Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The application of multivariate statistical techniques improves single-wavelength anomalous diffraction phasing

Navraj S. Pannu[a,b,c]* and Randy J. Read[b]

[a]Biophysical Structural Chemistry, Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands, [b]Department of Haematology, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Hills Road, Cambridge CB2 2XY, England, and [c]Trinity College, Cambridge CB2 1TQ, England

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

Received 16 July 2003
Accepted 22 September 2003

Recently, there has been a resurgence in phasing using the single-wavelength anomalous diffraction (SAD) experiment; data from a single wavelength in combination with techniques such as density modification have been used to solve macromolecular structures, even with a very small anomalous signal. Here, a formulation for SAD phasing and refinement employing multivariate statistical techniques is presented. The equation developed accounts explicitly for the correlations among the observed and calculated Friedel mates in a SAD experiment. The correlated SAD equation has been implemented and test cases performed on real diffraction data have revealed better results compared with currently used programs in terms of correlation with the final map and obtaining more reliable phase probability statistics.

1. Introduction

The power of the anomalous signal collected at a single wavelength to solve macromolecular structures was realised in the 1980s when Hendrickson & Teeter (1981) solved the structure of crambin using SAD data. Furthermore, Wang (1985) had shown using simulated data that the anomalous signal from two S atoms was sufficient to solve a structure of a small protein. Recently, data from a single wavelength have been used to solve structures, even with a very weak anomalous signal (Dauter *et al.*, 1999, 2002; Brodersen *et al.*, 2000; Rice *et al.*, 2000; Weiss *et al.*, 2001; Dauter & Adamiak, 2001; de Graaff *et al.*, 2001; Gordon *et al.*, 2001; Debreczeni, Bunkoczi, Girmann *et al.*, 2003; Debreczeni, Bunkoczi, Ma *et al.*, 2003). Furthermore, SAD can be preferable to MAD in a case where a crystal exhibits radiation decay during the course of a MAD experiment (Rice *et al.*, 2000).

Currently, to refine the anomalous substructure and phase a SAD data set, conventional techniques employ a Gaussian distribution either on the Bijvoet differences (North, 1965; Matthews, 1966) or the Friedel pairs (Blow & Rossmann, 1961). For example, the phasing package *SHARP* (de La Fortelle & Bricogne, 1997) employs a phase and amplitude integrated Gaussian distribution on Bijvoet differences for SAD data,

$$P_{SHARP} = \int_0^{\infty}\int_0^{2\pi} \exp\left[-\frac{(\Delta_{\rm obs} - \Delta_{\rm calc})^2}{2V}\right] {\rm d}\alpha\, {\rm d}|F|, \qquad (1)$$

where $\Delta_{\rm obs}$ is the Bijvoet difference of the observed Friedel pairs $|F^+|$ and $|F^-|$, V is the variance of the distribution and $\Delta_{\rm calc} = |F + H_c^+| - |F + H_c^-|$ is the Bijvoet difference determined from H_c^+ and H_c^-, the calculated structure factors from the anomalous substructure, and an assumed value of the true amplitude $|F|$ and phase α that is averaged/integrated out.

Since data from a SAD experiment come from the same crystal and share the same model of anomalous scatterers, explicitly accounting for the correlation of substructure model error may improve results further. Recently, the work on joint probability distributions of Hauptman (1982) and Giacovazzo (1983) has been further developed and generalized for substructure detection (Burla *et al.*, 2002) and phasing (Giacovazzo & Siliqi, 2001*a*,*b*) assuming a cumulative (Terwilliger & Eisenberg, 1987) and uncorrelated error term. The above-mentioned distributions, however, neglect some of the important correlations that occur in a SAD experiment.

Fig. 1 depicts a vector diagram of the bimodal phase information obtained in a SAD experiment. In this figure, F_A is defined as the structure factor only considering the f'' atomic scattering factor for the anomalously scattering atoms, $F_A = \sum_j f_j'' \exp(2\pi i \mathbf{h}\cdot\mathbf{x}j)$, and F is the structure factor not including the f'' contribution for the anomalous scatterers. The component F is shared between a reflection and the complex conjugate of its Friedel mate, whereas F_A contributes with opposite sign. The blue circle centred on $-F_A$ of radius $|F^+|$ with thickness corresponding to the measurement error of $|F^+|$ restricts all the possible values of F imposed by $|F^+|$. Similarly, the red circle centred on F_A of radius $|F^-|$ with thickness relating to the measurement error of $|F^-|$ restricts all possible values of F imposed by $|F^-|$. Therefore, the possible values of F consistent with both measurements are at the intersection of the two circles. The figure, however, only shows a SAD experiment considering the effect of measurement error of $|F^+|$ and $|F^-|$ and assumes no errors in the anomalous substructure. Obviously, since the anomalous scatterer(s) come from the same crystal, the anomalous (sub)structure factors and their errors are correlated and modelling these correlations will have an effect on the possible values of F.

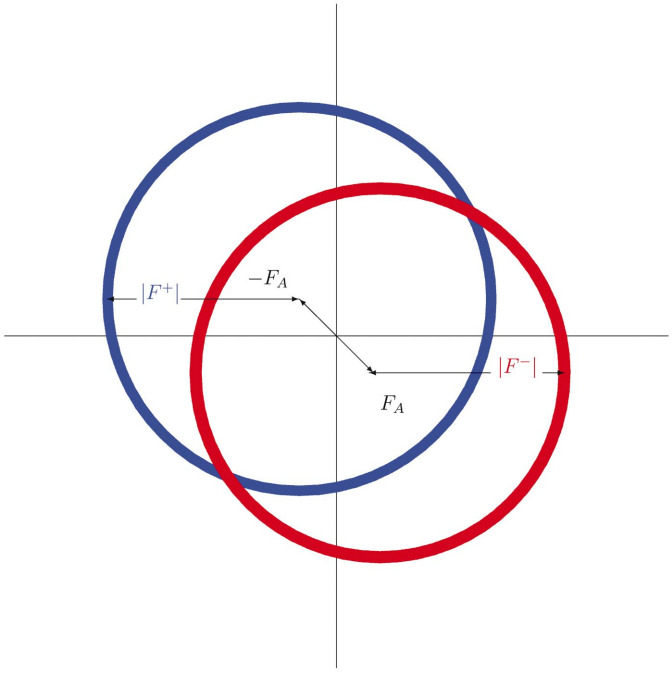

Figure 1
Phase information from a SAD experiment considering only measurement errors. The intersection of the two circles defines possible values for F, the structure factor not including the f_j'' contribution of atoms. $F_A = \sum_j f_j'' \exp(2\pi i \mathbf{h}\cdot\mathbf{x}j)$. The blue circle centred on $-F_A$ of radius $|F^+|$ with thickness corresponding to the measurement error of $|F^+|$ restricts all the possible values of F imposed by $|F^+|$. The red circle centred on F_A of radius $|F^-|$ with thickness relating to the measurement error of $|F^-|$ restricts all possible values of F imposed by $|F^-|$ [*i.e.* $F = F^+ - F_A = (F^-)^* + F_A$].

2. Implementation and test cases

An analysis of the complex multivariate distribution applied to many crystallographic experiments, including heavy-atom phasing by anomalous scattering, has been performed (Pannu *et al.*, 2003). The distribution discussed in this paper can be applied to explicitly account for the correlations in a SAD experiment. The multivariate distribution for the refinement of the two observed structure factors, $|F^+|$ and $|F^-|$, given the Friedel (sub)structure factors calculated from the model, $H_c^+ = |H_c^+|\exp(i\alpha_c^+)$, $(H_c^-)^* = |H_c^-|\exp(-i\alpha_c^-)$ with correlated errors is derived in Appendix *A* and shown below.

