

Integration of Patterson Information into Direct Methods. III. The Use of Harker Sections

BY A. ALTOMARE, G. CASCARANO AND C. GIACOVAZZO

Istituto di Ricerca per lo Sviluppo di Metodologie Cristallografiche CNR, c/o Dipartimento Geomineralogico, Campus Universitario, via Salvemini, 70124 Bari, Italy

(Received 18 October 1991; accepted 14 January 1992)

Abstract

A recently described probabilistic approach [Giacovazzo (1991). *Acta Cryst.* A47, 256–263; Altomare, Cascarano & Giacovazzo (1992). *Acta Cryst.* A48, 30–36] is aimed at integrating information contained in Patterson peaks with direct-methods procedures. The approach is here extended to the information contained in Harker sections. It is also shown that, when heavy atoms are present, the structure seminvariants can be readily estimated by a combination of the so-called symmetry function with a special least-squares procedure.

1. Symbols and abbreviations

The papers by Giacovazzo (1991) and Altomare, Cascarano & Giacovazzo (1992) are here referred to as papers I and II respectively. Symbols and abbreviations coincide with those used in those papers: for the sake of brevity they are not listed here. The following additional symbols will be used in this paper:

t_p : number of symmetry-independent atoms for which the location of at least one Harker peak is known;

$$p = mt_p;$$

$t_q = t - t_p$: number of symmetry-independent atoms for which the location of Harker peaks is unknown;

$$q = mt_q;$$

$$\sum_p(\mathbf{h}) = \sum_{j=1}^p f_j^2(\mathbf{h});$$

$$\sum_{3q}(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) = m \sum_{j=t_p+1}^t f_j(\mathbf{h}_1)f_j(\mathbf{h}_2)f_j(\mathbf{h}_3).$$

The atoms are assumed to be in general positions. This simplifies the mathematical treatment but does not involve loss of generality.

2. Introduction

In paper I of this series, a theoretical approach was described aiming at exploiting, by the method of joint probability distribution functions of structure factors, the information provided by a Patterson map. In the absence of any Patterson information, the symmetry-

independent atomic positional vectors were considered as random variables statistically independent of one another and uniformly distributed in the asymmetric unit. If an interatomic vector $\mathbf{u} = \mathbf{r}_{j_1} - \mathbf{r}_{j_2}$ is *a priori* known, \mathbf{r}_{j_2} is no longer considered to be independent since it is completely determined in terms of \mathbf{r}_{j_1} and \mathbf{u} . Standard probabilistic methods for triplet-invariant estimates were therefore modified to take prior information into account.

In paper II, the method was further developed to estimate triplets using the prior information on position and intensity of non-Harker peaks. The theory could not, however, be extended to Harker peaks: indeed in this case prior information about $\mathbf{u} = (\mathbf{C}_{s_1} - \mathbf{C}_{s_2})\mathbf{r}_j$ does not relate different random variables but restricts the single variable \mathbf{r}_j to a suitable domain (a point, line or plane).

The main aim of this paper is to describe a probabilistic approach that is able to exploit information contained in Harker sections for estimating both triplet invariants and one-phase seminvariants of first rank. For simplicity, we will suppose that the location and intensity of some Harker peaks are known, while no information is available about any other type of Patterson information (*i.e.* non-Harker peaks are assumed to be unknown).

3. The average value of $|F|^2$ when one or more Harker peaks are *a priori* known

Let

$$\mathbf{u}_j(s_1, s) = (\mathbf{C}_{s_1} - \mathbf{C}_s)\mathbf{r}_j \quad (1)$$

be the generic Harker vector generated by the atom in \mathbf{r}_j . By convention,

$$\mathbf{u}_j(s_1, 1) = (\mathbf{C}_{s_1} - \mathbf{I})\mathbf{r}_j = (\mathbf{R}_{s_1} - \mathbf{I})\mathbf{r}_j + \mathbf{T}_{s_1}. \quad (2)$$

The complete set of Harker vectors generated by the atom in \mathbf{r}_j will then be written as

$$\mathbf{u}_j(s_1, s), \quad s_1, s = 1, \dots, m. \quad (3)$$

For convenience, we will often make the substitution $\mathbf{C}_{s_1} = \mathbf{C}_s \mathbf{C}_n$ (and therefore $\mathbf{R}_{s_1} = \mathbf{R}_s \mathbf{R}_n$ and $\mathbf{T}_{s_1} = \mathbf{R}_s \mathbf{T}_n + \mathbf{T}_s$) so that

$$\begin{aligned} \mathbf{u}_j(s_1, s) &= \mathbf{C}_s(\mathbf{C}_n - \mathbf{I})\mathbf{r}_j \\ &= \mathbf{R}_s[(\mathbf{R}_n - \mathbf{I})\mathbf{r}_j + \mathbf{T}_n] = \mathbf{R}_s \mathbf{u}_j(n, 1). \end{aligned}$$

