On the Doublet Phase Sums of Isomorphous Data Sets

By Christos E. Kyriakidis, René Peschar and Henk Schenk

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

Abstract

The role of the doublet phase sum present among isomorphous data sets is investigated in connection with the triplet-phase-sum statistics. Several probabilistic and algebraic techniques are discussed to estimate the doublets. The combination of an algebraic estimation technique and a new difference Patterson synthesis, the maxima of which are used to improve iteratively the doublet phase sums, is shown to be successful. Test results for large model structures and idealized protein data show that this technique reduces the triplet-phase-sum errors to a level small enough for *ab initio* direct-methods applications.

1. Introduction

Although direct methods (DM) are used nowadays for the routine determination of structures of as many as 100 independent atoms, their role in solving protein structures *ab initio* seems to be quite limited. After the initial algebraically oriented approach of Kroon, Spek & Krabbendam (1977), the probabilistic integration of DM with the techniques to solve protein structures was undertaken. Expressions for the SIR-NAS* case have been derived by Hauptman (1982*a*) and Giacovazzo, Cascarano & Zheng (1988), those for the single-wavelength anomalous-scattering (SAS) case by Hauptman (1982*b*) and Giacovazzo (1983). Fortier & Nigam (1989) rationalized the similar expressions for the joint probability distributions (j.p.d.s) to be a result of isomorphous data sets. Recently, the full probabilistic integration of DM with any number and type of isomorphous data sets has been accomplished (Peschar & Schenk, 1991). Although test results (Furey, Chandrasekhar, Dyda & Sax 1990) exist which suggest that DM may be applicable in solving protein structures *ab initio*, the full potential of DM in this respect seems not to have been realized as yet.

An important characteristic of all probabilistic expressions in this field is the presence of doublet phase sums between isomorphous structure factors, the role of which has been the subject of some recent

studies. It has been pointed out that in the SAS case the doublets tend to have the same (positive) sign (Guo, 1990; Guo, Blessing & Hauptman, 1991). On the other hand, a doublet sign ambiguity is known to exist in the SIRNAS case. Fortier, Fraser & Moore (1986) analysed this sign ambiguity by cluster analysis. A different approach was followed by Fan Hai-fu and co-workers, who employed various techniques of introducing structural information to solve the sign ambiguity (Fan, Han, Qian & Yao 1984; Fan & Gu, 1985; Hao & Fan, 1988). An analysis of the conventional DM procedure suggests that, in spite of the studies mentioned, the importance of doublets and their use in *ab initio* DM has not been fully exploited.

Ab initio conventional DM rely essentially on the use of triplet phase sums, presumably concentrated around zero, and quartet phase sums, concentrated near zero or π (Schenk, 1973, 1974; Hauptman, 1975; Giacovazzo, 1977). It can be shown in various ways that for a correct estimation of a phase-sum invariant of order N^{-n}, invariants of lower order, $N^{-(n-1/2)}$, are essential (Peschar, 1987). For example, for a correct estimation of the quartet phase sum (order N^{-1}), the triplets (of order $N^{-1/2}$) that add up to the quartet are required. A recent investigation of two probabilistic formulae for the *ab initio* determination of protein structures (Peschar & Schenk, 1991) has revealed that correct doublet phase sums [which are of order $O(N^0)$] are important for a correct triplet-phase-sum evaluation. For this purpose, a new diffraction ratio (DR) was developed which shows a linear relationship with the ideal doublet phase sum as calculated from the atomic coordinates (Kyriakidis, Peschar & Schenk, 1993).

In this paper, only techniques for doublet estimation, which do not require the knowledge of a heavy-atom substructure, will be discussed and their influence on the triplet statistics will be assessed.

2. The j.p.d. estimation of the doublet phase sums

The general definition of a doublet phase sum is given by

$$\varphi_{Hm} + s_{mn}\varphi_{Hn} = \psi_2^{mn} \quad (m, n = 1, 2, \ldots, l-1, l$$

$$\text{isomorphous data sets}) \quad (1)$$

---

* SIR(N)AS: single isomorphous replacement including (neglecting) anomalous-scattering effects.

with

$$s_{mn} = \begin{cases} -1 & \text{if } H_m = H_n, \\ 1 & \text{if } H_m = -H_n. \end{cases} \quad (2)$$

Peschar & Schenk (1991) used the simplified expression

$$z_{mn} = |z_{mn}| \exp (i\Delta_{mn})$$

$$= \sum_{j=1}^{N} |f_{jm}||f_{jn}| \exp [-i(\delta_{jm} + s_{mn}\delta_{jn})]$$

$$\text{for } m \le n \, [m, n \in (1, \ldots, l)] \quad (3)$$

to estimate the doublets directly from the δ_{jm},

$$\delta_{jm} = \tan^{-1} [f''_{jm}/(f^o_{jm} + f'_{jm})], \quad (4)$$

where

$$f_{jm} = f^o_{jm} + f'_{jm} + if''_{jm}$$

$$= f^r_{jm} + if''_{jm}$$

$$= |f_{jm}| \exp (i\delta_{jm}). \quad (5)$$

The marginal j.p.d. of the phases and magnitudes of two isomorphous structure factors F_{Hm} and F_{Hn} up to $O(N^0)$ is obtained from equation (60) of Peschar & Schenk (1991) as

$$P(R_1, \Phi_1, R_2, \Phi_2) = C^{-1} \exp [2G_m G_n |L_{mn}|$$

$$\times \cos (\Phi_{Hm} + s_{mn}\Phi_{Hn} + \lambda_{mn})] \quad (6)$$

with C^{-1} a normalization constant.*

The conditional probability distribution of the doublets may be calculated from (6) in the usual way by fixing the magnitudes and integrating out the phases that do not take part in the doublet phase sums.

Based on the j.p.d. (6), the following doublet-estimation techniques will be used in the test procedures (§ 5):

(1) *ZER*. Consider $\lambda_{mn} = 0$ in (6). In this way, all the doublet estimates are set to zero.

(2) *JPDMOD*. The estimation based on the mode of the distribution (6).

(3) *JPDNUM*. Numerically† estimated $|\psi_2^{mn}|$ using

$$\langle |\psi_2^{mn}| \rangle = \int_0^\pi |\psi_2^{mn}| \exp [2G_m G_n |L_{mn}|$$

$$\times \cos (\psi_2^{mn} + \lambda_{mn})] \, d\psi_2^{mn} \left\{ \int_0^\pi \exp [2G_m G_n |L_{mn}| \right.$$

$$\left. \times \cos (\psi_2^{mn} + \lambda_{mn})] \, d\psi_2^{mn} \right\}^{-1}. \quad (7)$$

---

* For an explanation of the symbols used see Peschar & Schenk (1991).

