

Refinement of Macromolecular Structures by the Maximum-Likelihood Method

GARIB N. MURSHUDOV^a ALEXEI A. VAGIN^b AND ELEANOR J. DODSON^{a*}

^aChemistry Department, University of York, Heslington, York, England, and ^bUCMB-ULB, Free University of Brussels, avenue Paul Heger cp160/16 – P21050 Brussels, Belgium. E-mail: ccp4@yorvic.york.ac.uk

(Received 31 May 1996; accepted 24 October 1996)

Abstract

This paper reviews the mathematical basis of maximum likelihood. The likelihood function for macromolecular structures is extended to include prior phase information and experimental standard uncertainties. The assumption that different parts of a structure might have different errors is considered. A method for estimating σ_A using 'free' reflections is described and its effects analysed. The derived equations have been implemented in the program *REFMAC*. This has been tested on several proteins at different stages of refinement (bacterial α -amylase, cytochrome *c'*, cross-linked insulin and oligopeptide binding protein). The results derived using the maximum-likelihood residual are consistently better than those obtained from least-squares refinement.

1. Notations

I_h^o and I_h^c	Experimental and calculated intensities of the structure factor for reflection h . The subscript h will usually be omitted.
$ F_h^o $	Experimental amplitude of the structure factor
σ_{I^o}	Experimental uncertainty in intensity
σ_{F^o}	$\sigma_{ F :exp}$. Experimental uncertainty in amplitude of structure factor
s	Vector of position in reciprocal space
$ s $	$2 \sin \theta / \lambda$
ε	Multiplicity of diffracting plane
N_{part}	Number of partial structures. May be different subsets of atoms; for example protein atoms, H atoms, metals, or be derived from unparameterized electron density, such as the solvent continuum or a segment of poorly phased electron density which cannot be interpreted as an atomic model.
F_j^c	$(A_j^c, B_j^c) = F_j^c \exp i\varphi_j^c$. Calculated structure factor from j th partial structure. When $N_{part} = 1$, the subscript j will be dropped.
N_{atom}^j	Number of atoms in j th partial structure
E_j^c	Normalized partial calculated structure factor

x_{kj}	Vector of coordinates of k th atom of j th partial structure
B_{kj}	B value of k th atom of j th partial structure
Δx_j	Error in position of atoms in j th partial structure
ΔB_j	Error in B values of atoms in j th partial structure
$D_j(s)$	$D_j = \langle \exp [-(\Delta B_j s ^2 / 4)] \cos 2\pi s \Delta x_j \rangle$
F_{wc}	$\sum_{j=1}^{N_{part}} D_j F_j^c = (A_{wc}, B_{wc}) = F_{wc} \exp i\varphi_{wc}$, weighted sum of partial calculated structure factors
$f_k(s)$	Form factor of k th atom
\sum_N	$\sum_{k=1}^{N_{atom}} f_k^2(s)$
\sum_j	$\sum_j(s) = \sum_{k=1}^{N_{atom}^j} f_{kj}^2(s)$ for j th partial structure
\sum_{wc}	$\varepsilon \sum_{j=1}^{N_{part}} \sum_j (1 - D_j^2)$
$\sigma_{A:j}$	$(\sum_j / \sum_N)^{1/2} D_j$
E_{wc}^c	$\sum_{j=1}^{N_{part}} \sigma_{A:j} E_j^c$, σ_A weighted sum of normalized structure factors
$\sigma_{A:wc}$	$\varepsilon (1 - \sum_{j=1}^{N_{part}} \sigma_{A:j}^2)$
$\Delta\varphi$	Phase error of current model.
m	$(\langle \cos \varphi \rangle^2 + \langle \sin \varphi \rangle^2)^{1/2}$. Figure of merit of phases. In case of the uniform prior phase information $m = \langle \cos \Delta\varphi \rangle = I_0(X)/I_1(X)$ or $\tanh(X)$ for acentric and centric reflections, respectively.
X	$2 E^o E_{wc} / (2\sigma_{E^o}^2 + \sigma_{A:wc})$ for acentric reflections $ E^o E_{wc} / (\sigma_{E^o}^2 + \sigma_{A:wc})$ for centric reflections
$I_0(X)$ and $I_1(X)$	Zero and first-order modified Bessel functions of the first kind.
MIR	Multiple isomorphous replacement phasing
MAD	Multiple-wavelength anomalous dispersion phasing
$P(A, \dots; B, \dots)$	Conditional probability distribution of (A, \dots) when (B, \dots) are known
$P(A)$	$\int_B P(A, B) dB$, marginal probability distribution of A .

For independent experimental observations (A, \dots) and parameters (B, \dots) to be estimated using these, the log-likelihood function, LLK, is

$$LLK = \sum \log P(A, \dots; B, \dots)$$

The maximum-likelihood estimation of parameters (B, \dots) is achieved by minimization of this function. If $P(A, \dots; B, \dots)$ is the conditional probability distribution of the amplitude of the structure factor then LLK will be the amplitude-based maximum-likelihood (MLKF) residual.

If,

$$P(A, \dots; B, \dots) = c \exp -\{[A - T(B, \dots)]^2 / \sigma^2\},$$

i.e. $P(A, B, \dots)$ is the Gaussian distribution of A with expected value $T(B, \dots)$ and known uncertainties σ , then the log-likelihood is the least-squares residual,

$$LSQ = LLK = \sum [A - T(B, \dots)]^2 / \sigma^2.$$

If A are amplitudes of structure factors, B are parameters of atoms and T are amplitudes of calculated structure factors, LSQ is the amplitude-based least-squares residual (FLSQ). If A are intensities and T are calculated intensities of structure factors then LSQ is the intensity-based least-squares residual (ILSQ). $\omega = 1/\sigma^2$ is the weight for LSQ.

2. Introduction and historical survey

Crystallographers appreciated by the 1940's that the parameters they determined for the positions of atoms within a crystal could be improved by minimizing the differences between the observed amplitudes, $|F^o|$, and the $|F^c|$ calculated from those atomic parameters.

One of the first papers on the application of least-squares residuals to crystal structure refinement was written by Hughes (1941). Another approach to structural analysis – the difference Fourier method where corrections to atomic parameters were made in real space on the basis of the difference map at the atomic positions – was developed by Booth (1946, 1947). Cochran (1948) showed the similarity of these two approaches by comparing the first derivatives of the two functions. Cruickshank (1952, 1956) completed this comparative analysis and gave equations for the derivative calculation for atomic x_k and B_k for LSQ and modified Fourier methods.

Cruickshank (1952) also showed that in the final stages of a structure analysis, least-squares refinement against unweighted amplitudes (FLSQ) and against intensities (ILSQ) weighted by $w = 1/|F^o|^2$, are approximately the same. Wilson (1976) showed that in the theoretical case where only one parameter is to be estimated using ILSQ, then modifying the weights, w , as a function of $(I^o + 2I^c)$ will reduce refinement bias. He suggested using this type of weighting in the final

stages. Sheldrick (1995) uses ILSQ routinely for refining structures at high resolution, weighting each term with $w = 1/[\sigma_{F^o}^2 + a(I^o + 2I^c) + b(I^o + 2I^c)^2]$. Intensity-based refinement has some great advantages. Firstly, it is easy to use all observations whether they are strong, weak or even negative. Unlike amplitude-based least-squares ILSQ does not have a singularity when $F^c = 0$. Secondly, it is easier to estimate σ_{F^o} from the experiment than σ_{F^c} . Its principal disadvantages are firstly, when the initial model is far from the desired one and the residual is very high, minimization of the least-squares residual may not work well; secondly, ILSQ may work poorly when the model is incomplete; thirdly, adding several partial structures and refining different global parameters simultaneously may not be easy; and fourthly, adding prior phase information is very difficult.

Before using amplitude-based refinement it is necessary to address carefully two problems; that of estimating σ_{F^c} from σ_{F^o} , and that of assigning values to very small $|F^o|$. French & Wilson (1978) gave a reasonable method for estimating them based on expected distributions of weak amplitudes but such reflections must be used with care.

It was recognised in the 1960's that macromolecular refinement posed special problems. There were too few observations to refine the atomic parameters using least-squares minimization alone, and the calculation of the structure factors and derivatives from such a large number of coordinates challenged the computing resources available. Macromolecular refinement was first attempted using the real-space refinement program written by Diamond (1971). Atomic coordinates were fitted to a map, which could be generated from any available set of phases, experimental or calculated. Geometric constraints expressed as torsion angles were applied while building the model. This approach to refinement has advantages. Parts of a structure could be refined independently; and adding phase information is straightforward since this is used in the map calculation. This approach has been successfully used in different contexts by Jones, Zou, Cowan & Kjeldgaard (1991); Lamzin & Wilson (1993); Chapman (1995) and Oldfield (1996). Its disadvantages are that it depends on the quality of phases which may be poor; that weighting individual reflections is not straightforward; that maps using calculated phases almost always suffer from model bias; and that it is hard to combine derivatives from the map and from geometric parameters, and to weight the different classes of information.

[For reviews on refinement and a more complete set of references see Bricogne (1993); Watkin (1994).]

The basic assumption in least-squares minimization is that the conditional distribution of each $|F^o|$ or I^o when the model is known is Gaussian with expected value $|F^c|$ or I^c and known uncertainties. Beginning with Luzzati

(1952) it was shown by various authors that this assumption breaks down (Srinivasan & Ramachandran 1965; Read, 1990). Statistical theory shows that it is better to use maximum likelihood rather than least-squares to optimize the fit between observed and calculated quantities (Stuart & Ord, 1991). Maximum likelihood does not assume a Gaussian error distribution and instead estimates the conditional distribution of experimental data when the model is known. Luzzati showed that with the assumption that all errors are a consequence of coordinate errors, Δx , and that $\langle \cos(2\pi s \Delta x) \rangle$ and $\langle [\cos(2\pi s \Delta x)]^2 \rangle$ [*i.e.* the first and second moments of $\cos(2\pi s \Delta x)$] are the same for all atoms, then the conditional distribution of the structure factors themselves is Gaussian but not that of the amplitudes. The resulting marginal distribution of amplitudes of structure factors was given by Srinivasan & Ramachandran (1965). Later Bricogne (1988) noted that this distribution was used by other authors, the first being Rice (1954). Luzzati's results were extended by Read (1990) to include the assumption that there are errors in the atomic scattering factors which can be assigned to ΔB .

