References

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Acta Cryst.* A**36**, 344-350.

PAN, Z. & HONZATKO, R. B. (1987). *Comput. Appl. Chem.* In the press.

SERRA, M. A. & HONZATKO, R. B. (1986). *Acta Cryst.* C**42**, 1755-1757.

TEN EYCK, L. F. (1973). *Acta Cryst.* A**29**, 182-191.

TEN EYCK, L. F. (1977). *Acta Cryst.* A**33**, 486-492.

VAND, V., EILAND, P. F. & PEPINSKY, R. (1957). *Acta Cryst.* **10**, 303-306.

Acta Cryst. (1987). A**43**, 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.* A**41**, 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 ($G_{\mathbf{h}}$ and $S_{\mathbf{h}}$) of the translation-function phases. One such pair of formulae [Langs (1985), equations (13) and (14)] are

$$G_{\mathbf{h}}=A_{\mathbf{h}}[(G_{\mathbf{k}}G_{\mathbf{l}}-S_{\mathbf{k}}S_{\mathbf{l}})\cos\Phi_{\mathbf{h},\mathbf{k}} \\ -(G_{\mathbf{k}}S_{\mathbf{l}}+S_{\mathbf{k}}G_{\mathbf{l}})\sin\Phi_{\mathbf{h},\mathbf{k}}]/A_{\mathbf{k}}A_{\mathbf{l}}, \quad (1)$$

$$S_{\mathbf{h}}=A_{\mathbf{h}}[(G_{\mathbf{k}}G_{\mathbf{l}}-S_{\mathbf{k}}S_{\mathbf{l}})\sin\Phi_{\mathbf{h},\mathbf{k}} \\ +(G_{\mathbf{k}}S_{\mathbf{l}}+S_{\mathbf{k}}G_{\mathbf{l}})\cos\Phi_{\mathbf{h},\mathbf{k}}]/A_{\mathbf{k}}A_{\mathbf{l}}, \quad (2)$$

where the vector sum for the triple $\mathbf{h}+\mathbf{k}+\mathbf{l}=0$. The terms $A_{\mathbf{h}}$ and $\Phi_{\mathbf{h},\mathbf{k}}$ are *a priori* known values as defined in the previous paper and are not important in this exposition. The original set of $G_{\mathbf{h}}$ 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 $S_{\mathbf{h}}$ values which are initially estimated as

$$S_{\mathbf{h}}\simeq A_{\mathbf{h}}\langle G_{\mathbf{k}}G_{\mathbf{l}}\sin\Phi_{\mathbf{hk}}\rangle_{\mathbf{k}}/\langle A_{\mathbf{k}}A_{\mathbf{l}}\rangle_{\mathbf{k}}. \quad (3)$$

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 $G_{\mathbf{h}}$ 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 $G_{\mathbf{h}}$ and $S_{\mathbf{h}}$ terms for a molecular fragment p may be shown to be of the form

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

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

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 $\mathbf{h}+\mathbf{k}+\mathbf{l}=0$ vector-sum condition which pertains to the algebraic triples among the $G_{\mathbf{hkl}}$ and $S_{\mathbf{hkl}}$ terms applies only to those components linked to the translation vector contained in these sines and cosines. Thus the $\mathbf{h}+\mathbf{k}+\mathbf{l}=0$ 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 $G_{\mathbf{h}}$, $S_{\mathbf{h}}$ 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

$$\sum_i^n\sum_j^m X(h_1,k_i,l_1)Y(h_2,k_j,l_2)\cos/\sin\Phi_{\mathbf{h}_3,\mathbf{h}_1} \quad (6)$$

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

$$\sum X_{\mathbf{k}}Y_{\mathbf{l}}\cos\Phi_{\mathbf{h},\mathbf{k}}=\cos\varphi\varphi_{\mathbf{h}}(\sum X_{\mathbf{k}}\cos\varphi\varphi_{\mathbf{k}}\sum Y_{\mathbf{l}}\cos\varphi\varphi_{\mathbf{l}} \\ -\sum X_{\mathbf{k}}\sin\varphi\varphi_{\mathbf{k}}\sum Y_{\mathbf{l}}\sin\varphi\varphi_{\mathbf{l}}) \\ -\sin\varphi\varphi_{\mathbf{h}}(\sum X_{\mathbf{k}}\cos\varphi\varphi_{\mathbf{k}}\sum Y_{\mathbf{l}}\sin\varphi\varphi_{\mathbf{l}} \\ +\sum X_{\mathbf{k}}\sin\varphi\varphi_{\mathbf{k}}\sum Y_{\mathbf{l}}\cos\varphi\varphi_{\mathbf{l}}), \quad (7)$$

