Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

MAD phasing: probabilistic estimate of $|F_{oa}|$

Maria Cristina Burla,[a] Benedetta Carrozzini,[b] Giovanni Luca Cascarano,[b] Carmelo Giacovazzo,[b,c]* Giampiero Polidori[a] and Dritan Siliqi[c,d]

[a]Dipartimento di Scienze della Terra, Piazza Universita', 06100 Perugia, Italy, [b]IRMEC c/o Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, [c]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and [d]Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana, Albania

Correspondence e-mail: c.giacovazzo@area.ba.cnr.it

Received 19 November 2001
Accepted 4 March 2002

The method of the joint probability distribution function is applied in order to estimate the structure-factor moduli of the anomalous scatterer substructure. The two-wavelength case is examined: the prior knowledge of the moduli $|F_1^+|$, $|F_1^-|$, $|F_2^+|$, $|F_2^-|$ is used to predict the value of $|F_{oa}|$ arising from the normal scattering of the anomalous scatterers. The conclusive formula is applied to ideal and to real cases: evidence of the usefulness of the approach is obtained.

1. Notation

N: number of atoms in the unit cell.
a: number of anomalous scatterers in the unit cell.
$na = N - a$: number of non-anomalous scatterers.
ε: statistical Wilson coefficient.
$f_j = f_j^o + \Delta f_j + if_j'' = f_j' + if_j''$: scattering factor of the jth atom. f' is its real and f'' is its imaginary part. The thermal factor is included.
$\Sigma_{Np} = \sum_{j=1}^{N}(f_j'^2 + f_j''^2)$. The summation is calculated at the pth wavelength and is extended to all the atoms in the unit cell.
$\Sigma_o = \sum_{j=1}^{na}(f_j^o)^2$. The summation is extended to all the non-anomalous scatterers in the unit cell.
$\Sigma_{oa} = \sum_{j=1}^{a}(f_j^o)^2$. The summation is extended to all the anomalous scatterers in the unit cell.
$F^+ = |F^+| \exp(i\varphi^+) = F_{\mathbf{h}} = \sum_{j=1}^{N} f_j \exp(2\pi i \mathbf{hr}_j)$.
$E^+ = F^+/(\varepsilon\Sigma_N)^{1/2} = R\exp(i\varphi^+) = A^+ + iB^+$.
$F^- = |F^-| \exp(i\varphi^-) = F_{-\mathbf{h}} = \sum_{j=1}^{N} f_j \exp(-2\pi i \mathbf{hr}_j)$.
$E^- = F^-/(\varepsilon\Sigma_N)^{1/2} = G\exp(i\varphi^-) = A^- + iB^-$.
F_p^+, F_p^-, $E_p^+ = A_p^+ + iB_p^+$, $E_p^- = A_p^- + iB_p^-$ denote the values for the pth wavelength.
$F_{oa} = |F_{oa}| \exp(i\varphi_{oa}) = \sum_{j=1}^{a} f_j^o \exp(2\pi i \mathbf{hr}_j)$.
$E_{oa} = F_{oa}/(\varepsilon\Sigma_{oa})^{1/2} = R_{oa}\exp(i\varphi_{oa}) = A_{oa} + iB_{oa}$.
$\Delta_{\rm ano} = |F^+| - |F^-|$.

2. Introduction

In SAS (single-wavelength anomalous scattering) techniques $\Delta_{\rm ano} = |F^+| - |F^-|$ plays a central role in the recovery of the positions of the anomalous scatterers. Indeed, $\Delta_{\rm ano}$ is involved in the so-called anomalous difference Patterson synthesis (Rossmann, 1961) and, in the form $|F^+|^2 - |F^-|^2$, enters into the so-called 'odd' Patterson function (Okaya *et al.*, 1955). $\Delta_{\rm ano}$ has also been successfully used as an approximate value of $|F''|$ in direct-methods procedures aimed towards recovering the anomalous scatterer substructure (Mukherjee *et al.*, 1989).

Owing to the tunability of synchrotron light, the relevance of MAD techniques to the solution of protein structures has strongly increased in recent years. A crucial step for determining phases *via* MAD techniques is the determination of the anomalous scatterer substructure. In the approach suggested by Karle and Hendrickson (Karle, 1980; Hendrickson, 1985; Pähler *et al.*, 1990), the following quantities are derived *via* a least-squares procedure: the amplitudes for normal scattering of the anomalous scattering substructure ($|F_{oa}|$), the amplitude for normal scattering arising from all atoms ($|F_T|$) and the difference in phase between F_{oa} and F_T ($\alpha = \varphi_T - \varphi_{oa}$). A Patterson function is then calculated from which the anomalous scatterers can be located. The Karle and Hendrickson approach often leads to unrealistically large $|F_{oa}|$ values: these should be identified and rejected before the calculation of the Patterson synthesis (Pähler *et al.*, 1990; Yang *et al.*, 1990).

More recently, a Bayesian approach has been introduced (Terwilliger, 1994) which exploits the prior information on the number and type of anomalous scatterers in the asymmetric unit to derive probabilistic distributions for the expected values of $|F_{oa}|$. In particular, Bayes's rule is used to estimate the relative probability $P(F_{\rm oa}, F_T, \alpha)$ that each possible set of values F_{oa}, F_T, α are correct. The best average of each parameter is then calculated. The method reduces the tendency of the Karle and Hendrickson approach to overestimate $|F_{oa}|$ moduli.

The introduction of selenium into a protein as selenomethionine allowed the introduction of anomalous scatterers directly into the structure, thus facilitating the collection of diffraction data at useful wavelengths. This experimental advance encouraged some second-generation direct-methods programs (Miller *et al.*, 1994; Sheldrick, 1998; Burla *et al.*, 2001) for locating Se atoms in the asymmetric unit (Smith *et al.*, 1998; Howell *et al.*, 2000).

The *Shake-and-Bake* approach (Smith *et al.*, 1998) derives coordinates of the anomalous substructure from a single-wavelength set of data. As three or more sets of diffraction data are usually available from a MAD experiment, the positions obtained from the solutions from each set of data are used to identify and confirm the correct solution. Statistical criteria to normalize and select normalized difference structure factors for use in direct phasing have been suggested (Blessing & Smith, 1999).

The two-wavelength case benefited from the special attention of several authors. Singh & Ramaseshan (1968) described an algebraic technique using Bijvoet sums and differences allowing the calculation of $|F_{oa}|$. Unangst *et al.* (1967) proposed an alternative technique based on Bijvoet ratios: a modification of this method was proposed by Cascarano *et al.* (1982). Klop *et al.* (1989) showed that the ratio and sum-differences techniques are equivalent. They provide two solutions for $|F_{oa}|$: the ambiguity is resolved only if one of them is too large compared with the values allowed by the physical system.

In this paper, we apply the rigorous method of the joint probability distribution to estimate the amplitudes, for normal scattering, of the structure factors of the anomalous scattering substructure by simultaneously involving into the calculations the structure factors F^+, F^- at two wavelengths. The result may be considered the first step towards the application of the method to the n-wavelength case. The mathematical approach we use is able to take into account errors arising from different sources.