Recent interest in using these equations for refinement of macromolecular structures has been stimulated by their successful application in reducing model bias in map calculations (Lunin & Urzhumtsev, 1984; Read, 1986); by the theoretical results described by Bricogne (Bricogne, 1988) for their application with maximum entropy as a general tool in phase determination; by their successful application of his results in phase improvement (Carter, Crumley, Coleman, Hage & Bricogne, 1990); and by the increasing power of computers. The procedure described here is referred to as 'maximum-likelihood refinement' since all previous macromolecular refinement procedures are referred to as 'least-squares refinement' despite the fact that all these also utilize prior information.

As mentioned above the ratio of the number of experimental data to the number of parameters to be estimated is small. It is essential to use prior chemical or other information. This means that all macromolecular refinement can be seen as applications of Bayes' theorem. [For details of this, and of its application see Box & Tiao (1973).]

Assuming there are experimental data $|F|$ and parameters to be estimated x , then Bayes' theorem can be written as,

$$P(x; |F|) = p(x)P(|F|; x)/P(|F|) = p(x)L(x; |F|). \quad (1)$$

Here, P is the posterior probability distribution of the parameters when the experimental data are known; x is the parameter to be estimated; p is the prior probability distribution of parameters known before the experiment. It reflects the prior knowledge of the experimenter; L is the likelihood function which is

proportional to the conditional distribution of experimental data when the parameters are known.

To best estimate the parameters x the posterior must be forced to reach its maximum. To apply this theorem the form of the prior and likelihood is needed.

In this paper for convenience we refer to $P(|F|; F^c)$ and $L(F^c; |F|)$ instead of $P(|F|; x)$ and $L(x; |F|)$ since F^c is directly calculated from the x .

2.1. Prior knowledge

In principle, prior knowledge might contain information about bond lengths, bond angles, torsion angles *etc.* Two different methods of using this prior knowledge are well established. One of these methods is to express it as stereochemical restraints (Waser, 1963; Konnert, 1976) and another is to express it as energetic restraints (Levitt, 1974). The first idea has been incorporated into the various versions of *PROLSQ* (Konnert & Hendrickson, 1980), *FROG* (Lunin & Urzhumtsev, 1985), *TNT* (Tronrud, Ten Eyck & Matthews, 1987), *RESTRAINT* (Driessen *et al.*, 1989) and *SHELXL* (Sheldrick, 1995). The energy approach has been implemented in programs such as *EREF* (Jack & Levitt, 1978) and *X-PLOR* (Brünger, 1992). The examples described here incorporate the prior knowledge as stereochemical restraints.

2.2. Likelihood

Since likelihood is proportional to the conditional probability distribution of experimental data when the model is known, the form of this conditional probability distribution is needed. The best way would be to find the joint probability distribution of all structure factors. However, this task is not trivial and its implementation requires a large amount of computer memory and time. All existing refinement procedures assume that the errors in different reflections are independent, and this simplification still allows useful results to be obtained.

With this assumption, the required joint probability distribution of all amplitudes of structure factors has the form,

$$P[(|F_h^o|)^{\text{all reflections}}, (F_h^c)^{\text{all reflections}}] = \prod_{\text{all reflections}} P(|F_h^o|; F_h^c). \quad (2)$$

Thus, to describe the likelihood function, the conditional probability distribution of each reflection is generated and these are multiplied together to give the joint conditional probability distribution.

2.3. Posterior distribution

Since maximization of a function is equivalent to the minimization of its negative logarithm, (1) is equivalent to,

$$-\log P[x; (|F_h^o|)^{\text{all reflections}}] = -\log p(x) - \sum_{\text{all reflections}} \log L(F_h^c; |F_h^o|), \quad (3)$$

where $L(F_h^c; |F_h^o|) \propto P(|F_h^o|; F_h^c)$. $-\log p(x)$ can be written in a straightforward way from the stereochemical information, so only the second term needs to be considered.

3. Conditional distribution of structure-factor amplitudes split into separate components

Suppose the structure factors of interest are the sum of ideal structure factors derived from perfectly ordered atoms with an error,

$$F = \sum_{j=1}^{N_{\text{part}}} F_j + \text{experimental}_{\text{err}}, \quad (4)$$

where

$$F_j = \sum_{k=1}^{N_{\text{atom}}} f_{kj}(s) \exp -[(B_{kj}|s|^2)/4] \exp(2\pi i s x_{kj}). \quad (5)$$

Now suppose x_{kj} and B_{kj} have approximate values x_{kj}^c and B_{kj}^c .

If the errors in different atoms are independent and distributed with the same probability distribution {equality of the first and second moments of $\exp[-(\Delta B|s|^2)/4] \cos(2\pi s \Delta x)$ is required}, then according to Luzzati (1952) and Read (1990) we can write,

$$P(F_p^j; F_j^c) = \begin{cases} \frac{1}{\pi \varepsilon \sum_j (1 - D_j^2)} \exp - \left[\frac{|F_p^j - D_j F_j^c|^2}{\varepsilon \sum_j (1 - D_j^2)} \right] & \text{acentric} \\ \frac{1}{[2\pi \varepsilon \sum_j (1 - D_j^2)]^{1/2}} \exp - \left[\frac{|F_p^j - D_j F_j^c|^2}{2\varepsilon \sum_j (1 - D_j^2)} \right] & \text{centric.} \end{cases} \quad (6)$$

Suppose the distribution of experimental errors around F^o is normal with mean zero and variance $\sigma_{F^o}^2$, with errors distributed over two dimensions for acentric reflections and along one dimension for centric reflections.* The sum of independent random normally distributed variables will also be normally distributed with a mean equal to the sum of means and a variance equal to the sum of variances. Assuming that F_j 's are independent of each other and of the experimental errors, and using notations $F_{\text{wc}} = \sum_{j=1}^{N_{\text{part}}} D_j F_j^c$ and $\Sigma_{\text{wc}} = \varepsilon \sum_{j=1}^{N_{\text{part}}} \Sigma_j (1 - D_j^2)$,

* This assumption makes the addition of experimental uncertainties into the probability distribution straightforward. For a more elaborate treatment of experimental uncertainties in probability distributions see Pannu & Read (1996). Test cases show that this simple approach is helpful.

$$P[F; (F_j^c)^{j=1, N_{\text{part}}}] = \begin{cases} \frac{1}{\pi(2\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \exp - \left(\frac{|F^o - F_{\text{wc}}|^2}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) & \text{acentric} \\ \frac{1}{[2\pi(\sigma_{F^o}^2 + \Sigma_{\text{wc}})]^{1/2}} \exp - \left[\frac{|F^o - F_{\text{wc}}|^2}{2(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \right] & \text{centric.} \end{cases} \quad (7)$$

This implies that the distribution of a hypothetical vector of experimental error is normal. Bricogne & Gilmore (1990) suggested a similar way of adding σ_{F^o} . It is expected that the effect of experimental error will be considerable when its contribution is comparable to that from coordinate errors, and in that case the assumption that this hypothetical vector of experimental error is distributed by the normal law may not be justified. If there are many observations where I^o/σ_{I^o} is small it may be better to use an intensity-based residual, either ILSQ or intensity-based maximum likelihood.* It is more reasonable to assume a Gaussian distribution of experimental errors for weak intensities than for weak amplitudes (Pannu & Read, 1996).

Splitting the error terms for F^c and writing Σ_{wc} as a sum is convenient and reflects the different types of structure factors and errors. Different parts of a structure may need different treatments. For example only one domain of a structure may obey NCS. Another example is when a partial model has been built into an experimentally phased map. Often the remainder of the map is not clear enough to interpret, but it can generate a useful contribution to the calculated structure factor.

So far we have assumed that both amplitudes and phases of F^o are known. In reality this is usually not so, as only amplitudes can be measured. Hence, we require the distribution of the amplitudes rather than the structure factors (for centrosymmetric reflections the integral should be replaced by summation over the two possible values of the phase).

$$P[|F|^o, \varphi^o; (F_j^c)^{j=1, N_{\text{part}}}] = \int_0^{2\pi} P(\varphi^o; \varphi) P[|F^o|, \varphi; (F_j^c)^{j=1, N_{\text{part}}}] d\varphi, \quad (8)$$

where $P(\varphi^o; \varphi)$ is distribution of experimental phases. [In the following discussions we will shorten $P(\varphi^o; \varphi)$ to $P(\varphi)$.] (7) and (8) give the following distribution [as the Jacobian of the transformation from (A, B) to $(|F|, \varphi)$ is equal to $|F|$],

Some special cases of (9) are of interest. Suppose the exact phases of the structure factors were known. Then (9) would have the form,

* In any case the distribution of the experimental errors is only approximately Gaussian, and this begins to break down for small I^o/σ_{I^o} . The distribution of the amplitudes of structure factors is derived from that of intensities by a further approximation (French & Wilson, 1978).

$$P[|F^o|; (F_j^c)^{j=1, N_{\text{part}}}] = \begin{cases} \frac{|F^o|}{\pi(2\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \exp - \left(\frac{|F^o|^2 + |F_{\text{wc}}|^2}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) \int_0^{2\pi} P(\varphi) \exp \left[\left(\frac{2|F^o||F_{\text{wc}}|}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) \cos(\varphi - \varphi_{\text{wc}}) \right] d\varphi & \text{acentric} \\ \left[\frac{1}{2\pi(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \right]^{1/2} \exp - \left[\frac{|F^o|^2 + |F_{\text{wc}}|^2}{2(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \right] \sum_{l=0}^1 P(\varphi_l) \exp \left[\left(\frac{|F^o||F_{\text{wc}}|}{\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) \cos(\varphi_l - \varphi_{\text{wc}}) \right] & \text{centric.} \end{cases} \quad (9)$$

$$P[|F^o|; (F_{\text{wc}})] = \begin{cases} \frac{|F^o|}{\pi(2\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \exp - \left(\frac{|F - F_{\text{wc}}^c|^2}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) & \text{acentric} \\ \left[\frac{1}{2\pi(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \right]^{1/2} \exp - \left[\frac{|F - F_{\text{wc}}|^2}{2(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \right] & \text{centric,} \end{cases} \quad (10)$$

where $F = |F^o| \exp(i\varphi^o)$.

