

- HAUPTMAN, H. & DUAX, W. L. (1972). *Acta Cryst.* **A28**, 393–395.
- HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. (1969). *Acta Cryst.* **B25**, 811–814.
- HAUPTMAN, H., FISHER, J. & WEEKS, C. M. (1971). *Acta Cryst.* **B27**, 1550–1561.
- HOWELLS, E. R., PHILLIPS, D. C. & ROGERS, D. (1950). *Acta Cryst.* **3**, 210–214.
- HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* **A34**, 863–870.
- KARLE, J. (1970). *Acta Cryst.* **B26**, 1614–1617.
- KARLE, J. (1979). *Proc. Natl Acad. Sci. USA*, **76**, 2089–2093.
- KARLE, J. (1980). *Proc. Natl Acad. Sci. USA* **77**, 5–9.
- KARLE, J. (1982). *Acta Cryst.* **A38**, 327–333.
- KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.
- KARLE, J. & HAUPTMAN, H. (1957). *Acta Cryst.* **10**, 515–524.
- LANGS, D. A. (1988). *Science*, **241**, 188–191.
- LANGS, D. A. (1993). *Acta Cryst.* **D49**, 158–167.
- MAIN, P. (1977). *Acta Cryst.* **A33**, 750–757.
- MESSAGER, J. C. & TSOUCARIS, G. (1972). *Acta Cryst.* **A28**, 482–484.
- PLETNEV, V. Z., GALITSKII, N. M., IVANOV, V. T., SMITH, G. D., WEEKS, C. M. & DUAX, W. L. (1980). *Biopolymers*, **19**, 1517–1534.
- PLETNEV, V. Z., IVANOV, V. T., LANGS, D. A., STRONG, P. D. & DUAX, W. L. (1992). *Biopolymers*, **32**, 819–827.
- PLETNEV, V. Z., MIKHAILOVA, I. YU., IVANOV, V. T., LANGS, D. A., GROCHULSKI, P. & DUAX, W. L. (1991). *Biopolymers*, **31**, 409–415.
- SMITH, G. D. & KRISTENANSKY, J. (1991). Unpublished work.
- TSOUCARIS, G. (1970). *Acta Cryst.* **A26**, 492–499.
- VAUGHAN, P. A. (1958). *Acta Cryst.* **11**, 111–115.
- VITERBO, D. & WOOLFSON, M. M. (1973). *Acta Cryst.* **A29**, 205–208.
- WOOLFSON, M. M. (1954). *Acta Cryst.* **7**, 65–67.
- YAO, J. (1981). *Acta Cryst.* **A37**, 642–644.
- YAO, J. (1983). *Acta Cryst.* **A39**, 35–37.

*Acta Cryst.* (1993). **A49**, 557–569

## On Direct-Methods Phase Information from Differences Between Isomorphous Structure Factors

BY CHRISTOS E. KYRIAKIDIS, RENÉ PESCHAR AND HENK SCHENK

*Laboratory for Crystallography, University of Amsterdam, Nieuwe Achtergracht 166,  
1018 WV Amsterdam, The Netherlands*

(Received 4 June 1992; accepted 2 November 1992)

### Abstract

An efficient procedure is presented for the derivation of joint probability distributions of isomorphous data sets. The new technique is based on the use of the differences of isomorphous structure factors as random variables. It will be shown that the usual probabilistic techniques, applied to these random variables, finally result in the joint probability distribution of three single differences of isomorphous structure factors comprising three doublet and eight triplet phase sums. An advantage of the new technique is that the inherent correlation between the isomorphous data sets is removed if a probabilistic procedure is set up for the small difference itself. In this way, an enormous mathematical simplification is obtained while the final results are much better than those obtainable by previous probabilistic expressions. The final triplet distribution seems to be of sufficient quality to be used in a normal direct-methods procedure. In contrast to usual approaches, the heavy-atom substructure need not be solved first. The probabilistic expression will be explained in detail for one and three single differences. Applications for the cases of single anomalous scattering, two different wavelengths and single isomorphous replacement (excluding anomalous-scattering effects)

for both real and randomly generated data show the strength of the method.

### Abbreviations

c.f.	Characteristic function
j.p.d.	Joint probability distribution
c.p.d.	Conditional probability distribution
(p.)r.v.	(Primitive) random variable
s.f.	Structure factor
SD	Single difference
DM	Direct methods
DR	Diffraction ratio
SIR(N)AS	Single isomorphous replacement (neglecting) anomalous scattering
SAS	Single-wavelength anomalous scattering
2DW	Two different wavelengths

### 1. Introduction

The crystal structures of relatively small molecules with up to 100 independent atoms are readily determined from diffraction intensities by means of DM techniques relying on the mathematical application of a j.p.d. of complex-valued structure factors. DM estimate phases from the intensities and when these

phases are approximately correct the maxima in the Fourier summation based upon measured magnitudes and estimated phases correspond with the atomic coordinates. An increase of the size of the structure reduces the reliability of the phase estimates and consequently obstructs the structure determination *via* traditional DM (see, for example, Woolfson, 1987).

For large structures, the probability of imaging an incomplete structure or only a fragment increases. In macromolecular crystallography, the size problem is partly solved by introducing more data about the same structure, *e.g.* by SAS, SIRAS, SIRNAS, 2DW *etc.* (see, for example, Karle, 1989).

Various techniques exist to complete partial structural models provided the model contains a sufficient amount of scattering power. Techniques that have proved to be successful are the tangent recycling methods (Karle, 1970; Hull & Irwin, 1978), the Fourier recycling methods (Kinneking & de Graaff, 1984) and the tangent recycling methods applied to the difference s.f.s (Beurskens, Prick, Doesburg & Gould, 1979). This last technique, upon which the *DIRDIF* system of programs is based, has been very successful in completing heavy-atom models (Beurskens *et al.*, 1991). Also, extension of SIRNAS models (Sim, 1959) relies on an initial model (in general the heavy-atom substructure) from which partial structure factors can be calculated. The subsequent structure completion is based on the difference between the total and the model structure factors, this difference being a function of the non-model atoms while the phase model is usually derived from the substructure. The completion of the structure can be achieved in various ways (see, for example, Camalli, Giacovazzo & Spagna, 1985; Beurskens & Smykalla, 1991). A disadvantage of the above techniques is that a partial model must be available. Therefore, in this paper, a different approach will be followed. It will be shown that differences between isomorphous s.f.s (hereafter, these differences will be called SDs) can be defined to which the usual probabilistic machinery can be applied but which do not require an initial structural model. A major advantage of the new technique is that the inherent correlation between the isomorphous data sets is removed if a mathematical procedure is set up for the small difference itself. An important goal of the paper is the derivation of a new expression to estimate the triplet phase sums present among isomorphous data. It will be shown that the new procedure, supplemented by optimal doublet phase-sum estimates that use difference Patterson information [Kyriakidis, Peschar & Schenk (1993*b*), from now on referred to as KPS2], leads to far better results than obtainable by other j.p.d.-based expressions (Hauptman, 1982*a,b*; Giacovazzo, 1983; Giacovazzo, Cascarano & Zheng, 1988; Fortier & Nigam, 1989; Peschar &

Schenk, 1991; hereafter P&S), in particular if the DR is small (Kyriakidis, Peschar & Schenk, 1993*a*; from now on KPS1). In contrast with other DM techniques, the final triplet distributions in the SAS and 2DW cases seem to be of sufficient quality to be used in a normal DM procedure.

## 2. The single differences of isomorphous s.f.s

Hitherto, the use of DM to solve structures from single-crystal data seems to have been limited to small structures. The reason for this is clear: the j.p.d. of three structure factors depends in first approximation on  $N^{-1/2}$  so the j.p.d. gets increasingly flattened if  $N$  becomes large. On the other hand, large structures such as proteins have been solved using SIRNAS and/or SAS. This raises the question of why DM fails while other techniques succeed.

An efficient way to improve the applicability of DM is to reduce the number of variables ( $N$ ) involved. In the case of isomorphous data, as present in techniques such as SIRNAS, SIRAS, SAS and 2DW, this reduction can be achieved in a very simple way. It has been shown recently that the concept of isomorphous structure factors can be useful for estimation of the doublet and triplet phase sums present amongst them (KPS1; KPS2). From the tests, it appeared that for too low DRs, *i.e.* almost perfectly isomorphous structures, no useful estimates could be obtained, even for small structures. Analyses showed that in these cases the reliability indicators were no longer properly defined. If the differences between isomorphous structure factors become too small, the normal mathematical procedure more or less fails. It seems that the very small quantities cannot be expressed in terms of the usual variables.

