior* to entering the process? Can we speed up the process itself by employing different minimizing procedures? Can we develop similar but more robust functions of the phases that widen the already very wide radius of convergence of the procedures?

References

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473–478.
DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* **A41**, 286–290.
DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1988). *Acta Cryst.* **A44**, 353–357.
DETITTA, G., WEEKS, C., THUMAN, P., MILLER, R. & HAUPTMAN, H. (1991). Proceedings of the American Crystallographic Association Meeting, Toledo, OH, USA, Abstract 002.
FORTIER, S., DETITTA, G., FRONCKOWIAK, M., SMITH, D. & HAUPTMAN, H. (1979). *Acta Cryst.* **B35**, 2062–2066.
GIACOVAZZO, C. (1976). *Acta Cryst.* **A32**, 91–99.
HAN, F., DETITTA, G. & HAUPTMAN, H. (1991). *Acta Cryst.* **A47**, 484–490.
HAUPTMAN, H. (1975). *Acta Cryst.* **A31**, 680–687.
HAUPTMAN, H. A. (1988). Proceedings of the American Crystallographic Association Meeting, Philadelphia, PA, USA, p. 53, Abstract R4.
HAUPTMAN, H. A. (1989). Proceedings of the American Crystallographic Association Meeting, Seattle, WA, USA, Abstract U3.
HAUPTMAN, H. A. (1991). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 324–332. IUCr/Oxford Univ. Press.
HAUPTMAN, H. & HAN, F. (1993). *Acta Cryst.* **D49**, 3–8.
HAUPTMAN, H., VELMURUGAN, D. & HAN, F. (1990). Proceedings of the American Crystallographic Association Meeting, New Orleans, USA, Abstract T01.
MILLER, R., DETITTA, G., LANGS, D., WEEKS, C., HAUPTMAN, H. & JONES, R. (1993). *Science*, **259**, 1430–1433.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.
WEEKS, C., DETITTA, G., HAUPTMAN, H., THUMAN, P. & MILLER, R. (1994). *Acta Cryst.* **A50**, 210–220.
WEEKS, C. M., DETITTA, G. T., MILLER, R. & HAUPTMAN, H. A. (1993). *Acta Cryst.* **D49**, 179–181.
WEEKS, C., DETITTA, G., THUMAN, P., MILLER, R. & HAUPTMAN, H. (1992). Proceedings of the American Crystallographic Association Meeting, Toledo, OH, USA. Abstract 003.
WOOLFSON, M. M. & YAO, J. (1988). *Acta Cryst.* **A44**, 410–413.
WOOLFSON, M. M. & YAO, J. (1990). *Acta Cryst.* **A46**, 409–413.

Structure Solution by Minimal-Function Phase Refinement and Fourier Filtering. II. Implementation and Applications

BY CHARLES M. WEEKS, GEORGE T. DETITTA AND HERBERT A. HAUPTMAN

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

AND PAMELA THUMAN* AND RUSS MILLER

Department of Computer Science, State University of New York at Buffalo, Buffalo, NY 14260, USA

Abstract

The minimal function, $R(\varphi)$, has been used to provide the basis for a new computer-intensive direct-methods procedure that shows potential for providing fully automatic routine solutions for structures in the 200–400 atom range. This procedure, which has been called *shake-and-bake*, is an iterative process in which real-space filtering is alternated with phase refinement using a technique that reduces the

* Present address: Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA.