Acta Crystallographica Section D
**Biological Crystallography**
ISSN 0907-4449

# Fisher's information in maximum-likelihood macromolecular crystallographic refinement

**Roberto A. Steiner, Andrey A. Lebedev and Garib N. Murshudov***

Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, England

Correspondence e-mail: garib@ysbl.york.ac.uk

Received 13 June 2003
Accepted 21 August 2003

Fisher's information is a statistical quantity related to maximum-likelihood theory. It is a matrix defined as the expected value of the squared gradient of minus the log-likelihood function. This matrix is positive semidefinite for any parameter value. Fisher's information is used in the quasi-Newton scoring method of minimization to calculate the shift vectors of model parameters. If the matrix is non-singular, the scoring-minimization step is always downhill. In this article, it is shown how the scoring method can be applied to macromolecular crystallographic refinement. It is also shown how the computational costs involved in calculation of the Fisher's matrix can be efficiently reduced. Speed is achieved by assuming a continuous distribution of reciprocal-lattice points. Matrix elements calculated with this method agree very well with those calculated analytically. The scoring algorithm has been implemented in the program *REFMAC5* of the *CCP*4 suite. The Fisher's matrix is used in its sparse approximation. Tests indicate that the algorithm performs satisfactorily.

## 1. Introduction

The maximum-likelihood (ML) method together with Bayesian statistics have found successful application in macromolecular crystallography (Bricogne, 1997). Remarkably good results have already been achieved in the areas of experimental phasing and phase improvement (de La Fortelle & Bricogne, 1997; Lunin & Urzhumtsev, 1984; Read, 1990), map calculation (Read, 1986), density modification (Terwilliger, 2000, 2001), molecular replacement (Read, 2001) and structure refinement (Bricogne & Irwin, 1996; Pannu & Read, 1996; Murshudov *et al.*, 1997; Pannu *et al.*, 1998). Yet further advances can be expected with the development of more sophisticated probability distributions. In the long run, the natural ability of this statistical framework to incorporate information from different sources (Bricogne, 1997) can even be expected to lead to the substantial decompartimentalization of many of the traditional crystallographic computational steps. Such a unifying approach would properly represent the continuity existing in the process of structure determination. In principle, the ML method can also allow the joint treatment of data arising from related experimental techniques such as X-ray diffraction, electron diffraction and neutron diffraction.

The ever-increasing appreciation of the benefits and potential of the ML method in macromolecular crystallography is stimulating the introduction of related statistical quantities to the field. A well established concept in ML theory is that of Fisher's information (Fisher, 1922; Stuart *et al.*, 1999). Fisher's information is a matrix defined as the expected value of the squared gradient of minus the log-

likelihood function. This quantity is encountered in different fields. In processes of optimization, it is used in connection with the scoring method (Osborne, 1992; Smyth, 1996) to maximize a likelihood function (or minimize a minus log-likelihood function). It is also a measure of the ability to estimate a parameter (Stuart *et al.*, 1999). Algorithms based on Fisher's information are used in population genetics (Frieden *et al.*, 2001), image-quality assessment (Parra & Barrett, 1998; Barrett *et al.*, 1995), population pharmacokinetic studies (Tod *et al.*, 1998; Retout *et al.*, 2001) and time-series modelling (Klein & Melard, 1995). Frieden (1998) has also proposed a physical significance for this matrix as a measure of the state of disorder of a system.

Fisher's information and the related scoring method of minimization find natural application in ML macromolecular crystallographic refinement. The process of structure improvement is a large optimization problem which takes advantage of algorithms of different complexity and applicability. Generally, the choice of the most appropriate algorithm depends on the accuracy of the initial parameters. At early stages of crystal structure analysis, simulated annealing combined with molecular dynamics has been shown to improve the refinement behaviour and to increase the radius of convergence (Brünger *et al.*, 1987). At later stages, first- and second-order minimization methods give better results (Fletcher, 1987). In particular, the more demanding second-order methods (Newton and quasi-Newton methods), which use the second derivatives of the function in addition to the first derivatives, have better rates of convergence. The scoring method belongs to the quasi-Newton methods. It approximates the Hessian matrix used in Newton's method for the solution of the system of linear equations to obtain the shift vectors of atomic parameters with the Fisher's matrix.

In this account, we introduce the Fisher's information and the scoring method of minimization to macromolecular crystallographic refinement. In addition to working out the relevant equations, we show how the considerable costs associated with evaluation of the matrix can be dramatically reduced. To this end, we use a generalization of the method proposed by Templeton (1999) for the fast calculation of normal matrix elements in least-squares refinement. The scoring method of minimization has been implemented in the program *REFMAC*5 (Murshudov *et al.*, 1997) of the *CCP*4 (Collaborative Computational Project, Number 4, 1994) suite. It uses the sparse approximation of the Fisher's information. Tests indicate that the algorithm performs satisfactorily.

## 2. Fisher's information and the scoring method of minimization

Let $\mathbf{o} = (o_1, o_2, \ldots, o_n)^T$ be a random vector of observations with conditional probability distribution $\mathrm{P}(\mathbf{o}; \mathbf{p}) = \mathcal{L}(\mathbf{o}; \mathbf{p})$[1] where $\mathbf{p} = (p_1, p_2, \ldots, p_m)^T$ is a parameter vector. Let $L = -\log \mathcal{L}(\mathbf{o}; \mathbf{p})$ denote the minus log-likelihood function of $\mathbf{p}$. Let also $\mathbf{u}(\mathbf{p})$ be the gradient of $L$,

$$\mathbf{u}(\mathbf{p}) = \frac{\partial L}{\partial \mathbf{p}}. \tag{1}$$

The quantity $-\mathbf{u}(\mathbf{p})$ is called the score function (Stuart *et al.*, 1999).

Fisher's information $\mathcal{I}(\mathbf{p})$ is a $(m \times m)$ matrix defined as (Stuart *et al.*, 1999)

$$\begin{aligned}\mathcal{I}(\mathbf{p}) &= \mathcal{E}_{\mathbf{o}}[\mathbf{u}(\mathbf{p})\mathbf{u}(\mathbf{p})^T] \\ &= \int \ldots \int [\mathbf{u}(\mathbf{p})\mathbf{u}(\mathbf{p})^T] \exp(-L)\, \mathrm{d}o_1 \ldots \mathrm{d}o_n. \end{aligned} \tag{2}$$

The symbol $\mathcal{E}$ indicates the expectation operator. The subscript defines the variable(s) of integration. In the remainder of the paper it will be dropped, except when the variable(s) of integration need to be explicitly stated to avoid confusion.

If the second derivatives of $L$ exist and the operations of differentiation and integration can be exchanged, $\mathcal{I}$ is alternatively expressed as (Lee, 1997; Stuart *et al.*, 1999)

$$\mathcal{I}(\mathbf{p}) = \mathcal{E}_{\mathbf{o}}\left(\frac{\partial^2 L}{\partial \mathbf{p} \partial \mathbf{p}^T}\right). \tag{3}$$

The Hessian $(\partial^2 L/\partial \mathbf{p} \partial \mathbf{p}^T)$ calculated at the given observation point is also called the observed information matrix (Smyth, 1996).

The scoring method is a quasi-Newton method of minimization (Osborne, 1992; Smyth, 1996), whereby the maximum-likelihood estimator $\hat{\mathbf{p}}$ of $\mathbf{p}$ is sought in an iterative manner by solving the system of linear equations

$$\mathcal{I}_k \zeta_k = -\mathbf{u}_k, \tag{4}$$

where $\zeta_k$ represents the vector shift to be applied at cycle $k$ to the current estimate $\mathbf{p}_k$.

The scoring method has attractive properties compared with Newton's method, which represents the paradigm for the second-order minimization methods. The characteristic feature of the scoring algorithm is the replacement of the observed information matrix with its expectation. The latter is positive semidefinite for any parameter value. As a result, the scoring step is necessarily downhill if $\mathcal{I}$ is non-singular. This property gives the scoring method a distinct advantage over Newton's method as it removes one of the classical problems of the latter; that is, its inability to distinguish maxima from minima. According to Smyth (1996), the scoring method is, in general, linearly convergent at a rate which depends on the relative difference between the observed and expected information. In short runs the scoring method can sometimes be even superior to Newton's method. Kale (1961) has shown that although Newton's method ultimately converges faster in the vicinity of the minimum, the scoring method will often give better results for the first few iterations when the number of observations is large. This method therefore seems appropriate for macromolecular crystallographic refinement.

[1] In this account, we follow the notation of Stuart *et al.* (1999). When $\mathbf{p}$ is fixed and $\mathbf{o}$ varies, $\mathcal{L}(\mathbf{o}; \mathbf{p})$ represents the probability distribution of $\mathbf{o}$. When $\mathbf{o}$ is fixed and $\mathbf{p}$ varies, $\mathcal{L}(\mathbf{o}; \mathbf{p})$ instead represents the likelihood of $\mathbf{p}$. The interpretation will be clear from the context.

## 3. Fisher's information in crystallographic minimization

In order to use (4) in macromolecular crystallographic minimization, we need to find efficient ways of calculating $\mathcal{I}$ and $\mathbf{u}$. Equations for $\mathbf{u}$ in a form suitable for most residuals using fast Fourier transformation (FFT) have been given by various authors (Agarwal, 1978; Lunin & Urzhumtsev, 1985; Brünger, 1989; Bricogne, 1993; Murshudov *et al.*, 1997; Tronrud, 1999). The score vector $\mathbf{u}$ is therefore assumed to be known. In this account, we will focus on $\mathcal{I}$.

Let $L$ depend on all structure factors $F_{\mathbf{h}}$ and $F^*_{\mathbf{h}}$ and be invariant under the exchange of $F_{\mathbf{h}}$ and $F^*_{-\mathbf{h}}$. Let also $F_{\mathbf{h}}$ contain entirely the dependency on model parameters $p_i$s. If $F_{\mathbf{h}} = F^*_{-\mathbf{h}}$, *e.g.* in the absence of anomalous scatterers, the first derivative of $L$ with respect to a parameter $p_i$ can be expressed as

$$\frac{\partial L}{\partial p_i} = \sum_{\mathbf{h}} \left( \frac{\partial L}{\partial A_{\mathbf{h}c}} - i \frac{\partial L}{\partial B_{\mathbf{h}c}} \right) \frac{\partial F_{\mathbf{h}c}}{\partial p_i}, \tag{5}$$

where $A_{\mathbf{h}c}$ and $B_{\mathbf{h}c}$ are the real and imaginary parts of the calculated structure factor $F_{\mathbf{h}c}$, respectively, and the summation is over all reflections $\mathbf{h} = (h_1, h_2, h_3)$.[2]

As $L$ is real, the quantity $(\partial L/\partial p_i)$ is real. Thus, (5) can also be written using the complex conjugates

$$\frac{\partial L}{\partial p_i} = \sum_{\mathbf{h}} \left( \frac{\partial L}{\partial A_{\mathbf{h}c}} - i \frac{\partial L}{\partial B_{\mathbf{h}c}} \right)^* \left( \frac{\partial F_{\mathbf{h}c}}{\partial p_i} \right)^*; \tag{6}$$

therefore, by substituting (5) and (6) into (2), we obtain

$$\begin{aligned} \mathcal{I}(p_i, p_j) &= \mathcal{E}\Bigg[ \sum_{\mathbf{h}} \left( \frac{\partial L}{\partial A_{\mathbf{h}c}} - i \frac{\partial L}{\partial B_{\mathbf{h}c}} \right) \frac{\partial F_{\mathbf{h}c}}{\partial p_i} \\ &\quad \times \sum_{\mathbf{k}} \left( \frac{\partial L}{\partial A_{\mathbf{k}c}} - i \frac{\partial L}{\partial B_{\mathbf{k}c}} \right)^* \left( \frac{\partial F_{\mathbf{k}c}}{\partial p_j} \right)^* \Bigg] \\ &= \sum_{\mathbf{h}} \sum_{\mathbf{k}} \mathcal{E}\left[ \left( \frac{\partial L}{\partial A_{\mathbf{h}c}} - i \frac{\partial L}{\partial B_{\mathbf{h}c}} \right) \left( \frac{\partial L}{\partial A_{\mathbf{k}c}} - i \frac{\partial L}{\partial B_{\mathbf{k}c}} \right)^* \right] \\ &\quad \times \frac{\partial F_{\mathbf{h}c}}{\partial p_i} \left( \frac{\partial F_{\mathbf{k}c}}{\partial p_j} \right)^*. \end{aligned} \tag{7}$$

The conditional expectation (given $\mathbf{p}$) of the first derivatives of $L$ is zero. Therefore, if the derivatives $[(\partial L/\partial A_{\mathbf{h}c}) - i(\partial L/\partial B_{\mathbf{h}c})]$ for different reflections are uncorrelated (this is the case for most crystallographic likelihood functions), all terms in the above equation vanish except for $\mathbf{h} = \pm\mathbf{k}$. Thus, $\mathcal{I}$ can be expressed as

$$\begin{aligned} \mathcal{I}(p_i, p_j) &= \sum_{\mathbf{h}} \mathcal{E}\left[ \left| \frac{\partial L}{\partial A_{\mathbf{h}c}} - i \frac{\partial L}{\partial B_{\mathbf{h}c}} \right|^2 \right] \left( \frac{\partial F_{\mathbf{h}c}}{\partial p_i} \right) \left( \frac{\partial F_{\mathbf{h}c}}{\partial p_j} \right)^* \\ &\quad + \sum_{\mathbf{h}} \mathcal{E}\left[ \left( \frac{\partial L}{\partial A_{\mathbf{h}c}} - i \frac{\partial L}{\partial B_{\mathbf{h}c}} \right)^2 \right] \left( \frac{\partial F_{\mathbf{h}c}}{\partial p_i} \right) \left( \frac{\partial F_{\mathbf{h}c}}{\partial p_j} \right) \\ &= \mathcal{I}_1 + \mathcal{I}_2. \end{aligned} \tag{8}$$

[2] Without loss of generality, an orthogonal cell system is assumed throughout the paper. This can be generated from any cell system using a suitable matrix (McKie & McKie, 1986). Owing to the orthogonal setting, the reflection indices are therefore in general non-integer. For convenience, all equations are also given assuming a $P1$ space group.

The terms $\mathcal{I}_1$ and $\mathcal{I}_2$ generalize the terms $H_1$ and $H_2$, respectively, derived by Agarwal (1978) in the case of least-squares. Similarly to $H_2$, the term $\mathcal{I}_2$ can be predicted to be small as it involves the phase term. (8) can therefore be written as

$$\mathcal{I}(p_i, p_j) \cong \mathcal{I}_1 = \sum_{\mathbf{h}} \mathcal{E}(W_{\mathbf{h}}) \left( \frac{\partial F_{\mathbf{h}c}}{\partial p_i} \right) \left( \frac{\partial F_{\mathbf{h}c}}{\partial p_j} \right)^*, \tag{9}$$

where

$$W_{\mathbf{h}} = \left| \frac{\partial L}{\partial A_{\mathbf{h}c}} - i \frac{\partial L}{\partial B_{\mathbf{h}c}} \right|^2. \tag{10}$$

$\mathcal{E}(W_{\mathbf{h}})$ can be regarded as the likelihood-weighting factor. It modulates $\mathcal{I}$ depending on the particular likelihood function employed. If the observations are independent given $\mathbf{p}$ and normally distributed, then the likelihood-weighting factor becomes the least-squares weight.

Notably, $W_{\mathbf{h}}$ is the only term in (9) that depends on $L$. This is convenient for implementation purposes as the use of different likelihood functions only requires the estimation of a different $\mathcal{E}(W_{\mathbf{h}})$. (10) also shows that the derivation of $W_{\mathbf{h}}$ can be performed using only the first derivatives of $L$. Although an alternative formulation which uses the second derivatives is also possible (see Appendix *A*), the form of $W_{\mathbf{h}}$ given above has clear advantages in terms of simplicity of implementation.

## 4. The likelihood-weighting factor $\mathcal{E}(W_{\mathbf{h}})$

The particular form of $W_{\mathbf{h}}$ depends on the type of likelihood function used. In many applications $L$ for the acentric case has the form (Murshudov *et al.*, 1997; Bricogne, 1997)

$$\begin{aligned} L &= -\frac{1}{2} \sum_{\mathbf{h}} \log \int P(|F_{\mathbf{h}o}|, \varphi; F_{\mathbf{h}c})\, \mathrm{d}\varphi, \\ P(|F_{\mathbf{h}o}|, \varphi; F_{\mathbf{h}c}) &= \frac{1}{\pi \varepsilon \Sigma} \exp\left( -\frac{|F_{\mathbf{h}o} - F_{\mathbf{h}c}|^2}{\varepsilon \Sigma} \right) |F_{\mathbf{h}o}|, \end{aligned} \tag{11}$$

where $\varepsilon$ is the reflection multiplicity, $\Sigma = \mathcal{E}_{|F_{\mathbf{h}o}|,\varphi}(|F_{\mathbf{h}o} - F_{\mathbf{h}c}|^2)/\varepsilon$ is the conditional normalization coefficient and $\varphi$ is the phase. The notation $P(a; b)$ indicates the conditional probability for $a$ given $b$. In the above equation, $F_{\mathbf{h}c}$ is implicitly assumed to arise from the summation over all the available partial structures. It also incorporates the maximum-likelihood scale factor $D$.

The first derivatives of $L$ with respect to the real and imaginary parts of the calculated structure factor can be written as

$$\begin{aligned} \left( \frac{\partial L}{\partial A_{\mathbf{h}c}} - i \frac{\partial L}{\partial B_{\mathbf{h}c}} \right) &= -\frac{\mathcal{E}_{\varphi}[(F_{\mathbf{h}o} - F_{\mathbf{h}c})^*]}{\varepsilon \Sigma} \\ &= -\frac{\int P(\varphi; |F_{\mathbf{h}o}|, F_{\mathbf{h}c})(F_{\mathbf{h}o} - F_{\mathbf{h}c})^*\, \mathrm{d}\varphi}{\varepsilon \Sigma}, \end{aligned} \tag{12}$$

where we have used the fact that

$$P(\varphi; |F_{\mathbf{h}o}|, F_{\mathbf{h}c}) = \frac{P(|F_{\mathbf{h}o}|, \varphi; F_{\mathbf{h}c})}{P(|F_{\mathbf{h}o}|; F_{\mathbf{h}c})}. \tag{13}$$

As the integral $\int P(\varphi; |F_{\mathbf{h}o}|, F_{\mathbf{h}c})\,\mathrm{d}\varphi$ is equal to one, (12) can be rewritten as

$$\left(\frac{\partial L}{\partial A_{\mathbf{h}c}} - i\frac{\partial L}{\partial B_{\mathbf{h}c}}\right) = -\frac{|F_{\mathbf{h}o}| \int P(\varphi; |F_{\mathbf{h}o}|, F_{\mathbf{h}c}) \exp(-i\varphi)\mathrm{d}\varphi - (F_{\mathbf{h}c})^*}{\varepsilon\Sigma}. \tag{14}$$

(14) brings the calculation of the first derivatives to the calculation of the expected value of $\exp(i\varphi)$ given $F_{\mathbf{h}o}$ and $F_{\mathbf{h}c}$.

From (10) and (12) we have

$$\mathcal{E}(W_{\mathbf{h}}) = \frac{\mathcal{E}_{|F_{\mathbf{h}o}|}[|\mathcal{E}_{\varphi}(F_{\mathbf{h}o} - F_{\mathbf{h}c})|^2]}{(\varepsilon\Sigma)^2} \tag{15}$$

and remembering the expression for $\Sigma$ we can then write $\mathcal{E}(W_{\mathbf{h}})$ as

$$\mathcal{E}(W_{\mathbf{h}}) = \frac{\mathcal{E}_{|F_{\mathbf{h}o}|}[|\mathcal{E}_{\varphi}(F_{\mathbf{h}o} - F_{\mathbf{h}c})|^2]}{[\mathcal{E}_{|F_{\mathbf{h}o}|,\varphi}(|F_{\mathbf{h}o} - F_{\mathbf{h}c}|^2)]^2}. \tag{16}$$

In the above expressions for $\mathcal{E}(W_{\mathbf{h}})$, the term $\Sigma$ has been taken out of the expectation operator. In principle, this is legitimate as $\Sigma$ depends on the distribution of $(F_{\mathbf{h}o} - F_{\mathbf{h}c})$ but not on the observed structure factors themselves. In practice, however, $\Sigma$ is estimated by maximizing its likelihood given the observed reflections. If only free reflections are used, *i.e.* reflections not included in refinement (Brünger, 1992), then the dependence of $\Sigma$ on $F_{\mathbf{h}o}$ can be ignored. To employ all reflections, techniques such as the bootstrap method (Efron, 1979) or a marginal likelihood maximization (Lebedev *et al.*, 2003) need to be used.

(16) allows the appreciation of a practical advantage offered by the use of the Fisher's information. As the distribution of $(F_{\mathbf{h}o} - F_{\mathbf{h}c})$ is essentially the same within resolution shells, $\mathcal{E}(W_{\mathbf{h}})$ can be conveniently approximated by a one-dimensional function dependent only on the modulus of the scattering vector; that is, $\mathcal{E}(W_{\mathbf{h}}) \cong \mathcal{E}[W(s)]$. At present, $\mathcal{E}[W(s)]$ is calculated by numerical integration and fitted either by an exponential function or by a smoothening function with a Gaussian kernel depending on data quality. The latter type of function has been observed to work better for low-resolution data, at early stages of refinement, or when the number of free reflections is small. The exponential function is employed in all other cases. (16) also allows the conclusion that the knowledge of the likelihood itself is in principle not required for the estimation of the likelihood weighting factor. Only the first and the second moments of the differences between the observed and the calculated structure factors are needed.

An equation similar to (16) can be derived for intensity-based refinement. In this case, the form of the likelihood function is of the type (Pannu & Read, 1996)

$$L = -\frac{1}{2}\sum_{\mathbf{h}} \log \int\int P(I_{\mathbf{h}o}; F_{\mathbf{h}}) P(F_{\mathbf{h}}; F_{\mathbf{h}c})\ \mathrm{d}A_{\mathbf{h}}\mathrm{d}B_{\mathbf{h}},$$

$$P(F_{\mathbf{h}}; F_{\mathbf{h}c}) = \frac{1}{\pi\varepsilon\Sigma}\exp\left(-\frac{|F_{\mathbf{h}} - F_{\mathbf{h}c}|^2}{\varepsilon\Sigma}\right). \tag{17}$$

Following arguments similar to those leading to (16), $\mathcal{E}(W_{\mathbf{h}})$ for this type of likelihood can be expressed as

$$\mathcal{E}(W_{\mathbf{h}}) = \frac{\mathcal{E}_{I_{\mathbf{h}o}}[|\mathcal{E}_{F_{\mathbf{h}}}(F_{\mathbf{h}} - F_{\mathbf{h}c})|^2]}{[\mathcal{E}_{I_{\mathbf{h}o},F_{\mathbf{h}}}(|F_{\mathbf{h}} - F_{\mathbf{h}c}|^2)]^2}. \tag{18}$$

The approach presented in this section for the single-crystal case can be extended. An outline of a procedure applicable for multi-crystal refinement is given in Appendix *B*.

## 5. Integral approximation of Fisher's information

For practical purposes, equation (9) needs to be explicated further. Working out the derivatives of $F_{\mathbf{h}c}$ with respect to the various atomic parameters and indicating with $p_i(n)$ and $p_j(m)$ refinable parameters attached to atoms $n$ and $m$, respectively, $\mathcal{I}$ can be written as

$$\mathcal{I}[p_i(n), p_j(m)] = K_{p_ip_j} Q_{nm} \sum_{\mathbf{h}} \mathcal{E}[W(s)] H_{p_ip_j} F^o_{nm} T_{nm} \times \mathrm{trig}_{p_ip_j}(2\pi\mathbf{h}\mathbf{D}_{nm}), \tag{19}$$

where $Q_{nm}$, $F^o_{nm}$ and $T_{nm}$ are terms corresponding to the products of the occupancies, scattering factors at rest and thermal factors of atoms $n$ and $m$, respectively, whereas $K_{p_ip_j}$, $H_{p_ip_j}$ and $\mathrm{trig}_{p_ip_j}$, the latter of which represent a trigonometric function, are terms which depend on the particular combination of parameters considered as reported in Table 1. $\mathbf{D}_{nm}$ is the interatomic vector.

The direct calculation of $\mathcal{I}$ according to (19) is time-consuming. It requires a time which is proportional to the number of reflections ($N_{\mathrm{ref}}$) times the number of matrix elements ($N_{\mathrm{el}}$). A considerable improvement in speed can be achieved by using FFT methods (Agarwal, 1978; Murshudov *et al.*, 1997, 1999; Tronrud, 1999) or the fast differentiation algorithm proposed by Urzhumtsev & Lunin (2001) for the Hessian matrix. These methods require a number of operations ($c_1N_{\mathrm{el}} + c_2N_{\mathrm{ref}}\log N_{\mathrm{ref}}$), where $c_1$ and $c_2$ are constants, the values of which depend on the method considered. The dependence on the number of reflections and matrix elements can be reduced substantially, further improving the speed of the calculation. This can be performed by assuming that reciprocal-lattice points are dense enough that (19) can be legitimately made continuous (Agarwal, 1978; Dodson, 1981; Templeton, 1999). In mathematical terms, this is obtained by replacing the summation over the sphere defined by the measured reflections with an integration in reciprocal space,

$$\mathcal{I}[p_i(n), p_j(m)] \cong K_{p_ip_j} Q_{nm} \int_{\text{res. sphere}} \mathcal{E}[W(s)] H_{p_ip_j} F^o_{nm} T_{nm} \times \mathrm{trig}_{p_ip_j}(2\pi\mathbf{h}\mathbf{D}_{nm})\ \mathrm{d}h_1\ \mathrm{d}h_2\ \mathrm{d}h_3. \tag{20}$$

Elements of $\mathcal{I}$ can be evaluated very efficiently in a two-step procedure. In the first step, the *tabulation step*, a limited set of integrals are evaluated in a convenient coordinate system. In the second step, the *rotation step*, the tabulated values obtained in the first step and suitable rotation matrices are used to obtain the various elements of $\mathcal{I}$ in the original coordinate system.

**Table 1**
The explicit form of the terms $K_{p_ip_j}$, $H_{p_ip_j}$ and $\mathrm{trig}_{p_ip_j}$ in equation (19) for various types of matrix elements $\mathcal{I}(p_i, p_j)$.

