Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Likelihood-enhanced fast rotation functions

Laurent C. Storoni, Airlie J. McCoy and Randy J. Read*

Department of Haematology, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 2XY, England

Correspondence e-mail: rjr27@cam.ac.uk

Received 1 September 2003
Accepted 16 December 2003

Experiences with the molecular-replacement program *Beast* have shown that maximum-likelihood rotation targets are more sensitive to the correct orientation than traditional targets. However, this comes at a high computational cost: brute-force rotation searches can take hours or even days of computation time on current desktop computers. Series approximations to the full likelihood target have been developed that can be computed by fast Fourier transforms in minutes. These likelihood-enhanced targets are more sensitive to the correct orientation than the Crowther fast rotation function and they take advantage of information from partial solutions. The likelihood-enhanced rotation targets have been implemented in the program *Phaser*.

1. Introduction

As the database of known macromolecular structures expands, molecular replacement grows in importance as a method to solve new crystal structures. Traditionally, molecular replacement has been performed with Patterson-based methods but, in common with many areas of macromolecular crystallography, likelihood-based methods have been coming to the fore. Bricogne (1992, 1997) first suggested that likelihood could be the basis for improved molecular-replacement algorithms, also pointing out that series approximations of the likelihood target could be computed rapidly by fast Fourier transform (FFT) algorithms. We have built on these suggestions, devising a new rotation likelihood function, providing a statistical treatment for the use of multiple molecular-replacement models and showing that likelihood indeed provides a more sensitive target (Read, 2001). Here, we describe fast approximations to the rotation likelihood function and their implementation in the new program *Phaser*.

Maximum likelihood asserts that the best model is that which maximizes the probability of having made the set of observations in the experiment. For diffraction experiments, the observations are intensities, which can be transformed into structure-factor amplitudes. In a molecular-replacement search, the observed amplitudes can be predicted from the structure factors computed from a model in a trial orientation and/or position. For each reflection, the total model structure factor is the sum of the structure-factor components, or molecular transforms, from the symmetry-related copies of the model in the unit cell. In a rotation search, the amplitude, but not the phase, of the molecular transform contributed by each symmetry-related model in the unit cell is known. Because the relative phases are unknown, the molecular transforms cannot simply be summed to calculate the total structure-factor amplitude for comparison with the data.

However, probability functions for the amplitudes of the structure factors can still be derived. The form of the probability function varies with the assumptions made about the nature of the molecular-transform components contributing to the sum. If it is assumed that the central limit theorem applies, *i.e.* that there are many symmetry-related molecules in the unit cell, that they contribute independently to the sum and that none of the individual molecular transforms dominates the total structure factor, then the distribution can be approximated by a 'random walk', given in the acentric case by a two-dimensional Gaussian centred on the origin of the complex plane. This has the same functional form as that described by Wilson (1949) for the probability distribution of structure factors arising from a random distribution of atoms.

In a rotation search, the translation vector is unknown and can be considered to be a variable distributed randomly over the unit cell. The relative phases of the molecular transforms for symmetry-related molecules are determined by this vector, so the relative phases are not really independent random variables. Nonetheless, all possible relative phases between pairs of molecular transforms are sampled as the translation vector varies over the cell. Numerical simulations show that for most space groups the distribution of amplitudes agrees well with the distribution predicted for a random walk with truly independent relative phases. The poorest approximations are seen for the polar space groups with fourfold and sixfold axes. In spite of this, rotation functions based on the assumption of a Wilson distribution still work well. In addition, in these space groups the translation search need only be carried out over a plane, so it is computationally inexpensive to test a large number of potential orientations with a subsequent translation search.

In the following, the expressions presented previously (Read, 2001) are rearranged in order to make the approximations that will be developed more intuitive. For computational convenience, we compute the log of the likelihood, which has its maximum for the same values of the parameters as the likelihood. If the reflections are assumed to be independent, the total log-likelihood in the Wilson approximation is given in terms of the observed structure factors $F_o(\mathbf{h})$ by

$$\ln\left[\prod_{\mathbf{h}} p_a(F_o; F_{j,s})\right] = \sum_{\mathbf{h}} \ln\left[\frac{2F_o}{\varepsilon\Sigma_W}\exp\left(-\frac{F_o^2}{\varepsilon\Sigma_W}\right)\right] \quad (1)$$

