Acta Cryst. (1998). A**54**, 626–631

Iterative Low-Pass Filtering for Estimation of the Background in Fiber Diffraction Patterns

Magdalena I. Ivanova and Lee Makowski*

Institute of Molecular Biophysics, The Florida State University, Tallahassee, FL 32306, USA.
E-mail: lmakowsk@nsf.gov

(*Received 7 January 1998; accepted 23 February 1998*)

Abstract

Disorientation of particles in fibrous specimens results in diffraction patterns in which the layer-line intensity is arced about the center, causing layer lines to overlap at larger radii and, as a consequence, making correction of the background problematic. Taking advantage of the fact that the background is typically made up of spatial frequencies that are substantially lower than those constituting the diffracted intensities, estimates of background in fiber diffraction patterns were made by applying an iterative low-pass filter to the diffraction data. This procedure has been applied to X-ray fiber diffraction patterns from filamentous bacteriophages to calculate the intensities along layer lines as well as to correct for the background.

1. Introduction

Background in fiber diffraction patterns is caused by film fog, X-ray scattering from air, solvent, sample holder and components of the camera. Several methods have been proposed for estimation of the background, including the subtraction of an experimentally collected pattern of the background or computational simulation of the shape of the background.

Fraser *et al.* (1976) used base-line-corrected intensities assuming that background is linearly varying with the angle. The parameters for the background base line were determined by using experimentally and computationally estimated values for the size of the crystallites and disorientation parameter. Makowski (1978) approximated the background by a set of functions with one or more variable parameters and solved for these parameters at each radius, calculating them simultaneously with the layer-line intensities. Variations of this method are now routinely used for background extraction and are also known as a polynomial profile fitting (Lorenz & Holmes, 1993; Marvin *et al.*, 1987; Namba *et al.*, 1989; Wang & Stubbs, 1994).

Background can also be simulated by a two-dimensional linear expansion of cylinder functions with calculation of the background made through the use of sampled points between layer lines (Millane & Arnott, 1985). The background function is adjusted by varying the number of equations and parameters. However, this method requires data points known to be devoid of layer-line intensity, which makes correction for the background problematic at resolutions where the layer lines overlap. Even though this procedure allows easier adjustment of radial and azimuthal variations in the background, it is only suitable for diffraction from highly oriented specimens.

All current methods place implicit limits on the form of the background to which they may be applied. When the background does not conform to these assumptions, its magnitude is poorly estimated. In this paper, we describe an algorithm to estimate the background by exploiting the fact that the background in fiber diffraction patterns consists primarily of low frequencies. The estimation of the background depends on the frequency limit of the filter in both radial and azimuthal directions. These limits are determined empirically by examination of the estimated background as demonstrated here. The processing of the data is performed iteratively in such a way that estimates of the background are alternated with estimates of layer-line intensities.

2. Methods

Highly oriented fibers of filamentous bacteriophage M13 were prepared according to Nave *et al.* (1981) and Glucksman *et al.* (1992). X-ray diffraction patterns from an oriented fiber of mutant fd (fddx) were obtained using a rotating-anode X-ray source (Rigaku, RU 200) with double-mirror optics. The exposure time was 72 h and the dimensions of the resulting monochromatic X-ray beam were 0.30×0.30 mm at the film.

Optical densities on the film were measured on a 25 μm square raster with a Perkin–Elmer PDS microdensitometer. Each diffraction pattern was corrected for film nonlinearity, twist, tilt and polarization using techniques described by Fraser *et al.* (1976). The optical densities measured in this way were averaged onto a polar coordinate system in reciprocal space (Makowski, 1978).

