estimates of p and the variance of n are all n/N and $n(1-n/N)$ respectively (Kendall & Stuart, 1977, § 8.8; 1979, § 19.29). If n and $N-n$ are both large no problem arises, but if one is small or zero these estimates are unreasonable. In particular, if $n=0$ the estimate that $p=0$ with perfect certainty (variance zero) is unacceptable. For example, if $N=100$, $p=0{\cdot}01$, the probability of observing $n=0$ is practically equal to the probability of observing the expected value $n=1$; both are 0·36.... If one treats the likelihood function as a probability distribution and calculates the *mean*-likelihood (instead of the *maximum*-likelihood) values, one obtains the more reasonable estimates

$$\langle p\rangle=(n+1)/(N+2) \qquad (A1)$$

(the Laplace 'rule of succession'), and

$$\sigma^2(n)=N[(n+1)/(N+2)][1-(n+2)/(N+3)]. \qquad (A2)$$

These lead to reasonable values for $n=0$:

$$\langle n\rangle=1/(N+2)\simeq N^{-1}, \qquad (A3)$$

$$\sigma^2(n)=N(N+1)/(N+2)(N+3)\simeq 1. \qquad (A4)$$

Various neo-Bayesian estimates (Good, 1965) give expressions similar to ($A1$) and ($A2$), with different numerical values for the 'corrections' to n and N.

Recalculation of many of the values of $N_{\rm sg}$ in Tables 4 and 6 with variance given by ($A2$) instead of the *GLIM* default value gave only minor changes, usually in the direction of better agreement. For most space groups in the monoclinic and orthorhombic systems $N_{\rm sg}$ is not small, and the choice of the expression for the variance may be more important for the remaining systems.

References

ALLEN, F. H., BELLARD, S., BRICE, M. D., CARTWRIGHT, B. A., DOUBLEDAY, A., HIGGS, H., HUMMELINK, T., HUMMELINK-PETERS, B. G., KENNARD, O., MOTHERWELL, W. D. S., RODGERS, J. R. & WATSON, D. G. (1979). *Acta Cryst.* **B35**, 2331-2339.

BAKER, R. J. & NELDER, J. A. (1978). *The GLIM System.* Release 3. Oxford: Numerical Algorithms Group.

DONOHUE, J. (1985). *Acta Cryst.* **A41**, 203-204.

GOOD, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods.* Cambridge, MA: MIT Press.

HAHN, TH. (1987). Editor. *International Tables for Crystallography*, Vol. A: *Space-Group Symmetry*, 2nd ed. Dordrecht: Reidel. (Present distributor Kluwer Academic Publishers, Dordrecht.)

HAZELL, A. C. & HAZELL, R. G. (1977). *Acta Cryst.* **B33**, 360-365.

KENDALL, M. S. & STUART, A. (1977). *The Advanced Theory of Statistics*, Vol. 1, 4th ed. London: Griffin.

KENDALL, M. S. & STUART, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, 4th ed. London: Griffin.

KITAJGORODSKIJ, A. I. (1955). *Organic Crystallochemistry.* Moscow: Izd. Akad. Nauk SSSR.

MACKAY, A. L. (1967). *Acta Cryst.* **22**, 329-330.

MIGHELL, A. D., HIMES, V. L. & RODGERS, J. R. (1983). *Acta Cryst.* **A39**, 737-740.

WILSON, A. J. C. (1987*a*). *Acta Cryst.* **A43**, C289.

WILSON, A. J. C. (1987*b*). *Br. J. Criminol.* **27**, 294-301.

WOLFF, P. M. DE (1987). *Acta Cryst.* **A43**, C305.

Acta Cryst. (1988). A44, 724-729

Patterson-Oriented Automatic Structure Determination: Getting a Good Start

BY FRANTIŠEK PAVELČÍK

Department of Analytical Chemistry, Faculty of Pharmacy, J. A. Komensky University, Odbojárov 10, *CS*-832 32 *Bratislava, Czechoslovakia*

(*Received* 9 *December* 1987; *accepted* 25 *April* 1988)

Abstract

On the basis of a generalized symmetry minimum function several computer-oriented methods for interpreting Patterson functions and for locating the position of heavy-atom fragments in crystals belonging to space groups of higher symmetry than $P1$ have been developed. The methods utilize cross vectors for finding relationships among the peaks of the symmetry minimum function. This approach has the advantage of suppressing false peaks of the symmetry minimum function, in locating more than one atom and in revealing the correct solution with greater probability. The heavy-atom fragment can be extended by superposition or Fourier methods. The methods are valid for all space groups, are simple to apply and form the basis for fully automated structure determination. In contrast to many other Patterson methods no *a priori* structural information is necessary. A few selected examples demonstrate the power of the new version of the computer program *XFPS*.

0108-7673/88/050724-06$03.00

Introduction

For direct methods the theory and computer applications have both been developed to a high degree of sophistication, in contrast to Patterson-oriented *ab initio* methods where the theory was developed in the fifties and early sixties and relevant computer programs were underdeveloped in the past. Only recently can some renaissance of Patterson methods be observed. The programs *SHELXS* (Sheldrick, 1985*a*), *XFPS* (Pavelčík, 1986), *IMPAS* (Luger & Fuchs, 1986), *HASSP* (Terwilliger, Kim & Eisenberg, 1987) have appeared.

