*Acta Cryst.* (1997). A53, 643–648

# Evaluation of Reflection Intensities for the Components of Multiple Laue Diffraction Spots by the Maximum-Entropy Method

Y. XIE AND Q. HAO*

*Department of Applied Physics, De Montfort University, Leicester LE1 9BH, England. E-mail: qhao@dmu.ac.uk*

*(Received 23 January 1997; accepted 28 April 1997)*

## Abstract

In a Laue diffraction pattern, 10–20% of the spots result from the exact superposition of two or more reflections that are 'harmonics'; a high proportion of these are low-resolution reflections. For the solution of large or difficult structural problems, the intensities of the remaining 80–90% of the reflections, measurable as singles, may not be sufficient and thus the evaluation of the intensities of the components of the multiple spots is important. A new method for this deconvolution is presented that is based on maximizing the entropy of the Patterson function subject to the constraints imposed by the observed intensities of single and overlapping reflections. This method does not require data redundancy and therefore is of particular interest for time-resolved studies on a short time scale. A new computer program (*ME*) was implemented and tested with Laue diffraction data from hen egg white lysozyme. The *R* factor between the deconvoluted reflection intensities from Laue multiple spots and observed intensities from monochromatic data was 0.116.

## 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; Bartunik, Bartsch & Huang, 1992) and for time-resolved studies of crystal structures (Schlichting *et al.*, 1990; Johnson & Hajdu, 1990; Szebenyi, Bilderback, LeGrand, Moffat, Schildkamp, Smith Temple & Teng, 1992; Singer, Smalas, Carty, Mangel & Sweet, 1993). Usually, 80–90% of the spots in a single Laue diffraction correspond to single reflections, each with its values of *hkl* and associated *d* (plane spacing) and λ, which we describe 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, λ/2), ..., (*d*/*j*, λ/*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 observed intensities to give the individual reflection *F* values or the incomplete diffraction data set made up from the singles only must be used. Cruickshank, Helliwell & Moffat (1987) have examined the numbers of these multiples ('energy-overlapping' reflections) and their dependence on $d_{min}$, $\lambda_{min}$ and $\lambda_{max}$. 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, the reflections that cannot be straightforwardly measured as singles are not randomly distributed in reciprocal space (Cruickshank *et al.*, 1987); a high proportion of them are low-order reflections, axial reflections and reflections in special planes (*hk*0, *hhl etc.*). The absence of these reflections can be a serious drawback if the data are to be used for structure solution, for example using direct methods. In protein crystallography, the absence of low-order reflections has been shown to give electron-density maps that have poor connectivity (Duke, Hadfield, Walters, Wakatsuki, Bryan & Johnson, 1992). For reasons such as these, there has been increasing interest in methods to deconvolute reflection intensities of spots that are multiples. Helliwell *et al.* (1989) gave the first method using the intensities of spots on successive films in a film pack and the variation of film absorption with λ (program *UNSCRAM*). In a second method, when multiple spots have been recorded with redundancy, the different components of a spot may be deconvoluted at the wavelength-normalization stage (Campbell & Hao, 1993). A test on hen egg-white lysozyme showed that the multiples deconvoluted by this method did indeed contribute usefully to the continuity of the maps, owing to the improved completeness of the data (Campbell, Deacon, Habash, Helliwell, McSweeney, Hao, Raftery & Snell, 1994). Ren & Moffat (1995) presented a method similar in principle but using a different algorithm and obtained good-quality deconvoluted reflections. A third method that does not depend on the recording of data on multifilm packs or redundancy of the data was described by Hao, Campbell, Harding & Helliwell (1993); it uses relationships between structure-factor magnitudes like those that are used in direct methods. Program *DECONV*

by Hao, Harding & Campbell (1995*a*) is similar in principle to the third, but uses instead a real-space modification of the Patterson function. It avoids the difficulty in the third method of choosing a '*B* factor' for the normalization of structure factors and has the advantage that the values for the deconvoluted intensities are improved by iterations of the procedure. A fourth method was presented by Bourenkov, Popov & Bartunik (1996), which is based on the Bayesian approach and Wilson statistics; the contrast in electron-density maps was improved significantly after the inclusion of the multiple reflections deconvoluted by this method.