This suggested that a different type of r.v. should be defined: the single difference of isomorphous s.f.s,  $F_\nu^d$ , which is the difference between two isomorphous structure factors  $F_\nu^l$  and  $F_\nu^m$ . The subscript  $\nu$  refers to a particular reflection and the superscripts  $l$ ,  $m$  and  $d$  denote dependence on the isomorphous data sets  $l$ ,  $m$  and both  $l$  and  $m$ , respectively. We have

$$\begin{aligned} F_\nu^d &\equiv F_\nu^l - F_\nu^m = \sum_{j=1}^N f_{j\nu}^l \exp(2\pi i \mathbf{H}_\nu \cdot \mathbf{r}_j) \\ &\quad - \sum_{j=1}^N f_{j\nu}^m \exp(2\pi i \mathbf{H}_\nu \cdot \mathbf{r}_j) \\ &= \sum_{j=1}^n (f_{j\nu}^l - f_{j\nu}^m) \exp(2\pi i \mathbf{H}_\nu \cdot \mathbf{r}_j) \\ &= |F_\nu^d| \exp(i\varphi_\nu^d), \end{aligned} \quad (1)$$

where  $f_{j\nu}^l$  and  $f_{j\nu}^m$  represent the atomic scattering factors for a corresponding group of two isomorphous data sets in space group  $P1$  defined in a general way

including anomalous-scattering effects,

$$\begin{aligned} f_{j\nu}^l &\equiv f_j^l(\mathbf{H}_\nu) = f_j^0(\mathbf{H}_\nu) + f_j' + if_j'' \\ &= f_j^r(\mathbf{H}_\nu) + if_j'' \\ &= |f_{j\nu}^l| \exp(i\delta_{j\nu}^l). \end{aligned} \quad (2)$$

(Some of the  $f_{j\nu}^l$  may be zero or negative in the neutron diffraction case.)

Expression (1) shows that  $F_\nu^d$  depends only on the number of atoms ( $n$ ) for which the atomic scattering factors differ in  $F_\nu^l$  and  $F_\nu^m$  while, in contrast,  $F_\nu^l$  and  $F_\nu^m$  depend on all  $N$  variables. Both the magnitude  $|F_\nu^d|$  and the phase  $\varphi_\nu^d$  of  $F_\nu^d$  are functions of the magnitudes and phases of  $F_\nu^l$  and  $F_\nu^m$ . For our purpose, the following expression for  $|F_\nu^d|$  is important:

$$|F_\nu^d|^2 = |F_\nu^l|^2 + |F_\nu^m|^2 - 2|F_\nu^l||F_\nu^m| \cos \psi_\nu^d, \quad (3)$$

where  $\psi_\nu^d$  is the doublet phase sum between the data sets  $l$  and  $m$ ,

$$\varphi_\nu^l + s^d \varphi_\nu^m = \psi_\nu^d \quad (4)$$

with

$$s^d = \begin{cases} -1 & \text{for the cases SIR(N)AS, 2DW etc.} \\ 1 & \text{for SAS case.} \end{cases} \quad (5)$$

$F_\nu^d$  can be considered formally to be a 'structure factor' so all j.p.d.s developed so far apply. In the following we will focus on the j.p.d. of one single  $F_\nu^d$  (§ 2.1) and the j.p.d. of the three SDs of isomorphous s.f.s  $F_1^d$ ,  $F_2^d$  and  $F_3^d$  (§ 2.2).

### 2.1. The j.p.d. of a single difference of isomorphous s.f.s

The j.p.d. of a s.f. provides a good starting point to obtain the j.p.d. of the SD. The normalized j.p.d. of a single structure factor  $F_\nu$  in space group  $P1$  with random variables  $R_\nu$  for  $|F_\nu|$  and  $\Phi_\nu$  for the phase of  $F_\nu$  can be derived as (see, for example, the Appendix)

$$P(R_\nu, \Phi_\nu) = 2R_\nu(z_\nu)^{-1} \exp[-R_\nu^2(z_\nu)^{-1}] \quad (6)$$

with  $z_\nu$  defined as

$$z_\nu = \sum_{j=1}^N |f_{j\nu}|^2. \quad (7)$$

From (6),  $\langle R_\nu^2 \rangle$  can be calculated with a standard integral formula (Abramowitz & Stegun, 1970), which results in

$$\langle R_\nu^2 \rangle = \int_0^\infty R_\nu^2 P(R_\nu, \Phi_\nu) dR_\nu = z_\nu. \quad (8)$$

From the above, the j.p.d. for a single  $|F_\nu^d|$  (using the r.v.s  $R_\nu^d$  and  $\Phi_\nu^d$  for the magnitude and the phase of the  $F_\nu^d$ , respectively) is

$$P(R_\nu^d, \Phi_\nu^d) = 2R_\nu^d(z_\nu^d)^{-1} \exp[-(R_\nu^d)^2(z_\nu^d)^{-1}] \quad (9)$$

with

$$z_\nu^d = \sum_{j=1}^n |f_{j\nu}^l - f_{j\nu}^m|^2. \quad (10)$$

From (3) and (8), it follows that

$$\langle |F_\nu^l|^2 + |F_\nu^m|^2 - 2|F_\nu^l||F_\nu^m| \cos \psi_\nu^d \rangle = z_\nu^d. \quad (11)$$

With the assumption that  $|F_\nu^l|$  and  $|F_\nu^m|$  are known exactly, (11) can be rewritten

$$\begin{aligned} \langle \cos \psi_\nu^d \rangle \\ = (|F_\nu^l|^2 + |F_\nu^m|^2 - z_\nu^d) / (2|F_\nu^l||F_\nu^m|). \end{aligned} \quad (12)$$

Equation (12) is identical to (14), which is derived algebraically in KPS2. In KPS2, it has been shown that (12) may lead to accurate doublet estimates.

### 2.2. The j.p.d. of three SDs of isomorphous s.f.s

First, the j.p.d. of three s.f.s (with anomalous-scattering effects included) will be introduced and then, starting from the c.p.d. of three s.f.'s, the c.p.d. of three SDs will be obtained.

Recently, a general method has been proposed to derive j.p.d.s of s.f.s allowing the use of complex-valued atomic scattering factors (P&S). In this method, these atomic scattering factors are denoted as in (2). The j.p.d. in space group  $P1$  of the general-valued s.f.s  $F_H$ ,  $F_K$  and  $F_L$  can be expressed by means of the r.v.s  $R_1$ ,  $R_2$  and  $R_3$  for the magnitudes  $|F_H|$ ,  $|F_K|$  and  $|F_L|$ , respectively, and the r.v.s  $\Phi_1$ ,  $\Phi_2$  and  $\Phi_3$  for the phases  $\varphi_H$ ,  $\varphi_K$  and  $\varphi_L$ , respectively,

$$\begin{aligned} &P(R_1, R_2, R_3, \Phi_1, \Phi_2, \Phi_3) \\ &= R_1 R_2 R_3 (2\pi)^{-6} \int_0^\infty \int_0^\infty \int_0^\infty \int_0^{2\pi} \int_0^{2\pi} \int_0^{2\pi} \rho_1 \rho_2 \rho_3 \\ &\quad \times \exp\left(-i\left\{\sum_{\nu=1}^3 [\rho_\nu R_\nu \cos(\theta_\nu - \Phi_\nu)]\right\}\right) \\ &\quad \times C(\rho_1, \rho_2, \rho_3, \theta_1, \theta_2, \theta_3) \\ &\quad \times d\theta_1 d\theta_2 d\theta_3 d\rho_1 d\rho_2 d\rho_3. \end{aligned} \quad (13)$$