ISSN 0108-7673

3. Theory

3.1. *Deconvolution procedure*

In reciprocal space, diffraction from linear periodic structures is confined to parallel layer planes, which are separated by a distance equal to $1/c$, where c is the axial repeat distance. The intersection of the sphere of reflection with these layer planes gives rise to layer lines, which are seen as hyperbolas on a flat film. In practice, the particle axes are disoriented relative to the fiber axis z with an angular spread characterized by a half-width, σ_φ. As a consequence, the reflections on each layer line are tangentially broadened to become arcs in accordance with an angular distribution function $f(\varphi, \varphi_i)$. Holmes & Barrington Leigh (1974) approximated the intensity distribution function $f(\varphi, \varphi_i)$ by a Gaussian, assuming that disorientation of the specimens is caused by thermal displacement of the particles from their mean orientation. Their results were later generalized by Makowski (1978), who showed that the disorientation function can also be approximated by a Gaussian for many diffraction patterns with lower degrees of order than those caused by thermal vibrations. The intensity distribution function can be written as (Holmes & Barrington Leigh, 1974)

$$f(\varphi, \varphi_i) = \frac{1}{R_i \sigma_\varphi (2\pi)^{1/2}} \exp\left[\frac{-(\varphi - \varphi_i)^2}{2\sigma_\varphi{}^2}\right], \qquad (1)$$

where R_i is the distance from the meridian to the reflection i and σ_φ is the standard uncertainty. This approximation is valid only for those regions that are not in the immediate vicinity of the meridian. Near the meridian, additional geometric corrections are required (Makowski, 1978).

There is also spreading of the layer-line intensities due to convolution with the X-ray beam. Larger beams cause greater spreading of the reflections, which combined with angular disorientation can cause the overlapping of the layer lines at relatively low radii. In order to correct for the spread due to the beam size, we approximated the beam function $b(x, y)$ as a two-dimensional Gaussian, where x and y are the Cartesian reciprocal coordinates of the beam at the film

$$b(x, y) \sim \exp\left(x^2/2\sigma_x^2 + y^2/2\sigma_y^2\right). \qquad (2)$$

In general, the optical density $D(r, \varphi)$, measured on the film from a partially oriented specimen, can be written as a sum of contributions from reflections $I(r, \varphi)$ and the background $B(r, \varphi)$ (Makowski, 1978):

$$D(r, \varphi) = I(r, \varphi) + B(r, \varphi) \qquad (3)$$

and where

$$I(r, \varphi) = \sum_i I_i(r_i, \varphi_i) * b(r, r_i; \varphi, \varphi_i) * f(\varphi, \varphi_i), \qquad (4)$$

r is the distance from the center of the diffraction pattern and φ is the angle about the center of the diffraction pattern. The sum is over all reflections on all layer lines and $*$ denotes convolution. The angular distribution function $f(\varphi, \varphi_i)$ and the beam function $b(r, r_i; \varphi, \varphi_i)$ describe the spreading of the reflections on the layer line. The intensities of the reflections $I_i(r_i, \varphi_i)$ can be calculated from the measured optical densities $D(r, \varphi)$ by deconvolution if the background $B(r, \varphi)$ is known.

3.2. *Background estimate and simulation of diffraction patterns*

The background, $B(r, \varphi)$, in fiber diffraction patterns can be represented as a sum of low-frequency sinusoids separable at least partially from the reflections, which consist primarily of higher frequencies. The reflections also contain low-frequency terms and complete separation can only be implemented by simultaneously fitting the background to a gradually varying function and the data to Gaussians. An estimate of the background, $B(r, \varphi)$, may be made by applying a low-pass filter to the raw data, which separates the lower frequencies of the background from the higher frequencies of the diffracted intensities. The band width of the filter (which should correspond to the range of frequencies contained in the background) was specific for each data set and was empirically determined by examination of the calculated contribution from the layer lines and the calculated background, together with the raw data set.

