Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Estimating unobserved reflection intensities in Laue diffraction by the maximum-entropy method

Y. Xie and Q. Hao*

Department of Chemistry and Physics, De Montfort University, Leicester LE1 9BH, England

Correspondence e-mail: qhao@dmu.ac.uk

Received 20 March 1998
Accepted 23 July 1998

In protein crystallography, the use of low-resolution reflections is important in defining the molecular mask and polypeptide backbone. However, in Laue data collection, the loss of low-resolution reflection data ($>2d_{\rm min}$) can be as high as 40–50%, even after the deconvolution of multiples. To estimate the reflection intensities that are not recorded in data collection, a new method is presented based on maximizing the entropy of the Patterson function subject to the constraints imposed by the intensities of the observed reflections. The method has been tested with Laue diffraction data from hen egg-white lysozyme. All unobserved reflections within 5 Å resolution were estimated, and their inclusion in the electron-density-map calculation significantly improved the connectivity. This method could also be applied to improve the completeness of monochromatic data.

1. Introduction

Karle & Hauptman (1964) proposed that the Patterson function could be used for data extrapolation beyond the observed range of the actual measured data. In their paper, they presented an iterative procedure whereby a sharpened origin-removed Patterson function could be utilized to extrapolate data beyond the resolution limit of the observed data. In their test case, the final Patterson map showed remarkably improved peak resolution, but the actual precision of the extrapolated E values was not objectively quantified. One paper (Seeman *et al.*, 1976) reported using this method to obtain better $E(hkl)$ estimates within the observed set of measured data (without extrapolation). Langs (1998) revised the method and used new calculation protocols to extend observed data sets from 1 to 0.5 Å with reasonably good precision.

In some time-resolved Laue diffraction experiments, 40–50% of the reflection data, especially at low resolution, cannot be measured owing to the Laue geometry and the limited number of exposures. These reflections play an important role when an electron-density map is calculated from the Fourier transformation of the reflection data. In protein crystallography, the absence of low-resolution reflections can result in electron-density maps that have poor connectivity. To solve this problem, we propose using the maximum-entropy method to estimate the unmeasured reflections. This method is similar in principle to the method we have used in a previous study (Xie & Hao, 1997), in which the maximum-entropy method was used to deconvolute Laue multiple-diffraction spots.

Maximum-entropy theory was first presented by Shannon & Weaver (1949) and its applications were thoroughly discussed by Jaynes (1979). In crystallography, the theory has been used to solve the phase problem (Collins, 1982; Wilkins, 1983;

Bricogne, 1984; Navaza, 1985; Prince, 1993; Gilmore, 1996). A comprehensive review of its use in crystallography was presented by Gilmore (1996). The power of the maximum-entropy principle is that it yields a most probable solution consistent with experimental observations (*i.e.* the constraints) imposed on the solution. David (1987) has proposed a maximum-entropy method for deconvoluting overlapped intensities in powder diffraction. Bricogne (1991) has discussed the problem of decomposing Laue data in a maximum-entropy environment.

2. Mathematical analysis and implementation

Reflection intensities for the components of multiple Laue diffraction spots can be deconvoluted by the maximum-entropy method (Xie & Hao, 1997), direct methods (Hao *et al.*, 1993, 1995*a*) or the wavelength-normalization method (Campbell & Hao, 1993; Ren & Moffat, 1995). These reflection intensities, together with those of the single reflections, will later be used in the present study as constraints (called 'known reflections' hereafter). In the present context, the function to be maximized is the 'Patterson entropy'

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

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

$$\sum_{\mathbf{r}} p(\mathbf{r}) \cos(2\pi\mathbf{h}\cdot\mathbf{r}) = I_{\rm obs}(\mathbf{h}), \qquad (2)$$

where $I_{\rm obs}(\mathbf{h})$ is the observed intensity of the reflection.

