function. Finally, one obtains

$$\langle F_{PH} \rangle = \tfrac{1}{2}(\pi \Sigma_H)^{1/2} F\left(-\tfrac{1}{2}, 1; -Y\right), \qquad (37)$$

which is the result given in (12) with $Y = 1/4\alpha = F_P^2/\Sigma_H$.

References

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography.* London: Academic Press.

BRICOGNE, G. (1991) *Crystallographic Computing* 5, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 257–297. IUCr/Oxford Univ. Press.

COMARMOND, M. B., GIEGÉ, R., THIERRY, J. C. & MORAS, D. (1986). *Acta Cryst.* B42, 272–280.

CRICK, F. H. C. & MAGDOFF, B. S. (1956) *Acta Cryst.* 9, 901–908.

DUMAS, P. (1986). PhD thesis, Univ. Louis Pasteur, Strasbourg, France.

DUMAS, P. (1991). *Crystallographic Computing* 5, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, p. 479. IUCr/Oxford Univ. Press.

DUMAS, P. (1994). *Acta Cryst.* 50, 537–546.

DUMAS, P., BERGDOLL, M., CAGNON, C. & MASSON, J. M. (1994). *EMBO J.* 13. In the press.

GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Tables of Integrals, Series, and Products.* London: Academic Press.

International Tables for X-ray Crystallography (1974). Vol. IV. Birmingham: Kynoch Press. (Present distributor Kluwer Academic Publishers, Dordrecht.)

KRAUT, J., SIEKER, L. C., HIGH, D. F. & FREER, S. T. (1962). *Proc. Natl Acad. Sci. USA,* 48, 1417–1424.

LUZZATI, V. (1952). *Acta Cryst.* 5, 802–810.

NIKIFOROV, A. & OUVAROV, V. (1976). *Eléments de la Théorie des Fonctions Spéciales,* pp. 206–217. Moscow: Editions MIR.

NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A32, 325–333.

PERUTZ, M. F. (1956). *Acta Cryst.* 9, 867–873.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). *Numerical Recipes: the Art of Scientific Computing,* pp. 522–523. Cambridge Univ. Press.

READ, R. J. (1986). *Acta Cryst.* A42, 140–149.

READ, R. J. (1990). *Acta Cryst.* A46, 900–912.

READ, R. J. (1991). *Crystallographic Computing* 5, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 158–167. IUCr/Oxford Univ. Press.

RONDEAU, J. M., SAMANA, J. P. SAMANA, B., BARTH, P., MORAS, D. & BIELLMANN, J. F. (1987). *J. Mol. Biol.* 195, 945–948.

ROSSMANN, M. G. (1960). *Acta Cryst.* 13, 221–226.

SAMAMA, J. P., DELARUE, M., MOUREY, L., CHOAY, J. & MORAS, D. (1989). *J. Mol. Biol.* 210, 877–879.

SHMUELI, U. & WEISS, G. H. (1987). *Acta Cryst.* A43, 93–98.

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

TICKLE, I. J. (1991). *Isomorphous Replacement and Anomalous Scattering,* compiled by W. WOLF, P. R. EVANS & A. G. W. LESLIE, pp. 87–95. Proceedings of the CCP4 study weekend, 25–26 January 1991. SERC Daresbury Laboratory, Warrington, England.

WESTHOF, E., DUMAS, P. & MORAS, D. (1985). *J. Mol. Biol.* 184, 119–145.

WILSON, A. J. C. (1942). *Nature (London),* 150, 151–152.

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

The Heavy-Atom Problem: a Statistical Analysis. II. Consequences of the *A Priori* Knowledge of the Noise and Heavy-Atom Powers and use of a Correlation Function for Heavy-Atom-Site Determination

BY PHILIPPE DUMAS

*UPR de Biologie Structurale, Institut de Biologie Moléculaire et Cellulaire,
15 rue René Descartes, 67084 Strasbourg CEDEX, France*

Abstract

A preceding paper reported how to obtain *a priori* quantitative information on the lack of isomorphism (LOI), considered as noise corrupting the heavy-atom signal in a derivative data set. This related paper initially examines how additional *a priori* information can be drawn from the knowledge of the level of LOI. First, a corrected estimate of the coefficients necessary for a difference Patterson synthesis is derived. An estimate of their accuracy is also obtained. Then, individual and, independently, shell-averaged figures of merit that can be expressed in terms of the phasing power obtained in the preceding paper are determined. These afford an early estimate of the probable phase error on the heavy-atom structure factor. In a second and independent part of the paper, a correlation/translation function is proposed for the localization of the heavy-atom site(s). The results, bearing on both test and real cases, show that this method can be helpful in many situations.

1. Introduction

In a preceding paper (Dumas, 1994), from now on referred to as I, it was shown that a great deal of information can be obtained about the LOI corrupting a derivative data set before any heavy-atom sites are determined. This second paper is first devoted to drawing useful consequences from this knowledge, with regard first to re-estimating the best coefficients for a difference Patterson synthesis. All notation used in the paper is consistent with that used in I or is defined when necessary.

2. Estimation of the 'best' Patterson coefficients and of their quality

2.1. Previous results by Srinivasan

Generally, the coefficients used for calculating a difference Patterson function are $(F_{PH} - F_P)^2$ (Rossmann, 1960). It is well known that this is a crude approximation of F_H^2 for many acentric reflections. Srinivasan (1968) and Kalyanaraman & Srinivasan (1968) tried to improve on it by taking for F_H^2 its mean value $\langle F_H^2 \rangle$. Their work made the implicit assumption that the heavy atoms, considered as 'corrupting' the Fourier transform of the native crystal (Fig. 1a), are randomly distributed. They obtained (in a slightly different but equivalent form)

$$\langle F_H^2 \rangle = (F_{PH} - F_P)^2 + \Delta_{a,c}(X), \qquad (1)$$

where X is the reduced variable $2F_P F_{PH}/\Sigma_H$ [this notation is that used by Sim (1959, 1960); Srinivasan used half this value] and $\Delta_{a,c}(X)$ is a correction – always positive – for, respectively, acentric and centric terms. These are given by

$$\Delta_a(X) = \Sigma_H X\{1 - [I_1(X)/I_0(X)]\}$$
$$= \Sigma_H[X - G(X)] \qquad (2)$$
$$\Delta_c(X) = \Sigma_H X[1 - \tanh(X/2)]. \qquad (3)$$

When LOI is important, $\langle F_{HN}^2 \rangle$ and Σ_{HN}, not merely $\langle F_H^2 \rangle$ and Σ_H, have to be used in (1), (2) and (3). Incidentally, the noise component from LOI obeys the statistical premises of Wilson statistics more correctly than the heavy-atom component does. These results by Srinivasan and Kalyanaraman can readily be put into practice since, as shown in I, both Σ_H and Σ_{HN} can be satisfactorily estimated.