Consequently, the complete set of Harker vectors generated by the atom in \mathbf{r}_j will also be written as

$$\mathbf{R}_s \mathbf{u}_j(n, 1), \quad s, n = 1, \dots, m. \quad (4)$$

In (4), the Harker peaks are ordered first according to the n th Harker section [all the peaks $\mathbf{u}_j(n, 1)$ are on the same Harker section] and then according to the symmetry-equivalent sections. Obviously, if $\mathbf{u}_j(n, 1)$ is known, then $\mathbf{R}_s \mathbf{u}_j(n, 1)$, $s = 1, \dots, m$ are also known. However, if $\mathbf{u}_j(n_1, 1)$ is known, that does not necessarily imply that $\mathbf{u}_j(n_2, 1)$ is known, unless they are symmetry equivalent. The average value of $|F_{\mathbf{h}}|^2$ when some symmetry-independent Harker vectors are *a priori* known will be denoted by $\langle |F_{\mathbf{h}}|^2 | \{ \mathbf{u} \} \rangle$.

(a) *Non-centrosymmetric space groups*

In the absence of any information about non-Harker peaks,

$$\begin{aligned} \langle |F_{\mathbf{h}}|^2 | \{ \mathbf{u} \} \rangle &= \varepsilon_{\mathbf{h}} \left\langle \sum_{j_1, j_2=1}^l f_{j_1}(\mathbf{h}) f_{j_2}(\mathbf{h}) \right. \\ &\quad \times \left. \sum_{s_1, s_2=1}^m \exp [2\pi i \mathbf{h} (\mathbf{C}_{s_1} \mathbf{r}_{j_1} - \mathbf{C}_{s_2} \mathbf{r}_{j_2})] \right\rangle \\ &= \varepsilon_{\mathbf{h}} \left\langle \sum_{j=1}^l f_j^2(\mathbf{h}) \right. \\ &\quad \times \left. \sum_{s_1, s_2=1}^m \exp [2\pi i \mathbf{h} (\mathbf{C}_{s_1} - \mathbf{C}_{s_2}) \mathbf{r}_j] \right\rangle. \end{aligned}$$

This can be rearranged to emphasize the contribution from the Harker peaks:

$$\begin{aligned} \langle |F_{\mathbf{h}}|^2 | \{ \mathbf{u} \} \rangle &= \varepsilon_{\mathbf{h}} \left\{ \sum_N(\mathbf{h}) + \sum_{j=1}^{l_p} f_j^2(\mathbf{h}) \sum_n' \sum_{s=1}^m \cos [2\pi \mathbf{h} \mathbf{R}_s \mathbf{u}_j(n, 1)] \right\}. \quad (5) \end{aligned}$$

The primed summation over n runs over those Harker sections that contribute to the set $\{ \mathbf{u} \}$.

(b) *Centrosymmetric space groups*

Here,

$$\begin{aligned} \langle |F_{\mathbf{h}}|^2 | \{ \mathbf{u} \} \rangle &= \varepsilon_{\mathbf{h}} \left\{ 4 \left\langle \sum_{j=1}^l f_j^2(\mathbf{h}) \right. \right. \\ &\quad \times \left. \left. \sum_{s_1, s_2=1}^{m/2} \cos (2\pi \mathbf{h} \mathbf{C}_{s_1} \mathbf{r}_j) \cos (2\pi \mathbf{h} \mathbf{C}_{s_2} \mathbf{r}_j) \right\rangle \right\} \\ &= \varepsilon_{\mathbf{h}} \left\{ \sum_N(\mathbf{h}) + \sum_{j=1}^{l_p} f_j^2(\mathbf{h}) \right. \\ &\quad \times \left. \sum_n' \sum_{s=1}^m \cos [2\pi \mathbf{h} \mathbf{R}_s \mathbf{u}_j(n, 1)] \right\}. \quad (6) \end{aligned}$$

If $l_p = 0$ then $q = N$ and (5) and (6) reduce to the Wilson (1949) result. Information on Harker peaks

may significantly change the expected values of the squared moduli provided $\sum_p(\mathbf{h}) / \sum_N(\mathbf{h})$ is sufficiently large.

