Acta Cryst. (1996). A52, 797–811

A Bayesian Approach to Laue Diffraction Analysis and its Potential for Time-Resolved Protein Crystallography

GLEB P. BOURENKOV, ALEXANDER N. POPOV AND HANS D. BARTUNIK

Max-Planck Research Unit for Structural Molecular Biology, Protein Dynamics Group, MPG-ASMB, c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

(Received 26 January 1996; accepted 14 May 1996)

Abstract

A solution to the energy-overlap problem in Laue diffraction is described that does not require redundancy in the measurements. The new method follows a Bayesian approach with multidimensional probability density functions. The only assumption made is the validity of Wilson statistics. The intensity components of reflection multiplets are deconvoluted and estimates of their precision are obtained. The Laue patterns are processed to their physically relevant wavelength-dependent resolution limit; no 'soft parameters' are involved. The Bayesian method may also be applied to deconvoluting spatial overlaps. The power of the method is demonstrated by a test application to bovine trypsin. The completeness at low and medium resolution as well as at very high resolution (1.4 Å) is enhanced very substantially as compared with standard procedures; the 'low-resolution hole' problem is solved. As a consequence, the contrast in electron-density maps improves so far that they become comparable in quality with maps from monochromatic data at high resolution. The new method is of interest for all types of Laue diffraction experiments, in particular for single-shot time-resolved studies on short time scales. Simulation calculations for single-shot Laue conditions and for the disorder–order transition in trypsinogen as a model system demonstrate the potential power of applications in protein crystallography that combine high resolution and Bayesian processing.

1. Introduction

Laue diffraction from protein crystals permits simultaneous exploration of a large volume in reciprocal space (Hohlwein & Mason, 1981; Moffat, Szebenyi & Bilderback, 1984). Under favourable conditions characterized by a broad incident X-ray wavelength distribution (*e.g.* 0.5–2 Å), a high-symmetry space group and high crystalline order, the simultaneously excited reflection spots may yield structure-factor amplitudes for a high percentage of all possible independent reflections (Cruickshank, Helliwell & Moffat, 1987). Owing to the high spectral brightness in the hard X-ray wavelength range of a number of synchrotron-radiation sources, including in particular wigglers with short critical wavelengths, such a Laue diffraction pattern may be recorded within very short exposure times. This provides a potential for time-resolved investigation of protein molecules in structural states with lifetimes in the μs–s regime (Hajdu *et al.*, 1987). Even single-bunch exposures with an intrinsic time resolution < 100 ps are feasible (Szebenyi *et al.*, 1992).

Extraction of crystallographic information from Laue diffraction patterns is complicated by a number of fundamental and practical problems. This explains why Laue techniques until now have found only limited application to protein structure analysis (Hajdu & Andersson, 1993). A number of major problems are related to the energetic and spatial overlapping of reflection spots in Laue patterns. The spatial overlap problem may be solved by fitting analytical reflection profiles provided that the crystal mosaic spread is not too broad (Ren & Moffat, 1995*a*). However, the problem of harmonic overlaps is fundamental. A considerable percentage of all measurable reflection spots in Laue diffraction patterns are multiplets containing components from several different wavelengths. If the multiplets are not adequately deconvoluted into component intensities, the completeness in the diffraction data is substantially reduced, in particular at low to medium resolution. This 'low-resolution hole' causes errors in Fourier calculations, hence low contrast and quality of electron-density maps (Bartunik, Bartsch & Huang, 1992; Andersson *et al.*, 1992). A detailed discussion of such effects in terms of real-space point spread functions was given by Duke *et al.* (1992).

The truncation of low-reflection orders causes different density distributions as compared with density-sharpening operations, which involve deliberate down-weighting of low-resolution reflections. This is illustrated in Fig. 1 for the example of a one-dimensional structural model containing widely separated density peaks. Unlike sharpening, the low-frequency pertubation due to truncation of reflection orders at low resolution clearly reduces the contrast in the density distribution.

Maps that are affected by the low-resolution hole may be improved by including a small number of model structure-factor amplitudes (Bartunik, Bartsch & Huang, 1992). Such a procedure, however, would introduce a bias in the interpretation of the maps, which is not acceptable considering the aim of time-resolved Laue diffraction studies is to detect unknown conformational changes. In the specific case of a structure with high non-crystallographic symmetry, Hol and co-workers (Vellieux *et al.*, 1993) were able to obtain good electron-density maps through sixfold density averaging, despite low completeness (37%) in their Laue data. In general, however, without appropriate deconvolution of overlaps, Fourier maps calculated with Laue structure factors necessarily must be of poor quality.

A number of authors previously developed techniques aiming to reduce or solve the energy-overlap problem. Helliwell *et al.* (1989) suggested a method based on the use of a stack of photographic films and the dependence of the absorption in each film on the X-ray wavelength; the power of the method was limited by the poor energy resolution of photographic film (Wakatsuki, 1993). Another approach based on direct methods was proposed by Hao, Campbell, Harding & Helliwell (1993); the authors demonstrated the applicability to small-molecule structures. A closely related method involving a Patterson-function modification (squaring) was suggested for application to protein structures (Hao, Harding & Campbell, 1995*a*). This method produced roughly correct (*i.e.* better than random) estimates of multiplet component intensities; no measure of the uncertainty of the estimates could be obtained. Substantially more accurate estimates may be obtained from least-squares deconvolution of harmonics based on the measurement of symmetry-related reflection multiplets (Helliwell, 1992; Wakatsuki, 1993; Campbell & Hao, 1993; Bradbrook *et al.*, 1995; Ren & Moffat, 1995*b*). However, this method requires substantial redundancy in the observations. Furthermore, they have to be located in suitable ranges of the wavelength normalization curve. For example, if multiplet observations belong to an approximately linear region in the distribution of the normalization factor *versus* wavelength, the resulting least-squares systems tend to degenerate. The need for high redundancy in practice makes it necessary to collect Laue data at additional crystal settings. This excludes single-shot experiments and thus represents a severe drawback in all applications involving short time scales.

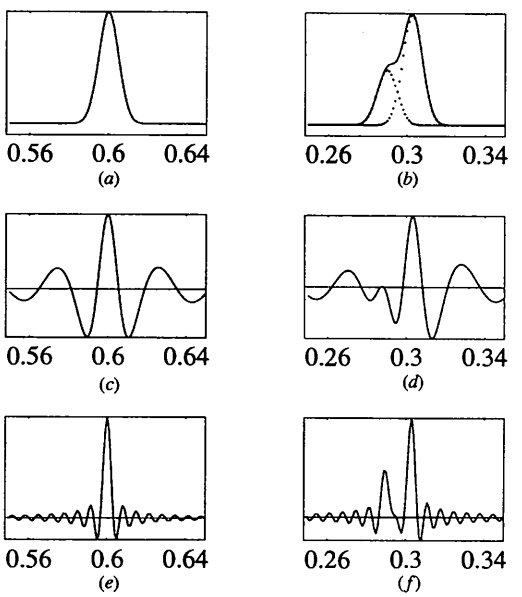


Fig. 1. Comparison of the effects of sharpening and low-order truncation on the density distribution for the example of a one-dimensional structure model. Original periodic density segments containing a single Gaussian peak (*a*) or a peak consisting of two overlapping Gaussians (*b*). The variance of the Gaussians is $u = 1/200$ of the period. The diffraction pattern is calculated for the reflection orders $h = 0$ to $h = 150$. (*c*),(*d*) Reconstructed density after omitting reflection orders $h = 1$ to $h = 30$. (*e*),(*f*) Sharpened density, obtained by weighting the structure factors with the function $\exp[2(\pi h u)^2]$.

All previously proposed methods of deconvoluting harmonics have in common that no solid justification is provided for treating a given reflection spot as single or multiple. In each of these procedures, the decision is made assuming a wavelength-independent diffraction limit as a 'soft parameter' that is empirically chosen to best predict spots that are visible in the Laue pattern (*e.g.* Hao, Harding & Campbell, 1995*b*). This is adequate in the case of monochromatic diffraction patterns. In Laue diffraction, however, this way of defining a diffraction limit has no physical justification. In fact, the incident X-ray intensity, the diffracting power of the crystal and other experimental factors vary with the wavelength. Therefore, the average signal-to-noise ratio for each resolution shell and thus the apparent diffraction limit clearly must be wavelength dependent. This was first demonstrated when analysing Laue exposures from yeast hexokinase P2 (Bartsch *et al.*, 1990; Rupp, 1991). In this study, the apparent diffraction limit was determined by comparing observed to predicted patterns and found to vary between 2.5 Å (for $\lambda = 0.95$ Å) and 3.2 Å ($\lambda = 0.65$ Å); the wavelength dependence was in good agreement with a theoretical estimate taking the signal-to-noise ratio into account. Neglecting the wavelength dependence and assuming a universal soft diffraction limit leads to systematical under- or overprediction of the Laue pattern. Incorrect assignment of the components of a given reflection, for example omission of some higher harmonics, introduces a bias; it affects both the completeness and the accuracy of data in all resolution shells.

We have developed a fundamentally new method of deconvoluting energy (and spatial) overlaps in Laue patterns from protein structures. The method provides a

complete and justified probabilistic solution. For each unique reflection that may be excited under the given experimental conditions defined by the wavelength bandpass and the geometries of the diffraction set-up and the detector, the best estimates of the structure-factor modulus, its square and their standard uncertainties are obtained. The results are independent of the multiplicity of the reflection; redundancy in the observation is not required. No soft parameters are involved. The only assumption made is the validity of Wilson statistics (Wilson, 1949). The new method follows a Bayesian statistical approach. The present paper describes the method together with the results of a test application to Laue data from orthorhombic trypsin which extend to the highest resolution, 1.4 Å, ever reached up to now in protein Laue diffraction.