2.2. Estimation of $\langle F_H^2 \rangle$ from $\langle F_{HN}^2 \rangle$

We must account for the fact that we have obtained $\langle F_{HN}^2 \rangle$, and not $\langle F_H^2 \rangle$. From the Parseval theorem, one has, on average,

$$\overline{\langle F_H^2 \rangle} = \varphi_{H/HN} \overline{\langle F_{HN}^2 \rangle}, \qquad (4)$$

with the correction term $\varphi_{H/HN} = \Sigma_H/(\Sigma_H + \Sigma_N) = 1/(1 + \varphi_{N/H})$ being known from I. Although this correction is only valid on average, we have no alternative but to use it for each reflection and, thus,

$$\langle F_H^2 \rangle = \varphi_{H/HN} \langle F_{HN}^2 \rangle. \qquad (5)$$

Finally, coefficients for an origin-removed Patterson function are obtained as $\langle F_H^2 \rangle > -\Sigma_H$. Interestingly, $\varphi_{H/HN} = W^2/(1 + W^2)$, with $W = (\Sigma_H/\Sigma_N)^{1/2}$ being the phasing power, and thus $\varphi_{H/HN}$ appears as a natural damping factor of the coefficients in shells of resolution where the phasing power is low. It was verified on a test case with significant LOI that the latter estimate for $\langle F_H^2 \rangle$ is slightly better correlated to the exact values than the classic estimate $(F_{PH} - F_P)^2$ (not shown). The difference observed on a Patterson function calculated with, respectively, exact, classic and 'dampened Srinivasan' coefficients is shown in Fig. 2. Two test cases (already used in I) have been examined with medium and high LOI. For medium LOI, no clear improvement is obtained, apart from a slight noise reduction. On the contrary, for high LOI, a significant difference is observed, especially for the Harker section $z = \frac{1}{2}$, where the noise reduction is very significant. However, as expected, this also results in a loss of sharpness due to the damping. Therefore, this Patterson synthesis is probably better used in conjunction with the classical one, in particular to observe which peaks are common to both.

2.3. Accuracy of $\langle F_H^2 \rangle$: calculation of $\sigma(F_H^2)$

As a straightforward continuation of Srinivasan's work, we calculate the variance of F_H^2: the smaller this variance, the better the estimate of $F_H^2 = \langle F_H^2 \rangle$ (or of $F_{HN}^2 = \langle F_{HN}^2 \rangle$) by (1). We are only concerned here with the uncertainty of $\langle F_H^2 \rangle$ of 'geometrical' origin, not with the uncertainty due specifically to LOI.

By definition,

$$\mathrm{var}(F_H^2) = \langle (F_H^2 - \langle F_H^2 \rangle)^2 \rangle$$
$$= \langle F_H^4 \rangle - \langle F_H^2 \rangle^2. \qquad (6)$$

2.3.1. The case of acentric reflections. $\langle F_H^2 \rangle^2$ is known from (1) and $\langle F_H^4 \rangle$ can be calculated in exactly the same way as used by Srinivasan for $\langle F_H^2 \rangle$ (see Appendix 1 for details of calculations). One finally obtains

$$\sigma(F_H^2) = [\mathrm{var}(F_H^2)]^{1/2}$$
$$= \Sigma_H[X^2 - G(X) - G^2(X)]^{1/2}. \qquad (7)$$

The dependence of $\Delta_a(X)$ on X is shown in Fig. 3(a). The dependence of the ratio $\Delta_a(X)/\sigma(F_H^2)$ on X, allowing us to grasp the relevance of the difference between the 'corrected' and the 'classical' estimates of
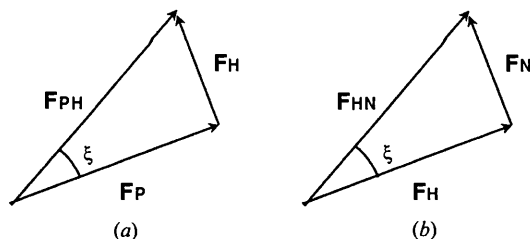

Fig. 1. Argand diagrams of the different structure factors discussed in the text. (a) Influence of heavy atoms 'corrupting' the Fourier transform of a macromolecular crystal. (b) Influence of noise originating from LOI 'corrupting' the Fourier transform of heavy atoms.

F_H^2, is shown in Fig. 3(b). With consideration of (2) and (7), the latter ratio is given by

$$\Delta_a(X)/\sigma(F_H^2) = [X - G(X)]$$
$$\times [X^2 - G(X) - G^2(X)]^{-1/2}. \quad (8)$$

This function runs from $2^{1/2}$ at $X = 0$ to $2^{-1/2}$ for large X, passing through a very weakly marked minimum at $X \simeq 3.59$. This result indicates that the correction $\Delta_a(X)$ is of statistical significance roughly for X lying in the range 0–2. This is rather limited since

$$\overline{X} = 2\overline{F_P F_{PH}}/\Sigma_H \simeq 2\overline{F_P^2}/\Sigma_H = 2\Sigma_P/\Sigma_H, \quad (9)$$

meaning that $\overline{X}$ commonly has a value well above 2. As a consequence, the terms $\langle F_H^2 \rangle / \sigma(F_H^2)$, which can be calculated for every reflection, reduce most often to the following limiting form corresponding to large X:

$$\langle F_H^2 \rangle / \sigma(F_H^2) = 2^{1/2}\{[(F_{PH} - F_P)^2/\Sigma_H] + \tfrac{1}{2}\}. \quad (10)$$

2.3.2. The case of centric reflections.

Calculation of $\langle F_H^4 \rangle$ for centric reflections is performed analogously and the result is obtained after easy calculations:

$$\sigma(F_H^2) = [\text{var}(F_H^2)]^{1/2}$$
$$= \Sigma_H X [1 - \tanh^2(X/2)]^{1/2}$$
$$= \Sigma_H X / \cosh(X/2). \quad (11)$$

This allows us to calculate the corresponding ratio $\Delta_c(X)/\sigma(F_H^2)$:

$$\Delta_c(X)/\sigma(F_H^2) = \exp(-X/2), \quad (12)$$

which indicates that the correction for centric terms is of very poor quality.

