

to compensate accurately for this spurious component, which was small in any case. A second source of error could be due to the assumption of an abrupt discontinuous step in the dielectric susceptibilities χ_0 , χ_{220} at $z=0$ (Batterman & Cole, 1964). A more accurate approximation for the fields near the surface would take into account the electron density for an actual crystal surface and perhaps even the presence of the chemisorbed species and surface roughness as well.

Although the sample was prepared chemically and thereafter kept in an inert He atmosphere, the coherent fraction was observed to degrade as a function of time.

In summary, we have obtained expressions that parametrize glancing-angle X-ray standing-wave fluorescence yields in the same manner as X-ray standing waves created in the Bragg geometry. The expressions permit fitting the fluorescent intensity from atoms above or below the surface of a crystal to obtain the coherent position of the atoms parallel to the surface and their coherent fraction. The fluorescence yield includes a dependence on the distance of the atoms from the surface because of a changing extinction of the wave fields above and below the surface.

The authors would like to acknowledge the assistance of the CHSS technical staff. CHSS is oper-

ated under grant DMR-90-21700 of the National Science Foundation. The National Institute of Standards and Technology is a part of the Technology Administration of the US Department of Commerce.

References

- AFANAS'EV, A. M., IMAMOV, R. M., MASLOV, A. V. & PASHAEV, E. M. (1984). *Phys. Status Solidi A*, **84**, 73-78.
- AFANAS'EV, A. M. & MELKONYAN, M. K. (1983). *Acta Cryst.* **A39**, 207-210.
- BATTERMAN, B. W. (1964). *Phys. Rev. A*, **133**, 759-764.
- BATTERMAN, B. W. (1969). *Phys. Rev. Lett.* **22**, 703-705.
- BATTERMAN, B. W. & COLE, H. (1964). *Rev. Mod. Phys.* **36**, 681-717.
- BEDZYK, M. J., SHEN, Q., O'KEEFFE, M. E., NAVROTSKI, G. & BERMAN, L. E. (1989). *Surf. Sci.* **220**, 419-426.
- COWAN, P. L. (1985). *Phys. Rev. B*, **32**, 5437-5439.
- COWAN, P. L., BRENNAN, S., JACH, T., BEDZYK, M. J. & MATERLIK, G. (1986). *Phys. Rev. Lett.* **57**, 2399-2402.
- EISENBERGER, P. & MARRA, W. C. (1981). *Phys. Rev. Lett.* **46**, 1081-1084.
- GOLOVCHENKO, J. A., BATTERMAN, B. W. & BROWN, W. L. (1974). *Phys. Rev. B*, **10**, 4239-4243.
- HASHIZUME, H. & SAKATA, O. (1989). *Rev. Sci. Instrum.* **60**, 2373-2375.
- HERTEL, N., MATERLIK, G. & ZEGENHAGEN, J. (1985). *Z. Phys.* **B58**, 199-204.
- JACH, T. & BEDZYK, M. J. (1990). *Phys. Rev. B*, **42**, 5399-5402.
- JACH, T., COWAN, P. L., SHEN, Q. & BEDZYK, M. J. (1989). *Phys. Rev. B*, **39**, 5739-5747.
- MARRA, W. C., EISENBERGER, P. & CHO, A. Y. (1979). *J. Appl. Phys.* **50**, 6927-6933.
- MATERLIK, G., FRAHM, A. & BEDZYK, M. J. (1984). *Phys. Rev. Lett.* **52**, 441-444.

Acta Cryst. (1993). **A49**, 350-358

On the Relation Between the Diffraction Ratio, the Doublet Values and the Reliability of the Triplet Estimates

BY CHRISTOS E. KYRIAKIDIS, RENÉ PESCHAR AND HENK SCHENK

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

(Received 29 April 1992; accepted 23 September 1992)

Abstract

A diffraction ratio is proposed that predicts the differences to be expected between the intensities of two-isomorphous data sets. This ratio is important for the *ab initio* structure determination of isomorphously related structures by means of direct methods. The diffraction ratio is shown to be linearly related to the average doublet phase sum of the isomorphous data. It is argued that the doublets are essential for correct triplet-phase-sum estimates. The

diffraction ratio and the idealized average triplet-phase-sum error, as calculated from a recent probabilistic theory, turn out to be related. A minimum diffraction ratio is required to obtain a triplet-phase-sum-error level comparable with that of small structures that are solved routinely by conventional direct methods. The diffraction ratio can be used to maximize the triplet-phase-sum reliability before collecting the data by choosing the optimal wavelength in a single anomalous-scattering experiment, selecting the most suitable heavy-atom derivative in a single-

isomorphous-replacement experiment or selecting the optimal wavelengths in a multiwavelength experiment.

1. Introduction

The crystal structure determination of large molecules such as proteins relies in general on the isomorphous-replacement (IR) technique (see, for example, Blundell & Johnson, 1976) combined with anomalous-scattering data (see for example, Ramaseshan & Abrahams, 1975). More recently, the multi-wavelength technique has also become popular (Hendrickson, 1988; Karle, 1989). An analysis of the above techniques from a probabilistic point of view leads readily to the definition of isomorphous data sets (Fortier & Nigam 1989). Recently, the full probabilistic integration of direct methods (DM) with any type and number of isomorphous data sets has been accomplished (Peschar & Schenk, 1991). In this latter study it is shown that various protein-solving techniques like SAS (single anomalous scattering) SIR(N)AS (single isomorphous replacement with or without anomalous scattering) and 2DW (two different wavelengths) all rely on isomorphous data that can be combined in a single probabilistic expression. Although tests suggest that DM may be applicable in solving protein structures *ab initio* (Furey, Chandrasekhar, Dyda & Sax, 1990), a standard DM structure determination has not yet proved to be possible. The success of the above-mentioned techniques relies on the determination of an initial small structural model, *i.e.* the heavy-atom substructure. In this respect it is interesting to note that DM are particularly successful in determining relatively small structures. An important prerequisite for a successful DM run is to have small estimation errors for the most reliably estimated triplet phase sums. Therefore, an (experimental) indicator to predict these errors will be useful, in particular if the wavelength(s) can be selected more or less freely as in a synchrotron environment or if a variety of heavy-atom derivatives may be prepared. In this way, it would be possible to choose beforehand the isomorphous data combination with the best chance of success. The knowledge of the difference in intensity due to various effects between isomorphous data sets plays a prominent role in the structure determination of large structures. The average change in intensity due to adding a heavy atom to a native protein was first established by Crick & Magdoff (1956) and has been employed more recently by Fortier, Weeks & Hauptman (1984) in the form of a so-called diffraction ratio (DR)

$$DR = (2\sum Z_H^2 / \sum Z_P^2)^{1/2} \quad (1)$$

where

$$\sum Z_H^2 = \sum_{k \in H} Z_k^2 \quad (2a)$$

$$\sum Z_P^2 = \sum_{j \in P} Z_j^2, \quad (2b)$$

with the assumption that the atomic content of the derivative equals the atomic content of the native protein (P) plus the heavy-atom content (H).