† For the numerical integration, the Simpson rule has been used with a step of 0.16 mc (1000 mc = 2π rad).

3. The algebraic estimation of the doublet phase sums

Let us define a set of p isomorphous structure factors, in accordance with Peschar & Schenk (1991), as follows,

$$F_{Hm} = \sum_{j=1}^{N} (f^r_{jm} + if''_{jm}) \exp (2\pi i \mathbf{H} \cdot \mathbf{r}_j),$$

$$m = 1, 2, \ldots, p. \quad (8)$$

According to this definition, two structure factors are termed isomorphous if the trigonometric parts of the structure factors are identical. In this way, the structure factors F_H and F^*_H, different because of anomalous scattering, can be considered to be isomorphously related. Formulated differently, the atom pairs $f_j(H)$ and $f^*_j(-H)$ are isomorphously related.

If we define

$$F_{Hm} - F_{Hn} = \sum_{j=1}^{N} [(f^r_{jm} - f^r_{jn})$$

$$+ i(f''_{jm} - f''_{jn})] \exp (2\pi i \mathbf{H} \cdot \mathbf{r}_j) \quad (9)$$

with $m, n = 1, 2, \ldots, p$, then

$$|F_{Hm} - F_{Hn}|^2 = (F_{Hm} - F_{Hn})(F^*_{Hm} - F^*_{Hn})$$

$$= |F_{Hm}|^2 + |F_{Hn}|^2$$

$$- 2|F_{Hm}||F_{Hn}| \cos (\varphi_{Hm} + s_{mn}\varphi_{Hn}). \quad (10)$$

On the other hand, (10) can be expressed alternatively as

$$|F_{Hm} - F_{Hn}|^2$$

$$= \sum_{j=1}^{N} t_{jmn} \exp (2\pi i \mathbf{H} \cdot \mathbf{r}_j) \sum_{k=1}^{N} t^*_{kmn} \exp (-2\pi i \mathbf{H} \cdot \mathbf{r}_k)$$

$$= \sum_{j=1}^{N} |t_{jmn}|^2 + \sum_{\substack{j=1 \\ j \ne k}}^{N} \sum_{k=1}^{N} t_{jmn}t^*_{kmn} \exp [2\pi i \mathbf{H} \cdot (\mathbf{r}_j - \mathbf{r}_k)]$$

$$\quad (11)$$

with

$$t_{jmn} = (f^r_{jm} - f^r_{jn}) + i(f''_{jm} - f''_{jn}). \quad (12)$$

Combination of (10) and (11) leads to

$$\langle \cos \psi_2^{mn} \rangle = (|F_{Hm}|^2 + |F_{Hn}|^2 - \gamma_{mn})/2|F_{Hm}||F_{Hn}| \quad (13)$$

with

$$\gamma_{mn} = \sum_{j=1}^{N} |t_{jmn}|^2 + \sum_{\substack{j=1 \\ j \ne k}}^{N} \sum_{k=1}^{N} t_{jmn}t^*_{kmn} \exp [2\pi i \mathbf{H} \cdot (\mathbf{r}_j - \mathbf{r}_k)]. \quad (14)$$

On the basis of (14), two cases can be distinguished.

Case 1. One pair of isomorphously related atoms with a non-zero difference between the atomic scattering factors. The double summation in (14) vanishes if the difference between two isomorphous structures

is caused by a single pair of non-identical isomorphously related atoms. In this case, γ becomes independent of the interatomic vectors and can be written as

$$\gamma_{mn} = \sum_{j=1}^{N} |t_{jmn}|^2. \qquad (15a)$$

For example, in the SAS case,

$$\gamma_{mn} = 4 \sum_{j=1}^{n} |f''_{jm}|^2, \qquad (15b)$$

which can be seen to depend on the n anomalously scattering atoms only. The expression in the SIRNAS case is

$$\gamma_{mn} = \sum_{j=1}^{n} |f^o_{jm}|^2, \qquad (15c)$$

where n is the number of the heavy atoms in the unit cell. Note that only $\cos \psi_2^{mn}$ can be estimated from (13), not ψ_2^{mn} itself.

Case 2. Several isomorphously related pairs of non-identical atoms. If several pairs of isomorphously related but non-identical atoms are responsible for the difference between two isomorphously related data sets, the double summation in (14) can only be omitted in a first approximation. The contribution of the interatomic vectors can be incorporated by calculating a special difference Patterson synthesis.

According to Rossmann (1960), a Patterson synthesis with coefficients

$$|F_{Hm} - F_{Hn}|^2 = |F_{Hm}|^2 + |F_{Hn}|^2 - 2|F_{Hm}||F_{Hn}| \cos \psi_2^{mn} \qquad (16)$$

is equivalent to a Patterson synthesis with coefficients

$$(|F_{Hm}| - |F_{Hn}|)^2 = |F_{Hm}|^2 + |F_{Hn}|^2 - 2|F_{Hm}||F_{Hn}|, \qquad (17)$$

provided that ψ_2^{mn} is small. The Patterson synthesis with (17) as coefficients can be looked upon as an approximation to the Patterson synthesis with (16) as coefficients using $\cos \psi_2^{mn} = 1$ (or equivalently $\varphi_{Hm} + s_{mn}\varphi_{Hn} = 0$). In analogy with Rossmann (1960), we define the difference Patterson function as

$$P(\mathbf{u}) = \sum_H |F_{Hm} - F_{Hn}|^2 \cos(2\pi \mathbf{H} \cdot \mathbf{u}) \qquad (18)$$

with

$$|F_{Hm} - F_{Hn}|^2 = |F_{Hm}|^2 + |F_{Hn}|^2 - 2|F_{Hm}||F_{Hn}|\langle\cos \psi_2^{mn}\rangle. \qquad (19)$$

An initial estimate for $\cos \psi_2^{mn} (\neq 1)$ is available from (13) if the double summation in (14) is omitted. An analysis of this P function should identify the interatomic vectors between the different isomorphously related atom pairs. The heights of the P peaks are expected to reveal the atomic scatterers involved. In this way, better approximation to $\cos \psi_2^{mn}$ can be calculated from (14).