3. The joint probability distribution $P(E_{oa}, E_1^+, E_2^+, E_1^-, E_2^-)$

The joint probability distribution $P(E^+, E^-)$ was independently obtained by Hauptman (1982) and by Giacovazzo (1983). From such a distribution, the expected value of $\cos(\varphi^+ + \varphi^-)$ could be estimated and therefore the approximation

$$4|F''|^2 \approx |F^+|^2 + |F^-|^2 - 2|F^+F^-|\langle\cos(\varphi^+ + \varphi^-)\rangle$$

was suggested as a possible estimate of the anomalous scattering substructure (Cascarano & Giacovazzo, 1984), where $F'' = \sum_{j=i}^{a} f_j'' \exp(2\pi\mathbf{h}\mathbf{r}_j)$. Since the information contained in single-wavelength data is rather poor, in this section we derive the more useful distribution $P(E_{oa}, E_1^+, E_2^+, E_1^-, E_2^-)$. We separate the normal scattering of the anomalous scatterer substructure from the other variables. E_i, $i = 1, 2$ denote the normalized structure factors at wavelengths λ_1 and λ_2, respectively. We will also suppose the following.

(i) The positions of all the atoms in the asymmetric unit are our primitive random variables.

(ii) For the pth wavelength

$$A_p^+ = \left[\sum_{j=1}^{na} f_j^o \cos(2\pi\mathbf{h}\mathbf{r}_j) + \sum_{j=1}^{a} f_{jp}' \cos(2\pi\mathbf{h}\mathbf{r}_j) - \sum_{j=1}^{a} f_{jp}'' \sin(2\pi\mathbf{h}\mathbf{r}_j) + |\mu_p^+| \cos\theta_p^+\right] \Big/ \left(\varepsilon\Sigma_{Np}\right)^{1/2},$$

$$B_p^+ = \left[\sum_{j=1}^{na} f_j^o \sin(2\pi\mathbf{h}\mathbf{r}_j) + \sum_{j=1}^{a} f_{jp}' \sin(2\pi\mathbf{h}\mathbf{r}_j) + \sum_{j=1}^{a} f_{jp}'' \cos(2\pi\mathbf{h}\mathbf{r}_j) + |\mu_p^+| \sin\theta_p^+\right] \Big/ \left(\varepsilon\Sigma_{Np}\right)^{1/2},$$

$$A_p^- = \left[\sum_{j=1}^{na} f_j^o \cos(2\pi\mathbf{h}\mathbf{r}_j) + \sum_{j=1}^{a} f_{jp}' \cos(2\pi\mathbf{h}\mathbf{r}_j) + \sum_{j=1}^{a} f_{jp}'' \sin(2\pi\mathbf{h}\mathbf{r}_j) + |\mu_p^-| \cos\theta_p^-\right] \Big/ \left(\varepsilon\Sigma_{Np}\right)^{1/2},$$

$$B_p^- = \left[-\sum_{j=1}^{na} f_j^o \sin(2\pi\mathbf{h}\mathbf{r}_j) - \sum_{j=1}^{a} f_{jp}' \sin(2\pi\mathbf{h}\mathbf{r}_j) + \sum_{j=1}^{a} f_{jp}'' \cos(2\pi\mathbf{h}\mathbf{r}_j) + |\mu_p^-| \sin\theta_p^-\right] \Big/ \left(\varepsilon\Sigma_{Np}\right)^{1/2},$$

$$A_{oa} = \sum_{j=1}^{a} f_j^o \cos(2\pi\mathbf{h}\mathbf{r}_j) \Big/ \left(\varepsilon\Sigma_{oa}\right)^{1/2},$$

$$B_{oa} = \sum_{j=1}^{a} f_j^o \sin(2\pi\mathbf{h}\mathbf{r}_j) \Big/ \left(\varepsilon\Sigma_{oa}\right)^{1/2}.$$

According to the above notation,

$$\mu^+ = |\mu^+| \exp(i\theta^+)$$

and

$$\mu^- = |\mu^-| \exp(i\theta^-)$$

are the measurement errors relative to F^+ and F^-, respectively. The error is assumed to be a complex quantity because its effects influence both the real and the imaginary components of the structure factors. Since $|F_{oa}|$ is not measured (it will be estimated from probabilistic considerations), no error will be associated with it.

(iii) $\langle \mu_1^+ \rangle = \langle \mu_2^+ \rangle = \langle \mu_1^- \rangle = \langle \mu_2^- \rangle = 0$.

(iv) $\langle \mu_1^+ \mu_2^+ \rangle = \langle \mu_1^+ \mu_1^- \rangle = \langle \mu_1^- \mu_2^- \rangle = \langle \mu_2^+ \mu_2^- \rangle = 0$. This implies that errors are uncorrelated.

The characteristic function of $P(A_{oa}, A_1^+, A_2^+, A_1^-, A_2^-, B_{oa}, B_1^+, B_2^+, B_1^-, B_2^-)$ is

$$\begin{aligned}
&C(u_a,u_1^+,u_2^+,u_1^-,u_2^-,v_a,v_1^+,v_2^+,v_1^-,v_2^-)\\
&\quad\approx \exp\big\{-\tfrac{1}{4}[k_{11}(u_a^2+v_a^2)+k_{22}(u_1^{+2}+v_1^{+2})+k_{33}(u_2^{+2}+v_2^{+2})\\
&\qquad+k_{44}(u_1^{-2}+v_1^{-2})+k_{55}(u_2^{-2}+v_2^{-2})+2k_{24}(u_1^+u_1^- - v_1^+v_1^-)\\
&\qquad+2k_{35}(u_2^+u_2^- - v_2^+v_2^-)+2k_{29}(u_1^+v_1^- + v_1^+u_1^-)\\
&\qquad+2k_{3,10}(u_2^+v_2^- + v_2^+u_2^-)+2k_{34}(u_2^+u_1^- - v_2^+v_1^-)\\
&\qquad+2k_{23}(u_1^+u_2^+ + v_1^+v_2^+)+2k_{25}(u_1^+u_2^- - v_1^+v_2^-)\\
&\qquad+2k_{2,10}(u_1^+v_2^- + v_1^+u_2^-)+2k_{28}(u_1^+v_2^+ - v_1^+u_2^+)\\
&\qquad+2k_{12}(u_au_1^+ + v_av_1^+)+2k_{14}(u_au_1^- - v_av_1^-)\\
&\qquad+2k_{17}(u_av_1^+ - v_au_1^+)+2k_{19}(u_av_1^- + v_au_1^-)\\
&\qquad+2k_{13}(u_au_2^+ + v_av_2^+)+2k_{15}(u_au_2^- - v_av_2^-)\\
&\qquad+2k_{18}(u_av_2^+ - v_au_2^+)+2k_{1,10}(u_av_2^- + v_au_2^-)\\
&\qquad+2k_{39}(u_2^+v_1^- + v_2^+u_1^-)+2k_{45}(u_1^-u_2^- + v_1^-v_2^-)\\
&\qquad+2k_{4,10}(u_1^-v_2^- - v_1^-u_2^-)]\big\}, \qquad (1)
\end{aligned}$$

where $u_a, u_1^+, u_2^+, \ldots, v_2^-$ are carrying variables associated with $A_{oa}, A_1^+, A_2^+, \ldots, B_2^-$, respectively.