This assumption is usually unrealistic. Phases can be

$$P[|F^o|; (F_j^c)^{j=1, N_{\text{part}}}] = \begin{cases} \frac{2|F^o|}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \exp - \left(\frac{|F^o|^2 + |F_{\text{wc}}|^2}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) I_0 \left(\frac{2|F^o||F_{\text{wc}}|}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) & \text{acentric} \\ \left[\frac{2}{\pi(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \right] \exp - \left[\frac{|F^o|^2 + |F_{\text{wc}}|^2}{2(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \right] \cosh \left(\frac{|F^o||F_{\text{wc}}|}{\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) & \text{centric.} \end{cases} \quad (11)$$

estimated with high accuracy for structures with rich NCS, such as viruses, and Burling, Weis, Flaherty & Brünger (1996) describe a MAD experiment which yielded very accurate phase information for many of the F^o . In such cases the phases could be used as observables. A more common situation is when there is no prior phase information available. Then (9) will become the familiar Rice distribution (see Bricogne, 1988; Rice, 1954),

This equation is the most popular one from the family of distributions given by (9). In crystallography a special case of this equation was first derived by Srinivasan & Ramachandran (1965). They also gave a normalized version of it which was later applied to reduce bias in map calculation by Read (1986).

Some special cases of (11) are interesting. If some atomic parameters are approximately known and it is assumed that the missing atoms are distributed uniformly over the asymmetric unit, then $N_{\text{part}} = 2$, $D_1 = D$, $D_2 = 0$, $\Sigma_1 = \Sigma_p$, $\Sigma_2 = \Sigma_q$ and $F_1^c = F_p^c$ where p denotes the partial structure and q denotes the missing part of the structure. This gives Srinivasan & Ramachandran's distribution which neglects experimental errors,

$$\begin{aligned} \Sigma_{\text{wc}} &= \Sigma_q + \Sigma_p(1 - D^2) \\ F_{\text{wc}} &= DF_p^c. \end{aligned} \quad (12)$$

Sim's distribution (1959) is a special case of this: he assumed the partial structure was correct and the only error was due to the missing atoms, *i.e.* $D = 1$ and $\Sigma_{\text{wc}} = \Sigma_q$. Wilson's distribution (1949) is a special case too: there it is assumed that no atomic parameters are known, and that the contents of the unit cell are distributed uniformly. So $F_{\text{wc}} = 0$, $I_0(0) = 1$ and $\Sigma_{\text{wc}} = \Sigma_N$.

Finally the log likelihood from (11) will have the form,

$$\text{LLK} = \sum_h \text{LLK}_h, \quad (13)$$

where,

$$\text{LLK}_h = \begin{cases} c_a - \log(|F^o|) + \log(2\sigma_{F^o}^2 + \Sigma_{\text{wc}}) \\ \quad + \frac{|F^o|^2 + |F_{\text{wc}}|^2}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} - \log I_0 \left(\frac{2|F^o||F_{\text{wc}}|}{2\sigma_{F^o}^2 + \Sigma_{\text{wc}}} \right) & \text{acentric} \\ c_c + \frac{1}{2} \log(\sigma_{F^o}^2 + \Sigma_{\text{wc}}) \\ \quad + \frac{|F^o|^2 + |F_{\text{wc}}|^2}{2(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} - \log \cosh \left[\frac{|F^o||F_{\text{wc}}|}{(\sigma_{F^o}^2 + \Sigma_{\text{wc}})} \right] & \text{centric.} \end{cases} \quad (14)$$

When there is some prior phase information the LLK is derived from (9).

4. Relationship between log-likelihood and least-squares residuals

Bricogne (1992), using Taylor's expansion of $I_0(t)$ and $\cosh(t)$ around $t = 0$, showed that when

$X(= 2|F^o||F_{wc}|/\Sigma_{wc})$ is small (either because Σ_{wc} is large or the $|F|$'s are small) the log-likelihood residual is similar to the normalized intensity-based Patterson correlation function (which is related to normalized intensity-based least squares). Here, we give the assumption under which the LLK and FLSQ residuals are similar. Assume that $X=2|F^o||F_{wc}|/(2\sigma_{F^o}^2 + \Sigma_{wc})$ for acentric and $|F^o||F_{wc}|/(\sigma_{F^o}^2 + \Sigma_{wc})$ for centric terms. When X is large $I_0(X)$ and $\cosh(X)$ have this asymptotic behaviour,

$$\begin{aligned} I_0(X) &\simeq \frac{1}{(2\pi X)^{1/2}} \exp X \\ \log[I_0(X)] &\simeq c - \frac{\log X}{2} + X, \\ \cosh(X) &\simeq \frac{1}{2} \exp X \\ \log[\cosh(X)] &\simeq c + X. \end{aligned} \quad (15)$$

It can be shown that (ignoring constants),

$$\text{LLK}_h \simeq \begin{cases} \frac{(|F^o| - |F_{wc}|)^2}{2\sigma_{F^o}^2 + \Sigma_{wc}} + \frac{1}{2} \log(2\sigma_{F^o}^2 + \Sigma_{wc}) \\ -\frac{1}{2} \log \frac{|F^o|}{|F_{wc}|} & \text{acentric} \\ \frac{(|F^o| - |F_{wc}|)^2}{2(\sigma_{F^o}^2 + \Sigma_{wc})} + \frac{1}{2} \log(\sigma_{F^o}^2 + \Sigma_{wc}) & \text{centric.} \end{cases} \quad (16)$$

Note, a similar approximation, although in a different context, was given by Rice (1954).

Assuming that $|F^o| \simeq |F_{wc}|$ so that $\log(|F^o|/|F_{wc}|)$ is small, the above equation is similar to FLSQ, the Gaussian-based likelihood function. This assumption is reasonable near the end of refinement. If all reflections are measured with equal σ_{F^o} (a very unreal assumption) and all $D_j \simeq 1$ then the above equation will become the unit-weighted LSQ residual.

The similarity between the LLK and FLSQ residuals can also be derived from (10), with the assumption that the phases of the calculated structure factors are always equal to the true phase.

5. Parameters of the likelihood function

In the applications of MLKF described here the overall parameters of likelihood and the atomic parameters are refined sequentially. We use the normalized version of (11) and deduce $\sigma_{A;wc}$ where $\sigma_{A;wc} = 1 - \sum_{j=1}^{N_{\text{part}}} \sigma_{A;j}^2$ and $E_{wc} = \sum_{j=1}^{N_{\text{part}}} \sigma_{A;j} E_j$. The normalization is done in resolution bins.

$$\begin{aligned} P[|E^o|; (E_j^c)^{j=1..N_{\text{part}}}] &= \begin{cases} \frac{2|E^o|}{2\sigma_{E^o}^2 + \sigma_{A;wc}} \exp - \left(\frac{|E^o|^2 + |E_{wc}|^2}{2\sigma_{E^o}^2 + \sigma_{A;wc}} \right) \\ \times I_0 \left[\frac{2|E^o||E_{wc}|}{(2\sigma_{E^o}^2 + \sigma_{A;wc})} \right] & \text{acentric} \\ \frac{2}{\pi(\sigma_{E^o}^2 + \sigma_{A;wc})} \exp - \left[\frac{|E^o|^2 + |E_{wc}|^2}{2(\sigma_{E^o}^2 + \sigma_{A;wc})} \right] \\ \times \cosh \left(\frac{|E^o||E_{wc}|}{\sigma_{E^o}^2 + \sigma_{A;wc}} \right) & \text{centric.} \end{cases} \end{aligned} \quad (17)$$

Read (1986) uses this equation in his program *SIGMA*. He estimates the σ_A parameters in reciprocal-space resolution shells, where each shell includes several hundred reflections.

Another way of representing $\sigma_{A;j}$ is to express them as some function of resolution. Crystallographers are used to expressing overall scale factors in this way. The one Gaussian overall scale approximation $k \exp(-B|s|^2)$ suggested by A. J. Wilson gives reasonable agreement between $|F^o|$ and $|F^c|$ for macromolecular structures at high resolution but not for low-resolution terms. To improve this fit Tronrud (1997) suggested using a two Gaussian approximation for the scale factor invoking the Babinet principle. Assuming that the contribution of the solvent and protein parts of the crystal to the structure factors are negatively correlated he showed that a scale derived from the following equation gives reasonable agreement between $|F^o|$ and $|F^c|$ for all resolution ranges,

$$k = k_0 \exp(-B_0|s|^2/4)[1 - k_1 \exp(-B_1|s|^2/4)]. \quad (18)$$

A similar approach can be used for $\sigma_{A;j}$ -s. The following expression can be fitted to the data,

$$\sigma_{A;j} = \sigma_{A;j,0} \exp(-C_{j,0}|s|^2/4)[1 - \sigma_{A;j,1} \exp(-C_{j,1}|s|^2/4)]. \quad (19)$$

Thus, to estimate all the $\sigma_{A;j}$ $2N_{\text{part}} + 2$ parameters are needed. For this a few hundred reflections is enough (in one of our test cases we estimated σ_A 's satisfactorily using only 200 reflections). So for the estimation of $\sigma_{A;j}$'s one could use reflections not included in the refinement of the atomic parameters, the 'free' reflections (Brünger, 1992). Care must be taken at this stage as the parameters are highly correlated and simultaneous estimation of them can cause problems. A singular value decomposition method (Press, Flannery, Teukolsky & Vetterling, 1986) to solve the linear equations has been used in the examples below. Another approach would be to refine the parameters $C_{j,i}$'s and $\sigma_{A;j,i}$'s consecutively.

Table 1. *Examples*

	BA2	Cytochrome <i>c'</i>	Insulin	OppA
Space group	<i>C</i> 222 ₁	<i>P</i> 6 ₃ 22	<i>P</i> 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Cell dimensions (Å, °)	52.5, 77.7, 238.2 90, 90, 90	54.5, 54.5, 181.0 90, 90, 120	53.9, 64.8, 48.9 90, 109.8, 90	109.2, 76.0, 70.3 90, 90, 90
Resolution (Å)	2.2	2.0	1.9	1.2
Number of residues	483	125	318	517
Method of solution	MIR	MR	MR	MAD/averaging
Completeness of the model	good model 85% of protein no waters	poor homology 25% homology 10 residues misfitted	good homology All protein and 20% of waters	good model 100%
Final <i>R</i> value/ <i>R</i> free (%)	11.9/20.9	16.7/NA	18.4/25.2	16/18

6. Examples of application

The maximum-likelihood equations have been implemented in the program *REFMAC* now available from CCP4 (Collaborative Computational Project, Number 4, 1994). It has been widely used in a β -test release, and to illustrate its performance results from four structures at different stages of refinement are described. In each test the MLKF procedure was run automatically from some starting point until convergence (*i.e.* until there was no significant increase in the figure of merit, m or decrease in the R values between cycles), without any manual rebuilding of the structure, and the resultant model compared with the final one submitted for publication. Details for the examples discussed are given in the Table 1.