for acentric reflections and

$$\ln\left[\prod_{\mathbf{h}} p_c(F_o; F_{j,s})\right] = \sum_{\mathbf{h}} \ln\left[F_o\left(\frac{2\pi}{\varepsilon\Sigma_W}\right)^{1/2}\exp\left(-\frac{F_o^2}{2\varepsilon\Sigma_W}\right)\right] \quad (2)$$

for centric reflections, where

$$\Sigma_W(\mathbf{h}) = \Sigma_{N'}(\mathbf{h}) + \Sigma_{\rm rot}(\mathbf{h}), \quad (3)$$

$$\Sigma_{N'}(\mathbf{h}) = \Sigma_N + \sum_{j_f} D_{j_f}^2 F_{j_f}^2(\mathbf{h}) - \langle D_{j_f}^2 F_{j_f}^2\rangle, \quad (4)$$

$$\Sigma_{\rm rot}(\mathbf{h}) = \sum_{j_r,s} D_{j_r,s}^2 F_{j_r,s}^2(\mathbf{h}) - \langle D_{j_r,s}^2 F_{j_r,s}^2\rangle, \quad (5)$$

where j_f and j_r refer to the fixed (*i.e.* non-rotating) and rotating molecules, respectively, and s labels the symmetry-related molecules in the unit cell. Σ_N is the bare variance of the Wilson distribution, in which nothing is known apart from the unit-cell content. Σ_W is a perturbatively corrected variance that takes into account the acquisition of extra information. The factor ε accounts for the statistical effect of symmetry on the expected intensity. The D factors are, intuitively, the fraction of the calculated structure-factor components that are correlated to the true values. Each F_{j_f} represents a structure-factor component with unknown relative phase compared with other components and may represent the sum of a number of molecular transforms with known relative phase.

For instance, if the orientation but not the position of a fixed molecule is known, each symmetry-related copy of the fixed molecule will contribute a separate F_{j_f} term. On the other hand, if the position is also known, those contributions have known relative phase and will be summed to give one combined F_{j_f} term.

If instead it is assumed that one component of the structure factor dominates the distribution, the probability function for the amplitude of the structure factor can be modelled as a random walk around the end of the dominating structure-factor component $F_{\rm big}$. For the acentric case, this gives a two-dimensional Gaussian centred on the end of the dominating structure factor. Integrating out the phase gives a probability function similar to that described by Sim (1959), denoted the Rice distribution in statistical literature,

$$\ln\left[\prod_{\mathbf{h}} p_a(F_o; F_{j,s})\right] = \sum_{\mathbf{h}} \ln\left[\frac{2F_o}{\varepsilon\Sigma_S}\exp\left(-\frac{F_o^2 + F_{\rm big}^2}{\varepsilon\Sigma_S}\right) I_0\left(\frac{2F_oF_{\rm big}}{\varepsilon\Sigma_S}\right)\right], \quad (6)$$

where

$$\Sigma_S(\mathbf{h}) = \Sigma_{N'}(\mathbf{h}) + \Sigma_{\rm rot}(\mathbf{h}) - F_{\rm big}^2(\mathbf{h}). \quad (7)$$

If the assumption that the central limit theorem holds is also relaxed, then the distribution can be described in terms of a Fourier–Bessel series, analogous to distributions obtained by Shmueli & Weiss (1995). The Rice distribution has been implemented in the program *Beast* and the Wilson and Rice target functions implemented in the program *Phaser*. The Fourier–Bessel series target function has not been implemented in the context of molecular replacement.

Brute-force searches using maximum-likelihood rotation functions are computationally intensive. They share this problem with the Rossmann & Blow (1962) rotation function, where several techniques have been used to reduce computation time. For instance, Tollin & Rossmann (1966) used the observation that the rotation-function values are dominated by strong reflections to approximate the function by only using these terms ('large terms') in the calculation. However, the most effective and most widely used method for reducing computation time is the fast rotation function developed by Crowther (1972), in which the Patterson is decomposed into orthogonal spherical harmonics and Bessel functions and the rotation function is calculated in Eulerian β sections with a fast Fourier transform (FFT). Here, we show how the Wilson likelihood target can be expanded as a Taylor series and the

initial terms in this series used for a similar fast Fourier transform calculation. This speeds up the calculations by several orders of magnitude, with minimal loss of sensitivity. If maximal sensitivity is required, the fast rotation functions can be used to generate a list of plausible orientations, which are then rescored with the full likelihood target. A similar time-saving strategy is used in *AMoRe* (Navaza, 1994). Although we have implemented our fast rotation function as a series of two-dimensional FFTs, following Crowther (1972), it could also have been implemented as a single three-dimensional FFT, as performed by Kovacs & Wriggers (2002).