(1) can also be written as

$$C = \exp[-\tfrac{1}{4}(\overline{\mathbf{U}}\mathbf{K}\mathbf{U})] \qquad (2)$$

where

$$\overline{\mathbf{U}} = (u_a, u_1^+, u_2^+, u_1^-, u_2^-, v_a, v_1^+, v_2^+, v_1^-, v_2^-)$$

and $\mathbf{K}$ is the symmetric square matrix, the elements of which are specified as follows:

$$\begin{aligned}
&k_{11} = k_{66} = 1\\
&k_{22} = k_{77} = e_1^+ = 1+\sigma_1^{+2} \text{ with } \sigma_1^{+2} = \langle|\mu_1^+|^2\rangle \big/ \big(\varepsilon\Sigma_{N1}\big)\\
&k_{33} = k_{88} = e_2^+ = 1+\sigma_2^{+2} \text{ with } \sigma_2^{+2} = \langle|\mu_2^+|^2\rangle \big/ \big(\varepsilon\Sigma_{N2}\big)\\
&k_{44} = k_{99} = e_1^- = 1+\sigma_1^{-2} \text{ with } \sigma_1^{-2} = \langle|\mu_1^-|^2\rangle \big/ \big(\varepsilon\Sigma_{N1}\big)\\
&k_{55} = k_{10,10} = e_2^- = 1+\sigma_2^{-2} \text{ with } \sigma_2^{-2} = \langle|\mu_2^-|^2\rangle \big/ \big(\varepsilon\Sigma_{N2}\big)\\
&k_{12} = k_{14} = k_{67} = -k_{69} = S_9 \big/ \big(\Sigma_{oa}\Sigma_{N1}\big)^{1/2}\\
&k_{13} = k_{15} = k_{68} = -k_{6,10} = S_{11} \big/ \big(\Sigma_{oa}\Sigma_{N2}\big)^{1/2}\\
&k_{16} = k_{27} = k_{38} = k_{49} = k_{5,10} = 0\\
&k_{17} = k_{19} = k_{46} = -k_{26} = S_{10} \big/ \big(\Sigma_{oa}\Sigma_{N1}\big)^{1/2}\\
&k_{18} = k_{1,10} = -k_{36} = k_{56} = S_{12} \big/ \big(\Sigma_{oa}\Sigma_{N2}\big)^{1/2}\\
&k_{23} = k_{45} = k_{9,10} = k_{78} = (\Sigma_o + S_5 + S_6) \big/ \big(\Sigma_{N1}\Sigma_{N2}\big)^{1/2}\\
&k_{24} = -k_{79} = (\Sigma_o + S_1) \big/ \Sigma_{N1}\\
&k_{25} = k_{34} = -k_{7,10} = -k_{89} = (\Sigma_o + S_5 - S_6) \big/ \big(\Sigma_{N1}\Sigma_{N2}\big)^{1/2}\\
&k_{28} = -k_{37} = +k_{4,10} = -k_{59} = (S_7 - S_8) \big/ \big(\Sigma_{N1}\Sigma_{N2}\big)^{1/2}\\
&k_{29} = k_{47} = S_3 \big/ \Sigma_{N1}\\
&k_{2,10} = k_{39} = k_{48} = k_{57} = (S_7 + S_8) \big/ \big(\Sigma_{N1}\Sigma_{N2}\big)^{1/2}\\
&k_{35} = -k_{8,10} = (\Sigma_o + S_2) \big/ \Sigma_{N2}\\
&k_{3,10} = k_{58} = S_4 \big/ \Sigma_{N2},
\end{aligned}$$

where

$$\begin{aligned}
&S_1 = \sum_a (f_{j1}'^2 - f_{j1}''), \quad S_2 = \sum_a (f_{j2}'^2 - f_{j2}''), \quad S_3 = 2\sum_a f_{j1}' f_{j1}'',\\
&S_4 = 2\sum_a f_{j2}' f_{j2}'', \quad S_5 = \sum_a f_{j1}' f_{j2}', \quad S_6 = \sum_a f_{j1}'' f_{j2}'',\\
&S_7 = \sum_a f_{j1}' f_{j2}'', \quad S_8 = \sum_a f_{j1}'' f_{j2}', \quad S_9 = \sum_a f_j^o f_{j1}',\\
&S_{10} = \sum_a f_j^o f_{j1}'', \quad S_{11} = \sum_a f_j^o f_{j2}', \quad S_{12} = \sum_a f_j^o f_{j2}''.
\end{aligned}$$

The joint probability distribution function $P(A_{oa}, A_1^+, A_2^+, A_1^-, A_2^-, B_{oa}, B_1^+, B_2^+, B_1^-, B_2^-)$ (P for short), is obtained by Fourier inversion of (2). We obtain

$$\begin{aligned}
P &\approx \pi^{-5}(\det \mathbf{K})^{-1/2} \exp(-\tfrac{1}{2}\overline{\mathbf{T}}\mathbf{K}^{-1}\mathbf{T})\\
&= \pi^{-5}(\det \mathbf{K})^{-1/2} \exp\{-[\lambda_{11}(A_{oa}^2 + B_{oa}^2) + \lambda_{22}(A_1^{+2} + B_1^{+2})\\
&\quad + \lambda_{33}(A_2^{+2} + B_2^{+2}) + \lambda_{44}(A_1^{-2} + B_1^{-2}) + \lambda_{55}(A_2^{-2} + B_2^{-2})\\
&\quad + 2\lambda_{24}(A_1^+A_1^- - B_1^+B_1^-) + 2\lambda_{35}(A_2^+A_2^- - B_2^+B_2^-)\\
&\quad + 2\lambda_{29}(A_1^+B_1^- + B_1^+A_1^-) + 2\lambda_{3,10}(A_2^+B_2^- + B_2^+A_2^-)\\
&\quad + 2\lambda_{23}(A_1^+A_2^+ + B_1^+B_2^+) + 2\lambda_{25}(A_1^+A_2^- - B_1^+B_2^-)\\
&\quad + 2\lambda_{2,10}(A_1^+B_2^- + B_1^+A_2^-) + 2\lambda_{28}(A_1^+B_2^+ - B_1^+A_2^+)\\
&\quad + 2\lambda_{12}(A_{oa}A_1^+ + B_{oa}B_1^+) + 2\lambda_{14}(A_{oa}A_1^- - B_{oa}B_1^-)\\
&\quad + 2\lambda_{17}(A_{oa}B_1^+ - B_{oa}A_1^+) + 2\lambda_{19}(A_{oa}B_1^- + B_{oa}A_1^-)\\
&\quad + 2\lambda_{13}(A_{oa}A_2^+ + B_{oa}B_2^+) + 2\lambda_{15}(A_{oa}A_2^- - B_{oa}B_2^-)\\
&\quad + 2\lambda_{18}(A_{oa}B_2^+ - B_{oa}A_2^+) + 2\lambda_{1,10}(A_{oa}B_2^- + B_{oa}A_2^-)\\
&\quad + 2\lambda_{34}(A_2^+A_1^- - B_2^+B_1^-) + 2\lambda_{39}(A_2^+B_1^- + B_2^+A_1^-)\\
&\quad + 2\lambda_{45}(A_1^-A_2^- + B_1^-B_2^-) + 2\lambda_{4,10}(A_1^-B_2^- - B_1^-A_2^-)]\}, \qquad (3)
\end{aligned}$$

where λ_{pq} are the elements of the matrix $\mathbf{K}^{-1}$. Owing to the large correlation between the variables of the distribution P,