It was shown by Prince (1993) that maximizing an entropy function under constraints is equivalent to finding the unconstrained minimum of its dual function. The dual function is defined by

$$\begin{aligned}\Phi(\chi) = &-\sum_{\mathbf{r}} p(\mathbf{r}) \ln[p(\mathbf{r})/p_0(\mathbf{r})] \\ &+ \sum_{\mathbf{h}} x_{\mathbf{h}}\Big[\sum_{\mathbf{r}} p(\mathbf{r})\cos(2\pi\mathbf{h}\cdot\mathbf{r}) - I_{\rm obs}(\mathbf{h})\Big], \qquad (3)\end{aligned}$$

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

$$\begin{aligned}\frac{\partial\Phi(\chi)}{\partial p(\mathbf{r})} &= -\ln p(\mathbf{r}) - 1 - \ln p_0(\mathbf{r}) + \sum_{\mathbf{h}} x_{\mathbf{h}} \cos(2\pi\mathbf{h}\cdot\mathbf{r}) \\ &= 0 \qquad (4)\end{aligned}$$

and

$$p(\mathbf{r}) = p_0(\mathbf{r}) \exp\Big[\sum_{\mathbf{h}} x_{\mathbf{h}} \cos(2\pi\mathbf{h}\cdot\mathbf{r})\Big]. \qquad (5)$$

The problem now is to solve the following non-linear multivariable equations in order to find the Lagrange multipliers $\chi_{\mathbf{h}}$:

$$\sum p_0(\mathbf{r}) \exp\Big[\sum_{\mathbf{h}'} x_{\mathbf{h}'} \cos(2\pi\mathbf{h}'\cdot\mathbf{r})\Big] \cos(2\pi\mathbf{h}\cdot\mathbf{r}) = I_{\rm obs}(\mathbf{h}). \qquad (6)$$

We have adopted a quasi-Newton algorithm (Prince, 1993) to solve these non-linear equations. The mathematical process is described as follows.

(i) Calculate the initial values of the Patterson function $p_0(\mathbf{r})$ using $I(000)$ *via* fast Fourier transform (FFT).

(ii) Calculate intensities of the known reflections *via* inverse FFT,

$$I_{\rm map}(\mathbf{h}) = \sum_{\mathbf{r}} p(\mathbf{r}) \cos(2\pi\mathbf{h}\cdot\mathbf{r}). \qquad (7)$$

(iii) Calculate deviation of the Lagrange multipliers $\Delta\boldsymbol{\chi}$ using

$$\Delta\boldsymbol{\chi} = \Delta\mathbf{I}\cdot\mathbf{H}(\chi)^{-1}, \qquad (8)$$

where $\Delta\boldsymbol{\chi}$ and $\Delta\mathbf{I}$ are both vectors, the dimension being the number of known reflections. For each reflection

$$\Delta I = I_{\rm map}(\mathbf{h}) - I_{\rm obs}(\mathbf{h}). \qquad (9)$$

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

$$\mathbf{H}_{ij}(\chi) = \sum_{\mathbf{r}} p(\mathbf{r}) \cos(2\pi\mathbf{h}_i\cdot\mathbf{r}) \cos(2\pi\mathbf{h}_j\cdot\mathbf{r}). \qquad (10)$$

Since

$$\begin{aligned}\cos(2\pi\mathbf{h}_i\cdot\mathbf{r})\cos(2\pi\mathbf{h}_j\cdot\mathbf{r}) = 0.5\{&\cos[2\pi(\mathbf{h}_i+\mathbf{h}_j)\cdot\mathbf{r}] \\ &+ \cos[2\pi(\mathbf{h}_i-\mathbf{h}_j)\cdot\mathbf{r}]\}, \qquad (11)\end{aligned}$$

(10) reduces to

$$\mathbf{H}_{ij}(\chi) = 0.5[I_{\rm map}(\mathbf{h}_i+\mathbf{h}_j) + I_{\rm map}(\mathbf{h}_i-\mathbf{h}_j)]. \qquad (12)$$

As the diagonal elements $\mathbf{H}_{kk}(\chi)$ 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

$$\boldsymbol{\chi}_{n+1} = \boldsymbol{\chi}_n + \Delta\boldsymbol{\chi}. \qquad (13)$$

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

(vi) Repeat from step (ii) until the process converges, *i.e.* a pre-set criterion is met. In the tests described below, this criterion was the requirement that the fractional change in intensity from input to output should be <0.03.

A flow chart of a program *ME*3.3, modified from the *ME* version described in Xie & Hao (1997) to implement the above procedure, is shown in Fig. 1.

3. Tests 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 imaging plate on Station 9.5 of the Daresbury Synchrotron

Table 1
Lysozyme: analysis of the reflection intensities estimated by the maximum-entropy method at different resolutions .

The *R* factor, as defined in equation (14), is calculated against monochromatic data.

d spacing (Å)	>4.0	>5.0	>6.0	>7.0
Number of estimated reflections	355	243	174	130
Number of estimated reflections in common with monochromatic data	229	119	52	17
R factor (on *I*)	0.444	0.405	0.391	0.346

Table 2
Lysozyme: detailed analysis of the estimated reflections as a function of resolution and intensities.