Equation (1) does predict differences in intensity but is valid only for isomorphous-replacement data. On the other hand, the so-called Bijvoet ratio B_r (Zachariasen, 1965), employed if anomalous-scattering effects are involved, defines the average change in intensity but lacks a predictive quality,

$$B_r = (I_H - I_{-H}) / \frac{1}{2}(I_H + I_{-H}). \quad (3)$$

Equations (1) and (3) are not suitable for our purpose since they are not expressed in normalized structure factors (n.s.f.s) and they are not generally applicable to all possible isomorphous data. Hence, the first purpose of this paper will be the development of a general intensity-difference indicator, a diffraction ratio, for two isomorphous data sets. In this paper, only the role of the intensity differences will be investigated; the quality of the isomorphism, though expected to be important (Srinivasan & Parthasarathy, 1976; Hauptman, 1982), will not be discussed. It will be shown that the DR is related to the average doublet phase sums and also, *via* the doublets, to the average triplet phase sums. Finally, some practical rules will be formulated, *e.g.* the optimal wavelength choice in a 2DW experiment, for potential DM applications.

2. A diffraction ratio for two isomorphous data sets

In DM all phase-sum estimations are calculated *via* the knowledge of (isomorphous) n.s.f.s. The latter can be defined in a general way by allowing the atomic scattering factor f_j to be complex valued, hence including possible anomalous scattering.

$$\begin{aligned} f_j &= f_j^o + f_j^r + if_j'' \\ &= f_j^r + if_j'' \\ &= |f_j| \exp(i\delta_j). \end{aligned} \quad (4)$$

To handle data from IR and anomalous-scattering experiments in the same way, the concept of isomorphous structure factors is used (Fortier & Nigam, 1989; Peschar & Schenk, 1991). Accordingly, two (normalized) structure factors E_{H1} and E_{H2} are considered to be isomorphously related provided their structure-factor expression is identical except for the atomic scattering factors.

$$\begin{aligned} E_{Hp} &= |E_{Hp}| \exp(i\varphi_{Hp}) \\ &= (z_{pp})^{-1/2} \sum_{j=1}^n \xi_{jH} f_{jp} \end{aligned} \quad (5)$$

with

$$\begin{aligned} z_{pq} &= |z_{pq}| \exp(i\Delta_{pq}) \\ &= \sum_{j=1}^N |f_{jp}| |f_{jq}| \exp[-i(\delta_{jp} + s_{pq}\delta_{jq})], \end{aligned} \quad (6)$$

where

$$s_{pq} = \begin{cases} -1 & \text{if } \mathbf{H}_p = \mathbf{H}_q \\ 1 & \text{if } \mathbf{H}_p = -\mathbf{H}_q \end{cases} \quad (7)$$

[\$\mathbf{H}_p\$ and \$\mathbf{H}_q\$ are the reciprocal-lattice vectors for the data sets \$p\$ (\$=1, 2\$) and \$q\$ (\$=1, 2\$), respectively];

$$\xi_{jH} = m_H \sum_{s=1}^{\tau_H} \exp[2\pi i(R_s^* \mathbf{H} \cdot \mathbf{r}_j + \mathbf{H} \cdot \mathbf{T}_s)], \quad (8)$$

where \$R_s\$ is the rotational matrix of symmetry operations, \$\mathbf{T}_s\$ is the translational vector of symmetry operations, \$m_H \ge 1\$ (multiplicity) and \$\tau_H\$ is the number of symmetry operations for which the trigonometric part is different;

$$n = N/m \quad (9)$$

$$m = \tau_H m_H \quad (10)$$

where \$N\$ is the number of atoms in the unit cell and \$m\$ is the total number of symmetry operations.

In space group \$P1\$ the normalized structure-factor expression becomes

$$E_{Hk} = \sum_{j=1}^N g_{jk} \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_j) \quad \text{for } k=1, 2 \quad (11)$$

with \$g_{jk} = (z_{kk})^{-1/2} f_{jk}\$, the normalized atomic scattering factors.

In this way, the usual isomorphous-replacement data are handled and it can also simply be shown that \$E_H\$ and \$E_H^*\$ are isomorphously related. The following relations between isomorphous n.s.f.s \$E_{H1}\$ and \$E_{H2}\$ are useful:

$$E_{H1} - E_{H2} = \sum_{j=1}^N (g_{j1} - g_{j2}) \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_j); \quad (12a)$$

$$E_{H1} + E_{H2} = \sum_{j=1}^N (g_{j1} + g_{j2}) \exp(2\pi i \mathbf{H} \cdot \mathbf{r}_j). \quad (12b)$$

In analogy with the Bijvoet ratio (3), the following diffraction ratio is introduced using (12):

$$\text{DR(ideal)} = \left[4 \frac{\langle (E_{H1} - E_{H2})(E_{H1} - E_{H2})^* \rangle}{\langle (E_{H1} + E_{H2})(E_{H1} + E_{H2})^* \rangle} \right]^{1/2}. \quad (13)$$

In a more explicit notation (13) becomes

$$\begin{aligned} \text{DR(ideal)} &= \left(4 \frac{\langle |E_{H1}|^2 + |E_{H2}|^2 - 2|E_{H1}||E_{H2}| \cos \psi_2^{12} \rangle}{\langle |E_{H1}|^2 + |E_{H2}|^2 + 2|E_{H1}||E_{H2}| \cos \psi_2^{12} \rangle} \right)^{1/2}, \end{aligned} \quad (14)$$

which shows that the DR is related to the doublet phase sum

$$\psi_2^{12} = \varphi_{H1} + s_{12} \varphi_{H2} \quad (15)$$

with \$\varphi_{H1}\$ and \$\varphi_{H2}\$ the phases of two isomorphously related n.s.f.s \$E_{H1}\$ and \$E_{H2}\$, respectively [\$s_{12}\$ is given by (7)].