It should be noted that (18) is identical to an expression proposed by Cascarano & Giacovazzo (1984), who aimed to determine the positions of the anomalous scatterers. Our goal is completely different: the *ab initio* phase determination without prior resort to atomic position determination. The optimization of the doublet-phase-sum estimation and, consequently, the estimation of the triplet phase sums requires interatomic vectors only and no identification or assignment of atomic positions.

In summary, the following scheme has been adopted:

(i) The calculation of $\cos \psi_2^{mn}$ from (13), omitting the double summation in (14). This gives a first, though in general not perfect, estimation for the magnitude of ψ_2^{mn}. In the test procedures, this step will be referred to as the ALG technique. Hence, the estimates are based on expression (13) [with γ defined in (15)].

(ii) The calculation of the Patterson synthesis (18). Assignment of products of atom types to the Patterson maxima. In the tests a visual inspection of the difference P function was used to assign the atomic scattering factors. Recalculation of (18) using additional terms in the double summation (14).

(iii) Step (ii) is repeated, if necessary, until the $\cos \psi_2^{mn}$ values do not change.

Estimation based on the complete three-step scheme will be called the PAT estimation technique, since it is based on the Patterson synthesis (18).

4. The use of doublet phase sums in the estimation of triplet phase sums

The main object of this paper is to show the influence of doublets on the estimation of triplets. Before showing the practical connection of doublets and triplets by use of test results, it is fruitful to show their theoretical relationship by examining carefully the main formula of Peschar & Schenk (1991), which gives the conditional probability distribution for each of the l^3 triplets when l isomorphous structure factors have been involved in the distibution (for the symbols used, consult that paper),

$$P(\Psi_3^{uvw}|R_1, \ldots, R_l)$$

$$= L^{-1} \exp[2|W_{uvw}| \cos(\Psi_3^{uvw} - \zeta_{uvw})] \qquad (20)$$

with

$$|W_{uvw}| \exp(-\zeta_{uvw})$$

$$= \sum_{a'=1}^{l} \sum_{b'=1}^{l} \sum_{c'=1}^{l} G_{a'}G_{b'}G_{c'}\left\{ \sum_{a=1}^{l} \sum_{b=1}^{l} \sum_{c=1}^{l} C_{aa}C_{bb}C_{cc} \right.$$

$$\times |e_{aa'}||e_{bb'}||e_{cc'}||Z_{abc}|B_{a'u}B_{b'v}B_{c'w}$$

$$\times \exp[-i(\varepsilon_{aa'}\varepsilon_{bb'}\varepsilon_{cc'} - \Delta_{abc} + s_{a'}\lambda_{a'u}g_{a'u}$$

$$\left. + s_{b'}\lambda_{b'v}g_{b'v} + s_{c'}\lambda_{c'w}g_{c'w})] \right\}. \qquad (21)$$

The terms ζ_{uvw} depend on the quantities Δ_{abc}, $\varepsilon_{aa'}$ and $\lambda_{a'u}$. The Δ_{abc} are functions of the phases δ_{jm} of the atomic scattering factors. The $\lambda_{a'u}$ are the doublet estimates themselves and the $\varepsilon_{aa'}$ are constructed solely from the $A_{nn'}$, $a_{nn'}$ and $D_{nn'}$. Hence all these quantities are functions of the doublet values. In other words, the triplet estimates depend on three kinds of terms. Except for the Δ_{abc}, which can be calculated readily using only the quantity and type of anomalous scatterers, the other two kinds of terms ($\varepsilon_{aa'}$ and $\lambda_{a'u}$) depend directly on the doublet values. This means that the better the doublet estimates, the more precise are the λ and ε calculations and, consequently, the more accurate the triplet estimates are expected to be. For $l = 2$, (20) reduces in the SAS case to the Hauptman (1982b) and Giacovazzo (1983) distributions. In the SIRNAS case for $l = 2$, it is identical to the Giacovazzo, Cascarano & Zheng (1988) distribution expression.

5. Test results and discussion

The following doublet-estimation techniques, defined in the previous paragraphs, have been tested extensively.

(1) ZER: all the doublet estimates set to zero.

(2) JPDMOD: estimation of ψ_2^{mn} based on the mode of the distribution (6).

(3) JPDNUM: j.p.d. estimation based on numerical integration of (7).

(4) ALG: algebraic estimation based on (13)

(5) PAT: improved algebraic estimation by means of the difference Patterson synthesis (18).

The test procedures focused on four criteria.

(A) *The kind of isomorphism.* From Peschar & Schenk (1991), it follows that the Friedel-related data {H} and {−H} should be considered to be separate data sets that are isomorphously related. To assess the influence of the type of isomorphous data, four different cases of two isomorphously related data sets have been tested.

(1) SAS. The isomorphous data sets are {$H(S_1)$} and {$-H(S_1)$}.

(2) SIRNAS. 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 which the atomic scattering factors are real valued.

(3) SIRAS. The isomorphous data sets are {$H(S_1)$} and {$H(S_2)$}, with S_1 the heavy-atom derivative and S_2 the native protein, in which the atomic scattering factors are complex valued. It should be noted that this definition of SIRAS may differ from that found in the literature. However, from our definition of isomorphism, it follows that the usual definition of

SIRAS (the complete Ewald sphere of data for two anomalously scattering isomorphously related structures) leads to four isomorphously related data sets: {$H(S_1)$}, {$-H(S_1)$}, {$H(S_2)$} and {$-H(S_2)$}.

(4) 2DW (two different wavelengths). The isomorphous data sets used are {$H(\lambda_1)$} and {$-H(\lambda_2)$}, with λ_1 and λ_2 two different wavelengths.

(B) *The quality of isomorphism.* The higher the DR, the lower the quality of isomorphism. To establish the practical limitations of the estimation techniques, structures with different values of DR (0.03–0.90) have been chosen.

(C) *The influence of the number of heavy atoms.* Structures with at least one heavy atom in the unit cell have been used to establish the importance of the difference Patterson information.

(D) *The behaviour of the doublet-estimation techniques.* This was investigated for the cases of randomly generated data and real protein data from the Protein Data Bank (PDB) at Brookhaven National Laboratory (Bernstein et al., 1977; Abola, Bernstein, Bryant, Koetzle & Weng, 1987).