In this paper, we demonstrate that the maximum-entropy technique can be used to evaluate individual reflection intensities from Laue multiple spots. The power of the maximum-entropy principle (Jaynes, 1979) is that it yields a most probable solution consistent with experimental observations (*i.e.* the constraints) imposed on the solution (David, 1987). The technique has been used in various areas including radio astronomy and crystallography. In crystallography, it has been applied to solve the phase problem (Collins, 1982; Wilkins, 1983; Bricogne, 1984; Navaza, 1985; Bricogne, 1991; Prince, 1993). Bricogne (1984) gave a comprehensive review of the maximum-entropy method covering aspects from the mathematical foundation to practical solutions. David (1987, 1990) proposed a maximum-entropy method for deconvoluting overlapped intensities in a powder diffraction environment. Furthermore, Bricogne (1991) discussed the problem of decomposing Laue data in a maximum-entropy environment.

## 2. Mathematical analysis and implementation

In the present context, the function to be maximized is the 'Patterson entropy'

$$S = -\sum_{\mathbf{r}} p(\mathbf{r}) \ln[p(\mathbf{r})/p_0(\mathbf{r})], \tag{1}$$

where the summation is taken over the entire unit cell, $p(\mathbf{r})$ and $p_0(\mathbf{r})$ are the Patterson function and its initial value, respectively. For each multiple spot, a constraint imposed upon the Patterson entropy is

$$\sum_j g(\lambda_j) \sum_{\mathbf{r}} p(\mathbf{r}) \cos(2\pi \mathbf{h}_j \cdot \mathbf{r}) = I_{\text{obs}}(\{\mathbf{h}_j\}), \tag{2}$$

where $g(\lambda_j)$ is the wavelength-normalization function at wavelength $\lambda_j$ and $I_{\text{obs}}$ is the measured intensity of the multiple spot.

It was shown by Jaynes (1979) that maximizing an entropy function under constraints is equivalent to finding the unconstrained minimum of its dual function. The dual function is defined by

$$\Phi(\mathbf{x}) = -\sum_{\mathbf{r}} p(\mathbf{r}) \ln[p(\mathbf{r})/p_0(\mathbf{r})] + \sum_{\{\mathbf{h}_j\}} x_{\{\mathbf{h}_j\}} \sum_j g(\lambda_j) \sum_{\mathbf{r}} p(\mathbf{r}) \cos(2\pi \mathbf{h}_j \cdot \mathbf{r}), \tag{3}$$

$\mathbf{x}$ is a vector of parameters that is related to the Lagrange multipliers. When $\Phi(\mathbf{x})$ reaches its minimum, the first derivative of $\Phi(\mathbf{x})$ should be zero, so that

$$\partial\Phi(\mathbf{x})/\partial p(\mathbf{r}) = -\ln p(\mathbf{r}) - 1 - \ln p_0(\mathbf{r}) + \sum_{\{\mathbf{h}_j\}} x_{\{\mathbf{h}_j\}} \sum_j g(\lambda_j) \cos(2\pi \mathbf{h}_j \cdot \mathbf{r}) = 0 \tag{4}$$

and then

$$p(\mathbf{r}) = p_0(\mathbf{r}) \exp\left[\sum_{\{\mathbf{h}_j\}} x_{\{\mathbf{h}_j\}} \sum_j g(\lambda_j) \cos(2\pi \mathbf{h}_j \cdot \mathbf{r})\right]. \tag{5}$$

The problem now is to find the Lagrange multipliers using equations (2) and (5). We have adopted a quasi-Newton algorithm (Prince, 1993) to solve these non-linear equations. The mathematical process is described as follows:

(i) Calculate initial values of Patterson function $p_0(\mathbf{r})$ using only single reflections *via* fast Fourier transform (FFT).