An alternative expression to (14) is arrived at when the dependence on the atomic coordinates is introduced in (13), with use of (11),

$$\begin{aligned} \text{DR(ideal)} &= \left(4 \left\{ \sum_{j=1}^N |g_{j1} - g_{j2}|^2 + 2 \sum_{j<i}^N \sum_{j<i}^N (g_{j1} - g_{j2}) \right. \right. \\ &\quad \times (g_{i1}^* - g_{i2}^*) \cos [2\pi \mathbf{H} \cdot (\mathbf{r}_j - \mathbf{r}_i)] \left. \left. \right\} \right. \\ &\quad \times \left. \left\{ \sum_{j=1}^N |g_{j1} + g_{j2}|^2 + 2 \sum_{j<i}^N \sum_{j<i}^N (g_{j1} + g_{j2}) \right. \right. \\ &\quad \times (g_{i1}^* + g_{i2}^*) \cos [2\pi \mathbf{H} \cdot (\mathbf{r}_j - \mathbf{r}_i)] \left. \left. \right\}^{-1} \right)^{1/2}. \end{aligned} \quad (16)$$

The theoretical expressions (13), (14) and (16) are labelled ideal since neither the doublets nor the inter-atomic vectors are completely known without prior knowledge of the atomic coordinates. Equation (14), calculated from the atomic coordinates, will be used as a reference. An approximation to the theoretical DR expression is readily obtained by neglect of the double summations in (16),

$$\text{DR2} = \left(4 \sum_{j=1}^N |g_{j1} - g_{j2}|^2 / \sum_{j=1}^N |g_{j1} + g_{j2}|^2 \right)^{1/2}. \quad (17)$$

Obviously, (17) can predict intensity differences before collecting the data. From the work of Peschar & Schenk (1991) it follows that the Friedel-related data sets \$\{\mathbf{H}\}\$ and \$\{-\mathbf{H}\}\$ should be considered to be separate data sets that are isomorphously related. In the following we will distinguish four cases of two isomorphously related data sets:

(i) *Single isomorphous replacement including anomalous-scattering effects (SIRAS)*. This is the most general case. The isomorphous data sets are: \$\{\mathbf{H}(S_1)\}\$ and \$\{\mathbf{H}(S_2)\}\$, with \$S_1\$ the heavy-atom(s) 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 those given in the literature (see, for example, Blundell & Johnson, 1976). However, from our definition of isomorphism it follows that the usual definition of SIRAS (a complete sphere of data for two anomalously scattering isomorphously related structures \$S_1\$ and \$S_2\$) leads to four isomorphously related data sets.

The complex normalized scattering factors

$$g_{j1} = g_{j1}^r + ig_{j1}^i,$$

$$g_{j2} = g_{j2}^r + ig_{j2}^i,$$

when substituted in (17) lead to

$$\text{DR2} = \left(\left\{ 4 \sum_{j=1}^N [(g_{j1}^r - g_{j2}^r)^2 + (g_{j1}^i - g_{j2}^i)^2] \right\} \times \left\{ \sum_{j=1}^N [(g_{j1}^r + g_{j2}^r)^2 + (g_{j1}^i + g_{j2}^i)^2] \right\}^{-1} \right)^{1/2}. \quad (18)$$

(ii) *Single isomorphous replacement neglecting anomalous-scattering effects (SIRNAS)*. The isomorphous data sets are $\{\mathbf{H}(S_1)\}$ and $\{\mathbf{H}(S_2)\}$, with S_1 the heavy-atom(s) derivative and S_2 the native protein for which the atomic scattering factors are real valued. In this case,

$$g_{j1} = g_{j1}^o$$

$$g_{j2} = g_{j2}^o,$$

so (17) reduces to

$$\text{DR2} = \left[4 \sum_{j=1}^N (g_{j1}^o - g_{j2}^o)^2 / \sum_{j=1}^N (g_{j1}^o + g_{j2}^o)^2 \right]^{1/2}. \quad (19)$$

(iii) *Single-wavelength anomalous scattering (SAS)*. In this case, the isomorphous data sets are $\{\mathbf{H}(S_1)\}$ and $\{-\mathbf{H}(S_1)\}$, so

$$g_{j2} = g_{j1}^*.$$

Hence, (17) reduces to

$$\text{DR2} = \left[4 \sum_{j=1}^N (g_{j1}^i)^2 / \sum_{j=1}^N (g_{j1}^o + g_{j1}^i)^2 \right]^{1/2}. \quad (20)$$

(iv) *Two different wavelengths (2DW)*. The isomorphous data sets are: $\{\mathbf{H}(\lambda_1)\}$ and $\{\mathbf{H}(\lambda_2)\}$. The same structure is measured for two different wavelengths (λ_1 and λ_2) using half of the Ewald sphere (only \mathbf{H} or $-\mathbf{H}$). In this case,

$$g_{j1} \neq g_{j2}$$

but

$$g_{j1}^o = g_{j2}^o,$$

hence, (17) becomes

$$\text{DR2} = \left(\left\{ 4 \sum_{j=1}^N [(g_{j1}^r - g_{j2}^r)^2 + (g_{j1}^i - g_{j2}^i)^2] \right\} \times \left\{ \sum_{j=1}^N [(g_{j1}^r + g_{j2}^r)^2 + (g_{j1}^i + g_{j2}^i)^2] \right\}^{-1} \right)^{1/2}. \quad (21)$$

In (18)–(21), the normalized scattering factors g_{j1} and g_{j2} may be replaced by the usual scattering factors f_{j1} and f_{j2} , respectively, since the normalization constants z_{11} and z_{22} are present in both numerator and denominator and can be omitted without significantly changing the results (as is obvious from § 3.1 and Tables 1–4). A second simplification, replacing the

θ -dependent f_j^o by Z_j , turns out to be useful. With this taken into account, (17) becomes

DR3

$$= \left(\left\{ 4 \sum_{j=1}^N [(Z_{j1} + f_{j1}^r + if_{j1}^i) - (Z_{j2} + f_{j2}^r + if_{j2}^i)]^2 \right\} \times \left\{ \sum_{j=1}^N [(Z_{j1} + f_{j1}^r + if_{j1}^i) + (Z_{j2} + f_{j2}^r + if_{j2}^i)]^2 \right\}^{-1} \right)^{1/2}, \quad (22)$$

where Z_{j1} and Z_{j2} are the atomic numbers of the j th atom from the first and the second data sets, respectively.

3. Test results and discussion

The formulae (18)–(22) have been tested extensively using a variety of randomly generated structures as well as real protein data from the Protein Data Bank (PDB) at Brookhaven National Laboratory (Bernstein *et al.*, 1977; Abola, Bernstein, Bryant, Koetzle & Weng, 1987). The predictive quality of (18)–(22) will be established first by comparing the data with the ideal diffraction ratio (14) (see § 3.1). The relation between the diffraction ratio and the doublet phase sum will be discussed in § 3.2. It will be shown in § 3.3 that the diffraction ratio is also related to the average error of the triplet phase sum.