4. **The conditional distribution $P(\Phi | \mathbf{R}_{\mathbf{h}_1}, \mathbf{R}_{\mathbf{h}_2}, \mathbf{R}_{\mathbf{h}_3}, \{ \mathbf{u} \})$**

If the set $\{ \mathbf{u} \}$ is *a priori* known then

$$\begin{aligned} \langle F_{\mathbf{h}_1} F_{\mathbf{h}_2} F_{\mathbf{h}_3} | \{ \mathbf{u} \} \rangle &= \left\langle \sum_{j=1}^l f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3) \right. \\ &\quad \times \sum_{s_1, s_2, s_3=1}^m \exp \{ 2\pi i [\mathbf{h}_1 (\mathbf{C}_{s_1} - \mathbf{C}_s) \mathbf{r}_j \\ &\quad \left. + \mathbf{h}_2 (\mathbf{C}_{s_2} - \mathbf{C}_s) \mathbf{r}_j] \} \right\rangle \\ &= \left\langle \sum_{j=1}^l f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3) \right. \\ &\quad \times \sum_{n, \nu, s=1}^m \exp \{ 2\pi i [\mathbf{h}_1 \mathbf{C}_s (\mathbf{C}_n - \mathbf{I}) \mathbf{r}_j \\ &\quad \left. + \mathbf{h}_2 \mathbf{C}_s (\mathbf{C}_\nu - \mathbf{I}) \mathbf{r}_j] \} \right\rangle, \quad (7) \end{aligned}$$

where

$$\begin{aligned} \mathbf{h}_1 + \mathbf{k}_1 + \mathbf{l}_1 &= 0, \\ \mathbf{C}_{s_1} &= \mathbf{C}_s \mathbf{C}_n \end{aligned}$$

and

$$\mathbf{C}_{s_2} = \mathbf{C}_s \mathbf{C}_\nu.$$

Equation (7) may be written in the more useful form

$$\begin{aligned} \langle F_{\mathbf{h}_1} F_{\mathbf{h}_2} F_{\mathbf{h}_3} | \{ \mathbf{u} \} \rangle &= \left\langle \sum_{j=1}^l f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3) \right. \\ &\quad \times \sum_{n, \nu, s=1}^m \exp \{ 2\pi i [\mathbf{h}_1 \mathbf{R}_s \mathbf{u}_j(n, 1) \\ &\quad \left. + \mathbf{h}_2 \mathbf{R}_s \mathbf{u}_j(\nu, 1)] \} \right\rangle. \quad (8) \end{aligned}$$

When $n = \nu = 1$ in (8), the Cochran contribution

$$\sum_3(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) \quad (9)$$

is obtained. The cases $\nu = 1$ (with $n \neq 1$), $n = 1$ (with $\nu \neq 1$) and $n = \nu$ (with $n \neq 1$) provide the supplementary contribution

$$\begin{aligned} &\sum_{j=1}^{l_p} f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3) \\ &\quad \times \sum_n' \sum_{s=1}^m \{ \cos [2\pi \mathbf{h}_1 \mathbf{R}_s \mathbf{u}_j(n, 1)] \\ &\quad + \cos [2\pi \mathbf{h}_2 \mathbf{R}_s \mathbf{u}_j(n, 1)] \\ &\quad + \cos [2\pi \mathbf{h}_3 \mathbf{R}_s \mathbf{u}_j(n, 1)] \}. \quad (10) \end{aligned}$$

A further contribution could arise if we were able to recognize the cases in which two symmetry-independent Harker peaks are generated by the same j th atom. In this case, we should add supplementary terms like

$$\sum'_n \sum'_\nu \sum_j \cos 2\pi[\mathbf{h}_1 \mathbf{R}_s \mathbf{u}_j(n, 1) + \mathbf{h}_2 \mathbf{R}_s \mathbf{u}_j(\nu, 1)].$$

Such a situation is related to the use of the symmetry sum function (Simpson, Dobrott & Lipscomb, 1965; Zimmermann, 1988; Pavelčík, 1989) and will not be considered here. In conclusion,

$$\begin{aligned} \langle F_{\mathbf{h}_1} F_{\mathbf{h}_2} F_{\mathbf{h}_3} | \{\mathbf{u}\} \rangle &= \sum_3 (\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) \\ &+ \sum_{j=1}^{t_p} f_j(\mathbf{h}_1) f_j(\mathbf{h}_2) f_j(\mathbf{h}_3) \\ &\times \sum'_n \sum_{s=1}^m \{ \cos [2\pi \mathbf{h}_1 \mathbf{R}_s \mathbf{u}_j(n, 1)] \\ &+ \cos [2\pi \mathbf{h}_2 \mathbf{R}_s \mathbf{u}_j(n, 1)] \\ &+ \cos [2\pi \mathbf{h}_3 \mathbf{R}_s \mathbf{u}_j(n, 1)] \}. \end{aligned}$$

