*Acta Cryst.* (1996). A**52**, 56–61

# Probabilistic Distribution of One-Phase Structure Seminvariants for an Isomorphous Pair of Structures: Theoretical Basis and Initial Applications

YONG-SHENG LIU AND NING-HAI HU

*Applied Spectroscopy Laboratory, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, People's Republic of China*

*(Received 3 April 1995; accepted 10 July 1995)*


## Abstract

Given a special type of triplet of reciprocal-lattice vectors in the monoclinic and orthorhombic systems, there exist eight three-phase structure seminvariants (3PSSs) for a pair of isomorphous structures. The first neighborhood of each of these 3PSSs is defined by the six magnitudes and the joint probability distribution of the corresponding six structure factors is derived according to Hauptman's neighborhood principle. This distribution leads to the conditional probability distribution of each of the 3PSSs, assuming as known the six magnitudes in its first neighborhood. The conditional probability distributions can be directly used to yield the reliable estimates (0 or $\pi$) of the one-phase structure seminvariants (1PSSs) in the favorable case that the variances of the distributions happen to be small [Hauptman (1975). *Acta Cryst.* A**31**, 680–687]. The relevant parameters in the formulas for the monoclinic and orthorhombic systems are given in a tabular form. The applications suggest that the method is efficient for estimating the 1PSSs with values of 0 or $\pi$.


## 1. Introduction

The procedures of crystal structure determination have been traditionally divided into two significantly different techniques, those for small molecules and those for macromolecules. In about the past ten or more years, it has been shown that both techniques, when properly integrated, could lead to more powerful methods of structure determination. Hauptman (1982*a*,*b*) successfully realized the fusion of direct methods with isomorphous replacement (IR) as well as anomalous scattering (AS), and presented the probabilistic theory of two- and three-phase structure invariants (2PSIs and 3PSIs) for both IR and AS cases. The first applications of the resultant formulas led to an enormous increase in the numbers of invariants whose values may be estimated reliably no matter what the fusion is. As a result, combination of the techniques of direct methods with IR or AS is increasingly facilitated. Many related papers have been published. Most of the problems in these studies concerned only structure-invariant estimates. For structure seminvariants, Velmurugan & Hauptman (1989), using the neighborhood principle for integrating direct methods with AS, derived the conditional probability distribution of 1PSSs having values 0 or $\pi$ through the joint probability distribution by embedding the 1PSSs into the 3PSIs and gave results of its applications (Velmurugan, Hauptman & Potter, 1989). Recently, we extended the theory of 2PSIs developed by Hauptman to two-phase structure seminvariants (2PSSs) in the AS case and, further, estimation of the 1PSSs (near 0 or $\pi$) was realized by combining the formula of the 2PSSs with Cochran's distribution (Liu & Hu, 1994). In addition, we also advanced another method to estimate 1PSSs by integrating Hauptman's theory of 3PSIs and the $\sum_1$ relationship for an isomorphous pair of structures (Hu & Liu, 1995). It should be noticed that the latter method is also suitable for the AS case. In the present paper, we show again that the integration of direct methods and IR would improve the procedures for structure seminvariant estimates.

Assume that the number of equivalent positions is $m$ for a given space group. Let

$$\Psi_n = \sum_{j=1}^{n} \Phi_{\mathbf{H}_j}$$

be an $n$-phase structure seminvariant, then the following formula has to hold:

$$\left(\sum_{j=1}^{n} \mathbf{H}_j\right) \cdot \mathbf{r}_0 = N, \qquad (1)$$

where $N$ is a positive, null or negative integer, $\mathbf{r}_0$ is the position of permissible origins (Hauptman & Karle, 1956), which depends on the rotation matrixes $\mathbf{R}_s$ of the space group (Giacovazzo, 1980):

$$(\mathbf{R}_s - \mathbf{I}) \cdot \mathbf{r}_0 = \mathbf{V}, \quad s = 1, 2, \ldots, m,$$

where $\mathbf{I}$ is a $3 \times 3$ identity matrix and $\mathbf{V}$ a vector with zero or integer components. It follows that the $n$-phase structure seminvariant $\Psi_n$ depends on the rotation matrixes $\mathbf{R}_s$. Therefore, in order to derive the formulas for estimating 1PSSs, it is possible first to construct a 3PSS and then to expand its probability distribution in terms of Hauptman's neighborhood principle, as mentioned below.



*Acta Crystallographica Section A*
ISSN 0108-7673 ©1996

For a pair of isomorphous structures in the monoclinic or orthorhombic systems, when all atoms are in general positions, the respective normalized structure factors $E_{\mathbf{H}}$ and $G_{\mathbf{H}}$ are defined by

$$E_{\mathbf{H}} = |E_{\mathbf{H}}|\exp(i\varphi_{\mathbf{H}}) = \alpha_{20}^{-1/2}\sum_{i=1}^{N/m} f_i \sum_{j=1}^{m}\exp(i2\pi\mathbf{H}\mathbf{C}_j\cdot\mathbf{r}_i), \tag{2}$$

$$G_{\mathbf{H}} = |G_{\mathbf{H}}|\exp(i\psi_{\mathbf{H}}) = \alpha_{02}^{-1/2}\sum_{i=1}^{N/m} g_i \sum_{j=1}^{m}\exp(i2\pi\mathbf{H}\mathbf{C}_j\cdot\mathbf{r}_i), \tag{3}$$

where

$$\alpha_{mn} = \sum_{j=1}^{N} f_j^m g_j^n, \tag{4}$$

$$\mathbf{C}_j\cdot\mathbf{r} = \mathbf{R}_j\cdot\mathbf{r} + \mathbf{T}_j, \quad j = 1, 2, \ldots, m, \tag{5}$$

$N$ is the number of atoms in the unit cell, $\mathbf{r}_i$ is the position vector of the atom labeled $i$, the $f_i$ and $g_i$ are zero-angle atomic scattering factors and therefore are equal to the atomic number $Z_i$ and $\mathbf{C}_j$ denotes the symmetry operator of the $j$th equivalent position in the space group that contains a corresponding rotational component $\mathbf{R}_j$ and a translational component $\mathbf{T}_j$.