0108-7673/87/050733-02$01.50

$$\begin{aligned}\sum X_{\mathbf{k}} Y_{\mathbf{l}} \sin \Phi_{\mathbf{h},\mathbf{k}} = \cos \varphi\varphi_{\mathbf{h}}(&\textstyle\sum X_{\mathbf{k}} \cos\varphi\varphi_{\mathbf{k}} \sum Y_{\mathbf{l}} \sin \varphi\varphi_{\mathbf{l}} \\ &+\textstyle\sum X_{\mathbf{k}} \sin \varphi\varphi_{\mathbf{k}} \sum Y_{\mathbf{l}} \cos \varphi\varphi_{\mathbf{l}}) \\ &+\sin \varphi\varphi_{\mathbf{h}}(\textstyle\sum X_{\mathbf{k}} \cos \varphi\varphi_{\mathbf{k}} \sum Y_{\mathbf{l}} \cos \varphi\varphi_{\mathbf{l}} \\ &-\textstyle\sum X_{\mathbf{k}} \sin \varphi\varphi_{\mathbf{k}} \sum Y_{\mathbf{l}} \sin \varphi\varphi_{\mathbf{l}}), \qquad (8)\end{aligned}$$

where $\varphi\varphi hkl = (\varphi hkl_p + \varphi h\bar{k}l_p)$ and the sums include all terms having a common h and l but different k index. These summations can be performed prior to generating the triples and stored in two-dimensional arrays. The contribution of the full collection of valid triples to the formulae is thus reduced to a two-dimensional permutation over these arrays.

Degeneracies among the refined phases

Since the translation components of (4) and (5) are independent of the value of the k index with regard to $P2_1$ symmetry, (1) and (2) can be rewritten as

$$\begin{aligned}\cos 4\pi\mathbf{h}\,.\,\mathbf{r}_p = \langle((&G_{\mathbf{k}}G_{\mathbf{l}} - S_{\mathbf{k}}S_{\mathbf{l}}) \cos \varphi\varphi_{\mathbf{k},\mathbf{l}} \\ &-(G_{\mathbf{k}}S_{\mathbf{l}} + S_{\mathbf{k}}G_{\mathbf{l}}) \sin \varphi\varphi_{\mathbf{k},\mathbf{l}}\rangle_{\mathbf{k}}/\langle A_{\mathbf{k}}A_{\mathbf{l}}\rangle_{\mathbf{k}}, \qquad (9)\\ \sin 4\pi\mathbf{h}\,.\,\mathbf{r}_p = \langle((&G_{\mathbf{k}}G_{\mathbf{l}} - S_{\mathbf{k}}S_{\mathbf{l}}) \sin \varphi\varphi_{\mathbf{k},\mathbf{l}} \\ &+(G_{\mathbf{k}}S_{\mathbf{l}} + S_{\mathbf{k}}G_{\mathbf{l}}) \cos \varphi\varphi_{\mathbf{k},\mathbf{l}}\rangle_{\mathbf{k}}/\langle A_{\mathbf{k}}A_{\mathbf{l}}\rangle_{\mathbf{k}}, \qquad (10)\end{aligned}$$

where $\mathbf{h}.\mathbf{r}_p = hx_p + lz_p$, $\varphi\varphi_{\mathbf{k},\mathbf{l}} = \varphi\varphi_{\mathbf{k}} + \varphi\varphi_{\mathbf{l}}$, and the particular vectors **k** and **l** include all terms with a common h and l index. The refinement result on the right-hand side of the equations is independent of the particular k index of the vector **h** since $y_p = 0$. It follows that the calculations need not be performed over the full set of $G_{\mathbf{h}}$ and $S_{\mathbf{k}}$ values, but only over the first unique two-dimensional component of such vectors, as the full set with a common h and l index may readily be obtained by adding the known values of $\varphi\varphi_{\mathbf{h}}$ to the first computed value of $4\pi\mathbf{h}.\mathbf{r}_p$.