In accordance with paper I [see equation (1.9)],

$p(\Phi | R_{\mathbf{h}_1}, R_{\mathbf{h}_2}, R_{\mathbf{h}_3}, \{\mathbf{u}\}) \approx [2\pi I_0(G)]^{-1} \exp(G \cos \Phi)$ where

$$G = 2 \langle F_{\mathbf{h}_1} F_{\mathbf{h}_2} F_{\mathbf{h}_3} | \{\mathbf{u}\} \rangle \times [\langle |F_{\mathbf{h}_1}|^2 | \{\mathbf{u}\} \rangle \langle |F_{\mathbf{h}_2}|^2 | \{\mathbf{u}\} \rangle \langle |F_{\mathbf{h}_3}|^2 | \{\mathbf{u}\} \rangle]^{-1/2}.$$

In terms of normalized structure factors and by taking into account the intensity of the Harker peaks, we can write [see paper II and equation (II.17)]

$$G = 2 R_{\mathbf{h}_1} R_{\mathbf{h}_2} R_{\mathbf{h}_3} N^{-1/2} \left\{ 1 + \sum_{\mathbf{u}} \left[I'_{\mathbf{u}} \left(\sum_{i=1}^3 \alpha_i \right) \right] \right\} \times \left\{ \prod_{i=1}^3 \varepsilon_{\mathbf{h}_i} \left[1 + \sum_{\mathbf{u}} (2I'_{\mathbf{u}} \alpha_i) \right] \right\}^{-1/2}, \quad (11)$$

where

$$R_{\mathbf{h}} = F_{\mathbf{h}} \times \langle |F_{\mathbf{h}}|^2 | \{\mathbf{u}\} \rangle^{-1/2},$$

$$\alpha_i = \sum_{s=1}^m \cos 2\pi \mathbf{h}_i \mathbf{R}_s \mathbf{u},$$

$$I'_{\mathbf{u}} = I_{\mathbf{u}} / (j_{\mathbf{u}} I_0).$$

It may therefore be concluded that triplet phases can be estimated *via* the concentration parameter (11) no matter whether prior information concerns Harker or non-Harker peaks. The usefulness of such information will depend on the existence of well localized and relatively large peaks. Thus structures with a few heavy atoms are the best candidates for the application of the theory, while equal-atom structures are out of its range, unless strong peak overlapping occurs.

Since ordinary heavy-atom structures are routinely solved by traditional Patterson or direct methods, the above theory seems best suited to structures having pseudotranslational symmetry of non-ideal type, as described in paper II.

5. The probabilistic estimation of the one-phase seminvariants of the first rank $F_{\mathbf{H}}$ *via* Harker sections

Let $F_{\mathbf{H}}$ be a one-phase structure seminvariant of the first rank (Giacovazzo, 1977, 1978): its vectorial index satisfies the condition $\mathbf{H} = \mathbf{h}(\mathbf{R}_n - \mathbf{I})$ for at least one rotation matrix \mathbf{R}_n . The value of $\varphi_{\mathbf{H}}$ may be estimated in two ways: (a) by applying the theory developed in § 4 to the special triplets (\sum_1 relationships) $\Phi_j = \varphi_{\mathbf{H}} - \varphi_{\mathbf{h}_1} + \varphi_{\mathbf{h}_2, \mathbf{R}_n}$; (b) by applying the algebraic relations between Harker sections and one-phase seminvariants of first rank described by Ardito, Cascarano, Giacovazzo & Luić (1985) and by Cascarano, Giacovazzo, Luić, Pifferi & Spagna (1987). In particular, it was shown that $F_{\mathbf{H}}$ is strictly related to the Fourier transform of the Harker section HS_n generated by the symmetry operator C_n ,

$$F_{\mathbf{H}} \approx \exp(-2\pi i \mathbf{h} \mathbf{T}_n) \tilde{\chi}[HS_n]. \quad (12)$$

Method (a) requires extended supplementary probabilistic considerations which will not be given here. We will focus our attention on (b) to provide the probabilistic background to that method.

Let some Harker vectors corresponding to t_p atoms be *a priori* known. Then

$$\begin{aligned} F_{\mathbf{H}} &= \sum_{j=1}^{t_p} f_j(\mathbf{H}) \sum_{s=1}^m \exp[2\pi i \mathbf{h}(\mathbf{R}_n - \mathbf{I}) C_s \mathbf{r}_j] \\ &+ \sum_{j=t_p+1}^l f_j(\mathbf{H}) \sum_{s=1}^m \exp(2\pi i \mathbf{H} C_s \mathbf{r}_j) \\ &= \exp(-2\pi i \mathbf{h} \mathbf{T}_n) \sum_{j=1}^{t_p} f_j(\mathbf{H}) \\ &\times \sum_{s=1}^m \exp\{2\pi i \mathbf{h} [\mathbf{U}_j(n, s)]\} \\ &+ \sum_{j=t_p+1}^l f_j(\mathbf{H}) \sum_{s=1}^m \exp(2\pi i \mathbf{H} C_s \mathbf{r}_j) \\ &= \exp(-2\pi i \mathbf{h} \mathbf{T}_n) F'_{\rho \mathbf{H}} + F_{q \mathbf{H}}, \quad (13) \end{aligned}$$