There are four kinds of 1PSS with values 0 or $\pi$ in the monoclinic and orthorhombic systems: $\varphi_{2h,0,2l}$, $\varphi_{2h,2k,0}$, $\varphi_{0,2k,2l}$, $\varphi_{0,2k,0}$, which together are indicated by $\varphi_{\mathbf{H}_s}$. For a special type of triplet of reciprocal-lattice vectors, $\mathbf{H} = hkl$, $\bar{\mathbf{H}} = \bar{h}k\bar{l}$ and $\mathbf{H}_s$, which satisfy (1), there are eight 3PSSs for an isomorphous pair of structures:

$$\begin{aligned}
\omega_1 &= \varphi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} + \varphi_{\mathbf{H}_s}, & \omega_5 &= \psi_{\mathbf{H}} + \psi_{\bar{\mathbf{H}}} + \psi_{\mathbf{H}_s},\\
\omega_2 &= \psi_{\mathbf{H}} + \psi_{\bar{\mathbf{H}}} + \varphi_{\mathbf{H}_s}, & \omega_6 &= \varphi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} + \psi_{\mathbf{H}_s},\\
\omega_3 &= \varphi_{\mathbf{H}} + \psi_{\bar{\mathbf{H}}} + \varphi_{\mathbf{H}_s}, & \omega_7 &= \psi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} + \psi_{\mathbf{H}_s},\\
\omega_4 &= \psi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} + \varphi_{\mathbf{H}_s}, & \omega_8 &= \varphi_{\mathbf{H}} + \psi_{\bar{\mathbf{H}}} + \psi_{\mathbf{H}_s}.
\end{aligned}$$

The first neighborhood of each of the eight 3PSSs is defined to consist of the six magnitudes $|E_{\mathbf{H}}|$, $|E_{\bar{\mathbf{H}}}|$, $|E_{\mathbf{H}_s}|$, $|G_{\mathbf{H}}|$, $|G_{\bar{\mathbf{H}}}|$, $|G_{\mathbf{H}_s}|$. The formulas for estimating 1PSSs with values 0 or $\pi$ are obtained ' y the derivation of the probabilistic distribution of these 3PSSs based on Hauptman's neighborhood principle (Hauptman, 1975*a*,*b*) with some differences in detail. Owing to limitations of space, only the basics of the derivations are given here.

## 2. The probabilistic distribution of the 3PSSs

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

It will again be assumed throughout that the number of equivalent positions is $m$ in a given monoclinic or orthorhombic space group for a pair of isomorphous structures. Their normalized structure factors $E$ and $G$ are defined by (2)–(5). Suppose that the independent atomic position vectors $\mathbf{r}_i$, $i = 1, 2, \ldots, N/m$, are fixed and that the primitive random variable is the ordered triple $(\mathbf{h}, \mathbf{k}, \mathbf{l})$ of reciprocal vectors, which is assumed to be uniformly distributed over the subset (1) of the threefold Cartesian product $S\times S\times S$ ($S$ denotes reciprocal space). Then the structure factors $E_{\mathbf{H}}$, $E_{\bar{\mathbf{H}}}$, $E_{\mathbf{H}_s}$, $G_{\mathbf{H}}$, $G_{\bar{\mathbf{H}}}$, $G_{\mathbf{H}_s}$ are functions of the primitive random variables $\mathbf{h}$, $\mathbf{k}$, $\mathbf{l}$, so that they are themselves random variables. Denote by $P = P(R_1, R_2, R_3, S_1, S_2, S_3; \Phi_1, \Phi_2, \Phi_3, \Psi_1, \Psi_2, \Psi_3)$ the joint probability distribution of the magnitudes $|E_{\mathbf{H}}|$, $|E_{\bar{\mathbf{H}}}|$, $|E_{\mathbf{H}_s}|$, $|G_{\mathbf{H}}|$, $|G_{\bar{\mathbf{H}}}|$, $|G_{\mathbf{H}_s}|$ and the phases $\varphi_{\mathbf{H}}$, $\varphi_{\bar{\mathbf{H}}}$, $\varphi_{\mathbf{H}_s}$, $\psi_{\mathbf{H}}$, $\psi_{\bar{\mathbf{H}}}$, $\psi_{\mathbf{H}_s}$ of the complex normalized structure factors $E_{\mathbf{H}}$, $E_{\bar{\mathbf{H}}}$, $E_{\mathbf{H}_s}$, $G_{\mathbf{H}}$, $G_{\bar{\mathbf{H}}}$, $G_{\mathbf{H}_s}$. Then $P$ is given by the 12-fold integral (Karle & Hauptman, 1958)