The simplified formula

With the following abbreviations to define the particular sums with a common h and l index,

$$B_{\mathbf{h}} = \sum_{k} (G_{\mathbf{hkl}} \cos \varphi\varphi_{\mathbf{hkl}} + S_{\mathbf{hkl}} \sin \varphi\varphi_{\mathbf{hkl}}), \qquad (11)$$

$$C_{\mathbf{h}} = \sum_{\mathbf{k}} (G_{\mathbf{hkl}} \sin \varphi\varphi_{\mathbf{hkl}} - S_{\mathbf{hkl}} \cos \varphi\varphi_{\mathbf{hkl}}), \qquad (12)$$

$$D_{\mathbf{h}} = \sum_{\mathbf{k}} 2|Ehkl_pEh\bar{k}l_p|, \qquad (13)$$

and with the use of the factoring scheme given by (7) and (8), equations (9) and (10) may be drastically simplified to

$$\cos 4\pi\mathbf{h}.\mathbf{r}_p = \langle B_{\mathbf{k}}B_{\mathbf{l}} - C_{\mathbf{k}}C_{\mathbf{l}}\rangle_{\mathbf{k}}/\langle D_{\mathbf{k}}D_{\mathbf{l}}\rangle_{\mathbf{k}}, \qquad (14)$$

$$\sin 4\pi\mathbf{h}.\mathbf{r}_p = \langle B_{\mathbf{k}}C_{\mathbf{l}} + C_{\mathbf{k}}B_{\mathbf{l}}\rangle_{\mathbf{k}}/\langle D_{\mathbf{k}}D_{\mathbf{l}}\rangle_{\mathbf{k}}. \qquad (15)$$

Comparison of results

Table 2 in the previous analysis summarized the refinement of the translation-function phases for various-sized misplaced fragments of the crystal structure of tetrahymanol hemihydrate (Langs, Duax, Carrell, Berman & Caspi, 1977), $2(C_{30}H_{52}O).H_2O$, monoclinic, $P2_1$. The degree of refinement convergence towards the known shift value was estimated as $\langle\cos 4\pi\mathbf{h}.[\mathbf{r}_p(\mathrm{obs}) - \mathbf{r}_p(\mathrm{cal})]\rangle$ given in the right-hand column of that Table 2. A fragment of 31 of the 63 non-hydrogen atoms was sufficiently large to permit a phase convergence to $\langle\cos \Delta\rangle \simeq 0{\cdot}99$ for three separate trials in which the basis sets comprised 759, 1499 and 2293 of the largest $G_{\mathbf{h}}$ values among the 5020 measured data. Each of these refined phase sets produced a translation map in which the peak height of the solution vector was ten times larger than the largest spurious peak.

Trials involving a smaller 14-atom fragment which represented 20% of the scattering power of the asymmetric unit were not as encouraging. A basis set of 1392 $G_{\mathbf{h}}$ values refined to $\langle\cos \Delta\rangle = 0{\cdot}373$; a large basis set of 1826 values showed a small improvement with $\langle\cos \Delta\rangle = 0{\cdot}427$ which represented an average phase error of 65° and produced a map in which the solution vector was as large as the two largest spurious peaks. Clearly there may have been a small advantage to further increasing the size of the basis set, but the expense of the calculations did not seem to warrant these efforts. Equations (14) and (15) now permit one to refine the full set of 5020 data in a fraction of the time previously required. This phase refinement converged to $\langle\cos \Delta\rangle = 0{\cdot}860$ and produced a translation map in which the solution vector was 13 times larger than the next spurious peak.

This work has been supported in part by NIH grant HL32303 from the National Heart, Lung and Blood Institute.

References

Langs, D. A. (1985). *Acta Cryst.* **A41**, 578-582.
Langs, D. A., Duax, W. L., Carrell, H. L., Berman, H & Caspi, E. (1977). *J. Org. Chem.* **42**, 2134-2137.

International Union of Crystallography

Acta Cryst. (1987). A**43**, 734

Commission on Crystallographic Apparatus

As a result of an expression of concern by the Commission on Journals with respect to the problem of the underestimation of standard deviations in the lattice constants reported in papers submitted to the journals published by the International Union of Crystallography, the Commission on Crystallographic Apparatus has decided to set up a project to assess the cause of the problem and to determine ways in which the problem may be solved.

This projecy (The Accuracy in Lattice Parameter Measurement Project) is to be organized by Professor Sagrario Martinez-Carrera of the Instituto de Quimica Physica 'Rocasolano'.

Any scientists who wish to participate in or to obtain further details about this project should contact Professor Sagrario Martinez-Carrera, Consejo Superior de Investigaciones Cientificas, Instituto de Quimica Physica 'Rocasolano', Serrano 119, Madrid, Spain.