3.1. Evaluation of the diffraction ratio

In Tables 1–4 some representative diffraction ratio results are listed for the four isomorphous data cases SIRAS, SIRNAS, SAS and 2DW for a variety of structures. The generated structures Pt-C₆₂N₁₅O₂₂, Pt-C₁₂₄N₃₁O₄₄, Pt-C₁₈₆N₄₇O₆₆, Pt-C₂₄₈N₆₃O₈₈, Pt-C₄₉₆N₁₂₇O₁₇₆ and Pt-C₇₄₄N₁₉₁O₂₆₄ belong to space group $P1$ whereas the proteins APP and C₅₅₀ crystallize in $C2$ and $P2_12_12_1$, respectively. The first six listed structures were 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 were chosen on similar grounds. APP, avian pancreatic polypeptide (Blundell, Pitts, Tickle, Wood & Wu, 1981), is a small protein crystallizing with Zn²⁺ in space group $C2$ with one molecule of 36 amino acid residues (302 atoms) in the asymmetric unit 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. In the test procedure, data up to 2.0 Å resolution and Cu $K\alpha$ radiation were used. In the PDB release of July 1991, this structure is referenced as 1PPT.

C₅₅₀, cytochrome *c* from *Paracoccus 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

Table 1. Comparison of the ideal diffraction ratio $DR(\text{ideal})$ (14) with the theoretical approximations $DR2$ (18) and $DR3$ (22).

SIRAS case; data sets: $\{\mathbf{H}(S_1)\}$ and $\{\mathbf{H}(S_2)\}$; strongest 250 $|E_H|$ values used. Randomly generated structures: space group $P1$, Cr $K\alpha$ radiation, resolution 2.3 Å.
APP: space group $C2$, resolution 2.0 Å, radiation Cu $K\alpha$.
 C_{550} : space group $P2_12_12_1$, resolution 2.5 Å, radiation Cu $K\alpha$.

| S1 | S2 | DR (ideal) | DR2 | DR3 |
|---|--|---------------|------|------|
| Pt-C ₆₂ N ₁₅ O ₂₂ | F-C ₆₂ N ₁₅ O ₂₂ | 0.80 | 0.82 | 0.84 |
| Pt-C ₁₂₄ N ₃₁ O ₄₄ | F-C ₁₂₄ N ₃₁ O ₄₄ | 0.64 | 0.67 | 0.64 |
| Pt-C ₁₈₆ N ₄₇ O ₆₆ | F-C ₁₈₆ N ₄₇ O ₆₆ | 0.55 | 0.58 | 0.53 |
| Pt-C ₂₄₈ N ₆₃ O ₈₈ | F-C ₂₄₈ N ₆₃ O ₈₈ | 0.49 | 0.52 | 0.47 |
| Pt-C ₄₉₆ N ₁₂₇ O ₁₇₆ | F-C ₄₉₆ N ₁₂₇ O ₁₇₆ | 0.37 | 0.38 | 0.34 |
| Pt-C ₇₄₄ N ₁₉₁ O ₂₆₄ | F-C ₇₄₄ N ₁₉₁ O ₂₆₄ | 0.29 | 0.32 | 0.28 |
| APP (derivative) | APP (native) | 0.54 | 0.60 | 0.53 |
| C_{550} (derivative) | C_{550} (native) | 0.32 | 0.34 | 0.31 |

Table 2. Comparison of the ideal diffraction ratio $DR(\text{ideal})$ (14) with the theoretical approximations $DR2$ (19) and $DR3$ (22).

SIRNAS case; data sets: $\{\mathbf{H}(S_1)\}$ and $\{\mathbf{H}(S_2)\}$; strongest 250 $|E_H|$ values used. Randomly generated structures: space group $P1$, Cr $K\alpha$ radiation, resolution 2.3 Å.
APP: space group $C2$, resolution 2.0 Å, radiation Cu $K\alpha$.
 C_{550} : space group $P2_12_12_1$, resolution 2.5 Å, radiation Cu $K\alpha$.

| S1 | S2 | DR (ideal) | DR2 | DR3 |
|---|--|---------------|------|------|
| Pt-C ₆₂ N ₁₅ O ₂₂ | F-C ₆₂ N ₁₅ O ₂₂ | 0.82 | 0.85 | 0.87 |
| Pt-C ₁₂₄ N ₃₁ O ₄₄ | F-C ₁₂₄ N ₃₁ O ₄₄ | 0.67 | 0.70 | 0.67 |
| Pt-C ₁₈₆ N ₄₇ O ₆₆ | F-C ₁₈₆ N ₄₇ O ₆₆ | 0.58 | 0.61 | 0.56 |
| Pt-C ₂₄₈ N ₆₃ O ₈₈ | F-C ₂₄₈ N ₆₃ O ₈₈ | 0.51 | 0.54 | 0.49 |
| Pt-C ₄₉₆ N ₁₂₇ O ₁₇₆ | F-C ₄₉₆ N ₁₂₇ O ₁₇₆ | 0.39 | 0.41 | 0.36 |
| Pt-C ₇₄₄ N ₁₉₁ O ₂₆₄ | F-C ₇₄₄ N ₁₉₁ O ₂₆₄ | 0.31 | 0.34 | 0.29 |
| APP (derivative) | APP (native) | 0.57 | 0.62 | 0.56 |
| C_{550} (derivative) | C_{550} (native) | 0.33 | 0.37 | 0.34 |

Table 3. Comparison of the ideal diffraction ratio $DR(\text{ideal})$ (14) with the theoretical approximations $DR2$ (20) and $DR3$ (22).

SAS case; data sets: $\{\mathbf{H}(S_1)\}$ and $\{-\mathbf{H}(S_1)\}$; strongest 250 $|E_H|$ values used. Randomly generated structures: space group $P1$, Cr $K\alpha$ radiation, resolution 2.3 Å.
APP: space group $C2$, resolution 2.0 Å, radiation Cu $K\alpha$.
 C_{550} : space group $P2_12_12_1$, resolution 2.5 Å, radiation Cu $K\alpha$.

| S1 | DR (ideal) | DR2 | DR3 |
|---|---------------|------|------|
| Pt-C ₆₂ N ₁₅ O ₂₂ | 0.29 | 0.31 | 0.26 |
| Pt-C ₁₂₄ N ₃₁ O ₄₄ | 0.25 | 0.26 | 0.21 |
| Pt-C ₁₈₆ N ₄₇ O ₆₆ | 0.22 | 0.23 | 0.19 |
| Pt-C ₂₄₈ N ₆₃ O ₈₈ | 0.20 | 0.21 | 0.17 |
| Pt-C ₄₉₆ N ₁₂₇ O ₁₇₆ | 0.11 | 0.16 | 0.13 |
| Pt-C ₇₄₄ N ₁₉₁ O ₂₆₄ | 0.09 | 0.13 | 0.10 |
| APP (derivative) | 0.10 | 0.14 | 0.11 |
| C_{550} (derivative) | 0.08 | 0.09 | 0.08 |

$Z = 4$. In addition to the anomalous scatterers Pt and Cl (PtCl₄⁻), the structure contains one Fe and six S atoms that also scatter anomalously at the wavelength used (Cu $K\alpha$). The structure was solved originally by SIRNAS to a resolution of 2.45 Å. In the test procedure, data up to 2.5 Å resolution and Cu $K\alpha$ radi-

Table 4. Comparison of the ideal diffraction ratio $DR(\text{ideal})$ (14) with the theoretical approximations $DR2$ (21) and $DR3$ (22).