After the terms up to  $O(N^{-1/2})$  have been collected and the moments-cumulants transformation has been performed, the c.f.  $C$  becomes

$$\begin{aligned} C = \exp\left[-\sum_{\nu=1}^3 (\rho_\nu^2/4)z_\nu - i|z_{123}|(\rho_1\rho_2\rho_3/4)\right. \\ \left.\times \cos(\theta_1 + \theta_2 + \theta_3 + \Delta_{123})\right], \end{aligned} \quad (14)$$

in which

$$z_\nu = \sum_{j=1}^N |f_{j\nu}|^2 \quad (\text{for } \nu = 1, 2, 3) \quad (15)$$

and

$$z_{123} = |z_{123}| \exp(i\Delta_{123}) \\ = \sum_{j=1}^N |f_{j1} f_{j2} f_{j3}| \exp[-i(\delta_{j1} + \delta_{j2} + \delta_{j3})]. \quad (16)$$

The evaluations of (13) and (14) can be performed in a standard way (see, for example, Peschar, 1991) and result in

$$P(R_1, R_2, R_3, \Phi_1, \Phi_2, \Phi_3) \\ = (R_1 R_2 R_3 / z_1 z_2 z_3 \pi^3) \\ \times \exp[-R_1^2/z_1 - R_2^2/z_2 - R_3^2/z_3 \\ + (2z_{123} R_1 R_2 R_3 / z_1 z_2 z_3) \\ \times \cos(\Phi_1 + \Phi_2 + \Phi_3 + \Delta_{123})]. \quad (17)$$

Finally, the r.v. for the triplet phase sum  $\psi_3$  is defined to be  $\Psi_{123}$ , so

$$\Psi_{123} = \Phi_1 + \Phi_2 + \Phi_3 \quad (18)$$

and the normalized c.p.d. of  $\Psi_{123}$  given the  $R_1$ ,  $R_2$  and  $R_3$  is readily obtained as

$$P(\Psi_{123} | R_1, R_2, R_3) \\ = L^{-1} \exp[(2|z_{123}|/z_1 z_2 z_3) R_1 R_2 R_3 \\ \times \cos(\Psi_{123} + \Delta_{123})] \quad (19)$$

with  $L$  the normalization constant,

$$L = 2\pi I_0(2z_{123} R_1 R_2 R_3 / z_1 z_2 z_3). \quad (20)$$

Finally, from (19), an expectation value for  $\Psi_{123}$  can be obtained,

$$\langle \exp(i\Psi_{123}) \rangle = [I_1(2w_{123}) / I_0(2w_{123})] \exp(-i\Delta_{123}) \quad (21)$$

with

$$w_{123} = z_{123} R_1 R_2 R_3 / z_1 z_2 z_3. \quad (22)$$

The distribution for  $\Psi_{123}$  is centred around  $-\Delta_{123}$  and  $I_1/I_0$  acts as a statistical weight ( $I_1$  and  $I_0$  are modified Bessel functions).

The j.p.d. theory recalled above is directly applicable to the SDs of isomorphous s.f.s. In analogy with (13)-(22), the c.p.d. of the triplet phase sum, present among the  $F_H^d$ ,  $F_K^d$  and  $F_L^d$ , can be expressed as

$$P(\Psi_{123}^d | R_1^d, R_2^d, R_3^d) \\ = (L^d)^{-1} \exp[(2|z_{123}^d|/z_1^d z_2^d z_3^d) R_1^d R_2^d R_3^d \\ \times \cos(\Psi_{123}^d + \Delta_{123}^d)], \quad (23)$$

where  $R_1^d$ ,  $R_2^d$ ,  $R_3^d$  are the r.v.s for the three magnitudes  $|F_H^d|$ ,  $|F_K^d|$  and  $|F_L^d|$ , respectively. The r.v. for the difference isomorphous s.f.s triplet  $\psi_3^d$  is

defined as

$$\Psi_{123}^d = \Phi_1^d + \Phi_2^d + \Phi_3^d, \quad (24)$$

furthermore

$$z_\nu^d = \sum_{j=1}^n |f_{j\nu}^l - f_{j\nu}^m|^2 \quad (\text{for } \nu = 1, 2, 3) \quad (25)$$

and

$$z_{123}^d = |z_{123}^d| \exp[i\Delta_{123}^d] \\ = \sum_{j=1}^n (f_{j1}^l - f_{j1}^m)(f_{j2}^l - f_{j2}^m)(f_{j3}^l - f_{j3}^m). \quad (26)$$

$L^d$  is the normalization constant, which is defined, in analogy with (20), by substitution of the variables  $R$  and  $z$  by  $R^d$  and  $z^d$ , respectively.

From (23), an expectation value for  $\Psi_{123}^d$  is readily obtained as

$$\langle \exp(i\Psi_{123}^d) \rangle = [I_1(2w_{123}^d) / I_0(2w_{123}^d)] \exp(-i\Delta_{123}^d) \quad (27)$$

with

$$w_{123}^d = z_{123}^d R_1^d R_2^d R_3^d / z_1^d z_2^d z_3^d. \quad (28)$$

With (1) taken into account, the product of the r.v.s  $R_1^d$ ,  $R_2^d$ ,  $R_3^d$  and  $\exp(i\Psi_{123}^d)$  may be written as

$$R_1^d R_2^d R_3^d \exp(i\Psi_{123}^d) \\ = (F_1^l - F_1^m)(F_2^l - F_2^m)(F_3^l - F_3^m). \quad (29)$$

The right-hand side of (29) can be expressed fully as

$$(F_1^l - F_1^m)(F_2^l - F_2^m)(F_3^l - F_3^m) \\ = |F_1^l F_2^l F_3^l| \exp[i(\Phi_1^l + \Phi_2^l + \Phi_3^l)] \\ - |F_1^l F_2^l F_3^m| \exp[i(\Phi_1^l + \Phi_2^l + \Phi_3^m)] \\ - |F_1^l F_2^m F_3^l| \exp[i(\Phi_1^l + \Phi_2^m + \Phi_3^l)] \\ + |F_1^l F_2^m F_3^m| \exp[i(\Phi_1^l + \Phi_2^m + \Phi_3^m)] \\ - |F_1^m F_2^l F_3^l| \exp[i(\Phi_1^m + \Phi_2^l + \Phi_3^l)] \\ + |F_1^m F_2^l F_3^m| \exp[i(\Phi_1^m + \Phi_2^l + \Phi_3^m)] \\ + |F_1^m F_2^m F_3^l| \exp[i(\Phi_1^m + \Phi_2^m + \Phi_3^l)] \\ - |F_1^m F_2^m F_3^m| \exp[i(\Phi_1^m + \Phi_2^m + \Phi_3^m)]. \quad (30)$$

With the right-hand side of (12) expressed as a cosine, it follows that

$$\Psi_\nu^d = \pm \lambda_\nu^d \quad (31)$$

with

$$\lambda_\nu^d = \cos^{-1} [(|F_\nu^l|^2 + |F_\nu^m|^2 - z_\nu^d) / (2|F_\nu^l||F_\nu^m|)]. \quad (32)$$

From (31) and (32), it is possible to express (30) completely in only one of the eight isomorphous triplets using

$$\Psi_{123}^{lmn} = \Phi_1^l + \Phi_2^m + \Phi_3^n \quad (l, m, n = 1, 2). \quad (33)$$

For example, for the triplet  $\Psi_{123}^{111}$ , (30) gives

$$\begin{aligned} & (F_1^1 - F_1^2)(F_2^1 - F_2^2)(F_3^1 - F_3^2) \\ &= \exp [i\Psi_{123}^{111}] \{ |F_1^1 F_2^1 F_3^1| \\ & \quad - |F_1^1 F_2^2 F_3^1| \exp[-i\lambda_3^d] \\ & \quad - |F_1^1 F_2^2 F_3^2| \exp[-i\lambda_2^d] \\ & \quad + |F_1^2 F_2^2 F_3^2| \exp[-i(\lambda_2^d + \lambda_3^d)] \\ & \quad - |F_1^2 F_2^1 F_3^1| \exp[-i\lambda_1^d] \\ & \quad + |F_1^2 F_2^1 F_3^2| \exp[-i(\lambda_1^d + \lambda_3^d)] \\ & \quad + |F_1^2 F_2^2 F_3^1| \exp[-i(\lambda_1^d + \lambda_2^d)] \\ & \quad - |F_1^2 F_2^2 F_3^2| \exp[-i(\lambda_1^d + \lambda_2^d + \lambda_3^d)] \}. \quad (34) \end{aligned}$$