2. Series expansion of Wilson likelihood function

The Rice likelihood function is slightly better for brute-force rotation searches than the Wilson likelihood function, but it requires one to pick out the largest contribution to each structure factor for each orientation. The Wilson likelihood function does not have this restriction and all contributions to the structure factor are treated equally, so it provides a more suitable starting point for the derivation of a fast rotation function. In practice, there is little difference in sensitivity between brute-force rotation searches carried out with the Wilson and Rice likelihood functions.

Starting from (1), we obtain for an acentric reflection

$$\ln\left[p_a(F_o; F_{j,s})\right] = \ln\left(\frac{2F_o}{\varepsilon\Sigma_W}\right) - \frac{F_o^2}{\varepsilon\Sigma_W} = \ln\left(\frac{2F_o}{\varepsilon\Sigma_{N'}}\right) - \ln(1+\xi) - \frac{F_o^2}{\varepsilon\Sigma_{N'}}(1+\xi)^{-1}, \quad (8)$$

where we have defined the quantity

$$\xi = \frac{\Sigma_{\rm rot}}{\Sigma_{N'}} \quad (9)$$

so that

$$\Sigma_W = \Sigma_{N'} + \Sigma_{\rm rot} = \Sigma_{N'}(1+\xi). \quad (10)$$

ξ is a perturbation term that will be distributed about zero, so that an approximation with a Taylor series is most accurate in the region that will be sampled in the search. The quantity ξ is assumed to be small so we can Taylor expand $\ln(1+\xi)$ and $(1+\xi)^{-1}$. Both Taylor series have a radius of convergence of 1, *i.e.* they converge when $\Sigma_{\rm rot} < \Sigma_{N'}$. We do not expect this series to converge for all orders in ξ, but numerical simulations show that the first few terms are a good approximation to the full likelihood. If we collect terms in powers of ξ we arrive at

$$\ln(p_a) = C_a + \sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{F_o^2}{\varepsilon\Sigma_{N'}} - \frac{1}{n}\right)\xi^n, \quad (11)$$

where

$$C_a = \ln\left(\frac{2F_o}{\varepsilon\Sigma_{N'}}\right) - \frac{F_o^2}{\varepsilon\Sigma_{N'}} \quad (12)$$

is independent of the rotation angle and thus only contributes an overall irrelevant term.

Similarly, for a centric reflection we obtain the expansion

$$\ln(p_c) = C_c + \frac{1}{2}\sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{F_o^2}{\varepsilon\Sigma_{N'}} - \frac{1}{n}\right)\xi^n, \quad (13)$$

where

$$C_c = \frac{1}{2}\left[\ln\left(\frac{2\pi F_o^2}{\varepsilon\Sigma_{N'}}\right) - \frac{F_o^2}{\varepsilon\Sigma_{N'}}\right]. \quad (14)$$

3. Likelihood-enhanced fast rotation functions

Ignoring the overall constants, the likelihood function becomes

$$\ln\left[\prod_{\mathbf{h}} p(F_o; F_{j,s})\right] \propto \sum_{\mathbf{h}} w_{\mathbf{h}} \sum_{n=1}^{\infty}(-1)^{n-1}\left(\frac{F_o^2}{\varepsilon\Sigma_{N'}} - \frac{1}{n}\right)\xi^n, \quad (15)$$

where $w_{\mathbf{h}} = 1$ for acentric and $w_{\mathbf{h}} = 0.5$ for centric reflections. This function could be used for a direct rotation function, in which the model contributions are computed as a function of orientation for each reflection, in terms of the unit cell for the observed data. By Parseval's theorem, such a direct rotation function would correspond to the integral of the product of the Fourier transforms of the observed and calculated terms. The first-order terms are intensities, so the first-order term would correspond to a Patterson overlap function. However, for the fast rotation function the model terms cannot be recomputed for every orientation. Pursuing the analogy with Patterson overlap functions, we can consider only the volume within a sphere of the origin of the real-space (Patterson or higher order 'Patterson of a Patterson') function. Expanding to $P1$ and transforming that overlap integral back into reciprocal space, we obtain the form