2DW case; data sets: $\{\mathbf{H}(\lambda_1)\}$ and $\{\mathbf{H}(\lambda_2)\}$, strongest 250 $|E_H|$ values used.
Cf-C₆₂N₁₅O₂₂: space group $P1$, resolution 2.3 Å.

| λ_1 | λ_2 | DR (ideal) | DR2 | DR3 |
|--------------|--------------|---------------|------|------|
| Cr $K\alpha$ | Fe $K\alpha$ | 0.06 | 0.07 | 0.06 |
| Cr $K\alpha$ | Co $K\alpha$ | 0.08 | 0.10 | 0.09 |
| Cr $K\alpha$ | Cu $K\alpha$ | 0.14 | 0.15 | 0.13 |
| Cr $K\alpha$ | Mo $K\alpha$ | 0.23 | 0.26 | 0.22 |
| Cr $K\alpha$ | Ag $K\alpha$ | 0.20 | 0.23 | 0.19 |
| Fe $K\alpha$ | Co $K\alpha$ | 0.03 | 0.03 | 0.03 |
| Fe $K\alpha$ | Cu $K\alpha$ | 0.07 | 0.08 | 0.07 |
| Fe $K\alpha$ | Mo $K\alpha$ | 0.17 | 0.19 | 0.16 |
| Fe $K\alpha$ | Ag $K\alpha$ | 0.14 | 0.16 | 0.13 |
| Co $K\alpha$ | Cu $K\alpha$ | 0.04 | 0.05 | 0.04 |
| Co $K\alpha$ | Mo $K\alpha$ | 0.15 | 0.17 | 0.14 |
| Co $K\alpha$ | Ag $K\alpha$ | 0.12 | 0.14 | 0.11 |
| Co $K\alpha$ | Mo $K\alpha$ | 0.11 | 0.12 | 0.10 |
| Cu $K\alpha$ | Ag $K\alpha$ | 0.08 | 0.09 | 0.08 |
| Mo $K\alpha$ | Ag $K\alpha$ | 0.03 | 0.03 | 0.03 |

ation were used. In the PDB release of July 1991, this structure is referenced as 155C.

In all cases, n.s.f.s were calculated from the atomic coordinates. Table 1 shows test results for the SIRAS case using as isomorphous data sets $\{\mathbf{H}(S_1)\}$ and $\{\mathbf{H}(S_2)\}$ with S_1 and S_2 the structures of the first and second data sets, respectively. In Table 2, the results are listed for the SIRNAS case. Table 3 gives results for the SAS case using as isomorphous data sets $\{\mathbf{H}(S_1)\}$ and $\{-\mathbf{H}(S_1)\}$. For this case, only S_1 for the Friedel pair is used (first column of Table 1). Finally, Table 4 illustrates 2DW test results for the hypothetical structure Cf-C₆₂N₁₅O₂₂. The isomorphous data sets are defined as $\{\mathbf{H}(\lambda_1)\}$ and $\{\mathbf{H}(\lambda_2)\}$ where λ_1 and λ_2 are two different wavelengths. The last three columns of Tables 1-4 give the diffraction-ratio values from (14), (18)-(21) and (22), respectively. The tables show that the ideal diffraction ratio (14) [$DR(\text{ideal})$] and the theoretical approximation (17) (DR2) are highly correlated. Nevertheless, in almost all instances, DR2 is slightly larger than $DR(\text{ideal})$. This can be explained as follows. The omission of the double summation makes the denominator too large and the numerator too small. Replacing f_j^o by Z_j corrects this (at least partly). We always have $f_j^o < Z_j$, so we may write

$$Z_j = f_j^o + \delta_j \quad \text{with } \delta_j > 0. \quad (23)$$

By substitution of (23) in (22), DR3 can be written as

$$\begin{aligned} DR3 = & \left[4 \sum_{j=1}^n |(f_{j1}^o + f'_{j1} + if''_{j1}) - (f_{j2}^o + f'_{j2} + if''_{j2})| \right. \\ & + (\delta_{j1} - \delta_{j2})^2 \left/ \sum_{j=1}^N |(f_{j1}^o + f'_{j1} + if''_{j1})| \right. \\ & \left. + (f_{j2}^o + f'_{j2} + if''_{j2}) + (\delta_{j1} + \delta_{j2})^2 \right]^{1/2}. \quad (24) \end{aligned}$$

Since $\delta_{j1} - \delta_{j2} < \delta_{j1} + \delta_{j2}$, DR3 is expected to be smaller than DR2. The diffraction ratio DR3 turns out to approximate DR(ideal) better than DR2 although DR3 tends to be slightly smaller than DR(ideal). In most practical cases to be discussed in connection with a possible DM application ($DR < 0.3$), DR3 is a good approximation to DR(ideal), so only DR3 will be used from now on.

3.2. The relation between the diffraction ratio and the average value of the doublet phase sum

The close relation between the diffraction ratio and the doublet phase sum (15), apparent from (14), has led to extensive tests to establish the functional form between the diffraction ratio (22) and the average doublet phase sum. For the sake of brevity, the results for the SIRAS case are not included since they are almost the same as those for the SIRNAS case.

Fig. 1 shows the relation between the diffraction ratio (22) and the mean absolute value of the doublet phase sum (15) in the SIRNAS case for a variety of randomly generated structures in space group $P1$ of various sizes ($N = 200, 300, 400, 500, 600, 700, 800$ and 1200) and their heavy-atom derivatives ($H = \text{Si, Ca, Ni, As, Rb, Rh, Ba, Er, Pt, At, Cf}$). In addition, the data for the proteins APP and C_{550} have been included as well. A linear relation between (22) and the mean absolute doublet value is apparently present (see Fig. 1). The smallest mean absolute doublet value is 1 mc (1000 mc = 2π rad) at $DR = 0.02$ while the largest value, 56 mc, is found at $DR = 0.82$.