$$
\begin{aligned}
P_{\rm SAD} = P(|F^+|, |F^-|; |H_c^+|, \alpha_c^+, |H_c^-|, \alpha_c^-) = {} & \frac{2|F^+||F^-|\det(\Sigma_2)}{\pi\det(\Sigma)} \\
& \times \exp\{-a_{11}|F^+|^2 - a_{22}|F^-|^2 - (a_{33} - c_{33})|H_c^+|^2 \\
& - (a_{44} - c_{44})|H_c^-|^2 - 2|H_c^+||H_c^-|[(a_{34} - c_{34})\cos(\alpha_c^+ - \alpha_c^-) \\
& - (b_{34} - d_{34})\sin(\alpha_c^+ - \alpha_c^-)]\} \\
& \times \int_0^{2\pi} \exp\{-2|F^-||H_c^+|[a_{23}\cos(\alpha^- - \alpha_c^+) - b_{23}\sin(\alpha^- - \alpha_c^+)]\} \\
& \times \exp\{-2|F^-||H_c^-|[a_{24}\cos(\alpha^- - \alpha_c^-) - b_{24}\sin(\alpha^- - \alpha_c^-)]\} \\
& \times I_0[\xi(|F^+|, |F^-|, |H_c^+|, \alpha_c^+, |H_c^-|, \alpha_c^-, \Sigma^{-1})^{1/2}]\,{\rm d}\alpha^-, \qquad (2)
\end{aligned}
$$

where

$$
\begin{aligned}
\xi(|F^+|, |F^-|, \alpha^-, |H_c^+|, \alpha_c^+, |H_c^-|, \alpha_c^-, \Sigma^{-1}) = {} & 4|F^+|^2 \\
& \times \{[a_{12}|F^-|\cos(\alpha^-) + b_{12}|F^-|\sin(\alpha^-) + a_{13}|H_c^+|\cos(\alpha_c^+) \\
& + b_{13}|H_c^+|\sin(\alpha_c^+) + a_{14}|H_c^-|\cos(\alpha_c^-) + b_{14}|H_c^-|\sin(\alpha_c^-)]^2 \\
& + [a_{12}|F^-|\sin(\alpha^-) - b_{12}|F^-|\cos(\alpha^-) + a_{13}|H_c^+|\sin(\alpha_c^+) \\
& - b_{13}|H_c^+|\cos(\alpha_c^+) + a_{14}|H_c^-|\sin(\alpha_c^-) - b_{14}|H_c^-|\cos(\alpha_c^-)]^2\}. \qquad (3)
\end{aligned}
$$

In the above equations, Σ is the (Hermitian) covariance matrix of the complex Gaussian distribution $P(F^+, F^-, H_c^+, H_c^-)$, with the elements of its inverse denoted $z_{jk} = a_{jk} + ib_{jk}$. Σ_2 is the covariance matrix of the bivariate Gaussian distribution $P(H_c^+, H_c^-)$, with the real and imaginary components of its inverse denoted c_{ij} and d_{ij}. The covariance matrices Σ and Σ_2 were calculated using the expressions derived previously (Pannu *et al.*, 2003) and account for experimental errors and the correlation between structure factors. The SAD likelihood function discussed below is the sum over all reflections of the minus natural logarithm of the derived probability distribution (equation 2, or equivalently equation 12 in Appendix *A*) 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 eigenvalues and eigenvectors

Table 1
Statistics for SAD refinement and phasing of lysozyme.

	MLPHARE	*SOLVE*	*SHARP*	SAD
Map correlation	0.433	0.493	0.529	0.546
Reported FOM	0.352	0.353	0.421	0.459
Mean cos(phase error)	0.335	0.413	0.447	0.451
Mean phase error (°)	65.59	59.86	57.11	56.80

using singular value decomposition (*e.g.* Golub & van Loan, 1996; Cowtan & Ten Eyck, 2000) from *LAPACK* routines (Anderson *et al.*, 1999) to remove negative eigenvalues. Automatic differentiation (Griewank *et al.*, 1996) was used to obtain the necessary partial derivatives of the SAD function.

