

in Edinburgh on which the MD work was done, and for the support of the ILL facility. Dr A. W. Hewat gave us invaluable assistance with the experimental work. Finally, KR wishes to acknowledge the financial support of the Carnegie Trust for the Universities of Scotland.

## References

- PAWLEY, G. S. (1980). *J. Appl. Cryst.* **13**, 630–633.  
 PAWLEY, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.  
 REFSON, K. & PAWLEY, G. S. (1986). *Acta Cryst.* **B42**, 402–410.  
 REFSON, K. & PAWLEY, G. S. (1987a). *Mol. Phys.* In the press.  
 REFSON, K. & PAWLEY, G. S. (1987b). *Mol. Phys.* In the press.

## SHORT COMMUNICATIONS

*Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.*

*Acta Cryst.* (1987). **A43**, 731–733

**Calculation of structure factors from molecular models with anisotropic atoms.** By ZHONGXIAO PAN, ATSUKO TAKAYA and RICHARD B. HONZATKO, *Department of Biochemistry and Biophysics, Iowa State University, Ames, IA 50011, USA*

(Received 6 October 1986; accepted 5 February 1987)

## Abstract

The mathematical basis and test results are presented for an efficient method of calculation of structure factors from large molecular models with anisotropic atoms.

The development of area detectors and synchrotron sources for X-rays has led to a significant improvement in both the quality and resolution of observed data from crystalline macromolecules. The existence of data to a resolution of 1.5 Å may ultimately permit the refinement of some protein structures using anisotropic thermal parameters, provided these parameters are subject to restraints in the least-squares refinement (Hendrickson & Konnert, 1980). Essential to any refinement routine is an efficient method for the generation of structure factors. Ten Eyck (1977) presented an efficient approach to the calculation of structure factors for isotropic atoms, which requires the generation of a model electron density map followed by Fourier inversion of the map. We present here the mathematics and trial results of an approach analogous to Ten Eyck's but using anisotropic atoms. We find that Ten Eyck's two-step process produces structure factors of high accuracy to a resolution of at least 1.0 Å.

Ten Eyck, based on earlier work of Vand, Eiland & Pepinsky (1957), gave a suitable analytical approximation to the form factor

$$f(s) = C_1 \exp[-\frac{1}{4}D_1s^2] + C_2 \exp[-\frac{1}{4}D_2s^2] + C_3. \quad (1)$$

Numerical values for  $C_1$ ,  $D_1$ ,  $C_2$ ,  $D_2$  and  $C_3$  derive from a least-squares fit of (1) to values of  $f(s)$  calculated from first principles. The quantity  $s$  is the length of the reciprocal-lattice vector associated with the indices  $h$ ,  $k$ ,  $l$ :

$$s^2 = h^2 a^{*2} + k^2 b^{*2} + l^2 c^{*2} + 2hka^*b^* \cos \gamma^* \\ + 2hla^*c^* \cos \beta^* + 2klb^*c^* \cos \alpha^*.$$

We define the anisotropic temperature factor as  $\exp[-\frac{1}{4}\sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j]$ , where the  $B^{ij}$  are the standard doubly contravariant components of an anisotropic thermal tensor and the  $s_i$  are the components of the reciprocal-lattice vector  $s$ . Explicitly,  $s_1 = ha^*$ ,  $s_2 = kb^*$  and  $s_3 = lc^*$ . Multiplication of (1) by the anisotropic temperature factor gives

$$f(s) \exp\left(-\frac{1}{4}\sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j\right) \\ = C_1 \exp\left[-\frac{1}{4}\left(D_1s^2 + \sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j\right)\right] \\ + C_2 \exp\left[-\frac{1}{4}\left(D_2s^2 + \sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j\right)\right] \\ + C_3 \exp\left(-\frac{1}{4}\sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j\right). \quad (2)$$

Ten Eyck demonstrated the reduction of error in calculating structure factors by inflating the isotropic thermal parameters of all atoms by a constant value. The appropriate form for the inflation factor is  $\exp[-\frac{1}{4}Bs^2]$ , where  $B$  is an arbitrary constant. Multiplication of (2) by the inflation factor gives

$$f(s) \exp\left(-\frac{1}{4}\sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j\right) \exp(-\frac{1}{4}Bs^2) \\ = C_1 \exp\left[-\frac{1}{4}\left(D_1s^2 + Bs^2 + \sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j\right)\right] \\ + C_2 \exp\left[-\frac{1}{4}\left(D_2s^2 + Bs^2 + \sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j\right)\right] \\ + C_3 \exp\left[-\frac{1}{4}\left(Bs^2 + \sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j\right)\right]. \quad (3)$$