The *R* factor, as defined in equation (14), is calculated against monochromatic data. The numbers in the second column are small due to the incompleteness of the monochromatic data at these *d* spacings.

Range of *d* spacings (Å)	Number of reflections in common with monochromatic data	*R* factor (on *I*)
>7.22	9	0.333
6.90–7.22	9	0.301
6.50–6.90	8	0.430
6.27–6.50	9	0.422
5.96–6.27	18	0.405
5.64–5.96	16	0.475
5.32–5.64	22	0.352
5.01–5.32	28	0.441
All	119	0.405

Intensity range	Number of reflections in common with monochromatic data	*R* factor (on *I*)
>2936	1	0.206
2686–2936	5	0.227
2446–2686	23	0.406
2218–2446	39	0.446
2001–2218	17	0.397
1795–2001	20	0.338
1600–1795	14	0.478
All	119	0.405

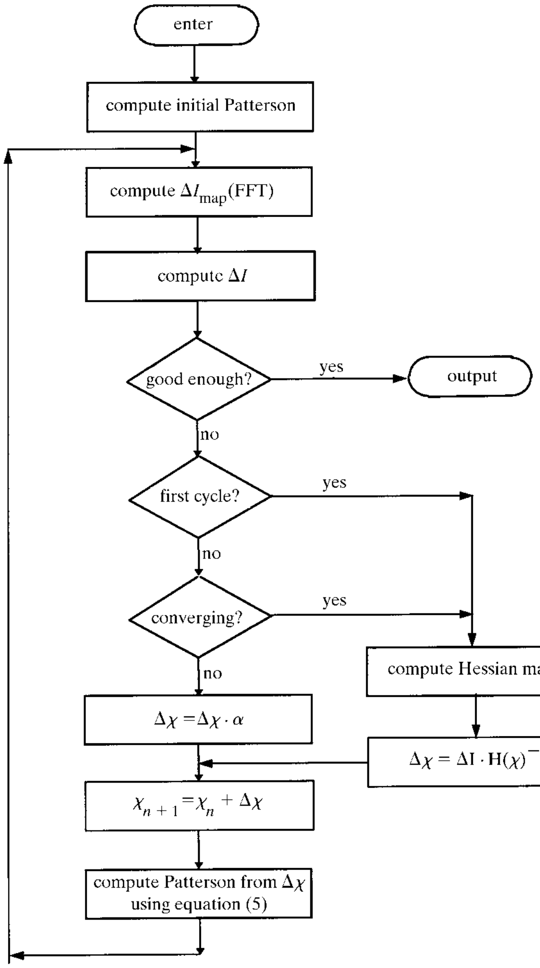

Figure 1
Flow chart of the program using the maximum-entropy method for estimating reflection intensities that are not recorded experimentally.

Radiation Source (SRS) were used. The intensity data were processed and normalized using the *LAUEGEN* and *LAUENORM* programs of the *Daresbury Laue Software Suite* (Campbell *et al.*, 1998; Helliwell *et al.*, 1989). The soft limits were estimated using the intensity histogram method (Hao *et al.*, 1995*b*) to be $\lambda_{\min} = 0.4$, $\lambda_{\max} = 1.55$ and $d_{\min} = 2.1$ Å. Intensity measurements for 17125 singles in the wavelength range 0.48–1.30 Å yielded 5296 unique reflections with

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

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 *et al.*, 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* (Xie & Hao, 1997). 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 common to the two sets (Xie & Hao, 1997). At 5 Å resolution, the number of single reflections was 272, representing a completeness of 43%; the combined number of single and deconvoluted multiple reflections was 383, representing 61% completeness.

Intensities of all unobserved reflections were then estimated using the program *ME*3.3 with the combined data set of single and deconvoluted multiple reflections being used as constraints. The reflections estimated in this way within resolutions of 4.0, 5.0, 6.0 and 7.0 Å were compared with monochromatic data (Table 1). It can be seen that the *R* factor for reflections within 7.0 Å is the lowest, but the number of reflections available in this range is very small. After taking into account the numbers of reflections estimated in different ranges of *d* spacings and their *R* factors, those with $d \geq 5.0$ Å were accepted for inclusion in calculated electron-density maps. Further *R*-factor analysis is shown in Table 2.