The background is initially calculated by setting all the intensities along the layer lines to be zero and applying a low-pass filter to the diffraction pattern in polar coordinates. The application of a low-pass filter is carried out by convolution of the data $D(r, \varphi)$ with a box-car function box(r, φ). The box-car function that we used was equal to unity for a specified angular and radial range and zero outside those limits:

$$B^1_{\rm av}(r, \varphi) = D(r, \varphi) * \mathrm{box}(r, \varphi). \qquad (5)$$

The result of this filtering is an overestimated background, which includes some density from reflections, but is adequate to use in calculating the first estimate of the layer-line intensities. The difference between the optical density, $D(r, \varphi)$, and the background estimate, $B^1_{\rm av}(r, \varphi)$ [see (3)], is the estimate of the intensity of the reflections falling on the layer line convoluted with the beam function and the angular disorientation function. From this estimate, the layer-line intensities are calculated by the deconvolution of this residual for beam size and disorientation. This is equivalent to fitting the data to Gaussian functions, which approximate the effects of the beam size and angular disorientation in both radial and azimuthal directions. The layer-line intensities resulting from this calculation are underestimated due to the initial overestimation of the background. Still,

these calculated intensities can be used to obtain the next (*e.g.* second) estimate of the background. This new estimate of the background is obtained by subtracting the calculated intensities from data followed by another convolution with the box-car function:

$$B^j(r,\varphi) = \left[D(r,\varphi) - I^{(j-1)}(r,\varphi)\right] * \mathrm{box}(r,\varphi), \quad (6)$$

where the intensities $I^{(j-1)}(r,\varphi)$ are as defined by (4) and were calculated at the previous iterative $(j-1)$ step. However, there are still residual intensities to be extracted in the new background estimate. These residual intensities are again fitted Gaussians so that new residual layer-line intensities may be calculated, which can be used to further improve the background estimate.

Residual images $\Delta(r,\varphi)$, calculated as the difference between the raw data and the simulated data, provide a means of gauging the correctness of the chosen parameters for the background function:

$$\Delta^j(r,\varphi) = D(r,\varphi) - D^j_{\mathrm{sim}}(r,\varphi), \quad (7)$$

where $D^j_{\mathrm{sim}}(r,\varphi)$ is calculated as a sum of the calculated intensities and the calculated background for each iteration,

$$D^j_{\mathrm{sim}}(r,\varphi) = I^j(r,\varphi) + B^j(r,\varphi). \quad (8)$$

Difference images calculated after processing the data file in Fig. 1 are shown in Figs. 5(*a*) and (*b*) and their use is discussed below.

3.3. *Iterative procedure for calculation of layer-line intensities and background estimate*

The process of calculating the intensities along the layer lines and the recursive estimate of the background is outlined in Fig. 2. In the initial step, the background estimate is made by setting all the layer-line intensities equal to zero. The data are low-pass filtered and an overestimate of the background results, see (5). The difference between the raw data and the estimated background is then used as the estimate of the scattered intensities due to the layer lines. The layer-line intensities are then calculated by angular deconvolution of this residual intensity data. In the first iteration, these intensities are underestimated.

The next iterative cycle begins by calculation of the simulated diffraction pattern $D_{\mathrm{sim}}(r,\varphi)$ using the current estimate of layer-line intensities convoluted by the angular spread and beam functions. A new background estimate is calculated as the difference between the raw data and the simulated diffraction pattern, see (6). There are still intensity data in this newly calculated background estimate, which is again subjected to low-pass filtering to remove more of the intensity features due to layer-line scattering. A residual $\Delta(r,\varphi)$ is calculated from the difference between the raw data and simulated pattern (background estimate plus calculated intensities for the current iteration) and this residual is used to calculate the intensities that are then added to the calculated layer-line intensities from the previous iterative step. This process converges when the difference between the raw data and the simulated diffraction