(ii) Calculate intensities for the components of each multiple spot *via* inverse FFT:

$$I_{\text{map}}(\mathbf{h}) = \sum_{\mathbf{r}} p(\mathbf{r}) \cos(2\pi \mathbf{h} \cdot \mathbf{r}). \tag{6}$$

(iii) Calculate deviation of the Lagrange multipliers $\Delta\mathbf{x}$:

$$\Delta\mathbf{x} = \Delta\mathbf{I} \cdot \mathbf{H}(\mathbf{x})^{-1}, \tag{7}$$

where $\Delta\mathbf{x}$ and $\Delta\mathbf{I}$ are both vectors, the dimension being the number of multiple spots. For each multiple spot,

$$\Delta I = \sum_j g(\lambda_j) I_{\text{map}}(\mathbf{h}_j) - I_{\text{obs}}(\{\mathbf{h}_j\}). \tag{8}$$

$\mathbf{H}(\mathbf{x})$ is the Hessian matrix. A typical element of $\mathbf{H}(\mathbf{x})$ is

$$\mathbf{H}_{kl}(\mathbf{x}) = \sum_j g(\lambda_j) \sum_i g(\lambda_i) \times \sum_{\mathbf{r}} p(\mathbf{r}) \cos(2\pi \mathbf{h}_{i,k} \cdot \mathbf{r}) \cos(2\pi \mathbf{h}_{j,l} \cdot \mathbf{r}), \tag{9}$$

where $\mathbf{h}_{i,k}$ is the $i$th component of the $k$th multiple spot. Since

$$\cos(2\pi \mathbf{h}_{i,k} \cdot \mathbf{r}) \cos(2\pi \mathbf{h}_{j,l} \cdot \mathbf{r}) = \tfrac{1}{2}\{\cos[2\pi(\mathbf{h}_{i,k} + \mathbf{h}_{j,l}) \cdot \mathbf{r}] + \cos[2\pi(\mathbf{h}_{i,k} - \mathbf{h}_{j,l}) \cdot \mathbf{r}]\}, \tag{10}$$

equation (9) reduces to

$$\mathbf{H}_{kl}(\mathbf{x}) = \tfrac{1}{2} \sum_j g(\lambda_i) \sum_i g(\lambda_i) [I_{\text{map}}(\mathbf{h}_{i,k} + \mathbf{h}_{j,l}) + I_{\text{map}}(\mathbf{h}_{i,k} - \mathbf{h}_{j,l})]. \tag{11}$$

As the diagonal elements $\mathbf{H}_{kk}(\mathbf{x})$ are much larger than the off-diagonal ones, the diagonal approximation (steepest

descents) is appropriate and this is how the computer program is implemented.

(iv) Compute the new Lagrange multipliers

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \Delta\mathbf{x}. \tag{12}$$

(v) Compute a new Patterson map using equation (5).

(vi) Repeat from step (ii) until a pre-set criterion is met. In the following test, the criterion was that the fractional intensity difference of the single reflections between input and output was $\leq 0.03$.

Combined standard uncertainties for deconvoluted multiples are derived from the division of $\sigma(I_{obs})$, evaluated from experimental measurements in an integration program, by $g(\lambda_j)$.

A flow chart of a program, *ME*, written to implement the above procedure, is shown in Fig. 1. To implement a maximum-entropy program, a few practical points need to be addressed:

(*a*) Good initial values are essential for the procedure to converge properly. Fortunately, in a Laue pattern, there are a large number of single reflections with accurately measured intensities. The initial Patterson map is calculated using those intensities. In the subsequent cycles, deconvoluted multiples are added to the singles to calculate new Patterson maps.