The inclusion of the estimated reflections that were not observed in the Laue experiment has improved the data completeness at 5 Å resolution from 61 to 100%. To see how the connectivity of the electron-density map is improved, one section of the map is superimposed on the refined model of Young *et al.* (1993) and shown in Fig. 2. The discontinuous electron densities between Cα5 and Cα6 of the map calculated using observed intensities only become continuous when unobserved but estimated intensities are added to the calculation. The map correlation coefficients (CC) were also calculated between the electron-density maps computed using Laue data and monochromatic data; for the resolution limits 7, 6 and 5 Å, the CCs were 0.243, 0.541 and 0.694, respectively, without the estimated reflections, and were enhanced to 0.585, 0.616 and 0.778, respectively, with these reflections.

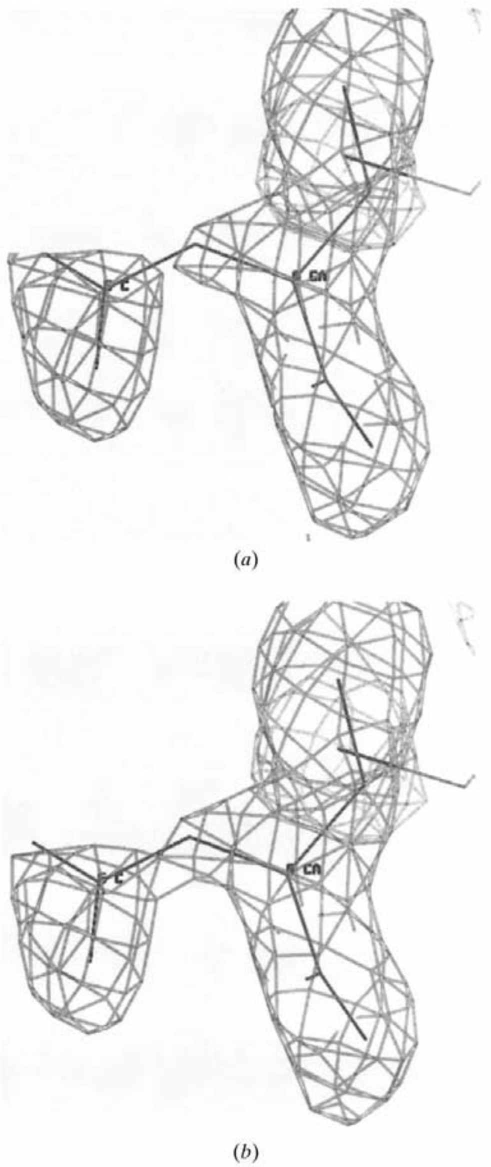

Figure 2
A region (residues 5 and 6) of the electron-density map of lysozyme calculated using (*a*) observed reflections (singles and deconvoluted multiples) only, (*b*) observed and estimated unobserved reflections. The contour level is 1σ. The connectivity has been improved significantly by the inclusion of the estimated data.

Table 3
Cytochrome *c* peroxidase (CCP): detailed analysis of the estimated reflections as a function of resolution and intensities.

The *R* factor, as defined in equation (14), is calculated against monochromatic data.

Range of *d* spacings (Å)	Number of reflections in common with monochromatic data	*R* factor (on *I*)
>9.25	339	0.448
8.50–9.25	85	0.406
7.75–8.50	133	0.409
7.00–7.75	173	0.410
6.25–7.00	264	0.407
5.50–6.25	429	0.397
4.75–5.50	727	0.376
4.00–4.75	796	0.410
All	2946	0.406

Intensity range	Number of reflections in common with monochromatic data	*R* factor (on *I*)
>214	8	0.487
201–214	47	0.391
189–201	99	0.391
177–189	200	0.464
165–177	374	0.410
154–165	675	0.405
143–154	1069	0.387
133–143	474	0.414
All	2946	0.406

Table 4
4Zn insulin: detailed analysis of the estimated reflections as a function of resolution and intensities.