$$\begin{aligned}
P = \prod_{k=1}^{3}\Bigg([R_kS_k/(2\pi)^4]\int_{\rho_k=0}^{\infty}\int_{\sigma_k=0}^{\infty}\int_{\theta_k=0}^{2\pi}\int_{\chi_k=0}^{2\pi}\rho_k\sigma_k\\
\times\exp\{-i[R_k\rho_k\cos(\theta_k-\Phi_k) + S_k\sigma_k\cos(\chi_k-\Psi_k)]\}\Bigg)\\
\times\prod_{j=1}^{N/m} q_j\prod_{k=1}^{3}(\mathrm{d}\rho_k\mathrm{d}\sigma_k\mathrm{d}\theta_k\mathrm{d}\chi_k),
\end{aligned} \tag{6}$$

where

$$\begin{aligned}
q_j &= q_j(\rho_1, \rho_2, \rho_3, \sigma_1, \sigma_2, \sigma_3; \theta_1, \theta_2, \theta_3, \chi_1, \chi_2, \chi_3)\\
&= \langle\exp\{(if_j/\alpha_{20}^{1/2})[\rho_1\cos(2\pi\mathbf{H}\cdot\mathbf{r}_j - \theta_1)\\
&\quad + \rho_1\cos(2\pi\mathbf{H}\mathbf{R}_2\cdot\mathbf{r}_j + 2\pi\mathbf{H}\cdot\mathbf{T}_2 - \theta_1) + \ldots\\
&\quad + \rho_1\cos(2\pi\mathbf{H}\mathbf{R}_m\cdot\mathbf{r}_j + 2\pi\mathbf{H}\cdot\mathbf{T}_m - \theta_1)\\
&\quad + \rho_2\cos(2\pi\bar{\mathbf{H}}\cdot\mathbf{r}_j - \theta_2) + \rho_2\cos(2\pi\bar{\mathbf{H}}\mathbf{R}_2\cdot\mathbf{r}_j\\
&\quad + 2\pi\mathbf{H}\cdot\mathbf{T}_2 - \theta_2) + \ldots + \rho_2\cos(2\pi\bar{\mathbf{H}}\mathbf{R}_m\cdot\mathbf{r}_j\\
&\quad + 2\pi\mathbf{H}\cdot\mathbf{T}_m - \theta_2) + \rho_3\cos(2\pi\mathbf{H}_s\cdot\mathbf{r}_j - \theta_3)\\
&\quad + \rho_3\cos(2\pi\mathbf{H}_s\mathbf{R}_2\cdot\mathbf{r}_j + 2\pi\mathbf{H}_s\cdot\mathbf{T}_2 - \theta_3) + \ldots\\
&\quad + \rho_3\cos(2\pi\mathbf{H}_s\mathbf{R}_m\cdot\mathbf{r}_j + 2\pi\mathbf{H}_s\cdot\mathbf{T}_m - \theta_3)]\\
&\quad + (ig_j/\alpha_{02}^{1/2})[\sigma_1\cos(2\pi\mathbf{H}\cdot\mathbf{r}_j - \chi_1)\\
&\quad + \sigma_1\cos(2\pi\mathbf{H}\mathbf{R}_2\cdot\mathbf{r}_j + 2\pi\mathbf{H}\cdot\mathbf{T}_2 - \chi_1) + \ldots\\
&\quad + \sigma_1\cos(2\pi\mathbf{H}\mathbf{R}_m\cdot\mathbf{r}_j + 2\pi\mathbf{H}\cdot\mathbf{T}_m - \chi_1)\\
&\quad + \sigma_2\cos(2\pi\bar{\mathbf{H}}\cdot\mathbf{r}_j - \chi_2) + \sigma_2\cos(2\pi\bar{\mathbf{H}}\mathbf{R}_2\cdot\mathbf{r}_j\\
&\quad + 2\pi\mathbf{H}\cdot\mathbf{T}_2 - \chi_2) + \ldots\\
&\quad + \sigma_2\cos(2\pi\bar{\mathbf{H}}\mathbf{R}_m\cdot\mathbf{r}_j + 2\pi\mathbf{H}\cdot\mathbf{T}_m - \chi_2)\\
&\quad + \sigma_3\cos(2\pi\mathbf{H}_s\cdot\mathbf{r}_j - \chi_3) + \sigma_3\cos(2\pi\mathbf{H}_s\mathbf{R}_2\cdot\mathbf{r}_j\\
&\quad + 2\pi\mathbf{H}_s\cdot\mathbf{T}_2 - \chi_3) + \ldots\\
&\quad + \sigma_3\cos(2\pi\mathbf{H}_s\mathbf{R}_m\cdot\mathbf{r}_j + 2\pi\mathbf{H}_s\cdot\mathbf{T}_m - \chi_3)]\}\rangle_{\mathbf{h},\mathbf{k},\mathbf{l}}.
\end{aligned}$$