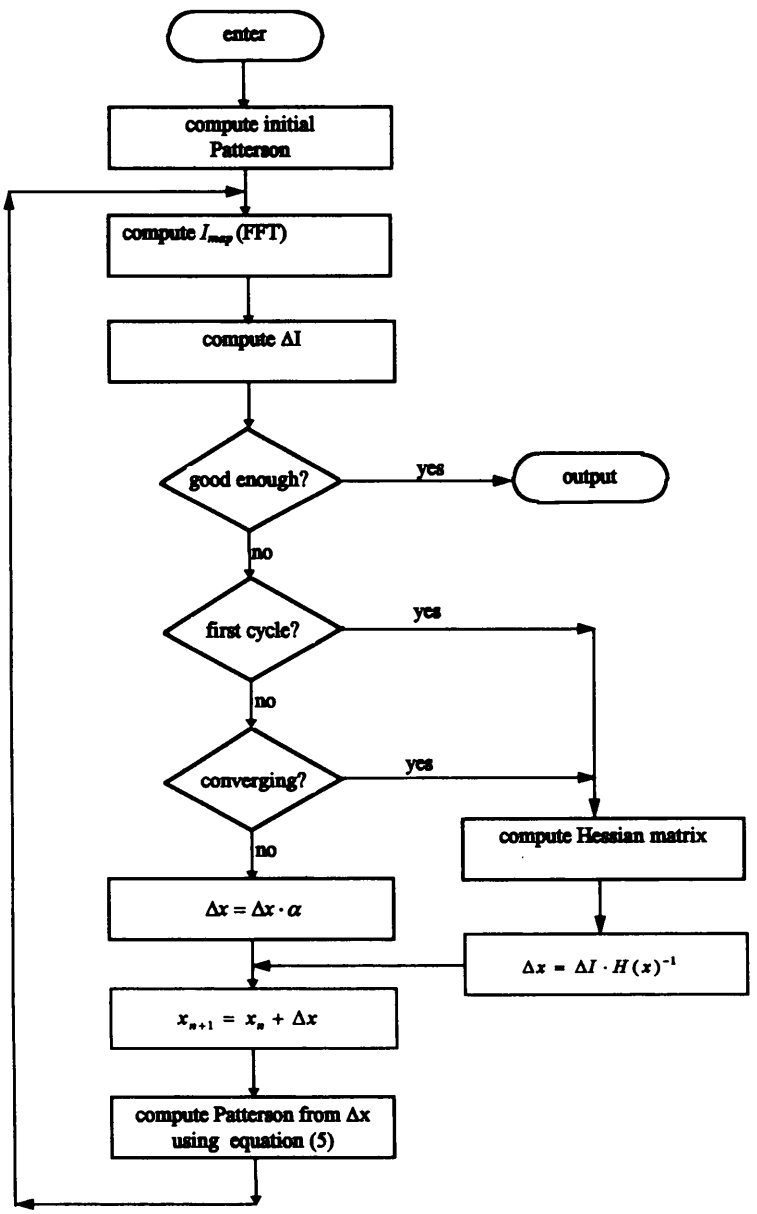



Fig. 1. Flow chart of the program for deconvoluting Laue multiple reflections using the maximum-entropy method.

(*b*) The scale factor of Patterson should be adjusted to satisfy the constraint

$$\sum_r p(r) = I(000), \qquad (13)$$

where the constant $I(000)$ can be estimated from the chemical composition of the unit cell.

(*c*) Since $\mathbf{x}$ appears in the exponential part of equation (5), any fluctuation in $\mathbf{x}$ could result in a large change in $P(\mathbf{r})$. A damping factor $\alpha$ ($< 1$) is introduced to smooth large changes and thus help the process converge.

(*d*) The method does not require data redundancy, however multiple experimental measurements are treated as extra constraints.

### 3. Test of the procedure

A test of the method was carried out using Laue data collected from tetragonal hen egg-white lysozyme (space group $P4_32_12$, $a = 79.19$, $c = 38.02$ Å). Seven Laue diffraction images recorded by Professor J. R. Helliwell's group using a MAR image plate on Station 9.5 of the Daresbury Synchrotron Radiation Source (SRS) were used. The intensity data were processed and normalized using the *LAUEGEN* and *LAUENORM* programs of the *Daresbury Laue Software Suite* (Helliwell *et al.*, 1989). The soft limits were estimated using the intensity histogram method (Hao, Harding & Campbell, 1995*b*) to be $\lambda_{min} = 0.4$, $\lambda_{max} = 1.55$ and $d_{min} = 1.9$ Å. Intensity measurements for 17 125 singles in the wavelength range 0.48–1.30 Å yielded 5296 unique reflections with

$$R = \sum_i |I_i - I_m| \Big/ \sum_i I_m = 0.067, \qquad (14)$$

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. These reflections were compared with high-quality monochromatic data (Young, Dewan, Nave & Tilton, 1993). The $R$ factor as defined in (14) between the Laue singles data and the reference monochromatic data was 0.061.