The SAD function derived above was compared with the programs *MLPHARE* (version 4.0; Otwinowski, 1991; Collaborative Computational Project, Number 4, 1994), *SOLVE* (version 2.03; Terwilliger & Berendzen, 1997) and *SHARP* (version 2.0.1; de La Fortelle & Bricogne, 1997) using the two test systems shown below. In all tests, the default or example scripts were used in running each program unless specified otherwise. The tests involved refining atomic parameters (coordinates, occupancies and isotropic B factors) and associated anomalous error parameters (if that option is available in the particular program) together. In the running of *SHARP*, the global and local imperfection parameters on anomalous differences were refined. In the *SOLVE* DNA-oligomer dataset test cases, the minimum atomic B factor was set to zero. In the running of *MLPHARE*, the occupancies were not refined, as holding them constant produced better results. All data sets used were scaled using *TRUNCATE* (French & Wilson, 1978) from *CCP*4 (Collaborative Computational Project, Number 4, 1994) to obtain $|F^+|$, $|F^-|$, the mean $|F|$ and Bijvoet differences. The Friedel pairs and their corresponding standard deviations were used in *SOLVE* and by the SAD function, whereas *MLPHARE* and *SHARP* employed the mean $|F|$ and Bijvoet differences with their corresponding standard deviations. The statistics in the tables shown were all computed with *SFTOOLS* (B. Hazes, unpublished work) from *CCP*4. The statistics for the SAD function were calculated using the standard definitions of the figure of merit (FOM) and 'best phase' (Blow & Crick, 1959; Drenth, 1999) with the probability distribution shown in (2).

2.1. Lysozyme in-house data set

The first test case used a lysozyme data set collected on an in-house source using the anomalous signal from the intrinsic S atoms and solvent chloride ions described further elsewhere (de Graaff *et al.*, 2001). The atomic parameters input to all programs were the 17 Cl and S atoms found by the program *CRUNCH* (de Graaff *et al.*, 2001), with the occupancies set to one and the B factors set to 20 Å^2 in all programs. Results of the test are shown in Table 1 where the phase errors and map correlations are calculated against phases generated from the final refined model coordinates from this data set (de Graaff *et al.*, 2001).

Table 2
Statistics for SAD refinement and phasing of the DNA oligomer: 360° pass.

	MLPHARE	*SOLVE*	*SHARP*	SAD
Map correlation	0.593	0.579	0.693	0.723
Reported FOM	0.563	0.494	0.564	0.640
Mean cos(phase error)	0.493	0.560	0.605	0.642
Mean phase error (°)	54.18	50.02	44.86	41.97

Table 3
Statistics for SAD refinement and phasing of the DNA oligomer: 90° pass.

	MLPHARE	*SOLVE*	*SHARP*	SAD
Map correlation	0.488	0.478	0.630	0.647
Reported FOM	0.403	0.349	0.459	0.563
Mean cos(phase error)	0.416	0.484	0.551	0.565
Mean phase error (°)	59.66	55.16	49.08	48.02

From the table, *SHARP* and the SAD function outperform the other functions. The SAD function gives essentially the same phase error as with *SHARP*, but a better estimate of the figure of merit (FOM). Thus, the SAD function gives a better map correlation with the final map than *SHARP*.

2.2. DNA-oligomer data set

The second test case is from a previously described DNA oligomer (Dauter & Adamiak, 2001). This data set was collected to 1.5 Å resolution at beamline X8C at the Brookhaven National Laboratory Synchrotron at a wavelength of 1.54 Å and had an anomalous signal from the intrinsic P atoms. Images were processed corresponding to 360, 270, 180 and 90° of the total rotation to test the effect of redundancy on the SAD structure-solution process. Tests from the 360 and 90° data sets are considered below. The atomic parameters input to all programs were the ten P atoms obtained by the program *CRUNCH* (de Graaff *et al.*, 2001). For the 360° data set, the coordinates found had only a 0.08 Å average absolute error from the final coordinates, while for the 90° data set the coordinates had a 0.124 Å error. Results of the tests from the 360 and 90° passes are shown in Tables 2 and 3, respectively, where the phase errors and map correlations are calculated against phases generated from the final refined coordinates from a model built at 0.95 Å resolution (Dauter & Adamiak, 2001).