$$\sum_{\mathbf{h}}\sum_{\mathbf{k}}\sum_{n=1}^{\infty}(-1)^{n-1}\left[\frac{F_o(\mathbf{h})^2}{\varepsilon\Sigma_{N'}} - \frac{1}{n}\right]\left[\frac{\Sigma_{\rm rot}(\mathbf{k})}{\Sigma_{N'}}\right]^n \chi_\Omega(\mathbf{h} - \mathbf{k}\mathbf{R}^{-1}), \quad (16)$$

where $\chi_\Omega(\mathbf{p}) = (1/V_\Omega)\int_\Omega d^3\mathbf{x}\exp(2\pi i\mathbf{p}\cdot\mathbf{x})$ is the Fourier transform of the sphere, $\Sigma_{\rm rot}$ is computed for the model in its initial orientation and $\mathbf{R}$ is the rotation matrix describing the rotation of the model. The factor $w_{\mathbf{h}}$ disappears because, in the expansion to $P1$, centric reflections contribute only half as many unique terms as acentric reflections. (16) is of the form

$$\sum_{\mathbf{h}}\sum_{\mathbf{k}}\left[I_t(\mathbf{h})I_s(\mathbf{k})\chi_\Omega(\mathbf{h} - \mathbf{k}\mathbf{R}^{-1})\right], \quad (17)$$

which, ignoring scalings, is similar to equation (15) in Navaza (2001). We can thus identify $I_t(\mathbf{h})$ and $I_s(\mathbf{k})$ with the terms in our series expansion and use the methods outlined in Navaza (2001) to compute a fast rotation function. The first term in this series becomes

$$I_1^t(\mathbf{h}) = \frac{1}{\Sigma_{N'}}\left[\frac{F_o^2(\mathbf{h})}{\varepsilon\Sigma_{N'}} - 1\right], \quad (18)$$

$$I_1^s(\mathbf{k}) = \Sigma_{\rm rot}(\mathbf{k}) = \sum_s\left[D^2F_{j_r,s}^2(\mathbf{k}) - \langle D^2F_{j_r,s}^2\rangle\right]. \quad (19)$$

In fact, each symmetry-related component of I_1^s contributes equally to the products with symmetry-related I_1^t terms, so it is sufficient to include only one symmetry-related contribution to I_1^s,

$$I_1^s(\mathbf{k}) = D^2F_{j_r}^2(\mathbf{k}) - \langle D^2F_{j_r}^2\rangle. \quad (20)$$

This is a scaled and variance-weighted version of the Patterson overlap function used in the Crowther target. We call this approximation LERF1 (likelihood-enhanced rotation function of order 1). The second term in the approximation series is

$$I_2^t(\mathbf{h}) = -\frac{1}{\Sigma_{N'}^2}\left[\frac{F_o^2(\mathbf{h})}{\varepsilon\Sigma_{N'}} - \frac{1}{2}\right], \quad (21)$$

$$I_2^s(\mathbf{k}) = \left[D^2F_{j_r}^2(\mathbf{k}) - \langle D^2F_{j_r}^2\rangle\right]^2. \quad (22)$$

The physical interpretation of this term is more difficult as it involves the square of the intensities for the rotated model. This can be thought of as the 'Patterson of a Patterson'. We refer to the second-order approximation that includes this term as LERF2.

It is also worth noting that this term does not include any cross-terms involving symmetry-related models with different symmetry operations. The term presented here is the best that can be performed in a 'fast' approximation.

In the case where there is no fixed model, LERF1 reduces to the fast rotation function proposed by Bricogne (1997), which does not have a means to exploit prior structural information. Such a rotation-function target is equivalent to one we proposed based on more heuristic arguments (Read, 1990*a*,*b*). In the absence of a fixed model, $\Sigma_{N'} = \Sigma_N$, which is no longer a function of **h** but just a radially symmetric function so that one factor of Σ_N^{-1} can be taken into the model Patterson coefficients. The coefficients for the observed Patterson map become $(E_0^2 - 1)$, corresponding to a sharpened origin-removed Patterson. The effect of the D values in the model Patterson coefficients is to make this the Patterson corresponding to the expected electron density, *i.e.* the Patterson corresponding to a model in which the atoms have been smeared over their possible positions (Read, 1990*a*). The normalization factor Σ_N converts the D-weighted amplitudes into σ_A-weighted E values, giving the coefficients $\sigma_A^2(E_C^2 - 1)$.