Taking into account the above four criteria, the structures tested have been selected from three structural types:

1. Randomly generated structures* with only one heavy atom in the unit cell and different DR: Pt-$C_{62}N_{15}O_{22}$, Pt-$C_{248}N_{63}O_{88}$, Pt-$C_{496}N_{127}O_{176}$, Pt-$C_{744}N_{191}O_{264}$;

2. Randomly generated structures with four atoms (two different heavy-atom types) in the unit cell and different DR: Hg_3Pt-$C_{59}N_{15}O_{22}$, Hg_3Pt-$C_{245}N_{63}O_{88}$, Hg_3Pt-$C_{493}N_{127}O_{178}$, Hg_3Pt-$C_{741}N_{191}O_{264}$;

3. Real protein structures: APP† and C_{550}.‡

All generated structures belong to space group $P1$. The protein APP crystallizes in $C2$ and C_{550} in

---

* The isomorphously related structures S_2 were constructed by replacing the heavy atoms Pt and Hg by F and Li, respectively.

---

* These structures have been constructed in such a way that the ratio of C, O and N atoms is comparable with that of known proteins. The resolution and the unit-cell parameters have been chosen on similar grounds.

† APP, avian pancreatic polypeptide (Blundell, Pitts, Tickle, Wood & Wu, 1981), is a small protein crystallizing with Zn^{2+} 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, $\beta = 105.30°$ 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 referred to as 1PPT.

‡ C_{550}, cytochrome c from *Paracoccos denitrificans* (Timkovich & Dickerson, 1976), is a protein with molecular weight $M_r = 14\,500$ (1017 atoms in the asymmetric unit), space group $P2_12_12_1$ and unit-cell parameters $a = 42.70$, $b = 82.17$, $c = 31.56$ Å and $Z = 4$. In addition to the anomalous scatterers Pt and Cl ($PtCl_4^{2-}$), the structure contains one Fe and six S atoms that also scatter anomalously at the wavelength used (Cu $K\alpha$). The structure was originally solved by SIRNAS to a resolution of 2.45 Å. In the test procedure, data of up to 2.5 Å resolution and Cu $K\alpha$ radiation were used. In the PDB release of July 1991, this structure is referred to as 155C.

Table 1. *Abbreviations used in Tables 2 to 11*

Procedures employed

JPDMOD	Doublet estimation using probabilistic technique (mode)
JPDNUM	Doublet estimation using probabilistic technique (numerical)
ALG	Doublet estimation using algebraic technique
PAT	Doublet estimation using difference Patterson synthesis
ZER	The doublet estimates are equal to zero
TRUE	The doublet estimates are equal to the true doublet values
hkl	Reflection with indices *hkl*
DR	Diffraction ratio (Kyriakidis, Peschar & Schenk, 1993)
E_{H1}	E values from the first data set
E_{H2}	E values from the second data set
W	Reliability factor of the distribution (Peschar & Schenk, 1991)
NTR	Number of the triplets involved in the statistics*
AER	Mean absolute doublet (triplet) error in mc [equations (22a), (23a)]
ERR	Mean doublet (triplet) error in mc [equations (22b), (23b)]

* Instead of one triplet (Cochran distribution), eight isomorphous triplets exist because of the distribution involving two isomorphous data sets. Hence, the real number of triplets involved in the statistics is eight times the NTR.

$P2_12_12_1$. In all cases, n.s.f.s have been calculated from the atomic coordinates.

In tests on the generated structures, data up to 2.3 Å resolution and Cr $K\alpha$ radiation were used; for the proteins APP and C_{550}, data up to 2.0 and 2.5 Å resolution, respectively, and Cu $K\alpha$ radiation were used. In the case of 2DW the wavelengths used were Cr $K\alpha$–Fe $K\alpha$ (Cr $K\alpha$–Cu $K\alpha$ for APP and C_{550})ᵢ

Table 1 lists abbreviations used in later tables. Tables 2, 3 and 4 show test results for a representative sample of ten reflections for the structures Pt-$C_{62}N_{15}O_{22}$ (structural type 1), Hg₃ Pt-$C_{493}N_{127}O_{176}$ (structural type 2) and APP (structural type 3), respectively. Each of the tables illustrates results for the isomorphous cases: 2DW (low DR), SAS (medium DR) and SIRAS† (high DR). The first column gives the reflection, *hkl*. Columns 2 and 3 list the magnitudes $|E_{H1}|$ and $|E_{H2}|$ from the isomorphous data sets involved. The last five columns contain the doublet values: (i) true value; (ii) estimated by JPDMOD; (iii) estimated by JPDNUM; (iv) estimated by ALG; (v) estimated by PAT.

An analysis of these tables shows that a quite good doublet estimation can be obtained for the structural type 1 by use of JPDNUM (for high DR) or JPDMOD (for medium and low DR). The estimation of the structural types 2 and 3 with these techniques is less accurate. In contrast, inspection of the tables shows that for all DR the quality of the ALG estimation of the doublets is good. Finally, the values obtained with the difference Patterson synthesis approach the true value more closely than the other techniques. In particular, Tables 3(c) and 4(c) show the enormous improvement gained if an estimate for the double summation of (14) is used. Obviously, only absolute doublet estimates are available in the SIRAS case.

† For the sake of brevity, the SIRNAS case is omitted since the results in this case are almost the same as the results of the SIRAS case.

Tables 5, 6 and 7 illustrate the overall error for the five different ways of doublet estimation as tested for the structural types 1, 2 and 3, respectively, in the four cases of isomorphous data sets. The first column indicates the relevant technique (SAS, SIRNAS, SIRAS and 2DW). The second column lists the theoretical DR (Kyriakidis, Peschar & Schenk, 1993) and columns 3 to 12 show the error in mc (1000 mc = 2π rad) of the mean absolute difference, AER,

$$\text{AER} = \langle ||\psi_2^{mn}|_{\text{true}} - |\psi_2^{mn}|_{\text{est}}| \rangle, \quad (22a)$$

and of the mean difference, ERR,

$$\text{ERR} = \langle |\psi_{2\text{true}}^{mn} - \psi_{2\text{est}}^{mn}| \rangle \quad (22b)$$

(with $\psi_{2\text{est}}^{mn} = |\psi_2^{mn}|_{\text{est}}$ because all the estimates are positive or zero), for the five different methods of doublet estimation.

The results in Table 5 illustrate the strength of ALG compared with the JPD estimation. The improved algebraic estimation PAT gives the same results as the normal algebraic estimation ALG since the two data sets differ only in one pair of non-identical anomalous scatterers so the double summation in (14) is not expected to contribute considerably. The estimation of the absolute doublet value is almost perfect for SIRNAS, SIRAS and 2DW, the only ambiguity being the sign. For SAS and 2DW, the doublets are almost always positive but for SIRAS and SIRNAS their sign cannot be predicted. Because of this, the difference between the mean absolute difference (22a) and the mean difference (22b) is very small for SAS and 2DW but quite large for SIRNAS and SIRAS.

