Evaluation of Reflection Intensities for the Components of Multiple Laue Diffraction Spots by Direct Methods

By Q. Hao

Department of Chemistry, University of Liverpool, Liverpool L69 3BX, England,
and SERC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England

J. W. Campbell

SERC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England

M. M. Harding

Department of Chemistry, University of Liverpool, Liverpool L69 3BX, England

and J. R. Helliwell

Department of Chemistry, University of Manchester, Manchester M13 9PL, England,
and SERC Daresbury Laboratory, Warrington, Cheshire WA4 4AD, England

Abstract

In a Laue diffraction pattern, 10–20% of the spots result from the exact superposition of two or more reflections that are 'harmonics', *e.g.* hkl, $2h,2k,2l$, ... For the solution of large or difficult structures, the intensities of the remaining 80–90% of the reflections, measurable as singles, may not be sufficient and thus evaluation of the intensities of the components of the multiple spots is important. A procedure for this deconvolution is given, based on the assumption of non-negativity and nonoverlapping peaks in the Patterson function. It has been tested with Laue diffraction data from an organic crystal, $C_{25}H_{20}N_2O_2$, where it allowed 275 reflection intensities to be evaluated from multiple spots, 140 of them with $|F|^2 > 3\sigma(|F|^2)$. For these 140 reflections, agreement with F_{calc} is reasonable ($R = 0.14$) and their addition to the 1129 singles made structure solution (by direct methods) significantly easier.

1. Introduction

Laue diffraction patterns, particularly those recorded with synchrotron radiation, have been increasingly used in recent years for the measurement of diffraction intensities (Helliwell, Habash, Cruickshank, Harding, Greenhough, Campbell, Clifton, Elder, Machin, Papiz & Zurek, 1989; Smith Temple & Moffat, 1987; Smith Temple, 1989; Bartunik, Bartsch & Huang, 1992). Usually, 80–90% of the spots in a single Laue diffraction pattern correspond to single reflections, each with its values of hkl and associated d (plane spacing) and λ; we describe these as singles. The remaining 10–20% of the spots are doubles, triples or higher multiples. If a crystal contains a plane of spacing d, then the spacings $d/2$, $d/3$ or, in general, d/j may

also occur, where j is any positive integer. Bragg's law is simultaneously satisfied by the sets of values (d, λ), $(d/2, \lambda/2)$, ..., $(d/j, \lambda/j)$, ... and the diffraction spots are exactly superposed. Thus, measurement of the spot intensity does not directly give the reflection intensity. Either a procedure must be devised to 'deconvolute' the measured intensities to give the individual reflection F values or the incomplete diffraction data set made up from the singles only must be used. The latter has been a common practice. Helliwell, Habash *et al.* (1989) describe one procedure for this deconvolution that uses the intensities of spots on successive films in a film pack and the variation of film absorption with λ; however, film as a detector is likely to be superseded by image plates or electronic devices, for which the method would not normally be applicable [unless two or more detectors placed one behind the other were used, as by Helliwell (1991)]. Cruickshank, Helliwell & Moffat (1987) have examined the numbers of these multiples ('energy-overlapping' reflections) and their dependence on d_{min}, λ_{min} and λ_{max}. For example, when $\lambda_{min} = 0.25$, $\lambda_{max} = 2.5$ Å and $d_{min} = 1.0$ Å, 16% of the reflections occur as multiples in the Laue diffraction pattern of a crystal with a fairly large unit cell and in a general orientation. Moreover, these reflections, which cannot be straightforwardly measured as singles, are not randomly distributed in reciprocal space; a high proportion of them are low-order reflections, axial reflections or reflections in special planes ($hk0$, hhl *etc.*). Direct-methods procedures for the solution of structures other than fairly simple ones [*e.g. MULTAN* (Germain, Main & Woolfson, 1970), *SAPI* (Yao, Zheng, Qian, Han, Gu & Fan, 1985)] are likely to fail with diffraction data that is incomplete in this way. In protein crystallography, the absence of low-order reflections has been shown to give electron-density maps

that are particularly difficult to interpret (Duke, Hadfield, Walters, Wakatsuki, Bryan & Johnson, 1992).

The procedure we describe here is based on the same principles and assumptions as 'direct methods' for phase determination (Sayre, 1952). It is similar to David's (1987) formula for extracting the relative contributions of overlapping reflections in powder diffraction.