The term  $\{ \dots \}$  does not depend on the triplet  $\Psi_{123}^{111}$  itself and can be expressed as  $A_{123}^{111} \exp(i\Lambda_{123}^{111})$ . In this way, combination of (29) with (34) yields

$$\begin{aligned} & |R_1^d R_2^d R_3^d| \exp(i\Psi_{123}^d) \\ &= A_{123}^{111} \exp [i(\Psi_{123}^{111} + \Lambda_{123}^{111})]. \quad (35) \end{aligned}$$

Obviously, (35) can be combined with (23) and (27). While  $\Psi_{123}^d$  is concentrated around  $-\Delta_{123}^d$ ,  $\Psi_{123}^{111}$  is concentrated around  $(-\Delta_{123}^d - \Lambda_{123}^d)$ . Insertion of (35) into (23) gives

$$\begin{aligned} & P(\Psi_3^{111} | R_1^d R_2^d R_3^d) \\ &= (L^{111})^{-1} \exp [2G_{123}^{111} \cos(\Psi_{123}^{111} + \Lambda_{123}^{111} + \Delta_{123}^d)] \quad (36) \end{aligned}$$

and, finally,

$$\begin{aligned} & \langle \exp(i\Psi_{123}^{111}) \rangle = [I_1(2G_{123}^{111}) / I_0(2G_{123}^{111})] \\ & \quad \times \exp[-i(\Lambda_{123}^{111} + \Delta_{123}^d)] \quad (37) \end{aligned}$$

with

$$G_{123}^{111} = z_{123}^d A_{123}^{111} / z_1^d z_2^d z_3^d \quad (38)$$

and  $L^{111}$  the normalization constant  $2\pi I_0(2G_{123}^{111})$ .

Expressions for the other triplets (33) can be set up in a similar way.

### 2.3. Test results and discussion

One of the main results of this paper is the c.p.d. of the triplet given the SD magnitudes  $|F_1^d|$ ,  $|F_2^d|$  and  $|F_3^d|$  [(36)]. The predictive quality of this expression has been assessed and will be compared with that of two previously published triplet estimating expressions:

(1) the well known Cochran distribution (Cochran, 1955); and

(2) the triplet distribution of P&S. The latter formula encompasses both the triplet expressions of Hauptman (1982b) and Giacovazzo (1983) in the SAS case and the triplet expression of Giacovazzo, Cascarano & Zheng (1988) in the SIRNAS case.

In KPS1 and KPS2, it was shown that correct doublet estimates are essential for correct triplet phase-sum estimates. In § 2.1, it was demonstrated that application of the j.p.d. theory to a single-difference structure factor leads to an expression for the doublet estimates that is identical with that derived by algebraic analysis in KPS2. In analogy with KPS2, the following estimates for the doublets have been considered:

(a) the ALG estimate based on the simple probabilistic expression (12);

(b) the PAT estimate - this modified ALG estimation technique employs interatomic vector information from a special difference Patterson synthesis;

(c) the true doublet values (denoted as the TRUE estimate).

To facilitate a comparison with the results of KPS1 and KPS2, the same test structures have been selected.

(1) Randomly generated structures with only one heavy atom in the unit cell and different DR: Pt-C<sub>62</sub>N<sub>15</sub>O<sub>22</sub>, Pt-C<sub>248</sub>N<sub>63</sub>O<sub>88</sub>, Pt-C<sub>496</sub>N<sub>127</sub>O<sub>176</sub>, Pt-C<sub>744</sub>N<sub>191</sub>O<sub>264</sub>.

(2) Randomly generated structures with four atoms (two different heavy-atom types) in the unit cell and different DR: Hg<sub>3</sub>Pt-C<sub>59</sub>N<sub>15</sub>O<sub>22</sub>, Hg<sub>3</sub>Pt-C<sub>245</sub>N<sub>63</sub>O<sub>88</sub>, Hg<sub>3</sub>Pt-C<sub>493</sub>N<sub>127</sub>O<sub>176</sub>, Hg<sub>3</sub>Pt-C<sub>741</sub>N<sub>191</sub>O<sub>264</sub>.

The structural types (1) and (2) belong to space group *P1* and have been constructed in such a way that the ratio of C, O and N atoms is comparable with that in known proteins. The resolution and the unit-cell parameters have been chosen on similar grounds.

(3) The real protein structure APP [avian pancreatic polypeptide (Blundell, Pitts, Tickle, Wood & Wu, 1981)] taken from the Protein Data Bank (PDB) at Brookhaven National Laboratory (Bernstein *et al.*, 1977; Abola, Bernstein, Bryant, Koetzle & Weng, 1987). APP is a small protein crystallizing with Zn<sup>2+</sup> in space group *C2* with one molecule of 36 amino-acid residues in the asymmetric unit (302 atoms) and unit-cell parameters  $a = 34.18$ ,  $b = 32.92$ ,  $c = 28.44$  Å,  $\beta = 105.30^\circ$  and  $Z = 4$ . The structure was solved originally by SIRAS. The heavy-atom derivative includes one Hg atom. In the PDB release of July 1991, this structure is known as 1PPT.

In all cases, s.f.s have been calculated from the atomic coordinates.

The test results to be presented here involve the following isomorphous data-set combinations.

(a) SAS case. The isomorphous data sets are the Friedel-related-index sets  $\{H(S_1)\}$  and  $\{-H(S_1)\}$  both from the same structure  $S_1$ .

(b) 2DW case. In this case, the isomorphous data sets used are  $\{H(\lambda_1)\}$  and  $\{-H(\lambda_2)\}$ , with  $\lambda_1 = \text{Cr } K\alpha$  radiation and  $\lambda_2 = \text{Fe } K\alpha$  radiation (Cr  $K\alpha$ -Cu  $K\alpha$  for APP).

Table 1. Cumulative statistics of the triplet phase sums for different probabilistic expressions

SAS case, space group  $P1$ , resolution 2.3 Å, radiation Cr  $K\alpha$ , strongest 250  $|E_p|$  values used.

Legend for this and other tables:

COCHRAN	Triplet estimation with the Cochran (1955) distribution
P&S (ALG)	Triplet estimation with the P&S probabilistic theory with doublet estimates by the algebraic technique
P&S (PAT)	Triplet estimation with the P&S probabilistic theory with doublet estimates by the improved algebraic technique
P&S (TRUE)	Triplet estimation with the P&S probabilistic theory when the doublet estimates are equal to the true doublet values
SD (ALG)	Triplet estimation with the presenting probabilistic theory with doublet estimates by the algebraic technique
SD (PAT)	Triplet estimation with the presenting probabilistic theory with doublet estimates by the improved algebraic technique
SD (TRUE)	Triplet estimation with the presenting probabilistic theory when doublet estimates are equal to the true doublet values

Abbreviations

<i>HKL</i>	Triplet indices	EST	Triplet estimation using (36)
DR	Diffraction ratio (KPS1)	TRUE	True triplet values
<i>W</i>	Reliability factor of the distribution (the <i>W</i> values for the SD distribution should be multiplied by $10^{-3}$ )		
NTR	Number of the triplets involved in the statistic (for the P&S and SD distributions this number should be multiplied by 8)		
AER	Mean absolute triplet error in mc (1000 mc = $2\pi$ rad) [(39)]		
ERR	Mean triplet error in mc [(40)]		

(a) Structure: Pt-C<sub>62</sub>N<sub>15</sub>O<sub>22</sub>; DR = 0.26; negative doublets: 5

COCHRAN			P&S (ALG)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.4	273	95	120	300	38	40	100	123	36	38
0.2	1335	113	90	1161	38	42	95	1104	36	39
0.0	3750	116	0	3750	50	56	0	3750	46	51

  

SD (ALG)				SD (TRUE)			
W	NTR	AER	ERR	W	NTR	AER	ERR
6.7	193	7	8	7.1	152	5	5
5.9	1121	13	14	6.2	1128	6	6
0.0	3750	14	17	0.0	3750	7	7

(b) Structure: Pt-C<sub>248</sub>N<sub>63</sub>O<sub>88</sub>; DR = 0.17; negative doublets: 20

COCHRAN			P&S (ALG)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.4	133	128	25	234	59	67	30	215	61	64
0.2	1019	145	15	1150	75	91	15	1093	63	69
0.0	2232	156	0	2232	84	103	0	2232	70	80