where $\mathbf{U}_j(n, s) = (\mathbf{C}_n - \mathbf{I}) C_s \mathbf{r}_j$ is the Harker vector in the section HS_n generated by the atom $C_s \mathbf{r}_j$. It is worthwhile emphasizing that the vectors $\mathbf{U}_j(n, s)$, $s = 1, \dots, m$, all lie in HS_n (they are the projections on HS_n of the vectors $C_s \mathbf{r}_j$). Thus the usual structure-factor algebra cannot be used for estimating $F'_{\rho \mathbf{H}}$ [$\mathbf{U}_j(n, 1)$ is not necessarily symmetry-equivalent to $\mathbf{U}_j(n, s)$ except when C_n commutes with all C_s operators. In this last case (see Appendix),

$$F'_{\rho \mathbf{H}} = \sum_{j=1}^{t_p} f_j(\mathbf{H}) \sum_{s=1}^m \exp[2\pi i \mathbf{h} C_s \mathbf{U}_j(n, 1)].$$

From now on, and in accordance with the above observations, F_{pH} represents the contribution from the t_p atoms generated by those symmetry operators for which $U_j(n, s)$ are actually known, while F_{qH} is the difference $F_H - \exp(-2\pi i \mathbf{h} \mathbf{T}_n) F_{pH}$. Under the above hypotheses, the characteristic function of F_H is

$$\begin{aligned} C(v_1, v_2) &= \langle \exp [i(v_1 A_H + v_2 B_H)] \rangle \\ &= \langle \exp \{i[v_1(A_{pH} + A_{qH}) + v_2(B_{pH} + B_{qH})]\} \rangle \\ &= \exp [i(v_1 A_{pH} + v_2 B_{pH})] \\ &\quad \times \langle \exp [i(v_1 A_{qH} + v_2 B_{qH})] \rangle, \end{aligned}$$

where v_1 and v_2 are two carrying variables; A_{pH} and A_{qH} are the real parts of $F_{pH} \exp(-2\pi i \mathbf{h} \mathbf{T}_n)$ and of F_{qH} respectively; B_{pH} and B_{qH} are the corresponding imaginary parts. In accordance with Giacovazzo (1983) we obtain

$$\begin{aligned} P(A_H, B_H | A_{pH}, B_{pH}) \\ \approx \frac{1}{\pi \sum_q (\mathbf{H})} \exp \left\{ -\frac{(A_H - A_{pH})^2}{\sum_q (\mathbf{H})} - \frac{(B_H - B_{pH})^2}{\sum_q (\mathbf{H})} \right\} \end{aligned}$$

or, in polar form,

$$\begin{aligned} P(|F_H|, \varphi_H | F_{pH}, \varphi_{pH}) \\ \approx \frac{F_H}{\pi \sum_q} \exp \left\{ -\frac{1}{\sum_q} [|F_H|^2 + |F_{pH}|^2 \right. \\ \left. - 2|F_H F_{pH}| \cos(\varphi_H - \varphi_{pH} + 2\pi \mathbf{h} \mathbf{T}_n)] \right\}, \end{aligned}$$

where $\sum_q (\mathbf{H})$ is the average scattering value corresponding to the q atoms. Then, in terms of pseudo-normalized structure factors $E' = F/\sum_q^{1/2}$,

$$\begin{aligned} p(\varphi_H | E'_{pH}, \varphi_{pH}) \\ \approx [2\pi I_0(G)]^{-1} \\ \times \exp \{ G \cos[\varphi_H - (\varphi_{pH} - 2\pi \mathbf{h} \mathbf{T}_n)] \}, \quad (14) \end{aligned}$$

$$G = 2|E'_H E'_{pH}|. \quad (15)$$

Equation (14) is a von Mises distribution and $\varphi_{pH} - 2\pi \mathbf{h} \mathbf{T}_n$ is the expected value of φ_H . Equation (14) agrees with the formula intuitively suggested by Cascarano, Giacovazzo, Luić, Pifferi & Spagna (1987).

6. A modified procedure

In the paper by Cascarano, Giacovazzo, Luić, Pifferi & Spagna (1987), the following practical procedure was proposed for the estimation of the one-phase structure seminvariants of the first rank

$$\mathbf{H} = \mathbf{h}(\mathbf{I} - \mathbf{R}_n). \quad (16)$$

(i) The Harker section HS_n is calculated and a suitable list of symmetry-independent vectors U_j are obtained by application of a peak-search routine.