From Tables 2 and 3, the results of these test cases mirror the results obtained from the lysozyme data. *SHARP* and the SAD function perform the best, but the SAD function gives better phase errors and considerably better figure-of-merit estimates, all of which lead to better correlations with the final refined map.

3. Discussion

The above test cases show that the SAD function produced better statistics in terms of map correlation and phase errors compared with the final refined structures. Furthermore, the

SAD function produced more reliable phase probability distribution estimates, as assessed by the agreement of the mean cosine of the phase difference with the final model and the reported figure of merit from the program. However, more test cases will be performed to determine whether this trend continues.

To maintain consistency with the previous reported results on these data sets (*e.g.* Dauter & Adamiak, 2001; Dauter *et al.*, 2002), the test cases were reported using only acentric data. However, centric data can and should be used in SAD phasing, as including reflections, even with low phasing power/figures of merit, is preferable to leaving them unphased (T. C. Terwilliger, personal communication)

The program *SHELXD* (Usón & Sheldrick, 1999; Schneider & Sheldrick, 2002) was also able to find accurate sites for the same (or very similar) data sets as were used here (Dauter *et al.*, 2002). Furthermore, *SHELXE* (Sheldrick, 2002), when provided with the coordinates determined by *CRUNCH*, was able to generate phases with no heavy-atom refinement for the 360° oligomer data set, giving a map correlation of 0.685, a phase error of 46.30 and a mean cosine of the phase error of 0.583 when compared with the final model. However, *SHELXE* gave a figure of merit of 0.485, an underestimation of the mean cosine of the phase error. The map correlation is comparable to that of *SHARP*. It should be noted that the SAD likelihood function was also able to generate phase information with no atomic or error-parameter refinement, giving similar values of the map correlation. However, by not refining parameters, suboptimal figures of merit resulted. In all the test cases described, the refinement of atomic and error parameters was necessary to obtain the best map correlation combined with the most accurate phase probability statistics in the SAD likelihood function.

This equation can not only be used for anomalous structure factor phasing with $|F^+|$ and $|F^-|$, but also for model refinement using a SAD data set, an idea initially proposed by Garib Murshudov (personal communication). Efforts are currently under way to test the SAD function for model refinement as a direct way of incorporating prior experimental phase information into refinement, rather than using information derived from phasing probabilities (*e.g.* Hendrickson–Lattman coefficients) as described previously (Pannu *et al.*, 1998). Similarly, the SAD function applied to refinement would only require a one-dimensional numerical integration.

APPENDIX *A*
Derivation of the required distribution

To apply a maximum-likelihood analysis to a SAD phasing experiment, the probability distribution of the observations ($|F^+|$, $|F^-|$) given the calculated Friedel (sub)structure factors [H_c^+, (H_c^-)*] is required. For generality, F^+ will be denoted by $F_1 = |F_1|\exp(i\alpha_1)$ and (F^-)* (the complex conjugate of F^-) will be labelled by $F_2 = |F_2|\exp(i\alpha_2)$. For the structure factors calculated from the anomalous substructure, H_c^+ will be denoted $F_3 = |F_3|\exp(i\alpha_3)$ and (H_c^-)* (the complex conjugate of H_c^-) will be labelled $F_4 = |F_4|\exp(i\alpha_4)$. This distribution can be obtained from the formula

$$P(|F_1|,|F_2|;\,|F_3|,\alpha_3,|F_4|,\alpha_4) = \frac{\int_0^{2\pi}\int_0^{2\pi} P(|F_1|,\alpha_1,|F_2|,\alpha_2,|F_3|,\alpha_3,|F_4|,\alpha_4)\,\mathrm{d}\alpha_1\,\mathrm{d}\alpha_2}{P(|F_3|,\alpha_3,|F_4|,\alpha_4)}. \quad (4)$$