  

SD (ALG)				SD (TRUE)			
W	NTR	AER	ERR	W	NTR	AER	ERR
2.0	237	30	38	2.4	248	11	11
1.6	1223	35	45	1.8	1287	13	13
0.0	2232	45	60	0.0	2232	14	14

(c) Structure: Pt-C<sub>496</sub>N<sub>127</sub>O<sub>176</sub>; DR = 0.13; negative doublets: 25

COCHRAN			P&S (ALG)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.5	103	149	15	177	88	128	20	103	60	75
0.3	552	151	10	488	100	131	10	452	79	97
0.0	900	163	0	900	100	130	0	900	84	106

  

SD (ALG)				SD (TRUE)			
W	NTR	AER	ERR	W	NTR	AER	ERR
0.8	190	41	60	1.1	168	17	17
0.7	417	55	73	0.9	407	18	19
0.0	900	65	93	0.0	900	27	30

(d) Structure: Pt-C<sub>744</sub>N<sub>191</sub>O<sub>264</sub>; DR = 0.11; negative doublets: 24

COCHRAN			P&S (ALG)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.4	107	182	10.0	220	93	131	10.0	273	75	90
0.2	389	178	7.5	384	93	131	5.5	453	86	107
0.0	880	184	0.0	880	101	135	0.0	880	91	115

  

SD (ALG)				SD (TRUE)			
W	NTR	AER	ERR	W	NTR	AER	ERR
0.5	125	61	81	0.7	161	23	23
0.4	507	61	87	0.6	320	25	25
0.0	880	65	91	0.0	880	35	37

Table 2. *Cumulative statistics of the triplet phase sums for different probabilistic expressions*SAS case, space group  $P1$ , resolution 2.3 Å, radiation Cr  $K\alpha$ , strongest 250  $|E_v|$  values used.(a) Structure:  $Hg_3Pt-C_{59}N_{15}O_{22}$ ; DR = 0.34; negative doublets: 0

COCHRAN			P&S (ALG)				P&S (PAT)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.45	251	45	900	181	38	44	1500	149	15	16	2500	123	14	15
0.30	1020	50	650	770	34	40	1000	1041	23	25	1000	1313	27	31
0.00	3750	80	0	3750	49	60	0	3750	48	58	0	3750	48	58
SD (ALG)			SD (PAT)				SD (TRUE)							
W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR			
6.1	184	58	72	25	178	4	4	28	171	3	3			
5.5	1131	53	65	14	1032	9	10	17	1001	9	9			
0.0	3750	51	62	0	3750	47	60	0	3750	42	51			

(b) Structure:  $Hg_3Pt-C_{245}N_{63}O_{88}$ ; DR = 0.28; negative doublets: 1

COCHRAN			P&S (ALG)				P&S (PAT)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.40	159	76	90	162	34	38	250	108	26	28	250	108	26	28
0.30	610	85	70	639	41	48	150	671	33	36	150	682	33	36
0.00	2555	100	0	2555	51	60	0	2555	53	61	0	2555	52	60
SD (ALG)			SD (PAT)				SD (TRUE)							
W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR			
4.2	135	73	81	20	159	6	6	20	115	4	4			
3.5	778	59	69	14	638	9	10	13	721	7	7			
0.0	2555	63	75	0	2555	38	47	0	2555	35	44			

(c) Structure:  $Hg_3Pt-C_{493}N_{127}O_{176}$ ; DR = 0.23; negative doublets: 2

COCHRAN			P&S (ALG)				P&S (PAT)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.5	164	121	35	179	65	80	80	138	41	46	80	173	41	46
0.3	598	110	25	525	70	84	25	578	62	72	30	549	57	66
0.0	1172	118	0	1024	69	85	0	1024	70	85	0	1024	68	83
SD (ALG)			SD (PAT)				SD (TRUE)							
W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR			
2.6	109	73	83	10	107	17	18	10	120	7	7			
2.0	498	77	92	5	497	22	23	6	468	12	13			
0.0	1172	75	92	0	1172	51	65	0	1172	42	52			

(d) Structure:  $Hg_3Pt-C_{741}N_{191}O_{264}$ ; DR = 0.20; negative doublets: 4

COCHRAN			P&S (ALG)				P&S (PAT)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.25	289	136	20.0	213	77	105	20.0	344	58	72	25.0	311	56	68
0.20	637	134	8.5	561	83	107	8.0	560	79	99	8.0	557	77	96
0.00	809	135	0.0	809	84	108	0.0	809	87	111	0.0	809	87	110
SD (ALG)			SD (PAT)				SD (TRUE)							
W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR			
1.7	110	71	84	5.6	120	13	13	5.3	164	11	11			
1.4	386	74	98	2.1	413	34	42	2.5	418	28	33			
0.0	809	89	119	0.0	809	72	99	0.0	809	67	89			

(c) SIRNAS. In this case, the isomorphous data sets are defined as  $\{H(S_1)\}$  and  $\{H(S_2)\}$  with  $S_1$  the heavy-atom derivative and  $S_2$  the native protein.

In all performed tests, data up to 2.3 Å resolution and Cr  $K\alpha$  radiation were used (for the APP structure, data up to 2.0 Å resolution and Cu  $K\alpha$  radiation).

Cumulative statistics of the triplet phase-sum estimates for three different probabilistic expressions (Cochran, P&S and SDs) are shown in Tables 1, 2 and 3 for the SAS case for the structural types 1, 2 and 3, respectively. In each table, seven sets of four

variables are present. In Table 1, the ALG and PAT estimation techniques are identical (only one heavy atom in the unit cell, see KPS2) and because of this the number of statistics sets is reduced to five. The first set lists the Cochran-distribution statistics. The next three sets give the P&S statistics when the doublets are estimated by the ALG and PAT estimation techniques and when TRUE doublets are used in the probabilistic expression. The last three sets illustrate the SD statistics for the ALG, PAT and TRUE doublet-estimation techniques. The four variables involved in each set are: the reliability factor of the distribution,

Table 3. *Cumulative statistics of the triplet phase sums for different probabilistic expressions*

SAS case, space group  $C2$ , resolution  $2.0 \text{ \AA}$ , radiation  $\text{Cu } K\alpha$ , strongest 250  $|E_v|$  values used, structure APP, DR=0.11, negative doublets: 7.

COCHRAN			P&S (ALG)				P&S (PAT)				P&S (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.25	315	140	20.0	185	71	97	20.0	255	48	52	30.0	255	48	52
0.20	906	141	15.0	894	71	92	15.0	1225	55	64	20.0	1021	51	57
0.00	3750	153	0.0	3750	77	97	0.0	3750	80	100	0.0	3750	79	98

  

SD (ALG)				SD (PAT)				SD (TRUE)			
W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.37	122	52	56	0.56	110	17	18	0.80	197	11	11
0.28	1230	58	70	0.40	1018	25	27	0.55	1049	12	12
0.00	3750	69	89	0.00	3750	58	76	0.00	3750	35	52

Table 4. *SD theory – a representative sample of 6 ( $\times 8$ ) triplets*

SAS case, structure APP, space group  $C2$ , resolution  $2.0 \text{ \AA}$ , radiation  $\text{Cu } K\alpha$ , strongest 250  $|E_v|$  values used, DR=0.11.