The symbols $x_i$ and $U_{ij}$ indicate a positional and anisotropic atomic displacement parameter (ADP), respectively. The symbols $B$ and $q$ indicate the isotropic atomic displacement parameter and the occupancy, respectively. The symbol $h_i$ indicates a reflection index matching (in the sense of Laue) a certain parameter. For example, $p_ip_j = xz$ corresponds to $H_{p_ip_j} = hl$ and $p_ip_j = U_{11}U_{23}$ corresponds to $H_{p_ip_j} = h^2kl$. The symbol $s$ is the length of the scattering vector. The quantity $c_{ij}$ which is present when anisotropic ADPs are involved is a coefficient defined as $c_{ij} = 1$ if $i = j$ and $c_{ij} = 2$ if $i \neq j$, which takes into account the symmetry of the **U** tensor.

| $p_i$ | $p_j$ | $K_{p_ip_j}$ | $H_{p_ip_j}$ | $\mathrm{trig}_{p_ip_j}$ |
|---|---|---|---|---|
| $x_i$ | $x_j$ | $2\pi^2$ | $h_ih_j$ | cos |
| $B$ | $B$ | 1/32 | $s^4$ | cos |
| $U_{ij}$ | $U_{kl}$ | $2\pi^4c_{ij}c_{kl}$ | $h_ih_jh_kh_l$ | cos |
| $q$ | $q$ | 1 | 1 | cos |
| $x_i$ | $B$ | $\pi/4$ | $h_is^2$ | sin |
| $x_i$ | $U_{kl}$ | $2\pi^3c_{kl}$ | $h_ih_kh_l$ | sin |
| $x_i$ | $q$ | $\pi$ | $h_i$ | sin |
| $B$ | $U_{kl}$ | $c_{kl}/4$ | $s^2h_kh_l$ | cos |
| $B$ | $q$ | 1/8 | $s^2$ | cos |
| $U_{ij}$ | $q$ | $\pi^2c_{ij}$ | $h_ih_j$ | cos |

### 5.1. Tabulation step

Let us write out the expression of the product $F^o_{nm}T_{nm}$ in an explicit form. Using the five-Gaussian approximation for the atomic scattering factor at rest $f^o(n, s) = \sum_{u=1}^{5} a_u \exp(-b_us^2/4)$, $b_5 = 0$ and the isotropic description of the atomic thermal factor $t(n, s) = \exp[-B(n)s^2/4]$, this product can be expressed by the double summation[3]

$$F^o_{nm}T_{nm} = \sum_{u=1}^{5}\sum_{v=1}^{5} a_u(n)a_v(m) \times \exp\left\{-\frac{[b_u(n) + b_v(m) + B(n) + B(m)]s^2}{4}\right\}. \quad (21)$$

Equation (20) can therefore be written in the form

$$\mathcal{I}[p_i(n), p_j(m)] \cong \sum_{u=1}^{5}\sum_{v=1}^{5} K_{p_ip_j}Q_{nm}a_u(n)a_v(m) \times \int_{\text{res. sphere}} \mathcal{E}[W(s)]H_{p_ip_j}\exp\left[-\frac{B_{uv}(n, m)s^2}{4}\right] \times \mathrm{trig}_{p_ip_j}(2\pi\mathbf{h}\mathbf{D}_{nm})\, \mathrm{d}h_1\, \mathrm{d}h_2\, \mathrm{d}h_3, \quad (22)$$

where $B_{uv}(n, m) = b_u(n) + b_v(m) + B(n) + B(m)$.

Let $I_{p_ip_j}$ denote integrals of the form

$$I_{p_ip_j} = \int_{\text{res. sphere}} \mathcal{E}[W(s)]H_{p_ip_j}\exp(-Bs^2/4) \times \mathrm{trig}_{p_ip_j}(2\pi\mathbf{h}\mathbf{D})\, \mathrm{d}h_1\, \mathrm{d}h_2\, \mathrm{d}h_3. \quad (23)$$

If general tables for $I_{p_ip_j}$ were available over a range of $B$ and $D$ values wide enough to cover all $B_{uv}$ and $D_{nm}$, the evaluation of the various elements of $\mathcal{I}$ could be performed very efficiently using interpolation techniques.

The construction of such a family of tables requires the choice of a coordinate system. Using standard spherical polar coordinates ($\rho$, $\psi$, $\theta$) (see Appendix *C*), the tabulation of the integrals $I_{p_ip_j}$ can be conveniently performed in a coordinate system $x'$, $y'$, $z'$ oriented such that the $z'$ axis is parallel to the interatomic vector **D**.

Let $\mathbf{h}'$ represent **h** in this system. Integrals $I'_{p_ip_j}$ (the prime refers to the 'convenient' system) can be written as

$$I'_{p_ip_j} = \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)]\exp(-B\rho^2/4)\int_0^{\pi}\int_0^{2\pi} H'_{p_ip_j} \times \mathrm{trig}_{p_ip_j}(2\pi\rho D\cos\psi)J\, \mathrm{d}\theta\, \mathrm{d}\psi\, \mathrm{d}\rho, \quad (24)$$

where $J$ is the Jacobian for the coordinate transformation applied. How these integrals are evaluated is exemplified for positional parameters in Appendix *C*.

The use of the 'convenient' coordinate system substantially reduces the number of integrals to be tabulated, as most of the $I'_{p_ip_j}$ vanish. In the case of anisotropic refinement, for example, only 12 tables of integrals are required instead of 55. This is beneficial for the efficiency of the method, especially in the case of ML refinement. In least-squares refinement the weighting factor $\mathcal{E}(W_{\mathbf{h}})$ does not change while updating the model, whereas in ML refinement the weight is revised at each refinement cycle. ML refinement therefore requires that all tables are recalculated accordingly. Since the time taken by the tabulation step is proportional to the number of integrals $I'_{p_ip_j}$ that need to be evaluated, it is important to reduce it to the minimum.

### 5.2. Rotation step

In general, integrals $I'_{p_ip_j}$ are different from $I_{p_ip_j}$. They coincide only when $\mathbf{D}_{nm}$ is aligned with the orthogonal $z$ axis or in the case of $I_{BB}$, $I_{qq}$ and $I_{Bq}$ matrix elements. For them the equivalence holds because their $H$ term does not depend on the orientation (see Table 1). For all other elements, $H$ is orientation dependent.

To obtain expressions of $H_{p_ip_j}$ in terms of $H'_{p_ip_j}$, it is sufficient to consider that

$$\mathbf{h} = \mathbf{R}^T\mathbf{h}', \quad (25)$$

where **R** is the rotation matrix that transforms ($x$, $y$, $z$) to ($x'$, $y'$, $z'$). A way to construct the matrix **R** is given in Appendix *C*.

Using the relation (25) integrals, $I_{p_ip_j}$ can be calculated as a linear combination of the reduced set of tabulated quantities. For example, in the case of positional parameters

$$I_{x_ix_j} = \sum_{u=1}^{3}\sum_{v=1}^{3} t_{iu}t_{jv}I'_{x_u,x_v}, \quad (26)$$

where $t_{iu}$ are the elements of the $\mathbf{R}^T$ matrix. The number of summations that need to be performed depends on the type of parameter pair being evaluated. At most, in the case of $I_{U_{ij}U_{kl}}$ matrix elements, four summations are required.

An example of this procedure is given in Appendix *C*.

[3] Note that in the case of anisotropic refinement the isotropic approximation is used only for $T_{nm}$. The trigonometric terms in (19) and (20) are exact.

## 6. Results

### 6.1. Validity of the integral approximation

The validity of the integral approximation, *i.e.* the replacement of equation (19) by equation (20), relies on a distribution of reciprocal-lattice points dense enough to be considered continuous. This approximation is therefore expected to increase in accuracy the larger the molecule considered and the higher the resolution available. Such a dependency on the molecular size makes this method particularly attractive in the case of large macromolecular problems, which incidentally are those for which the calculation of even a sparse second-derivative matrix by other methods is prohibitively expensive.

In order to obtain a feeling of the limitations of the integral approximation, a number of tests have been carried out on the relatively small protein bovine pancreatic phospholipase $A_2$ (bpPLA2; 123 amino acids; one molecule in the asymmetric unit), for which complete data are available to 0.97 Å (Steiner *et al.*, 2001). A comparison of the values for positional and *BB* elements obtained by summation according to (19) with those obtained by integration using (20) has been performed using different resolution limits. As a measure of agreement within a distance limit $D_{\max}$ we have used the diagonal normalized quantity $A(\%)$ defined as

$$A_{p_ip_j}(\%) = \left\langle 100 \sum_{D=0}^{D_{\max}} \frac{|\mathcal{I}^{\rm sum}_{p_ip_j}(D) - \mathcal{I}^{\rm int}_{p_ip_j}(D)|}{[\mathcal{I}^{\rm sum}_{p_ip_i}(0)\mathcal{I}^{\rm sum}_{p_jp_j}(0)]^{1/2}} \right\rangle, \qquad (27)$$

where $\mathcal{I}^{\rm int}_{p_ip_j}$ represents an $\mathcal{I}$ element calculated with (20) and $\mathcal{I}^{\rm sum}_{p_ip_j}$ represents an $\mathcal{I}$ element calculated with (19). The notation $\langle x \rangle$ is used here to indicate the average of $x$. The diagonal normalization is used in order to bring the error $|\mathcal{I}^{\rm sum}_{p_ip_j}(D) - \mathcal{I}^{\rm int}_{p_ip_j}(D)|$ onto a scale relative to the elements of highest