The above distribution requires a two-dimensional integration. However, one integral can be computed analytically, leaving only a one-dimensional integration to be carried out numerically. To show this, first consider only the numerator in the above expression. The distribution $P(|F_1|, \alpha_1, |F_2|, \alpha_2, |F_3|, \alpha_3, |F_4|, \alpha_4)$ can be obtained from $P(F_1, F_2, F_3, F_4)$, a distribution that will be approximated by a complex multivariate normal of mean zero and covariance Σ (Pannu *et al.*, 2003),

$$\begin{aligned}
&P(F_1,F_2,F_3,F_4)\\
&= \frac{1}{\pi^4\det(\Sigma)}\exp\left[-\begin{pmatrix}F_1^*\\F_2^*\\F_3^*\\F_4^*\end{pmatrix}^T\begin{pmatrix}z_{11}&z_{12}&z_{13}&z_{14}\\z_{12}^*&z_{22}&z_{23}&z_{24}\\z_{13}^*&z_{23}^*&z_{33}&z_{34}\\z_{14}^*&z_{24}^*&z_{34}^*&z_{44}\end{pmatrix}\begin{pmatrix}F_1\\F_2\\F_3\\F_4\end{pmatrix}\right]\\
&= \frac{1}{\pi^4\det(\Sigma)}\exp(-z_{11}F_1F_1^* - z_{22}F_2F_2^* - z_{33}F_3F_3^* - z_{44}F_4F_4^*)\\
&\quad\times\exp[-z_{12}F_1^*F_2 - (z_{12}F_1^*F_2)^* - z_{13}F_1^*F_3 - (z_{13}F_1^*F_3)^*]\\
&\quad\times\exp[-z_{14}F_1^*F_4 - (z_{14}F_1^*F_4)^* - z_{[}23F_2^*F_3 - (z_{23}F_2^*F_3)^*]\\
&\quad\times\exp[-z_24F_2^*F_4 - (z_{24}F_2^*F_4)^* - z_{34}F_3^*F_4 - (z_{34}F_3^*F_4)^*]. \quad (5)
\end{aligned}$$

In the above expression, the jkth component of the inverse of the Hermitian covariance matrix, Σ^{-1}, is denoted $z_{jk} = a_{jk} + ib_{jk}$. To obtain the desired distribution of the numerator in (4) requires integrating, after transforming to polar coordinates, over the unknown phases α_1 and α_2. Manipulation of the above expression gives

$$\begin{aligned}
&P(|F_1|,|F_2|,|F_3|,\alpha_3,|F_4|,\alpha_4)\\
&= \frac{|F_1||F_2||F_3||F_4|}{\pi^4\det(\Sigma)}\exp(-a_{11}|F_1|^2 - a_{22}|F_2|^2 - a_{33}|F_3|^2 - a_{44}|F_4|^2)\\
&\quad\times\exp\{-2|F_3||F_4|[a_{34}\cos(\alpha_3-\alpha_4) - b_{34}\sin(\alpha_3-\alpha_4)]\}\\
&\quad\times\int_0^{2\pi}\Bigg[\int_0^{2\pi}\exp\big(-2|F_1|\{|F_2|[a_12\cos(\alpha_1-\alpha_2) - b_{12}\sin(\alpha_1-\alpha_2)]\}\\
&\quad+|F_3|[a_{13}\cos(\alpha_1-\alpha_3) - b_{13}\sin(\alpha_1-\alpha_3)]\\
&\quad+|F_4|[a_{14}\cos(\alpha_1-\alpha_4) - b_{14}\sin(\alpha_1-\alpha_4)]\}\big)\,\mathrm{d}\alpha_1\Bigg]\\
&\quad\times\exp\{-2|F_2||F_3|[a_{23}\cos(\alpha_2-\alpha_3) - b_{23}\sin(\alpha_2-\alpha_3)]\}\\
&\quad\times\exp\{-2|F_2||F_4|[a_{24}\cos(\alpha_2-\alpha_4) - b_{24}\sin(\alpha_2-\alpha_4)]\}\,\mathrm{d}\alpha_2. \quad (6)
\end{aligned}$$