In Tables 6 and 7, the importance of the inclusion of the double summation in (14) is shown. Compared with the ALG and JPD estimation techniques, the estimation error is reduced considerably in all cases.

Cumulative statistics of the triplet-phase-sum estimates are shown in Tables 8, 9 and 10 for the different methods of doublet estimation and the three structural types. Only the SAS results are shown because in that case the sign ambiguity is absent to a great extent. Column 1 of each group of columns shows the reliability factor W of the triplet distribution (20), column 2 gives the number of the triplets involved in the distribution and column 3 the mean absolute difference AER,

$$\text{AER} = \langle ||\psi_3|_{\text{true}} - \psi_3|_{\text{est}}| \rangle. \quad (23a)$$

The mean difference ERR is given in column 4,

$$\text{ERR} = \langle |\psi_{3\text{ true}} - \psi_{3\text{ est}}| \rangle. \quad (23b)$$

The first group of four columns refers to doublet estimation estimated by means of the JPDMOD* expression. The remaining three groups of columns

* For the SAS case, the JPDMOD and JPDNUM techniques give almost the same results, so only the first case is reported here.

Table 2. *Doublet values for ten reflections*; *structural type* 1: $Pt-C_{62}N_{15}O_{22}$

Space group $P\bar{1}$; resolution 2.3 Å.

(a) 2DW case, Cr $K\alpha$-Fe $K\alpha$ radiations, $DR = 0.03$

					Doublet values (mc)				
h	k	l	E_{H1}	E_{H2}	TRUE	JPDMOD	JPDNUM	ALG	PAT
2	3	1	2.225	2.217	2	4	4	2	2
2	−2	2	1.539	1.542	3	4	4	3	3
−1	3	1	1.423	1.430	4	4	4	4	4
6	0	1	1.321	1.332	5	5	5	5	5
−2	1	0	1.133	1.107	1	4	4	0	0
−3	4	1	0.960	0.981	6	5	5	6	6
0	0	3	0.911	0.899	3	4	4	3	3
4	3	2	0.890	0.887	7	5	5	7	7
5	−1	2	0.754	0.759	8	5	5	8	8
2	0	0	0.729	0.704	2	4	4	2	2

(b) SAS case, Cr $K\alpha$ radiation, $DR = 0.26$

					Doublet values (mc)				
h	k	l	E_{H1}	E_{H2}	TRUE	JPDMOD	JPDNUM	ALG	PAT
1	0	0	4.150	4.231	11	32	32	9	9
0	1	0	3.774	3.530	4	32	32	4	4
0	3	0	2.195	1.983	15	37	35	15	15
5	2	1	1.548	1.763	24	44	39	24	24
5	−2	2	1.221	0.960	30	45	40	30	30
2	2	2	1.047	1.057	45	38	41	44	44
−1	−2	4	0.989	1.238	28	42	40	28	28
−4	1	4	0.918	0.876	60	45	40	59	59
−2	0	3	0.854	0.789	56	38	40	56	56
2	−3	2	0.663	0.956	−8	41	40	19	19

(c) SIRAS case, Cr $K\alpha$ radiation, $DR = 0.84$

					Doublet values (mc)				
h	k	l	E_{H1}	E_{H2}	TRUE	JPDMOD	JPDNUM	ALG	PAT
2	3	1	2.225	2.559	−23	5	31	23	23
1	2	2	2.114	2.346	26	4	32	26	26
1	−2	3	1.679	2.142	65	5	39	65	65
−5	0	1	1.603	1.480	1	5	51	0	0
6	0	0	1.251	1.041	−63	5	75	63	63
0	1	0	0.893	0.886	−132	3	85	132	132
2	0	1	0.863	0.312	−47	4	165	47	47
2	−1	4	0.767	0.298	187	5	188	187	187
5	−1	2	0.754	0.249	205	5	200	205	205
0	2	4	0.754	0.998	175	5	103	175	175

illustrate the results when the doublets are estimated by the algebraic cases ALG and PAT and if the true doublets are used.

The triplet-estimation results for the structures of type 1 are listed in Table 8. It appears that the overall triplet estimation error is appreciably larger if JPDMOD is used rather than ALG or PAT. As expected, the ALG and PAT estimations lead to identical results since for these structures the double summation in (14) is negligible. Both ALG and PAT are seen to yield increasingly better results if the DR becomes larger than 0.10–0.15 and, for $DR = 0.26$, they attain almost the same error level as the true doublet values. However, if the DR becomes smaller than 0.1 [e.g. compare the $DR = 0.13$ for Table 8(c) and $DR = 0.11$ for Table 8(d)], even the most reliable triplets are estimated with an average error too large for DM applications. Apparently, the almost correct but small doublet estimates are not sufficient to determine the triplet-phase-sum values alone.

Table 9 shows the enormous error reduction to be gained if the double summation is included in (14) (ALG versus PAT). This applies especially to the triplets that are estimated to be the most reliable. The same conclusion also holds for the small protein APP and the more typically sized protein C_{550} [see Tables

Table 3. *Doublet values for ten reflections*; *structural type* 2: $Hg_3Pt\text{-}C_{493}N_{127}O_{176}$

Space group $P1$; resolution 2.3 Å.

(a) 2DW case, Cr $K\alpha$-Fe $K\alpha$ radiations, DR=0.03

h	k	l	E_{H1}	E_{H2}	Doublet values (mc)				
					TRUE	JPDMOD	JPDNUM	ALG	PAT
1	1	0	3.148	3.173	1	3	3	1	1
0	−1	1	2.635	2.610	0	3	3	0	0
−13	5	3	2.428	2.418	4	4	4	2	4
4	1	5	2.258	2.259	3	3	3	2	3
−7	8	6	2.158	2.155	5	4	4	2	5
7	5	0	2.025	2.035	3	3	3	2	3
−6	−6	5	1.921	1.900	2	4	4	1	2
−5	−5	8	1.912	1.903	5	4	4	3	5
9	−2	7	1.848	1.866	6	4	4	3	6
−7	2	4	1.740	1.742	1	3	3	3	1