(ii) The atomic pseudo-positions \mathbf{r}'_j are calculated by solving the equation

$$\mathbf{r}'_j = \mathbf{D}_n^*(U_j + \mathbf{T}_n) + (\mathbf{I} - \mathbf{D}_n^* \mathbf{D}_r) \mathbf{v},$$

where \mathbf{v} is a free vector, $\mathbf{D}_n = (\mathbf{I} - \mathbf{R}_n)$, \mathbf{D}_n^* is the reflexive generalized inverse of \mathbf{D}_n . \mathbf{r}'_j cannot be considered to be the true positional vector of the j th atom but only a representative of the subset of the positional atomic vectors compatible with U_j .

(iii) the function

$$F'_H = \sum_j w_j f_j \sum_s \exp(2\pi i \mathbf{H} \mathbf{C}_s \mathbf{r}'_j) \quad (17)$$

is calculated, where \mathbf{H} satisfies (16) via the same operator \mathbf{C}_n used for the choice of the Harker section, f_j is the scattering factor of the heaviest atomic species and w_j is an occupancy factor, the role of which is described below.

(iv) A special least-squares procedure is applied that minimizes

$$s = \sum (|F_H|_{\text{obs}} - |F'_H|)^2$$

by optimizing, for each atom, w , \mathbf{r}' and the isotropic thermal parameter B . It is assumed that $w_j = (I_j/I_{\text{max}})^{1/2}$, where I_j is the intensity of the j th peak and I_{max} is the maximum intensity on the Harker section. Small residual values

$$\text{Res} = \sum ||F_H|_{\text{obs}} - |F'_H|| / |F_H|_{\text{obs}}$$

at the end of the procedure denote high proficiency of the process.

(v) The reliability of each structure seminvariant is calculated on the basis of the concentration parameter (15).

The above method has been successfully applied to crystal structures with few dominant heavy atoms and to structures with medium-weight atoms (S, Cl, Fe etc.).

A more efficient procedure (see also some suggestions by Pavelčík, 1990) can be introduced according to the scheme below.

(a) The steps (i) and (ii) of the old procedure are replaced by the calculation of the symmetry minimum function (Simpson, Dobrott & Lipscomb, 1965; Zimmermann, 1988; Pavelčík, 1988, 1990)

$$Q(\mathbf{r}) = \left[\min_{n=1}^m \frac{1}{p_n} P(\mathbf{r} - \mathbf{C}_n \mathbf{r}) \right]^{1/2},$$

where P is the Patterson function and p_n is the multiplicity of the symmetry interaction. Q is a combination of all the information contained in the Harker sections (including inversion and improper rotation peaks, which were not taken into account in the old procedure). The maxima of Q should provide the possible atomic positions which satisfy the Harker 'sections'.

(b) The special least-squares process described in step (iv) of the procedure is applied to the largest

peaks to recognize spurious atoms and is followed by Fourier syntheses to introduce new atomic positions.

The pseudo-structure so obtained may be used for the more accurate estimation of the structure seminvariants (to be used in direct-methods approaches) or for a safer application of the image-seeking functions to elucidate the correct crystal structure.

7. Applications

In Table 1, the main data for six crystal structures are given. In Table 2, for each structure, the following data are shown: the atomic species that are considered heavy, the number of reflections with largest value of $|E|$ chosen by SIR88 (Burla, Camalli, Cascarano, Giacovazzo, Polidori, Spagna & Viterbo, 1989) as active in the phasing process, the number of one-phase structure seminvariants of the first rank included in them and the corresponding number of one-phase structure seminvariants of second rank.

At the end of the procedure, the Res values given in Table 3 are obtained. All of them are sufficiently small, denoting high proficiency of the method, except for TRICL where Res = 0.46. Accordingly, all the SS1(1) included in the NRIF reflections for RH5, FREIES, CUPP, AGI and BAVO were correctly estimated, while four errors were made for TRICL (see error in Table 3).

A more complete grasp of the efficiency of the procedure can be obtained by considering in Table 3 the number of errors relative to all the one-phase seminvariants included in the overall set of measured reflections. Again, the results are quite satisfactory, except for TRICL where the large Res value hinders reliable estimates.

The method also allows the estimation in $P2_12_1$ of the non-centrosymmetric one-phase structure seminvariants of second rank, provided the residual Res is sufficiently small. Since the probabilistic formula (14) is of von Mises type, reliable probabilistic estimates can be obtained for phases marked by large values of G [as given by (15)]. In Table 4, the list of SS1(2) seminvariants for BAVO is given, together with their true and estimated phases.