2. Method

Sayre (1952) developed the so-called squaring method, which was useful as a method of phase determination. An analogous argument may be used for the deconvolution of overlapping reflections (multiples). Consider a Patterson function $P(\mathbf{r})$ and its square $P^2(\mathbf{r})$. Their Fourier transforms are, respectively,

$$\int P(\mathbf{r}) \cos(2\pi \mathbf{h} \cdot \mathbf{r}) \, dV = |E(\mathbf{h})|^2 \qquad (1)$$

and

$$\int P^2(\mathbf{r}) \cos(2\pi \mathbf{h} \cdot \mathbf{r}) \, dV = V^{-1} \sum_{\mathbf{k}} |E(\mathbf{k})|^2 |E(\mathbf{h} - \mathbf{k})|^2, \qquad (2)$$

where $|E(\mathbf{h})|$ is the modulus of the normalized structure factor of reflection $\mathbf{h}$. V is the unit-cell volume and the latter equation is obtained by the convolution theorem. Consider a simple Patterson function with nonoverlapping non-negative interatomic vectors: the square of that function looks similar to the function itself. Manipulation of (1) and (2) yields

$$|E(\mathbf{h})|^2 = C(\mathbf{h}) \sum_{\mathbf{k}} |E(\mathbf{k})|^2 |E(\mathbf{h} - \mathbf{k})|^2, \qquad (3)$$

where $C(\mathbf{h})$ is a scale factor (Shiono & Woolfson, 1991). For structures containing equal atoms,

$$C(\mathbf{h}) = f^2(\mathbf{h}) \Big/ \Big[\sum_{\mathbf{k}} f^2(\mathbf{k}) f^2(\mathbf{h} - \mathbf{k}) \Big], \qquad (4)$$

where f is a scattering factor appropriate to normalized structure factors; f is independent of $\mathbf{h}$ and equals $N^{-1/2}$ for an equal-atom structure where N is the number of atoms in the unit cell. [Therefore, in a structure with N equal atoms, we have $C(\mathbf{h}) = N/M_{\mathbf{h}}$, where $M_{\mathbf{h}}$ is the number of terms in the summation in (3).] Equation (3) makes the further approximation that the summation over all the measurable singles is equal to a summation over all reciprocal-lattice points. [It might be expected that the use of an origin-removed Patterson function, with $|E(\mathbf{h})|^2$ replaced in (1) by $|E(\mathbf{h})|^2 - 1$, would be more appropriate; we have made tests with this function and have found that in practice it gives worse results.]

Initially, the intensities I_{Laue} are derived for all spots on a set of film packs by integration and Lp correction. For the singles, $|F|^2$ can be derived by $|F|^2 = I_{\text{Laue}}/g(\lambda)$, where $g(\lambda)$ is the wavelength-normalization function that

has to be determined (see Helliwell, Habash et al., 1989). $|E(\mathbf{h})|$ is then derived,

$$|E(\mathbf{h})| = |F(\mathbf{h})| \exp[B(\sin^2\theta)/\lambda^2] \Big[\sum_{l=1}^{M} f_l^2(\mathbf{h}) \Big]^{-1/2}, \qquad (5)$$

where B is a temperature factor and M is the number of atoms in the unit cell. The E values for all available single reflections may then be used on the right-hand side of (3) to estimate $|E(\mathbf{h})|$ values for the reflections in the multiple spots. We then have the fractional intensity contribution of the nth reflection in an N-multiple spot,

$$\begin{aligned} r_n &= \frac{|F(\mathbf{h}_n)|^2}{\sum_{j=1}^{N} |F(\mathbf{h}_j)|^2} \\ &= \frac{|E(\mathbf{h}_n)|^2 \exp[-2B(\sin^2\theta_n)/\lambda_n^2] \sum_{l=1}^{M} f_l^2(\mathbf{h}_n)}{\sum_{j=1}^{N} |E(\mathbf{h}_j)|^2 \exp[-2B(\sin^2\theta_j)/\lambda_j^2] \sum_{l=1}^{M} f_l^2(\mathbf{h}_j)} \end{aligned} \qquad (6)$$