2. Outline of the Bayesian method

2.1. *Bayesian analysis of monochromatic X-ray diffraction data*

The use of probability theory following a Bayesian approach in the analysis of X-ray diffraction data was first proposed by Kheiker & Nekrasov (1970) for the case of monochromatic single-crystal diffraction. A similar concept formed the basis of the program *TRUNCATE* (French & Wilson, 1978) in the *CCP*4 program suite. These developments aimed to find the best estimates of the structure-factor amplitude F and its square I from the intensity observation J, when J is small compared with the noise or even negative. *A priori* given information about I being positive and obeying Wilson's distributions was taken into account through probability density functions (PDFs),

$$p(X) \propto \exp(-X/\langle I\rangle) \tag{1a}$$

$$p(X) \propto X^{-1/2}\exp(-X/2\langle I\rangle). \tag{1b}$$

Expression (1*a*) corresponds to acentric reflections (having phase φ, $0 \le \varphi \le 2\pi$), (1*b*) to centric reflections ($\varphi = n\pi/2, n = 0, 1, 2, 3$). The symbol $\propto$ denotes 'follows'. X is an I-associated random variable and $\langle I\rangle$ is a mean intensity at given resolution.

Experimental observation of the reflection J provides parameters (J, σ_J) of the normal probability density $p(J|X)$ for the intensity. This information must be associated with the prior PDF $p(X)$ through the Bayes theorem, which states

$$p(X|J) \propto p(X)p(J|X) \tag{2}$$

and can be read as follows: the posterior PDF, *i.e.* probability of the observation J due to elementary hypothesis X, follows the probability of the hypothesis multiplied with the probability of the hypothesis to produce the observation. The best estimates for F, I and their uncertainties are given by the moments of the resulting posterior PDF.

2.2. *The Bayesian approach for the case of Laue diffraction*

The previously developed Bayesian method (Kheiker & Nekrasov, 1970) is limited to non-overlapping reflections and independent observations. In Laue diffraction, an identical approach may be applied to non-overlapping single reflections only. Analysis of Laue reflection multiplets, however, requires a Bayesian approach that deals with *multidimensional* PDFs.

An n-multiple Laue reflection intensity is given by the experimental observation in the form of

$$J = \mathbf{aX}. \tag{3}$$

Here, $\mathbf{a} = \{a_i\}$ is a vector whose elements contain products of the Lorentz-polarization and wavelength normalization factors for the nX_i components of the variable vector $\mathbf{X}$. If several exposures are made, the products further contain overall scale and temperature factors from frame-to-frame scaling. The components of $\mathbf{a}$ are always positive. We assume that the uncertainties of a_i are negligible compared with the observation uncertainty σ_J (this is discussed below).

With the assumption of a normal distribution of J, the observation of a reflection multiplet yields a degenerate multivariate normal PDF over the n-dimensional space of $\mathbf{X}$:

$$p(J|\mathbf{X}) \propto \exp[-(\mathbf{aX} - J)^2/2\sigma^2]. \tag{4}$$

Fig. 2(*a*) shows an example of such a function in the case of two variables. If a set of $m \ge 1$ redundant observations $\mathbf{J} = \{J_1, \ldots, J_m\}$ is present, insertion of their product in the Bayes formula will produce a more generalized form of the normal PDF (4):

$$p(\mathbf{J}|\mathbf{X}) = \prod_{i=1}^{m} p(J_i|\mathbf{X}) \propto \exp[-(\mathbf{X} - \mathbf{R})^T\mathbf{C}(\mathbf{X} - \mathbf{R})/2]. \tag{5}$$

The expression contains the normal $n \times n$ matrix $\mathbf{C} = \mathbf{A}^T\mathbf{WA}$. The rows of the matrix $\mathbf{A}$ are formed by $\mathbf{a}_j$ $(j = 1, \ldots, m)$. $\mathbf{W}$ is a diagonal matrix with elements $W_{jj} = \sigma_{Jj}^{-2}$. The vector $\mathbf{R}$ may be found as any vector that satisfies the equation

$$\mathbf{CX} = \mathbf{A}^T\mathbf{WJ}. \tag{6}$$

Owing to the positivity of the components of $\mathbf{X}$, the resultant PDF is a truncated multivariate normal over the positive hyperquadrant of the $\mathbf{X}$ space. This PDF is never degenerate, even if the $\mathbf{C}$ matrix is. In a next step, we associate this PDF through the Bayes theorem with an appropriate Wilson prior for the components and obtain one of the following posterior PDFs on the positive hyperquadrant:

$$p(\mathbf{X}|\mathbf{J}) \propto \exp[-\mathbf{hx} - (\mathbf{X} - \mathbf{R})^T\mathbf{C}(\mathbf{X} - \mathbf{R})/2] \tag{7}$$

for a set of acentric reflections, $h_i = 1/\langle I\rangle_i$, and

$$p(\mathbf{X}|\mathbf{J}) \propto \prod_{i=1}^{n} X_i^{-1/2} \exp[-\mathbf{hx} - (\mathbf{X}-\mathbf{R})^T \mathbf{C}(\mathbf{X}-\mathbf{R})/2] \tag{8}$$

for a set of centric reflections, $h_i = 1/2\langle I\rangle_i$. Figs. 2(*b*) and (*c*) show $p(\mathbf{X}|\mathbf{J})$ for the example of two variables. Within a given Laue multiplet, the reflection components are either all centric or all acentric.

When the n-dimensional PDF is found, the resulting posterior probability density for the component X_k is the projection of $p(\mathbf{X}|\mathbf{J})$ onto the $(p, 0, X_k)$ plane:

$$p(X_k|\mathbf{J}) \propto \int_0^\infty \ldots \int_0^\infty \int_0^\infty p(\mathbf{X}|\mathbf{J})\, dX_1 \ldots dX_{k-1} dX_{k+1} \ldots dX_n \tag{9}$$

(Figs. 2*d*,*e*). The best estimates for the structure-factor amplitude $(F_o)_k$, it's square $(I_o)_k$ and their standard uncertainties are given by the moments of the posterior PDF (9):

$$(F_o)_k = \int_0^\infty X_k^{1/2} p(X_k|\mathbf{J})\, dX_k \tag{10}$$

$$(I_o)_k = \int_0^\infty X_k p(X_k|\mathbf{J})\, dX_k \tag{11}$$

$$\sigma^2_{(F_o)_k} = (I_o)_k - (F_o)_k^2 \tag{12}$$

$$\sigma^2_{(I_o)_k} = \int_0^\infty X_k^2 p(X_k|\mathbf{J})\, dX_k - (I_o)_k^2. \tag{13}$$

2.3. *Example illustrating Bayesian deconvolution*

Let us consider the example of an acentric reflection with multiplicity 2 and only one observation. This situation refers to a great part of the Laue data in an experiment aimed to obtain maximum crystallographic information from a minimum number of crystal settings. The PDF (9) in this case has the form (Fig. 2*d*)

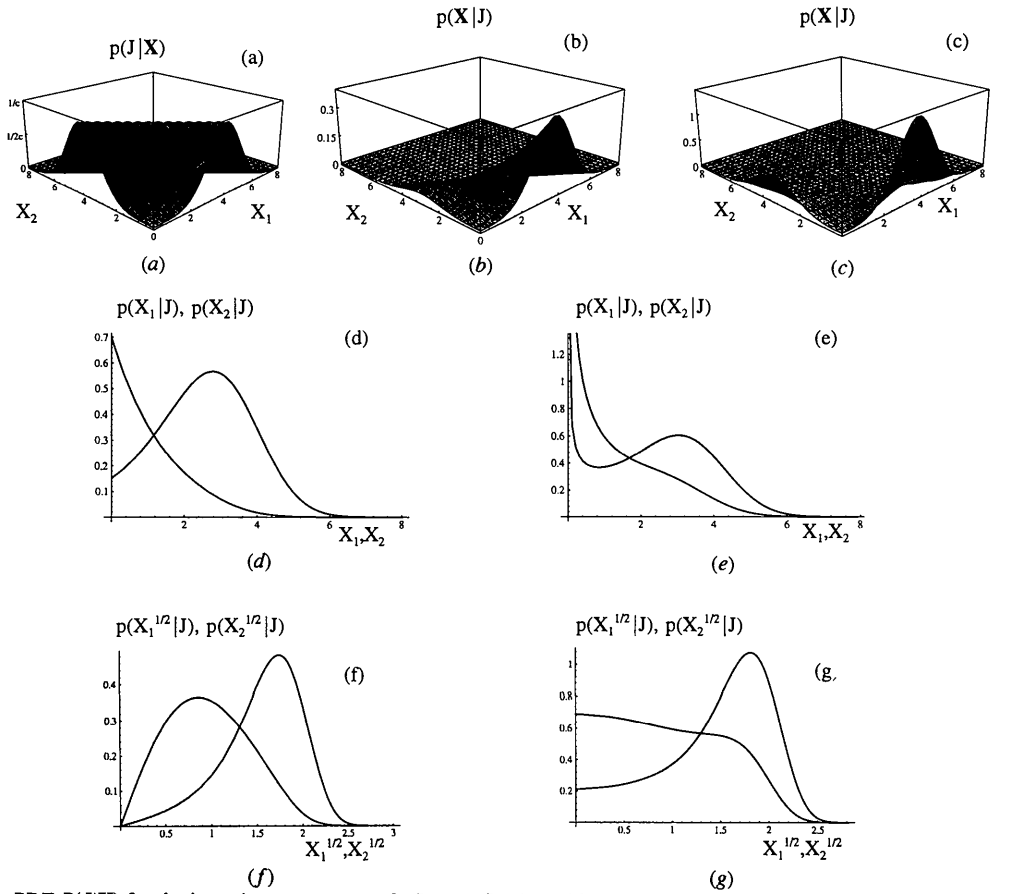


Fig. 2. (*a*) Joint PDF $P(J|\mathbf{X})$ for the intensity components of a Laue reflection with multiplicity 2 obtained by a single hypothetical observation with intensity $J = 5$, $\sigma_J = 1$. (*b*),(*c*) Joint posterior PDFs with (*b*) acentric or (*c*) centric Wilson priors $\langle I\rangle_1 = 3$, $\langle I\rangle_2 = 1$. (*d*),(*e*) Marginal posterior PDFs for intensities in the (*d*) acentric or (*e*) centric case. (*f*),(*g*) Marginal posterior PDFs for amplitudes in the (*f*) centric or (*g*) acentric case.