From the work of Hauptman (1982*a*),

$$\begin{aligned}
q_j \simeq 1 - \big((m/4)\{(f_j^2/\alpha_{20})[\rho_1^2 + \rho_2^2 + \rho_3^2 + 2\rho_1\rho_2\\
\times\cos(\theta_1+\theta_2) + \rho_3^2\cos 2\theta_3]
\end{aligned}$$

$$+ (g_j^2/\alpha_{02})[\sigma_1^2 + \sigma_2^2 + \sigma_3^2 + 2\sigma_1\sigma_2\cos(\chi_1 + \chi_2)$$
$$+ \sigma_3^2\cos 2\chi_3] + (2f_jg_j/\alpha_{20}^{1/2}\alpha_{02}^{1/2})[\rho_1\sigma_1\cos(\theta_1 - \chi_1)$$
$$+ \rho_2\sigma_2\cos(\theta_2 - \chi_2) + \rho_3\sigma_3\cos(\theta_3 - \chi_3)$$
$$+ \rho_1\sigma_2\cos(\theta_1 + \chi_2) + \rho_2\sigma_1\cos(\theta_2 + \chi_1)$$
$$+ \rho_3\sigma_3\cos(\theta_3 + \chi_3)]\} + (im/4)\{(f_j^3/\alpha_{20}^{3/2})$$
$$\times [\rho_1^2 + \rho_2^2 + 2\rho_1\rho_2\cos(\theta_1 + \theta_2)]\rho_3\cos(\theta_3 + n\pi)$$
$$+ (g_j^3/\alpha_{02}^{3/2})[\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\cos(\chi_1 + \chi_2)]\sigma_3$$
$$\times \cos(\chi_3 + n\pi) + (f_j^2g_j/\alpha_{20}\alpha_{02}^{1/2})[\rho_1^2 + \rho_2^2$$
$$+ 2\rho_1\rho_2\cos(\theta_1 + \theta_2)]\sigma_3\cos(\chi_3 + n\pi)$$
$$+ (f_jg_j^2/\alpha_{20}^{1/2}\alpha_{02})[\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\cos(\chi_1 + \chi_2)]\rho_3$$
$$\times \cos(\theta_3 + n\pi) + (f_j^2g_j/\alpha_{20}\alpha_{02}^{1/2})$$
$$\times [2\rho_1\sigma_2\cos(\theta_1 + \chi_2) + 2\rho_2\sigma_1\cos(\theta_2 + \chi_1)$$
$$+ 2\rho_1\sigma_1\cos(\theta_1 - \chi_1) + 2\rho_2\sigma_2\cos(\theta_2 - \chi_2)]\rho_3$$
$$\times \cos(\theta_3 + n\pi) + (f_jg_j^2/\alpha_{20}^{1/2}\alpha_{02})[2\rho_1\sigma_2\cos(\theta_1 + \chi_2)$$
$$+ 2\rho_2\sigma_1\cos(\theta_2 + \chi_1) + 2\rho_1\sigma_1\cos(\chi_1 - \theta_1)$$
$$+ 2\rho_2\sigma_2\cos(\chi_2 - \theta_2)]\sigma_3\cos(\chi_3 + n\pi)\}),$$

where $n$ is an integer that depends on the space group as listed in Table 1. From (4) and

$$\prod_{j=1}^{N/m} q_j \simeq \exp\left[(1/m)\sum_{j=1}^{N}(q_j - 1)\right],$$

the expression for $\prod_{j=1}^{N/m} q_j$ can be easily obtained. Substitution of this expression into (6) and completion of the 12-fold integral give

$$P \simeq c_0\exp([1/(1-\alpha^2)](R_3^2\cos 2\Phi_3 + S_3^2\cos 2\Psi_3)$$
$$+ d_1\cos(\Phi_3 + n\pi) + d_2\cos(\Psi_3 + n\pi)$$
$$+ 4\beta R_3S_3\sin\Phi_3\sin\Psi_3 + \{[2R_1R_2/(1-\alpha^2)]$$
$$+ 2\beta_0R_1R_2[\beta_1R_3\cos(\Phi_3 + n\pi)$$
$$+ \beta_4S_3\cos(\Psi_3 + n\pi)]\}\cos(\Phi_1 + \Phi_2)$$
$$+ \{[2S_1S_2/(1-\alpha^2)] + 2\beta_0S_1S_2[\beta_2R_3\cos(\Phi_3 + n\pi)$$
$$+ \beta_5S_3\cos(\Psi_3 + n\pi)]\}\cos(\Psi_1 + \Psi_2)$$
$$+ \{2\beta + 2\beta_0[\beta_4R_3\cos(\Phi_3 + n\pi) + \beta_2S_3$$
$$\times \cos(\Psi_3 + n\pi)]\}[R_1S_1\cos(\Phi_1 - \Psi_1)$$
$$+ R_2S_2\cos(\Phi_2 - \Psi_2)] + \{-2\beta + 2\beta_0[\beta_4R_3\cos(\Phi_3$$
$$+ n\pi) + \beta_2S_3\cos(\Psi_3 + n\pi)]\}[R_1S_2\cos(\Phi_1 + \Psi_2)$$
$$+ R_2S_1\cos(\Phi_2 + \Psi_1)]), \tag{7}$$

where

$$c_0 = \left[\prod_{j=1}^{3}(R_jS_j)/\pi^6(1-\alpha^2)^3\right]$$
$$\times \exp\left[-\sum_{j=1}^{3}(R_j^2 + S_j^2)/(1-\alpha^2)\right],$$