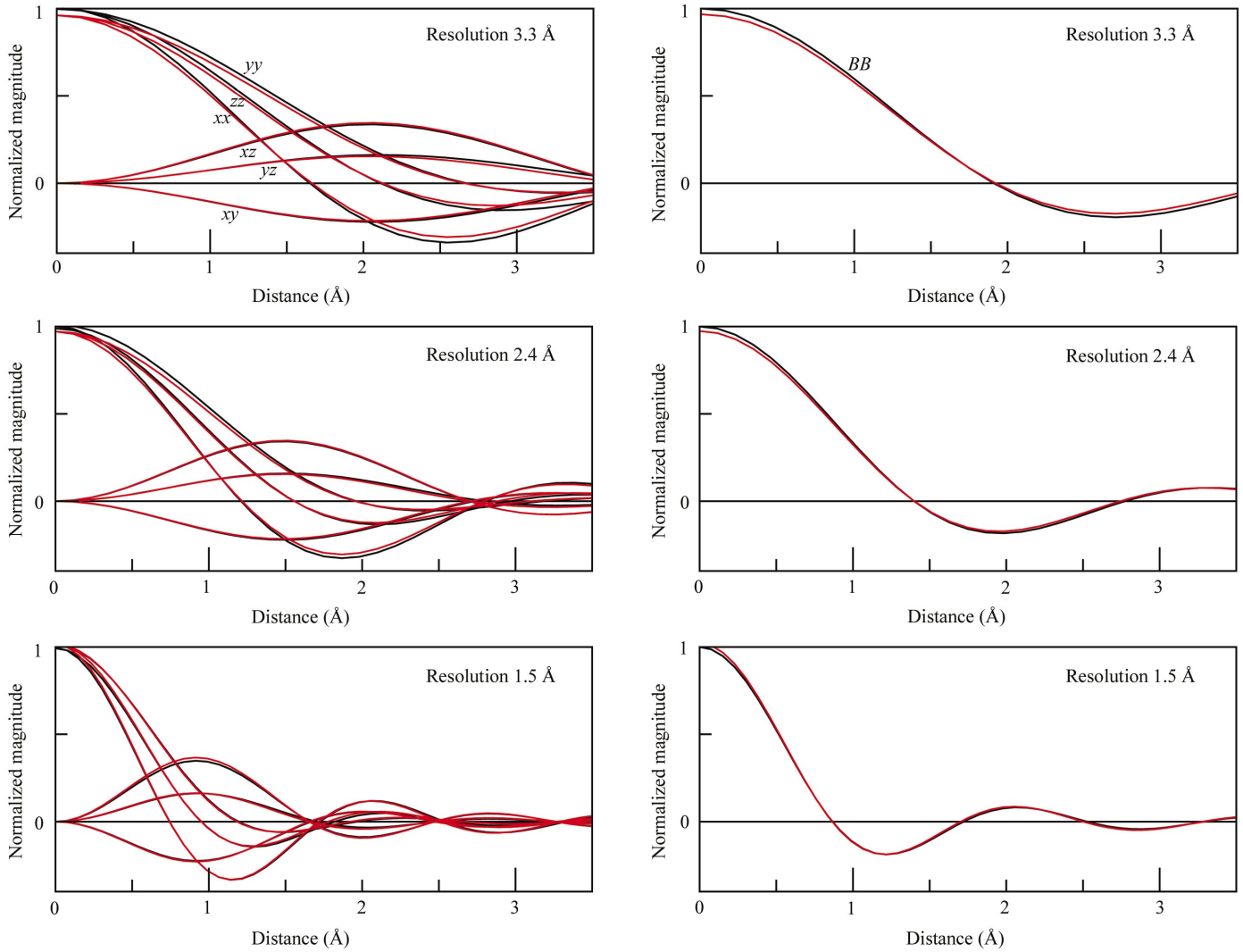


**Figure 1**
Comparison of normalized Fisher's information matrix elements for positional and *BB* parameters calculated with the summation method (black lines) and the integration method (red lines) as a function of the interatomic distance and resolution limit. The atoms forming the pair are an N atom and a C atom. The orientation of the interatomic vector corresponds to that of the vector $N-C^{\beta}$ of the first amino acid of the structure bpPLA2 (PDB code 1g4i; Steiner *et al.*, 2001). This vector has been chosen for display for its random orientation. The ADPs of the N atom and C atom are 7.26 and 7.94 $Å^2$, respectively.

magnitude which dominate the matrix (Agarwal, 1978; Tronrud, 1992; Jelsch, 2001). This type of normalization is similar to that employed for matrix conditioning prior to solving the system of linear equations (4). The best agreement is for $A(\%)$ equal zero.

Using a sample of 100 randomly oriented interatomic vectors and a distance cutoff $D_{\max}$ of 3.5 Å, a decrease in resolution from 1.0 to 3.3 Å causes $A_{xx}(\%)$ to increase from 0.2 to about 2.8. At the resolution of 2.4 Å, $A_{xx}(\%)$ is only 1.8. Comparable values are obtained for $A_{yy}(\%)$ (2.0) and $A_{zz}(\%)$ (1.2). Owing to the diagonal normalization, $A_{xy}(\%)$, $A_{xz}(\%)$ and $A_{yz}(\%)$ have much lower values. They are 0.3, 0.2 and 0.2, respectively. $A_{BB}(\%)$ is 1.1. The very good accord between the two methods can also be appreciated visually in Fig. 1. A further decrease in protein size does not worsen the statistics appreciably. In the case of human insulin-like growth factor (IGF), whose deposited structure (PDB code 1gzr; Brzozowski *et al.*, 2002) has 61 amino acids in the asymmetric unit, the average $A_{ii}(\%)$ and $A_{ij}(\%)$ indices at 2.4 Å are 2.2 and 0.3, respectively. $A_{BB}(\%)$ is 1.4. As expected, for a protein of larger size the agreement improves. In the case of the PknB Ser/Thr kinase from *Mycobacterium tubercolosis* (PDB code 1phk; 260 amino acids in the asymmetric unit) the average $A_{ii}(\%)$ and $A_{ii}(\%)$ at 2.4 Å are 0.9 and 0.2. $A_{BB}(\%)$ is 0.6.

Matrix elements calculated with the integration method therefore represent remarkably well those obtained by summation. This result supports the use of the fast integration method for the purpose of minimization.

### 6.2. Examples of refinement

The scoring algorithm which uses the sparse approximation of $\mathcal{I}$ has been implemented in the latest version of the program *REFMAC*5 (version 5.2.0). As in previous versions of the program, the solution of the system of linear equations (4) is carried out using a preconditioned conjugate-gradient method (Golub & Van Loan, 1996). This technique does not require the inversion of the matrix $\mathcal{I}$ to calculate the parameter shifts. At present, only $\mathcal{I}$ elements for positional and ADPs (isotropic, mixed or anisotropic) parameters are used in the refinement. In future versions, however, the list can easily be extended, since all equations for the calculation of the various elements have already been implemented. In principle, this method can also be adapted to multipole refinement (Coppens, 1997).

The algorithm for the calculation of the sparse $\mathcal{I}$ is very efficient. Using an Acer laptop model 603TER equipped with a Pentium III 550 MHz processor and 256 MB memory and the default options (see below) for $B$–$D$ tabulation, $\mathcal{I}$ elements are calculated on average in fractions of a second.

The range of $B$ values for tabulation is automatically determined by the program on the basis of the current range of ADPs and the chemical elements present in the structure. The maximum distance $D_{\max}$ can be chosen by the user within a radius which depends on the total number of atoms in the structure. Depending on the molecular size, in the current version distances up to $\leq$7.0 Å can be considered. This limit is large enough for minimization applications, as the magnitude of off-diagonal elements falls rapidly with the increase in the interatomic distances of their parent atoms (Fig. 1). Thus, off-

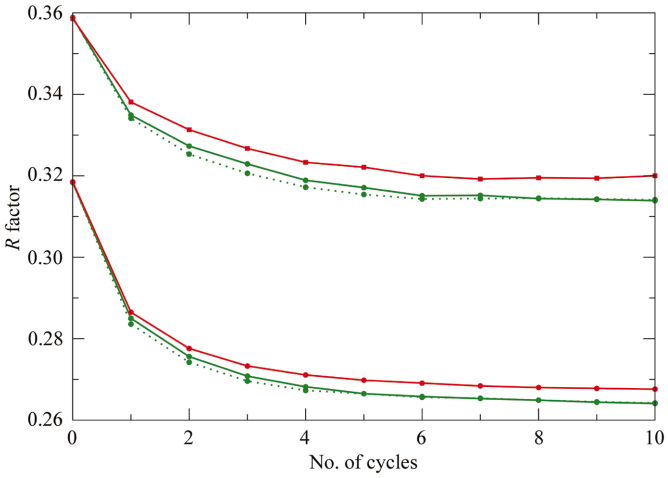


**Figure 2**
The behaviour of the $R$ statistics in identical parallel refinement runs of PknB (Ortiz-Lombardia *et al.*, 2003). The drops in $R$ and $R_{\rm free}$ obtained using the version 5.1.24 of the program *REFMAC*5 are shown by red circles and squares, respectively. The drops in $R$ and $R_{\rm free}$ achieved using the latest version of the program (5.2.0) are shown in by green circles and squares, respectively. The continuous line refers to the refinement carried out using only diagonal elements of $\mathcal{I}$. The dotted line refers to the refinement performed using the sparse approximation of $\mathcal{I}$.

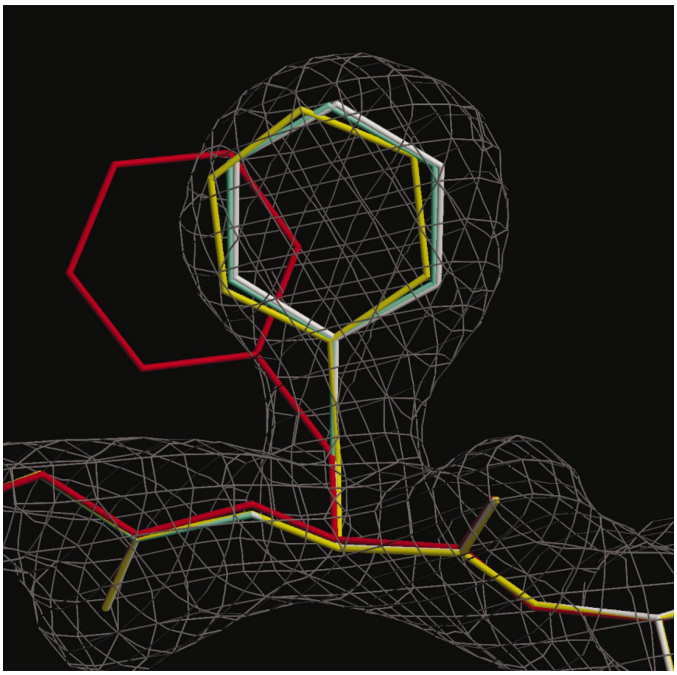

**Figure 3**
The effect of the inclusion of off-diagonal elements of $\mathcal{I}$ in the refinement of an aromatic residue. Phe23 of human IGF (Brzozowski *et al.*, 2002) was positioned out of density (shown in red). 15 cycles of positional refinement using only the diagonal elements of the matrix led to the structure shown in yellow. The same number of cycles with the inclusion of the off-diagonal elements of $\mathcal{I}$ up to 1.6 Å and up to 3.5 Å produced the structures displayed in cyan and white, respectively. Electron density is shown at the 1.5$\sigma$ level.

diagonal matrix arising from atoms far from each other can in practice be safely neglected. Currently, the default value is set at 3.5 Å. This cutoff value allows the correlations between bonded atoms, which are the most significant, to be taken into consideration and also the correlations between atoms related by non-bonding interactions.