In the following discussion m and D are derived from σ_A . $\Delta\varphi$ is the phase difference between phases calculated from the current and final models, $\langle \cos \Delta\varphi \rangle$ the average cosine phase error which should equal m , the estimated figure of merit, at all stages of refinement. If σ_A and, therefore, m are overestimated the likelihood function is biased towards accepting the current model. Better results were obtained when σ_A was estimated from the reflections not included in the refinement. R values and R free are given for convenient comparison with other refinement procedures. They were always calculated using the Tronrud scaling technique.

Another indicator of how successfully a refinement procedure is driving the model towards the correct solution is the correlation between maps generated from the current model and the F^c map derived from the final coordinates. *REFMAC* writes out map coefficients $(m|F^o| - D|F^c|)$ and $(2m|F^o| - D|F^c|)$ using the current m and D at the end of each refinement pass, taking care to 'restore' unmeasured data. Tronrud (1996) and Cowtan (1996) show that absent reflections cause unpredictable noise in map calculations which sometimes may lead to errors in interpretation. Assuming that absent reflections are best approximated by their expected value (in the case of maximum likelihood setting $m|F^o| = D|F^c|$) then the difference contribution is zero, and the $2m|F^o| - D|F^c|$ contribution is $D|F^c|$. The same approximation is used for the 'free' reflec-

tions, to avoid biasing the R free and σ_A statistics in any way.

These coefficients are,

$$\begin{aligned}
 \text{FWT} &= \begin{cases} (2m|F^o| - D|F^c|) \exp(i\varphi_c) & \text{if reflection was} \\ & \text{included in refinement} \\ D|F^c| \exp(i\varphi_c) & \text{otherwise} \end{cases} \\
 \text{DELFWT} &= \begin{cases} (m|F^o| - D|F^c|) \exp(i\varphi_c) & \text{if reflection was} \\ & \text{included in refinement} \\ 0 & \text{otherwise.} \end{cases} \quad (20)
 \end{aligned}$$

These should generate maps which are similar to those derived from the *SIGMAA* program (Read, 1986). Any differences will be because of the estimates of m and D which in the *SIGMAA* program are based on all the data and can be biased towards the current model, and to the effect of 'restoring' the unmeasured and 'free' reflections. This will reduce noise but may also introduce bias. *REFMAC* also includes the σ_{E^o} in the derivation of these terms which usually leads to improved behaviour. In fact in several cases when this has not been so it has been shown that the σ_{E^o} were wrongly estimated during data processing.

We tabulate the correlation for maps with coefficients $(2|F^o| - |F^c|)$, $(2m|F^o| - D|F^c|)$ where the values of m and D are derived by the *SIGMAA* program from all data, and for the $(2m|F^o| - D|F^c|)$ coefficients from *REFMAC*, where the values of m and D are derived from the 'free' reflections, and unmeasured reflections are 'restored'.

In general MLKF refinement gives more improvement at high resolution than at low resolution. The reason for this is that the fit of the high-resolution structure factors depends on the accurate position of atoms but the low-resolution data fit depends on large movements of structure or a more complete description of the model. To improve the fit of the data at low resolution the model needs to be rebuilt, or new features

Table 2. Map correlations and statistics

m free, estimated figure of merit based on σ_A values fitted to 'free' reflections. m all, estimated figure of merit based on σ_A values fitted to all reflections. FLSQ, unweighted least-squares refinement using *PROTIN/PROLSQ/SFALL* implemented in *CCP4*. MLKF, maximum likelihood refinement using *REFMAC* implemented in *CCP4*. SigmaA, Coefficients generated using *SIGMAA* program implemented in *CCP4*. mapc, map correlation between named map, and the F^c map from the deposited coordinates.

	R value	R free	$\langle \Delta\varphi \rangle$	$\langle \cos \Delta\varphi \rangle$	m free	m all	mapc
(a) BA2. Beginning refinement from MIR model							
Initial	0.47	0.46	56.4	0.44	0.46	0.45	$2F^o - F^c$ 0.55
FLSQ	0.30	0.40	409	0.64	—	0.76	$2F^o - F^c$ 0.73
	—	—	—	—	—	—	SigmaA 0.81
MLKF	0.28	0.35	35.6	0.70	0.69	0.80	Refmac 0.85
(b) Cytochrome c' : when FLSQ fails							
FLSQ	0.31	0.32	41.7	0.62	—	0.78	$2F^o - F^c$ 0.65
	—	—	—	—	—	—	SigmaA 0.73
MLKF	0.29	0.38	36.3	0.68	0.69	0.81	Refmac 0.78
(c) Cross-linked insulin: FLSQ apparently converged							
FLSQ	0.26	0.32	27.2	0.62	0.80	0.79	$2F^o - F^c$ 0.84
	—	—	—	—	—	—	SigmaA 0.87
MLKF	0.21	0.26	15.9	0.91	0.81	0.87	Refmac 0.90

such as waters added. The erratic behaviour of R factors and $\langle \cos \Delta\varphi \rangle$ at low resolution may be because of the small number of observations in these ranges.

6.1. Bacterial α -amylase. Beginning refinement from an excellent MIR model

The structure of BA2 was solved by Brzozowski *et al.* (1997). The initial model for 80% of the residues had been built very carefully into a good MIR map and any refinement program could have performed satisfactorily. 5% of the reflections were reserved for R free estimation. This model was subjected to both FLSQ and MLKF refinement. The MLKF results were significantly better, with lower phase error and also R value and R free. The overall m estimated from the 'free' reflections remained very close to the calculated $\langle \cos \Delta\varphi \rangle$ throughout the MLKF refinement.

The maps generated using the MLKF coefficients gave higher map correlations. The m estimated by *SIGMAA* from all reflections overestimated the phase reliability and hence the *SIGMAA* maps were a little less effective at removing model bias. The map correlations and other statistics are tabulated in Table 2.

6.2. Cytochrome c' . Preparing to rebuild from a molecular replacement solution

The structure was solved by Baker, Anderson, Dobbs & Dodson (1995). This starting model was based on a molecular replacement (MR) solution where the model used had only 25% homology to cytochrome c' . Although a solution was found ten residues had out-of-register errors and another ten were completely misplaced. In such cases where an extensive rebuilding

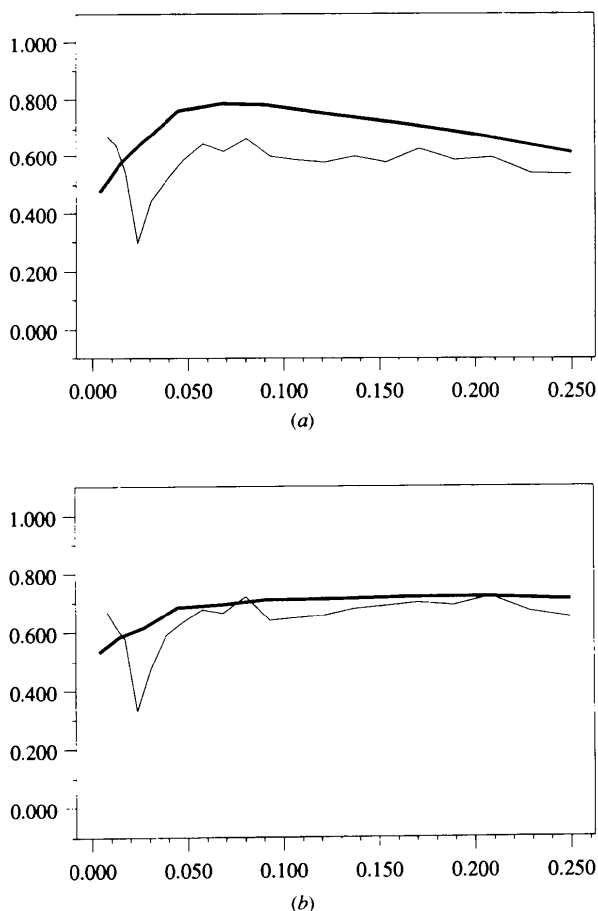


Fig. 1. Cytochrome c' : behaviour of m and real $\langle \cos \Delta\varphi \rangle$ (a) before and (b) after MLKF refinement. Bold lines show estimated m , thin lines show the true $\langle \cos \Delta\varphi \rangle$.

is necessary, the problem of map bias is very serious. Automatic cycles of FLSQ refinement using all reflections had reduced the R value to 34.8, and this partially refined model was used as a starting point for MLKF.

At this point 5% of reflections were assigned as 'free' and were used for estimation of the overall likelihood parameters. At first the σ_A weighting was overestimated

and the initial m was much higher than the $\langle \cos \Delta\varphi \rangle$, indicating the importance of assigning 'free' reflections at the beginning of refinement. *REFMAC* was able to refine the FLSQ model further and the phase error was reduced by 6° . The behaviour of $\langle \cos \Delta\varphi \rangle$ versus resolution (Fig. 1) shows that during refinement the phases for the high-resolution data were improved most, and m and $\langle \cos \Delta\varphi \rangle$ converged.