and also

$$\sum_{n=1}^{N} g(\lambda) |F(\mathbf{h}_n)|^2 = I_{\text{Laue}}. \qquad (7)$$

Equation (7) is used when I_{Laue} is measured on the top film; otherwise, wavelength-dependent interfilm scale factors are allowed for when the intensity has to be measured on a second or subsequent film. Equations (6) and (7) have sufficient information to evaluate individual $|F(\mathbf{h}_n)|^2$, $n = 1, ..., N$. In principle, there is sufficient information to deconvolute all higher multiples; in practice, because of experimental errors in I_{Laue} and $g(\lambda)$ and the approximation made in the derivation of (3), it does not appear to be worth going beyond triples.

Estimated standard deviations (e.s.d.s) for $|F(\mathbf{h}_n)|^2$ values have been derived from the division of $\sigma(I_{\text{Laue}})$, evaluated in the program INTLAUE (Helliwell, Habash et al., 1989), by $g(\lambda)$. They do not take any account of the fact that (3) is an approximation, although errors from this source should be uniformly distributed over all the component E values in a multiple; nor do they take account of any errors in $g(\lambda)$ or B. B must be estimated from a Wilson plot of the single reflections, which may not give the same value as a Wilson plot using all reflections. The estimate of B will also be in error if the structure is not one of randomly distributed equal atoms.

3. A test of the procedure

Synchrotron-radiation Laue diffraction patterns were recorded (at the Daresbury 9.7 SRS station) on films for an organic crystal, $C_{25}H_{20}N_2O_2$, by Helliwell, Gomez de Anderez, Habash, Helliwell & Vernon (1989) (hereafter HGHHV). They were used for structure solution and for comparison with Mo $K\alpha$ diffraction data. The crystal was

Table 1. *Analysis of reflections as a function of resolution*

All the numbers of refections refer to one unique set. Multiples refer to the doubles and triples deconvoluted by direct methods. R factors are against F_{calc} for this (known) structure.

Resolution (Å)	All reflections measured				Reflections with $|F|^2 > 3\sigma(|F|^2)$				Theory no.
	Singles		Multiples		Singles		Multiples		
	No.	R	No.	R	No.	R	No.	R	
10.00–3.000	5	0.13	6	0.07	5	0.13	6	0.07	55
3.000–2.046	33	0.05	72	0.14	32	0.05	72	0.14	107
2.046–1.552	129	0.05	55	0.12	129	0.05	52	0.09	190
1.552–1.250	231	0.05	24	0.74	224	0.05	5	0.79	294
1.250–1.047	341	0.06	45	0.85	331	0.06	5	0.77	424
1.047–0.900	390	0.12	73	0.80	330	0.09	0	–	572
All	1129	0.07	275	0.29	1051	0.06	140	0.14	1642

orthorhombic, with $a = 11.879$, $b = 17.178$, $c = 9.711$ Å, space group $P2_12_12_1$, $Z = 4$. Eight film packs of the set of ten (two were misplaced) were rescanned (50 μm raster) and the data processed with the current Daresbury software (Campbell, Clifton, Elder, Machin, Zurek, Helliwell, Habash, Hajdu & Harding, 1987; Helliwell, Habash *et al.*, 1989), with the values $\lambda_{min} = 0.25$, $\lambda_{max} = 2.6$ Å, $d_{min} = 0.9$ Å. With *LAUENORM*, 5846 intensity measurements for singles in the wavelength range 0.35 to 1.35 Å gave 1129 independent reflections (1099 in common with the 1125 reflections of HGHHV). A wavelength-normalization curve $g(\lambda)$ was established with

$$R_{merge} = \sum_i |I_i - I_m| \Big/ \sum_i I_m = 0.089, \qquad (8)$$

where I_i represents the wavelength-normalized intensity and I_m is the mean of two or more measurements of the same or symmetry-equivalent reflections (including Friedel opposites). E values defined by (5) were derived for these singles with $B = 4.0$ Å^2, which is in agreement with the average B of the known structure (HGHHV). Then, with use of (3) and (6), r_n values for the reflections in the double and triple spots were derived; there were 792 doubles and 66 triples. Equation (7) then allowed 858 $|F|^2$ values (for the wavelength range 0.35 to 1.35Å) to be derived from the double and triple spots.