Tests on various structures show that the latest version of the program performs generally better than version 5.1.24, which used an approximation of the diagonal Hessian. For example, the behaviour of the $R$ statistics in the case of the restrained refinement of PknB using data extending to 1.95 Å resolution is shown in Fig. 2. The starting model for refinement was a PknB model taken at an intermediate stage of refinement after a round of model rebuilding. The initial $R$ factor and $R_{\rm free}$ for this structure were 0.318 and 0.359, respectively. In this case, version 5.2.0 using the diagonal approximation of $\mathcal{I}$ gives in ten cycles an improvement in $R$ factor/$R_{\rm free}$ of 0.35/0.6% compared with version 5.1.24. The final $R$ factor and $R_{\rm free}$ are 0.264 and 0.314, respectively. The effect of the inclusion of off-diagonal $\mathcal{I}$ in the refinement has been tested carrying out parallel refinement runs with the latest version of the program using either only the diagonal elements of $\mathcal{I}$ or the default sparse $\mathcal{I}$ matrix. Although the improvements attainable with the inclusion of the off-diagonal elements are intrinsically limited owing to strong diagonality of the matrix, we have observed that the use of the sparse approximation marginally improves the convergence of the refinement (Fig. 2). At the early stages, the use of off-diagonal matrix elements produces better agreement statistics for a given number of cycles of refinement.

The rate of convergence is correlated to the amount of off-diagonal elements employed. To show this effect, residue Phe23 of IGF was displaced from its position by adding a 45° rotation to both $\chi_1$ and $\chi_2$ (in red in Fig. 3). 15 cycles of positional refinement were then performed by using only the diagonal elements of $\mathcal{I}$, the diagonal elements plus the off-diagonal elements with a cutoff value $D_{\rm max}$ of 1.6 Å and the diagonal elements plus the off-diagonal elements with a cutoff value of 3.5 Å. Fig. 3 shows that the refinement which uses the 3.5 Å cutoff has converged within the number of minimization cycles (in white), whereas that which uses only the diagonal has not yet brought the residue into density (in yellow). The use of less off-diagonal elements (1.6 Å cutoff) gives an intermediate side-chain position (in cyan) which is close to the converged one.

## 7. Conclusions and future developments

Fisher's information $\mathcal{I}$ and the related scoring method of minimization have been applied to maximum-likelihood macromolecular crystallographic refinement and implemented in the program *REFMAC*5. The minimization algorithm performs satisfactorily. The sparse-matrix approximation of $\mathcal{I}$ is currently used.

The present work has an immediate development. This relates to the concept of the Fisher's information in statistics as an estimation indicator. The Cramer–Rao inequality (Rao, 1945; Cramer, 1946) states the inverse proportionality between $\mathcal{I}$ and the variance $\mathrm{var}(\hat{t})$ of the unbiased estimate $\hat{t}$ of the parameter $t$,

$$\mathcal{I} \geq \frac{1}{\mathrm{var}(\hat{t})}. \qquad (28)$$

This relation[4] explains why $\mathcal{I}$ is said to be a measure of information. Its value increases as the estimation quality increases. At the validation stage, Fisher's information therefore represents an adequate statistical tool for the estimation of errors associated with crystallographic parameters. The routine estimation of parameter error requires the fast estimation of both $\mathcal{I}$ and of its inverse. In this paper, we have shown how $\mathcal{I}$ can be efficiently calculated. Efforts now need to be put into finding rapid ways of calculating $\mathcal{I}^{-1}$ and into the assessment of how the approximations introduced by the integration algorithm are propagated in the inversion or pseudo-inversion step. With respect to the latter point, different inversion procedures will perform differently. Numerical tests will be required to evaluate the best algorithm.

Fisher's information can also play a role in devising a correct weighting scheme in macromolecular refinement. Lebedev *et al.* (2003) have used the full covariance matrix to show that the weights between the X-ray and the geometrical terms can be estimated using an approximation of a marginal likelihood function. A practical application of this technique can be sought having a good approximation of the full $\mathcal{I}$ to hand.

Owing to the extremely low CPU requirements of our method of calculation of $\mathcal{I}$, second-order minimization could also be applied to real-space refinement used by molecular-graphics applications. This can be performed because owing to Parseval's theorem the minimization in real-space can be performed using the derivatives with respect to the reciprocal-space residual (Diamond, 1971). The application to real-space refinement is in principle even simpler and faster than that in reciprocal-space, as $\mathcal{E}(W_{\mathbf{h}})$ is constant. This eliminates the disadvantage of table update.

## APPENDIX *A*
## An equivalent expression for $W_{\mathbf{h}}$

The definition of $\mathcal{I}$ according to (3) allows the derivation of an alternative expression for $W_{\mathbf{h}}$. If the likelihood has the form

$$L = -\frac{1}{2}\sum_{\mathbf{h}} \log P(|F_{\mathbf{h}o}|; F_{\mathbf{h}c}) = \frac{1}{2}\sum\nolimits_{\mathbf{h}} L_{\mathbf{h}}, \qquad (29)$$

where $L_{\mathbf{h}}$ depends only on the reflection $\mathbf{h}$, $F_{\mathbf{h}} = F^*_{-\mathbf{h}}$ and $L_{\mathbf{h}} = L_{-\mathbf{h}}$, then by applying the chain rule twice the second-derivative matrix can be written as

[4] Equation (28) is valid for a mono-dimensional function. An extension for the multivariate case can be found in Stuart *et al.* (1999).

$$\frac{\partial^2 L}{\partial p_i(n)\partial p_j(m)} = \delta_{nm}\sum_{\mathbf{h}}\left(\frac{\partial L}{\partial A_{\mathbf{h}c}} - i\frac{\partial L}{\partial B_{\mathbf{h}c}}\right)\left[\frac{\partial^2 F_{\mathbf{h}c}}{\partial p_i(n)\partial p_j(m)}\right] + \frac{1}{2}\sum_{\mathbf{h}}\left(\frac{\partial^2 L}{\partial A_{\mathbf{h}c}^2} + \frac{\partial^2 L}{\partial B_{\mathbf{h}c}^2}\right)\left[\frac{\partial F_{\mathbf{h}c}}{\partial p_i(n)}\right]\left[\frac{\partial F_{\mathbf{h}c}}{\partial p_j(m)}\right]^* + \frac{1}{2}\sum_{\mathbf{h}}\left(\frac{\partial^2 L}{\partial A_{\mathbf{h}c}^2} - \frac{\partial^2 L}{\partial B_{\mathbf{h}c}^2} - 2i\frac{\partial^2 L}{\partial A_{\mathbf{h}c}\partial B_{\mathbf{h}c}}\right) \times \left[\frac{\partial F_{\mathbf{h}c}}{\partial p_i(n)}\right]\left[\frac{\partial F_{\mathbf{h}c}}{\partial p_j(m)}\right], \quad (30)$$

where $\delta_{nm}$ is the Kronecker delta.

The expected value of the first term is zero. That of the last term is small, as it depends on the phases. $\mathcal{I}$ can therefore be approximated by

$$\mathcal{I}(p_i, p_j) \cong \frac{1}{2}\sum_{\mathbf{h}}\mathcal{E}\left(\frac{\partial^2 L}{\partial A_{\mathbf{h}c}^2} + \frac{\partial^2 L}{\partial B_{\mathbf{h}c}^2}\right)\left(\frac{\partial F_{\mathbf{h}}}{\partial p_i}\right)\left(\frac{\partial F_{\mathbf{h}}}{\partial p_j}\right)^*. \quad (31)$$

From equations (2) and (3), we obtain the equivalence

$$\mathcal{E}\left(\left|\frac{\partial L_{\mathbf{h}}}{\partial A_{\mathbf{h}c}} - i\frac{\partial L_{\mathbf{h}}}{\partial B_{\mathbf{h}c}}\right|^2\right) = \mathcal{E}\left(\frac{\partial^2 L_{\mathbf{h}}}{\partial A_{\mathbf{h}c}^2} + \frac{\partial^2 L_{\mathbf{h}}}{\partial B_{\mathbf{h}c}^2}\right). \quad (32)$$

Thus, according to (29), the approximation (31) coincides with that given by (9) and (10). Note that $[(\partial^2 L_{\mathbf{h}}/\partial A_{\mathbf{h}c}^2) + (\partial^2 L_{\mathbf{h}}/\partial B_{\mathbf{h}c}^2)]$ can be either positive or negative, whereas its expectation is always non-negative.