$$d_1 = \beta_0[\beta_1(R_1^2 + R_2^2) + \beta_2(S_1^2 + S_2^2) + \beta_3]R_3,$$
$$d_2 = \beta_0[\beta_4(R_1^2 + R_2^2) + \beta_5(S_1^2 + S_2^2) + \beta_6]S_3,$$
$$\alpha = \alpha_{11}/(\alpha_{20}\alpha_{02})^{1/2}, \tag{8}$$
$$\beta = \alpha/(1-\alpha^2), \tag{9}$$
$$\beta_0 = 2/[(1-\alpha^2)(\alpha_{20}\alpha_{02})^{1/2}]^3, \tag{10}$$
$$\beta_1 = \alpha_1 - \alpha^3\alpha_2 + 3\alpha^2\alpha_3 - 3\alpha\alpha_4, \tag{11}$$
$$\beta_2 = \alpha^2\alpha_1 - \alpha\alpha_2 + (1 + 2\alpha^2)\alpha_3 - (2 + \alpha^2)\alpha\alpha_4, \tag{12}$$
$$\beta_3 = 2(1-\alpha^2)[\alpha\alpha_2 - \alpha_1 + 3\alpha\alpha_4 - (1 + 2\alpha^2)\alpha_3], \tag{13}$$
$$\beta_4 = \alpha^2\alpha_2 - \alpha\alpha_1 + (1 + 2\alpha^2)\alpha_4 - (2 + \alpha^2)\alpha\alpha_3, \tag{14}$$
$$\beta_5 = \alpha_2 - \alpha^3\alpha_1 + 3\alpha^2\alpha_4 - 3\alpha\alpha_3, \tag{15}$$
$$\beta_6 = 2(1-\alpha^2)[\alpha\alpha_1 - \alpha_2 + 3\alpha\alpha_3 - (1 + 2\alpha^2)\alpha_4], \tag{16}$$
$$\alpha_1 = \alpha_{02}^{3/2}\alpha_{30}, \tag{17}$$
$$\alpha_2 = \alpha_{20}^{3/2}\alpha_{03}, \tag{18}$$
$$\alpha_3 = \alpha_{20}\alpha_{02}^{1/2}\alpha_{12}, \tag{19}$$
$$\alpha_4 = \alpha_{20}^{1/2}\alpha_{02}\alpha_{21}. \tag{20}$$

### 2.2. *The conditional probability distribution of the 3PSSs, given the six magnitudes $|E_{\mathbf{H}}|$, $|E_{\bar{\mathbf{H}}}|$, $|E_{\mathbf{H}_s}|$, $|G_{\mathbf{H}}|$, $|G_{\bar{\mathbf{H}}}|$, $|G_{\mathbf{H}_s}|$ in its first neighborhood*

Refer to §2.1 for the probabilistic background. Let

$$|E_{\mathbf{H}}| = R_1,\ |E_{\bar{\mathbf{H}}}| = R_2,\ |E_{\mathbf{H}_s}| = R_3, \tag{21}$$
$$|G_{\mathbf{H}}| = S_1,\ |G_{\bar{\mathbf{H}}}| = S_2,\ |G_{\mathbf{H}_s}| = S_3. \tag{22}$$

Then, the 3PSS $\omega_1 = \varphi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} + \varphi_{\mathbf{H}_s}$, as a function of the primitive random variable $(\mathbf{h}, \mathbf{k}, \mathbf{l})$, is itself a random variable. Denote by $P_1 = P_1(\Omega_1|R_1, R_2, R_3, S_1, S_2, S_3)$ the conditional probability distribution of $\omega_1$, given (21) and (22), then $P_1$ is derived from (7) by fixing $R_1$, $R_2$, $R_3$, $S_1$, $S_2$, $S_3$, integrating $P$ with respect to $\Psi_1$, $\Psi_2$, $\Psi_3$ from 0 to $2\pi$ and multiplying by a suitable normalizing factor. The final formula is

$$P_1 \simeq (1/K_1)\exp[(-1)^nA_1\cos\Omega_1], \tag{23}$$

where

$$K_1 = 2\pi I_0(A_1), \tag{24}$$

$$A_1 = \beta\{R_3[\beta_1(R_1 + R_2)^2 + \beta_2(S_1^2 + S_2^2) + \beta_3]$$
$$+ 2\beta_4(R_1 + R_2)R_3(S_1T_1 + S_2T_2) + S_3[\beta_4(R_1 + R_2)^2$$
$$+ \beta_5(S_1^2 + S_2^2) + \beta_6]T_3 + 2\beta_2S_1S_2R_3T_1T_2$$
$$+ 2\beta_2(R_1 + R_2)S_3(S_1T_1 + S_2T_2)T_3$$
$$+ 2\beta_5S_1S_2S_3T_1T_2T_3\}, \tag{25}$$

the function $T_j$ is the ratio of two modified Bessel functions of order one and zero:

$$T_j = I_1(2\beta R_jS_j)/I_0(2\beta R_jS_j),\quad j = 1, 2, 3, \tag{26}$$

and the $\beta$ and $\beta_j$ $(j = 0, 1, 2, 3, 4, 5, 6)$ are defined by (8)–(20).