Using the definitions above for  $s_i$  and  $s^2$  in the exponent of the first term of the right-hand side of (3), we have, after collecting common terms,

$$D_1s^2 + Bs^2 + \sum_{i=1}^3\sum_{j=1}^3 B^{ij}s_i s_j \\ = (B^{11} + B + D_1)a^{*2}h^2 \\ + (B^{22} + B + D_1)b^{*2}k^2 + (B^{33} + B + D_1)c^{*2}l^2 \\ + 2[B^{12} + (B + D_1)\cos \gamma^*]a^*b^*hk \\ + 2[B^{13} + (B + D_1)\cos \beta^*]a^*c^*hl \\ + 2[B^{23} + (B + D_1)\cos \alpha^*]b^*c^*kl.$$

Table 1. Numerical values for the five-parameter approximation to the form factor

	$C_1$	$D_1$	$C_2$	$D_2$	$C_3$
C	1.9110	8.3180	2.6424	31.6550	1.4364
N	2.9764	7.8582	2.5000	26.4242	1.5169
O	4.0150	6.8187	2.3748	22.5909	1.5957

Similar expressions follow for the exponents of the second and third terms of (3). The expression above for the exponent of the first term is more compact in tensor notation:

$$D_1 s^2 + B s^2 + \sum_{i=1}^3 \sum_{j=1}^3 B^{ij} s_i s_j = \sum_{i=1}^3 \sum_{j=1}^3 {}^1R^{ij} h_i h_j,$$

where the  $h_i$  are the indices  $h, k, l$  and the symmetric tensor of second rank  ${}^1R$  has the elements  ${}^1R^{ij}$ :

$$\begin{aligned} {}^1R^{11} &= (B^{11} + B + D_1) a^* a^* \\ {}^1R^{22} &= (B^{22} + B + D_1) b^* b^* \\ {}^1R^{33} &= (B^{33} + B + D_1) c^* c^* \\ {}^1R^{12} &= [B^{12} + (B + D_1) \cos \gamma^*] a^* b^* = {}^1R^{21} \\ {}^1R^{13} &= [B^{13} + (B + D_1) \cos \beta^*] a^* c^* = {}^1R^{31} \\ {}^1R^{23} &= [B^{23} + (B + D_1) \cos \alpha^*] b^* c^* = {}^1R^{32}. \end{aligned}$$

One may define symmetric tensors  ${}^2R$  and  ${}^3R$  by analogy to the above for the second and third terms of (3). Thus, (3) becomes

$$\begin{aligned} f(s) \exp \left[ -\frac{1}{4} \sum_{i=1}^3 \sum_{j=1}^3 B^{ij} s_i s_j \right] \exp \left[ -\frac{1}{4} B s^2 \right] \\ = C_1 \exp \left[ -\frac{1}{4} \sum_{i=1}^3 \sum_{j=1}^3 {}^1R^{ij} h_i h_j \right] \\ + C_2 \exp \left[ -\frac{1}{4} \sum_{i=1}^3 \sum_{j=1}^3 {}^2R^{ij} h_i h_j \right] \\ - C_3 \exp \left[ -\frac{1}{4} \sum_{i=1}^3 \sum_{j=1}^3 {}^3R^{ij} h_i h_j \right]. \end{aligned}$$

The Fourier transform of the above gives the electron density of the atom:

$$\begin{aligned} \rho(x_1, x_2, x_3) &= (4\pi)^{3/2} V^{-1} \det({}^1R)^{-1/2} C_1 \\ &\times \exp \left[ -4\pi^2 \sum_{i=1}^3 \sum_{j=1}^3 {}^1Q_{ij} x_i x_j \right] \\ &+ (4\pi)^{3/2} V^{-1} \det({}^2R)^{-1/2} C_2 \\ &\times \exp \left[ -4\pi^2 \sum_{i=1}^3 \sum_{j=1}^3 {}^2Q_{ij} x_i x_j \right] \\ &+ (4\pi)^{3/2} V^{-1} \det({}^3R)^{-1/2} C_3 \\ &\times \exp \left[ -4\pi^2 \sum_{i=1}^3 \sum_{j=1}^3 {}^3Q_{ij} x_i x_j \right], \quad (4) \end{aligned}$$

where  $V$  is the volume of the direct basis, the  $x_i$  are the components of a vector  $\mathbf{x}$  expressed in the direct basis (fractional coordinates) and the  ${}^kQ_{ij}$  are the doubly covariant components of a tensor obtained by taking the inverse of the matrices  ${}^kR$ . Equation (4) gives the value for the electron density of an anisotropic atom in any general-