It is interesting to consider how the information from a fixed model is introduced into the coefficients of LERF1. The structure-factor contributions from a fixed model comprise part of $\Sigma_{N'}$, which appears in both the numerator and the denominator of the observed Patterson coefficient. The effect of the $\Sigma_{N'}$ term in the numerator is to subtract from the observed Patterson coefficient the contributions to the Patterson of the fixed molecules. The effect of the $\Sigma_{N'}$ term in the denominator is to downweight structure factors for which there is a large relative contribution from the fixed model components, reflecting the fact that much of the observed intensity has already been accounted for.

It has been proposed previously that the component of the observed Patterson explained by the partial model could be removed by using the coefficients $|F_o|^2 - |F_c|^2$ (Nordman, 1994; Zhang & Matthews, 1994). This accounts for part of the effect of $\Sigma_{N'}$. Alternatively, Dauter *et al.* (1991) replaced the observed Patterson coefficients $|F_o|^2$ by $(|F_o| - |F_c|)^2$ in searching for the eglin-C component of a complex with subtilisin, after finding and refining the subtilisin component.

4. Implementation

The target functions described in the previous section were implemented in the program *Phaser*. For efficiency, the calculations were performed in terms of normalized structure factors (E values). In order to reduce the errors in the calculation of spherical harmonics the method of Navaza (2001) was used. The e_{lmn} values were given by

$$e_{lmn}^{t,s\ (i)} = s(n)\sum_{\mathbf{h}} I_i^{t,s}(\mathbf{h})Y_l^m(\hat{\mathbf{h}})\frac{j_{l+2n-1}(2\pi hb)}{2\pi hb}, \quad (23)$$

where $s(n) = \{12\pi[2(l+2n)-1]\}^{1/2}$. The (i) refers to either the first- or second-order approximation. The radial integral (in hkl space) is then given by the sum over n,

$$C_{m,m'}^{l\ (i)} = \sum_{n=1}^{N} \overline{e_{lmn}^{t\ (i)}} e_{lm'n}^{s\ (i)}. \quad (24)$$

The search target LERF1 only uses the first term in the expansion

$$\mathcal{R}_{\mathrm{LERF1}} = \mathrm{FT}\left[\sum_{l=2}^{L} C_{m,m'}^{l\ (1)} d_{m,m'}^l(\theta)\right], \quad (25)$$

where FT is the Fourier transform of the indices m and m' and $d_{m,m'}^l$ are the irreducible rotation group matrices about the y axis.

Similarly, the LERF2 target is given by

$$\mathcal{R}_{\mathrm{LERF2}} = FT\left\{\sum_{l=2}^{L}\left[C_{m,m'}^{l\ (1)} + C_{m,m'}^{l\ (2)}\right] d_{m,m'}^l(\theta)\right\}. \quad (26)$$

For test purposes, we have also implemented the Crowther target. As in the *AMoRe* implementation, the low-order spherical harmonics can be omitted. By default, all terms with $l \geq 2$ are used.

Phaser reports the fast rotation function score as a fraction of the maximum value and optionally rescores the top solutions with a log-likelihood gain (LLG) score. The LLG score is the difference between the log-likelihood score for the molecular-replacement trial and the log-likelihood from the Wilson distribution. It thus measures how much better the data can be predicted from the model than from a set of random atoms.

5. Test cases

5.1. β-Lactamase and β-lactamase inhibitor protein complex

The structure of the complex between β-lactamase (BETA) and β-lactamase inhibitor protein (BLIP) was solved (Strynadka *et al.*, 1996) using the program *AMoRe* (Navaza, 1994). Although the BETA (62% of content of asymmetric unit) component was easily the top peak in the rotation function, the BLIP component (38% of asymmetric unit) did not produce a clear signal. It was only by scoring a large number of

rotation peaks with the translation function that the correct solution was found in the original structure determination.

We used anisotropically corrected data for our tests (McCoy, Storoni & Read, work to be published). Even though the LERF1 and LERF2 targets are able to find the correct orientations without the anisotropic corrections, the rescoring with the full likelihood target does not give the best score for the correct peak. This is because the full likelihood target is more sensitive to the effects of anisotropy than are its low-order approximations.