3. A priori determination of various figures of merit

The aim of this section is the calculation of the probable phase error on the heavy-atom structure factors that can
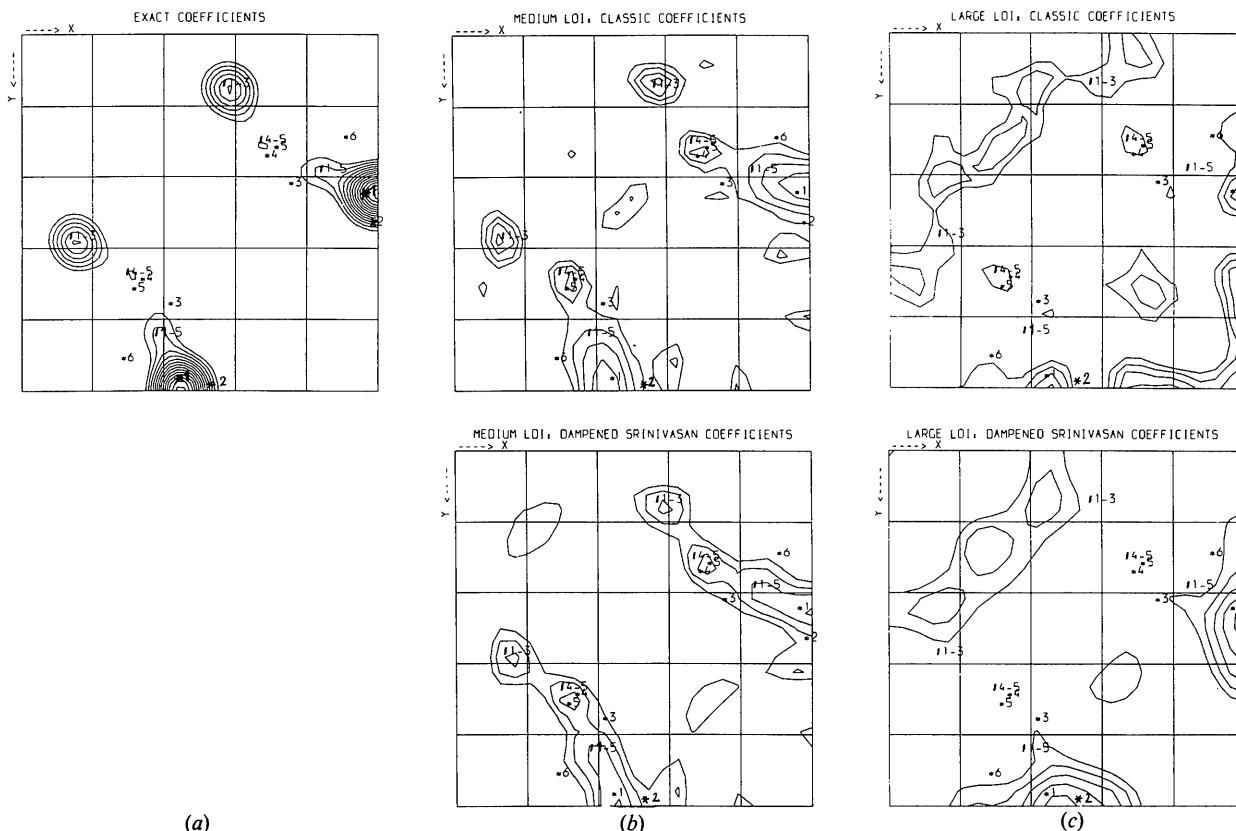

Fig. 2. Comparison of the Harker sections $z = \tfrac{1}{2}$ for a test Patterson function calculated between 12 and 3 Å with: (a) exact coefficients (i.e. F_H^2 calculated from heavy-atom coordinates); (b) medium LOI; (c) high LOI. For both (b) and (c), two estimates for the coefficients have been compared: the classical estimate [$F_H^2 = (F_{PH} - F_P)^2$] and the dampened Srinivasan estimate [$\langle F_H^2 \rangle = \varphi_{H/HN}\langle F_{HN}^2 \rangle$, (5)]. The contours are at $\tfrac{1}{2}$ e.s.d. from each other and start at 1 e.s.d. of the section. The * (singly numbered) correspond to the self-peaks that should appear in a Harker section and the # (doubly numbered) correspond to those cross peaks that happen to fall in the immediate vicinity of this section. The grid spacing corresponds to 0.1 crystallographic units.

be predicted from our knowledge of the level of noise from LOI. In the mulitple-isomorphous-replacement case, this quantity is an important component of the usual figure of merit (FOM) derived after an analysis as first performed by Blow & Crick (1959). In the single-isomorphous-replacement case, the phase is taken as the average of the two possible phases $\varphi_H + \delta\varphi$ and $\varphi_H - \delta\varphi$, φ_H being the phase of $\mathbf{F}_H$. The FOM, in such a case, is thus usually taken as $\cos\delta\varphi$. This estimate, however, is grossly overoptimistic if φ_H is ill known and $\delta\varphi \simeq 0$. Such a calculation may thus be of interest for obtaining a more realistic weighting in the single-isomorphous-replacement case. We will determine both a shell-averaged and an individual estimate of this FOM. Here, we restrict ourselves to the theoretical aspect and additional work is required to practically assess these results.

3.1. Determination of a shell-averaged figure of merit

We want to estimate $\overline{m_a} = \overline{\langle\cos\xi\rangle}$, for acentric reflections, with ξ defined as the angle $(\mathbf{F}_H, \mathbf{F}_{HN})$ (Fig. 1b). For centric reflections, the analogous quantity is

$\Delta_a(X)/\Sigma_H$

(a)

$\Delta_a(X)/\sigma(F_H^2)$

(b)

Fig. 3. (a) Variation of the normalized Srinivasan correction $\Delta_a(X)/\Sigma_H$. (b) Variation of $\Delta_a(X)/\sigma(F_H^2)$.

$\overline{m_c} = \overline{\langle\mathrm{sign}\rangle}$ with sign being the sign of $\mathbf{F}_H$ relative to $\mathbf{F}_H + \mathbf{F}_N = \mathbf{F}_{HN}$. In both cases, the brackets denote the average over all possible values of $\mathbf{F}_H$ and $\mathbf{F}_N$ suitably weighted from Wilson statistics. The horizontal bar denotes the average over all reflections in a given shell of resolution where Σ_H and Σ_N are known approximately from I.

3.1.1. *The case of acentric reflections.* From Sim's (1959, 1960) results, we know that for a given reflection $\langle\cos\xi\rangle = I_1(X)/I_0(X)$, with $X = 2F_H F_{HN}/\Sigma_N$. In order to perform the second averaging (the ensemble average corresponding to the horizontal bar), we need to calculate $\langle I_1(X)/I_0(X)\rangle$ and, thus, we need to know $P_a(X|\Sigma_H, \Sigma_N)$, the density probability of the variable X when Σ_H and Σ_N are known. Its calculation, detailed in Appendix 2A, gives

$$P_a(X|\Sigma_H, \Sigma_N) = \varphi_{N/H} X I_0(X) K_0[(1 + \varphi_{N/H})^{1/2} X], \tag{13}$$

with K_0 the Bessel function of imaginary argument and zeroth order [Gradshteyn & Ryzhik (1980) (from now on referred to as GR), pp. 951-952]. Obviously, this result is not limited to the pair F_H, F_{HN}: it is valid for any pair of structure-factor moduli F_S and F_{ST}, provided that both F_S and F_T follow Wilson statistics. Use of this expression gives

$$\overline{m_a}(\varphi_{N/H}) = \varphi_{N/H} \int_0^{\infty} X I_1(X)$$
$$\times K_0[(1 + \varphi_{N/H})^{1/2} X]\mathrm{d}X, \tag{14}$$

which yields (GR, p. 693)

$$\overline{m_a}(\varphi_{N/H}) = (\pi/4)[\varphi_{N/H}/(1 + \varphi_{N/H})^{3/2}]$$
$$\times {}_2F_1(\tfrac{3}{2}, \tfrac{3}{2}; 2; 1/(1 + \varphi_{N/H})), \tag{15}$$

with ${}_2F_1(\alpha, \beta; \gamma; z)$ the hypergeometric function (Nikiforov & Ouvarov, 1976, p. 196; GR, p. 1045). If one uses the phasing power W as the variable, the latter equation reads

$$\overline{m_a}(W) = (\pi/4)[W/(1 + W^2)^{3/2}]$$
$$\times {}_2F_1(\tfrac{3}{2}, \tfrac{3}{2}; 2; W^2/(1 + W^2)). \tag{16}$$

It is possible to express the hypergeometric function in terms of complete elliptical functions but this does not seem worthwhile. The variation of $\overline{m_a}(W)$ is shown in Fig. 4.