(det **K**) is expected to be close to zero (ill-conditioned matrix). Values between 10^{-20} and 10^{-40} are usual for cases commonly found in practice. *Vice versa*, (det $\mathbf{K}^{-1}$) is expected to be quite large. The ill-conditioning effect will be large when the anomalous signal is very small, *e.g.* this occurs when $\Delta f_1' \approx \Delta f_2'$ or when $f_1'' \approx f_2''$.

The change of variables

$$A_{oa} = R_{oa}\cos\varphi_a, \quad B_{oa} = R_{oa}\sin\varphi_a,$$
$$A_j^+ = R_j\cos\varphi_j^+, \quad B_j^+ = R_j\sin\varphi_j^+,$$
$$A_j^- = G_j\cos\varphi_j^-, \quad B_j^- = G_j\sin\varphi_j^-$$

for $j = 1, 2$ transforms (3) into

$$\begin{aligned}
&P(R_{oa}, R_1, G_1, R_2, G_2, \varphi_a, \varphi_1^+, \ldots, \varphi_2^-)\\
&\simeq \pi^{-5}R_{oa}R_1R_2G_1G_2(\det\mathbf{K})^{-1/2}\\
&\quad\times\exp\{-[\lambda_{11}R_{\rm oa}^2 + \lambda_{22}R_1^2 + \lambda_{33}R_2^2 + \lambda_{44}G_1^2 + \lambda_{55}G_2^2\\
&\quad+ 2\lambda_{23}R_1R_2\cos(\varphi_1^+ - \varphi_2^+) - 2\lambda_{28}R_1R_2\sin(\varphi_1^+ - \varphi_2^+)\\
&\quad+ 2\lambda_{24}R_1G_1\cos(\varphi_1^+ + \varphi_1^-) + 2\lambda_{29}R_1G_1\sin(\varphi_1^+ + \varphi_1^-)\\
&\quad+ 2\lambda_{25}R_1G_2\cos(\varphi_1^+ + \varphi_2^-) + 2\lambda_{2,10}R_1G_2\sin(\varphi_1^+ + \varphi_2^-)\\
&\quad+ 2\lambda_{34}G_1R_2\cos(\varphi_1^- + \varphi_2^+) + 2\lambda_{39}G_1R_2\sin(\varphi_1^- + \varphi_2^+)\\
&\quad+ 2\lambda_{35}R_2G_2\cos(\varphi_2^+ + \varphi_2^-) + 2\lambda_{3,10}R_2G_2\sin(\varphi_2^+ + \varphi_2^-)\\
&\quad+ 2\lambda_{45}G_1G_2\cos(\varphi_1^- - \varphi_2^-) - 2\lambda_{4,10}G_1G_2\sin(\varphi_1^- - \varphi_2^-)\\
&\quad+ 2\lambda_{12}R_{oa}R_1\cos(\varphi_a - \varphi_1^+) - 2\lambda_{17}R_{oa}R_1\sin(\varphi_a - \varphi_1^+)\\
&\quad+ 2\lambda_{13}R_{oa}R_2\cos(\varphi_a - \varphi_2^+) - 2\lambda_{18}R_{oa}R_2\sin(\varphi_a - \varphi_2^+)\\
&\quad+ 2\lambda_{14}R_{oa}G_1\cos(\varphi_a + \varphi_1^-) - 2\lambda_{19}R_{oa}G_1\sin(\varphi_a + \varphi_1^-)\\
&\quad+ 2\lambda_{15}R_{oa}G_2\cos(\varphi_a + \varphi_2^-) + 2\lambda_{1,10}R_{oa}G_2\sin(\varphi_a + \varphi_2^-)]\}. \qquad (4)
\end{aligned}$$

Distribution (4) is the main result of this paper. It may be noticed that (i) all the λ_{ii}s are positive numbers (to comply with the classical Wilson statistics) and (ii) the moduli of the coefficients

$$\lambda_{28}, \lambda_{29}, \lambda_{2,10}, \lambda_{39}, \lambda_{3,10}, \lambda_{4,10}$$

are found to be several orders of magnitude smaller than the others. Indeed, such coefficients cooperate to define the probability of the expected sign of the invariant sinuses, which are weak phase relationships (see Hauptman, 1982, and Giacovazzo, 1983, for the single-wavelength case). They could also be omitted from (4) without losing valuable information.

4. The conditional probability $P(R_{oa}\,|\,R_1, R_2, G_1, G_2)$

No measured value is available for R_{oa}. We will estimate it by first calculating the conditional probability

$$P(R_{oa}|R_1, R_2, G_1, G_2) \qquad (5)$$

and then the conditional mean value $\langle(R_{oa}|R_1, R_2, G_1, G_2)\rangle$. The derivation of (5) requires first the calculation of

$$\begin{aligned}
&P(R_{oa}, R_1, \ldots, G_2) =\\
&\quad\int_o^{2\pi}\ldots\int_o^{2\pi} P(R_{oa}, R_1, \ldots, G_2, \varphi_a, \ldots, \varphi_2^-)\,{\rm d}\varphi_a\ldots{\rm d}\varphi_2^- \qquad (6)
\end{aligned}$$

and then estimation of the ratio

$$P(R_{oa}, R_1, R_2, G_1, G_2)\Big/\int_o^\infty P(R_{oa}, R_1, R_2, G_1, G_2)\,{\rm d}R_{oa}.$$

The estimate of (6) cannot be performed *via* exact calculations. The results obtained in our previous papers (Giacovazzo & Siliqi, 2001*a*,*b*,*c*) suggest the usefulness of the approximation

$$\varphi_1^+ \approx \varphi_2^+ \approx -\varphi_1^- \approx -\varphi_2^-. \qquad (7)$$

We, therefore, introduce (7) into (4). We then obtain

$$\begin{aligned}
&P(R_{oa}, \varphi_a, \varphi_1^+|R_1, R_2, G_1, G_2) \qquad (8)\\
&\quad\approx LR_{oa}\exp[-\lambda_{11}R_{oa}^2 - 2R_{oa}Q_1\cos(\varphi_1^+ - \varphi_a)\\
&\quad\quad- 2R_{oa}Q_2\sin(\varphi_1^+ - \varphi_a)],
\end{aligned}$$

where L is a suitable normalizing constant,

$$Q_1 = \lambda_{12}R_1 + \lambda_{13}R_2 + \lambda_{14}G_1 + \lambda_{15}G_2$$
$$Q_2 = \lambda_{17}R_1 + \lambda_{18}R_2 - \lambda_{19}G_1 - \lambda_{1,10}G_2.$$

From (8) the conditional distribution

$$\begin{aligned}
P(R_{oa}|R_1, R_2, G_1, G_2) &= 2\lambda_{11}R_{\rm oa}\exp(-\lambda_{11}R_{oa}^2)\\
&\quad\times\exp(-X^2/\lambda_{11})I_0(2R_{oa}X) \qquad (9)
\end{aligned}$$

is obtained, where $X^2 = Q_1^2 + Q_2^2$.