$$p(X_1) \propto \exp\{[1/(\langle I\rangle_1 a_1) - 1/(\langle I\rangle_2 a_2)]a_1 X_1/\sigma_J\} \times (1 + \mathrm{erf}\{[J - (\sigma_J^2/\langle I\rangle_2 a_2) - X_1 a_1]/2^{1/2}\sigma_J\}). \quad (14)$$

The estimates (11) and (13) yield exact analytical expressions:

$$(I_o)_1 = (\sigma_J/a_1)\left[\Delta_{12} + \frac{(1+\xi_1\chi_1)}{(\chi_1 - \chi_2)}\right], \quad (15)$$

$$\sigma^2_{(I_o)_1} = (\sigma_J^2/a_1^2)\{2\Delta_{12}^2 + [2(1+\xi_1\chi_1)\Delta_{12} + (1+\xi_1^2)\chi_1 - \xi_1\chi_1](\chi_1 - \chi_2)^{-1}\} - I^2_{o_1}, \quad (16)$$

where

$$\xi_i = J/\sigma_J - \sigma_J/\langle I\rangle_i a_i,$$
$$\chi_i = (\pi/2)^{1/2}\exp(\xi_i^2/2)\{1 + \mathrm{erf}(\xi_i/2^{1/2})\},$$
$$\Delta_{ij} = (1/\langle I\rangle_i a_i - 1/\langle I\rangle_j a_j)^{-1}.$$

The correlation coefficient is

$$c_{12} = \left((\sigma_J^2/a_1^2a_2^2)\{2\Delta_{12}^2 + [(2 + \xi_1\chi_1 + \xi_2\chi_2)\Delta_{12}] \times (\chi_1 - \chi_2)^{-1}\} - I_{o_1}I_{o_2}\right)(\sigma_{I_{o_1}}\sigma_{I_{o_2}})^{-1/2}. \quad (17)$$

In the special case $\langle I\rangle_1 a_1 = \langle I\rangle_2 a_2$ ($\xi_1 = \xi_2 = \xi$, $\chi_1 = \chi_2 = \chi$), the formulae (15)–(17) transform in the following way:

$$(I_o)_i = (\sigma_J/a_1)[(1+\xi^2)\chi + \xi]/[2(1+\xi\chi)], \quad (18)$$

$$\sigma^2_{(I_o)_i} = (\sigma_J^2/a_1^2)[2 + \xi^2 + (3\xi + \xi^2)\chi]/[3(1+\xi\chi)] - I^2_{o_i}, \quad (19)$$

$$c_{12} = \{(\sigma_J^2/a_1a_2)[2 + \xi^2 + (3\xi + \xi^2)\chi] \times [3(1+\xi\chi)]^{-1} - I_{o_1}I_{o_2}\}(\sigma_{I_{o_1}}\sigma_{I_{o_2}})^{-1/2}. \quad (20)$$

In order to illustrate the effect of the analysis, the posterior moments for a series of hypothetical observations are listed in Table 1. The values of J/σ_J range from −3 to 100, $\langle I\rangle_2 a_2$ from 1 to 20; $\langle I\rangle_1 a_1$ was chosen to be 20.

When the observed sum of component intensities is small, one notices as a rather trivial result that both terms are small and the estimates are practically uncorrelated. The case of $J/\sigma_J \geq 3$ is more interesting. For the first term, which *a priori* is supposed to be stronger, one gets useful estimates even when the ratio $\langle I\rangle_2 a_2/\langle I\rangle_1 a_1$ is as high as ~ 75%. However, this does not mean at all that the second term may be ignored, not even when $\langle I\rangle_2 a_2/\langle I\rangle_1 a_1$ is only ~ 5%, since this would cause a shift in the estimates by more than a standard deviation.

In most cases, the procedure affects both terms of a doublet. The posterior moments for the weaker term in general provide less accurate information, depending on the relative size of the components and on the signal-to-noise ratio. Only in the case when $\langle I\rangle_2 a_2 \ll \langle I\rangle_1 a_1$ and $J \gg \sigma_J$ is no information obtained about the second term (*i.e.* the posterior moments do not significantly differ from the prior ones). Of course, there is always a finite probability that the true value of intensity of the second term is even higher than the intensity of the first one. This is expressed by a joint posterior PDF, which results in a modified posterior expected value and an increase in the standard uncertainty.

When $\langle I\rangle_2 a_2$ approaches the value $\langle I\rangle_1 a_1$, *i.e.* when prior preference may not be given to any of the components, then evidently the probability for both components to have any intensity between 0 and J/a_1 will be uniform. As $J/\sigma_J \to \infty$, $I_{o_i} \to J/2a_i$ and $I_{o_i}/\sigma_{I_{o_i}} \to 3^{1/2}$. For the components of the real Laue doublet, a substantial difference between priors may always be expected owing to the nature of the wavelength normalization curve and the dependence of $\langle I\rangle$ on $\sin\theta/\lambda$. This forms the main difference of our approach to the Bayesian approach that Sivia & David (1994) proposed for an analysis of overlapping reflections in monochromatic powder diffraction patterns. In their method, there is no difference in the priors; thus, use of the uniform positive prior is permitted.

Table 1. *Posterior expected values, standard uncertainties and correlation coefficients for the intensities of components derived with Wilson's prior from a set of hypothetical observations of a double acentric Laue reflection with unit standard uncertainty*

The prior for the first component is assumed $\langle I\rangle_1 a_1 = 20$. In each cell, the moments are arranged in the scheme:

$a_1\langle I_o\rangle_1 \quad a_1\sigma_{a1\langle I_o\rangle 1}$
$a_2\langle I_o\rangle_2 \quad a_2\sigma_{a2\langle I_o\rangle 2}$
c_{12}

Observation J	Prior for the second component $\langle I\rangle_2 a_2$ = 1	5	10	15	20
−3	0.27 0.25 0.21 0.21 −0.048	0.26 0.25 0.25 0.24 −0.055	0.26 0.25 0.26 0.25 −0.056	0.26 0.25 0.26 0.25 −0.057	0.26 0.25 0.25 0.24 −0.057
0	0.66 0.54 0.43 0.39 −0.170	0.62 0.52 0.58 0.50 −0.204	0.62 0.52 0.60 0.50 −0.209	0.62 0.52 0.61 0.51 −0.211	0.62 0.52 0.62 0.62 −0.212
3	2.21 1.1 0.84 0.73 −0.521	1.75 1.1 1.47 1.0 −0.618	1.68 1.1 1.59 1.1 −0.626	1.66 1.1 1.62 1.1 −0.628	1.64 1.1 1.64 1.1 −0.630
6	4.93 1.38 1.02 1.0 −0.694	3.49 1.85 2.56 1.77 −0.852	3.21 1.86 2.89 1.83 −0.857	3.11 1.86 3.00 1.85 −0.859	3.06 1.86 3.06 1.86 −0.859
10	8.90 1.45 1.05 1.05 −0.732	6.21 2.83 3.78 2.75 −0.937	5.44 2.95 4.59 2.92 −0.943	5.16 2.96 4.88 2.95 −0.943	5.03 2.96 5.03 2.96 −0.943
20	18.9 1.45 1.05 1.05 −0.725	14.3 4.81 5.61 4.72 −0.937	11.6 5.67 8.35 5.64 −0.984	10.6 5.79 9.44 5.78 −0.985	10.0 5.8 10.0 5.8 −0.985
50	48.9 1.45 1.05 1.05 −0.725	43.3 6.64 6.54 6.56 −0.989	34.4 12.5 15.5 12.5 −0.997	28.4 14.2 21.6 14.2 −0.998	25.0 14.4 25.0 14.4 −0.998
70	68.9 1.45 1.05 1.05 −0.725	63.3 6.73 6.66 6.66 −0.989	52.1 15.6 17.8 15.6 −0.998	41.6 19.5 28.3 19.5 −0.999	35.0 20.2 35.0 20.0 −0.999
100	98.9 1.45 1.05 1.05 −0.725	93.3 6.74 6.67 6.67 −0.989	80.6 18.2 19.3 18.2 −0.998	63.3 27.0 36.7 27.0 −0.999	50.0 28.9 50.0 28.9 −0.999

2.4. *Relationship between least-squares and Bayesian solutions*

When redundant observations are available (*e.g.* at different crystal settings) and **C** is a well defined matrix, (6) has a unique solution **R**. The posterior PDF for acentric reflections (7) and the exponential term in (8) both convert to a multivariate normal through a shift of the centre to the point

$$\mathbf{R}' = \mathbf{R} - \mathbf{C}^{-1}\mathbf{h}. \qquad (21)$$

If all values $R_i'/(\mathbf{C}^{-1})_{ii}^{1/2} > 3$, then (10)–(13) can well be expressed by an expansion of the integral to the whole **X** space, and $(F_o)_k = R_k'^{1/2}$, $(I_o)_k = R_k'$, $\sigma_{(I_o)_k} = (\mathbf{C}^{-1})_{kk}^{1/2}$, $\sigma_{(F_o)_k} = (\mathbf{C}^{-1})_{kk}^{1/2}/2(F_o)_k$ represent fairly good approximations to the moments. In this case, our approach converges to the least-squares solution, which takes the components of **R** as the best estimates of intensities with the covariance matrix $\mathbf{C}^{-1}$ (if proper weights are used).