3.1.2. *The case of centric reflections.* We want to derive $\overline{m_c} = \overline{\langle\mathrm{sign}\rangle}$, with sign being the sign of $\mathbf{F}_H$ relative to $\mathbf{F}_H + \mathbf{F}_N = \mathbf{F}_{HN}$. For this one-dimensional problem, the double averaging can easily be done at once by considering all possible values of $\mathbf{F}_H$ and, for each of them, all possible values of $\mathbf{F}_N$ such that $\mathbf{F}_H(\mathbf{F}_H + \mathbf{F}_N)$
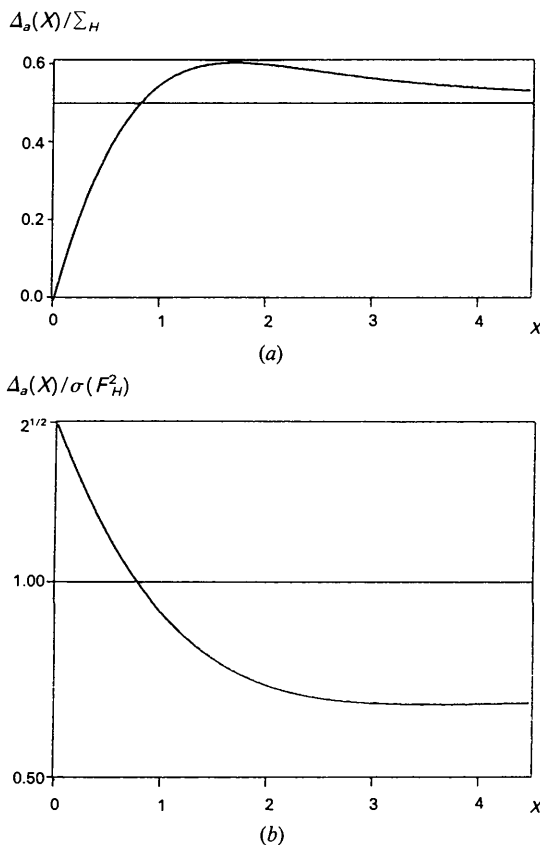
is either positive or negative. This leads to

$$\overline{m_c} = 2\int_0^\infty P_{\Sigma_H}(F_H)\left[\int_{-F_H}^\infty P_{\Sigma_N}(F_N)\mathrm{d}F_N\right.$$
$$\left. - \int_{-\infty}^{-F_H} P_{\Sigma_N}(F_N)\mathrm{d}F_N\right]\mathrm{d}F_H, \qquad (17)$$

with $P_{\Sigma_{H,N}}(F_{H,N})$ being the Wilson centric distribution with variances Σ_H and Σ_N. The factor 2 arises from the integration over $\mathbf{F}_H$ being performed between 0 and ∞ instead of between $-\infty$ and ∞. Taking care of the normalization of a density probability function, one obtains

$$\overline{m_c} = 4\int_0^\infty P_{\Sigma_H}(F_H)\left[\int_0^{F_H} P_{\Sigma_N}(F_N)\mathrm{d}F_N\right]\mathrm{d}F_H. \qquad (18)$$

The calculations necessary for this integration are detailed in Appendix 2B; they give

$$\overline{m_c}(W) = (2/\pi)[W/(1+W^2)]$$
$$\times {}_2F_1(1,1;\tfrac{3}{2};W^2/(1+W^2)). \qquad (19)$$

In this case, a remarkable simplification occurs (see Appendix 2B for full details), yielding

$$\overline{m_c}(W) = (2/\pi)\arctan W, \qquad (20)$$

with $W = (\Sigma_H/\Sigma_N)^{1/2}$ as the variable. It is worth noting the similarity of the two results for $\overline{m_a}(W)$ [(16)] and $\overline{m_c}(W)$ [(19)], which can both be expressed in terms of a hypergeometric function.* The variation of $\overline{m_c}(W)$ is shown in Fig. 4.

3.2. Determination of a figure of merit for each reflection

We turn now to the determination of individual figures of merit.

---

* Incidentally, one may also stress that the two multiplying factors, $\pi/4$ for acentric reflections and $2/\pi$ for centric reflections, correspond to the respective values of $\rho = \langle F\rangle^2/\langle F^2\rangle$ given by Wilson (1949).
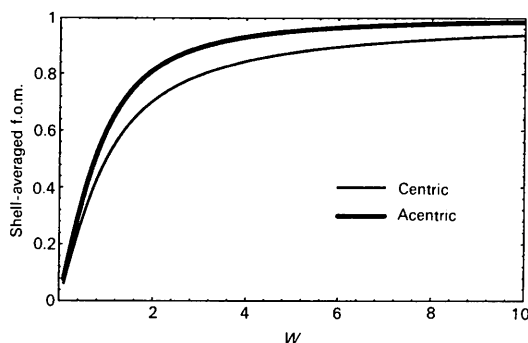

Fig. 4. Variation of the acentric and centric shell-averaged figures of merit $\overline{m_a}(W)$ (thick line) and $\overline{m_c}(W)$ (thin line) *versus* the phasing power W.

3.2.1. The case of acentric reflections.
Analogously to the previous section, one is led to determining $P_a(X|\Sigma_H,\Sigma_N,F_{HN})$, the density probability of $X = 2F_H F_{HN}/\Sigma_N$ when F_{HN} is considered fixed. $P_a(X|\Sigma_H,\Sigma_N,F_{HN})$ is readily obtained from the acentric Wilson distribution considering that, for a given X, F_H must be equal to $\Sigma_N X/(2F_{HN})$. Thus,

$$P_a(X|\Sigma_H,\Sigma_N,F_{HN}) = 2\alpha X\exp(-\alpha X^2) \qquad (21)$$

with $\alpha = \Sigma_N^2/(4\Sigma_H F_{HN}^2)$. Interestingly, F_{HN}^2, not F_{HN}, is required and the known quantity $\langle F_{HN}^2\rangle$ is therefore useful to evaluate α. The individual figure of merit is thus obtained as

$$m_a(\alpha) = 2\alpha\int_0^\infty X[I_1(X)/I_0(X)]\exp(-\alpha X^2)\mathrm{d}X. \qquad (22)$$

Contrary to the previous shell-averaged estimate, no closed form seems obtainable and the integral must be numerically evaluated.