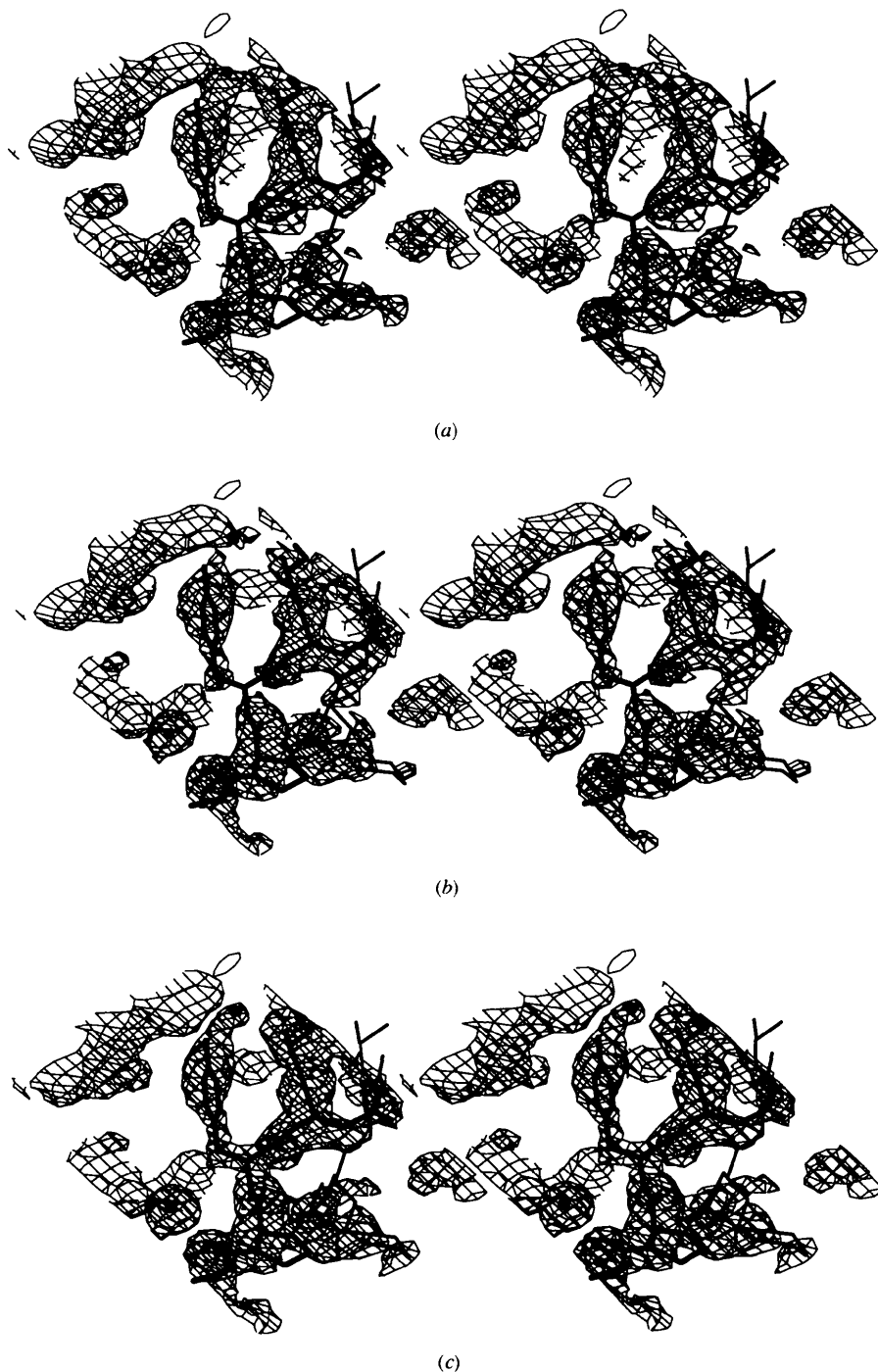


Fig. 2. Cytochrome c' : electron densities calculated after FLSQ and MLKF. (a) $2F_o - F_c$ map, (b) *SIGMAA* map and (c) *REFMAC* map. Bold lines show final coordinates, thin lines show coordinates included in refinement.

It is interesting to inspect the maps (Fig. 2) which were available for correcting the model. These illustrate a section where the initial model was completely wrong. The $(2F^o - F^c)$ map (Fig. 2a) and $(3F^o - 2F^c)$ map (not shown) based on the FLSQ model are seriously biased and noisy and it would be easy to trace the chain perpendicular to its true direction. The *SIGMA* maps for the initial FLSQ model (Fig. 2b) correlated with the final F^c model map better but still there is a break in the main chain and the electron density could be interpreted wrongly. The map after *REFMAC* had a map correlation coefficient 5% higher than that for the map calculated by *SIGMA* coefficients, showed less ambiguous connectivity, and density for side chains and water molecules had appeared (Fig. 2c).

6.3. Cross-linked insulin. End stages of refinement when FLSQ has apparently converged

This structure is not yet submitted but refinement is virtually complete (Edwards, personal communication). The solution for this cross-linked insulin was found by molecular replacement using a model with 95% homology. Least-squares refinement and rebuilding cycles had given an R value of 24% and free R value 34%. Maximum-likelihood refinement using *REFMAC* reduced the R value by 4% but the free R value dropped

even more – by 6% (Table 2c). The geometric parameters such as root-mean-square deviation of bond lengths from ideality also improved. The maps using *REFMAC* coefficients showed regions where there were multiple conformations (Fig. 3), and a serious error in interpretation at the C terminus of one chain. After rebuilding and further refinement by *REFMAC* the R value was reduced to 18% and free R value to 25%. The plot of R value versus resolution (Fig. 4) again shows that there is more improvement at high resolution than at low resolution.

6.4. Oligopeptide binding protein (*OppA*). Refinement of heavy atoms estimating error terms in different ways

OppA is co-crystallized with up to eight uranium ions and these pose special problems for refinement procedures. (Tame *et al.*, 1994). Several different liganded forms have been refined and one of these structures was used as a starting model for this form which diffracted to 1.2 Å. The refinement by FLSQ converged at the worryingly high R value of 22.2% with R free at 24.5%, probably because of problems in modelling the U atoms (Table 3). The first application of MLKF assuming that all atoms have same expected errors (14 and 15) reduced the R value and R free to 21.3 and 23.5%, respectively. (Adding H atoms in their

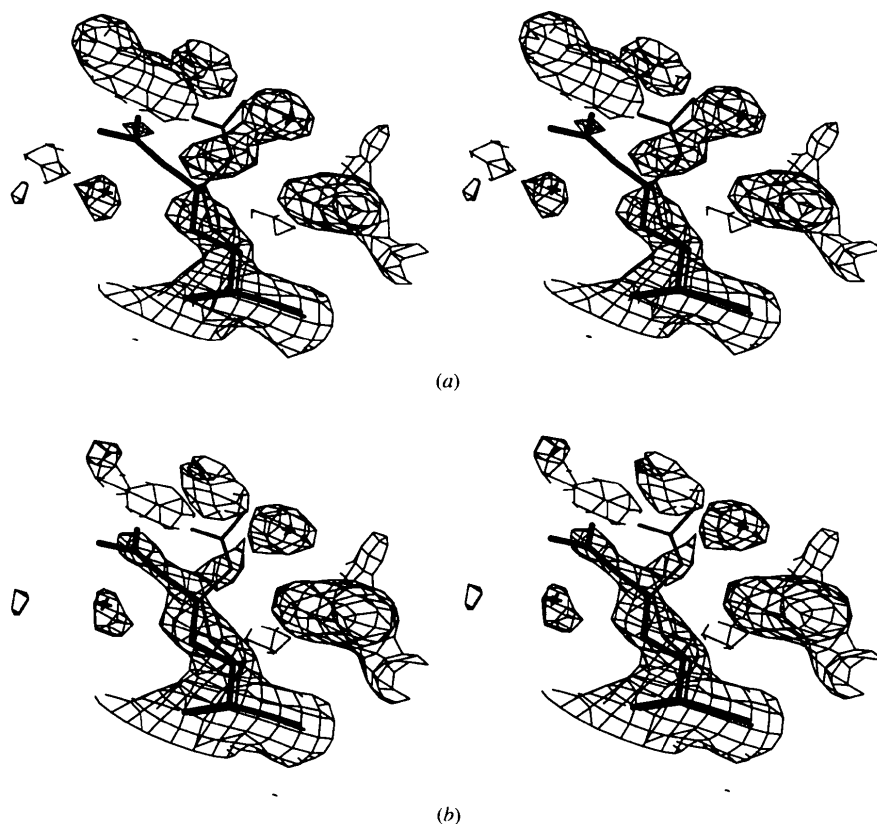


Fig. 3. Insulin: electron density corresponding to ArgB22 (a) before *REFMAC*; (b) after *REFMAC*. Thin lines show coordinates included in refinement, thick lines show coordinates of corrected model.

riding positions improved R value and R free to 20.8% and 23.1%.) After allowing the U atoms and the light atoms to have different expected coordinate errors (14), the R value and R free dropped further to 19.9 and 22.1%, respectively. Further refinement of this protein including anisotropic B factors is under way.

7. Conclusions and future perspectives

The equations and techniques described here give a foundation for using maximum likelihood in many ways. The program *REFMAC* has been coded to allow modular development and results using (14) are very encouraging; the test cases show that results derived using the maximum-likelihood implementation in the program *REFMAC* are consistently better than those derived using least squares.

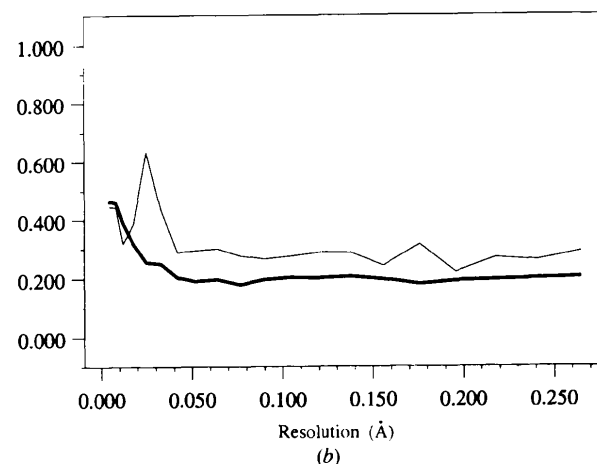
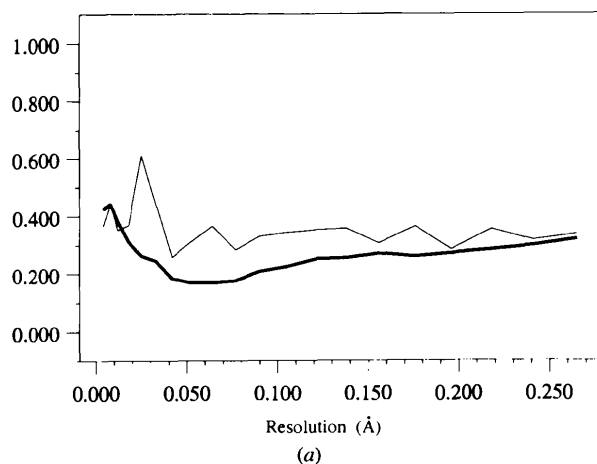


Fig. 4. Insulin: behaviour of R value and R free versus resolution (a) before and (b) after *REFMAC*. Bold lines show R value, thin lines show R free.