The use of the least-squares method becomes problematic not only when **C** is degenerate but also when some of the R_i (at least one) are negative. Application of the *TRUNCATE* procedure (French & Wilson, 1978) to the estimates that are extracted in such a way would be completely invalid. Wakatsuki (1993) and Campbell & Hao (1993) proposed to solve the problem by eliminating outlying terms. The non-linear solution of (6) for the $R_i^{1/2}$, as proposed by Ren & Moffat (1995*b*), yields identical results. From the point of view of the underlying PDFs, these methods use the coordinates for a point where the truncated PDF (5) adopts its maximum as estimates for the intensity and their square roots as estimates for the amplitude. Setting a negative observation to zero introduces a maximal bias to the data (French & Wilson, 1978), even in the case of independent measurements. In the case of a multiple reflection, negativity of R_i for a component does not necessarily imply that the corresponding term is negligible. Rather, it may indicate the existence of strong correlations between the variables; under such conditions, the least-squares system is poorly defined. The joint PDF in this case is similar to that in the case of a completely degenerate system that is combined with an acentric prior (Fig. 2*b*). In both cases, although the weaker term has a maximum-likelihood value at zero, the contribution of the corresponding component to the observed intensity is significant. Eliminating this component yields systematically overestimated values and underestimated standard uncertainties of the intensities of the other terms.

In the powder diffraction method of Sivia & David (1994), which only considers the amplitudes, the estimates were proposed to be taken as the maxima of the PDF (4) after transformation to the amplitude space. We stress the point that in general the best estimates are not provided by the maxima of the PDF but by its moments. For example, if we consider a Fourier transform of the resulting data, then its value at each point yields an expected value (approaching the maximum-likelihood value through the central limiting theorem) of the electron density only in the case when the first moments are taken as coefficients.

3. Parameters influencing the precision in Laue data processing

3.1. *Wavelength dependence in the apparent diffraction limit*

Protein Laue data evaluation by standard procedures involves an uncertainty in the choice of the diffraction limit. Up to now, it has never been reported that a protein Laue data set was processed to the actual geometrical diffraction limit $d_{\min} = \lambda_{\min}/(2\sin\theta_{\max})$. Rather, a 'soft' diffraction limit at poorer resolution was chosen from inspections of the diffraction pattern or its gnomonic projection. An analogous procedure was proposed by Hao, Harding & Campbell (1995*b*) for introducing a 'soft' minimum wavelength $\lambda_{\min}$, which may exceed the experimentally defined value. However, there is no physical justification for assuming a constant diffraction limit for a broad wavelength range if one considers the wavelength dependencies in the incident intensity, crystal reflectivity and hence the signal-to-noise ratio. The assumption of a constant diffraction limit causes under- or overprediction of the Laue patterns; thus, it may strongly affect the results of the evaluation (Bartunik, Bartsch & Huang, 1992).

Using our test data (see below for details), one can demonstrate the effect of a truncation introduced by choosing a wavelength-independent diffraction limit. Careful visual analysis of the patterns showed reflections up to *ca* 1.8 Å. We processed the data with the *Daresbury Laue Software Suite* (Campbell, 1993) assuming a constant $d_{\min}$ value of 1.8 Å and 1.3 Å, respectively. The results are presented in Fig. 3. One can clearly see from the processing to 1.8 Å that omission of 'invisible' high-resolution reflections yields data of lower accuracy. On the other hand, processing to 1.3 Å resolution with a standard (non-Bayesian) procedure results in a dramatically reduced completeness, since more reflections are designated as multiples that cannot be deconvoluted.

The wavelength dependence in the apparent resolution limit in Laue diffraction patterns was first investigated for the example of Laue exposures of hexokinase on photographic film (Bartsch *et al.*, 1990; Rupp, 1991). With the assumption of a smooth variation in $d_{\min}$ on λ, an empirical distribution $d_{\min}(\lambda)$ was derived from a comparison of predicted and observed diffraction patterns.

Our approach solves the problem in a most general way. The formulation involves the parameter $\langle I\rangle a/\sigma_I$, which in fact is a wavelength-dependent signal-to-noise

ratio. If the pattern has been evaluated using an obviously overestimated diffraction limit (*e.g.* the geometrical diffraction limit) and unshifted estimates taken from (10)–(13), not a single reflection will be lost due to overprediction. Furthermore, no bias is introduced in each particular observation. In the case when the contribution of a high-resolution component to a given multiplet intensity is negligible, then both Wilson's prior (1*a*) and (1*b*) approach the $\delta^+(0)$ function, and the posterior moments for the strongest term approach the expressions derived by Kheiker & Nekrasov (1970) for the case of a single reflection.

The estimates and their uncertainties will produce a modified signal-to-noise ratio for each reflection or for entire resolution shells. This ratio may be used in the same way as with monochromatic data for truncating data at an appropriate resolution or for excluding individual reflections.

3.2. *Importance of precise wavelength normalization*

All formulae derived in §2 are based on the assumption that the coefficients a_i are known accurately. Explicit consideration of errors in these coefficients leads to a highly multidimensional (n^2m instead of n) integral in the calculation of the moments. In order to minimize the effects of these errors and to improve the accuracy for the whole data set, the wavelength-normalization factors should be as precise as possible. When this is achieved, the remaining influence of the errors in a_i may be neglected. This is demonstrated by the example given below.

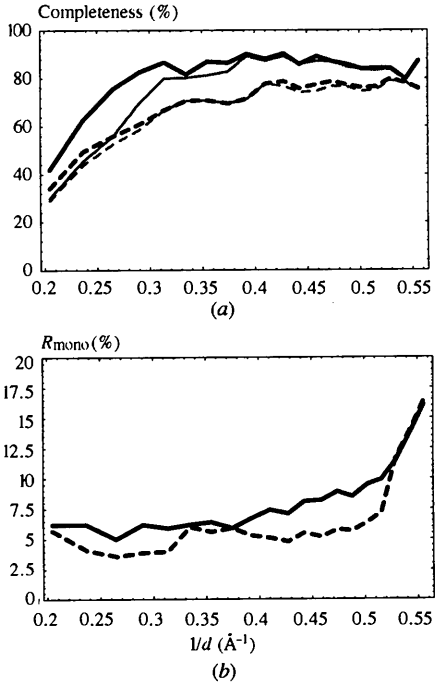


Fig. 3. (*a*) Completeness and (*b*) R_{mono} of BPT Laue data processed with the standard procedure. Wavelength-independent diffraction limits were assumed at 1.8 Å (solid lines) or 1.3 Å (dashed lines) resolution. The completeness for single reflections is indicated by grey lines. R_{mono}, *i.e.* the R factor of scaling to monochromatic data, is shown for singles only.

Previously developed wavelength-normalization techniques used a wavelength binning followed by power-series smoothing (Helliwell *et al.*, 1989) or modelling by Tchebychev polynomials (Smith Temple, 1989). However, if the wavelength normalization is based on measurements of symmetry-equivalent single reflections only, high redundancy in the measurements (*e.g.* at different crystal settings) is needed to achieve sufficient accuracy. Furthermore, there are difficulties in extending such normalization curves to the long-X-ray-wavelength range. If a reflection from a plane with spacing d is observed at $\lambda > 2\lambda_{min}$, then the reflection spot on the detector contains a further contribution corresponding to the plane spacing $d/2$ unless there is systematic extinction. The contribution of the second harmonic might be negligible and, thus, the reflection spot effectively single. However, this may only be verified if both the ratio of the mean intensities at d and $d/2$ and the ratio between the wavelength normalization factors at λ and $\lambda/2$ are known. Extrapolation of the polynomial to higher λ values is not justified since its coefficients are physically meaningless.

An alternative is to derive the wavelength-normalization curve from a reference crystal, not necessarily for the same structure, and to scale the Laue data to external monochromatic data. Multiple reference reflections may be included. We subdivide the wavelength range into a suitable number of bins. Reliable bin scale factors require a sufficient number (not less than say 40) reference data points in each bin. The bin scale factors g_i are obtained by least-squares solution of equations of the general form $J^{ref} = \sum I^{mono}{}_i g_i$. The reference patterns should be processed with an overestimated wavelength range; the effective limits λ_{min} and λ_{max} are defined on the basis of the resulting wavelength-normalization curve. Fig. 4 shows an example of a normalization curve that was derived in this way from measurements on beamline BW6 at DORIS. The discontinuities in the curve correspond to absorption edges of heavy atoms that are contained in the beam-defining components of the instrument (Au-coated planar and toroidal mirrors; Fe contamination of Be windows) and the imaging plate used as a detector (Ba, Br, Eu). Wavelengths in the close vicinity of such discontinuities were omitted during the binning procedure. Corrections for wavelength-dependent absorption by the sample crystal and its environment relative to the reference crystal and its environment are applied by fitting a single parameter. If the sample contains a resonant absorber with an absorption edge within the incident bandwidth, two (or more) parameters have to

be fitted for two (or more) different segments of the entire wavelength range.