(9) may be rewritten

$$P(R_{oa}|R_1, R_2, G_1, G_2) = LW(R_{oa})M(R_{oa}, \lambda_{11}, X),$$

where $W(R_{oa}) = 2R_{oa}\exp(-R_{oa}^2)$ is the classical Wilson distribution for acentric crystals, $M(R_{oa}, \lambda_{11}, X) = \exp[(1 - \lambda_{11})R_{oa}^2]I_o(2R_{oa}X)$ and $L = \lambda_{11}\exp(-X^2/\lambda_{11})$.

The new form of P suggests that prior knowledge of R_1, R_2, G_1 and G_2 modifies the Wilson distribution through the function M, which is the product of two functions: a rapidly decreasing exponential function ($\lambda_{11} > 1$) and the monotonic increasing function I_o. The form and the location of P will depend of the λ_{11} and X parameters.

The location may be calculated as follows. Since

$$\begin{aligned}
\int_o^\infty x^\mu\exp(-\alpha x^2)I_o(bx) &= \Gamma[(\mu+1)/2][2\alpha^{(\mu+1)/2}]^{-1} \qquad (10)\\
&\quad\times\exp[b^2/(4\alpha)]_1F_1\left(\frac{1-\mu}{2}; 1; \frac{-b^2}{4\alpha}\right),
\end{aligned}$$

where Γ and ${}_1F_1$ are the gamma and the confluent hypergeometric function, respectively, from (10) we obtain

$$\langle R_{oa}|R_1, R_2, G_1, G_2\rangle = 2^{-1}(\pi/\lambda_{11})_1F_1\left(-\frac{1}{2}; 1; -\frac{X^2}{\lambda_{11}}\right).$$

Since ${}_1F_1(-\frac{1}{2}; 1; -z^2)$ is well approximated by the hyperbole $y = [1 + 2z^2/\pi^{1/2}]^{1/2}$ in the full range $(0, \infty)$, the expected value of R_{oa} may be calculated *via* the simpler expression

$$\langle R_{oa}|R_1, R_2, G_1, G_2\rangle = \frac{1}{2}(\pi/\lambda_{11})^{1/2}\left(1 + \frac{4}{\pi}\frac{X^2}{\lambda_{11}}\right)^{1/2}. \qquad (11)$$

(9) is shown in Fig. 1 for selected pairs (λ_{11}, X). The different location of the maxima and the different sharpness of the

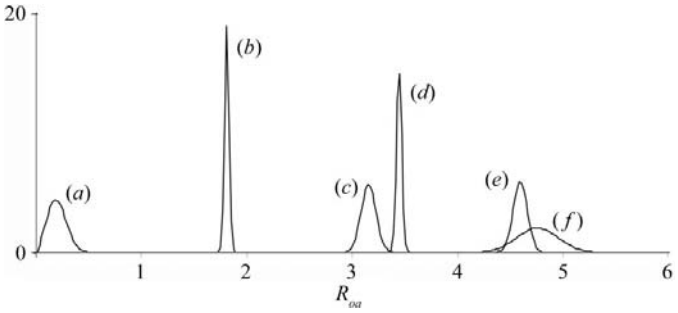

Figure 1
The probability distribution (9) is plotted for some selected values of λ_{11} and X. (*a*) $R_{oa}=0.21, \lambda_{11}=49.8, X=7.5$; (*b*) $R_{oa}=1.80, \lambda_{11}=1136.4, X=2043.6$; 2043.6; (*c*) R_{oa} = 3.15, λ_{11} = 101.4, X = 318.8; (*d*) R_{oa} = 3.44, λ_{11} = 724.5, X = 2488.9; (*e*) R_{oa} = 4.59, λ_{11} = 110.3, X = 505.7; (*f*) R_{oa} = 4.75, λ_{11} = 13.3, X = 63.0

curves suggest that high values of $\langle R_{oa}\rangle$ do not necessarily correlate with sharp distributions.

The standard deviation associated with the estimate (11) may be calculated as follows. Since

$$\langle R_{oa}^2 | R_1, R_2, G_1, G_2\rangle^2 = (\pi/4)\lambda_{11}^{-1} + X^2\lambda_{11}^{-2}$$

and

$$\langle R_{oa}^2 | R_1, R_2, G_1, G_2\rangle = \lambda_{11}^{-1} + X^2\lambda_{11}^{-2}$$

then

$$\sigma_{R_{oa}} = [\langle R_{oa}^2|\ldots\rangle - \langle R_{oa}|\ldots\rangle^2]^{1/2} = \left[\left(1-\frac{\pi}{4}\right)\lambda_{11}^{-1}\right]^{1/2},$$

from which

$$\frac{\langle R_{oa}|\ldots\rangle}{\sigma_{R_{oa}}} = \left[\frac{\pi/4 + X^2/\lambda_{11}}{1-\pi/4}\right]^{1/2}. \qquad (12)$$

Since λ_{11} is always expected to be positive, the expected values of $\langle R_{oa}\rangle$ and of $\sigma_{R_{oa}}$ are always positive.

As well as the reflections with the largest values of $F_{\rm obs}/\sigma_F$, which play a central role in most crystallographic calculations, the reflections with the largest value of the ratio (12) are likely to be the most useful ones. Fig. 2 shows that (12) is a monotonic function of $z = X^2/\lambda_{11}$: accordingly, large z values will characterize accurate estimates. Since large values of X^2 define large values of R_{oa}^2, a large percentage of accurate estimates are expected to be associated with large values of $\langle R_{oa}^2|\ldots\rangle$. This result is quite encouraging; indeed, the efficiency of direct methods (see §6) increases when they are applied to large normalized structure-factor moduli.