Table 3. *OppA*: treatment of heavy atoms

FLSQ, least-squares refinement; MLKF1, maximum-likelihood refinement assuming all atoms have same expected errors; MLKF2, same as MLKF1 with H atoms added in their riding positions; MLKF3, maximum-likelihood refinement assuming light atoms and each of the eight U atoms have different expected errors. Atomic parameters of U atoms were not refined.

	FLSQ	MLKF1	MLKF2	MLKF3
R value (%)	22.1	21.3	20.8	19.9
R free (%)	24.5	23.5	23.1	22.1

Partitioning structure factors allows us to assign different expected errors to each component. The approximations used to estimate the coefficients of Σ_{wc} seem sufficiently robust to give satisfactory answers. This should allow a more flexible approach to refinement of macromolecular structure, where the quality of diffraction from different parts can vary a great deal. Proteins which incorporate metal ions lend themselves to this approach. Another unsolved problem is that of the contribution to the structure factor from bulk solvent. The scaling technique used by Tronrud only improves the agreement between structure-factor amplitudes and does not change the phases. If a better modelling procedure for the bulk solvent or for a missing part of the structure can be devised, these contributions can be added to the structure factor, and weighted appropriately, although for doing this (9) may need to be modified.

Incorporating experimental σ_{Fo} seems to improve the course of the refinement, provided the experimental values are realistic. More work needs to be carried out in this area and the findings fed back into data-processing packages. The conversion of σ_{Fo} to σ_{Fo} may also be a source of error. The implementation of intensity-based likelihood and its least-squares approximation (*Appendix C*) will allow the optimal use of experimental uncertainties for weak reflections.

One more approximation which might cause problems is the normalization procedure used to derive σ_A . The calculation of an appropriate normalization factor is a tricky problem for many macromolecular techniques, such as direct methods and density modification. It may be better to work with structure factors on the absolute scale for refinement.

If phase information is available with reliable probability distributions, (9) could be used. Using this phase information should improve the reliability of derived coordinates. At the same time the map coefficients calculated within *REFMAC* would combine experimental and model phases.

A further extension to the program will allow the derivation of standard uncertainties of the atomic parameters. For this it is necessary to generate the non-diagonal terms of the second derivative matrix (see *Appendix B*).

APPENDIX A

Derivatives of $-\log$ -likelihood with respect to parameters. Estimation of σ_A

These appendices are included to help any programmers who may wish to develop the technique further.

A1. Chain rule

Most estimation procedures can be treated as the minimization of some function with respect to parameters and most minimization procedures use the first and second derivatives of the function with respect to these parameters. Before giving details of derivatives of $-\log$ -likelihood it is useful to introduce the chain rule for the relevant function type. Most crystallographic residuals to date have the form,

$$f = \sum_h f(A_h, B_h, \Sigma_h) = \sum_h f_h, \quad (21)$$

where A_h, B_h are the real and imaginary parts of the structure factor, Σ_h is 'uncertainty'. Note that summation is over all reciprocal space. For example in (17) $\Sigma_h = 2\sigma_e^2 + \sigma_{A:wc} = 2\sigma_e^2 + \varepsilon(1 - \sum \sigma_{A:j}^2)$. For LSQ residuals Σ_h is assumed to be known and is not refined.

This functional form is needed for the fast calculation of the second derivatives. Assume all parameters to be estimated are within A_h, B_h or Σ_h . Then using the chain rule, for the first derivatives,

$$\frac{\partial f}{\partial p} = \sum_h \frac{\partial f}{\partial A_h} \frac{\partial A_h}{\partial p} + \frac{\partial f}{\partial B_h} \frac{\partial B_h}{\partial p} + \frac{\partial f}{\partial \Sigma_h} \frac{\partial \Sigma_h}{\partial p}, \quad (22)$$

and for the second derivatives, assuming that $\partial^2 f / \partial A_h \partial B_h = \partial^2 f / \partial B_h \partial A_h$ and so on,

$$\begin{aligned} \frac{\partial^2 f}{\partial p_i \partial p_j} = & \sum \left[\frac{\partial f}{\partial A_h} \frac{\partial A_h^2}{\partial p_i \partial p_j} + \frac{\partial f}{\partial B_h} \frac{\partial B_h^2}{\partial p_i \partial p_j} + \frac{\partial f}{\partial \Sigma_h} \frac{\partial \Sigma_h^2}{\partial p_i \partial p_j} \right] \\ & + \sum \left[\frac{\partial^2 f}{\partial A_h^2} \frac{\partial A_h}{\partial p_i} \frac{\partial A_h}{\partial p_j} + \frac{\partial^2 f}{\partial B_h^2} \frac{\partial B_h}{\partial p_i} \frac{\partial B_h}{\partial p_j} \right. \\ & + \frac{\partial^2 f}{\partial \Sigma_h^2} \frac{\partial \Sigma_h}{\partial p_i} \frac{\partial \Sigma_h}{\partial p_j} \\ & + \frac{\partial^2 f}{\partial A_h \partial B_h} \left(\frac{\partial A_h}{\partial p_i} \frac{\partial B_h}{\partial p_j} + \frac{\partial B_h}{\partial p_i} \frac{\partial A_h}{\partial p_j} \right) \\ & + \frac{\partial^2 f}{\partial \Sigma_h \partial A_h} \left(\frac{\partial \Sigma_h}{\partial p_i} \frac{\partial A_h}{\partial p_j} + \frac{\partial \Sigma_h}{\partial p_j} \frac{\partial A_h}{\partial p_i} \right) \\ & \left. + \frac{\partial^2 f}{\partial \Sigma_h \partial B_h} \left(\frac{\partial \Sigma_h}{\partial p_i} \frac{\partial B_h}{\partial p_j} + \frac{\partial \Sigma_h}{\partial p_j} \frac{\partial B_h}{\partial p_i} \right) \right]. \quad (23) \end{aligned}$$

This equation is general and can be applied to any function of the form (21) and for any parameters. Here we give its application for σ_A estimation (17).

A2. Derivatives of $-\log$ -likelihood

Derivatives of the $-\log$ -likelihood function when there is no phase information with respect to Σ have been given by Bricogne & Gilmore (1990). The derivatives of normalized $-\log$ -likelihood (again without phase information) with respect to σ_A when they are to be estimated in reciprocal-space resolution shells have been given by Read (1986). Here, we give derivatives of $-\log$ -likelihood in general form. In the following discussion only the acentric case is considered. The centric case can be derived by analogy. For simplicity Miller indices will be dropped. It is useful to rewrite the normalized version of (13-14) in the form,

$$\text{LLK} = \log \Sigma + \sum \frac{|E^\circ|^2 + |E_{wc}|^2}{\Sigma} - \log \int_0^{2\pi} P(\varphi) \times \exp 2|E^\circ| \frac{A_{E:wc} \cos \varphi + B_{E:wc} \sin \varphi}{\Sigma} d\varphi, \quad (24)$$

where $\Sigma = 2\sigma_{E:\varepsilon}^2 + \varepsilon(1 - \sigma_{wc}^2)$ and $A_{E:wc}$ and $B_{E:wc}$ are the real and imaginary parts of the weighted normalized structure factors E_{wc} . For convenience we use the following notation: if $t(\varphi)$ is a function of φ and $\mathcal{P}(\varphi)$ is the distribution of φ then $\langle t(\varphi) \rangle_{\mathcal{P}(\varphi)} = \int_0^{2\pi} t(\varphi) \mathcal{P}(\varphi) d\varphi$. If a prior distribution of phases is known $\mathcal{P}(\varphi)$ will be replaced by *prior*. In the above equation $P(\varphi)$ is the prior distribution,

$$P(\varphi) = \frac{P(\varphi) \exp 2|E^\circ| \frac{A_{E:wc} \cos \varphi + B_{E:wc} \sin \varphi}{\Sigma}}{\int_0^{2\pi} P(\varphi) \exp 2|E^\circ| \frac{A_{E:wc} \cos \varphi + B_{E:wc} \sin \varphi}{\Sigma} d\varphi}, \quad (25)$$

is the posterior distribution of phases, taking into account the prior distribution and the calculated phase of E_{wc} .

Only derivatives involving $A_{E:wc}$ are considered. Those involving $B_{E:wc}$ can be obtained by simple replacement of $A_{E:wc}$ by $B_{E:wc}$. Now using the facts that $\partial \log[y(x)] / \partial x = [\partial y(x) / \partial x] / y(x)$, $|E_{wc}|^2 = A_{wc}^2 + B_{wc}^2$ and changing the order of integration and differentiation we have [the subscript $\mathcal{P}(\varphi)$ will be dropped for brevity in some equations],

$$\frac{\partial \text{LLK}}{\partial A_{E:wc}} = \frac{2[A_{E:wc} - |E^\circ| \langle \cos \varphi \rangle_{\mathcal{P}(\varphi)}}{\Sigma} \quad (26)$$

$$\frac{\partial^2 \text{LLK}}{\partial A_{E:wc}^2} = \frac{2}{\Sigma} - \frac{4|E^\circ|^2}{\Sigma^2} [\langle \cos^2 \varphi \rangle_{\mathcal{P}(\varphi)} - \langle \cos \varphi \rangle_{\mathcal{P}(\varphi)}^2] \quad (27)$$

$$\frac{\partial^2 \text{LLK}}{\partial B_{E:wc} \partial A_{E:wc}} = -\frac{4|E^\circ|^2}{\Sigma^2} (\langle \sin \varphi \cos \varphi \rangle - \langle \sin \varphi \rangle \langle \cos \varphi \rangle) \quad (28)$$

$$\frac{\partial \text{LLK}}{\partial \Sigma} = \frac{1}{\Sigma} - \frac{[|E^o|^2 + |E_{\text{wc}}|^2 - 2|E^o|(A_{E:\text{wc}}\langle \cos \varphi \rangle + B_{E:\text{wc}}\langle \sin \varphi \rangle)]}{\Sigma^2} \quad (29)$$

$$\begin{aligned} \frac{\partial^2 \text{LLK}}{\partial \Sigma \partial A_{E:\text{wc}}} &= -\frac{2(A_{E:\text{wc}} - |E^o|\langle \cos \varphi \rangle)}{\Sigma^2} \\ &+ \frac{4|E^o|^2}{\Sigma^3} [A_{E:\text{wc}}^2(\langle \cos^2 \varphi \rangle - \langle \cos \varphi \rangle^2) \\ &+ A_{E:\text{wc}} B_{E:\text{wc}}(\langle \sin \varphi \cos \varphi \rangle - \langle \sin \varphi \rangle \langle \cos \varphi \rangle)]. \end{aligned} \quad (30)$$

$$\begin{aligned} \frac{\partial^2 \text{LLK}}{\partial \Sigma^2} &= -\frac{1}{\Sigma^2} \\ &+ \frac{2[|E_o|^2 + |E_c|^2 - 2|E^o|(A_{E:\text{wc}}\langle \cos \varphi \rangle + B_{E:\text{wc}}\langle \sin \varphi \rangle)]}{\Sigma^3} \\ &- \frac{4|E_o|^2}{\Sigma^4} [A_{E:\text{wc}}^2(\langle \cos^2 \varphi \rangle - \langle \cos \varphi \rangle^2) \\ &+ 2A_{E:\text{wc}} B_{E:\text{wc}}(\langle \sin \varphi \cos \varphi \rangle - \langle \sin \varphi \rangle \langle \cos \varphi \rangle) \\ &+ B_{E:\text{wc}}^2(\langle \sin^2 \varphi \rangle - \langle \sin \varphi \rangle^2)]. \end{aligned} \quad (31)$$