The relation between the DR and the mean value of the doublet phase sum in the SAS case is visualized in Fig. 2 for a set of randomly generated structures in space group $P1$ of various sizes ($N = 200, 300, 400, 500, 600, 700, 800$ and 1200) and their single heavy-atom derivatives ($H = \text{Ni, As, Rb, Rh, Ba, Er,$

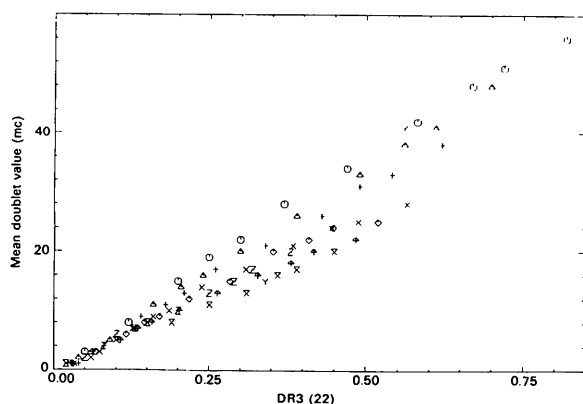


Fig. 1. The relation between DR3 (22) and the mean doublet value (in mc) in the SIRNAS case. Symbols used: \circ H -200; \triangle H -300; $+$ H -400; \times H -500; \diamond H -600; ∇ H -700; \times H -800; Σ H -1200; Υ PDB. $H = \text{Si, Ca, Ni, As, Rb, Rh, Ba, Er, Pt, At, Cf}$ and PDB = APP and C_{550} .

Ta, Pt, Hg, At, Rn, U, Bk, Cf). The DR and the average doublet phase sum again show a linear relationship. The minimum mean doublet value is 1 mc at $DR = 0.015$; the maximum is 54 mc at $DR = 0.45$. This conclusion also holds for multiple heavy-atom derivatives as can be seen from the structures $\text{Hg}_3\text{Ta-}$, Hg_4- , $\text{Hg}_3\text{Rn-}$, $\text{Hg}_3\text{U-}$ and $\text{Hg}_3\text{Bk-800}$ (symbol Υ) in Fig. 2.

Finally, in Fig. 3, the DR-doublet-phase-sum relation is shown in the 2DW* case for three small heavy-atom structures in combination with 15 possible combinations of two wavelengths from the set ($\text{Cr } K\alpha$, $\text{Fe } K\alpha$, $\text{Co } K\alpha$, $\text{Cu } K\alpha$, $\text{Mo } K\alpha$, $\text{Ag } K\alpha$). It appears that this relation is also linear. The minimum doublet value is 2 mc at $DR = 0.04$, the maximum 44 mc at $DR = 0.22$.

* For the case of 2DW, only small structures have been used since, for larger structures, the doublets became too small to be useful.

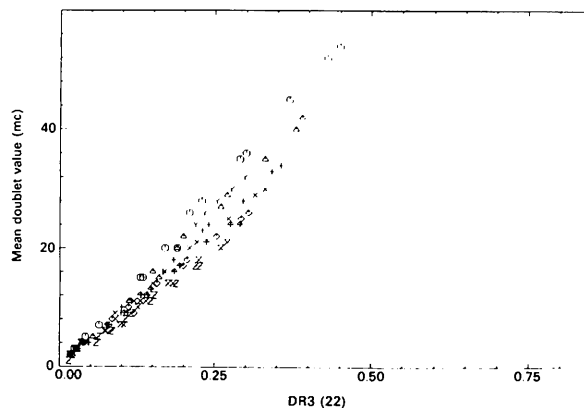


Fig. 2. The relation between the DR3 (22) and the mean doublet value (in mc) in the SAS case. Symbols used: \circ H -200; \triangle H -300; $+$ H -400; \times H -500; \diamond H -600; ∇ H -700; \times H -800; Σ H -1200; Υ H -800. $H = \text{Ni, As, Rb, Rh, Ba, Er, Ta, Pt, Hg, At, Rn, U, Bk, Cf}$; $H4 = H\text{-Hg}_3\text{-800}$ ($H = \text{Ta, Hg, Rn, U, Bk}$).

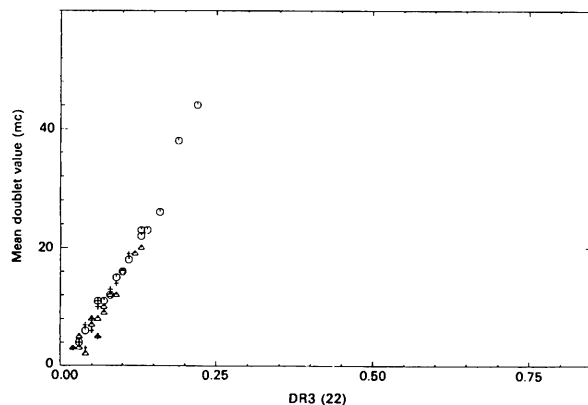


Fig. 3. The relation between the DR3 (22) and the mean doublet value (in mc) in the 2DW case. Symbols used: \circ Rn-100; \triangle U-100; $+$ Cf-100.

In conclusion, the DR-doublet-phase-sum relation is more or less linear, in particular for DR smaller than 0.20, though the linear relationship is different for the various data combinations (SIRNAS, 2DW and SAS).

3.3. The relation between the diffraction ratio and the average error of the estimation of triplet phase sums

The main purpose of this paper is to investigate the role of the diffraction ratio in potential DM applications. In practice, an average error of 30° for the most reliable triplet phase sums is usually sufficient to solve structures without difficulty *via* conventional direct methods. 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. Therefore, it is expected that the doublets, which are of the order $O(N^0)$, play a similar role with respect to the estimation of triplets amongst phases of isomorphous data sets.

In a first assessment of this hypothesis, ideal doublet values have been calculated from the atomic coordinates and used as estimates in a recently secured probabilistic expression for the estimation of triplet phase sums amongst isomorphous data [equation (84) in Peschar & Schenk, 1991]. In this way, it has been determined to what extent and how the DR and the (idealized) triplet-phase-sum estimation errors are correlated.

The calculations have been performed for the same structures as in Figs. 1–3 for the cases SIR(N)AS, SAS and 2DW. For brevity, the results for the SIRAS case have been omitted from the figures since they are almost the same as those in the SIRNAS case.

In Fig. 4(a), the relation between the DR given by (22) and the overall estimated triplet-phase-sum error is shown for the same structures as in Fig. 1. In the DR range of 0.0–0.07, the average error of the estimated triplets decreases only slowly from the random 250 mc (DR = 0.0) to 229 mc, *e.g.* for Si- n with $n = 200$ –1200 atoms and Ca- n with $n = 700$ –1200 atoms. An enormous reduction of the estimation error (from 218 to 129 mc) takes place if the DR increases from 0.07 to 0.10, *e.g.* for Ca- n with $n = 300$ –600 atoms, Ni-800, Ni-1200 and As-1200. Finally, from DR = 0.1 to 0.82, the average error reduces steadily to 52 mc (the rest of the test structures). In practical DM, only the most reliable triplets are used in the initial stages of phase determination. Figs. 4(b) and (c) show the results corresponding to Fig. 4(a) for the 50 and 20% most reliable triplets only. It is apparent from Fig. 4 that once again a DR of approximately 0.1 is a critical point. Above this limit, the average error decreases rapidly to an encouragingly low value of 60 mc.