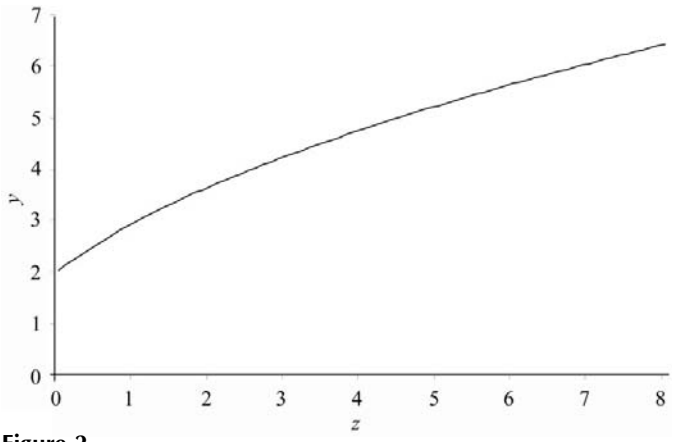

Figure 2
Plot of y = $[(\pi/4 + z)/(1 - \pi/4)]^{1/2}$.

Table 1
Expected $\Delta f'$ and f'' values for each chosen λ value.

Cyanase				1srv			
	Wavelength (Å)	$\Delta f'$	f''		Wavelength (Å)	$\Delta f'$	f''
λ_1	1.0781	−2.112	0.595	λ_1	1.1271	−1.805	0.646
λ_2	0.9795	−9.643	0.499	λ_2	0.9793	−8.582	3.843
λ_3	0.9793	−8.582	3.843	λ_3	0.9791	−7.663	3.841
λ_4	0.9465	−2.618	3.578	λ_4	0.9465	−2.618	3.578

5. The centric case

In the centric case $|F_p^+| = |F_p^-| = |F_p|$. We therefore study the joint probability distribution function

$$P(E_{oa}, A_1, A_2, B_1, B_2) \qquad (13)$$

where

$$E_{oa} = 2\sum_{j=1}^{a/2} f_j^o \cos 2\pi\mathbf{hr}_j \Big/ \Sigma_{\rm oa}^{1/2}$$

$$A_p = \left(2\sum_{j=1}^{na/2} f_j^o \cos 2\pi\mathbf{hr}_j + 2\sum_{j=1}^{a/2} f'_{jp} \cos 2\pi\mathbf{hr}_j + |\mu_p|\cos\theta_p\right) \Big/ (\varepsilon\Sigma_{Np})^{1/2}$$

$$B_p = \left(2\sum_{j=1}^{a/2} f''_{jp} \cos 2\pi\mathbf{hr}_{jp} + |\mu_p|\sin\theta_p\right) \Big/ (\varepsilon\Sigma_{Np})^{1/2}.$$

The characteristic function of (13) is

$$C(u_a, u_1, u_2, v_1, v_2) \approx \exp(-\tfrac{1}{2}\overline{\mathbf{U}}\mathbf{K}\mathbf{U}), \qquad (14)$$

where $\mathbf{K}$ is the symmetric matrix

$$\mathbf{K} = \begin{vmatrix} 1 & S_9/(\Sigma_{oa}\Sigma_{N1})^{1/2} & S_{11}/(\Sigma_{oa}\Sigma_{N1})^{1/2} & S_{10}/(\Sigma_{oa}\Sigma_{N1})^{1/2} & S_{12}/(\Sigma_{oa}\Sigma_{N1})^{1/2} \\ & e'_1 & (\Sigma_o + S_5)/(\Sigma_{N1}\Sigma_{N2})^{1/2} & S_3/(2\Sigma_{N1})^{1/2} & S_7/(\Sigma_{N1}\Sigma_{N2})^{1/2} \\ & & e'_2 & S_8/(\Sigma_{N1}\Sigma_{N2})^{1/2} & S_4/(2\Sigma_{N1})^{1/2} \\ & & & (S_{13} + \sigma_1^2/2) & S_6/(\Sigma_{N1}\Sigma_{N2})^{1/2} \\ & & & & (S_{14} + \sigma_2^2/2) \end{vmatrix},$$

$$e_1' = 1 + \sigma_1^2/2, \quad e_2' = 1 + \sigma_2^2/2,$$
$$S_{13} = \left(\sum_{j=1}^{a} f_{j1}''^2\right) \Big/ \Sigma_{N1}, \quad S_{14} = \left(\sum_{j=1}^{a} f_{j2}''^2\right) \Big/ \Sigma_{N2}$$

and

$$\overline{\mathbf{U}} = (u_a, u_1, u_2, v_1, v_2).$$

The distribution (14) is the Fourier transform of (14). We obtain

$$\begin{aligned} P(E_{oa},A_1, A_2, B_1, B_2) \approx (2\pi)^{-5/2}(\det \mathbf{K})^{-1/2} \\ \times \exp[-\tfrac{1}{2}(\lambda_{11}E_{\rm oa}^2 + \lambda_{22}A_1^2 + \lambda_{33}A_2^2 + \lambda_{44}B_1^2 + \lambda_{55}B_2^2 \\ + 2\lambda_{12}E_{oa}A_1 + 2\lambda_{13}E_{oa}A_2 + 2\lambda_{14}E_{oa}B_1 + 2\lambda_{15}E_{oa}B_2 \\ + 2\lambda_{23}A_1A_2 + 2\lambda_{24}A_1B_1 + 2\lambda_{25}A_1B_2 + 2\lambda_{34}A_2B_1 \\ + 2\lambda_{35}A_2B_2 + 2\lambda_{45}B_1B_2)]. \end{aligned}$$

The change of variables

$$A_1 = R_1 \cos\varphi_1, \quad A_2 = R_2 \cos\varphi_2$$
$$B_1 = R_1 \sin\varphi_1, \quad B_2 = R_2 \sin\varphi_2$$

and the approximation $\varphi_1 \approx \varphi_2$ led to

$$\begin{aligned} P(E_{oa},R_1, R_2, \varphi_1) \approx (2\pi)^{-3/2} R_1R_2(\det \mathbf{K})^{-1/2} \\ \times \exp\{-\tfrac{1}{2}\lambda_{11}R_{oa}^2 + \tfrac{1}{4}(\lambda_{22} + \lambda_{44})R_1^2 + \tfrac{1}{4}(\lambda_{33} + \lambda_{55})R_2^2 \\ + \tfrac{1}{2}(\lambda_{23} + \lambda_{45})R_1R_2 + \tfrac{1}{4}X_1 \cos[2(\varphi_1 - \xi_1)] \\ + E_{oa}X_{oa}\cos(\varphi_1 - \xi_{oa})\}, \end{aligned}$$

where

$$\begin{aligned} X_1^2 &= \{[R_1^2(\lambda_{22} - \lambda_{44}) + R_2^2(\lambda_{33} - \lambda_{55}) + 2R_1R_2(\lambda_{23} - \lambda_{45})]^2 \\ &\quad + 4[R_1^2\lambda_{24} + R_2^2\lambda_{35} + R_1R_2(\lambda_{25} + \lambda_{34})]^2\} \\ &= X_{1B}^2 + X_{1T}^2 \\ \tan 2\xi_1 &= X_{1T}/X_{1B} \\ X_{\rm oa}^2 &= (\lambda_{12}R_1 + \lambda_{13}R_2)^2 + (\lambda_{14}R_1 + \lambda_{15}R_2)^2 \\ &= X_{oaB}^2 + X_{oaT}^2 \\ \tan\xi_{\rm oa} &= X_{oaT}/X_{oaB}. \end{aligned}$$