4. Test application of the Bayesian method to Laue diffraction from trypsin at 1.4 Å resolution

4.1. *Detailed procedure of data processing*

The Bayesian method was tested using Laue data from benzamidine-inhibited β-trypsin from bovine pancreas (BPT) as a model structure. The crystal form is orthorhombic (space group $P2_12_12_1$) with cell dimensions $a = 54.8$, $b = 58.4$, $c = 67.8$ Å. The crystal structure was previously refined at 1.6 Å resolution (Marquart, Walter, Deisenhofer, Bode & Huber, 1983), and recently further refined at 1.0 Å resolution (Popov & Bartunik, 1996). Laue data were collected on the double-focusing wiggler beamline BW6 at DORIS using a MAR imaging plate detector (diameter 300 mm). Ten exposures were recorded from a single crystal at a temperature of 283 K. For the first exposure, the crystal was oriented with its *a* axis close to the beam direction and the *b* axis close to the spindle axis. Subsequent exposures were taken at 10° intervals in the spindle rotation angle. A wavelength-normalization curve was derived from a reference data set. The wavelength range was determined as $0.52 \leq \lambda \leq 2.1$ Å. The crystal mosaicity was *ca* 0.1°. The detected Laue reflection spots had a size $\simeq 400$ μm (at FWHM) in both the radial and tangential directions.

Processing of the data started by refining the crystal cell, orientation and detector setting parameters with *LAUEGEN* (Campbell, 1993). A final prediction was made assuming a geometrical diffraction limit $d_{\min} = 1.0$ Å. Spot integration was performed with the program *INTLAUE* (Shrive, Clifton, Hajdu & Greenhough, 1990). Negative observations were kept throughout the processing. Considering spatially overlapping reflections, we treated observations as uncorrelated if they resulted from a deconvolution of reflections

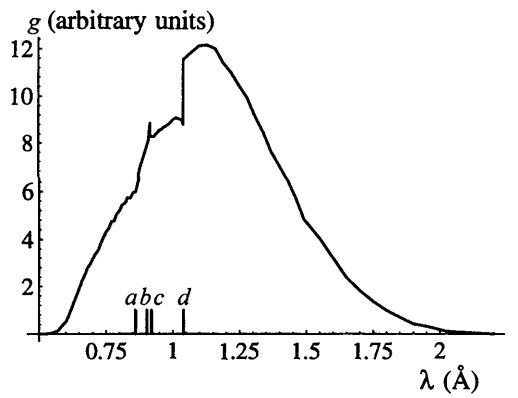


Fig. 4. Wavelength-normalization curve calculated from scaling to reference monochromatic data. *a*, *b*, *c* and *d* denote the absorption edges Au L_I, L_{II}, L_{III} and Br *K*.

that were separated by more than one FWHM. The intensities of the reflections with centroid-to-centroid distances less than one FWHM were added up and further treated in the same way as multiplets.

The exponent index, α, of the absorption correction coefficient of the form $\exp(\alpha\lambda^3)$, frame-to-frame scale factors and the temperature factors were fitted as individual parameters for each of the exposures on the basis of redundant measurements of single reflections. Scaled singlet data were divided into appropriate resolution ranges and $\langle I \rangle$ values were calculated for the ranges $d > 1.27$ Å. Low completeness (47%) of the data in this step in general does not strongly affect the accuracy in the $\langle I \rangle$ values. ~ 40 reflections in each bin are sufficient for reliable estimates (French & Wilson, 1978). This is in good agreement with the theory of structure-factor statistics (Stanley, 1955). At $d < 1.27$ Å, the $\langle I \rangle$ values were strongly affected by a poor signal-to-noise ratio. Therefore, we used an extrapolation assuming that $\langle I \rangle$ in the high-resolution range followed the distribution in the mean squared scattering factor (Wilson, 1949). We interpolated the dependence over the range 2.0–1.27 Å and then extrapolated to 1.0 Å resolution. Prior moments for the intensities at $d < 2.0$ Å were obtained by linear interpolation between the centres of neighbouring bins.

In view of the fact that overprediction may introduce additional noise (not a systematical bias) in the case of non-deconvolutable spatial overlaps, predicted reflections were removed if the prior probability of the reflection intensity being higher than the noise did not exceed 1%. An apparent resolution limit was obtained that varied, for $\lambda = \lambda_{\min}$, between 1.10 Å for the first to 1.22 Å for the last pattern; this reflects an increase in the overall temperature factor by 1.3 Å^2 during the entire data collection. Spot reintegration was then repeated with a correspondingly modified prediction of the pattern.

In the further processing, all observations were treated in the same way, whether they were single or multiple. The redundant observations were found and the data were divided into subsets of cross-correlated reflections in such a way that each unique reflection appeared in one subset only. Normal matrices **C** were calculated for each subset, vectors **R** were found through the singular-value decomposition of **C** and **h** were formed with the prior.

In the program *TRUNCATE* (French & Wilson, 1978), single reflections are considered as negligible if $[J - (\sigma_J^2 h)] \leq -4\sigma_J$. By analogy, the intensity contribution of the *k*th component of a multiple reflection may be treated as negligible when $[(\mathbf{A}^T\mathbf{W}\mathbf{J}_k - h_k/C_{kk}] \leq -4C_{kk}^{-1/2}$. Thus, we omit components on the basis of an estimate of an 'upper limit' of its significance. The estimates for the structure factors and the intensities were derived according to (10)–(13). Integration was carried out numerically, since we could

not derive an analytical expressions for the general case. This part of the procedure is the most complicated step. A special algorithm was implemented for sampling the multi-dimensional PDFs [(7), (8)], which permits the moments to be integrated in a reasonable time with a precision of approximately three significant digits in up to ten dimensions. A description of the details of this algorithm is outside the scope of the present paper.

4.2. *Results*

Bayesian estimates of the intensities and the structure factors were obtained for subsets containing up to ten unique reflections; they included multiplets up to order 10. Resolution shells were rejected if $\langle I\rangle/\langle\sigma_{I_o}\rangle \leq 2$. This limited the resolution of the final data set ('Bayesian Laue data', BLD) to 1.4 Å. Further reflections were omitted if the estimates of σ_{I_o} exceeded $\langle I\rangle$; this limit corresponds to the unmodified prior. BLD contained in total 36 044 reflections.

In order to illustrate the power of the new method, we produced an alternative data set ('standard Laue data', SLD) following a conventional procedure that is basically identical to the procedure usually applied, *e.g.* with the *Daresbury Laue Software Suite* (Campbell, 1993). A constant 'soft' resolution limit of 1.65 Å was used; use of a higher resolution limit would have led to even more serious overprediction problems. The same scale and wavelength-normalization factors were taken as for BLD. All multiple reflections were included that could be processed with a least-squares procedure. This yielded a total of 20 796 reflections corresponding to 78% of the recorded unique reflections. At the same nominal resolution of 1.65 Å, BLD yielded 95% of the recorded reflections.

The completeness of both data sets is compared in Fig. 5. At low and medium resolution ($d > 3.3$ Å), the completeness improves substantially (from 55% to 84%), and, at high resolution ($d > 1.65$ Å), from 77% to 90% in going from the standard processing to the Bayesian procedure. The completeness drops to or below 50% (Bradbrook *et al.*, 1995) at $d \geq 6.3$ and 3.7 Å for BLD and SLD, respectively. In addition, the Bayesian approach provides a substantial number of reflections at high resolution $1.65 \geq d \geq 1.4$ Å (completeness 68%).

For further comparison, agreement factors $R_{\rm mono} = \sum |I_o - I_{\rm mono}| / \sum (I_o + I_{\rm mono})$ were calculated with respect to monochromatic data (Table 2). Consideration of those reflections that are present in both data sets SLD and BLD shows that the improvement in the estimates is due to the different treatment of higher harmonics. It is further due to the fact that observations that introduce degenerate terms into the least-squares matrices were omitted during the standard processing but not in the Bayesian processing. For reflections that are present in BLD only, the accuracy in the estimates is lower, since all these correspond to degenerate components.

Table 2. *R factor versus resolution from a comparison of Laue data to monochromatic data*

Resolution shell (Å)	$R_{\rm mono}$					
	SLD*	SLD	BLD*	BLD†	BLD‡	BLD
5.00	0.045	0.050	0.046	0.031	0.15	0.099
3.02	0.046	0.055	0.037	0.043	0.092	0.067
2.39	0.053	0.066	0.051	0.058	0.16	0.073
2.09	0.061	0.070	0.060	0.066	0.26	0.079
1.90	0.079	0.081	0.076	0.073	0.41	0.107
1.76	0.12	0.12	0.10	0.11	0.39	0.14
1.65	0.17	0.18	0.16	0.17	0.39	0.20
1.58			0.19			0.23
1.51			0.23			0.27
1.45			0.27			0.30
1.40			0.28			0.31

* Single reflections only. † Reflections contained in both SLD and BLD. ‡ Reflections contained in BLD only.

Hajdu *et al.* (1991) showed that Patterson methods are very sensitive in revealing the effects of lack of completeness in data sets. We compared Patterson functions calculated with the Laue data with monochromatic Patterson functions (at 1.0 Å resolution). The correlation factor was found to increase from 0.60 to 0.79 in going from SLD to BLD.

