Direct incorporation of experimental phase information in model refinement

Pavol Skubák,[a] Garib N. Murshudov[b] and Navraj S. Pannu[a]*

[a]Biophysical Structural Chemistry, Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands, and [b]Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, England

Correspondence e-mail:
raj@chem.leidenuniv.nl

The incorporation of prior phase information into a maximum-likelihood formalism has been shown to strengthen model refinement. However, the currently available likelihood refinement target using prior phase information has shortcomings; the 'phased' refinement target considers experimental phase information indirectly and statically in the form of Hendrickson–Lattman coefficients. Furthermore, the current refinement target implicitly assumes that the prior phase information is independent of the calculated model structure factor. This paper describes the derivation of a multivariate likelihood function that overcomes these shortcomings and directly incorporates experimental phase information from a single-wavelength anomalous diffraction (SAD) experiment. This function, which simultaneously refines heavy-atom and model parameters, has been implemented in the refinement program REFMAC5. The SAD function used in conjunction with the automated model-building procedures of ARP/wARP leads to a successful solution when current likelihood functions fail in a test case shown.

1. Introduction

A great deal of information is gained in experimentally phasing a molecule, yet the default procedure for automated model-building procedures combined with iterative structure refinement (Perrakis *et al.*, 1999; Terwilliger, 2003) only considers the diffraction data obtained from the native crystal and neglects any available experimental phase information.

Previously, the incorporation of prior phase information has been shown to strengthen model refinement (Pannu *et al.*, 1998). However, the functional form of the likelihood refinement target encodes the prior phase information statically in the form of Hendrickson–Lattman coefficients (Hendrickson & Lattman, 1970). The likelihood function is then dependent on the reliability and accuracy of the phasing program used to generate the coefficients and does not allow the simultaneous refinement of the associated heavy-atom and model parameters. Finally, the derivation of the current refinement target incorporating prior phase information assumes that the prior phase distribution is independent of the model. This assumption is incorrect, as the phase information is used to build the model. All of the above shortcomings of the likelihood function utilizing prior phase information from Hendrickson–Lattman coefficients probably contributed to the reluctance to include prior information in automated model-building procedures.

To overcome these assumptions, a multivariate analysis directly modelling the correlations and errors in a phasing experiment and model refinement should be applied; the

resulting multivariate function would directly consider the diffraction data collected in the experiment. Multivariate statistics have played an important role in crystallography (*e.g.* see Bricogne, 2000) and joint probability distributions have recently led to promising results in substructure detection (Burla *et al.*, 2002) and phasing (Giacovazzo & Siliqi, 2001*a*,*b*, 2004; Pannu *et al.*, 2003; Pannu & Read, 2004).

Below, we derive a multivariate single-wavelength anomalous diffraction (SAD) likelihood function that directly incorporates the measured Friedel pairs, $|F^+|$ and $|F^-|$, and the associated calculated model structure factors into structure refinement. The function allows for the simultaneous refinement of the heavy-atom and model parameters and thus directly and dynamically considers the experimental phase information from a SAD experiment.

The SAD likelihood function has been implemented in the program *REFMAC*5 (Murshudov *et al.*, 1997) from the *CCP*4 package (Collaborative Computational Project, Number 4, 1994). The newly implemented SAD function was compared with other refinement targets and performs favourably. In a test case, the SAD function in conjunction with the automated model-building procedures implemented in *ARP/wARP* (Perrakis *et al.*, 1999) leads to a correctly built model when current likelihood functions fail.

2. Implementation and test cases

An analysis of the complex multivariate distribution applied to many crystallographic experiments, including heavy-atom phasing by anomalous scattering and model refinement in the presence of multiple data sets and models, has been performed (Pannu *et al.*, 2003). The distribution discussed in this paper can be applied to account explicitly for the correlations and errors in a SAD experiment when applied to both SAD phasing and model refinement. The multivariate conditional probability distribution for the refinement of the two observed structure-factor amplitudes, $|F^+|$ and $|F^-|$, given the Friedel structure factors calculated from the model, $F_c^+ = |F_c^+| \exp(i\alpha_c^+)$ and $(F_c^-)^* = |F_c^-| \exp(-i\alpha_c^-)$, is shown below,

$$
P_{SAD} = P(|F^+|, |F^-|; |F_c^+|, \alpha_c^+, |F_c^-|, \alpha_c^-)
$$

$$
= \frac{2|F^+||F^-| \det(\Sigma_2)}{\pi \det(\Sigma_4)}
$$

$$
\times \exp[-a_{11}|F^+|^2 - a_{22}|F^-|^2 - (a_{33} - c_{33})|F_c^+|^2]
$$

$$
\times \exp\{-(a_{44} - c_{44})|F_c^-|^2 - 2|F^+||F_c^-|
$$

$$
\times [(a_{34} - c_{34})\cos(\alpha_3 - \alpha_4) - (b_{34} - d_{34})\sin(\alpha_3 - \alpha_4)]\}
$$

$$
\times \int_0^{2\pi} \big(\exp\{-2|F^-||F_c^+|[a_{23}\cos(\alpha^- - \alpha_c^+)
$$

$$
- b_{23}\sin(\alpha^- - \alpha_c^+)]\} \exp\{-2|F^-||F_c^-|
$$

$$
\times [a_{24}\cos(\alpha^- - \alpha_c^-) - b_{24}\sin(\alpha^- - \alpha_c^-)]\}
$$

$$
\times I_0[\xi(|F^+|, |F^-|, |F_c^+|, \alpha_c^+, |F_c^-|, \alpha_c^-, \Sigma_4^{-1})]^{1/2} \big) \, d\alpha^-,
$$

$$(1)$$

where

$$
\xi(|F^+|, |F^-|, \alpha^-, |F_c^+|, \alpha_c^+, |F_c^-|, \alpha_c^-, \Sigma_4^{-1})
$$

$$
= 4|F^+|^2\{[a_{12}|F^-|\cos(\alpha^-) + b_{12}|F^-|\sin(\alpha^-)
$$

$$
+ a_{13}|F_c^+|\cos(\alpha_c^+) + b_{13}|F_c^+|\sin(\alpha_c^+)
$$

$$
+ a_{14}|F_c^-|\cos(\alpha_c^-) + b_{14}|F_c^-|\sin(\alpha_c^-)]^2
$$

$$
+ [a_{12}|F^-|\sin(\alpha^-) - b_{12}|F^-|\cos(\alpha^-)
$$

$$
+ a_{13}|F_c^+|\sin(\alpha_c^+) - b_{13}|F_c^+|\cos(\alpha_c^+)
$$

$$
+ a_{14}|F_c^-|\sin(\alpha_c^-) - b_{14}|F_c^-|\cos(\alpha_c^-)]^2\}. \quad (2)
$$

In the above equations, Σ_4 is the (Hermitian) covariance matrix of the complex Gaussian distribution $P(F^+, F^-, F_c^+, F_c^-)$, with the elements of its inverse denoted $z_{jk} = a_{jk} + ib_{jk}$. Σ_2 is the covariance matrix of the bivariate Gaussian distribution $P(F_c^+, F_c^-)$, with real and imaginary components of its inverse denoted c_{ij} and d_{ij}. The covariance matrices Σ_4 and Σ_2 were calculated using the expressions derived by Pannu *et al.* (2003) and account for experimental errors and the correlation between structure factors.

In Appendix *A*, the probability distribution of two observations given N models is derived. The SAD likelihood function shown above is a special case of this general distribution when there are only two models (*i.e.* $N = 2$). A likelihood function conditional on N models may be applied to multiple models output from an NMR experiment or from simulated-annealing optimization techniques (Rice & Brünger, 1994) or when refining N related models obtained from conditional dynamics (Scheres & Gros, 2001).

The 'SAD' likelihood function discussed below is the sum over all reflections of the minus natural logarithm of the derived probability distribution (1) to obtain a function suitable for minimization. To ensure that the matrix remains positive definite, the inverse of the covariance matrix was calculated from the eigenvectors using only positive eigenvalues with *LAPACK* routines (Anderson *et al.*, 1999).