Table 2.  $R$  factor in shells between structure factors derived from model electron density and an analytical expression

Range in resolution ( $\text{\AA}$ )	$R$ factor for the isotropic case	$R$ factor for the anisotropic case
10.0-5.0	0.013	0.008
5.0-3.33	0.006	0.005
3.33-2.5	0.015	0.012
2.5-2.0	0.026	0.012
2.0-1.67	0.025	0.006
1.67-1.43	0.013	0.009
1.43-1.25	0.011	0.012
1.25-1.11	0.021	0.010
1.11-1.0	0.028	0.011

Table 3. The variation of the overall  $R$  factor to 1.0  $\text{\AA}$  resolution between structure factors derived from model electron density and an analytical expression as a function of  $B$  in equation (3)

Value of $B$ in equation (3)	$R$ factor for the isotropic case	$R$ factor for the anisotropic case
0.0	0.020	0.019
1.0	0.016	0.011
2.0	0.016	0.009
3.0	0.017	0.009
5.0	0.017	0.010
10.0	0.026	0.016

ized coordinate frame and therefore is applicable to all space groups.

We performed test calculations with (4) incorporated in Ten Eyck's (1977) program using the small-molecule crystal of 1-(*p*-nitrobenzylideneamino)guanidinium chloride (Serra & Honzatko, 1986), space group  $P2_1/c$ ,  $a = 8.406$ ,  $b = 11.490$ ,  $c = 11.510 \text{ \AA}$  and  $\beta = 101.89^\circ$ . In order to reflect more precisely the electron density levels of a protein, we omitted the hydrogens and the chloride ion from the model. The constants ( $C_1$ ,  $D_1$ ,  $C_2$ ,  $D_2$  and  $C_3$ ) for atoms N, C and O are from Pan & Honzatko (1987) and are reproduced in Table 1. Model electron density maps had 32 divisions along each axis. Fourier inversion of the map employed Ten Eyck's (1973) routines for space group  $P1$ . Structure factors calculated on the basis of an analytical expression came from programs of Dr Robert Jacobson, Ames Laboratory. All calculations were performed on a VAX 11/780 located in Ames Laboratory.

The  $R$  factor in shells between structure factors calculated from the model density and from the analytical expression appears in Table 2. The  $R$  factor is essentially constant as a function of resolution to at least 1.0  $\text{\AA}$ . The residual of 1 to 2% is due, in part, to the truncation of atoms beyond a limiting radius and to the lower accuracy of the five-parameter equation for the form factor (1) as opposed to the nine-parameter equation of the analytical calculation. The data of Table 3 imply that only a modest inflation (2.0  $\text{\AA}^2$ ) of the temperature factor minimizes the overall  $R$  factor. Although in theory the more inflated the temperature factor, the more accurate the result, round-off error of the computer also becomes more significant as temperature factors increase.

We thank the Office of Naval Research N00014-84-G-0094, the Petroleum Research Fund 16269-G4, the National Institutes of Health GM33828, and Ames Laboratory for support.

## References

- HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Acta Cryst.* **A36**, 344–350.  
 PAN, Z. & HONZATKO, R. B. (1987). *Comput. Appl. Chem.* In the press.  
 SERRA, M. A. & HONZATKO, R. B. (1986). *Acta Cryst.* **C42**, 1755–1757.  
 TEN EYCK, L. F. (1973). *Acta Cryst.* **A29**, 182–191.  
 TEN EYCK, L. F. (1977). *Acta Cryst.* **A33**, 486–492.  
 VAND, V., EILAND, P. F. & PEPINSKY, R. (1957). *Acta Cryst.* **10**, 303–306.

*Acta Cryst.* (1987). **A43**, 733–734

**Translation functions. Note on the use of symmetry in the minimization of structure-independent spurious maxima.** By DAVID A. LANGS, *Medical Foundation of Buffalo, Inc., 73 High Street, Buffalo, NY 14203, USA*

(Received 21 February 1987; accepted 18 May 1987)

## Abstract

In a previous paper [Langs (1985). *Acta Cryst.* **A41**, 578–582] formulae were derived to reduce the magnitudes of structure-independent spurious peaks which appear in translation syntheses. The present note describes how crystallographic symmetry may be exploited to simplify these calculations by a factor of a hundredfold. The solution maximum produced by this new formulation is often more than ten times larger than the largest spurious background peak.