(b) SAS case, Cr $K\alpha$ radiation, DR=0.23

h	k	l	E_{H1}	E_{H2}	Doublet values (mc)				
					TRUE	JPDMOD	JPDNUM	ALG	PAT
1	0	1	3.261	3.005	15	24	24	0	15
0	−1	1	2.635	2.394	0	24	25	0	0
12	1	1	2.255	2.173	39	31	28	19	38
9	−3	5	1.991	2.068	43	33	29	21	43
5	−3	2	1.935	1.984	30	27	29	20	30
1	−1	4	1.875	1.814	14	27	29	21	13
−4	12	1	1.851	1.636	46	37	30	20	46
9	5	4	1.764	1.749	23	33	30	26	23
−10	−3	3	1.724	1.889	21	31	30	19	22
11	2	3	1.716	1.396	40	32	30	0	38

(c) SIRAS case, Cr $K\alpha$ radiation, DR=0.69

h	k	l	E_{H1}	E_{H2}	Doublet values (mc)				
					TRUE	JPDMOD	JPDNUM	ALG	PAT
6	2	0	2.720	2.065	9	1	26	0	9
5	−2	2	2.443	2.415	7	1	26	10	9
2	−4	6	2.185	1.324	−34	1	41	0	34
−9	−3	4	2.115	1.684	41	2	36	0	45
3	3	3	2.056	2.120	−9	1	30	38	9
−9	−8	2	2.001	2.337	0	1	33	51	0
−5	−5	7	1.819	2.094	112	2	37	57	108
11	0	4	1.736	1.787	−93	1	40	52	102
−2	4	6	1.719	1.260	−65	2	45	0	73
11	2	3	1.716	1.425	−120	1	44	0	114

10(a) and (b)]. It appears that in APP the error level of the most reliable triplets according to the PAT technique is only slightly larger than the ideal results for the true doublet values (100 and 98 mc, respectively, for all the triplets, only 52 mc in both cases for the most reliable triplets). The error levels in the JPD and ALG cases are much higher. The general error level in C_{550} is higher than in APP but once again the error level of the most reliable triplets is comparable for the PAT and TRUE doublet cases. This leads to the conclusion that, provided the DR is large enough, the PAT technique correctly estimates not only the doublets but also the triplet phase sums (in particular, those estimated most reliably). As a result, the PAT estimation may be used in a DM procedure applicable to large structures.

Finally, the cumulative statistics in Table 11 show that, provided the correct signs of the doublets are known, low error level can be obtained in the SIRNAS case as well. The influence of the doublet estimates in the SIRNAS case is stronger than in the SAS case, however, the sign ambiguity in the former case still limits its applicability in direct methods.

Table 4. *Double values for ten reflections*; *structural type* 3: *APP*

Space group $C2$; resolution 2.0 Å.

(a) 2DW case, Cr $K\alpha$-Cu $K\alpha$ radiations, DR=0.05

					Doublet values (mc)				
h	k	l	E_{H1}	E_{H2}	TRUE	JPDMOD	JPDNUM	ALG	PAT
1	1	0	3.014	2.999	2	4	4	2	2
3	3	0	2.456	2.456	3	5	5	3	3
-7	3	5	1.926	1.898	4	6	6	5	4
2	2	4	1.840	1.807	5	5	5	4	5
6	2	7	1.719	1.708	8	7	7	6	7
7	1	9	1.634	1.625	10	8	8	7	8
-8	6	8	1.602	1.589	7	7	7	6	7
2	12	6	1.580	1.642	-1	8	8	0	0
-10	10	4	1.534	1.514	10	7	7	7	9
7	11	3	1.330	1.280	9	7	8	7	9

(b) SAS case, Cu $K\alpha$ radiation, DR=0.11

					Doublet values (mc)				
h	k	l	E_{H1}	E_{H2}	TRUE	JPDMOD	JPDNUM	ALG	PAT
1	1	0	3.014	3.025	2	10	10	6	4
1	3	0	2.214	2.176	1	11	11	8	5
-7	3	5	1.926	1.971	8	13	13	10	8
-4	12	3	1.691	1.590	10	17	13	11	10
-7	7	9	1.642	1.628	22	17	13	15	18
2	12	6	1.580	1.413	-6	18	14	0	8
-1	7	6	1.480	1.505	3	14	14	12	10
-7	7	10	1.419	1.466	12	18	14	17	16
0	6	5	1.395	1.482	0	13	14	11	7
4	12	0	1.345	1.355	16	16	14	18	17

(c) SIRÁS case, Cu $K\alpha$ radiation, DR=0.53

					Doublet values (mc)				
h	k	l	E_{H1}	E_{H2}	TRUE	JPDMOD	JPDNUM	ALG	PAT
-1	1	2	2.263	2.807	33	1	18	32	33
-3	7	9	1.685	1.551	15	1	35	34	17
10	4	6	1.614	1.707	54	1	35	54	53
12	2	2	1.586	1.308	26	1	40	0	27
1	9	1	1.414	1.670	62	1	34	60	61
-1	1	8	1.384	1.113	-39	1	42	29	40
-1	1	5	1.341	1.795	-56	1	31	53	56
4	10	7	1.340	1.924	93	1	37	65	80
2	6	10	1.170	1.420	-99	1	46	82	93
3	9	2	1.157	1.260	109	1	44	76	94

Table 5. *Overall error* (*in* mc) *for the doublet estimates for structural type* 1

Space group $P1$; resolution 2.3 Å; radiation Cr $K\alpha$; strongest 250 $|E_H|$ values used.

Case	DR	ZER AER	ERR	JPDMOD AER	ERR	JPDNUM AER	ERR	ALG AER	ERR	PAT AER	ERR
(a) Pt-$C_{62}N_{15}O_{22}$											
SAS	0.26	42	42	12	13	11	12	1	2	1	2
SIRNAS	0.87	87	87	87	87	46	97	0	66	0	66
SIRAS	0.84	87	87	83	86	42	96	0	70	0	70
2DW	0.03	4	4	1	1	0	1	0	0	0	0
(b) Pt-$C_{248}N_{63}O_{88}$											
SAS	0.17	16	16	4	6	5	7	1	2	1	2
SIRNAS	0.49	31	31	31	31	13	39	0	33	0	33
SIRAS	0.47	28	28	27	28	12	33	0	29	0	29
2DW	0.02	2	2	1	1	0	0	0	0	0	0
(c) Pt-$C_{496}N_{127}O_{176}$											
SAS	0.13	10	10	4	6	6	7	1	2	1	2
SIRNAS	0.36	16	16	16	16	10	22	0	13	0	13
SIRAS	0.34	16	16	15	16	9	24	0	18	0	18
2DW	0.01	1	1	1	1	0	0	0	0	0	0
(d) Pt-$Cr_{744}N_{191}O_{264}$											
SAS	0.11	8	8	3	3	6	6	1	2	1	2
SIRNAS	0.29	15	15	15	15	7	17	0	13	0	13
SIRAS	0.28	14	14	13	14	7	16	0	12	0	12
2DW	0.01	1	1	0	0	0	0	0	0	0	0

Table 6. *Overall error* (*in* mc) *for the doublet estimates for structural type* 2

Space group $P1$; resolution 2.3 Å; radiation Cr $K\alpha$; strongest 250 $\|E_H|$ values used.