3.2.2. The case of centric reflections.
For centric reflections, we deal again with the average value $\langle \mathrm{sign}\rangle$ of the two possible signs of $\mathbf{F}_H$ relative to $\mathbf{F}_P$. First, let us suppose that $\Sigma_N = 0$; it can immediately be shown from Woolfson's (1956) result that $\langle \mathrm{sign}\rangle = \tanh X/2$ if $F_{PH} > F_P$ and $\langle \mathrm{sign}\rangle = -1$ if $F_{PH} < F_P$. Now, if $\Sigma_N > 0$, the previous result can be used if F_P is replaced by $|\mathbf{F}_P + \mathbf{F}_N|$ and if all possible values of $\mathbf{F}_N$ are properly weighted by the centric Wilson probability density. After simple calculations, one obtains

$$\langle\mathrm{sign}\rangle = (2\pi\Sigma_N)^{-1/2}\int_{f_{\min}}^{f_{\max}}\exp(-f^2/2\Sigma_N)$$
$$\times\{1 + \tanh[((|F_P + f|F_{PH})/\Sigma_H]\}\mathrm{d}f - 1 \qquad (23)$$

with $f_{\min} = -F_P - F_{PH}$ and $f_{\max} = F_{PH} - F_P$. Again, this integral must be evaluated numerically. In the particular case where $\Sigma_N \to 0$, the Gaussian weighting term is equivalent to the Dirac distribution $\delta(f)$ and one retrieves correctly the previous result.

4. Using a correlation function to determine heavy-atom positions

In this second part of the paper, a simple method is described whose goal is heavy-atom localization without making explicit use of the Patterson function. We state clearly that it does not pretend to be a substitute for the latter in all cases; however, it is simple and very often yields directly the coordinates of at least the major site without need for further interpretation. In particular, it was successfully used alone for the determination of heavy-atom sites of the bleomycin resistance protein (Dumas, Bergdoll, Cagnon & Masson, 1994).

4.1. Principle of the method

The method is based on the use of a well known and robust tool, namely a correlation function (CF), calculated at each point of a grid, between the estimate of F_H^2 for each reflection and its calculated value if the heavy-atom were located at this point. This is certainly not new, since it has been used (or analogously with the classic R factor) many times for small molecules. However, to our knowledge, it has not been systematically used in the field of macromolecules for that purpose, but only as a checking tool as in the program *PROTEIN* (Steigemann, 1982). The correlation function is defined as

$$C_F(\mathbf{r}) = \sum_{\mathbf{h}} \left(F_{He}^2 - \overline{F_{He}^2} \right) \left(F_{Hc}^2 - \overline{F_{Hc}^2} \right)$$
$$\times \left[\sum_{\mathbf{h}} \left(F_{He}^2 - \overline{F_{He}^2} \right)^2 \right]^{-1/2}$$
$$\times \left[\sum_{\mathbf{h}} \left(F_{Hc}^2 - \overline{F_{Hc}^2} \right)^2 \right]^{-1/2}, \qquad (24)$$

in which the subcripts e and c stand for estimated and calculated [F_{Hc}^2 is in fact a short notation for $F_{Hc}^2(\mathbf{r})$]. The position where this function takes its maximum value is likely to correspond to a heavy-atom site. In fact, it is more convenient to make use of the analogous function $C_E(\mathbf{r})$ bearing on normalized intensities $E_H^2 = F_H^2/\Sigma_H$. This ensures that a single value for $\overline{E_{He,c}^2}$ (ideally 1) is relevant at any resolution, as it should be. The results obtained in I (for Σ_H and Σ_N) and in this paper [(5)] are used for estimating F_{He}^2 and E_{He}^2.

In fact, $C_F(\mathbf{r})$ is nothing other than the translation function as proposed by Fujinaga & Read (1987). It is easily shown that it can be understood as the coefficient of correlation between two 'origin-removed' Patterson functions, namely the one calculated with terms F_{He}^2 and the theoretical one calculated with terms F_{Hc}^2. Therefore, it is proposed here to use the method of Terwilliger & Eisenberg (1983) not only to refine heavy-atom parameters but also to try to find at least the major heavy-atom site.*

4.2. Discussion on the discernability of the heavy-atom site(s)

There are clearly two symmetrical drawbacks attached to this method. The first one concerns the limited accuracy of the terms F_{He}^2 and is therefore identical to the one attached to a difference Patterson. The second one concerns the limited accuracy of the terms F_{Hc}^2 owing to the assumption that there exists one, and only one, site accounting for the whole difference between the terms

---

* It should be mentioned, however, that Terwilliger & Eisenberg did not use a coefficient of correlation to be maximized but rather a sum of squared residuals to be minimized.

F_P and F_{PH}. We deal in this section with the second drawback.

Let us divide Σ_H *a priori* into two parts:

$$\Sigma_H = \Sigma_{H_Q} + \Sigma_{H_R}, \qquad (25)$$

where the index Q stands for one particular site and R for the other sites or, in terms of a model for a translation function, for the missing atoms of the model. Clearly, the signal-to-noise ratio of the proposed correlation function depends critically on the ratio $\varphi_{Q/H} = \Sigma_{H_Q}/\Sigma_H$ lying between 0 and 1. This can be approximately expressed in a quantitative manner with consideration of a result of Hauptman (1982) (see also Fujinaga & Read, 1987), allowing one to relate $C_E(\mathbf{r})$ to $\varphi_{Q/H}$:

$$C_E(\mathbf{r}_Q) \simeq \varphi_{Q/H} D^2, \qquad (26)$$

where $\mathbf{r}_Q$ corresponds to the location of the site Q. In fact, there is no strict correspondence for the term $D = \langle \cos(2\pi \mathbf{h} \cdot \Delta \mathbf{r}) \rangle$, this only being meaningful in the case of a standard translation function with $\Delta \mathbf{r}$ standing for the coordinate error of the model being used. However, it is easy, at least intuitively, to make the transposition in the scope of our problem by replacing D^2 by C_{Emax}, the maximum value that can be attained by the CF. This maximum value depends on the quality of the estimated terms F_{He}^2 and, therefore, on all kinds of errors (including those due to LOI) and on the number of centric terms relative to others.* The exact value of this maximum is unimportant, only the following relationship being meaningful for our purpose:

$$C_E(\mathbf{r}_Q) \simeq \varphi_{Q/H} C_{Emax}, \qquad (27)$$

which states that the sum of the values of the CF at the positions of the heavy-atom sites should be constant and equal to C_{Emax}. This is verified quantitatively in § 4.4. When put into words, this means that whatever the absolute 'strength' of one or several sites, they have to 'share a given amount of information'. This clearly points to the fact that a weak single site can be as distinguishable, or even more so, than several equally strong sites (Table 1).

4.3. Symmetry of the correlation function

The paper by Hirshfeld (1967), in which is defined the so-called 'Cheshire' group of a space group, gives all theoretical and practical indications for defining the

---

* It might be thought that the term D^2 has another meaning, namely that it reflects the error due to grid size in the position of the trial heavy atom. This is certainly incorrect as it is assumed in this study that the heavy-atom sites are considered at their exact locations. See Table 1 for a clear proof of this statement.