We tested LERF1 and LERF2 by comparing the relative score of the correct peak and the highest incorrect peak. The incorrect scores represent orientations that are not clustered to the correct peak. As the results in Table 1 show, the fast rotation function in *Phaser* using either LERF1 or LERF2 targets is able to distinguish the correct solution.

We also tested the approximations by rescoring the peaks produced by the Crowther target and LERF1 with the full likelihood target. Figs. 1 and 2 show the correlation between the fast rotation targets and the full likelihood targets. In Fig. 1 the Crowther target was used and the correlation coefficient was 0.788. Fig. 2 shows the same results with LERF1, where we obtained the significantly higher correlation coefficient of 0.927. The results for LERF2 (Fig. 3) were comparable to those for LERF1 and the correlation coefficient was also 0.927. The figures only show the rescoring of clustered peaks. A similar test involving the rescoring of all rotation function values gave very similar results, with correlation coefficients of 0.815 and 0.938 for the Crowther target and LERF1, respectively.

Phaser also allows the information from the known partial model of BETA to be used to increase the signal to noise. When the known orientation of BETA is added to the search, the correct solution has an even more obvious peak in the function, as can be seen in Table 1. When both the orientation and the position of BETA are fixed, the solution becomes even clearer.

Table 1
Rotation-function results for the BLIP component.

Results are expressed as *Z* scores, *i.e.* r.m.s. deviations above the mean score.

Prior information	No prior information	Fix orientation of BETA	Fix position of BETA
Crowther score			
Correct	3.89	—	—
Top incorrect	4.50	—	—
LERF1 score			
Correct	4.23	4.54	5.21
Top incorrect	3.77	3.69	3.87
LERF2 score			
Correct	4.26	4.53	5.00
Top incorrect	3.98	3.95	4.06
LLG score			
Correct	4.50	4.92	6.08
Top incorrect	3.94	3.62	3.96

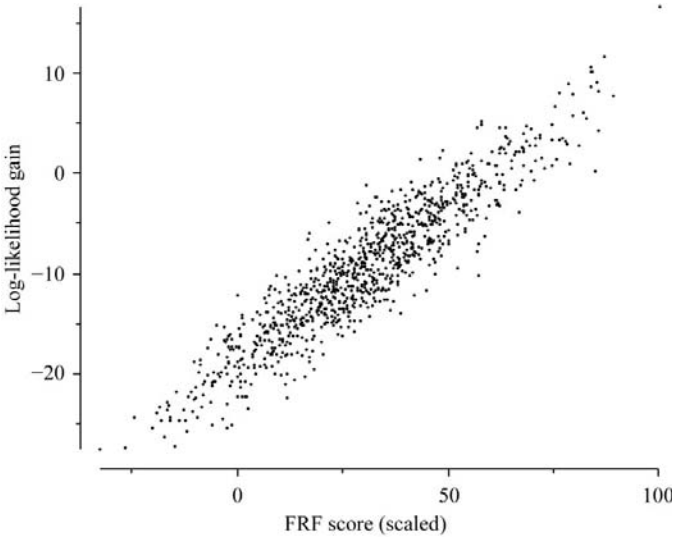

Figure 2
Scatter plot showing correlation between LERF1 and LLG for the BLIP component.

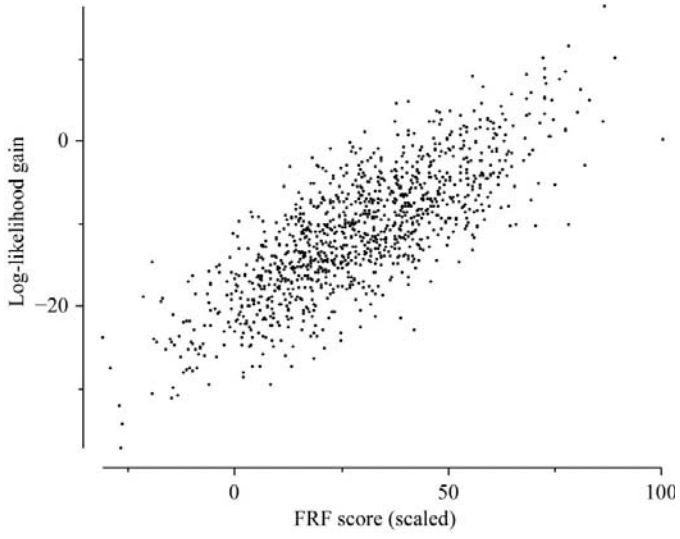

Figure 1
Scatter plot showing correlation between Crowther target and LLG for the BLIP component.