All intensities of the multiple spots were then processed by the maximum-entropy program *ME*. As a result of the deconvolution process, 1442 unique reflections were obtained. The $R$ factor between the deconvoluted multiples and monochromatic data was 0.116 for the 1257 reflections in common. Details are given in Table 1. There is a strong resolution dependence of the accuracy of measurements because of the nature of the normalization function $g(\lambda)$. At high resolution (low wavelength) and very low resolution (long wavelength), the scale factor to be applied to the data, $1/g(\lambda)$, is large and changes rapidly with $\lambda$, resulting in large errors in these regions. It is also clear that very weak reflections have large errors.

To see how the deconvoluted multiples could improve connectivity of the electron-density map, one section of the map superimposed on the refined model of Young, Dewan, Nave & Tilton (1993) is shown in Fig. 2. The discontinuous electron densities near Cα 9 of the $2F_o - F_c$ map calculated using singles only (Fig. 2*a*) become continuous when deconvoluted multiples are added in the calculation (Fig. 2*b*).

Other deconvolution methods were also tested with the same Laue data. The direct-methods program *DECONV*

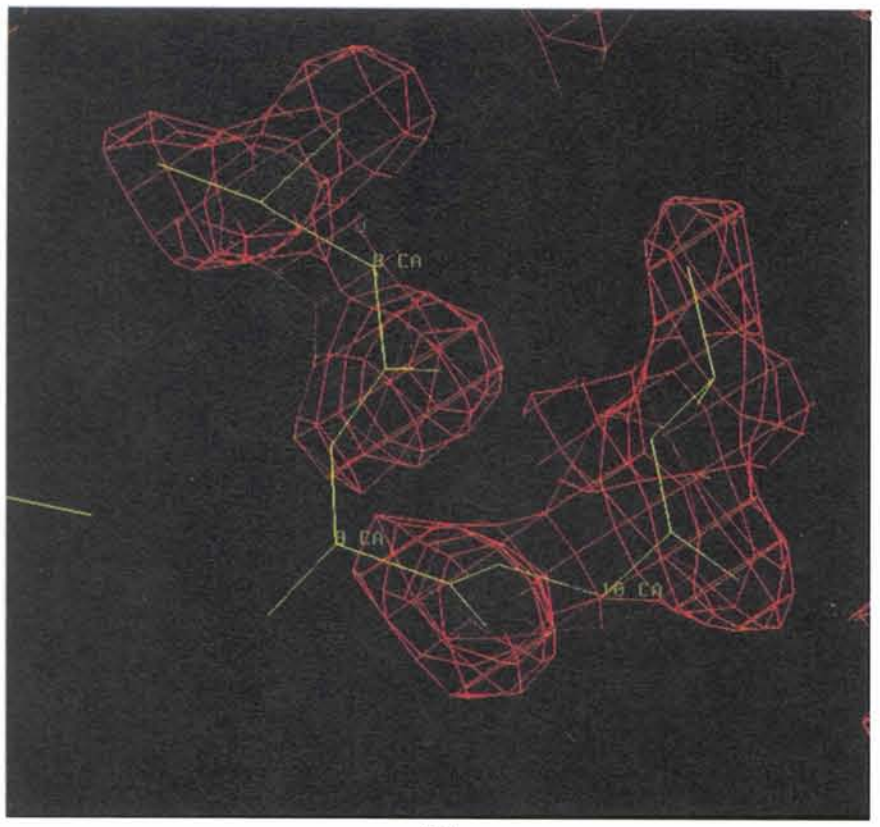
Table 1. *Analysis of the deconvoluted multiple reflections obtained by the maximum-entropy method as a function of resolution and intensities; the R factor as defined in equation (14) is calculated against monochromatic data*