$H$	$K$	$L$	EST	TRUE	ERR	W	$E_H - E_{-H}$	$E_K - E_{-K}$	$E_L - E_{-L}$
11	60	235	-161	-134	27	0.316	-0.042	-0.115	-0.013
			-177	-149	28				
			-165	-142	23				
			-181	-157	24				
			-147	-117	30				
			-163	-132	31				
			-152	-125	27				
			-168	-140	28				
			-95	-95	0				
			-113	-116	3				
16	120	195	-81	-78	3	0.446	-0.020	0.021	-0.095
			-99	-99	0				
			-82	-80	2				
			-100	-101	2				
			-68	-63	5				
			-86	-84	2				
			-95	-95	0				
			-113	-116	3				
			-81	-78	3				
			-99	-99	0				
19	180	191	38	42	4	0.400	-0.057	-0.064	-0.034
			58	66	8				
			20	21	1				
			40	45	5				
			47	53	6				
			67	77	10				
			29	32	3				
			49	56	7				
			379	350	29				
			368	335	33				
46	120	173	365	333	32	0.277	-0.131	0.021	0.117
			354	318	36				
			383	357	26				
			372	342	30				
			369	340	29				
			358	325	33				
			379	350	29				
			368	335	33				
			365	333	32				
			354	318	36				
47	90	161	231	194	37	0.254	-0.099	-0.030	0.056
			217	179	38				
			244	210	34				
			231	195	36				
			238	205	33				
			225	190	35				
			252	221	31				
			239	206	33				
			314	256	58				
			314	250	64				
119	120	178	328	273	55	0.222	-0.069	0.021	0.140
			328	267	61				
			323	267	56				
			323	261	62				
			337	284	53				
			337	278	59				
			314	256	58				
			314	250	64				
			328	273	55				
			328	267	61				



Table 5. Cumulative statistics of the triplet phase sums for different probabilistic expressions

2DW case, space group  $P1$ , resolution 2.3 Å, radiations Cr  $K\alpha$ -Fe  $K\alpha$ , strongest 250  $|E_v|$  values used.(a) Structure: Pt-C<sub>62</sub>N<sub>15</sub>O<sub>22</sub>; DR = 0.029; negative doublets: 2

COCHRAN			SD (ALG)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.3	292	88	0.011	130	23	25	0.012	121	20	22
0.2	1222	97	0.008	1052	25	26	0.009	1042	22	24
0.0	3750	113	0.000	3750	25	27	0.000	3750	25	26

(b) Structure: Pt-C<sub>248</sub>N<sub>63</sub>O<sub>88</sub>; DR = 0.019; negative doublets: 10

COCHRAN			SD (ALG)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.3	133	128	0.0033	140	47	56	0.0047	138	30	32
0.2	1019	145	0.0020	1087	52	65	0.0026	1176	34	36
0.0	2232	156	0.0000	2232	55	71	0.0000	2232	36	39

(c) Structure: Pt-C<sub>496</sub>N<sub>127</sub>O<sub>176</sub>; DR = 0.014; negative doublets: 10

COCHRAN			SD (ALG)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.5	103	149	0.0012	114	80	113	0.0020	133	53	56
0.3	552	151	0.0007	513	88	127	0.0012	427	57	66
0.0	900	163	0.0000	900	91	131	0.0000	900	63	76

(d) Structure: Pt-C<sub>744</sub>N<sub>191</sub>O<sub>264</sub>; DR = 0.012; negative doublets: 9

COCHRAN			SD (ALG)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.4	107	182	0.0008	117	88	118	0.0015	114	59	63
0.2	389	178	0.0005	435	88	125	0.0008	385	63	71
0.0	880	184	0.0000	880	91	131	0.0000	880	70	88

$W$ ;<sup>\*</sup> the number of triplets involved in the statistics, NTR; the absolute mean difference in mc (1000 mc = 2π rad),

$$\text{AER} = \langle |\psi_3|_{\text{true}} - |\psi_3|_{\text{est}} \rangle; \quad (39)$$

and the mean difference in mc,

$$\text{ERR} = \langle |\psi_3|_{\text{true}} - \psi_3|_{\text{est}} \rangle. \quad (40)$$

The data in Table 1 show that an enormous error reduction is gained when (36) is applied, compared with the P&S expression. Table 1(d) shows that, even for a small DR (0.11) and a relatively large number of negative doublets, an acceptable overall error can still be obtained. Another striking difference between (36) and the P&S expression is the much lower triplet error for the former if the true doublet values are used. This demonstrates that, provided the doublet estimates are correct, the triplet estimation *via* (36) is much better than can be achieved by the P&S formula.

The importance of the inclusion of interatomic-vector information from the difference Patterson synthesis is illustrated in Table 2. An enormous error reduction occurs with a change from ALG to PAT in the SD case, in particular for the most reliably estimated triplets. A second conclusion is that although

the SD (ALG) technique yields slightly worse results than the P&S (ALG) technique, in particular for larger DR, the SD (PAT) technique is invariably much better than both the P&S (ALG) and P&S (PAT) techniques. Similar trends are present in the data for the small protein APP, listed in Tables 3 and 4.

The construction of Tables 5 to 8 (2DW case) is similar to that of Tables 1 to 4. The 2DW data have very small diffraction ratios in the range 0.012–0.047. In this range, much smaller than 0.1, the P&S expression fails to give reasonable estimates because of the highly correlated data, so only the Cochran estimates are listed. From the tables, it can be judged that even for DR as small as 0.02, the SD (PAT) technique results in reliable estimates. For DR smaller than 0.02 (see Tables 5c, d and e), the average error of the most reliably estimated triplets increases rapidly. This may be caused partly by the relatively large number of negative doublets. However, if true doublets are employed in the triplet estimation, the average error in the SD (PAT) technique also increases so the deterioration cannot be attributed to the doublets alone. The APP data in Table 7 show that, even for small DR (0.047), an acceptable overall triplet phase-sum error is possible. In Table 8, a representative set of triplets of APP show that, in spite of the very small doublet estimates (concentrated near zero owing to the small DR), the triplet phase-sum estimates may be found anywhere and are correctly estimated in the interval  $(-\pi, \pi)$ .

\* For the Cochran distribution,  $W = 2w_{123}$  [ $w_{123}$  is defined in (22)]. For the P&S distribution,  $W = 2W_{uvw}$  with  $u, v, w = 1, 2$  [ $W_{uvw}$  is defined in equation (21) of KPS2]. For the SD distribution,  $W = 2G_{123}^{lmn}$  with  $l, m, n = 1, 2$  [ $G_{123}^{lmn}$  is defined in (38)].

Table 6. Cumulative statistics of the triplet phase sums for different probabilistic expressions

2DW case, space group  $P1$ , resolution 2.3 Å, radiations Cr  $K\alpha$ -Fe  $K\alpha$ , strongest 250  $|E_\nu|$  values used.(a) Structure:  $Hg_3Pt-C_{59}N_{15}O_{22}$ ; DR = 0.04; negative doublets: 0

COCHRAN			SD (ALG)				SD (PAT)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.45	251	45	0.0088	194	54	71	0.035	214	23	26	0.040	201	24	27
0.30	1020	50	0.0079	1065	57	70	0.017	1371	25	28	0.020	1303	24	28
0.00	3750	80	0.0000	3750	53	65	0.000	3750	51	66	0.000	3750	48	60

(b) Structure:  $Hg_3Pt-C_{245}N_{63}O_{88}$ ; DR = 0.04; negative doublets: 0

COCHRAN			SD (ALG)				SD (PAT)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.40	159	76	0.0063	127	67	76	0.029	143	21	23	0.029	138	23	24
0.25	1050	89	0.0046	1018	62	72	0.015	1021	22	24	0.015	1074	24	26
0.00	2555	100	0.0000	2555	68	79	0.000	2555	46	56	0.000	2555	44	54

(c) Structure:  $Hg_3Pt-C_{493}N_{127}O_{176}$ ; DR = 0.03; negative doublets: 1

COCHRAN			SD (ALG)				SD (PAT)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.60	134	125	0.0037	113	70	80	0.014	109	25	27	0.015	108	26	27
0.30	598	110	0.0026	572	75	92	0.006	585	32	34	0.007	591	31	34
0.00	1172	118	0.0000	1172	77	95	0.000	1172	56	69	0.000	1172	52	64

(d) Structure:  $Hg_3Pt-C_{741}N_{191}O_{264}$ ; DR = 0.02; negative doublets: 2

COCHRAN			SD (ALG)				SD (PAT)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.35	97	145	0.0024	136	81	104	0.008	156	29	31	0.0093	103	30	31
0.20	637	134	0.0019	366	89	117	0.003	392	40	50	0.0026	497	44	53
0.00	809	135	0.0000	809	95	128	0.000	809	76	105	0.0000	809	72	101

Table 7. Cumulative statistics of the triplet phase sums for different probabilistic expressions

2DW case, space group  $C2$ , resolution 2.0 Å, radiations Cu  $K\alpha$ -Cr  $K\alpha$ , strongest 250  $|E_\nu|$  values used, structure: APP, DR = 0.047, negative doublets: 5.