For $-\log$ -likelihood the singularity of derivatives of f with respect to A , B when $F^c \rightarrow 0$ which is characteristic for FLSQ, does not arise. For FLSQ it is assumed that $\langle \cos \varphi \rangle$ is identical to $\cos \varphi_c$ which is undetermined when F^c is zero. It can be verified that when $F^c \rightarrow 0$ then $\langle \cos \varphi \rangle \rightarrow \langle \cos \varphi \rangle_{\text{prior}}$. When no prior phase information is available $\langle \cos \varphi \rangle_{\text{prior}} = 0$. This is also true for $\langle \sin \varphi \rangle$, $\langle \cos 2\varphi \rangle$ and $\langle \sin 2\varphi \rangle$.

To calculate derivatives one needs to have the form of $\langle \cos \varphi \rangle$ and the other terms. Let us denote $m_{\text{comb}} = (\langle \cos \varphi \rangle^2 + \langle \sin \varphi \rangle^2)^{1/2}$, φ_{comb} as combination of calculated and prior phases,

$$\langle \cos \varphi \rangle = m_{\text{comb}} \cos \varphi_{\text{comb}} \quad (32)$$

$$\langle \sin \varphi \rangle = m_{\text{comb}} \sin \varphi_{\text{comb}} \quad (33)$$

$$\langle \cos^2 \varphi \rangle - \langle \cos \varphi \rangle^2 = \frac{1 + \langle \cos 2\varphi \rangle}{2} - m_{\text{comb}}^2 \cos^2 \varphi_{\text{comb}} \quad (34)$$

$$\langle \sin^2 \varphi \rangle - \langle \sin \varphi \rangle^2 = \frac{1 - \langle \cos 2\varphi \rangle}{2} - m_{\text{comb}}^2 \sin^2 \varphi_{\text{comb}} \quad (35)$$

$$\begin{aligned} &\langle \sin \varphi \cos \varphi \rangle - \langle \sin \varphi \rangle \langle \cos \varphi \rangle \\ &= \frac{\langle \sin 2\varphi \rangle}{2} - m_{\text{comb}}^2 \sin \varphi_{\text{comb}} \cos \varphi_{\text{comb}}. \end{aligned} \quad (36)$$

For the uniform phase information m_{comb} and φ_{comb} should be replaced by m_c and φ_c . In this case equations (32)–(36) can be simplified further.

To set up the Newton–Raphson equations used to estimate the parameters $\sigma_{A,j,0}$ and $C_{j,0}$ the first derivatives of $A_{E:\text{wc}}$, $B_{E:\text{wc}}$ and Σ must be calculated with respect to the parameters to be refined. $A_{E:\text{wc}} = \sum_j \sigma_{A,j} A_{E,j}$ and $B_{E:\text{wc}} = \sum_j \sigma_{A,j} B_{E,j}$. For simplicity we give the results for $\sigma_{A,j} = \sigma_{A,j,0} \exp -C_{j,0}|s|^2/4$.

$$\frac{\partial A_{\text{wc}}}{\partial \sigma_{A,j,0}} = A_{E,j} \exp -(C_{j,0}|s|^2/4) \quad (37)$$

$$\frac{\partial A_{\text{wc}}}{\partial C_{j,0}} = -\frac{|s|^2}{4} A_{E,j} \sigma_{A,j,0} \exp -(C_{j,0}|s|^2/4) \quad (38)$$

$$\frac{\partial \Sigma}{\partial \sigma_{A,j,0}} = -2\varepsilon \sigma_{A,j,0} \exp -(2C_{j,0}|s|^2/4) \quad (39)$$

$$\frac{\partial \Sigma}{\partial C_{j,0}} = \varepsilon \frac{|s|^2}{2} \sigma_{A,j,0}^2 \exp -(2C_{j,0}|s|^2/4). \quad (40)$$

Substituting the partial derivatives given in (26)–(31) into (23) it is seen that the terms involving the first derivatives of f with respect to $A_{E:\text{wc}}$, $B_{E:\text{wc}}$ and Σ become small as the gradient tends to zero and at the end stages of refinement will be much smaller than terms involving second derivatives. In the early stages of minimisation they might be large enough to make the second derivative matrix non-positive definite and so by analogy with the normal equation construction (Press *et al.*, 1986) we will drop these terms. The second derivatives can be simplified even more, but the equations given here are easily programmable.

Changing the residual would lead to different equations (26–31) and changing the parametric representation would lead to different equations (37)–(40).

APPENDIX B

Derivatives of $-\log$ -likelihood with respect to atomic parameters

Beginning with Agarwal (1978) various authors (Agarwal, Lifschitz & Dodson, 1980; Lunin & Urzhumtzev, 1985; Tronrud *et al.*, 1987; Brünger, 1989; Bricogne, 1993) showed how to calculate gradients using fast Fourier transformation and convolution of the difference map and atomic density. Here we restate these results in a form which is convenient and general for most residuals.

Before giving equations of derivatives let us define the notations. ρ is the atomic electron density expressed as a sum of Gaussians. ρ_{nm} is convolution of atoms n and m (ρ_{nm} centred at $r_n - r_m$ and ρ'_{nm} centred at $r_n + r_m$). For derivation of equations we use the following properties of the structure factors,

$$\begin{aligned}
F_h &= A_h + iB_h = \sum_{n=1}^{N_{\text{atom}}} [f_n \exp(2\pi i r_n h_i)] \\
F_h^* &= F_{-h} \\
A_h &= \frac{F_h + F_h^*}{2} = \frac{F_h + F_{-h}}{2} = A_{-h} \\
B_h &= \frac{F_h - F_h^*}{2i} = \frac{F_h - F_{-h}}{2i} = -B_{-h}
\end{aligned} \tag{41}$$

Now assume $r_n = (x_{n1}, x_{n2}, x_{n3})$ and f_n are the coordinate and form factor of the n th atom and $h = (h_1, h_2, h_3)$ is the Miller index of the reflecting plane. Here we write derivatives for coordinates. Writing the equation derived in *Appendix A* for reciprocal space: remember $\partial f / \partial A_h = \partial f / \partial A_{-h}$ and $\partial f / \partial B_h = -\partial f / \partial B_{-h}$.

$$\begin{aligned}
\frac{\partial f}{\partial x_{ni}} &= \sum \left(\frac{\partial f}{\partial A_h} - i \frac{\partial f}{\partial B_h} \right) 2\pi i h_i f_n \exp(2\pi i r_n h) \\
\frac{\partial^2 f}{\partial x_{ni} \partial x_{mj}} &= -\delta_{nm} \sum \left(\frac{\partial f}{\partial A_h} - i \frac{\partial f}{\partial B_h} \right) (2\pi)^2 h_i h_j f_n f_m \exp(2\pi i r_n h) \\
&\quad + \frac{1}{2} \sum \left(\frac{\partial^2 f}{\partial A_h^2} + \frac{\partial^2 f}{\partial B_h^2} \right) (2\pi)^2 h_i h_j f_n f_m \\
&\quad \times \exp[2\pi i h(r_n - r_m)] \\
&\quad - \frac{1}{2} \sum \left(\frac{\partial^2 f}{\partial A_h^2} - \frac{\partial^2 f}{\partial B_h^2} - 2i \frac{\partial^2 f}{\partial A_h \partial B_h} \right) \\
&\quad \times (2\pi)^2 h_i h_j f_n f_m \exp[2\pi i h(r_n + r_m)] \\
&= H_0(x_{ni}, x_{mj}) + H_1(x_{ni}, x_{mj}) + H_2(x_{ni}, x_{mj}).
\end{aligned} \tag{42}$$

The first derivative can be written in terms of a convolution of Fourier syntheses as,

$$\frac{\partial f}{\partial x_{ni}} = \mathcal{F} \left(\frac{\partial f}{\partial A_h} + i \frac{\partial f}{\partial B_h} \right) * \frac{\partial \rho_n}{\partial x_{ni}}. \tag{44}$$

First write diagonal terms in reciprocal space (only H_1 will be written as H_2 and H_0 are smaller than H_1),

$$H_1(x_{ni}, x_{nj}) = \frac{1}{2} \sum \left(\frac{\partial^2 f}{\partial A_h^2} + \frac{\partial^2 f}{\partial B_h^2} \right) (2\pi)^2 h_i h_j f_n^2. \tag{45}$$

Agarwal (1978) discussed the fast calculation of these terms for FLSQ and the same approach can be applied here.

Fast Fourier methods are needed to calculate the non-diagonal terms. The equations necessary for this are (H_0 will be dropped from consideration for the reason described in *Appendix A*),

$$\begin{aligned}
H_1(x_{ni}, x_{mj}) &= \frac{1}{2} \mathcal{F} \left(\frac{\partial^2 f}{\partial A_h^2} + \frac{\partial^2 f}{\partial B_h^2} \right) * \frac{\partial^2 \rho_{nm}}{\partial x_{ni} \partial x_{nj}} \\
H_2(x_{ni}, x_{mj}) &= -\frac{1}{2} \mathcal{F} \left(\frac{\partial^2 f}{\partial A_h^2} - \frac{\partial^2 f}{\partial B_h^2} + i2 \frac{\partial^2 f}{\partial A_h \partial B_h} \right) * \frac{\partial^2 \rho_{nm}}{\partial x_{ni} \partial x_{nj}}.
\end{aligned} \tag{46}$$

For terms close to the diagonal which involve atoms close to each other, the contribution of H_1 will be much larger than that of H_2 .