Case	DR	ZER AER	ERR	JPDMOD AER	ERR	JPDNUM AER	ERR	ALG AER	ERR	PAT AER	ERR
(a) Hg_3Pt-$C_{59}N_{15}O_{22}$											
SAS	0.34	58	58	11	11	11	11	19	19	9	9
SIRNAS	1.41	163	163	163	163	96	198	140	214	75	178
SIRAS	1.36	163	163	155	162	92	193	137	211	72	180
2DW	0.04	4	4	1	1	1	1	2	2	0	0
(b) Hg_3Pt-$C_{245}N_{63}O_{88}$											
SAS	0.28	39	39	9	9	10	10	14	14	3	3
SIRNAS	0.96	62	62	62	62	43	85	54	68	28	72
SIRAS	0.90	61	61	59	61	42	83	52	68	27	75
2DW	0.04	4	4	1	1	1	1	1	1	0	0
(c) Hg_3Pt-$C_{493}N_{127}O_{176}$											
SAS	0.23	25	25	8	9	8	8	10	10	3	3
SIRNAS	0.73	36	36	36	36	22	50	30	42	16	45
SIRAS	0.69	33	33	32	33	21	45	28	39	14	41
2DW	0.03	3	3	1	1	1	1	1	1	0	0
(d) Hg_3Pt-$C_{741}N_{191}O_{264}$											
SAS	0.20	21	21	7	8	8	8	9	9	2	3
SIRNAS	0.61	29	29	29	29	19	39	26	35	12	35
SIRAS	0.57	28	28	27	28	17	36	25	34	12	32
2DW	0.02	2	2	1	1	0	0	1	1	0	0

Table 7. *Overall error* (*in* mc) *for the doublet estimates for structural type* 3

Radiation Cu $K\alpha$; strongest 250 $\|E_H|$ values used.

Case	DR	ZER AER	ERR	JPDMOD AER	ERR	JPDNUM AER	ERR	ALG AER	ERR	PAT AER	ERR
(a) APP; space group $C2$; resolution 2.0 Å											
SAS	0.11	13	13	4	5	5	5	5	5	3	3
SIRNAS	0.56	41	41	41	41	25	50	30	53	19	48
SIRAS	0.53	39	39	38	39	23	48	28	50	18	46
2DW	0.05	5	5	2	2	2	2	2	2	1	1
(b) C_{550}; space group $P2_12_12_1$; resolution 2.5 Å											
SAS	0.09	7	7	3	4	3	4	4	5	2	3
SIRNAS	0.33	15	15	15	15	12	21	15	23	3	15
SIRAS	0.32	15	15	13	15	10	19	14	21	2	14
2DW	0.04	2	2	1	1	1	2	1	2	1	1

Table 8. *Cumulative statistics of the triplet phase sums for different doublet estimations: structural type* 1

Space group $P1$; resolution 2.3 Å; radiation Cr $K\alpha$; strongest 250 $|E_H|$ values used; SAS case.

	JPD				ALG				PAT				TRUE		
W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
(a) Pt-$C_{62}N_{15}O_{22}$; DR=0.26															
350	93	49	72	150	72	38	40	150	72	38	40	100	123	36	38
100	1743	70	88	90	1161	38	42	90	1161	38	42	95	1104	36	39
0	3750	63	76	0	3750	50	56	0	3750	50	56	0	3750	46	51
(b) Pt-$C_{248}N_{63}O_{88}$; DR=0.17															
45	135	73	98	25	233	59	67	25	233	59	67	30	185	61	64
20	1237	89	114	15	1149	75	91	15	1149	75	91	15	1093	63	69
0	2232	92	115	0	2232	84	103	0	2232	84	103	0	2232	70	80
(c) Pt-$C_{496}N_{127}O_{176}$; DR=0.13															
25	135	109	162	15	177	88	128	15	177	88	128	20	103	60	75
15	461	104	142	10	488	100	131	10	488	100	131	10	452	79	97
0	900	105	139	0	900	100	130	0	900	100	130	0	900	84	106
(d) Pt-$C_{744}N_{191}O_{264}$; DR=0.11															
15	139	126	202	10	220	93	131	10	220	93	131	10	273	75	90
9.5	390	112	161	7.5	384	93	131	7.5	384	93	131	5.5	453	86	107
0	880	108	143	0	880	101	135	0	880	101	135	0	880	91	115

Table 9. *Cumulative statistics of the triplet phase sums for different doublet estimations: structural type* 2

Space group $P1$; resolution 2.3 Å; radiation Cr $K\alpha$; strongest 250 $|E_H|$ values used; SAS case.

	JPD				ALG				PAT				TRUE		
W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
(a) Hg$_3$Pt-$C_{59}N_{15}O_{22}$; DR=0.34															
2000	219	15	17	850	235	30	35	2000	149	15	16	2500	123	14	15
1000	1281	29	34	650	770	34	40	1000	1041	23	25	1000	1313	27	31
0	3750	48	57	0	3750	49	60	0	3750	48	58	0	3750	48	57
(b) Hg$_3$Pt-$C_{245}N_{63}O_{88}$; DR=0.28															
250	103	37	43	90	162	34	38	250	108	26	28	150	108	26	28
150	752	45	52	70	639	41	48	150	671	33	36	150	682	33	36
0	2555	56	65	0	2555	51	60	0	2555	53	61	0	2555	52	60
(c) Hg$_3$Pt-$C_{493}N_{127}O_{176}$; DR=0.23															
85	178	78	97	35	179	65	80	80	138	41	46	80	173	41	46
55	560	76	91	25	525	70	84	25	578	62	72	30	549	57	66
0	1024	74	91	0	1024	69	85	0	1024	70	85	0	1024	68	83
(d) Hg$_3$Pt-$C_{741}N_{191}O_{264}$; DR=0.20															
45	156	86	120	20	213	77	105	35	166	44	51	40	148	42	49
20	560	91	117	8.5	561	83	107	8	560	79	99	8	557	77	96
0	809	91	116	0	809	84	108	0	809	87	111	0	809	87	110

Table 10. *Cumulative statistics of the triplet phase sums for different doublet estimations; structural type* 3

Cu $K\alpha$ radiation; SAS case; strongest 250 $|E_H|$ values used.