However, below DR = 0.10, the average error of the triplets rises sharply to values above 100 mc, even for the most reliable triplets.

The same conclusions also hold in the SAS case, as can be judged from Fig. 5, but at a fixed DR level

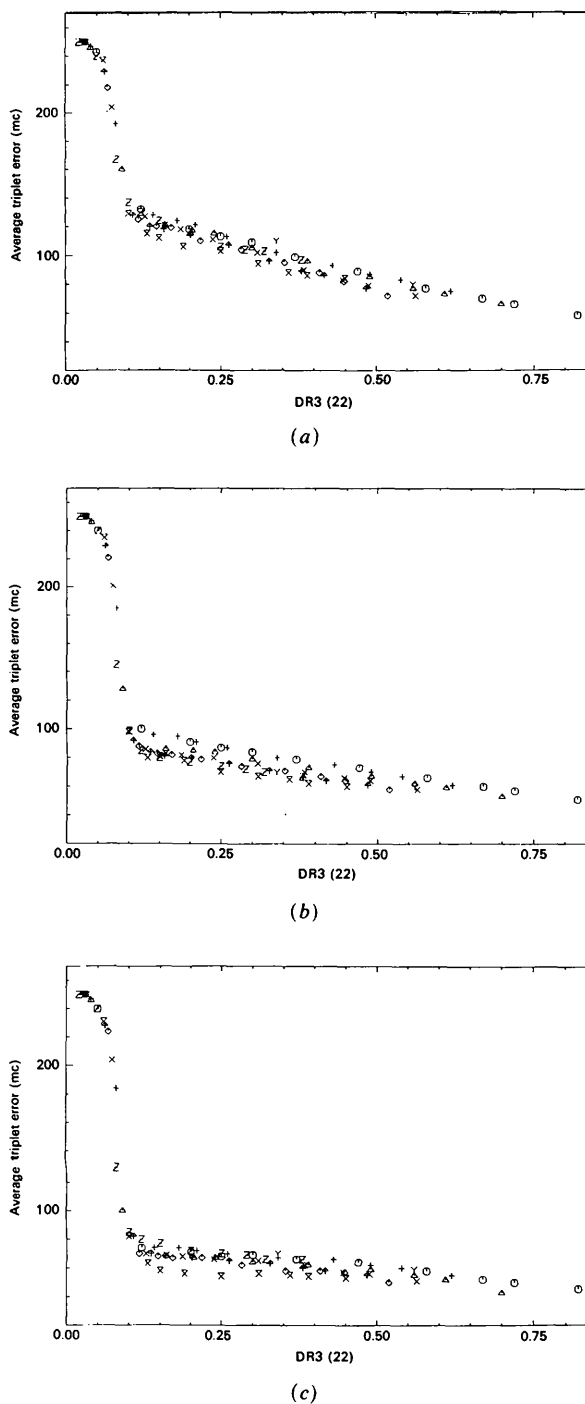


Fig. 4. The relation between the DR3 (22) and the average triplet error (in mc) in the SIRNAS case. Symbols used: as in Fig. 1. For (a) all, (b) 50% and (c) 20% of the triplets.

the average errors show a larger variation than in the SAS case. This variation results mainly from differences in structural size. An interesting anomaly in the predictions exists for the structures Ba- n and Er- n with $n = 200-600$. For Ba-600 with a DR of 0.09, the overall triplet error is 108 mc while for Er-600 with a larger DR (0.11), the error is 120 mc. The statistics for the 50 and 20% most reliable triplets show similar trends to those for the SIRNAS case.

Finally, in Fig. 6, triplet statistics are presented for the structures for which the DR-doublet relations were given in Fig. 3. For most structure-double-wavelength combinations, the average triplet errors are larger than 200 mc, corresponding to a DR smaller than 0.07. However, for each of the three structures, at least one wavelength combination can be found such that the average triplet error reduces enormously. This figure illustrates the importance of selecting the most profitable wavelength combination.

The test results have led to the formulation of two rules.

Rule 1. The larger the DR the better the estimation of the triplets for a specified pair of isomorphous structures.

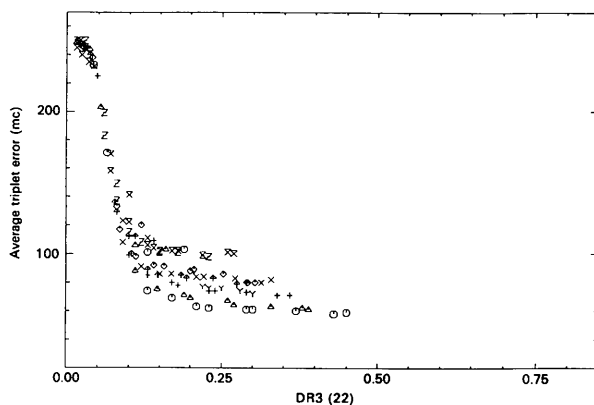


Fig. 5. The relation between DR3 (22) and the average triplet error (in mc) in the SAS case. Symbols used: as in Fig. 2.

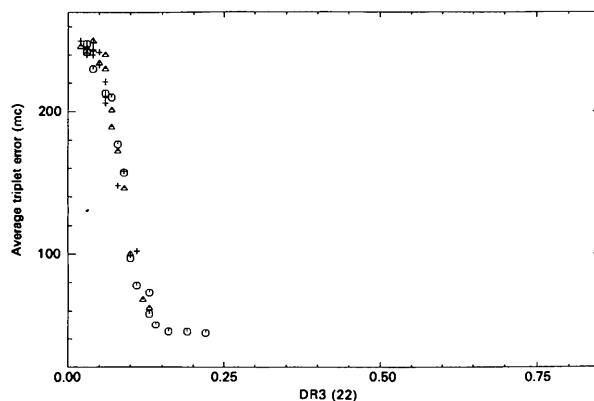


Fig. 6. The relation between the DR3 (22) and the average triplet error (in mc) in the 2DW case. Symbols used: as in Fig. 3.

In summary, a DR of 0.1 seems to be a lower limit for a DM application, at least as judged from the ideally calculated doublet phase values together with the probabilistic expression of Peschar & Schenk (1991). Structures with a $DR < 0.1$ have average errors that are too large and that may, therefore, be regarded as unsuitable for a standard DM application. Nevertheless, only estimates for the doublets will be available in practice instead of true values, which, statistically, will influence the estimates in a negative way. Hence, it can tentatively be concluded that a successful routine DM run for two isomorphous data sets is not likely to occur if the DR is smaller than 0.1. If the DR increases from 0.1 onwards, the estimation of the triplets gets better. It should be noted that a too large DR in the case of SIR(N)AS would lead to nonisomorphism. However, as was mentioned in the *Introduction*, the quality of isomorphism is not a subject of this paper. For structures with a $DR > 0.1$, the most reliable triplets have a theoretical error level that is acceptable for DM. Whether the practical error level is also small enough remains to be seen.