## APPENDIX *B*
## Extension of $\mathcal{I}$ to multi-crystal refinement

Here, we outline a method to use the Fisher's information in the case of observations originating from more than one crystal. For example, it could be applied to MAD/MIR cases or unliganded/liganded cases.

Let us assume that the form of the likelihood function is

$$L = -\frac{1}{2}\sum_{\mathbf{h}}\log P(\mathbf{I}_{\mathbf{h}o}; \mathbf{F}_{\mathbf{h}c}) = \frac{1}{2}\sum_{\mathbf{h}} L_{\mathbf{h}}, \quad (33)$$

where $\mathbf{I}_{\mathbf{h}o}$ and $\mathbf{F}_{\mathbf{h}c}$ are random vectors of observations and calculated structure factors, respectively. Let also assume the invariance $L_{\mathbf{h}} = L_{-\mathbf{h}}$ and that $\mathbf{F}_{\mathbf{h}} = \mathbf{F}_{-\mathbf{h}}^*$.

The first derivative of this function with respect to the model parameters can be expressed as the sum of the scalar products

$$\frac{\partial L}{\partial p_i} = \sum_{\mathbf{h}}\left(\frac{\partial L}{\partial \mathbf{A}_{\mathbf{h}c}} - i\frac{\partial L}{\partial \mathbf{B}_{\mathbf{h}c}}\right)\frac{\partial \mathbf{F}_{\mathbf{h}c}}{\partial p_i}. \quad (34)$$

This equation can be used to calculate the derivatives of the likelihood with respect to the atomic parameters using FFT as described for the single-crystal case by Agarwal (1978). In the multi-crystal case, the number of FFTs required is equal at most to the number of crystals treated.

If the first derivatives of the likelihood function with respect to the real and imaginary parts of the structure factors are uncorrelated for different reflections, then we can write Fisher's information in the quadratic forms

$$\mathcal{E}\left(\frac{\partial L}{\partial p_i}\frac{\partial L}{\partial p_j}\right) = \sum_{\mathbf{h}}\left(\frac{\partial \mathbf{F}_{\mathbf{h}c}}{\partial p_j}\right)^*\mathcal{E}\left[\left(\frac{\partial L}{\partial \mathbf{A}_{\mathbf{h}c}} - i\frac{\partial L}{\partial \mathbf{B}_{\mathbf{h}c}}\right)^* \times \left(\frac{\partial L}{\partial \mathbf{A}_{\mathbf{h}c}} - i\frac{\partial L}{\partial \mathbf{B}_{\mathbf{h}c}}\right)\right]\left(\frac{\partial \mathbf{F}_{\mathbf{h}c}}{\partial p_i}\right) + \sum_{\mathbf{h}}\left(\frac{\partial \mathbf{F}_{\mathbf{h}c}}{\partial p_j}\right)\mathcal{E}\left[\left(\frac{\partial L}{\partial \mathbf{A}_{\mathbf{h}c}} - i\frac{\partial L}{\partial \mathbf{B}_{\mathbf{h}c}}\right) \times \left(\frac{\partial L}{\partial \mathbf{A}_{\mathbf{h}c}} - i\frac{\partial L}{\partial \mathbf{B}_{\mathbf{h}c}}\right)\right]\left(\frac{\partial \mathbf{F}_{\mathbf{h}c}}{\partial p_i}\right), \quad (35)$$

where $p_i$ and $p_j$ are parameters belonging to one or different crystals. Owing to the presence of the phase term, the second summation term can be predicted to be very small.

Let us now analyse $[(\partial L/\partial A_{\mathbf{h}c}) - i(\partial L/\partial B_{\mathbf{h}c})]$. Most likelihood functions for multi-crystal cases have the form (here we use the assumption that the intensities but not the amplitudes of the structure factors are the observables)

$$P(\mathbf{I}_{\mathbf{h}o}; \mathbf{F}_{\mathbf{h}c}) = (\pi^k|\varepsilon\Sigma|)^{-1}\int\ldots\int P(\mathbf{I}_{\mathbf{h}o}, \mathbf{F}_{\mathbf{h}}; \mathbf{F}_{\mathbf{h}c}) \times \exp[-(\mathbf{F}_{\mathbf{h}} - \mathbf{F}_{\mathbf{h}c})^{*T}(\varepsilon\Sigma)^{-1}(\mathbf{F}_{\mathbf{h}} - \mathbf{F}_{\mathbf{h}c})] \times \mathrm{d}A_1\ \mathrm{d}B_1\ldots\mathrm{d}A_k\ \mathrm{d}B_k, \quad (36)$$

the derivatives then become

$$\left(\frac{\partial L}{\partial \mathbf{A}_{\mathbf{h}c}} - i\frac{\partial L}{\partial \mathbf{B}_{\mathbf{h}c}}\right) = (\varepsilon\Sigma)^{-1}\mathcal{E}_{\mathbf{F}_{\mathbf{h}}}(\mathbf{F}_{\mathbf{h}} - \mathbf{F}_{\mathbf{h}c}), \quad (37)$$

where the expectation is taken using the probability distribution of $P(\mathbf{F}_{\mathbf{h}}; I_{\mathbf{h}o}, F_{\mathbf{h}c})$. In the above equation $[(\partial L/\partial \mathbf{A}_{\mathbf{h}c}) - i(\partial L/\partial \mathbf{B}_{\mathbf{h}c})]$ is a vector and the right-hand side should be understood as a matrix-to-vector multiplication.

Thus, the derivative of the likelihood for each structure factor is a linear combination of the moments for the various crystals. If the first moments can be evaluated, then the derivatives can be calculated as a linear combination and $\mathcal{I}$ can be calculated using set of resolution-dependent functions. The actual calculation of the moments is outside the scope of this work and will not be considered here.

It is worth noting that although the above equations are not valid when structure factors arise from anomalous scatterers, *i.e.* for combinations of $F_+$ and $F_-^*$, they can equally be used with suitable modifications. The derivatives can be taken with respect to $F$ and $F''$. Then, using the fact that $F_+$ and $F_-^*$ are linear combination of $F$ and $F''$ ($F_+ = F + iF''$, $F_-^* = F - iF''$), their derivatives can also be written.

## APPENDIX *C*
## Fast calculation procedure

In this appendix, some details of the method used for the fast calculation of Fisher's information matrix elements as integrals are given. As an example, positional elements will be used. For any other type of element, the necessary equations can be derived by analogy.

### C1. Tabulation step

Reflection indices in the rotated coordinate system defined in §5.1 are expressed in terms of spherical polar coordinates according to

$$\begin{aligned} h'_1 &= \rho \sin\psi \cos\theta \\ h'_2 &= \rho \sin\psi \sin\theta \\ h'_3 &= \rho \cos\psi, \end{aligned} \tag{38}$$

where

$$\begin{aligned} 0 &\le \rho < \infty \\ 0 &\le \psi \le \pi \\ 0 &\le \theta \le 2\pi. \end{aligned} \tag{39}$$

Considering that the Jacobian $J$ for the coordinate transformation is

$$J = \rho^2 \sin\psi, \tag{40}$$

$I'_{x_i x_j}$ integrals of equation (24) can be written as

$$I'_{xx} = \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)] \exp(-B\rho^2/4)\rho^4 \int_0^{\pi}\int_0^{2\pi} \sin^2\psi \cos^2\theta \times \cos(2\pi\rho D\cos\psi)\sin\psi \; \mathrm{d}\theta \; \mathrm{d}\psi \; \mathrm{d}\rho, \tag{41}$$

$$I'_{yy} = \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)] \exp(-B\rho^2/4)\rho^4 \int_0^{\pi}\int_0^{2\pi} \sin^2\psi \sin^2\theta \times \cos(2\pi\rho D\cos\psi)\sin\psi \; \mathrm{d}\theta \; \mathrm{d}\psi \; \mathrm{d}\rho, \tag{42}$$

$$I'_{zz} = \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)] \exp(-B\rho^2/4)\rho^4 \int_0^{\pi}\int_0^{2\pi} \cos^2\psi \times \cos(2\pi\rho D\cos\psi)\sin\psi \; \mathrm{d}\theta \; \mathrm{d}\psi \; \mathrm{d}\rho, \tag{43}$$

$$I'_{xy} = \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)] \exp(-B\rho^2/4)\rho^4 \int_0^{\pi}\int_0^{2\pi} \sin^2\psi \sin\theta\cos\theta \times \cos(2\pi\rho D\cos\psi)\sin\psi \; \mathrm{d}\theta \; \mathrm{d}\psi \; \mathrm{d}\rho, \tag{44}$$

$$I'_{xz} = \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)] \exp(-B\rho^2/4)\rho^4 \int_0^{\pi}\int_0^{2\pi} \sin\psi\cos\psi\cos\theta \times \cos(2\pi\rho D\cos\psi)\sin\psi \; \mathrm{d}\theta \; \mathrm{d}\psi \; \mathrm{d}\rho, \tag{45}$$

$$I'_{yz} = \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)] \exp(-B\rho^2/4)\rho^4 \int_0^{\pi}\int_0^{2\pi} \sin\psi\cos\psi\sin\theta \times \cos(2\pi\rho D\cos\psi)\sin\psi \; \mathrm{d}\theta \; \mathrm{d}\psi \; \mathrm{d}\rho. \tag{46}$$

An inspection of the above integrals gives that (42) is equal to (41) and that (44), (45) and (46) are zero. In the case of positional elements, therefore, the rotation of the coordinate system reduces the number of independent elements from six to two. Similar simplifications are obtained for the other elements.