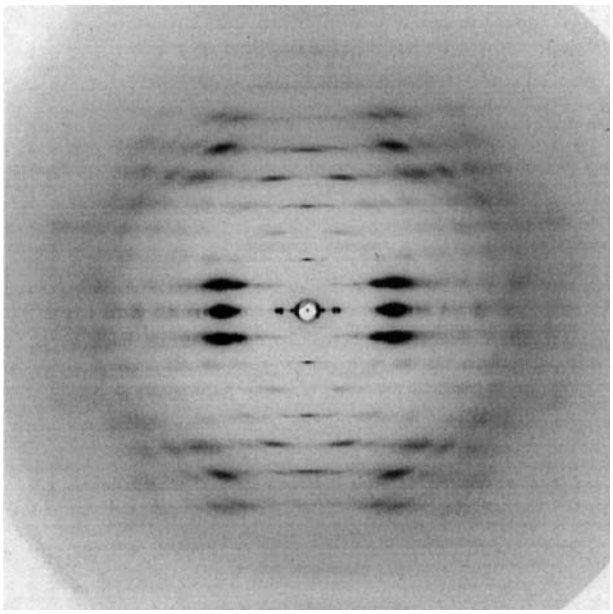


Fig. 1. Diffraction pattern of fddx at pH = 8.0. The equator is the 0th layer line, which runs horizontally though the center of the pattern; the meridian is the vertical through the center.

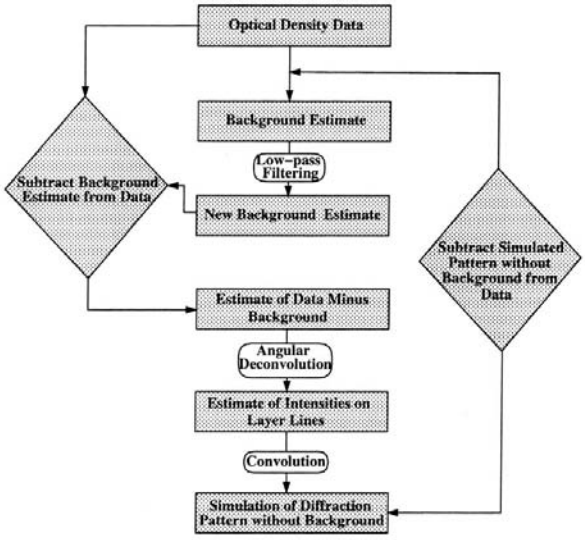


Fig. 2. Flow chart of the background correction algorithm. In the initial step, all layer-line intensities are set equal to zero, resulting in a blank simulated diffraction pattern and, consequently, the use of the measured data as the first estimate of the background.

data (calculated intensities plus the estimated background) is less than some small specified value.

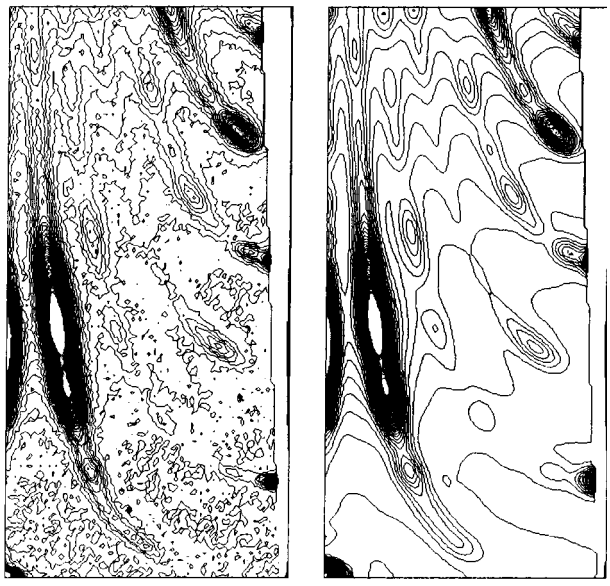
4. Results

We applied our background correction method to an X-ray diffraction pattern of a mutant phage (fddx), which has five amino acids inserted at the amino terminus (Petrenko *et al.*, 1996) and is shown in Fig. 1. The distance between layer lines in this diffraction pattern is approximately $\frac{1}{32}$ Å^{-1}, which indicates that the insert does not change the axial repeat of the major coat protein of the native phage.