The SAD function derived above was implemented in the program *REFMAC*5 (version 5.1.24; Murshudov *et al.*, 1997) and compared with the 'Rice' likelihood function lacking prior phase information (Bricogne & Irwin, 1996; Murshudov *et al.*, 1997; Pannu & Read, 1996), denoted below as Rice, and the likelihood function encoding prior phase information with Hendrickson–Lattman coefficients (Pannu *et al.*, 1998), denoted below as MLHL.

For the two test cases described below, the automated model-building program *ARP/wARP* (version 6.0; Perrakis *et al.*, 1999) employing the modified *REFMAC*5 program from the *CCP*4 suite (version 4.2.2; Collaborative Computational Project, Number 1994) for iterative model refinement was used. The sequence information for the protein was not supplied to *ARP/wARP*. Furthermore, the default parameters were used in the running of the program, unless otherwise stated below. In particular, the same low-resolution cutoff, obtained by viewing a Wilson plot as suggested by *ARP/wARP*, was used in both test cases for all target functions. We have also re-run the test cases using no low-resolution cutoff. While using the whole resolution range, there were no significant changes over the results presented below for the

SAD function, while the MLHL function produced significantly poorer results in both test cases and the Rice function produced poorer results in the thioesterase test case.

For both test cases, the anomalous substructure was determined, refined and phased automatically using the *CRANK* suite (Ness *et al.*, 2004). *CRANK* used the programs *CRUNCH*2 (de Graaff *et al.*, 2001) for substructure detection, *BP*3 for substructure refinement and phasing (Pannu & Read, 2004) and *DM* for density modification (Cowtan, 1994). The Hendrickson–Lattmann coefficients required for the MLHL function were obtained from *BP*3. Furthermore, the refined anomalous substructure from *BP*3 was input into *ARP/wARP* and *REFMAC*5 for all target functions in order to allow further refinement. For all likelihood functions and test cases, 250 cycles of automated model building with iterative structure refinement were performed and the results of each likelihood function are compared with the final refined structure using the program *SFTOOLS* (Bart Hazes, unpublished work).

2.1. Subtilisin test-data set

The first test case used was the protein subtilisin, which contains an anomalous signal from three calcium ions. The data were collected using synchrotron radiation at a wavelength of 1.54 Å. More information on this data set can be obtained from Dauter *et al.* (2002). The resolution range 1.77–8.2 Å was used for all likelihood functions and the starting map had a relatively high phase error of about 58°. The quality of the starting map and the performance of the three likelihood functions in the automated model-building test are shown in Table 1.
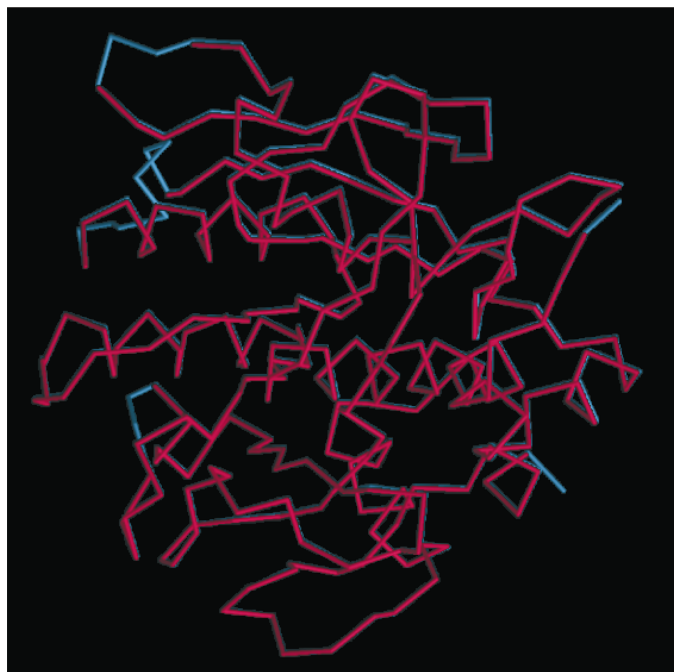

Figure 1
Subtilisin models constructed by *ARP/wARP* using the SAD function (shown in red) superimposed on the final refined structure (shown in blue).