So far, the form of the function to be used has not been given. The coefficients for the $-\log$ -likelihood function assuming that σ_A 's are already estimated and will be held fixed (again only given for acentric reflections).

$$\begin{aligned}
\frac{\partial f}{\partial A_h} + i \frac{\partial f}{\partial B_h} &= 2 \frac{\sigma_A |E_c| \exp(i\varphi_c) - m_{\text{comb}} |E_o| \exp(i\varphi_{\text{comb}})}{\Sigma_h} \\
&\quad \times \frac{\sigma_A}{(\Sigma_c)^{1/2}}.
\end{aligned} \tag{48}$$

This function has the same symmetry as the original space group.

These coefficients resemble those given by Read (1986) to reduce map bias. However, since the σ_A are estimated using reflections excluded from the refinement a less biased map at any stage of refinement is obtained. This was found to be so in our test cases.

For the second derivative map (we give only for coefficients for H_1),

$$\frac{\partial^2 f}{\partial A_h} + \frac{\partial^2 f}{\partial B_h} = \left[\frac{4}{\Sigma} - \frac{4|E_o|^2}{\Sigma^2} (1 - m_{\text{comb}}^2) \right] \frac{\sigma_A^2}{\Sigma_c}. \tag{49}$$

This function has the same symmetry as the Patterson function of the original space group.

APPENDIX C

Moments of conditional distribution

For a probability distribution $P(F_1; F_2)$ the moments of function $T(F_1)$ can be written as,

$$\langle T(F_1) \rangle = \int_{\text{all values of } F_1} T(F_1) P(F_1; F_2) dF_1.$$

In our case $|F| > 0$ and the probability distribution of interest is a modified version of the Rice distribution. If phase information is available the analytical expression for moments is difficult to derive. Rice (1954) gave $\langle |F|^n \rangle$ for acentric reflections. Pannu & Read (1996) give first and second moments for acentric and centric reflections and use them to build a Gaussian approximation to the conditional distribution of amplitudes of the structure factors. Using Rice's and generalizing Pannu and Read's results then (here experimental uncertainties are ignored),

$$\langle |F|^n \rangle = \begin{cases} (\Sigma_{\text{wc}})^{\frac{n}{2}} \Gamma\left(\frac{n}{2} + 1\right) {}_1F_1\left(-\frac{n}{2}; 1; -\frac{|F_{\text{wc}}|^2}{\Sigma_{\text{wc}}}\right) & \text{acentric} \\ \left(\frac{1}{\pi}\right)^{1/2} (2\Sigma_{\text{wc}})^{\frac{n}{2}} \Gamma\left(\frac{n+1}{2}\right) {}_1F_1\left(-\frac{n}{2}; \frac{1}{2}; -\frac{|F_{\text{wc}}|^2}{2\Sigma_{\text{wc}}}\right) & \text{centric.} \end{cases}$$

${}_1F_1(a; b; c)$ is known as a confluent hypergeometric function or Kummer's function [for properties of these functions see Wang & Guo (1989); Rice (1954)]. Moments of even order of amplitudes of structure factors (consequently moments of intensities) are polynomials of $|F_{wc}|^2$ and Σ_{wc} . Moments of second and fourth order are interesting as they give the first and second moments of intensities,

$$\langle I \rangle = \langle |F|^2 \rangle = |F_{wc}|^2 + \Sigma_{wc} \quad \text{acentric and centric,}$$

and

$$\langle I^2 \rangle = \langle |F|^4 \rangle = \begin{cases} |F_{wc}|^4 + 4|F_{wc}|^2 \Sigma_{wc} + 2(\Sigma_{wc})^2 & \text{acentric} \\ |F_{wc}|^4 + 6|F_{wc}|^2 \Sigma_{wc} + 3(\Sigma_{wc})^2 & \text{centric.} \end{cases}$$

These moments could be used to build a Gaussian approximation for conditional distribution of intensities.

We thank the members of the protein structure group of Chemistry Department of University of York, for cheerfully testing the program, and Dr Phil Evans, Dr Misha Isupov and Dr Kostya Polyakov for useful comments. This work has been supported by EU BIOTECH grant BIO2CT-92-0524.

References

- Agarwal, R. C. (1978). *Acta Cryst.* **A34**, 791–809.
- Agarwal, R. C., Lifchitz, A. & Dodson, E. (1980). *Refinement of Protein structures. Proceedings of Daresbury Study Weekend*, edited by P. A. Machin, J. W. Campbell & M. Elder, pp. 36–39. Warrington: Daresbury Laboratory.
- Baker, E. N., Anderson, B. F., Dobbs, A. J. & Dodson, E. J. (1995). *Acta Cryst.* **D51**, 282–289.
- Booth, A. D. (1946). *Proc. Roy. Soc. London Ser. A*, **188**, 77–92.
- Booth, A. D. (1947). *Proc. Roy. Soc. London Ser. A*, **190**, 482–489.
- Box, G. E. P. & Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis*. Addison-Wesley.
- Bricogne, G. (1988). *Acta Cryst.* **A44**, 517–545.
- Bricogne, G. (1992). *Molecular Replacement. Proceedings of the CCP4 Study Weekend*, edited by E. J. Dodson, S. Gover & W. Wolf, pp. 62–75. Warrington: Daresbury Laboratory.
- Bricogne, G. (1993). *International Tables for Crystallography*, Vol. B, edited by U. Shmueli, pp. 23–182. Dordrecht: Kluwer Academic Press.
- Bricogne, G. & Gilmore, C. J. (1990). *Acta Cryst.* **A46**, 284–297.
- Brünger, A. T. (1989). *Acta Cryst.* **A45**, 42–50.
- Brünger, A. T. (1992). *X-PLOR manual*, Version 3.1. New Haven: Yale University Press.
- Brzozowski, A. M., Lawson, D. M., Frandsen, H. B., Svendsen, A., Borchert, T., Dauter, Z. & Dodson, G. G. (1997). *J. Mol. Biol.* In the press.
- Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Nature Struct. Biol.* **3**, 54–58.
- Carter, C. W. Jr, Crumley, K. V., Coleman, D. E., Hage, F. & Bricogne, G. (1990). *Acta Cryst.* **A46**, 57–68.
- Chapman, M. S. (1995). *Acta Cryst.* **A51**, 69–80.
- Cochran, W. (1948). *Acta Cryst.* **1**, 138–142.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cowtan, K. D. (1996). In *Macromolecular Refinement. Proceedings of CCP4 Study Weekend*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 23–28. Warrington: Daresbury Laboratory.
- Cruickshank, D. W. (1952). *Acta Cryst.* **5**, 511–518.
- Cruickshank, D. W. (1956). *Acta Cryst.* **9**, 747–753.
- Diamond, R. (1971). *Acta Cryst.* **A27**, 436–452.
- Driessen, H., Haneef, M. I. J., Harris, G. W., Howlin, B., Khan, G. & Moss, D. S. (1989). *J. Appl. Cryst.* **22**, 510–516.
- French, S. & Wilson, K. (1978). *Acta Cryst.* **A34**, 517–525.
- Hughes, E. W. (1941). *J. Am. Chem. Soc.* **63**, 1737–1752.
- Jack, A. & Levitt, M. (1978). *Acta Cryst.* **A34z**, 931–935.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Konnert, J. H. (1976). *Acta Cryst.* **A32**, 614–617.
- Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* **A36**, 344–350.
- Lamzin, V. S. & Wilson, K. S. (1993). *Acta Cryst.* **D49**, 129–147.
- Levitt, M. (1974) *J. Mol. Biol.* **82**, 393–420.
- Lunin, V. Yu. & Urzhumtsev, A. G. (1984). *Acta Cryst.* **A40**, 269–277.
- Lunin, V. Yu. & Urzhumtsev, A. G. (1985). *Acta Cryst.* **A41**, 327–333.
- Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.
- Oldfield, T. (1996). *Macromolecular Refinement. Proceedings of CCP4 Study Weekend*, pp. 67–74. Warrington: Daresbury Laboratory.
- Pannu, N. S. & Read, R. J. (1996). *Acta Cryst.* **A52**, 659–668.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986). *Numerical Recipes. The Art of Scientific Computing*. Cambridge University Press.
- Read, R. J. (1986) *Acta Cryst.* **A42**, 140–149.
- Read, R. J. (1990). *Acta Cryst.* **A46**, 900–912.
- Rice, S. O. (1954). *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax, pp. 133–195. New York: Dover.
- Sheldrick, G. M. (1995). *SHELXL93, a Program for the Refinement of Crystal Structures from Diffraction Data*. Institut für Anorganische Chemie, Göttingen, Germany.
- Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.
- Srinivasan, R. & Ramachandran, G. N. (1965). *Acta Cryst.* **19**, 1008–1014.
- Stuart, A & Ord, K. J. (1991). *Kendall's Advanced Theory of Statistics*, Vol. 2, 5th ed. London/Melbourne/Auckland: Edward Arnold.
- Tame, J. R. H., Murshudov, G. N., Dodson, E. J., Neil, T. K., Dodson, G. G., Higgins, C. F. & Wilkinson, A. J. (1994). *Science*, **264**, 1578–1581.
- Tronrud, D. (1996). *Macromolecular Refinement. Proceedings of CCP4 Study Weekend*, edited by E. Dodson, M.

- Moore, A. Ralph & S. Bailey, pp. 1–10. Warrington: Daresbury Laboratory.
- Tronrud, D. E. (1997). *Methods Enzymol.* In the press.
- Tronrud, D. E. Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* **A43**, 483–501.
- Wang, Z. X. & Guo, D. R. (1989). *Special Functions.* Singapore: World Scientists Publishing Co.
- Waser, J. (1963). *Acta Cryst.* **16**, 1091–1094.
- Watkin, D. (1994). *Acta Cryst.* **A50**, 411–437.
- Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
- Wilson, A. J. C. (1976). *Acta Cryst.* **A32**, 994–996.