	JPD				ALG				PAT				TRUE		
W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR	W	NTR	AER	ERR
(a) APP; space group $C2$; resolution 2.0 Å; DR=0.11															
30	198	71	95	20	185	71	97	20	515	48	52	30	255	48	52
20	1206	83	103	15	894	71	92	15	1225	55	64	20	1021	51	57
0	3750	90	111	0	3750	77	97	0	3750	80	100	0	3750	79	98
(b) C_{550}; space group $P2_12_12_1$; resolution 2.5 Å; DR=0.9															
50	106	135	162	40	136	183	204	25	121	96	103	65	124	83	86
15	1094	146	191	15	776	140	192	8	876	125	158	15	835	103	122
0	3750	146	200	0	3750	138	200	0	3750	137	199	0	3750	131	191

Table 11. *Cumulative statistics of the triplet phase sums for different doublet estimations: structural type* 1

SIRNAS case (true sign for the doublets); structure Pt-$C_{744}N_{191}O_{264}$; space group $P1$; resolution 2.3 Å; radiation Cr $K\alpha$; strongest 250 $|E_H|$ values used; DR=0.29.

	JPD				PAT		
W	NTR	AER	ERR	W	NTR	AER	ERR
1.0	103	88	88	2.0	135	66	70
0.5	413	126	126	1.0	520	71	80
0.0	860	169	169	0.0	864	88	104

References

ABOLA, E. E., BERNSTEIN, F. C., BRYANT, S. H., KOETZLE, T. F. & WENG, J. (1987). *Crystallographic Databases - Information Content, Software Systems, Scientific Applications*, edited by F. H. ALLEN, G. BERGERHOFF & S. SIEVERS, pp. 107-132. Bonn, Cambridge, Chester: Data Commission of the International Union of Crystallography.
BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.
BLUNDELL, T. L., PITTS, J. E., TICKLE, I. J., WOOD, S. P. & WU, C. W. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4175-4179.
CASCARANO, G. & GIACOVAZZO, C. (1984). *Acta Cryst.* **A40**, 305-306.

FAN, H.-F. & GU, Y.-X. (1985). *Acta Cryst.* **A41**, 280-284.
FAN, H.-F., HAN, F.-S., QIAN, J.-Z. & YAO, J.-X. (1984). *Acta Cryst.* **A40**, 489-495.
FORTIER, S., FRASER, M. E. & MOORE, N. J. (1986). *Acta Cryst.* **A42**, 149-156.
FORTIER, S. & NIGAM, G. D. (1989). *Acta Cryst.* **A45**, 247-254.
FUREY, W. JR, CHANDRASEKHAR, K., DYDA, F. & SAX, M. (1990). *Acta Cryst.* **A46**, 560-567.
GIACOVAZZO, C. (1977). *Acta Cryst.* **A33**, 933-944.
GIACOVAZZO, C. (1983). *Acta Cryst.* **A39**, 585-592.
GIACOVAZZO, C., CASCARANO, G. & ZHENG, C.-D. (1988). *Acta Cryst.* **A44**, 45-51.
GUO, D. Y. (1990). *Acta Cryst.* **A46**, 942-944
GUO, D. Y., BLESSING, R. H. & HAUPTMAN, H. (1991). *Acta Cryst.* **A47**, 340-345.
HAO, Q. & FAN, H.-F. (1988). *Acta Cryst.* **A44**, 379-382.
HAUPTMAN, H. (1975). *Acta Cryst.* **A31**, 671-679.
HAUPTMAN, H. (1982a). *Acta Cryst.* **A38**, 289-294.
HAUPTMAN, H. (1982b). *Acta Cryst.* **A38**, 632-641.
KROON, J., SPEK, A. L. & KRABBENDAM, H. (1977). *Acta Cryst.* **A33**, 382-385.
KYRIAKIDIS, C. E., PESCHAR, R. & SCHENK, H. (1993). *Acta Cryst.* **A49**, 350-358.
PESCHAR, R. (1987). Thesis, Univ of Amsterdam, The Netherlands.
PESCHAR, R. & SCHENK, H. (1991). *Acta Cryst.* **A47**, 428-440.
ROSSMANN, M. G. (1960). *Acta Cryst.* **13**, 221-226.
SCHENK, H. (1973). *Acta Cryst.* **A29**, 77-82.
SCHENK, H. (1974). *Acta Cryst.* **A30**, 477-482.
TIMKOVICH, R. & DICKERSON, R. E. (1976). *J. Biol. Chem.* **251**, 4033-4046.

SHORT COMMUNICATIONS

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 *words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.*

The asymmetric unit of X-ray intensity data of the seven crystal systems.

By IL-HWAN SUH, KWANG-JU KIM, GEUM-HONG CHOO and JIN-HO LEE, *Department of Physics, Chungnam National University, Daejeon 305-764, Korea,* SUNG HO CHOH, *Department of Physics, Korea University, Seoul 136-702, Korea,* and MOON-JIB KIM, *Department of Physics, Soonchunhyang University, Onyang 336-600, Korea*

Abstract

A crystal structure can be determined from the X-ray intensity data of one asymmetric unit. As the function $I(hkl)$ of the X-ray intensity has a center of symmetry if it is assumed that anomalous scattering is negligible, $I(hkl)$ has the symmetry of a centrosymmetric point group, *i.e.* a Laue group. The asymmetric units of the intensity data are derived here for all Laue groups.

1. Introduction

The Laue group, together with a corresponding asymmetric unit of X-ray intensity data, must be known to determine a crystal structure. The asymmetric units of intensity data reported so far are incomplete and even contain errors

(Sakurai, 1986; Stout & Jensen, 1989). In the present paper, an explanation of how to derive the equivalent intensities for each of the 11 Laue groups is given. The exact extents of the 11 asymmetric units of intensity data are shown.

2. Theory

There is only one Laue group in each of the triclinic, monoclinic and orthorhombic systems but two Laue groups in each of the tetragonal, trigonal, hexagonal and cubic systems. Thus there are 11 Laue groups altogether.

In the trigonal system there exist two Laue groups, $\bar{3}$ and $\bar{3}m$. Both are compatible with a rhombohedral lattice as well as with a hexagonal lattice. Therefore, in Table 1, the point groups $\bar{3}$ and $\bar{3}m$ are described with rhombohedral