Table 1
Statistics for the automated model building of subtilisin.

	Starting map	Rice	MLHL	SAD
Map correlation	0.5403	0.5581	0.7159	0.9435
Mean cos(phase error)	0.43	0.33	0.54	0.87
Mean phase error	57.78	65.34	48.98	19.81
No. residues built	N/A	3	40	256

The results show that *ARP/wARP* in combination with the SAD function in *REFMAC*5 was able to build the vast majority of the model (256 of 275 residues), while the other refinement targets failed. We were unable to improve the performance of the other target functions by changing any option in *ARP/wARP* or *REFMAC*5. The large difference in phase error and map correlation between the likelihood functions highlights the success of the SAD function. Fig. 1 shows the models built by *ARP/wARP* using the SAD refinement target (shown in red) superimposed on the final refined model (shown in blue).

Fig. 2 shows the change in the phase error as a function of the automated model-building cycle for all three likelihood functions. The Rice function was unable to lower the phase error. The MLHL target shows a similar phase-error improvement as the SAD function in the first cycles, but does not build or improve the model in subsequent cycles. In contrast, the SAD function continues to improve the phases allowing *ARP/wARP/REFMAC*5 to build the model to near-completion.

2.2. Thioesterase test-data set

The second test case used is a selenomethionine thioesterase peak-wavelength data set collected at beamline X9B at Brookhaven National Laboratory with anomalous signal from the eight Se atoms (Li *et al.*, 2000; Dauter *et al.*, 2002). The resolution range 2.52–8.3 Å was used for all likelihood functions in the procedure and the starting map input into the automated model-building procedure had a relatively low phase error of about 42°. The results of this automated model-
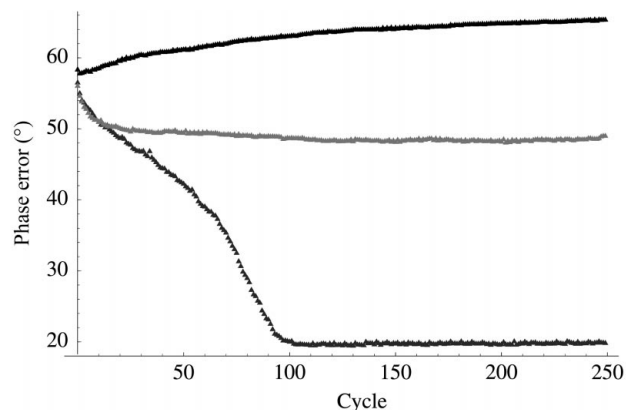

Figure 2
The dependence of phase error *versus* automated model-building cycle for all three likelihood functions in the subtilisin test case. The black curve shows the change in the phase error for the Rice function, the light grey curve is for MLHL and the dark grey curve is for the SAD function.

building test case for the three likelihood functions are shown in Table 2.

Table 2 shows that both likelihood functions incorporating prior phase information performed equally well and to the same phase error, while the likelihood function lacking prior phase information (*i.e.* the Rice function) performed significantly more poorly.

3. Discussion

The derived and implemented multivariate likelihood function directly incorporates diffraction data collected from a SAD experiment and models the correlations and errors that occur in the experiment and refinement process. As a result, the SAD likelihood function can refine and improve model and heavy-atom parameters together, allowing direct and dynamic incorporation of the experimental phase information. The simultaneous refinement of the available parameters combined with a multivariate analysis in the SAD likelihood function appeared to result in a synergic effect that enabled *ARP/wARP* and *REFMAC*5 to build the subtilisin molecule successfully when current likelihood targets failed. From the thioesterase test case, it appears that if the starting experimental phase information is of sufficient quality, the existing likelihood function incorporating prior phase information can be used to construct automatically a model of similar quality to the SAD function.