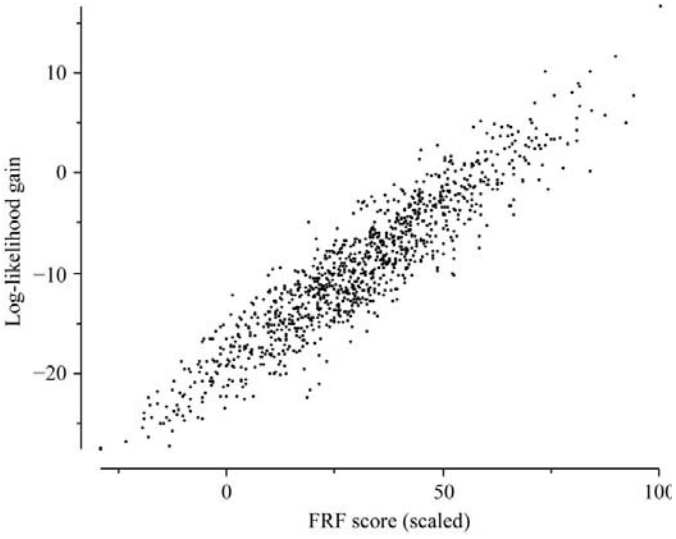

Figure 3
Scatter plot showing correlation between LERF2 and LLG for the BLIP component.

Table 2
Rotation-function results for the third dimer of the MoaE subunit of molybdopterin synthase.

Results are expressed as *Z* scores, *i.e.* r.m.s. deviations above the mean score.

Prior information	No prior information	Fix orientations of dimers 1 and 2	Fix positions of dimers 1 and 2
Crowther score			
Correct	5.72	—	—
Top incorrect	4.76	—	—
LERF1 score			
Correct	5.52	8.36	9.69
Top incorrect	4.34	4.21	4.28
LLG score			
Correct	5.62	7.44	6.52
Top incorrect	4.57	3.92	3.31

The knowledge of the placement of the BETA component is a very powerful piece of information. With the additional information, we were able to solve the BETA and BLIP complex by using only data to 6 Å resolution. Without the known solution for BETA, there is no signal in the BLIP search at this resolution.

5.2. Molybdopterin synthase MoaE subunit deletion mutant

Escherichia coli molybdopterin synthase is a tetramer comprising two copies each of subunits MoaD and MoaE. A deletion mutant of MoaE forms a stable homodimer that can be crystallized. When its structure was determined, two of three dimers in the asymmetric unit could be located by molecular replacement, but not the third (Rudolph *et al.*, 2003). In retrospect, the third dimer is probably more difficult to locate because it has significantly higher overall *B* factors (55 compared with 42 Å^2 for the other two dimers). When the structure was solved, the third dimer was found rather inventively using *ARP*/*wARP* (Perrakis *et al.*, 1999) to build into unexplained density based on phases from the first two dimers and then superimposing the molecular-replacement model on the partial model. We investigated this structure to see whether *Phaser* would have been capable of finding the third dimer more easily, taking advantage of the ability of LERF1 and LERF2 to exploit the information from a partial solution.

In the published molecular-replacement searches, only data from a restricted resolution shell (8–4 Å) were used, presumably to reduce CPU requirements. Because LERF1 and LERF2 can be computed much more efficiently than the brute-force likelihood functions in *Beast*, we used data to 3 Å resolution in these tests. As in the published structure determination, the molecular-replacement model was a MoaE dimer constructed from the structure of molybdopterin synthase (Rudolph *et al.*, 2000), entry 1fm0 in the Protein Data Bank (Berman *et al.*, 2000). As reported in the original structure determination, the solution for the first two dimers is extremely clear, with *Z* scores above 8 for the rotation search. In fact, when data to 3 Å resolution are used, there is also a clear signal for the third dimer, although it is considerably weaker (Table 2). Using only the restricted 8–4 Å resolution shell, the orientation for the third dimer is lost in the noise (results not shown). As the results in Table 2 demonstrate, the addition of prior information about the first two dimers dramatically increases the signal in both the LERF1 fast rotation function and in the full rotation likelihood function. Even with the restricted resolution shell, the correct orientation for the third dimer is unambiguous when information about the first two is added (results not shown). Using the new methods in *Phaser*, this structure solution would have been straightforward.