With the same procedure as described above, the conditional probability distributions $P_i = P_i(\Omega_i | R_1, R_2, R_3, S_1, S_2, S_3)$ of the 3PSSs $\omega_i$ $(i = 2, 3, 4, 5, 6, 7, 8)$ are obtained. These results can be expressed as

$$P_i \simeq (1/K_i)\exp[(-1)^n A_i \cos \Omega_i], \qquad (27)$$

$$K_i = 2\pi I_0(A_i), \quad i = 1, 2, \ldots, 8, \qquad (28)$$

where $A_1$ is given by (25) and

$$\begin{aligned}
\omega_2 &= \psi_{\mathbf{H}} + \psi_{\bar{\mathbf{H}}} + \varphi_{\mathbf{H}_s},\\
A_2 &= \beta_0\{R_3[\beta_1(R_1^2 + R_2^2) + \beta_2(S_1 + S_2)^2 + \beta_3]\\
&\quad + 2\beta_4(S_1 + S_2)R_3(R_1T_1 + R_2T_2) + S_3[\beta_4(R_1^2 + R_2^2)\\
&\quad + \beta_5(S_1 + S_2)^2 + \beta_6]T_3 + 2\beta_1R_1R_2R_3T_1T_2\\
&\quad + 2\beta_2(S_1 + S_2)S_3(R_1T_1 + R_2T_2)T_3\\
&\quad + 2\beta_4R_1R_2S_3T_1T_2T_3\}, \qquad (29)\\
\omega_3 &= \varphi_{\mathbf{H}} + \psi_{\bar{\mathbf{H}}} + \varphi_{\mathbf{H}_s},\\
A_3 &= \beta_0\{R_3[\beta_1(R_1^2 + R_2^2) + \beta_2(S_1^2 + S_2^2) + \beta_3 + 2\beta_4R_1S_2]\\
&\quad + 2S_1R_3(\beta_2S_2 + \beta_4R_1)T_1 + 2R_2R_3(\beta_1R_1 + \beta_4S_2)T_2\\
&\quad + S_3[\beta_4(R_1^2 + R_2^2) + \beta_5(S_1^2 + S_2^2) + \beta_6\\
&\quad + 2\beta_2R_1S_2]T_3 + 2\beta_4S_1R_2R_3T_1T_2\\
&\quad + 2S_1S_3(\beta_2R_1 + \beta_5S_2)T_1T_3\\
&\quad + 2R_2S_3(\beta_2S_2 + \beta_4R_1)T_2T_3 + 2\beta_2S_1R_2S_3T_1T_2T_3\}, \qquad (30)\\
\omega_4 &= \psi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} + \varphi_{\mathbf{H}_s},\\
A_4 &= \beta_0\{R_3[\beta_1(R_1^2 + R_2^2) + \beta_2(S_1^2 + S_2^2) + \beta_3 + 2\beta_4S_1R_2]\\
&\quad + 2R_1R_3(\beta_1R_2 + \beta_4S_1)T_1 + 2S_2R_3(\beta_2S_1 + \beta_4R_2)T_2\\
&\quad + S_3[\beta_4(R_1^2 + R_2^2) + \beta_5(S_1^2 + S_2^2)\\
&\quad + \beta_6 + 2\beta_2S_1R_2]T_3 + 2\beta_4R_1S_2R_3T_1T_2\\
&\quad + 2R_1S_3(\beta_2S_1 + \beta_4R_2)T_1T_3 + 2S_2S_3(\beta_2R_2\\
&\quad + \beta_5S_1)T_2T_3 + 2\beta_2R_1S_2S_3T_1T_2T_3\}. \qquad (31)
\end{aligned}$$

It is easy to obtain the formulas of $A_i$ for the other $\omega_i$ $(i = 5, 6, 7, 8)$ from (25), (29), (30) and (31), *e.g.* $A_5$ from (25) and $A_6$ from (29), by means of substituting $R_j$ by $S_j$ and $S_j$ by $R_j$ $(j = 1, 2, 3)$ and interchanging $\beta_1$ and $\beta_5$, $\beta_2$ and $\beta_4$, $\beta_3$ and $\beta_6$.

Equations (23) and (27) are obtained by making use of the following conditions:

$$\varphi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} = 0, \quad \psi_{\mathbf{H}} + \psi_{\bar{\mathbf{H}}} = 0, \qquad (32)$$

$$\varphi_{\mathbf{H}} + \psi_{\bar{\mathbf{H}}} \simeq 0, \quad \psi_{\mathbf{H}} + \varphi_{\bar{\mathbf{H}}} \simeq 0, \qquad (33)$$

where the validity of (33) requires that $[2\alpha/(1 - \alpha^2)]R_1S_1$ is large (Hauptman, 1982*a*). It is also because of (32) and (33) that the formula (27) can be used to estimate the 1PSSs $\varphi_{\mathbf{H}_s}$ or $\psi_{\mathbf{H}_s}$.

Table 1. *The types of seminvariant vectors and n values for monoclinic and orthorhombic systems*

| Type | Space group and $n$ | | | | |
|---|---|---|---|---|---|
| | | $P2$ | $P2_1$ | | |
| $2h\ 0\ 2l$ | | 0 | $k$ | | |
| | | $Pm$ | $Pc$ | | |
| $0\ 2k\ 0$ | | 0 | $l$ | | |
| | $P222$ | $P222_1$ | $P2_12_12$ | $P2_12_12_1$ | |
| $2h\ 2k\ 0$ | 0 | $l$ | 0 | $l+h$ | |
| $0\ 2k\ 2l$ | 0 | 0 | $k+l$ | $h+k$ | |
| $2h\ 0\ 2l$ | 0 | $l$ | $k+l$ | $k+l$ | |
| | $Pmm2$ | $Pmc2_1$ | $Pcc2$ | $Pma2$ | $Pca2_1$ |
| $2h\ 2k\ 0$ | 0 | $l$ | 0 | 0 | $l$ |
| | $Pnc2$ | $Pmn2_1$ | $Pba2$ | $Pna2_1$ | $Pnn2$ |
| $2h\ 2k\ 0$ | 0 | $l+h$ | 0 | $l$ | 0 |