COCHRAN			SD (ALG)				SD (PAT)				SD (TRUE)			
W	NTR	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
0.25	309	136	0.026	181	47	52	0.038	164	32	35	0.064	184	21	22
0.20	890	138	0.019	1070	62	76	0.026	950	36	38	0.040	906	24	25
0.00	3750	153	0.000	3750	78	99	0.000	3750	68	88	0.000	3750	49	66

At this point, it should be mentioned that, although in the Cochran and P&S distributions the reliability factors may take large values, in (36) they can be very small, even for very accurate triplet estimation. This behaviour for  $W$  occurs because of the different kind of r.v.s involved in SDs. The maximum value for the  $W$  of the SD distribution is given approximately as  $W_{\max} = n^{-1/2} R_1^d R_2^d R_3^d$ .  $n$  and the  $R_\nu^d$  ( $\nu = 1, 2, 3$ ) are very small quantities and consequently  $W_{\max}$  is also small. For example, if  $n = 1$  (one heavy atom in the unit cell) and the  $R_\nu^d$  values (expressed in terms of  $E$  values) range from 0.01-0.1,  $W_{\max}$  will be of the order  $10^{-6}$ - $10^{-3}$ , respectively.

Although the P&S expressions and (36) contain apparently similar terms and use the same data, i.e. the magnitudes of the observed structure factors and the contents of the unit cell, they are not identical, as indicated by the differences shown in the tables and those in  $W$ . An explanation for this seemingly paradoxical difference in results given the same data

may be found in the different starting points for the j.p.d.s. Apparently, the *ab initio* definition of SDs of isomorphous s.f.s to be r.v.s, on the one hand, and the simultaneous use of isomorphous s.f.s to be r.v.s, on the other hand, exploit the present data in a conceptually different way.

In the SAS and 2DW cases, the new approach is able to reduce the error level of the triplet phase-sum estimates such that DM may be applied. However, in the SIRNAS and SIRAS cases, a sign ambiguity still exists. Once this sign problem is solved, the SD theory may lead to protein structure determination by means of DM in the SIRNAS/SIRAS cases as well. This is illustrated in Table 9 by the cumulative triplet statistics of APP in the SIRNAS case. Half of the doublet population is negative so the triplet statistics are incorrect. However, if the doublet signs are assumed to be known, the triplet estimates are almost perfect: in total, 2495 ( $\times 8$ ) out of 3750 ( $\times 8$ ) triplets are estimated with an average error of 1 mc. It is

Table 8. *SD theory – a representative sample of 6 (×8) triplets*

2DW case, structure APP, space group  $C2$ , resolution 2.0 Å, radiations Cr  $K\alpha$ -Cu  $K\alpha$ , strongest 250  $|E_\nu|$  values used, DR = 0.05.

$H$	$K$	$L$	EST	TRUE	ERR	$W$	$E_{H(\lambda_1)} - E_{H(\lambda_2)}$	$E_{K(\lambda_1)} - E_{K(\lambda_2)}$	$E_{L(\lambda_1)} - E_{L(\lambda_2)}$
10	150	228	101	14	87	0.0114	0.0005	-0.060	-0.001
			94	9	85				
			106	20	86				
			100	15	85				
			102	15	87				
			96	10	86				
			109	21	88				
			102	16	86				
19	180	191	43	37	6	0.0349	-0.025	-0.031	-0.020
			52	47	5				
			34	28	6				
			43	38	5				
			47	41	6				
			56	51	5				
			39	32	7				
			48	42	6				
26	30	47	-155	-159	4	0.0398	-0.011	-0.033	-0.044
			-159	-164	5				
			-161	-166	5				
			-165	-171	6				
			-149	-152	3				
			-153	-157	4				
			-155	-159	4				
			-159	-164	5				
31	120	207	205	242	37	0.0190	-0.062	0.009	0.000
			199	235	36				
			200	235	35				
			193	228	35				
			208	244	36				
			202	237	35				
			203	237	34				
			196	230	34				
36	150	170	421	441	20	0.0313	-0.053	-0.060	0.049
			417	436	19				
			427	447	20				
			423	442	19				
			425	445	20				
			421	440	19				
			431	451	20				
			427	446	19				
53	60	173	327	284	43	0.0148	-0.012	-0.045	0.043
			324	278	46				
			331	288	43				
			328	282	46				
			332	289	43				
			329	283	46				
			336	293	43				
			333	287	46				

Table 9. *Cumulative statistics of the triplet phase sums for different probabilistic expressions*

Structure APP, SIRNAS case, space group  $C2$ , resolution 2.0 Å, radiation Cu  $K\alpha$ , strongest 250  $|E_\nu|$  values used, DR = 0.56, negative doublets: 105.

COCHRAN			P&S (PAT)				P&S (TRUE)				
$W$	NTR	ERR	$W$	NTR	AER	ERR	$W$	NTR	AER	ERR	
0.25	289	132	1.0	424	144	221	1.5	253	68	73	
0.15	1770	152	0.5	1798	139	212	0.5	1685	71	79	
0.00	3750	176	0.0	3750	144	217	0.0	3750	104	126	
			SD (PAT)				SD (TRUE)				
$W$	NTR	AER	ERR	$W$	NTR	AER	ERR	$W$	NTR	AER	ERR
	200	102	79	158		230	375	0.7	0.7		
						140	1393	0.8	0.8		
	100	2099	111	201		40	2495	1.0	1.0		
						30	2680	6.9	12.7		
						20	2896	14.6	26.1		
						10	3194	24.3	46.3		
	0	3750	129	213		0	3750	40.6	76.1		

Table 10. *SD theory for very low DRs - cumulative statistics of the triplet phase sums*

Structure Si-C<sub>744</sub>N<sub>191</sub>O<sub>264</sub>, SIRNAS case, space group P1, resolution 2.3 Å, radiation Cr Kα, strongest 250 |E<sub>ν</sub>| values used, DR = 0.02, negative doublets: 78.

W	SD (TRUE)		
	NTR	AER	ERR
320	100	45	46
190	465	46	49
0	806	54	57

promising that this high-reliability triplet estimation can be achieved even for very low DR or very 'light' heavy atoms (Table 10). Fan, Han, Qian & Yao (1984) and Klop, Krabbendam & Kroon (1990) proposed expressions to solve this sign problem. However, the expressions developed in these papers depend on  $N^{-1/2}$  and consequently the reliability of these formulae is low for large structures. In view of the  $n$  dependence ( $n \ll N$ ) of the SD expressions, the difference structure factors are expected to play an important role in this sign determination. The extension of the SD theory to the estimation of other types of (sem-)invariants is obvious. In particular, quartet phase sums are expected to be of importance for the solution of large molecules (Sheldrick, 1990). Research on these subjects is already in progress. Finally, it is worth mentioning that the SD theory can also be applied to the case of (calculated) partial structure factors and structure factors of the associated complete structure (with both considered as isomorphous data sets). This may be helpful not only for protein structures but also for difficult small structures.

One of us (CEK) gratefully acknowledges financial support by the Commission of the European Communities, under the project B/BIOT 900103-001.