Tests made on several equal-atom structures were in general not successful. The main reason is the unsatisfactory quality of their Patterson maps: false or misplaced peaks were present in too high percentage among the subset of the largest intensity peaks lying in the Harker section. Several methods for improving the location of peaks and for excluding false ones were tested but none proved successful. When this problem is solved, the procedure described here should have a greater impact on the phasing process.

Table 1. Code name, space group, crystallochemical data and references for test structures

Structure code	Space group	Molecular formula	Z	Reference
RH5	$P4/n$	$Rh_2AgP_6C_3H_{84}O_3F_3S$	4	(1)
FREIES	$P2_1/a$	$PbAgSbS_3$	4	(2)
CUPP	$P2_1/n$	$Cu_2Br_2P_4C_{88}H_{68}$	4	(3)
AGI	$P2_1/c$	$Ag_2I_2P_4C_{64}H_{56}$	4	(4)
TRICL	$P2_12_1$	$C_{24}H_{16}Cl_3N_3O_3$	4	(5)
BAVO	$P2_12_1$	$Ba(VO_3)H_2O$	4	(6)

References: (1) Bachechi, Ott & Venanzi (1986); (2) Ito & Novacki (1974); (3) Camalli, Caruso & Venanzi (1986); (4) Camalli, Caruso & Venanzi (1985); (5) Kobelt, Paulus & Kunstmann (1974); (6) Ulická, Pavelčík & Hulm (1987).

Table 2. Test-structure data

For each test structure, the following data are shown: the code name; the atomic species (S_p) that are considered heavy; the number (NRIF) of reflections with largest $|E|$ chosen by SIR88 as active in the process; the number of one-phase structure seminvariants of the first rank [SS1(1)] included in NRIF; the corresponding number of one-phase structure seminvariants of second rank [SS1(2)].

Code	S_p	NRIF	SS1(1)	SS1(2)
RH5	Rh, Ag	249	92	-
FREIES	Pb, Ag, Sb	175	48	-
CUPP	Cu, Br	499	86	-
AGI	Ag, I	468	60	-
TRICL	Cl	196	22	19
BAVO	Ba, V	144	28	6

8. Concluding remarks

The integration of Patterson information into direct-methods procedures has been extended to include the information contained in the Harker sections. It has been shown that the same formalism introduced in papers I and II for taking into account information provided by non-Harker peaks is also valid for Harker peaks. A combination of the symmetry function with direct methods has also been suggested, which proved able to estimate reliably one-phase structure seminvariants when some heavy atoms are present.

Thanks are due to Professor G. D. Andreotti, Dr C. Rizzoli and Dr F. Pavelčík who allowed us to use the PC version of the program *XFPS* originally written by Dr F. Pavelčík.

APPENDIX

In (13), F'_{pH} can be approximated as

$$F'_{pH} \approx \langle f(\mathbf{H})/f(\mathbf{h}) \rangle \sum_{j=1}^{l_p} f_j(\mathbf{h}) \sum_{s=1}^m \exp[2\pi i \mathbf{h} \cdot \mathbf{U}_j(n, s)], \quad (\text{A.1})$$

where $\langle f(\mathbf{H})/f(\mathbf{h}) \rangle = \langle f^0(\mathbf{H})/f^0(\mathbf{h}) \rangle \exp[-B(s_H^2 - s_h^2)]$ is an average value that can be estimated as soon as the overall thermal factor is available from a Wilson

Table 3. *Test-structure results*

For each structure we give: the value of the final residual Res; the number of wrong estimates (error) for the one-phase seminvariants of first rank [SS1(1)] included in the NRIF reflections; the number of errors (Error_{tot}) for one-phase seminvariants of first rank [SS1(1)_{tot}] is their number: if SS1(1)_{tot} > 1000, the calculations are limited to the first 1000] included in the NRIF_{tot} measured reflections.

	Res	SS1(1)	Error	NRIF _{tot}	SS1(1) _{tot}	Error _{tot}
RH5	0.25	92	0	1420	441	13
FREIES	0.26	48	0	972	148	44
CUPP	0.27	86	0	7455	1000	47
AGI	0.27	60	0	11463	1000	45
TRICL	0.46	22	4	2425	196	51
BAVO	0.07	28	0	620	57	1

Table 4. *Results for BAVO*

For each one-phase seminvariant of the second rank for BAVO the $|E|$ values, the true and the estimated phases and the G values are given.