Table 1. *Results of test calculations between* 10 *and* 3 Å *resolution on the evolution of the value of the CF for one to four heavy-atom sites and various relative importances of each*

The heavy atom of the ith site has an occupancy equal to Q_i and an isotropic temperature factor equal to B_i (Å^2). 920 atoms are present in the asymmetric unit, the space group is $P2_12_12_1$ and the cell parameters are $a = 50$, $b = 60$ and $c = 70$ Å. In column C_i are given on each line two values of the CF corresponding to site i. The upper value is the one at the exact site position, the lower value is the one at the closest grid point from this site. The grid size is 0.75 Å (one quarter of the resolution). It is seen that there can exist a very significant difference between the two values. In the last column is given the value of the sum of the CF at the different sites, which is reasonably, if not exactly, constant. In the last line, the third value (in parentheses) is the theoretical one obtained by use of (27) with $C_{E\max} = 0.702$: the agreement with the observed value is good.

Q_1,B_1	Q_2,B_2	Q_3,B_3	Q_4,B_4	C_1	C_2	C_3	C_4	$\sum C_i$
0.3,30				$\begin{cases} 0.758 \\ 0.583 \end{cases}$				0.758
	0.5,30				$\begin{cases} 0.769 \\ 0.618 \end{cases}$			0.769
		0.7,30				$\begin{cases} 0.771 \\ 0.661 \end{cases}$		0.771
			0.9,30				$\begin{cases} 0.732 \\ 0.531 \end{cases}$	0.732
0.3,30	0.3,30			$\begin{cases} 0.343 \\ 0.263 \end{cases}$	$\begin{cases} 0.366 \\ 0.298 \end{cases}$			0.709
0.3,30	0.3,30	0.3,30		$\begin{cases} 0.243 \\ 0.155 \end{cases}$	$\begin{cases} 0.205 \\ 0.162 \end{cases}$	$\begin{cases} 0.266 \\ 0.251 \end{cases}$		0.714
0.3,30	0.3,30	0.3,30	0.3,30	$\begin{cases} 0.175 \\ 0.106 \end{cases}$	$\begin{cases} 0.159 \\ 0.130 \end{cases}$	$\begin{cases} 0.187 \\ 0.175 \end{cases}$	$\begin{cases} 0.158 \\ 0.112 \end{cases}$	0.679
0.3,30	0.3,40	0.3,50	0.3,60	$\begin{cases} 0.256 \\ 0.172 \\ (0.249) \end{cases}$	$\begin{cases} 0.179 \\ 0.149 \\ (0.190) \end{cases}$	$\begin{cases} 0.163 \\ 0.171 \\ (0.147) \end{cases}$	$\begin{cases} 0.104 \\ 0.072 \\ (0.116) \end{cases}$	0.702

new symmetry and new unit cell relevant for the CF. We recall here a few important points. All considerations on the symmetry of the CF derive from the fact that only the Harker vectors are needed for the calculation of each F_{Hc}^2. Therefore, any symmetry operation leaving unchanged the set of Harker vectors will belong to the set of symmetries of the CF. In particular, the CF gives rise to the same ambiguity of origin as the Patterson function.

It is worth recalling that, at variance with the Patterson function, the CF is not always centrosymmetric. This stems from the fact that the two constellations of heavy atoms corresponding to a generic one at either r_Q or $-r_Q$ are not always the mirror images of each other. Indeed, this is the case for all space groups that are their own enantiomorph but *not* for all chiral space groups.

In such a situation, *i.e.* for chiral space groups, the departure from centrosymmetry can be more or less pronounced for two opposite reasons. On the one hand, all centric reflections will fully contribute to centrosymmetry; on the other hand, only those symmetry operations responsible for chirality will contribute to the lack of centrosymmetry of the resulting CF. As a consequence, a group like $P4_12_12$, with as many as three independent centric zones and only four, out of eight, symmetry operations responsible for chirality, will produce a weak departure from centrosymmetry (see Fig. 5). In contrast, a group like $P3_1$, with no centric zones and as many as two out of three symmetry operations responsible for chirality, will produce an important departure from centrosymmetry.

4.4. Results

4.4.1. Test cases.
Numerous test calculations have been performed to investigate the intrinsic limitations of the method. In particular, the effect of the existence of several sites has been examined. Quantitative results are given in Table 1. One can effectively verify that, when passing from one site to several sites, the values of the CF at the different sites sum to approximately the same value as stated by (27). Also, the theoretical values of the CF at the different sites are in good agreement with the observed ones.

Another important fact emerges from these results and concerns the importance of the sampling. Indeed, in the present test case it appears very clearly that the usual rule of thumb of a grid size equal to one quarter of the resolution may not be sufficient. In practice, one may perform calculations with such a grid size for the sake of saving CPU time, but one should afterwards explore the highest peaks in more detail.

4.4.2. Real cases.
Several structures solved in our laboratory have been used to test the method. These structures represent a significant sample of problems in terms of the nature of the molecule (proteins and RNA), molecule size (from 14 to 120 kDa in the asymmetric unit) and space group ($C2$, $C222_1$, $P2_12_12$, $P2_12_12_1$, $P4_12_12$, $P4_32_12$, $P3_121$ and $P6_122$). In all but one of these cases, the first peak of the CF did correspond to a heavy-atom site.

The method has in particular been successfully used to solve all heavy-atom derivatives of the bleomycin

resistance protein (space group $P4_12_12$, $a = b = 48.4$, $c = 111.5$ Å), whose structure was recently determined (Dumas, Bergdoll, Cagnon & Masson, 1994). It must be emphasized that the Patterson function was not necessary, at any stage, for any of the derivatives. In particular, the method proved to be fully efficient in solving a europium derivative with as many as six different sites. The major site was unambiguously identified as the first peak of the correlation map (10–4 Å) with height 0.315, well detached from the next peak of height 0.25, the average and e.s.d. values of the map being 0.047 and 0.043, respectively (Fig. 5). The position obtained for this site was 0.28 Å from the position after refinement. All other sites were readily determined by successive residual Fourier maps. Refinement of their parameters between 12 and 3.3 Å resolution (program *REFINE* from the *CCP*4 package) finally converged to the following set of relative occupancies and temperature factors $Q_1 = 6.2$, $B_1 = 57$ Å^2; $Q_2 = 2.7$, $B_2 = 92$ Å^2; $Q_3 = 2.4$, $B_3 = 80$ Å^2; $Q_4 = 1.2$, $B_4 = 27$ Å^2; $Q_5 = 0.9$, $B_5 = 59$ Å^2; $Q_6 = 0.5$, $B_6 = 40$ Å^2. These values allow one to roughly estimate the ratio $\varphi_{Q/H}$ as 0.75. According to (27), this gives ca 0.41 as the maximum possible value for the CF, which should also be the sum of the values of the CF at the different sites. In fact, this value (0.41) is significantly lower than the one observed (0.717). This discrepancy is certainly the result of errors due to LOI but also of those due to a refinement procedure that allowed too much freedom in the highly correlated occupancies and temperature factors of the different sites. This points to a possible improvement of such a refinement procedure that would explicitly restrain the individual 'strength' of each site to be in reasonable agreement with the observed CF value following (27).