In many time-resolved Laue diffraction studies, the most important criterion will be the contrast and quality in electron-density maps calculated with model phases. Figs. 6–8 demonstrate that the quality of electron-density maps of BPT improved dramatically when Bayesian estimates were included. In order to simulate a near-to-real Laue study, we used phases from a low-temperature (223 K) structure of BPT in an aqueous solution containing 70% methanol (Viehmann & Bartunik, 1993); all solvent molecules were removed from the model. Ten cycles of isotropic refinement against F_o were carried out for both data sets with the program *SHELXL*93 (Sheldrick, 1993). This yielded final $R = \sum |F_o - F_c| / \sum F_c$ values of 0.23 for SLD and 0.24 for BLD at 1.65 Å resolution ($R = 0.26$ for BLD

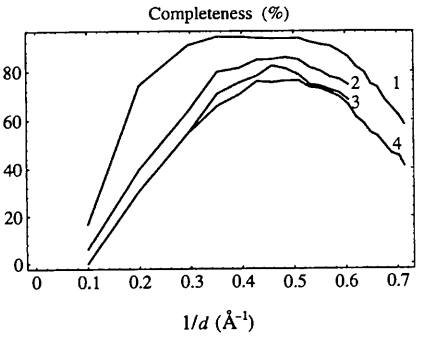
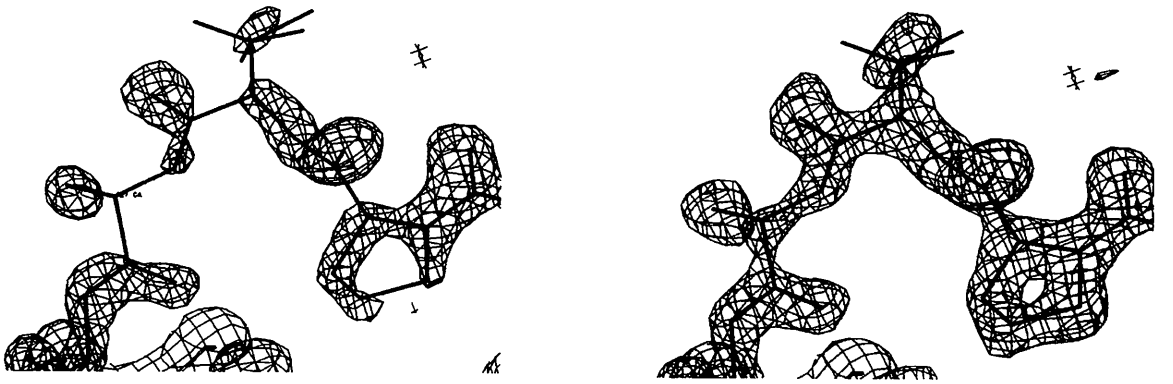
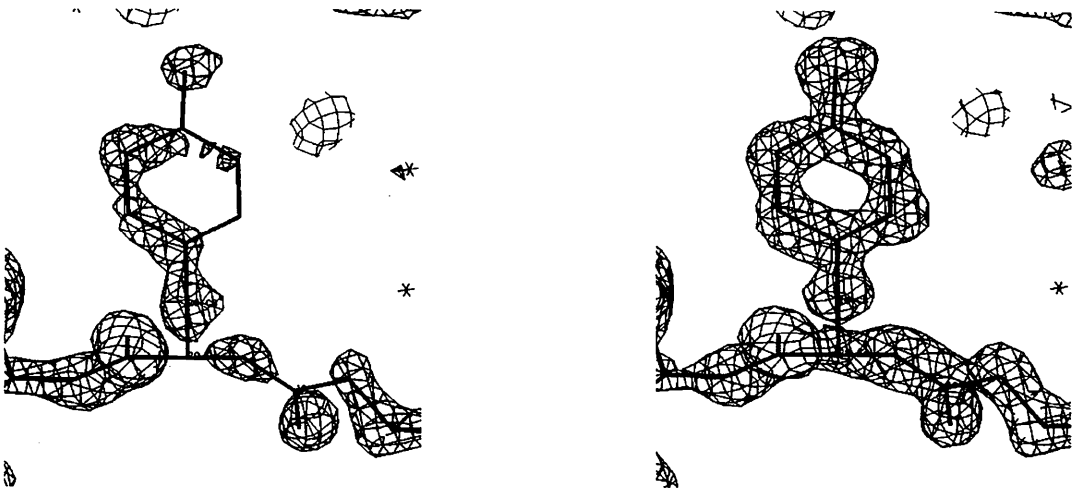


Fig. 5. Completeness in Laue data *versus* resolution. 1 denotes all data in BLD, 2 all data in SLD, 3 single reflections in SLD, 4 single reflections in BLD.



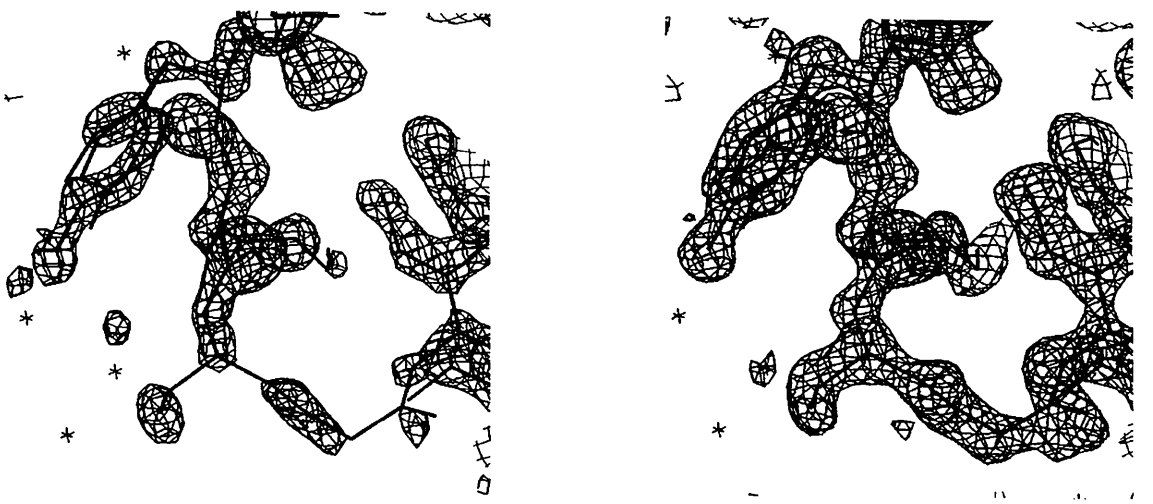
(*a*) (*b*)

Fig. 6. $(2F_o - F_c)$ electron-density maps for Pro 124, Thr 125 and Ser 127. Structure-factor amplitudes from (*a*) SLD and (*b*) BLD.



(*a*) (*b*)

Fig. 7. $(2F_o - F_c)$ electron-density maps for Tyr 20. Structure-factor amplitudes from (*a*) SLD and (*b*) BLD.



(*a*) (*b*)

Fig. 8. $(2F_o - F_c)$ electron-density maps for Tyr 59, Lys 60, Ser 61 and Gly 62. Structure-factor amplitudes from (*a*) SLD and (*b*) BLD.

at 1.4 Å resolution). $(2F_o - F_c)$ difference Fourier maps that were calculated with SLD showed quite low contrast; about 20 discontinuities were visible along the main chain. With BLD, the number of main-chain cuts was reduced to one. In total, the Bayesian procedure resulted in $(2F_o - F_c)$ maps of comparable quality to monochromatic maps that were calculated at the same nominal resolution.

In order to quantitatively assess the improvement in the quality of the maps, (F_o, φ_{calc}) maps that were calculated with structure-factor amplitudes from both Laue data sets were analysed in their correlation with monochromatic maps. In this case, phases were taken from an atomic resolution BPT model (Popov & Bartunik, 1996). An overall correlation factor, c, was calculated:

$$c = (\langle \rho_m \rho_L \rangle - \rho_m \langle \rho_L \rangle) \times (\langle \rho_m^2 \rangle - \langle \rho_m \rangle^2)^{-1/2} (\langle \rho_L^2 \rangle - \langle \rho_L \rangle^2)^{-1/2}. \quad (22)$$

ρ_m refers to the monochromatic (1.0 Å resolution) and ρ_L to Laue electron density. We obtained $c = 0.79$ and 0.90 for the maps with SLD and BLD coefficients, respectively. This corresponds to real-space R factors (Bränden & Jones, 1990) of 0.34 and 0.16, respectively. Further, all atoms (protein and solvent) in the model were subdivided into bins according to the atomic B factors. The bin width was set to 5 Å^2. Correlation factors were calculated separately for areas surrounding the atoms from each bin. Fig. 9 shows the correlation factors as a function of the atomic temperature factors. As one would expect, the improvement in the map quality is most evident in parts of the structure with high B values.

5. Estimate of the completeness of Laue data from a single-shot experiment

The Bayesian method of Laue data processing does not require redundancy in the measurement of structure-factor amplitudes. This is of particular interest for structural studies of non-cyclic reactions. In many applications, adequately short time scales may only be reached if all information that is required for structural analysis is obtained in a single Laue exposure. The time needed for changing the crystal orientation and for reading out an area detector is long compared with, for example, the typical lifetimes (1–10 ms) of enzymatic reaction intermediates at ambient temperature. Even in the case of reactions that may be cycled (*e.g.* photodissociation of ligands from the iron in heme proteins), single-shot experiments may yield better results since they will be less affected by radiation damage or by variations in the excitation conditions.