## APPENDIX

### The j.p.d. of a single structure factor

The derivation of the j.p.d. of a single structure factor  $F_\nu$ , where

$$F_\nu = |F_\nu| \exp(i\varphi_\nu) = \sum_{j=1}^N f_{j\nu} \exp(2\pi i \mathbf{H}_\nu \cdot \mathbf{r}_j), \quad (A1)$$

comes down to the derivation of the j.p.d. of the r.v.s  $R_\nu$  for the magnitude  $|F_\nu|$  and  $\Phi_\nu$  for the phase  $\varphi_\nu$ . The j.p.d.  $P(R_\nu, \Phi_\nu)$  can be written as (see, for example, Karle & Hauptman, 1958)

$$P(R_\nu, \Phi_\nu) = R_\nu (2\pi)^{-2} \int_0^\infty \int_0^{2\pi} \rho \times \exp[-i\rho R_\nu \cos(\theta - \Phi_\nu)] \times C(\rho, \theta) d\theta d\rho, \quad (A2)$$

where  $C$  is the c.f.,

$$C = \left\langle \exp \left[ i \left( \sum_{j=1}^N \rho f_{j\nu} \cos(2\pi \mathbf{H}_\nu \cdot \mathbf{r}_j - \theta) \right) \right] \right\rangle. \quad (A3)$$

With the assumption that the p.r.v.s for the atomic coordinates are independent, (A3) becomes

$$C = \exp \left( \sum_{j=1}^N \{ \ln \langle \exp [i\rho f_{j\nu} \cos(2\pi \mathbf{H}_\nu \cdot \mathbf{r}_j - \theta)] \rangle_{\mathbf{r}_j} \} \right). \quad (A4)$$

Evaluation of (A4) (see, for example, Peschar, 1991) gives

$$C = \exp(-z_\nu \rho^2 / 4) \quad (A5)$$

with

$$z_\nu = \sum_{j=1}^N |f_{j\nu}|^2. \quad (A6)$$

With the variable transformation  $\rho = tz_\nu^{-1/2}$  and substitution of (A5) into (A2), one gets

$$P(R_\nu, \Phi_\nu) = R_\nu (2\pi)^{-2} z_\nu^{-1} \int_0^\infty \int_0^{2\pi} t \times \exp[-itR_\nu(z_\nu)^{-1/2} \cos(\theta - \Phi_\nu) - t^2/4] dt. \quad (A7)$$

With the integral formula (Giacovazzo, 1980, equation E.16)

$$(2\pi)^{-1} \int_0^\infty \int_0^{2\pi} \exp(-p^2 t^2 - iat \cos \varphi) t dt d\varphi = (2p^2)^{-1} \exp[-a^2(4p^2)^{-1}], \quad (A8)$$

(A7) becomes

$$P(R_\nu, \Phi_\nu) = 2R_\nu(z_\nu)^{-1} \exp[-R_\nu^2(z_\nu)^{-1}]. \quad (A9)$$

## References

- ABOLA, E. E., BERNSTEIN, F. C., BRYANT, S. H., KOETZLE, T. F. & WENG, J. (1987). In *Crystallographic Databases - Information Content, Software Systems, Scientific Applications*, edited by F. H. ALLEN, G. BERGERHOFF & S. SIEVERS. Bonn/Cambridge/Chester: Data Commission of the International Union of Crystallography.
- ABRAMOWITZ, M. & STEGUN, I. A. (1970). *Handbook of Mathematical Functions*. New York: Dover.
- BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.
- BEURSKENS, P. T., BEURSKENS, G., BOSMAN, W. P., GARCIA-GRANDA, S., GOULD, R. O., SMITS, J. M. M. & SMYKALLA, C. (1991). *Direct Methods for Solving Crystal Structures*, edited by H. SCHENK, pp. 369-378. New York: Plenum.
- BEURSKENS, P. T., PRICK, A. J., DOESBURG, H. M. & GOULD, R. O. (1979). *Acta Cryst.* **A35**, 765-772.
- BEURSKENS, P. T. & SMYKALLA, C. (1991). *Direct Methods for Solving Crystal Structures*, edited by H. SCHENK, pp. 281-290. New York: Plenum.
- BLUNDELL, T. L., PITTS, J. E., TICKLE, I. J., WOOD, S. P. & WU, C. W. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4175-4179.

- CAMALLI, M., GIACOVAZZO, C. & SPAGNA, R. (1985). *Acta Cryst.* **A41**, 605–613.
- COCHRAN, W. (1955). *Acta Cryst.* **8**, 473–478.
- FAN, H., HAN, F., QIAN, J. & YAO, J. (1984). *Acta Cryst.* **A40**, 489–495.
- FORTIER, S. & NIGAM, G. D. (1989). *Acta Cryst.* **A45**, 247–254.
- GIACOVAZZO, C. (1980). *Direct Methods in Crystallography*. London: Academic.
- GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 585–592.
- GIACOVAZZO, C., CASCARANO, G. & ZHENG, C. (1988). *Acta Cryst.* **A44**, 45–51.
- HAUPTMAN, H. (1982a). *Acta Cryst.* **A38**, 289–294.
- HAUPTMAN, H. (1982b). *Acta Cryst.* **A38**, 632–641.
- HULL, S. E. & IRWIN, M. J. (1978). *Acta Cryst.* **A34**, 863–870.
- KARLE, J. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 37–40. Copenhagen: Munksgaard.
- KARLE, J. (1989). *Acta Cryst.* **A45**, 765–781.
- KARLE, J. & HAUPTMAN, H. (1958). *Acta Cryst.* **11**, 264–269.
- KINNEGING, A. J. & DE GRAAFF, R. A. G. (1984). *J. Appl. Cryst.* **17**, 364–366.
- KLOP, E. A., KRABBENDAM, H. & KROON, J. (1990). *Acta Cryst.* **A46**, 517–520.
- KYRIAKIDIS, C. E., PESCHAR, R. & SCHENK, H. (1993a). *Acta Cryst.* **A49**, 350–358.
- KYRIAKIDIS, C. E., PESCHAR, R. & SCHENK, H. (1993b). *Acta Cryst.* **A49**, 359–369.
- PESCHAR, R. (1991). *Direct Methods for Solving Crystal Structures*, edited by H. SCHENK, pp. 111–119. New York: Plenum.
- PESCHAR, R. & SCHENK, H. (1991). *Acta Cryst.* **A47**, 428–440.
- SHELDRIK, G. M. (1990). *Acta Cryst.* **A46**, 467–473.
- SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
- WOOLFSON, M. M. (1987). *Acta Cryst.* **A43**, 593–612.

*Acta Cryst.* (1993). **A49**, 569–572

## How to Obtain Easily the Induced Representations of Point Groups: the Icosahedral Point Groups

BY YVES BILLIET

*Département de Chimie, Faculté des Sciences, Boîte Postale N° 825, Niamey, Niger*

(Received 22 June 1992; accepted 5 November 1992)

### Abstract

With tables of subduced representations as a starting point and use of the Frobenius reciprocity theorem, a simple method to obtain induced representations is given. Tables are given of the 22 induced representations of  $532 (I)$  and of the 84 induced representations of  $\bar{5}32/m (I_h)$ .

### Introduction

The crystallographic and molecular point groups are prime candidates to exemplify various group-theoretical properties. They are well known to both physicists and chemists. Their orders are not large so they lend themselves to non-computerized calculations but neither are they small so they can be used to illustrate the distinct possible cases of group-theoretical properties.

The properties we wish to emphasize here are those of subduced and induced representations of groups: their dimension, additivity, transitivity and how they are related by the Frobenius reciprocity theorem. These properties will be used to construct the induced representations of the icosahedral point groups  $532 (I)$  and  $\bar{5}32/m (I_h)$  from the subduced representations of these groups onto their subgroups.

We have chosen the icosahedral groups because of the interest in them in many fields: electronic (Boyle, 1972) and vibrational (Boyle & Parker, 1980) properties of molecules, coupling coefficients (Fowler &

Ceulemans, 1985), the Jahn–Teller effect (Ceulemans & Fowler, 1989, 1990), inorganic (Pitochelli & Hawthorne, 1960) and biological molecules (Litvin, 1975), and quasicrystals (Schechtman, Blech, Gratias & Cahn, 1984; Jaric, 1988). Recently, Litvin (1991) tabulated many of the basic group-theoretical properties of the icosahedral point groups. Their irreducible representations and character tables are well known (Griffith, 1964; Backhouse & Gard, 1974). We do not give here the fundamental principles of group theory or group representations and instead refer the reader to classic works (Lomont, 1959; Murnaghan, 1963; Gorenstein, 1968; Kirillov, 1976; Serre, 1978; Mal'liavin, 1981).

### I. Notation and basic properties

Consider a finite group  $G$ , a subgroup  $H$ , a representation  $\pi(G)$  of  $G$  and a representation  $\rho(H)$  of  $H$ .

(i) The representation of  $H$  subduced from  $\pi(G)$  is denoted  $\pi(G)\downarrow H$ , while the representation of  $G$  induced by  $\rho(H)$  is denoted  $\rho(H)\uparrow G$ .

(ii) The dimensions of these representations are related:

$$\dim [\pi(G)\downarrow H] = \dim [\pi(G)];$$

$$\dim [\rho(H)\uparrow G] = \dim [\rho(H)] \times |G|/|H|;$$

where  $|G|$  and  $|H|$  are the orders of  $G$  and  $H$ , respectively.