Another very significant test was the one made on anti-thrombin III (ATIII), recently solved in our laboratory (Samama, Delarue, Mourey, Choay & Moras, 1989; Delarue, Samama, Mourey & Moras, 1990; Mourey, 1991). This protein crystallizes with a dimer in the asymmetric unit in the tetragonal space group $P4_32_12$, $a = b = 91.3$, $c = 383$ Å. One platinum derivative of moderate quality was obtained that resisted interpretation by the use of the difference Patterson function. It could only be interpreted with the use of molecular-replacement phases. Retrospectively, one major difficulty was the occurrence of several cross peaks in the Harker sections. Tests subsequently made with the CF revealed two sites, out of five, as the first and the third peaks in the list. The second spurious peak had in fact x and y of the first site and z of the second site. The heights of the two correct peaks were very low: 0.078 and 0.062 (overall average = 0, e.s.d. = 0.015). Such low values for a correlation coefficient could be considered, at first sight, insignificant. They are, however, in agreement with (27) because of the significant level of LOI ($B_N = 2.1$ Å^2; see Table 3 of I), which decreases D^2, and of the great number of sites, which decreases $\varphi_{Q/H}$. It is fair to state that probably to a large extent chance allowed the second site to emerge out of noise.

The only case encountered so far that did not give one unambiguous major site as the first peak of the CF concerns the gold derivative of one crystal form of tRNA^{Asp} from yeast (Moras *et al.*, 1980). tRNA^{Asp} crystallized in two independent yet closely related crystal forms A and B. The gold induced a substantial LOI owing to interconvertibility from form A to a more B-like form and its phasing power was rather limited (Comarmond, Giegé, Thierry & Moras, 1986). The values obtained for
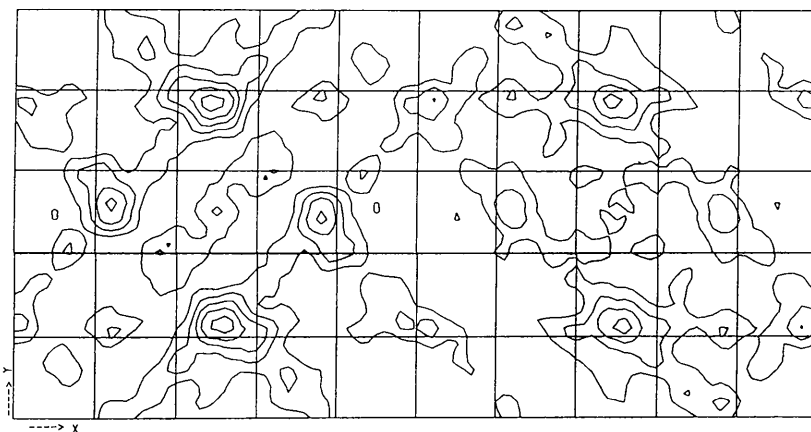

Fig. 5. Section $Z = 0.028$ of the CF calculated between 10 and 4 Å resolution for an europium derivative of the bleomycin resistance protein. X runs from 0 to 1 and Y from 0 to $\frac{1}{2}$ (grid size equal to 0.1 crystallographic units). The first contour is at 2.5 times the e.s.d. of the map and the interval between contours is equal to 1 e.s.d. The quasimirror about the diagonal is clear and corresponds to the weak departure from centrosymmetry for a chiral space group like $P4_12_12$ as discussed in the text. The highest peak at $X = 0.26$, $Y = 0.12$ corresponds to the major europium site. The difference in height between the correct site and its mirror image about the diagonal is 1.3 times the e.s.d of the map.

B_N and B_H were 1.4 and 90 Å^2, respectively (paper I). The high B_H value is in agreement with the fact that the temperature factor for one of the two gold sites could not be refined and was held fixed at 100 Å^2 during the early stage of the solution of the structure.

APPENDIX 1
Calculation of $\langle F_H^4 \rangle$ for acentric reflections

One can calculate $\langle (F_H^2)^n \rangle$ for any value of n and consider the particular case $n = 2$. Using the binomial formula with $F_H^2 = F_P^2 + F_{PH}^2 - 2F_P F_{PH} \cos \xi$ (Fig. 1a), one obtains

$$\langle (F_H^2)^n \rangle = \Sigma_H^n \sum_{k=0}^{n} (-1)^k C_n^k Z^{n-k} X^k \langle \cos^k \xi \rangle \quad (28)$$

with $Z = (F_P^2 + F_{PH}^2)/\Sigma_H$ and $\langle \cos^k \xi \rangle$ to be calculated with the probability density $p(\xi) = \exp(X \cos \xi)/[2\pi I_0(X)]$ (Sim, 1959, 1960). Using the integral representation of $I_0(X)$,

$$2\pi I_0(X) = \int_0^{2\pi} \exp(X \cos \xi) d\xi \quad (29)$$

and, by successive derivations with respect to X, one obtains

$$2\pi I_0^{(k)}(X) = \int_0^{2\pi} \cos^k(\xi) \exp(X \cos \xi) d\xi, \quad (30)$$

which, considering the expression for $p(\xi)$, gives

$$\langle \cos^k \xi \rangle = I_0^{(k)}(X)/I_0(X). \quad (31)$$

We can now go back to the particular case $n = 2$. First, noting that $I_0'(X) = I_1(X)$ and $I_1'(X) = I_0(X) - I_1(X)/X$ and, finally, considering (28) and (31), we obtain the result

$$\langle (F_H^2)^2 \rangle = \Sigma_H^2 [X^2 + Z^2 - (1 + 2Z)G(X)]. \quad (32)$$

APPENDIX 2

A. Calculation of the density probability of X

We seek to derive the density probability $P_a(X|\Sigma_H, \Sigma_N)$ of $X = 2F_H F_{HN}/\Sigma_N$ for acentric reflections, F_H and F_N obeying Wilson statistics. The density probability of F_H is

$$P(F_H) = (2F_H/\Sigma_H) \exp(-F_H^2/\Sigma_H) \quad (33)$$

and the conditional density probability of F_{HN}, given

F_H, is (Sim, 1959)

$$P(F_{HN}/F_H) = (2F_{HN}/\Sigma_N) \exp[-(F_H^2 + F_{HN})/\Sigma_N]$$
$$\times I_0(2F_H F_{HN}/\Sigma_N). \quad (34)$$

Now, for given F_H and X, F_{HN} must be equal to $\Sigma_N X/(2F_H)$ and, thus,

$$P_a(X|\Sigma_H, \Sigma_N)$$
$$= (\Sigma_N/2) \int_0^{\infty} P(F_H) P(\Sigma_N X/2F_H|F_H) dF_H/F_H \quad (35)$$

which is easily transformed from (33) and (34) into

$$P_a(X|\Sigma_H, \Sigma_N) = \varphi_{N/H} X I_0(X)$$
$$\times \int_0^{\infty} \exp(-AF_H^2 - B/F_H^2) dF_H/F_H \quad (36)$$

with $A = 1/\Sigma_H + 1/\Sigma_N$ and $B = \Sigma_N X^2/4$. By making the change of variable $u = 1/F_H^2$, one obtains

$$\int_0^{\infty} \exp(-AF_H^2 - B/F_H^2) dF_H/F_H$$
$$= \frac{1}{2} \int_0^{\infty} \exp(-Au - B/u) du/u \quad (37)$$

and (Prudnikov, Brychkov & Marichev, 1990, p. 344)

$$\int_0^{\infty} u^{\alpha-1} \exp(-pu - q/u) du = 2(q/p)^{\alpha/2} K_\alpha[2(pq)^{1/2}]. \quad (38)$$

In our case, $\alpha = 0$ and the desired result is obtained as

$$P_a(X|\Sigma_H, \Sigma_N) = \varphi_{N/H} X I_0(X)$$
$$\times K_0 \left[(1 + \varphi_{N/H})^{1/2} X \right]. \quad (39)$$

It can be immediately verified that this function is correctly normalized to 1 by integrating it between 0 and ∞ (GR, p. 672).