Looking only at the inner integral (the integration with respect to α_1), expanding the trigonometic functions and simplifying gives

$$\int_0^{2\pi} \exp(-2|F_1|\{\cos(\alpha_1) \times [a_{12}|F_2|\cos(\alpha_2) + b_{12}|F_2|\sin(\alpha_2) + a_{13}|F_3|\cos(\alpha_3) + b_{13}|F_3|\sin(\alpha_3) + a_{14}|F_4|\cos(\alpha_4) + b_{14}|F_4|\sin(\alpha_4)] + \sin(\alpha_1)[a_{12}|F_2|\sin(\alpha_2) - b_{12}|F_2|\cos(\alpha_2) + a_{13}|F_3|\sin(\alpha_3) - b_{13}|F_3|\cos(\alpha_3) + a_{14}|F_4|\sin(\alpha_4) - b_{14}|F_4|\cos(\alpha_4)]\}\, d\alpha_1. \quad (7)$$

The above integral has an analytic solution,

$$\int_0^{2\pi} \exp[a\sin(x) + b\cos(x)]\, dx = 2\pi I_0[(a^2 + b^2)^{1/2}]. \quad (8)$$

Therefore, the required distribution can be expressed in the form

$$P(|F_1|, |F_2|, |F_3|, \alpha_3, |F_4|, \alpha_4) = \frac{2|F_1||F_2||F_3||F_4|}{\pi^3 \det(\Sigma)} \times \exp\{-a_{11}|F_1|^2 - a_{22}|F_2|^2 - a_{33}|F_3|^2 - a_{44}|F_4|^2 - 2|F_3||F_4|[a_{34}\cos(\alpha_3 - \alpha_4)] - b_{34}\sin(\alpha_3 - \alpha_4)]\} \times \int_0^{2\pi} \exp\{-2|F_2||F_3|[a_{23}\cos(\alpha_2 - \alpha_3) - b_{23}\sin(\alpha_2 - \alpha_3)]\} \times \exp\{-2|F_2||F_4|[a_{24}\cos(\alpha_2 - \alpha_4) - b_{24}\sin(\alpha_2 - \alpha_4)]\} \times I_0[\xi(|F_1|, |F_2|, |F_3|, \alpha_3, |F_4|, \alpha_4, \Sigma^{-1})^{1/2}]\, d\alpha_2, \quad (9)$$

where the argument of the Bessel function squared is

$$\xi(|F_1|, |F_2|, \alpha_2, |F_3|, \alpha_3, |F_4|, \alpha_4, \Sigma^{-1}) = 4|F_1|^2 \times \{[a_{12}|F_2|\cos(\alpha_2) + b_{12}|F_2|\sin(\alpha_2) + a_{13}|F_3|\cos(\alpha_3) + b_{13}|F_3|\sin(\alpha_3) + a_{14}|F_4|\cos(\alpha_4) + b_{14}|F_4|\sin(\alpha_4)]^2 + [a_{12}|F_2|\sin(\alpha_2) - b_{12}|F_2|\cos(\alpha_2) + a_{13}|F_3|\sin(\alpha_3) - b_{13}|F_3|\cos(\alpha_3) + a_{14}|F_4|\sin(\alpha_4) - b_{14}|F_4|\cos(\alpha_4)]^2\}. \quad (10)$$

Now, the conditional probability can be formed. The distribution of $P(|F_3|, \alpha_3, |F_4|, \alpha_4)$ is

$$P(|F_3|,\alpha_3, |F_4|, \alpha_4) = |F_3||F_4|\pi^2 \det(\Sigma_2) \exp\{-c_{33}|F_3|^2 - c_{44}|F_4|^2 - 2|F_3||F_4|[c_{34}\cos(\alpha_3 - \alpha_4) - d_{34}\sin(\alpha_3 - \alpha_4)]\} \quad (11)$$

where c_{ij} and d_{ij} are the real and imaginary components (respectively) of the inverse of covariance matrix (Σ_2^{-1}) of the bivariate complex Gaussian distribution $P(F_3, F_4)$.