Analytic integration over $\mathrm{d}\theta$ and $\mathrm{d}\psi$ simplifies the above non-zero expressions to

$$I'_{xx} = I'_{yy} = 4\pi \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)]\rho^4 \exp(-B\rho^2/4) \times \left[\frac{\sin(2\pi\rho D)}{(2\pi\rho D)^3} - \frac{\cos(2\pi\rho D)}{(2\pi\rho D)^2}\right] \mathrm{d}\rho \tag{47}$$

and

$$I'_{zz} = 4\pi \int_{s_{\min}}^{s_{\max}} \mathcal{E}[W(\rho)]\rho^4 \exp(-B\rho^2/4) \times \left\{\frac{[(2\pi\rho D)^2 - 2]\sin(2\pi\rho D)}{(2\pi\rho D)^3} + \frac{2\cos(2\pi\rho D)}{(2\pi\rho D)^2}\right\} \mathrm{d}\rho. \tag{48}$$

These integrals are evaluated numerically in the chosen $B$–$D$ grid.

### C2. Rotation step

**C2.1. The rotation matrix**. As discussed in §5.2, in order to calculate the various matrix elements for a particular atom pair $mn$ the explicit form of the rotation matrix $\mathbf{R}$ which brings the $z$ axis parallel to the interatomic vector $\mathbf{D}_{mn}$ is needed.

This matrix can be defined by the Euler angles $\Phi$, $\Theta$ and $\Psi$ where $\Phi$, $\Theta$ and $\Psi$ represent the angles corresponding to sequential anticlockwise rotations around $z$, $y'$ (the $y$ axis after the first rotation) and $x''$ (the $x$ axis after the second rotation), respectively.

Such a matrix can be constructed from the angles $\alpha$, $\beta$ and $\gamma$ formed by $\mathbf{D}_{mn}$ with the directions of the $x$, $y$ and $z$ axes, respectively, considering that

$$\Phi = \begin{cases} \arcsin(\cos\beta/\sin\gamma) & \text{if } \cos\alpha \ge 0 \\ -\arcsin(\cos\beta/\sin\gamma) & \text{if } \cos\alpha < 0 \end{cases},$$

$$\Theta = \begin{cases} \gamma & \text{if } \cos\alpha \ge 0 \\ -\gamma & \text{if } \cos\alpha < 0 \end{cases}.$$

The angle $\Psi$ can take any value since there is no restriction on the position of the $xy$ plane. Let, for convenience, $\Psi = 0$. The rotation matrix has then the form

$$\mathbf{R} = \begin{pmatrix} \cos\Theta\cos\Phi & \cos\Theta\sin\Phi & -\sin\Theta \\ -\sin\Phi & \cos\Phi & 0 \\ \sin\Theta\cos\Phi & \sin\Theta\sin\Phi & \cos\Theta \end{pmatrix}.$$

**C2.2. Evaluation of elements in the original setting**. Let $X$ and $Z$ be the integrals in (47) and (48), respectively. If $t_{x_i x_j}$ denotes an element of the matrix $\mathbf{R}^T$, application of (26) leads to

$$
\begin{aligned}
I_{xx} &= (t_{11}^2 + t_{12}^2)X + t_{13}^2 Z \\
I_{yy} &= (t_{21}^2 + t_{22}^2)X + t_{23}^2 Z \\
I_{zz} &= (t_{31}^2 + t_{32}^2)X + t_{33}^2 Z \\
I_{xy} &= (t_{11}t_{21} + t_{12}t_{22})X + t_{13}t_{23} Z \\
I_{xz} &= (t_{11}t_{31} + t_{12}t_{32})X + t_{13}t_{33} Z \\
I_{yz} &= (t_{21}t_{31} + t_{22}t_{32})X + t_{23}t_{33} Z \\
I_{yx} &= I_{xy} \\
I_{zx} &= I_{xz} \\
I_{zy} &= I_{yz}. \qquad (49)
\end{aligned}
$$

In matrix form, the same set of equations can also be written as

$$I = \mathbf{R}^T I' \mathbf{R}. \qquad (50)$$

RAS and GNM thank the Wellcome Trust for financial support. AAL was funded by the BBSRC (contract No. 87/B17320). The authors thank Dr Miguel Ortiz-Lombardia for providing the coordinates and structure factors of PknB and the York Structural Biology Laboratory for the stimulating research environment.

## References


Agarwal, R. C. (1978). *Acta Cryst.* A**34**, 791–809.

Barrett, H. H., Denny, J. L., Wagner, R. F. & Myers, K. J. (1995). *J. Opt. Soc. Am. A*, **12**, 834–852.

Bricogne, G. (1993). *International Tables for X-ray Crystallography*, Vol. B, edited by U. Shmueli, pp. 25–106. Dordrecht: Kluwer Academic Publishers.

Bricogne, G. (1997). *Methods Enzymol.* **276**, 361–423.

Bricogne, G. & Irwin, J. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 85–92. Warrington: Daresbury Laboratory.

Brünger, A. T. (1989). *Acta Cryst.* A**45**, 42–50.

Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.

Brzozowski, A., Dodson, E. J., Dodson, G. G., Murshudov, G. N., Verma, C., Turkenburg, J. P., de Bree, F. & Dauter, Z. (2002). *Biochemistry*, **41**, 9389–9397.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding.* Oxford: IUCr/Oxford Science Publications.

Cramer, H. (1946). *Mathematical Models of Statistics.* Princeton, NJ, USA: Princeton University Press.

Diamond, R. (1971). *Acta Cryst.* A**27**, 436–452.

Dodson, E. J. (1981). *Proceeding of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 29–39. Warrington: Daresbury Laboratory.

Efron, B. (1979). *Ann. Stat.* **7**, 1–26.

Fisher, R. A. (1922). *Philos. Trans. R. Soc. London Ser. A*, **222**, 309–368.

Fletcher, R. (1987). *Practical Methods of Optimization.* Chichester: John Wiley & Sons.

Frieden, B. R. (1998). *Physics From Fisher Information. A Unification.* Cambridge University Press.

Frieden, B. R., Plastino, A. & Soffer, B. H. (2001). *J. Theor. Biol.* **208**, 49–64.

Golub, G. H. & Van Loan, C. H. (1996). *Matrix Computations.* Baltimore, MD, USA: Johns Hopkins University Press.

Jelsch, C. (2001). *Acta Cryst.* A**57**, 558–570.

Kale, B. K. (1961). *Biometrika*, **48**, 452–455.

Klein, A. & Melard, G. (1995). *J. Comput. Appl. Math.* **64**, 57–68.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

Lebedev, A. A., Tickle, I. J., Laskowski, R. A. & Moss, D. S. (2003). *Acta Cryst.* D**59**, 1557–1566.

Lee, P. M. (1997). *Bayesian Statistics. An Introduction.* London: Arnold.

Lunin, V. & Urzhumtsev, A. G. (1984). *Acta Cryst.* A**40**, 269–277.

Lunin, V. Y. & Urzhumtsev, A. G. (1985). *Acta Cryst.* A**41**, 327–333.

McKie, D. & McKie, C. (1986). *Essentials Of Crystallography.* Oxford: Blackwell.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.

Ortiz-Lombardia, M., Pompeo, F., Boitel, B. & Alzari, P. M. (2003). *J. Biol. Chem.* **278**, 13094–13100.

Osborne, M. R. (1992). *Int. Stat. Rev.* **60**, 99–117.

Pannu, N. S., Murshudov, G. N., Dodson, E. J. & Read, R. J. (1998). *Acta Cryst.* D**54**, 1285–1294.

Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* A**52**, 659–668.

Parra, L. & Barrett, H. H. (1998). *IEEE Trans. Med. Imaging*, **17**, 228–235.

Rao, C. R. (1945). *Bull. Calcutta Math. Soc.* **37**, 81–91.

Read, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

Read, R. J. (1990). *Acta Cryst.* A**46**, 900–912.

Read, R. J. (2001). *Acta Cryst.* D**57**, 1373–1382.

Retout, S., Duffull, S. & Mentre, F. (2001). *Comput. Methods Programs Biomed.* **65**, 141–151.

Smyth, G. K. (1996). *Stat. Comput.* **6**, 201– 216.

Steiner, R. A., Rozeboom, H. J., de Vries, A., Kalk, K. H., Murshudov, G. N., Wilson, K. S. & Dijkstra, B. W. (2001). *Acta Cryst.* D**57**, 516–526.

Stuart, A., Ord, J. K. & Arnold, S. (1999). *Kendall's Advanced Theory of Statistics*, Vol. 2A. London: Arnold.

Templeton, D. H. (1999). *Acta Cryst.* A**55**, 695–699.

Terwilliger, T. (2000). *Acta Cryst.* D**56**, 965–972.

Terwilliger, T. (2001). *Acta Cryst.* D**57**, 1755–1762.

Tod, M., Mentre, F., Merle, Y. & Mallet, A. (1998). *J. Pharmokinet. Biopharm.* **26**, 689–716.

Tronrud, D. E. (1992). *Acta Cryst.* A**48**, 912–916.

Tronrud, D. E. (1999). *Acta Cryst.* A**55**, 700–703.

Urzhumtsev, A. G. & Lunin, V. Y. (2001). *Acta Cryst.* A**57**, 451–460.