The *R* factor, as defined in equation (14), is calculated against monochromatic data

Range of *d* spacings (Å)	Number of reflections in common with monochromatic data	*R* factor (on *I*)
>9.13	67	0.509
8.25–9.13	24	0.401
7.38–8.25	31	0.345
6.50–7.38	61	0.374
5.63–6.50	93	0.318
4.75–5.63	164	0.389
3.88–4.75	234	0.346
3.00–3.88	467	0.400
All	1141	0.387

Intensity range	Number of reflections in common with monochromatic data	*R* factor (on *I*)
>1434	2	0.243
1320–1434	2	0.187
1211–1320	10	0.193
1107–1211	343	0.401
1007–1107	408	0.391
912–1007	186	0.370
822–912	134	0.361
736–822	56	0.434
All	1141	0.387

A second test was carried out with Laue data from cytochrome *c* peroxidase (CCP; Fulop *et al.*, 1994). The crystal is orthorhombic, space group $P2_12_12_1$, with unit-cell parameters $a = 107.4$, $b = 76.8$, $c = 51.4$ Å and $Z = 4$. The Laue diffraction patterns recorded on four film packs had been processed with the Laue evaluation and analysis program *LEAP* (Fulop *et al.*, 1994), with $\lambda_{\min} = 0.25$, $\lambda_{\max} = 2.06$ and $d_{\min} = 2.2$ Å. The overall completeness of 9428 unique single reflections processed from four film packs to 2.2 Å resolution was 53%. The *R* factor, as defined in (14), between the Laue singles data and the reference monochromatic data was 0.130. Intensities of all remaining reflections to 4.0 Å were then estimated using the program *ME*3.3 while the single reflections were used as constraints. These estimated reflections were then compared with monochromatic data (Table 3), giving an overall $R = 0.406$.

The *ME* procedure was also tested with Laue data from 4Zn insulin (Hao *et al.*, 1995*a*). The crystal is trigonal, space group *R*3, with unit-cell parameters $a = b = 80.7$, $c = 37.6$ Å and $Z = 9$. The Laue diffraction patterns recorded on two film packs had been processed with the *Daresbury Laue Software Suite* (Helliwell *et al.*, 1989), with $\lambda_{\min} = 0.30$, $\lambda_{\max} = 2.0$ and $d_{\min} = 2.3$ Å. The overall completeness of 2053 unique single reflections processed from two film packs to 2.3 Å resolution was 49%. The *R* factor between the Laue singles data and the reference monochromatic data was 0.134. Intensities of all remaining reflections to 3.0 Å were then estimated using the program *ME*3.3 with the single reflections used as constraints. These estimated reflections were then compared with monochromatic data (Table 4) giving an overall $R = 0.387$.

4. Concluding remarks

It is well known that high data completeness at low resolution is essential in calculating electron-density maps; however, in Laue experiments a high proportion of low-resolution reflection data are often not recorded. We have demonstrated that the maximum-entropy method can be used to estimate these reflection intensities. The starting map is calculated using only the $I(000)$ term to ensure minimum bias, and all the observed data are used as constraints in the entropy-maximization process. The test results have shown that useful data can be obtained from the computer program *ME*3.3 which is based on this method. The inclusion of these estimated data increases the data completeness and therefore improves the connectivity of the electron-density map. It is worth noting that this method is not restricted to use with Laue data, as it could equally well be applied to data collected with monochromatic radiation.

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, Dr Wakatsuki and Dr S. Arzt for providing the Laue data and discussions.

References

Bricogne, G. (1984). *Acta Cryst.* A**40**, 410–445.

Bricogne, G. (1991). *Acta Cryst.* A**47**, 803–829.

Campbell, J. W. & Hao, Q. (1993). *Acta Cryst.* A**49**, 889–893.

Campbell, J. W., Hao, Q., Harding, M. M., Nguti, N. D. & Wilkinson, C. (1998). *J. Appl. Cryst.* **31**, 496–502.

Collins, D. M. (1982). *Nature (London)*, **298**, 49–51.

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

Fulop, V., Phizackerley, R. P., Solits, S. M., Clifton, I. J., Wakatsuki, S., Erman, J., Hadju, J. & Edwards, S. L. (1994). *Structure*, **2**(3), 201–208.

Gilmore, C. J. (1996). *Acta Cryst.* A**52**, 561–589.

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.

Karle, J. & Hauptman, H. (1964). *Acta Cryst.* **17**, 392–396.

Langs, D. A. (1998). *Acta Cryst.* A**54**, 44–48.

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.

Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P. & Rich, A. (1976). *J. Mol. Biol.* **104**, 109–144.

Shannon, C. E. & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Wilkins, S. W. (1983). *Acta Cryst.* A**39**, 892–896.

Xie, Y. & Hao, Q. (1997). *Acta Cryst.* A**53**, 643–648.

Young, A. C. M., Dewan, J. C., Nave, C. & Tilton, R. F. (1993). *J. Appl. Cryst.* **26**, 309–319.