B. Calculation of the double integral for centric reflections

One wants to calculate

$$\overline{m_c} = 4 \int_0^{\infty} P_{\Sigma_H}(F_H) \left[\int_0^{F_H} P_{\Sigma_N}(F_N) dF_N \right] dF_H. \quad (40)$$

By replacing $P_{\Sigma_H}(F_H)$ and $P_{\Sigma_N}(F_N)$ by their respective expressions, one obtains

$$\overline{m_c} = [2/\pi(\Sigma_H \Sigma_N)^{1/2}] I(\alpha, \beta) \quad (41)$$

with $\alpha = 1/(2\Sigma_H)$, $\beta = 1/\Sigma_N$ and

$$I(\alpha,\beta) = \int_0^\infty \exp(-\alpha x^2)\left[\int_0^x \exp(-\beta y^2)dy\right]dx. \quad (42)$$

The inner integral is expressed by the probability integral function Φ (GR, p. 930). If one uses its series representation (GR, p. 931), $I(\alpha,\beta)$ reads

$$I(\alpha,\beta) = (1/\beta)\int_0^\infty \exp(-\gamma X^2)$$
$$\times \sum_{k=0}^\infty [2^k X^{2k+1}/(2k+1)!!]dX \quad (43)$$

with $X = \beta^{1/2}x$, $\gamma = 1 + \alpha/\beta$ and $(2k+1)!!$ being equal to the product of all odd numbers up to $2k+1$. The Gaussian term, ensuring the absolute convergence, authorizes the inversion of summation and integration that leads to (GR, p. 337)

$$I(\alpha,\beta) = (1/\beta)\sum_{k=0}^\infty \int_0^\infty \exp(-\gamma X^2)$$
$$\times [2^k X^{2k+1}/(2k+1)!!]dX$$
$$= (1/\beta)\sum_{k=0}^\infty [2^k/(2k+1)!!](k!/2\gamma^{k+1}). \quad (44)$$

After elementary manipulations, this equation is transformed into

$$I(\alpha,\beta) = (1/2\gamma\beta)\sum_{k=0}^\infty \left[k!/(k+\tfrac{1}{2})!\right]\gamma^{-4} \quad (45)$$

with $(k+\tfrac{1}{2})!$ the product of all half-integers from $\tfrac{1}{2}$ to $k+\tfrac{1}{2}$. The latter is immediately expressed in terms of the Euler Γ function as $(k+\tfrac{1}{2})! = \Gamma(\tfrac{3}{2}+k)/\Gamma(\tfrac{3}{2}) = (\tfrac{3}{2})_k$ where the notation $(a)_k = \Gamma(a+k)/\Gamma(a)$ has been introduced [in particular $(1)_k = k!$]. With this notation, the previous equation can be rewritten as

$$I(\alpha,\beta) = (1/2\gamma\beta)\sum_{k=0}^\infty \left[(1)_k(1)_k/(\tfrac{3}{2})_k\right](\gamma^{-k}/k!). \quad (46)$$

The infinite sum corresponds to the exact definition of the hypergeometric function $_2F_1(1,1;\tfrac{3}{2};1/\gamma)$ (Nikiforov

& Ouvarov (1976), p. 196; GR, p. 1045). One thus obtains, by using $W = \varphi_{H/N}^{1/2}$ as the variable,

$$\overline{m_c}(W) = (2/\pi)[W/(1+W^2)]$$
$$\times\ _2F_1\left(1,1;\tfrac{3}{2};W^2/(1+W^2)\right). \quad (47)$$

At this point, a rather fortunate relationship greatly simplifies the result. One has $_2F_1(1,1;\tfrac{3}{2};\sin^2\varphi) = \varphi/(\sin\varphi\cos\varphi)$ (GR, p. 1041) and, for $W = \tan\varphi$, the argument of the hypergeometric function can be exactly transformed into $\sin^2\varphi$. Transforming $\sin(\arctan W)$ and $\cos(\arctan W)$ back into their algebraically equivalent forms, one finally obtains the remarkably simple result

$$\overline{m_c}(W) = (2/\pi)\arctan W. \quad (48)$$

References

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
COMARMOND, M. B., GIEGÉ, R., THIERRY, J. C. & MORAS, D. (1986). *Acta Cryst.* **B42**, 272–280.
DELARUE, M., SAMAMA, J. P., MOUREY, L. & MORAS, D. (1990). *Acta Cryst.* **B46**, 550–556.
DUMAS, P. (1994) *Acta Cryst.* **A50**, 526–537.
DUMAS, P., BERGDOLL, M., CAGNON, C. & MASSON, J. M. (1994). *EMBO J.* **13**. In the press.
FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517–521.
GRADSHTEYN, I. S. & RYZHIK, I. M. (1980). *Tables of Integrals, Series, and Products.* London: Academic Press.
HAUPTMAN, H. (1982). *Acta Cryst.* **A38**, 289–294
HIRSHFELD, F. L. (1967). *Acta Cryst.* **A24**, 301–311.
KALYANARAMAN, A. R. & SRINIVASAN, R. (1968). *Z. Kristallogr.* **126**, 262–267.
MORAS, D., COMARMOND, M. B., FISCHER, J., WEISS, R., THIERRY, J. C., EBEL, J. P. & GIEGÉ, R. (1980). *Nature (London)*, **288**, 669–674.
MOUREY, L. (1991). Thesis, Univ. Louis Pasteur, Strasbourg, France.
NIKIFOROV, A. & OUVAROV, V. (1976). *Eléments de la Théorie des Fonctions Spéciales.* Moscow: Editions MIR.
PRUDNIKOV, A. P., BRYCHKOV, Y. A. & MARICHEV, O. I. (1990). *Integrals and Series*, Vol. 1. New York: Gordon and Breach.
ROSSMANN, M. G. (1960). *Acta Cryst.* **13**, 221–226.
SAMAMA, J. P., DELARUE, M., MOUREY, L., CHOAY, J. & MORAS, D. (1989). *J. Mol. Biol.* **210**, 877–879.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.
SRINIVASAN, R. (1968). *Z. Kristallogr.* **126**, 175–181.
STEIGEMANN, W. (1982). *PROTEIN. A Package of Crystallographic Programs for the Analysis of Proteins.* Martinsried, Germany.
TERWILLIGER, T. C. & EISENBERG, D. (1983). *Acta Cryst.* **A39**, 813–817.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.
WOOLFSON, M. M. (1956). *Acta Cryst.* **9**, 804–810.