In the calculation of $A_i$, the parameters $\beta_j$ have an important effect. Based on some preliminary calculations, it appears that $\beta_1 < 0$, $\beta_2 < 0$, $\beta_3 > 0$, $\beta_4 > 0$, $\beta_5 > 0$, $\beta_6 < 0$. Formula (27) is completely general and includes the special case of a native protein and a heavy-atom isomorphous derivative. When the $f$ structure is a native protein and the $g$ structure a heavy-atom derivative, the parameters described above can be further simplified to

$$\begin{aligned}
\alpha &= 1/(1 + p_2)^{1/2},\\
\beta &= (1 + p_2)^{1/2}/p_2,\\
\beta_0 &= 2(1 + p_2)^{3/2}/(\alpha_{20}p_2)^3,\\
\beta_1 &= -\alpha_{20}^{3/2}\alpha_{30}(p_3 - p_2^3)/(1 + p_2)^{3/2},\\
\beta_2 &= -\alpha_{20}^{3/2}\alpha_{30}p_3/(1 + p_2)^{1/2},\\
\beta_3 &= 2\alpha_{20}^{3/2}\alpha_{30}p_2(p_3 - p_2^2)/(1 + p_2)^{3/2},\\
\beta_4 &= \alpha_{20}^{3/2}\alpha_{30}p_3/(1 + p_2),\\
\beta_5 &= \alpha_{20}^{3/2}\alpha_{30}p_3,\\
\beta_6 &= -2\alpha_{20}^{3/2}\alpha_{30}p_2p_3/(1 + p_2),
\end{aligned}$$

where $p_2 = (\alpha_{02} - \alpha_{20})/\alpha_{20}$, $p_3 = (\alpha_{03} - \alpha_{30})/\alpha_{30}$ and $\alpha_{02} - \alpha_{20}$ and $\alpha_{03} - \alpha_{30}$ can be easily calculated by the summations over the heavy atoms only.

## 3. Applications to the 1PSS estimates

### 3.1. *The formulas for the 1PSS estimates*

When (32) holds, the probability distribution of the 1PSS $\varphi_{\mathbf{H}_s}$ is directly obtained from those of $\omega_1$ and $\omega_2$:

$$P_i(\varphi_{\mathbf{H}_s}) \simeq (1/K_i)\exp[(-1)^n A_i \cos \varphi_{\mathbf{H}_s}], \quad i = 1, 2. \qquad (34)$$

Table 2. *Estimated results of the 1PSSs $\varphi_{\mathbf{H}_s}$ accumulated in groups according to given minimum values of $|A|$ for the three pairs of isomorphous structures*

I: rubredoxin. II: ferrocytochrome c. III: cytochrome $c_{550}$. $N$: the number of $\varphi_{\mathbf{H}_s}$ in group. $\langle|A|\rangle$: average of $|A|$ values over the 1PSSs in the group. %: the percentage of $\varphi_{\mathbf{H}_s}$ correctly estimated.

| | I | | | | II | | | | III | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | $\langle\lvert A\rvert\rangle$ | $\langle\lvert E-G\rvert\rangle$ | % | $N$ | $\langle\lvert A\rvert\rangle$ | $\langle\lvert E-G\rvert\rangle$ | % | $N$ | $\langle\lvert A\rvert\rangle$ | $\langle\lvert E-G\rvert\rangle$ | % |
| Total | 78 | 46.0 | 0.26 | 92.3 | 371 | 22.5 | 0.23 | 78.7 | 236 | 24.1 | 0.31 | 84.7 |
| $\lvert A\rvert > 1.0$ | 73 | 49.1 | 0.26 | 95.9 | 296 | 28.1 | 0.27 | 85.5 | 204 | 27.9 | 0.34 | 89.2 |
| $\lvert A\rvert > 3.0$ | 68 | 52.6 | 0.27 | 97.1 | 224 | 36.5 | 0.32 | 93.3 | 162 | 34.6 | 0.38 | 92.0 |
| $\lvert A\rvert > 5.0$ | 63 | 56.5 | 0.28 | 98.4 | 190 | 42.3 | 0.35 | 97.4 | 140 | 39.5 | 0.41 | 97.1 |

For a given $\mathbf{H}_s$, considering all the reflections $\mathbf{H}$ and the corresponding centrosymmetric reflections $\bar{\mathbf{H}}$ so that each of the triplets ($\mathbf{H}$, $\bar{\mathbf{H}}$, $\mathbf{H}_s$) satisfies (1), (34) becomes

$$P_i(\varphi_{\mathbf{H}_s}) \simeq C_i \exp\left[\sum_{\mathbf{H},\bar{\mathbf{H}}}(-1)^n A_i \cos\varphi_{\mathbf{H}_s}\right], \quad i = 1, 2, \tag{35}$$

where $C_i$ is a normalizing constant. In the practical application, an average $A$ value over $A_1$ and $A_2$ can be used. Then we have

$$P(\varphi_{\mathbf{H}_s}) \simeq [2\pi I_0(A)]^{-1} \exp(A\cos\varphi_{\mathbf{H}_s}), \tag{36}$$

where

$$A = \tfrac{1}{2}\sum_{\mathbf{H},\bar{\mathbf{H}}}(-1)^n(A_1 + A_2). \tag{37}$$

Equation (36) has a unique maximum at $\varphi_{\mathbf{H}_s} = 0$ or $\pi$ according as $A > 0$ or $A < 0$, respectively. Therefore, a reliable estimate of the seminvariant $\varphi_{\mathbf{H}_s}$ can be obtained by calculating the sign of $A$ when $|A|$ is large. Similarly, the seminvariant $\psi_{\mathbf{H}_s}$ can be estimated from the distributions of $\omega_5$ and $\omega_6$ with the same approach.