The above results are promising, but further test cases will be performed to determine whether the trend continues. In particular, test cases with diffraction data at lower resolution will be performed to determine whether the additional information provided by the direct and dynamic incorporation of experimental phase information will push the resolution limits needed for automated building techniques.

In the future, likelihood-based gradient difference maps (*e.g.* de La Fortelle & Bricogne, 1997) will be considered in order to identify any previously undetected anomalous sites. In addition, a multivariate likelihood function will be implemented that incorporates the experimental diffraction data from any variety of phasing experiments [*i.e.* S/MIR(AS) and/or MAD].

The distribution derived above can also be used for the refinement of structure and structure–ligand complexes that directly model the correlation between the observations and the models. Combining the diffraction information from a native structure and a structure in a complex may help emphasize the differences between them, which is usually of major interest to structural biologists, and considering all available information directly may lead to more efficient structure determinations.

APPENDIX A
A1. Derivation of the required distribution

The conditional probability distribution of two observed structure-factor amplitudes, given M model structure factors,

Table 2
Statistics for the automated model building of thioesterase.

	Starting map	Rice	MLHL	SAD
Map correlation	0.7672	0.4878	0.8713	0.8703
Mean cos(phase error)	0.63	0.29	0.75	0.75
Mean phase error	41.81	67.99	31.73	31.68
No. residues built	N/A	22	539	544

will be derived. The SAD likelihood function is the special case when $M = 2$.

The starting point for the derivation will be the multivariate complex Gaussian probability distribution of structure factors (Pannu *et al.*, 2003). Below, N structure factors will be considered, $F_1, F_2, F_3, \ldots, F_N$, where F_1 and F_2 represent the 'observed' structure factors and $N = M + 2$. The amplitude of a structure factor F_i will be denoted $|F_i|$ and its phase α_i.

$$P(F_1, F_2, \ldots, F_N) = \frac{1}{\pi^N \det(\Sigma_N)} \exp\left(-\sum_{i=1}^{N}\sum_{j=1}^{N} F_i^* z_{ij} F_j\right). \quad (3)$$

In the above expression, Σ_N is the Hermitian covariance matrix of this N-dimensional probability distribution and z_{ij} denotes the ijth element of the inverse matrix of Σ_N. The equation can be rewritten by separately summing over the diagonal and off-diagonal terms,

$$P(F_1, F_2, \ldots, F_N) = \frac{1}{\pi^N \det(\Sigma_N)} \quad (4)$$
$$\times \exp\left\{-\sum_{i=1}^{N}\left[|F_i|^2 z_{ii} + \sum_{j=i+1}^{N} 2\Re\left(F_i^* F_j z_{ij}\right)\right]\right\}.$$

After transforming to polar coordinates and simplifying, we obtain

$$P(|F_1|, \alpha_1, |F_2|, \alpha_2, \ldots, |F_N|, \alpha_N) \quad (5)$$
$$= \frac{\prod_{i=1}^{N} |F_i|}{\pi^N \det(\Sigma_N)} \exp\left[-\sum_{i=1}^{N}\left(|F_i|^2 a_{ii}\right.\right.$$
$$\left.\left. + \sum_{j=i+1}^{N} \{2|F_i||F_j|[a_{ij}\cos(\alpha_j - \alpha_i) - b_{ij}\sin(\alpha_j - \alpha_i)]\}\right)\right].$$

In the above equation, a_{ij} and b_{ij} represent the real and imaginary components of the inverse covariance matrix. The unknown phase angles α_1 and α_2 are now integrated out.