It is interesting to note that the LLG score performs less well than LERF1 in this case when information about fixed molecules is provided. In the current implementation of the likelihood function and its approximations, it is assumed that all molecules have similar overall *B* factors and thus that they contribute in proportion to their molecular weight at all resolutions. In this case, it is assumed that in fixing the first two dimers two-thirds of the scattering has been accounted for, but at high resolution the proportion will be significantly higher. The LLG score is more sensitive than its low-order approximations LERF1 and LERF2 to errors in the estimate of the proportion of scattering that has been accounted for, because it is more sensitive to larger deviations from expected values. In future work, we will explore the possibility of refining the relative *B* factors of partial solutions and the missing components.

6. Conclusions

The likelihood-enhanced fast rotation functions are an effective approximation to the full likelihood target, with the advantage of being several orders of magnitude faster. Where the fast rotation function fails to find the correct orientation, the correct solution is higher in the list than with the Crowther function. Therefore, this list is a better starting point for subsequent translation searches or for rescoring with the full likelihood than the Crowther method. The lack of improvement of LERF2 over LERF1, the increased computation time arising from the extra calculations and the possible lack of convergence of the expansion series suggest that there is no real advantage in using even higher order terms in the series, *i.e.* implementing LERF*n*. In fact, a variety of tests (not shown) demonstrate that LERF2 is not systematically better or worse than LERF1.

The likelihood target for the translation function is equally as computationally intensive as the rotation function and also takes hours to days to compute on current desktop computers. It is possible to formulate a likelihood-enhanced fast translation function that can also be calculated with the fast Fourier transform. We expect that combining the likelihood-enhanced fast rotation function with a likelihood-enhanced fast translation function will provide a powerful and fast method for molecular-replacement structure solution in macromolecular crystallography. The program *Phaser* will be released as part of the *CCP*4 and *PHENIX* software suites and is also available from the authors (see http://www-structmed.cimr.cam.ac.uk/phaser for details).

We are grateful to Michael James and Natalie Strynadka for supplying the data for the β-lactamase complex test case and to Rudolph and coworkers for depositing their diffraction data in the PDB. This work was funded by NIH/NIGMS under grant No. 1P01GM063210 and by a Principal Research Fellowship from the Wellcome Trust (RJR).

References

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Bricogne, G. (1992). *Proceedings of the CCP4 Study Weekend. Molecular Replacement*, edited by W. Wolf, E. J. Dodson & S. Gover, pp. 62–75. Warrington: Daresbury Laboratory.

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

Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.

Dauter, Z., Betzel, C., Genov, N., Pipon, N. & Wilson, K. S. (1991). *Acta Cryst.* B**47**, 707–730.

Kovacs, J. A. & Wriggers, W. (2002). *Acta Cryst.* D**58**, 1282–1286.

Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.

Navaza, J. (2001). *Acta Cryst.* D**57**, 1367–1372.

Nordman, C. E. (1994). *Acta Cryst.* A**50**, 68–72.

Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

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

Read, R. J. (1990*b*). Am. Crystallogr. Assoc. Annu. Meet., Abstract C08.

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

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Rudolph, M. J., Wuebbens, M. M., Rajagolpalan, K. V. & Schindelin, H. (2000). *Nature Struct. Biol.* **8**, 42.

Rudolph, M. J., Wuebbens, M. M., Turque, O., Rajagopalan, K. V. & Schindelin, H. (2003). *J. Biol. Chem.* **278**, 14514–14522.

Shmueli, U. & Weiss, G. H. (1995). *Introduction to Crystallographic Statistics.* Oxford University Press.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813–815.

Strynadka, N. C., Jensen, S. E., Alzari, P. M. & James, M. N. (1996) *Nature Struct. Biol.* **3**, 290–297.

Wilson, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Tollin, P. & Rossmann, M. G. (1966). *Acta Cryst.* **21**, 872–876.

Zhang, X.-J. & Matthews, B. W. (1994). *Acta Cryst.* D**50**, 675–686.