| Resolution range (Å) | No. of reflections | $R$ |
|---|---|---|
| 6.83–7.54 | 16 | 0.165 |
| 6.13–6.83 | 47 | 0.075 |
| 5.43–6.13 | 76 | 0.080 |
| 4.73–5.43 | 134 | 0.097 |
| 4.03–4.73 | 215 | 0.060 |
| 3.32–4.03 | 166 | 0.129 |
| 2.62–3.32 | 208 | 0.197 |
| 1.92–2.62 | 395 | 0.287 |
| All | 1257 | 0.116 |
| **Intensity range** | **No. of reflections** | $R$ |
| >30647 | 7 | 0.046 |
| 22537–30647 | 11 | 0.046 |
| 15672–22537 | 28 | 0.092 |
| 10050–15672 | 81 | 0.078 |
| 5672–10050 | 153 | 0.098 |
| 2538–5672 | 269 | 0.111 |
| 647–2538 | 461 | 0.216 |
| <647 | 247 | 0.476 |
| All | 1257 | 0.116 |

Table 2. *Analysis of the deconvoluted multiple reflections obtained by the direct method (Hao, Harding & Campbell, 1995a) as a function of resolution and intensities; the R factor as defined in equation (14) is calculated against monochromatic data*

| Resolution range (Å) | No. of reflections | $R$ |
|---|---|---|
| 6.83–7.54 | 16 | 0.210 |
| 6.13–6.83 | 46 | 0.233 |
| 5.43–6.13 | 76 | 0.217 |
| 4.73–5.43 | 134 | 0.210 |
| 4.03–4.73 | 214 | 0.183 |
| 3.32–4.03 | 166 | 0.209 |
| 2.62–3.32 | 204 | 0.294 |
| 1.92–2.62 | 386 | 0.429 |
| All | 1242 | 0.220 |
| **Intensity range** | **No. of reflections** | $R$ |
| >15391 | 10 | 0.210 |
| 11385–15391 | 16 | 0.163 |
| 7981–11385 | 43 | 0.193 |
| 5181–7981 | 100 | 0.211 |
| 2983–5181 | 180 | 0.196 |
| 1388–2983 | 269 | 0.236 |
| 396–1388 | 333 | 0.300 |
| <396 | 291 | 0.497 |
| All | 1242 | 0.220 |

(Hao, Harding & Campbell, 1995*a*) yielded 1417 unique reflections from the multiple spots but the *R* factor of 0.220 (see Table 2) against the monochromatic data was much higher than the maximum-entropy result. The wavelength-normalization-curve method (Campbell & Hao, 1993) produced only 696 unique reflections as many multiple spots were rejected due to insufficient number of equations, *i.e.* lack of data redundancy. However, the *R* factor of 0.089 between the deconvoluted reflections and monochromatic data (see Table 3) is slightly better than that from the maximum-entropy method.



(*a*)

(*b*)

Fig. 2. A region (residues 8, 9 and 10) of the $2F_o-F_c$ electron-density map calculated using (*a*) single reflections only, (*b*) singles + deconvoluted multiples. The contour level is $1\sigma$. The connectivity has been improved significantly by the inclusion of the multiples.

Table 3. *Analysis of the deconvoluted multiple reflections obtained by the wavelength-normalization-curve method (Campbell & Hao, 1993) as a function of resolution and intensities; the R factor as defined in equation (14) is calculated against monochromatic data*

| Resolution range (Å) | No. of reflections | *R* |
|---|---|---|
| 6.83–7.54 | 11 | 0.050 |
| 6.13–6.83 | 32 | 0.042 |
| 5.43–6.13 | 48 | 0.065 |
| 4.73–5.43 | 73 | 0.047 |
| 4.03–4.73 | 103 | 0.052 |
| 3.32–4.03 | 74 | 0.076 |
| 2.62–3.32 | 109 | 0.138 |
| 1.92–2.62 | 178 | 0.300 |
| All | 628 | 0.089 |

| Intensity range | No. of reflections | *R* |
|---|---|---|
| >9224 | 9 | 0.044 |
| 6812–9224 | 11 | 0.129 |
| 4766–6812 | 33 | 0.057 |
| 3083–4766 | 45 | 0.056 |
| 1766–3083 | 102 | 0.090 |
| 813–1766 | 140 | 0.097 |
| 226–813 | 188 | 0.170 |
| <226 | 100 | 0.337 |
| All | 628 | 0.089 |