### 3.2. *Test calculations*

All the test calculations were made using error-free diffraction data. The normalized structure factors $E_{\mathbf{H}}$ and $G_{\mathbf{H}}$ were calculated from the known atomic coordinates of three protein structures and their heavy-atom derivatives as follows.

I: rubredoxin (Adman, Sieker, Jensen, Bruschi & LeGall, 1977), space group $P2_1$, 389 atoms in the asymmetric unit (not including water molecules). A derivative was made by replacing the Fe atom in the native protein by a Pt atom. A set of reflections at 1.5 Å resolution was calculated.

II: ferrocytochrome c (Takano & Dickerson, 1981), space group $P2_12_12$, 900 atoms in the asymmetric unit. A heavy-atom derivative was constructed by replacing the O atom of a water molecule by a Pt atom. The resolution of the reflection set is 2.0 Å.

III: cytochrome $c_{550}$ and its $PtCl_4^{2-}$ isomorphous derivative (Timkovich & Dickerson, 1976), space group $P2_12_12_1$, 1017 atoms in the asymmetric unit for the native protein. Intensity data were generated to a resolution of 2.5 Å.

The seminvariants $\varphi_{\mathbf{H}_s}$ were estimated based on (36) and (37) for each pair of isomorphous structures. The test results are given in Table 2. As is shown in Table 1, $\varphi_{\mathbf{H}_s}$ is of one form, $\varphi_{2h,0,2l}$, for I, and three forms, $\varphi_{2h,2k,0}$, $\varphi_{0,2k,2l}$, $\varphi_{2h,0,2l}$, for II and III. The first row of Table 2 gives the total number of $\varphi_{\mathbf{H}_s}$ at the given resolution and the percentage of $\varphi_{\mathbf{H}_s}$ estimated correctly. The results for those seminvariants having $|A|$ values larger than 1.0, 3.0 and 5.0 are shown in rows 2, 3 and 4, respectively. From Table 2, the following conclusions may be made.

(i) Formulas (36) and (37) yield reliable estimates of the 1PSSs having values 0 or $\pi$ for all three structures. The larger the $|A|$ values, the more reliable are the estimates.

(ii) Reliability of the estimation increases with the increase of the differences between the normalized structure-factor magnitudes of an isomorphous pair.

(iii) Efficiency of the estimation is related to the complexity of structure. Especially for those seminvariants with smaller $|A|$ values, the results of I are obviously better than those of both II and III. When $|A| > 3.0$, more than 90% of the seminvariants are correctly estimated for the three structures. Therefore, for a pair of isomorphous structures consisting of a native protein having as many as 1000 atoms and a Pt-atom derivative, the present method is expected to be useful.

## 4. Concluding remarks

The integration of direct methods with isomorphous replacement has been applied to the estimation of the 1PSSs. The derivations of the conditional probability distributions of a special type of 3PSSs have been stated. These distributions directly lead to an approach to estimating the 1PSSs. The analysis also includes the special case that one member of the pair is a native protein and the other member is a heavy-atom isomorphous derivative. The initial applications of this work, using error-free data from the three protein structures, have shown satisfactory results.

This work was supported in part by Jilin Aodong Pharmaceutical Ltd.

### References

Adman, E. T., Sieker, L. C., Jensen, L. H., Bruschi, M. & LeGall, J. (1977). *J. Mol. Biol.* **112**, 113–120.
Giacovazzo, C. (1980). *Direct Methods in Crystallography.* London: Academic Press.
Hauptman, H. (1975*a*). *Acta Cryst.* A**31**, 671–679.
Hauptman, H. (1975*b*). *Acta Cryst.* A**31**, 680–687.
Hauptman, H. (1982*a*). *Acta Cryst.* A**38**, 289–294.
Hauptman, H. (1982*b*). *Acta Cryst.* A**38**, 632–641.
Hauptman, H. & Karle, J. (1956). *Acta Cryst.* **9**, 45–55.
Hu, N.-H. & Liu, Y.-S. (1995). *Acta Cryst.* A**51**, 520–524.
Karle, J. & Hauptman, H. (1958). *Acta Cryst.* **11**, 264–269.
Liu, Y.-S. & Hu, N.-H. (1994). *Acta Cryst.* A**50**, 595–601.
Takano, T. & Dickerson, R. E. (1981). *J. Mol. Biol.* **153**, 79–94.
Timkovich, R. & Dickerson, R. E. (1976). *J. Biol. Chem.* **251**, 4033–4046.
Velmurugan, D. & Hauptman, H. (1989). *Acta Cryst.* A**45**, 158–163.
Velmurugan, D., Hauptman, H. & Potter, S. A. (1989). *Acta Cryst.* A**45**, 163–165.