The desired distribution, shown in (2) and below, is obtained by dividing (9) by (11),

$$P_{\rm SAD} = P(|F_1|, |F_2|; |F_3|, \alpha_3, |F_4|, \alpha_4) = \frac{2|F_1||F_2|\det(\Sigma_2)}{\pi \det(\Sigma)} \times \exp\{-a_{11}|F_1|^2 - a_{22}|F_2|^2 - (a_{33} - c_{33})|F_3|^2 - (a_{44} - c_{44})|F_4|^2 - 2|F_3||F_4|[(a_{34} - c_{34})\cos(\alpha_3 - \alpha_4) - (b_{34} - d_{34})\sin(\alpha_3 - \alpha_4)]\} \times \int_0^{2\pi} \exp\{-2|F_2||F_3|[a_{23}\cos(\alpha_2 - \alpha_3) - b_{23}\sin(\alpha_2 - \alpha_3)]\} \times \exp\{-2|F_2||F_4|[a_{24}\cos(\alpha_2 - \alpha_4) - b_{24}\sin(\alpha_2 - \alpha_4)]\} \times I_0[\xi(|F_1|, |F_2|, \alpha_2, |F_3|, \alpha_3, |F_4|, \alpha_4, \Sigma^{-1})^{1/2}]\, d\alpha_2. \quad (12)$$

We thank G. N. Murshudov, A. J. McCoy, E. J. Dodson and G. M. Sheldrick for useful discussions, J. P. Abrahams, R. A. G. de Graaff, A. G. W. Leslie, J. R. Plaisier and T. C. Terwilliger for critical reading of this work, Z. Dauter and M. Hilge for the diffraction data used in the test cases and R. A. G. de Graaff for running *CRUNCH*. Funding for this work was provided by Leiden University, Trinity College (Cambridge, UK) and the Wellcome Trust.

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: Third Edition.* Philadelphia, USA: Society for Industrial and Applied Mathematics.
Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
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. & Ten Eyck, L. F. (2000). *Acta Cryst.* D**56**, 842–856.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.
Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.
Debreczeni, J. É., Bunkoczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 393–395.
Debreczeni, J. É., Bunkoczi, G., Ma, Q. Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.
Drenth, J (1999). *Principles of Protein X-ray Crystallography*, 2nd ed. New York: Springer–Verlag.
French, S. & Wilson, K. (1978). *Acta Cryst.* A**34**, 517–525.
Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 585–592.
Giacovazzo, C. & Siliqi, G. (2001*a*). *Acta Cryst.* A**57**, 40–46.
Giacovazzo, C. & Siliqi, G. (2001*b*). *Acta Cryst.* A**57**, 700–707.
Golub, G. H. & van Loan, C. F. (1996). *Matrix Computations*, 3rd ed. London: Johns Hopkins University Press.
Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Graaff, R. A. G. de, Hilge, M., van der Plas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* D**57**, 1857–1862.
Griewank, A., Juedes, D., Mitev, H., Utke, J., Vogel, O. & Walther, A. (1996). *ACM TOMS*, **22**, 131–167.
Hauptman, H. (1982). *Acta Cryst.* A**38**, 632–641.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Matthews, B. W. (1966). *Acta Cryst.* **20**, 82–86.
North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.
Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
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.
Rice, L. M, Earnest, T. N. & Brünger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.

Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Terwilliger, T. C. & Eisenberg, D. (1987). *Acta Cryst.* A**43**, 6–13.
Terwilliger, T. C. & Berendzen, J. (1997). *Acta Cryst.* D**53**, 571–579.
Usón, I. & Sheldrick, G. M. (1999). *Curr. Opin. Struct. Biol.* **9**, 643–648.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.