Our test experiment was carried out with an orthorhombic crystal. In this case, a maximum of 55% of the asymmetric unit may be covered with a single exposure (Clifton, Elder & Hajdu, 1991). The actual maximum coverage by a single exposure in this experiment was 27%. Most of the reflections were observed twice. From the doublets in the data that are truncated at 1.65 Å, the redundancy was sufficient for least-squares deconvolution of 28 systems containing four observations. For all of the 466 doublets represented by two observations, the relative cross correlation was −98 to −100%, indicating complete degeneration. The standard procedure yielded a completeness of 9% ($R_{mono} = 0.065$) at $d > 3.3$ Å, and of 18.1% ($R_{mono} = 0.091$) at $3.3 < d < 1.65$ Å. The Bayesian processing up to the geometrical resolution limit yielded 24.8% of the unique reflections (*i.e.* 90.5% of the recorded reflections) at $d > 3.3$ Å and 21.3% (85.2%) at $3.3 < d < 1.65$ Å. R_{mono} was 0.125 at $d > 3.3$ Å and 0.108 at $3.3 < d < 1.65$ Å.

In the following, we estimate the completeness in the Laue data that may be obtained in the case of sufficiently high crystal symmetry from a single-shot experiment if the data are processed with the Bayesian method. Simultaneous exploration of a high percentage (> 80%) of an asymmetric unit in theory is feasible only for crystals belonging to the Laue class $4/mmm$ or higher symmetries (Clifton, Elder & Hajdu, 1991). We take tetragonal hen-egg-white lysozyme ($P4_32_12$; $a = b = 79.1$ Å, $c = 37.9$ Å) as a hypothetical model structure. The maximal fraction of simultaneous symmetry-unique observations is obtained for a crystal orientation defined by the rotation angles $\varphi_x = 0$, $\varphi_y = 25$, $\varphi_z = 22.5°$ relative to the incident beam. Hereby, the X axis is parallel to the incident beam direction; the crystal axes a and c are along **X** and **Z**, respectively. With the assumption of a crystal-to-detector distance of 200 mm and a wavelength band of 0.52–2.1 Å, the asymmetric unit is covered to 85.3% at $d > 2$ Å for this crystal orientation. In order to estimate the percentage of the reflection intensities that may be measured with

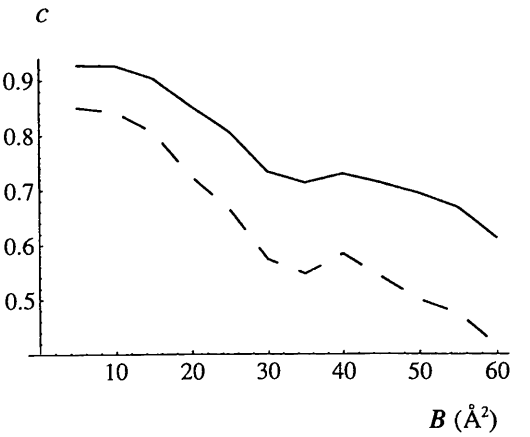


Fig. 9. Correlation factor derived from a comparison of Laue (F_o, φ_{calc}) density maps to a monochromatic map as a function of the atomic temperature factor. Solid line for BLD, dashed line for SLD.

sufficient accuracy, we request $I/\sigma_I > 2$ at this resolution. This criterion is related to the interpretability of the electron-density map. For a well ordered and sufficiently large crystal of such a small- to medium-size structure, reflections up to this resolution on average will have such high peak-to-background ratios that incoherent scattering from the crystal and its environment may be neglected. The essential source of errors, besides the counting statistics, will then be the limited accuracy in deconvoluted harmonic components of energetically overlapping spots. For doublets, the Bayesian method will yield the strong intensity components I_1 with $I/\sigma_1 > 2$ when the ratio of the prior moments is $\langle I_1\rangle a_1/\langle I_2\rangle a_2 > 0.75$ (see §2.3). In the case of higher-order multiplets, the denominator is replaced by the sum of the prior moments for the remaining terms. We assume the same wavelength-normalization curve as in the test application to trypsin. For the calculation of the distribution of $\langle I\rangle$ *versus* resolution, we assume an overall temperature factor $B = 9\,\text{Å}^2$. As a final result, we obtain a completeness of 75% in the set of structure-factor amplitudes to 2 Å resolution (56% at 4 Å). Thus, a single-shot Laue exposure may in fact provide sufficiently high completeness for Fourier summation. The estimated completeness corresponds to 87.9% of the simultaneously recorded reflections; this is in good agreement with the percentage of reflections extracted from a single exposure in the test trypsin application. For higher-resolution ranges, *e.g.* $2.0 > d > 1.5$ Å, the incoherent scattering in general will substantially affect the signal-to-noise ratio; as a consequence, the completeness will be further reduced to < 65%.

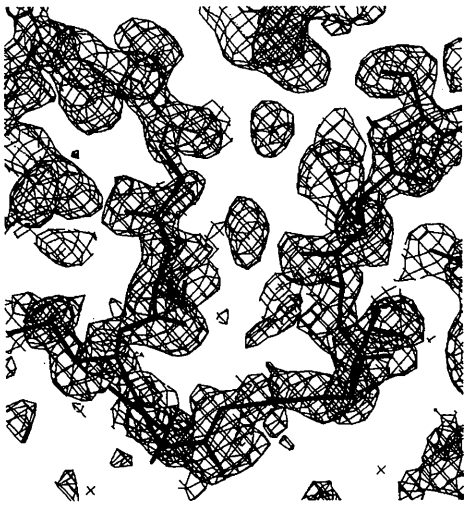


Fig. 10. $(2F_o - F_c)$ electron-density map calculated with (BLD) structure-factor amplitudes from BPT and phases from trypsinogen for a section containing the residues 184–188. Bold lines indicate the residues that were omitted from the model.

6. Contrast in Laue density maps and detection of conformational changes

The contrast in electron-density maps that are calculated with a given set of Laue data depends on the quality of the phases. In time-resolved Laue studies, phases will be known for the initial structural state. The interpretability of Laue density maps for an excited state, *e.g.* a reaction intermediate, will depend both on the differences in the phases between the two structures and the quality and completeness of the Laue data.

We investigated the contrast in electron-density maps that may result from Laue studies of protein structures undergoing extended conformational changes. As a hypothetical model, we considered a structural transition from bovine trypsinogen to a trypsin-like conformation. The activation domain comprising 33 residues (*i.e.* 15% of the molecule) is disordered in trypsinogen and ordered in trypsin (Bode & Huber, 1986). We aligned (using the program *O*; Jones, Zou, Cowan & Kjeldgaard, 1991) the structural model of trypsinogen to that of trypsin; the disordered residues were omitted. The remaining residues have a r.m.s. deviation of 0.26 Å. The r.m.s. phase difference between the two models is 48°. We take trypsinogen as the 'initial state' and trypsin as the 'final state' of the disorder–order reaction. We assume that Laue diffraction patterns were recorded from the final state and that Bayesian processing yielded the Laue data ('BLD') as described above in the test application to the trypsin. The BLD structure-factor amplitudes were combined with the phases from the initial state model. Fig. 10 shows an example of a $(2F_o - F_c)$ difference Fourier map at 1.4 Å resolution covering a segment of the final state structure that is disordered in the initial state. In this 'omit map', the backbone is quite well defined; in fact, it would be possible to trace the main chain for the whole activation domain despite the lacking phase information. Furthermore, most side chains (except of Gly and Ala) can be identified. This simulation demonstrates the feasibility of detecting even very extended conformational changes in Laue density maps at high resolution if the quality and completeness of Laue structure-factor amplitudes are as high as obtained in the Bayesian processing of the trypsin data.

7. Discussion

In the test application to BPT, the Bayesian procedure yielded estimates for the intensity components of reflection multiplets. The accuracy of these estimates is necessarily lower than the accuracy that can be achieved from independent measurement of (single) reflections under given white-beam conditions. Nevertheless, the quality of electron-density maps improved substantially, essentially due to much higher completeness at low and medium resolution. This represents a

significant gain from the Bayesian analysis as compared with the standard procedure. Most of the reflections in both data sets SLD and BLD at 1.65 Å resolution were represented by four observations; this redundancy did not yield well defined systems for least-squares deconvolution. This is partly due to a symmetric crystal setting. Under different crystal settings, the standard procedure can yield higher completeness. The improvement in the map quality demonstrates that the Bayesian method successfully copes with the degenerate systems. In the case of a single exposure, the least-squares method failed completely whereas the Bayesian procedure still was efficient. This demonstrates the potential power of the method for applications where highest possible completeness must be obtained from a minimum number of exposures. In general, redundant observations will also be present and the Bayesian procedure makes use of them. For example, corrections for wavelength-dependent absorption must be defined from redundant observations of single reflections. Redundancy in the measurements of multiple reflections, even if they are recorded at such conditions that their wavelength normalization coefficients are linearly dependent, reduces the standard uncertainty in the deconvoluted components. The accuracy in the Bayesian estimates noticeably increases when redundancy removes, at least partly, the degeneracy of the corresponding least-squares system. However, the existence of redundant measurements is not a necessary condition for obtaining statistically significant estimates from the Bayesian approach.