On assuming $2\varphi_1 \approx 0$ and $\varphi_1 \approx 0, \pi$, we obtain the conditional distribution

$$P(E_{oa}|R_1, R_2) \approx L \exp(-\tfrac{1}{2}\lambda_{11}R_{\rm oa}^2)\cosh(R_{oa}X_{oa}), \qquad (15)$$

where L is a suitable normalization constant. By applying

$$\int_o^\infty \exp(-\beta x^2)\cosh(ax)\,{\rm d}x = \frac{1}{2}\left(\frac{\pi}{\beta}\right)^{1/2}\exp\left(\frac{a^2}{4\beta}\right)$$

to normalize (15) and

$$\int_o^\infty x^2\exp(-\beta x^2)\cosh(ax)\,{\rm d}x = \pi^{1/2}\frac{(2\beta + a^2)}{8\beta^2\beta^{1/2}}\exp\left(\frac{a^2}{4\beta}\right)$$

to obtain the expected value of R_{oa}^2, we obtain

$$\langle R_{oa}^2|R_1, R_2\rangle = \lambda_{11}^{-1} + X_{oa}^2\lambda_{11}^{-2}. \qquad (16)$$

(16) is the desired relation in centric space groups.

Considerations similar to those described for the acentric case lead to

$$\langle R_{oa}|R_1, R_2\rangle = \frac{\pi}{4}\lambda_{11}^{-1} + X_{oa}^2\lambda_{11}^{-2}.$$

6. Experimental applications

To check the usefulness of our probabilistic approach we applied our conclusive formula (11) to the following two test structures (selenomethionine MAD data).

(i) 1srv (Walsh *et al.*, 1999), space group $C222_1$, unit-cell parameters $a = 63.47$, $b = 65.96$, $c = 75.03$ Å, 1186 non-H atoms and three methionines in 145 amino acids. Multiwavelength data to 2.27 Å resolution were collected. This crystal structure was originally undertaken to check the feasibility of ultrafast protein crystal structure determination *via* MAD techniques.

(ii) Cyanase (Walsh *et al.*, 2000), space group $P1$, unit-cell parameters $a = 76.34$, $b = 81.03$, $c = 82.30$ Å, $\alpha = 70.3$, $\beta = 72.2$, $\gamma = 66.4^\circ$, 13965 non-H atoms and 40 methionines in 1560 residues. Multiwavelength data to 2.4 Å resolution were collected.

(11) was first applied to calculated data without errors (*i.e.* $|E|$ without errors, $\sigma_{|E|}/|E| = 0.01$). 7303 reflections (to 2.27 Å resolution) for 1srv and 64 022 reflections for cyanase (to 2.4 Å resolution) were computed for the various wavelengths quoted in Table 1. To check the relative efficiency of (11) we calculated for our test structures the values of

$$R_A = \frac{\sum|(R_{oa})_t - \langle R_{oa}\rangle|}{\sum(R_{oa})_t},$$

where $(R_{oa})_t$ is the true value of R_{oa}. The estimates are quite good for some pairs of wavelengths. However, in spite of the fact that we are using data without errors, the R_A values are dispersed over a rather wide range (between 0.09 and 0.41). Their trend can be correlated with the values of the pair (Δ'^2, Δ''^2), where

$$\Delta'^2 = [(\Delta f')_{\lambda i} - (\Delta f')_{\lambda j}]^2 \text{ and } \Delta''^2 = (f''_{\lambda i} - f''_{\lambda j})^2.$$

Pairs of wavelengths showing sufficiently large values of both Δ'^2 and Δ''^2 are characterized by low values of R_A; *vice versa*, cases for which at least one of Δ'^2 and Δ''^2 is very small show large values of R_A. While this result may be reasonable when observed data are used, this is unexpected in the case of Table 2, where calculated (without error) data are used.

Two questions are therefore still open: (i) why do we end with different R_A values for different wavelength pairs and (ii) why is the correlation between R_A and the pair (Δ'^2, Δ''^2) not very high [*i.e.* for cyanase $R_A = 0.14$ when $(\Delta'^2, \Delta''^2) = (35.6, 0.1)$, while $R_A = 0.19$ when $(\Delta'^2, \Delta''^2) = (49.4, 9.5)$]? We cannot conclusively answer the two questions, but we suggest the following limiting factors for our approach: (i) our joint probability distributions neglect terms of high order (a simple Gaussian distribution is proposed) and (ii) the use of the approximation (7) [the anomalous scatterer contribution is responsible for the small phase differences among Friedel pairs, which are assumed to be zero]. In spite of the previous limitations, Table 2 suggests that our approach is able,

provided a favourable pair (Δ'^2, Δ''^2) is used, to estimate the R_{oa} values with quite high accuracy, given the four magnitudes F_i^+, F_i^-, F_j^+ and F_j^-.

To check how sensitive the method is to the assumed content and scattering factors of the anomalous structure, we used the cyanase calculated data and (i) varied the number of anomalous scatterers (up to ±40% of the expected content) and (ii) modified f' and f'' for each wavelength (up to ±40% of their expected values). The resulting R_A values did not vary for more then a few units percent (typically from 0.36 to 0.39, from 0.15 to 0.18).

Table 2
Calculated data.

R_A values for the different pairs of wavelengths (λ_i, λ_j) and for each test structure. The values of the quantities $\Delta'^2 = [(\Delta f')_{\lambda i} - (\Delta f')_{\lambda j}]^2$ and $\Delta''^2 = (f''_{\lambda i} - f''_{\lambda j})^2$ are also given.

		(λ_1, λ_2)	(λ_1, λ_3)	(λ_1, λ_4)	(λ_2, λ_3)	(λ_2, λ_4)	(λ_3, λ_4)
Cyanase	R_A	0.37	0.11	0.41	0.40	0.19	0.14
	(Δ'^2, Δ''^2)	(56.7, 0.0)	(42.9, 10.5)	(0.3, 8.9)	(1.1, 11.2)	(49.4, 9.5)	(35.6, 0.1)
1srv	R_A	0.14	0.09	0.34	0.39	0.35	0.37
	(Δ'^2, Δ''^2)	(45.9, 10.2)	(34.3, 10.2)	(0.7, 8.6)	(0.8, 0.0)	(35.6, 0.1)	(25.5,0.1)

We now check the average efficiency of our formula (11) *versus* the observed data. The results are shown in Table 3. While the R_A values are of moderate size for 1srv, the larger experimental errors in cyanase data lead to larger R_A values. Since the experimental errors (in the statistical sense) increase the dispersion of the anomalous differences and of the dispersive differences, the expected consequence for our probabilistic approach is the overestimation of the large R_{oa} values. For example, in Table 4 we give for the pair (λ_1, λ_2) of cyanase the number of reflections ($N_{\rm ref}$) for which $\langle R_{oa}\rangle >$ TRH and the respective value of R_A. To check whether (11) is overestimating or underestimating the R_{oa} values in the set, we calculated for each set of reflections the residual

$$\mathrm{SR}_A = \frac{\sum(R_{oa})_t - \langle R_{oa}\rangle}{\sum(R)_t}.$$

A negative value of SR_A indicates overestimation and a positive value underestimation. Table 4 clearly suggests that at large TRH values (11) overestimates, while it underestimates at low TRH value. The same trend has been verified for 1srv.