There are several ways to judge the correctness of these deconvoluted $|F|$ values.

(a) When merged, they yielded 275 independent reflections wtih $R_{merge} = 0.136$ (*LAUENORM*).

(b) They were compared with the structure factors calculated from the published coordinates of HGHHV, after refinement of a scale factor; details are given in Table 1. There is a strong wavelength and resolution dependence of the accuracy of the measurements because of the errors described in the previous section and the nature of the normalization function $g(\lambda)$. At low wavelength (high resolution), the scale factor to be applied to the data, $1/g(\lambda)$, is very large and changes rapidly with λ, resulting in large errors in this region. Thus, in small-molecule cases, the lowest-wavelength component for a multiple spot is usu-

ally the least accurate. $1/g(\lambda)$ would cause large errors at both ends of the λ range for proteins.

(c) Structure solution by the direct-methods programs *MULTAN* (Debaerdemaeker, Tate & Woolfson, 1985) and *SAPI* (Yao *et al.*, 1985) was attempted with and without the deconvoluted reflections; with use of the 1129 singles only, neither program yielded a structure within 200 trials. When the 275 deconvoluted reflections were added, *SAPI* straightforwardly gave a solution (28 atoms out of 29) at the 49th trial and this solution had the highest combined figure of merit. When only the 140 deconvoluted reflections with $|F|^2 > 3\sigma(|F|^2)$ were added to the singles, a similar solution was found at trial 61, which had the highest combined figure of merit. Of these 140 reflections, 129 had $d \geq 1.8$ Å and we consider these low-resolution reflections to be important in facilitating solution by direct methods.

We note, however, that HGHHV were able to solve the structure with *MULTAN*87 using their Laue diffraction data, 1125 singles, although with considerable difficulty (Gomez de Anderez, 1990): the default run did not prove successful. However, *MULTAN* did give the solution when the starting phases were generated by *RANTAN* and the phases were determined by *SAYTAN* (Debaerdemaeker *et al.*, 1985) and the number of phase sets was increased to 200. The solution appeared at the 110th trial; it had only the third-highest combined figure of merit and 24 out of 29 atoms were found. Small differences in the two data sets presumably result in different paths in the phase-determining procedure.

As a comparison, *UNSCRAM* (Helliwell, Habash *et al.*, 1989; Zurek, Papiz, Machin & Helliwell, 1985), a program to deconvolute multiples using the differential absorption of film with wavelength, was also used. The merging R factor (*LAUENORM*) for multiples (only doubles were included in practice) was 0.30, involving 763 measurements that merged to 227 independent reflections. The R_{merge} was considerably higher than the values found for pea lectin [between 0.128 and 0.157; see Table 7 of Helliwell, Habash *et al.* (1989)]. The R factor against structure factors calculated from the published coordinates of HGHHV was 0.26, which became 0.25 for 218 reflections with $|F|^2 > 3\sigma|F|^2$. There is no strong resolution

Table 2. *Analysis of reflections (doubles) deconvoluted by
UNSCRAM as a function of resolution*

R factors are against F_{calc}.

Resolution (Å)	Reflections measured		Reflections with $\|F\|^2 > 3\sigma(\|F\|^2)$	
	No.	R	No.	R
10.00–3.000	3	0.32	3	0.32
3.000–2.046	55	0.23	52	0.21
2.046–1.552	51	0.20	51	0.20
1.552–1.250	13	0.14	13	0.14
1.250–1.047	46	0.24	44	0.24
1.047–0.900	59	0.41	55	0.41
All	227	0.26	218	0.25

dependence of the accuracy, which is shown by Table 2. After addition of these 227 reflections to the 1129 singles, *MULTAN*87 and *SAPI* found no solution within 200 trials, perhaps because of the lack of good reflections at low resolution.

4. Concluding remarks

This new procedure for deconvolution, applied to an organic crystal of medium complexity, has given useful reflection intensities for components of the multiple spots in the Laue diffraction pattern. It is slightly better than the older procedure, *UNSCRAM* (Helliwell, Habash *et al.*, 1989), in giving the low-resolution $|F|$ values, which may be crucial for structure solution by direct methods and for the interpretation of difference electron-density maps for larger structures. It has the great advantage that it is applicable to intensity measurements derived from image plates, whereas *UNSCRAM* required measurements from successive films in film packs. [Image plates are rapidly replacing films for intensity measurements and one image plate has sufficient dynamic range to replace a pack of six films (Amemiya, Matsushita, Nakagawa, Satow, Miyahara & Chikawa, 1988).] The method should not require atomic-resolution data since Sayre's (1952) equation, which is based on the same assumptions as (3), holds even with data to a resolution no better than 3 Å. However, the presence of heavy atoms could cause large errors when the squaring method is used (Shiono & Woolfson, 1991).