We tested the method in an application to a well ordered crystal. In the case of crystals exhibiting broadened mosaicity, Laue patterns contain streaky and spatially overlapping reflections. Such broadening is often observed during or shortly after initiation of a reaction in the crystal (Bartunik, Bartunik & Viehmann, 1992). With the new Bayesian method, there is no formal difference between the processing of partially overlapping reflections and the processing of a multiplet measured with redundancy provided that profile fitting is applied. Our formulation extends to this case without any modification simply by considering each individual pixel as an observation and incorporating the model profile value in the coefficient a_i; such a further development of our software is being completed. In the present test application, we took observations as uncorrelated if reflections were separated by more than one FWHM. Observations resulting from more closely spaced reflections were taken as -100% correlated, simply because the correlation coefficients were not calculated by the peak-integration software of the Daresbury package.

Structure refinement using Laue data should not take the squared structure-factor amplitudes derived from deconvoluted intensity components as target values but rather their linear superpositions corresponding to the experimental observations. The individual reflection intensity components that are extracted from energy or spatial overlaps may not be treated as independent observations.

The previously discussed method is based on the validity of Wilson distributions, *i.e.* on the assumption that atoms are uniformly distributed over the unit cell. The Bayesian method may be extended if additional *a priori* information is available on the crystal structure under investigation. In time-resolved studies, the protein structure will be known approximately. For example, when investigating a reaction in the crystal, the structure corresponding to the initial state will have been determined previously. The initial model may be taken into account through conditional probability densities $p(F_o|F_{\rm model})$ (Srinivasan & Chandrasekharan, 1966). Replacing the Wilson prior by this function in the Bayesian approach results in substantially sharper posterior PDFs for the components. Hence, more accurate estimates may be obtained. Software development for calculating moments for the case of conditional PDFs is in progress.

8. Conclusions

The Bayesian procedure of deconvoluting multiplets in Laue diffraction patterns represents a novel approach to solving the fundamental problems of Laue protein crystallography. As compared with all previously suggested methods of deconvolution, the Bayesian approach provides reliable estimates of the uncertainties in the deconvoluted reflection intensity components and the resulting structure-factor amplitudes. A meaningful weighting scheme may be derived from these estimates. Furthermore, the new method does not require redundant measurement of reflection intensities. Hence, Laue data collection may be limited to a minimum number of crystal orientations that are needed for a given space group in order to achieve sufficiently high completeness. This is of importance for time-resolved applications, in particular for time scales that are short (< 10 ms) compared with the time needed with presently available technology for changing the crystal setting. Model calculations for the example of a tetragonal protein structure showed that a single-shot Laue experiment combined with Bayesian processing may yield a completeness of 75% in the structure-factor amplitudes to 2.0 Å resolution. Such time-resolved applications represent the main aim of Laue protein crystallography.

A test application of the Bayesian method to Laue diffraction data from orthorhombic trypsin demonstrated its power. The completeness in the Laue data set improved strongly compared with the standard procedure followed up to now. The low-resolution hole was filled up significantly and structure-factor amplitudes were obtained up to substantially higher

resolution. In fact, much higher resolution (1.4 Å) was reached than in any previous Laue data processing. No 'soft' diffraction limit was assumed; rather, data were evaluated to the physically relevant (wavelength-dependent) diffraction limit without loss of completeness at low and medium resolution. The contrast in electron-density maps calculated with Laue structure-factor amplitudes and model phases improved dramatically. $(2F_o - F_c)$ difference Fourier maps became well comparable in quality to typical monochromatic high-resolution maps of the same structure.

Structure-factor amplitudes of similar quality and completeness as obtained from the Bayesian processing of the trypsin Laue data may produce interpretable density maps at high resolution, even if the structure under investigation differs substantially from the structural model that provided the phases. The theoretical example of a disorder–order transition in trypsinogen, involving a r.m.s. difference in phases of almost 50°, indicates the wide range of potential applications of Laue techniques.

The power of the Bayesian method may be further enhanced. Spatial overlaps may be treated in a formally identical way as energy overlaps if profile fitting is included in the spot-integration routines. Furthermore, since approximate knowledge of the protein structure (*e.g.* in an initial state) may be assumed for time-resolved investigations, we may use conditional probability densities that should provide even more accurate estimates of the structure factors derived from the deconvolution. We are presently carrying out such further developments. The new approach may provide a breakthrough in the applicability of Laue diffraction methods that until now has mostly been limited by poor contrast and quality of the resulting electron-density maps.

The authors gratefully acknowledge helpful discussions with Professor D. M. Kheiker, A. Lebedev and M. Kozin, and the support of C. Reissner in the crystal preparation.

References

Andersson, I., Clifton, I. J., Edwards, S. L., Fülöp, V., Hadfield, A. T., Nordlund, P., Phizackerley, R. P., Soltis, S. M., Wakatsuki, S. & Hajdu, J. (1992). *Time-Resolved Macromolecular Crystallography*, edited by D. W. J. Cruickshank, J. R. Helliwell & L. N. Johnson, pp. 95–102. Oxford University Press.

Bartsch, H. H., Huang, Q., Koelln, I., Rupp, S., Summers, L. J. & Bartunik, H. D. (1990). *Acta Cryst.* A**46**, C21.

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

Bartunik, H. D., Bartunik, L. J. & Viehmann, H. (1992). *Philos. Trans. R. Soc. London*, **340**, 209–220.

Bode, W. & Huber, R. (1986). *Molecular and Cellular Basis of Digesion*, edited by P. Desnuelle, H. Sjöström & O. Norén, pp. 213–234. Amsterdam: Elsevier.

Bradbrook, G., Deacon, A., Habash, J., Helliwell, J. R., Helliwell, M., Nieh, Y. P., Snell, E. H., Trapani, S., Thompson, A. W., Campbell, J. W., Allinson, N. M., Moon, K., Ursby, T. & Wulff, M. (1995), *SPIE*, **2521**, 161–177.

Bränden, C. & Jones, A. (1990). *Nature (London)*, **343**, 687–689.

Campbell, J. W. (1993). *Daresbury Laue Software Suite*. SERC Daresbury Laboratory, Warrington, England.

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

Clifton, I. J., Elder, M. & Hajdu, J. (1991). *J. Appl. Cryst.* **24**, 267–277.

Cruickshank, D. W. J., Helliwell, J. R. & Moffat, K. (1987). *Acta Cryst.* A**43**, 656–674.

Duke, E. M. H., Hadfield, A., Walters, S., Wakatsuki, S., Bryan, R. K. & Johnson, L. N. (1992). *Philos. Trans. R. Soc. London*, **340**, 245–261.

French, S. & Wilson, K. S. (1978). *Acta Cryst.* A**34**, 517–525.

Hajdu, J., Almo, S. C., Farber, G. K., Prater, J. K., Petsko, G. A., Wakatsuki, S., Clifton, I. J. & Fülöp, V. (1991). *Crystallographic Computing 5: from Chemistry to Biology*, edited by D. Moras, A. D. Pojarny & J. C. Thierry, pp. 29–49. Oxford University Press.

Hajdu, J. & Andersson, I. (1993). *Annu. Rev. Biophys. Biomol. Struct.* **22**, 467–498.

Hajdu, J., Machin, P., Campbell, J. W., Greenhough, T. J., Clifton, I. J., Zurek, S., Gover, S., Johnson, L. N. & Elder, M. (1987). *Nature (London)*, **329**, 178–181.

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. (1992). *Macromolecular Crystallography* with Synchrotron Radiation. Cambridge University Press.

Helliwell, J. R., Habash, J., Cruickshank, D. W. J., Harding, 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.

Hohlwein, D. & Mason, S. A. (1981). *J. Appl. Cryst.* **14**, 24–27.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.

Kheiker, D. M. & Nekrasov, Yu. V. (1970). *Appar. Metody Rentgenovskogo Anal.* **7**, 23–28.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). *Acta Cryst.* B**39**, 480–490.

Moffat, K., Szebenyi, D. M. E. & Bilderback, D. H. (1984). *Science*, **223**, 1423–1425.

Popov, A. N. & Bartunik, H. D. (1996). In preparation.

Ren, Z. & Moffat, K. (1995*a*). *J. Appl. Cryst.* **28**, 461–481.

Ren, Z. & Moffat, K. (1995*b*). *J. Appl. Cryst.* **28**, 482–483.
Rupp, S. (1991). Diploma thesis, University of Hamburg, Germany.
Sheldrick, G. W. (1993). *SHELXL93. Program for the Refinement of Crystal Structures.* University of Göttingen, Germany.
Shrive, A. C., Clifton, I. J., Hajdu, J. & Greenhough, T. G. (1990). *J. Appl. Cryst.* **23**, 169–174.
Sivia, D. S. & David, W. I. F. (1994). *Acta Cryst.* A**50**, 703–714.
Smith Temple, B. (1989). PhD thesis, Cornell University, USA.
Srinivasan, R. & Chandrasekharan, R. (1966). *Indian J. Pure Appl. Phys.* **4**, 178.
Stanley, E. (1955). *Acta Cryst.* **8**, 356–357.
Szebenyi, D. M., Bilderback, D., LeGrand, A., Moffat, K., Schildkamp, W., Temple, B. S. & Teng, T. (1992). *J. Appl. Cryst.* **25**, 414–423.
Vellieux, F. M. D., Hajdu, J., Verlinde, C. L. M. J., Groendijk, H., Read, R. J., Greenhough, T. J., Campbell, J. W., Kalk, K. H., Littlechild, J. A., Watson, H. C. & Hol, W. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 2344–2359.
Viehmann, H. & Bartunik, H. D. (1993). Unpublished results.
Wakatsuki, S. (1993). Proc. CCP4 Study Weekend, 29–30 January 1993, pp. 71–79. SERC Daresbury Laboratory, Warrington, England.
Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.