An analysis of the 2DW experiments has led to the definition of a second rule

Rule 2. The larger the differences of the f' and f'' values for a pair of wavelengths, the better will be the estimation of the triplets. If these differences for pairs of wavelengths are the same, the best estimation is achieved by that wavelength pair for which the sum of the f' and the f'' values is the smallest.

For example, according to this rule, the optimal pair of wavelengths for the artificial structure $Cf-C_{62}N_{15}O_{22}$ is $Cr K\alpha-Mo K\alpha$. This second rule incorporates and extends a rule for the optimal wavelength choice as formulated by Mitchel (1957) and Caticha-Ellis (1962). According to these authors, the optimal wavelength depends only on the difference of the f' values. In Fig. 7, the average errors from Fig. 6 are shown plotted *versus* the absolute difference of the

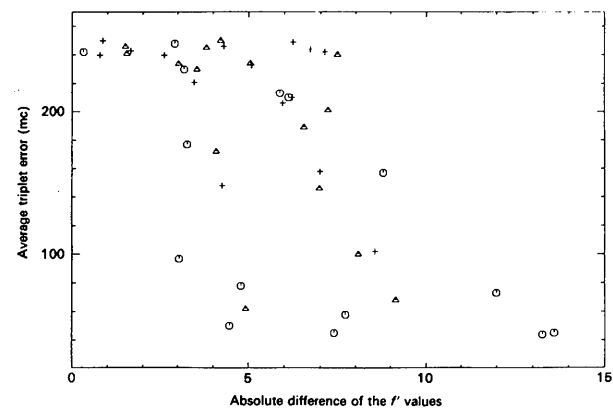


Fig. 7. The relation between the absolute differences of the values and the average triplet error (in mc) in the 2DW case. Symbols used: as in Fig. 3.

values. Obviously, the correlation present in Fig. 6 is absent in Fig. 7.

In addition to the DR already discussed, several alternative DR expressions have been investigated. Amongst these (see, for example, Srinivasan & Parthasarathy, 1976; Hendrickson & Teeter, 1981) a ratio introduced recently by Ralph & Woolfson (1991) is

$$R = \frac{\sum_{i=1}^N [(f_i^o + f_i')^2 + (f_i'')^2 - (f_i^o)^2]}{\sum_{j=1}^N [f_j^2]} \quad (25)$$

(The summation over i is for anomalous scatterers only and that over j is for all atoms.) This ratio may be modified slightly to be comparable with DR3,

$$R' = \left\{ 2 \sum_{i=1}^N [(Z_i + f_i')^2 + (f_i'')^2 - (Z_i)^2] / \sum_{j=1}^N Z_j^2 \right\}^{1/2} \quad (26)$$

Equation (26) has been tested for the same structures as in Fig. 5. The results, presented in Fig. 8, show a reduced DR-triplet-error correlation compared with Fig. 5.

In conclusion, the DR as defined in this paper can be used to predict whether a structure solution *via* a pair of isomorphously related data by means of direct methods may, in principle, be successfully attempted. In this paper only ideal (calculated) doublets have

been used. As a logical next step, to be discussed in the next paper, estimated doublets will be used instead of ideal ones (Kyriakidis, Peschar & Schenk, 1993).

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

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., SHIMANOCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535-542.
- BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*. New York: Academic Press.
- BLUNDELL, T. L., PITTS, J. E., TICKLE, I. J., WOOD, S. P. & WU, C. W. (1981). *Proc. Natl Acad. Sci. USA*, **78**, 4175-4179.
- CATICHA-ELLIS, S. (1962). *Acta Cryst.* **15**, 863-865.
- CRICK, F. H. C. & MAGDOFF, B. S. (1956). *Acta Cryst.* **9**, 901-908.
- FORTIER, S. & NIGAM, G. D. (1989). *Acta Cryst.* **A45**, 247-254.
- FORTIER, S., WEEKS, C. M. & HAUPTMAN, H. (1984). *Acta Cryst.* **A40**, 544-548.
- FUREY, W. JR, CHANDRASEKHAR, K., DYDA, F. & SAX, M. (1990). *Acta Cryst.* **A46**, 560-567.
- HAUPTMAN, H. (1982). *Acta Cryst.* **A38**, 289-294.
- HENDRICKSON, W. A. (1988). *Crystallographic Computing 4*, edited by N. W. ISSACS & M. R. TAYLOR, pp. 97-108. Oxford Univ. Press.
- HENDRICKSON, W. A. & TEETER, N. M. (1981). *Nature (London)*, **290**, 107-113.
- KARLE, J. (1989). *Acta Cryst.* **A45**, 765-781.
- KYRIAKIDIS, C. E., PESCHAR, R. & SCHENK, H. (1993). *Acta Cryst.* **A49**, 359-369.
- MITCHEL, C. M. (1957). *Acta Cryst.* **10**, 475-476.
- PESCHAR, R. (1987). Thesis, Univ. of Amsterdam, The Netherlands.
- PESCHAR, R. & SCHENK, H. (1991). *Acta Cryst.* **A47**, 428-440.
- RALPH, A. C. & WOOLFSON, M. M. (1991). *Acta Cryst.* **A47**, 533-537.
- RAMASESHAN, S. & ABRAHAMS, S. C. (1975). Editors. *Anomalous Scattering*. Copenhagen: Munksgaard.
- SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography*. New York: Pergamon.
- TIMKOVICH, R. & DICKERSON, R. E. (1976). *J. Biol. Chem.* **251**, 4033-4046.
- ZACHARIASEN, W. H. (1965). *Acta Cryst.* **18**, 714-716.

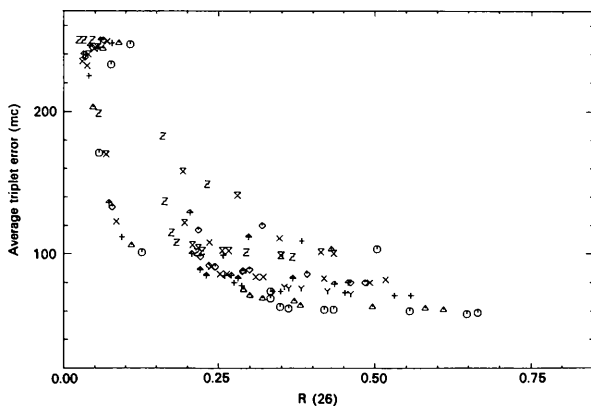


Fig. 8. The relation between R (26) and the average triplet error (in mc) in the SAS case. Symbols used: as in Fig. 2.