To obtain an insight into the efficiency of our probabilistic approach, we have applied the program *REVISE* (Fan *et al.*, 1993; Collaborative Computational Project, Number 4, 1994) to the observed data of cyanase and 1srv. *REVISE* provides estimates of quantities

$$\mathrm{FM} = \frac{|F'|^2 + |F''|^2}{\Delta f'^2 + f''^2} \qquad (17)$$

via a least-squares procedure, minimizing undesirable fluctuations of the experimental data. F' and F'' are the real and imaginary parts of the anomalous scattering structure factor, respectively. The discrepancy index R_A calculated for the quantity (17) (all reflections included), using all four wavelengths simultaneously, is 0.46 for cyanase and 0.51 for 1srv.

As soon as the $\langle R_{oa}\rangle$ estimates were available, the crystal structure solution of 1srv and cyanase was attempted by traditional direct-methods procedures. About 600 reflections for 1srv and 2000 for cyanase (those with the largest values of $\langle R_{oa}\rangle$ were selected, triplet invariants were calculated and the tangent formula applied in a multisolution random approach. For both the test structures, the traditional figures of merit clearly indicated the good solutions. The results are as follows.

Table 3
Experimental data: the values of R_A for the acentric reflections of the test structures.

		(λ_1, λ_2)	(λ_1, λ_3)	(λ_1, λ_4)	(λ_2, λ_3)	(λ_2, λ_4)	(λ_3, λ_4)
Cyanase	R_A	0.39	0.41	0.46	0.42	0.42	0.44
1srv	R_A	0.39	0.33	0.36	0.38	0.35	0.34

Table 4
Experimental (λ_1, λ_2) data of cyanase.

$N_{\rm ref}(i)$ is the number of reflections for which $\langle R_{oa}\rangle$ is larger than TRH(i) and smaller than TRH(i + 1) (*e.g.* $N_{\rm ref}$ = 658 is the number of reflections for which $\langle R_{oa}\rangle$ is larger than 0.3 and smaller than 0.6); R_A is the corresponding residual value.

TRH	0.00	0.30	0.60	0.90	1.20	1.50	1.80	2.10
$N_{\rm ref}$	27	658	26359	19634	5048	1616	485	236
R_A	2.39	0.79	0.45	0.38	0.32	0.28	0.25	0.27
SR_A	2.32	0.54	0.02	−0.01	−0.02	−0.03	−0.07	−0.12

For 1srv, the crystal structure solution succeeded for each pair of wavelengths. The average number of good solutions among 40 trials was three; each of them correctly locating all the three Se atoms.

For cyanase, the crystal structure solution was tried for the pairs (λ_1, λ_2) and (λ_2, λ_3). Eight good solutions over 40 trials were found for the first pair, locating (on average) 38 of the 40 Se atoms. For the second pair of wavelengths, 36 good solutions over 40 trials were found, locating (on average) 34 of the 40 Se atoms.

7. Conclusions

A probabilistic theory has been described which is able to obtain, *via* two-wavelength diffraction data, reliable estimates of the structure-factor moduli of the anomalous scatterer substructure. The approach has been successfully applied to the experimental data of two test structures. The method has a large power reserve. Indeed, it may be easily generalized to more than two wavelengths, thus becoming able to exploit the full information of a MAD experiment; also, our tests did not use any filtering criterion. These statistical criteria play a central role in excluding unreliable data and are usually based on the normalized differences at the given wavelength (Blessing & Smith, 1999). The nature of our probabilistic approach can profit by filtering criteria which use dispersive differences in addition to anomalous differences.

Our next efforts will be devoted to devise a probabilistic procedure able to exploit the opportunities described in the three points above.

The authors thank one of the referees for useful comments.

References

Blessing, R. H. & Smith, G. D. (1999). *J. Appl. Cryst.* **32**, 664–670.
Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G. L., Giacovazzo, C., Polidori, G. & Spagna, R. (2001). *J. Appl. Cryst.* **34**, 523–526.
Cascarano, G. & Giacovazzo, C. (1984). *Acta Cryst.* A**40**, 305–306.
Cascarano, G., Giacovazzo, C., Peerdeman, A. F. & Kroon, J. (1982). *Acta Cryst.* A**38**, 710–720.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Fan, H.-F., Woolfson, M. M. & Yao, J.-X. (1993). *Proc. R. Soc. London A*, **442**, 13–32.
Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 585–592.
Giacovazzo, C. & Siliqi, D. (2001*a*). *Acta Cryst.* A**57**, 40–46.
Giacovazzo, C. & Siliqi, D. (2001*b*). *Acta Cryst.* A**57**, 414–419.
Giacovazzo, C. & Siliqi, D. (2001*c*). *Acta Cryst.* A**57**, 700–707.
Hauptman, H. A. (1982). *Acta Cryst.* A**38**, 632–641.
Hendrickson, W. A. (1985). *Trans. Am. Crystallogr. Assoc.* **21**, 11–21.
Howell, P. L., Blessing, R. H., Smith, G. D. & Weeks, C. M. (2000). *Acta Cryst.* D**56**, 604–617.
Karle, J. (1980). *Int. J. Quantum Chem. Quantum Biol. Symp.* **7**, 357–367.
Klop, E. A., Krabbendam, H. E. & Kroon, J. (1989). *Acta Cryst.* A**45**, 203–208.
Miller, R., Gallo, S. M., Khala, M. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
Mukherjee, A. K, Helliwell, J. R. & Main, P. (1989). *Acta Cryst.* A**45**, 715–718.
Okaya, Y., Saito, Y. & Pepinsky, R. (1955). *Phys. Rev.* **98**, 1857–1858.
Pähler, A., Smith, J. L. & Hendrickson, W. A. (1990). *Acta Cryst.* A**46**, 537–540.
Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.
Sheldrick, G. (1998). In *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier. Dordrecht: Kluwer.
Singh, A. K. & Ramaseshan, S. (1968). *Acta Cryst.* B**24**, 35–40.
Smith, G. D., Nagar, B., Rini, J. M., Hauptman, H. A. & Blessing, R. H. (1998). *Acta Cryst.* D**54**, 799–804.
Terwilliger, T. C. (1994). *Acta Cryst.* D**50**, 11–16.
Unangst, D., Müller, E., Müller, J. & Keinert, B. (1967). *Acta Cryst.* **23**, 898–901.
Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimiak, A. (1999). *Acta Cryst.* D**55**, 1168–1173.
Walsh, M. A., Otwinowski, Z., Perrakis, A., Anderson, P. M. & Joachimiak, A. (2000). *Structure*, **8**, 505–514.
Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. (1990). *Science*, **249**, 1398–1405.