Fig. 3(*a*) shows contour maps of the diffraction pattern of fddx in polar coordinates in reciprocal space. These are compared to the simulated pattern (Fig. 3*b*) calculated by using angular deconvolution with the background-correction algorithm. The patterns (data and calculated) are nearly identical, with the greatest deviation being due to noise in the data. The calculated layer-line intensities are shown in Fig. 4.

The difference files between the data and the simulated sets (Figs. 5*a* and *b*) include high-frequency noise components present in the original data but not in the pattern simulated by a slowly varying background plus Gaussian layer lines. In addition, there are difference features located near the second layer-line meridional reflection. This is due to a breakdown in the approximation used to calculate the disorientation function $f(\varphi, \varphi_i)$ at the meridian (Holmes & Barrington Leigh, 1974) and to the approximation of the beam shape used. Additional residual intensities on the equator are in the vicinity of strong crystalline reflections caused by diffraction from the pseudocrystalline lattice arrangement of the phage particles. These are also due to the approximate nature of the beam profile. Estimated errors for layer-line intensities in the immediate vicinity of these residuals are slightly higher than those for the remainder of the pattern.

The estimated background for this pattern (Fig. 6*a*) clearly contains a residual contribution from the layer-line intensities at the position of the strong 10 Å peak on the first layer line. This appears to indicate that the background was overestimated. However, we further analyzed this background assuming it is due to actual background plus a poorly oriented subpopulation of particles with a disorientation angle σ_φ equal to 4.5° (more than twice that of the one used in the initial layer-line calculations). The background estimate from this second layer-line extraction is shown in Fig. 6(*b*) and still contains a small residual contribution from layer-line intensities. We conclude that the first background estimate included intensity from a subpopulation of particles with orientations much worse than most of the specimen and that, once the contribution from these particles is removed, a weak residual from a very poorly oriented subpopulation of particles still remains. This low-frequency contribution from poorly oriented diffracting particles can be treated as background since a complete extraction of the intensity corresponding to



(*a*) (*b*)

Fig. 3. Plots of the (*a*) scattered and (*b*) calculated intensities as a function of the radius (vertical) and the angle (horizontal) in reciprocal space. The left sides of these plots correspond to the equator and the right ones to the meridian. The layer lines start at the right and curve up to the left. (*b*) gives the simulated data set calculated by adding the calculated intensities and background. The simulated data have reproduced all the features from the recorded data except the high-frequency noise.

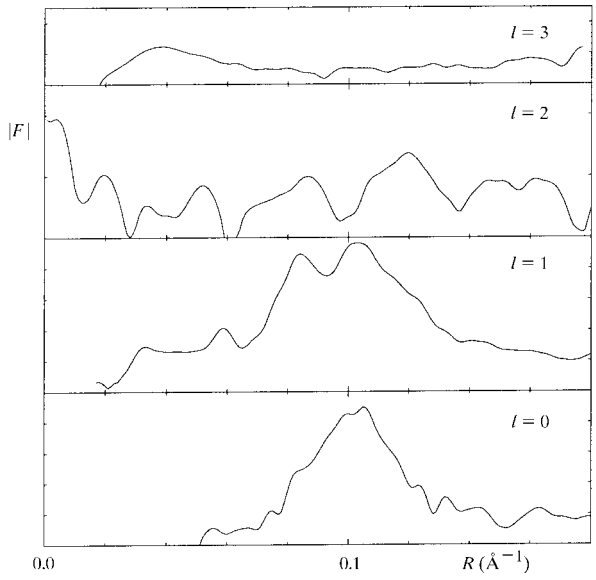
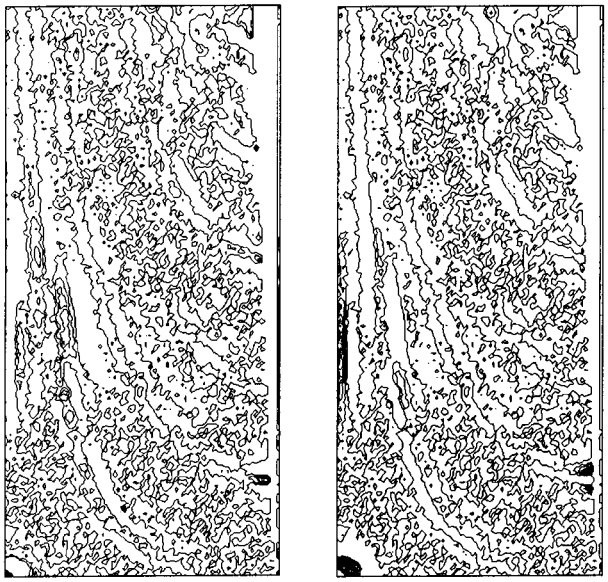