## 4. Concluding remarks

Evaluation of reflection intensities for the components of the multiple Laue diffraction spots using the maximum-entropy method has been achieved without the requirement of data redundancy. The deconvolution has been most successful with the low-resolution reflections. The inclusion of these deconvoluted reflections increases the completeness of the Laue data and therefore improves the connectivity of the electron-density map. The test results have shown substantial improvement in data quality over the previously published direct method (Hao, Harding & Campbell, 1995*a*). The maximum-entropy method complements the wavelength-normalization-curve method (Campbell & Hao, 1993; Ren & Moffat, 1995), which yields good-quality data but requires high data redundancy.

The Biotechnology and Biological Sciences Research Council (BBSRC) is thanked for a research grant to QH. We are grateful to Professor J. R. Helliwell, Dr J. W. Campbell, Dr M. M. Harding and Dr S. Arzt for providing the lysozyme Laue data and discussions. We also thank Dr E. Prince for useful discussions.

## References

Bartunik, H. D., Bartsch, H. H. & Huang, Q. (1992). *Acta Cryst.* A**48**, 180–188.

Bourenkov, G. P., Popov, A. N. & Bartunik, H. D. (1996). *Acta Cryst.* A**52**, 797–811.

Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.
Bricogne, G. (1991). *Acta Cryst.* A**47**, 803–829.
Campbell, J. W., Deacon, A., Habash, J., Helliwell, J. R., McSweeney, S., Hao, Q., Raftery, J. & Snell, E. (1994). *Bull. Mater. Sci.* **17**(1), 1–18.
Campbell, J. W. & Hao, Q. (1993). *Acta Cryst.* A**49**, 889–893.
Collins, D. M. (1982). *Nature (London)*, **298**, 49–51.
Cruickshank, D. W. J., Helliwell, J. R. & Moffat, K. (1987). *Acta Cryst.* A**43**, 656–674.
David, W. I. F. (1987). *J. Appl. Cryst.* **20**, 316–319.
David, W. I. F. (1990). *Nature (London)*, **346**, 731–734.
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.
Hao, Q., Campbell, J. W., Harding, M. M. & Helliwell, J. R. (1993). *Acta Cryst.* A**49**, 528–531.
Hao, Q., Harding, M. M. & Campbell, J. W. (1995*a*). *J. Synchrotron Rad.* **2**, 27–30.
Hao, Q., Harding, M. M. & Campbell, J. W. (1995*b*). *J. Appl. Cryst.* **28**, 447–450.
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.
Jaynes, E. T. (1979). *The Maximum Entropy Formalism*, pp. 15–118. Cambridge, MA: MIT Press.
Johnson, L. N. & Hajdu, J. (1990). *Synchrotron Radiation and Biophysics*, edited by S. S. Hasnain, ch. 6. Chichester: Ellis Horwood.
Navaza, J. (1985). *Acta Cryst.* A**41**, 232–244.
Prince, E. (1993). *Acta Cryst.* D**49**, 61–65.
Ren, Z. & Moffat, K. (1995). *J. Appl. Cryst.* **28**, 482–493.
Schlichting, I., Almo, S. C., Rapp, G., Wilson, K., Petratos, K., Lentfer, A., Wittinghofer, A., Kabsch, W., Pai, E. E., Petsko, G. A. & Goody, R. S. (1990). *Nature (London)*, **345**, 309–315.
Singer, P. T., Smalas, A., Carty, R. P., Mangel, W. F. & Sweet, R. M. (1993). *Science*, **259**, 669–673.
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. SERC Daresbury Laboratory, Warrington, England.
Szebenyi, D. M. E., Bilderback, D. H., LeGrand, A., Moffat, K., Schildkamp, W., Smith Temple, B. & Teng, T.-Y. (1992). *J. Appl. Cryst.* **25**, 414–423.
Wilkins, S. W. (1983). *Acta Cryst.* A**39**, 892–896.
Young, A. C. M., Dewan, J. C., Nave, C. & Tilton, R. F. (1993). *J. Appl. Cryst.* **26**, 309–319.