The application of the method to Laue intensity data from proteins is being explored. Further development is being undertaken in an approach using real space and density modification (Woolfson, personal communication). This is roughly equivalent to the approach described here using reciprocal space and relationships between E values but it conveniently allows the values for the deconvoluted F values to be improved by iteration of the procedure. It

may also allow values of F to be derived for those low-resolution reflections that have not been recorded at all in the Laue experiment.

References

AMEMIYA, Y., MATSUSHITA, T., NAKAGAWA, A., SATOW, Y., MIYAHARA, J. & CHIKAWA, J. (1988). *Nucl. Instrum. Methods Phys. Res.* A266, 644–653.

BARTUNIK, H. D., BARTSCH, H. H. & HUANG, Q. (1992). *Acta Cryst.* A48, 180–188.

CAMPBELL, J. W., CLIFTON, I. J., ELDER, M., MACHIN, P. A., ZUREK, S., HELLIWELL, J. R., HABASH, J., HAJDU, J. & HARDING, M. M. (1987). *Biophysics and Synchrotron Radiation*, edited by A. BIANCONI & A. CONGIU CASTELLANO, pp. 53–60. Heidelberg: Springer Verlag.

CRUICKSHANK, D. W. J., HELLIWELL, J. R. & MOFFAT, K. (1987). *Acta Cryst.* A43, 656–674.

DAVID, W. I. F. (1987). *J. Appl. Cryst.* 20, 316–319.

DEBAERDEMAEKER, T., TATE, C. & WOOLFSON, M. M. (1985). *Acta Cryst.* A41, 286–290.

DUKE, E. M. H., HADFIELD, A., WALTERS, S., WAKATSUKI, S., BRYAN, R. K. & JOHNSON, L. N. (1992). *Philos. Trans. R. Soc. London Ser. A*, 340, 255–261.

GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B26, 274–285.

GOMEZ DE ANDEREZ, D. (1990). PhD thesis, Univ. of York, England.

HELLIWELL, J. R. (1991). *Nucl. Instrum. Methods Phys. Res.* A308, 260–266.

HELLIWELL, J. R., HABASH, J., CRUICKSHANK, D. W. J, HARDING, M. M., GREENHOUGH, T. J., CAMPBELL, J. W., CLIFTON, I. J., ELDER, M., MACHIN, P. A., PAPIZ, M. Z. & ZUREK, S. (1989). *J. Appl. Cryst.* 22, 483–497.

HELLIWELL, M., GOMEZ DE ANDEREZ, D., HABASH, J., HELLIWELL, J. R. & VERNON, J. (1989). *Acta Cryst.* B45, 591–596.

SAYRE, D. (1952). *Acta Cryst.* 5, 60–65.

SHIONO, M. & WOOLFSON, M. M. (1991). *Acta Cryst.* A47, 526–533.

SMITH TEMPLE, B. (1989). PhD thesis, Cornell Univ., Ithaca, New York, USA.

SMITH TEMPLE, B. & MOFFAT, K. (1987). *Computational Aspects of Protein Crystal Data Analysis*, edited by J. R. HELLIWELL, P. A. MACHIN & M. Z. PAPIZ, pp. 84–89. Proceedings of a Daresbury Study Weekend. SERC Daresbury Laboratory, Warrington, England.

YAO, J., ZHENG, C., QIAN, J., HAN, F., GU, Y. & FAN, H. (1985). *SAPI85. A Computer Program for the Automatic Solution of Crystal Structures from X-ray Data*. Institute of Physics, Academia Sinica, Beijing, China.

ZUREK, S., PAPIZ, M. Z., MACHIN, P. A. & HELLIWELL, J. R. (1985). *Information Quarterly for Protein Crystallography*, No. 16. SERC Daresbury Laboratory, Warrington, England.