Fig. 4. Layer lines extracted from mutant phage fddx, whose diffraction pattern is shown in Fig. 1.

the majority of highly oriented diffracting particles has been made. The calculated intensity along layer lines due to the poorly oriented ($\sigma \sim 4.5$) subpopulation is consistent with the intensity due to the more highly oriented population but it is substantially noisier. Use of a disorientation function consisting of the sum of several Gaussians can also utilize diffraction from these subpopulations of particles while making the estimate of layer-line intensities. This multi-Gaussian approach would give similar results to those described in this paper for diffraction patterns with a disorientation function not well approximated by a single Gaussian.

The described procedures for background correction have been applied to numerous diffraction patterns similar to those shown in Fig. 1. Second background extraction was also applied to all of the initially obtained background estimates using a disorientation angle σ_φ twice that of the first extraction. In every case, our results suggest that we are able to separate the diffraction from particle populations having different degrees of orientation. The different degrees of orientation can be explained by the presence of broken particles in the specimen, uneven drying of different regions of the fiber, or surface effects during the orientation of the specimen.

In order to experimentally evaluate the computationally estimated background, a 'blank' (background image) was obtained. That is, a diffraction pattern was recorded by exposing an empty part of the glass capillary to X-rays under the same conditions used for collection of the fiber pattern. The resulting background pattern was transformed to polar coordinates. It was then scaled to the fiber diffraction data from the virus specimen and subtracted from that pattern. Angular deconvolution was then used to calculate the intensities along the layer lines assuming the blank to be a perfect estimate for the background (zero residual background). The results were compared to the calculated intensities obtained using the background-estimation procedure described here. No qualitative difference between the layer-line intensities estimated in the two ways was observed. However, the one obtained using the recorded background pattern has a higher residual, probably due to errors intrinsic to using a 'blank' pattern to estimate background and to variations in the film fog.



(*a*) (*b*)

Fig. 5. Difference patterns calculated by subtracting the simulated data from the original data (*a*) and by subtracting the original data from the simulated data (*b*). Plot (*a*) gives the part of the raw diffraction data that has not been accounted for by the simulated pattern and plot (*b*) gives the part of the raw data that has been overestimated in the simulated pattern. These residual patterns contain mostly high-frequency components, which have been removed from the original data. There are still intensity features appearing where the equatorial (caused by diffraction from the pseudocrystalline lattice arrangement of the phage particles) and meridional (second layer line) reflections are located. The approximation of the angular disorientation function to a Gaussian breaks down at these regions and, consequently, a complete calculation of the intensities in close vicinity to these regions cannot be achieved.