## Introduction

Multiple-angle trigonometric expansions were shown to provide the basis for an unlimited number of exact algebraic relationships among the sine and cosine components ( $G_h$  and  $S_h$ ) of the translation-function phases. One such pair of formulae [Langs (1985), equations (13) and (14)] are

$$G_h = A_h[(G_k G_l - S_k S_l) \cos \Phi_{h,k} - (G_k S_l + S_k G_l) \sin \Phi_{h,k}] / A_k A_l, \quad (1)$$

$$S_h = A_h[(G_k G_l - S_k S_l) \sin \Phi_{h,k} + (G_k S_l + S_k G_l) \cos \Phi_{h,k}] / A_k A_l, \quad (2)$$

where the vector sum for the triple  $h+k+l=0$ . The terms  $A_h$  and  $\Phi_{h,k}$  are *a priori* known values as defined in the previous paper and are not important in this exposition. The original set of  $G_h$  values obtained from the translation-function coefficients will have inherent errors due to the limitations of the search model and the accuracy of the data, as will the set of associated  $S_h$  values which are initially estimated as

$$S_h \approx A_h \langle G_k G_l \sin \Phi_{hk} \rangle_k / \langle A_k A_l \rangle_k. \quad (3)$$

Averaging over the above triple relationships will allow these values to be refined to minimize those errors. This analysis must necessarily be restricted to those  $G_h$  values which have the largest magnitudes, as the full data set will generate an unmanageable number of triples for these calculations.

## Crystallographic symmetry

The accuracy of the above triple formulae is greatly improved by the redundancy which results as a consequence of crystallographic symmetry. For example, in the space group  $P2_1$ , the  $G_h$  and  $S_h$  terms for a molecular fragment  $p$  may be shown to be of the form

$$G_{hkl} = [ |E_{hkl}|^2 - |E_{hkl} p|^2 - |E_{h\bar{k}l} p|^2 ] \\ = 2(-1)^k |E_{hkl} p E_{h\bar{k}l} p| \\ \times \cos[4\pi(hx_p + lz_p) + \phi hkl_p + \phi h\bar{k}l_p], \quad (4)$$

$$S_{hkl} = 2(-1)^k |E_{hkl} p E_{h\bar{k}l} p| \\ \times \sin[4\pi(hx_p + lz_p) + \phi hkl_p + \phi h\bar{k}l_p]. \quad (5)$$

It is important to notice that the translation portion of the sine and cosine terms for axially related data having common  $h$  and  $l$  indices is independent of the value of  $k$ . The  $h+k+l=0$  vector-sum condition which pertains to the algebraic triples among the  $G_{hkl}$  and  $S_{hkl}$  terms applies only to those components linked to the translation vector contained in these sines and cosines. Thus the  $h+k+l=0$  condition applies only to the  $h$  and  $l$  indices for each of the three reciprocal-lattice vectors. This permits the use of numerous other valid 'triples' for which the sum of the  $k$  indices of the three vectors does not equal zero in these calculations. It can furthermore be shown as a consequence of this symmetry that: (1) these calculations need not be summed over the total number of independent triples as this summation can be factored, (2) not every  $G_h, S_h$  pair need be refined to determine their refinement values, and (3) the dimensions of the data arrays employed in these calculations may be greatly reduced. These advantages will allow the refinement to be about 100 times faster and use half of the computer memory previously required. The resultant phasing accuracy will be better because the full set of diffraction data may now be incorporated into the refinement.

## Factoring the triples

In space group  $P2_1$  the contribution to (1) and (2) for a common family of 'triples' of the sort

$$\sum_i^n \sum_j^m X(h_1, k_i, l_1) Y(h_2, k_j, l_2) \cos/\sin \Phi_{h_1, h_2} \quad (6)$$

need not be summed over the  $n \times m$  independent triples. Each of these contributors may be factorized to separate the  $h, k$  and  $l$  vectors such that

$$\sum X_k Y_l \cos \Phi_{h,k} = \cos \varphi \varphi_h (\sum X_k \cos \varphi \varphi_k \sum Y_l \cos \varphi \varphi_l \\ - \sum X_k \sin \varphi \varphi_k \sum Y_l \sin \varphi \varphi_l) \\ - \sin \varphi \varphi_h (\sum X_k \cos \varphi \varphi_k \sum Y_l \sin \varphi \varphi_l \\ + \sum X_k \sin \varphi \varphi_k \sum Y_l \cos \varphi \varphi_l), \quad (7)$$