research papers

$$P(|F_1|, |F_2|, |F_3|, \alpha_3, \ldots, |F_N|, \alpha_N)$$

$$= \frac{\prod_{i=1}^{N} |F_i|}{\pi^N \det(\Sigma_N)} \exp\left[- |F_1|^2 a_{11} - |F_2|^2 a_{22} - \sum_{i=3}^{N}\left(|F_i|^2 a_{ii} \right.\right.$$

$$\left.\left. + \sum_{j=i+1}^{N} \{2|F_i||F_j|[a_{ij}\cos(\alpha_j - \alpha_i) - b_{ij}\sin(\alpha_j - \alpha_i)]\}\right)\right]$$

$$\times \int_{0}^{2\pi} \exp\left(- \sum_{i=2}^{N} \{2|F_2||F_i| \right.$$

$$\times [a_{2i}\cos(\alpha_i - \alpha_2) - b_{2i}\sin(\alpha_i - \alpha_2)]\}\Big)$$

$$\times \int_{0}^{2\pi} \exp\left(- \sum_{i=2}^{N} \{2|F_1||F_i| \right.$$

$$\times [a_{1i}\cos(\alpha_i - \alpha_1) - b_{1i}\sin(\alpha_i - \alpha_1)]\}\Big) \, d\alpha_1 \, d\alpha_2. \quad (6)$$

The inner integral can be solved analytically,

$$\int_{0}^{2\pi} \exp\left(- \sum_{i=2}^{N} \{2|F_1||F_i| \right.$$

$$\times [a_{1i}\cos(\alpha_i - \alpha_1) - b_{1i}\sin(\alpha_i - \alpha_1)]\}\Big) \, d\alpha_1$$

$$= \int_{0}^{2\pi} \exp\left\{ - 2\cos\alpha_1 |F_1| \sum_{i=2}^{N} [|F_i|(a_{1i}\cos\alpha_i - b_{1i}\sin\alpha_i)] \right.$$

$$\left. - 2\sin\alpha_1 |F_1| \sum_{i=2}^{N} [|F_i|(a_{1i}\sin\alpha_i + b_{1i}\cos\alpha_i)] \right\} \, d\alpha_1$$

$$= 2\pi I_0\left(4|F_1|^2 \left\{ \left[\sum_{i=2}^{N} |F_i|(a_{1i}\cos\alpha_i - b_{1i}\sin\alpha_i) \right]^2 \right.\right.$$

$$\left.\left. + \left[\sum_{i=2}^{N} |F_i|(a_{1i}\sin\alpha_i + b_{1i}\cos\alpha_i) \right]^2 \right\} \right)^{1/2}. \quad (7)$$

$I_0(x)$ is the modified Bessel function of zeroth order. The marginal distribution can now be written as a function involving only one integral,

$$P(|F_1|, |F_2|, |F_3|, \alpha_3, \ldots, |F_N|, \alpha_N) = 2\frac{\prod_{i=1}^{N} |F_i|}{\pi^{N-1} \det(\Sigma_N)}$$

$$\times \exp\left[- |F_1|^2 a_{11} - |F_2|^2 a_{22} - \sum_{i=3}^{N}\left(|F_i|^2 a_{ii} \right.\right.$$

$$\left.\left. + \sum_{j=i+1}^{N} \left\{2|F_i||F_j|[a_{ij}\cos(\alpha_j - \alpha_i) - b_{ij}\sin(\alpha_j - \alpha_i)]\right\}\right)\right]$$

$$\times I_0(2|F_1|\xi) \int_{0}^{2\pi} \exp\left(- \sum_{i=2}^{N} \{2|F_2||F_i| \right.$$

$$\times [a_{2i}\cos(\alpha_i - \alpha_2) - b_{2i}\sin(\alpha_i - \alpha_2)]\}\Big) \, d\alpha_2, \quad (8)$$

where

$$\xi = \left(\sum_{i=2}^{N}\left\{ |F_i|^2(a_{1i}^2 + b_{1i}^2) + \sum_{j=i+1}^{N} 2|F_i||F_j|[(a_{1i}a_{1j} + b_{1i}b_{1j}) \right.\right.$$

$$\left.\left. \times \cos(\alpha_j - \alpha_i) + (a_{1j}b_{1i} - a_{1i}b_{1j})\sin(\alpha_j - \alpha_i)] \right\} \right)^{1/2}. \quad (9)$$