5. Discussion

Most fiber diffraction processing methods estimate the background from sampled points measured between the layer lines. This method works for well oriented specimens that have well defined gaps between the layer

(*a*) (*b*)

Fig. 6. (*a*) The background calculated by using our background algorithm. There is substantial layer-line information appearing in this background estimate, which suggests that the background has been overestimated. (*b*) The new background estimate recalculated by applying the background procedure onto the calculated background shown in (*a*), but using a doubled spread angle σ_φ.

lines. In practice, only a few macromolecular assemblies can be oriented to a high degree and in every case angular disorientation will cause the layer lines to overlap at sufficiently high resolution, with the limiting resolution depending on the angular spread σ_φ and the repeat distance *c*. Larger repeat distances, *c*, place the layer lines in reciprocal space closer, causing them to spread into each other at lower resolution. Thus, the effects of angular disorientation, beam size and symmetry make the estimate of the background from sampled points biased and resolution limited.

An alternative method is to use a properly scaled experimental estimate of the background. In many cases, however, the recorded background pattern does not adequately simulate the background scatter, owing to differences in scatter from the sample matrix, film and so on. Furthermore, they can provide no information about diffraction from poorly oriented or damaged particles present in the specimen. Consequently, a computationally estimated background combined with deconvolution procedures to calculate the intensities along the layer lines is potentially more accurate than an experimentally derived background.

The method described here also makes possible processing of data from less well oriented specimens and maximizes the resolution to which data may be estimated. Using low-pass filtering, we were able to demonstrate that the background can be estimated even when portions of it come from diffraction from the specimen itself. The parameters used to estimate the background were chosen in such a way as to give the smallest difference between the simulated data set (calculated as a sum of the background estimate and the convolution of the calculated layer-line intensities with the beam and disorientation functions) and the real data set. The intensities along the layer lines were calculated by an iterative method employing the weighted optical densities at points in the neighborhood of a reflection. Thus, we were able to separate (deconvolve) the intensities at different places, where the layer lines were overlapping owing to angular disorientation and large beam size.

These deconvolution procedures allow the analysis of fiber diffraction patterns to extend to the highest possible resolution limits as determined by the intrinsic properties of the particles, such as diameter, symmetry and orientation of the particles in the specimen. Removal of high-frequency noise and the resulting smoothing of intensities along layer lines also influences the process of the structure refinement by making possible a better fit between the calculated intensities from the model structures to the scattered amplitudes extracted from the data.

We are grateful to Valery Petrenko, who sent us the mutant phage fddx. We also thank Alla Korepanova and T. Somasundaram for discussions and help in sample preparation. Diffraction experiments at the Rigaku RU 200 station were carried out with the kind assistance of T. Somasundaram. This work is supported by a grant from the National Science Foundation.

References

Fraser, R. D. B., Macrae, T. P., Miller, A. & Rowlands, R. J. (1976). *J. Appl. Cryst.* **9**, 81–94.

Glucksman, M. J., Bhattacharjee, S. & Makowski, L. (1992). *J. Mol. Biol.* **226**, 455–470.

Holmes, K. C. & Barrington Leigh, J. (1974). *Acta Cryst.* A**30**, 635–638.

Lorenz, M. & Holmes, K. C. (1993). *J. Appl. Cryst.* **26**, 82–91.

Makowski, L. (1978). *J. Appl. Cryst.* **11**, 273–283.

Marvin, D. A., Bryan, R. K. & Nave, C. (1987). *J. Mol. Biol.* **193**, 315–343.

Millane, R. P. & Arnott, S. (1985). *J. Appl. Cryst.* **18**, 419–423.

Namba, K., Pattanayek, R. & Stubbs, G. (1989). *J. Mol. Biol.* **208**, 307–325.

Nave, C., Brown, R. S., Fowler, A. G., Lander, J. E., Marvin, D. A., Provencher, S. W., Tsugita, A., Armstrong, J. & Perham, R. N. (1981). *J. Mol. Biol.* **149**, 675–707.

Petrenko, V. A., Smith, G. P., Gong, X. & Quinn, T. (1996). *Protein Eng.* **9**, 797–801.

Wang, H. & Stubbs, G. (1994). *J. Mol. Biol.* **239**, 371–384.