$h\ k\ l$	$ E $	φ_{true}	φ_{calc}	G
2 6 2	1.8	192	171	2.95
2 2 4	1.41	6	9	1.92
8 2 2	1.30	179	181	1.84
2 4 4	1.29	6	-19	1.71
8 4 4	1.25	333	18	1.45
4 6 2	1.23	357	8	1.38

scaling procedure ($s = \sin \theta / \lambda$). Thus, any reflection \mathbf{h} satisfying $\mathbf{h}(\mathbf{I} - \mathbf{R}_n) = \mathbf{H}$ and the Harker vectors \mathbf{U}_j can be used directly for calculating $F'_{\rho\mathbf{H}}$ and then for estimating $F_{\mathbf{H}}$.

If \mathbf{C}_n commutes with all \mathbf{C}_s operators then $(\mathbf{R}_n - \mathbf{I})\mathbf{C}_s\mathbf{r}_j \equiv \mathbf{C}_s(\mathbf{R}_n - \mathbf{I})\mathbf{r}_j = \mathbf{C}_s\mathbf{U}_j(n, 1)$ and the usual algebra of the structure factors can be used. Thus (A.1) reduces to

$$F'_{\rho\mathbf{H}} \approx \langle f(\mathbf{H}) / f(\mathbf{h}) \rangle F'_{\rho\mathbf{h}}, \quad (\text{A.2})$$

where $F'_{\rho\mathbf{h}}$ is a structure factor of index \mathbf{h} in which the atomic positions are replaced by the interatomic vectors \mathbf{U}_j .

Acta Cryst. (1992). **A48**, 500–508

The Enumeration and Symmetry-Significant Properties of Derivative Lattices

BY JOHN S. RUTHERFORD

Department of Chemistry, University of Transkei, Private Bag X1, Umtata, Transkei, South Africa

(Received 20 January 1991; accepted 20 January 1992)

Abstract

For a lattice in two dimensions, the number of distinct derivative lattices of index n is given by the arithmetic function $\sigma_1(n)$ which is the sum of the divisors of n , including 1 and n . The function $\sigma_1(n)$ has as its generating function the Dirichlet series $\zeta(s)\zeta(s-1)$

References

- ALTOMARE, A., CASCARANO, G. & GIACOZZO, C. (1992). *Acta Cryst.* **A48**, 30–36.
- ARDITO, G., CASCARANO, G., GIACOZZO, C. & LUIĆ, M. (1985). *Z. Kristallogr.* **172**, 25–34.
- BACHECHI, F., OTT, J. & VENANZI, L. M. (1986). Unpublished.
- BURLA, M. C., CAMALLI, M., CASCARANO, G., GIACOZZO, C., POLIDORI, G., SPAGNA, R. & VITERBO, D. (1989). *J. Appl. Cryst.* **22**, 389–393.
- CAMALLI, M., CARUSO, F. & VENANZI, L. M. (1985). Unpublished.
- CAMALLI, M., CARUSO, F. & VENANZI, L. M. (1986). Unpublished.
- CASCARANO, G., GIACOZZO, C., LUIĆ, M., PIFFERI, A. & SPAGNA, R. (1987). *Z. Kristallogr.* **179**, 113–125.
- GIACOZZO, C. (1977). *Acta Cryst.* **A33**, 933–944.
- GIACOZZO, C. (1978). *Acta Cryst.* **A34**, 562–572.
- GIACOZZO, C. (1983). *Acta Cryst.* **A39**, 685–692.
- GIACOZZO, C. (1991). *Acta Cryst.* **A47**, 256–263.
- ITO, T. & NOVACKI, W. (1974). *Z. Kristallogr.* **139**, 85–102.
- KOBELT, D., PAULUS, E. F. & KUNSTMANN, W. (1974). *Z. Kristallogr.* **139**, 15–32.
- PAVELČÍK, F. (1988). *Acta Cryst.* **A44**, 724–729.
- PAVELČÍK, F. (1989). *J. Appl. Cryst.* **22**, 181–182.
- PAVELČÍK, F. (1990). *J. Appl. Cryst.* **23**, 225–227.
- SIMPSON, P. G., DOBROTT, D. & LIPSCOMB, W. N. (1965). *Acta Cryst.* **18**, 169–179.
- ULICKÁ, L., PAVELČÍK, F. & HULM, K. (1987). *Acta Cryst.* **C43**, 2266–2268.
- WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
- ZIMMERMANN, H. (1988). *Z. Kristallogr.* **183**, 113–122.

where $\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$ is the Riemann zeta function. That is, $\zeta(s)\zeta(s-1) = \sum_{n=1}^{\infty} \sigma_1(n)n^{-s}$. The probability that s points chosen at random on the two-dimensional lattice do not lie on any of the derivative lattices so enumerated is therefore $[\zeta(s)\zeta(s-1)]^{-1}$. The equivalent results in three dimensions are: the arithmetic function $\sum_{d|n} [n/d]^2 \sigma_1(d)$, where the sum