Using the definition of conditional probability, the required probability distribution can be obtained as

$$P(|F_1|, |F_2|; |F_3|, \alpha_3, \ldots, |F_N|, \alpha_N)$$

$$= \frac{P(|F_1|, |F_2|, |F_3|, \alpha_3, \ldots, |F_N|, \alpha_N)}{P(|F_3|, \alpha_3, \ldots, |F_N|, \alpha_N)}. \quad (10)$$

$P(|F_1|, |F_2|, |F_3|, \alpha_3, \ldots, |F_N|, \alpha_N)$ is given in (8) and $P(|F_3|, \alpha_3, \ldots, |F_N|, \alpha_N)$ can be obtained from (5), denoting the corresponding covariance matrix by Σ_{N-2} and the ijth element of its inverse by $c_{ij} + id_{ij}$. Thus, the required distribution can be expressed as

$$P(|F_1|, |F_2|; |F_3|, \alpha_3, \ldots, |F_N|, \alpha_N) = \frac{2|F_1||F_2| \det(\Sigma_{N-2})}{\pi \det(\Sigma_N)}$$

$$\times \exp\left[- |F_1|^2 a_{11} - |F_2|^2 a_{22} - \sum_{i=3}^{N}\left(|F_i|^2(a_{ii} - c_{ii}) \right.\right.$$

$$\left. + \sum_{j=i+1}^{N} \{2|F_i||F_j|[(a_{ij} - c_{ij})\cos(\alpha_j - \alpha_i) \right.$$

$$\left.\left. - (b_{ij} - d_{ij})\sin(\alpha_j - \alpha_i)]\}\right)\right] \int_{0}^{2\pi} \exp\left(- \sum_{i=2}^{N} \{2|F_2||F_i| \right.$$

$$\times [a_{2i}\cos(\alpha_i - \alpha_2) - b_{2i}\sin(\alpha_i - \alpha_2)]\}\Big) I_0(2|F_1|\xi) \, d\alpha_2. \quad (11)$$

References

Anderson, E., Bai, Z., Bischof, C., Blackford, S., Demmel, J., Dongarra, J., Du Croz, J., Greenbaum, A., Hammarling, S., McKenney, A. & Sorensen, D. (1999). *LAPACK Users' Guide*, 3rd ed. Philadelphia: Society for Industrial and Applied Mathematics.

Bricogne, G. (2000). *Advanced Special Functions and Applications: Proceedings of the Melfi School on Advanced Topics in Mathematics and Physics*, edited by D. Cocolicchio, G. Dattoli & H. M. Srivastava, pp. 315–323. Rome: Aracne Editrice.

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

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Siliqi, G. (2002). *Acta Cryst.* D**58**, 928–935.

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

Cowtan, K. (1994). *Jnt CCP4/ESF–EAMCB Newsl. Protein Crystallogr.* **31**, 34–38.

Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.

Giacovazzo, C. & Siliqi, G. (2001a). *Acta Cryst.* A**57**, 40–46.

Giacovazzo, C. & Siliqi, G. (2001b). *Acta Cryst.* A**57**, 414–419.

Giacovazzo, C. & Siliqi, G. (2004). *Acta Cryst.* D**60**, 73–82.

Graaff, R. A. G. de, Hilge, M., van der Plas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* D**57**, 1857–1862.

Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* B**26**, 136–143.

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

Li, J., Derewenda, U., Dauter, Z., Smith, S. & Derewenda, Z. S. (2000). *Nature Struct. Biol.* **7**, 555–559.

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

Ness, S. R., de Graaff, R. A. G., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*, **12**, 1753–1761.

Pannu, N. S., McCoy, A. J. & Read, R. J. (2003). *Acta Cryst.* D**59**, 1801–1808.

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

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

Pannu, N. S. & Read, R. J. (2004). *Acta Cryst.* D**60**, 22–27.

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

Rice, L. M. & Brünger, A. T. (1994). *Proteins*, **19**, 277–290.

Scheres, S. H. W. & Gros, P. (2001). *Acta Cryst.* D**57**, 1820–1828.

Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 1174–1182.