For automated Patterson deconvolution by superposition methods reliable starting atoms are necessary. The symmetry minimum function, SMF (Simpson, Dobrott & Lipscomb, 1965), gives in general reliable starting atoms, particularly for higher-symmetry space groups in which the origin is fixed by space-group symmetry. For computer applications see Hubbard, Babich & Jacobson (1977) and Pavelčík (1986). But usually only one atom can be used because of origin and enantiomorph ambiguities. These problems are encountered in lower-symmetry space groups. For example, in $P\bar{1}$ with more than one heavy atom in the asymmetric unit, the double-weighted cross vectors are the source of the highest peaks of the SMF. In $P2$, $P2_1$ *etc.* every piece of information is stored in one section of the SMF and because only one peak can be picked out the resulting superposition or Fourier map does not distinguish between enantiomorphs. Similar problems are met in structures with heavy atoms in special or quasi-special positions.

Many early experiences with superposition methods show that knowledge of two symmetry-independent atoms improves the chances for the solution of the phase problem (Rabinowitz & Kraut, 1964; Alden, Stout, Kraut & High, 1964; High & Kraut, 1966; Kim, Jeffrey, Rosenstein & Corfield, 1967). In the last paper one can also trace the use of symmetry considerations among cross vectors for finding a pair of atoms. But until recently the use of cross vectors for computerized interpretation of the Patterson function was rather infrequent. Jacobson & Beckman (1979) turned attention to the use of overlapping cross vectors for the determination of positions of two atoms by the superposition method. More recently Luger & Fuchs (1986) have utilized symmetry considerations among cross vectors for the determination of Harker vectors. Because many interatomic vectors are used, the determination of a single Harker vector is more reliable than the determination from Harker regions alone.

In this paper several useful procedures are described for determining the positions of two or more atoms, which are based either on SMF peaks or are generalizations of the SMF. In all the procedures the cross vectors play a very important role. The main conceptual difference between the approach presented here and that of Sheldrick (1985*b*) or Luger & Fuchs (1986) is that a stored Patterson function is used instead of the peak list only. Whereas Sheldrick uses the Harker peaks to determine one atom and Patterson peaks in general positions to generate other atoms, here the concept of equivalent origins and consideration of both enantiomorphs is applied to peaks of the SMF. These new procedures are part of the fully automated Patterson interpretation computer program *XFPS* (version '88).

Procedures

(i) *Second-order symmetry minimum function, SOSMF*

The symmetry minimum function can be generalized to use the full space of the Patterson function (not only the Harker regions)

$$H(\mathbf{r})=\int_{V}\min_{s=1}^{S} P(\mathbf{R}_s\,.\,\mathbf{x}+\mathbf{t}_s-\mathbf{r})\,.\,d\mathbf{x} \qquad (1)$$

where the integration is over the unit cell. In cases where both $\mathbf{x}$ and $\mathbf{r}$ are true atomic positions the high values of the Patterson function recorded make important contributions to the final value of the integral. For the sampled Patterson function we obtain

$$H(\mathbf{r}_i)=\sum_{j=1}^{G}\min_{s=1}^{S} P(\mathbf{R}_s\,.\,\mathbf{x}_j+\mathbf{t}_s-\mathbf{r}_i) \qquad (2)$$

where the summation is over all grid points $\mathbf{x}_j$ of the Patterson function. The calculation of such a function is computationally expensive. A saving of computer time can be achieved if maxima on the SMF are fairly well resolved and only these maxima are considered as the $\mathbf{x}_j$ and $\mathbf{r}_i$. In the actual procedure all maxima higher than the atomic number of the lightest heavy atom (on an atomic number scale) of the asymmetric part of the SMF are tested for all equivalent origins and enantiomorphs. Let us first introduce a definition. The symmetry minimum interaction, SMI, for i and j atoms (or potential atoms) is defined as the minimal value of the Patterson function (by grid-point lookup) for vectors formed by atom i and all symmetry-equivalent positions of atom j. Because of the necessity for origin and enantiomorph specification the definition is

$$I^{oe}(\mathbf{r}_{ij})=\min_{s=1}^{S} P\{\mathbf{R}_s[q(e)\,.\,\mathbf{r}_j+\mathbf{t}_0]+\mathbf{t}_s-\mathbf{r}_i\}. \qquad (3)$$

For simple cases without specification of origin and enantiomorph the superscript will be omitted and for cases where it is necessary to consider a change of origin and enantiomorph for atom i also, the superscript will be extended to $I^{oepf}(\mathbf{r}_{ij})$. $q(e)$ defines the enantiomorph $q(1)=1$, $q(2)=-1$, $\mathbf{t}_0$ is an origin shift vector [*e.g.* $(0,0,0)$, $(\frac{1}{2},0,0),\ldots,(\frac{1}{2},\frac{1}{2},\frac{1}{2})$ for $P\bar{1}$]. The SMI is taken as zero if the value of the Patterson function is smaller than a preselected value [$I^{oe}(\mathbf{r}_{ij})=0$ for $P(\mathbf{r})<P_{\text{lim}}$]. $\mathbf{R}_s$ and $\mathbf{t}_s$ are the rotation and translation parts of the symmetry operator, respectively.

The symmetry minimum function of second order is defined as

$$Q^2(\mathbf{r}_i)=\sum_{\substack{j=1\\ j\neq i}}^{N}\max_{o=1}^{O}\max_{e=1}^{E} I^{oe}(\mathbf{r}_{ij}),\quad i=1,2,\ldots,N \qquad (4)$$

where N is the number of peaks used, $E=1$ for centro- and $E=2$ for non-centrosymmetric space groups, O is the number of equivalent origins (*e.g.* for $P\bar{1}$ $O=8$). For polar space groups there is a one-dimensional search for maxima in the polar direction over all grid points. For groups like *Pm* it is difficult to apply this procedure because of the necessity for two-dimensional search.

(ii) *Third-order symmetry minimum function, TOSMF*

The principles outlined in the previous section can be extended to a simultaneous search for three atoms (interaction triangle)

$$Q^3(\mathbf{r}_i)=\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{\substack{k=1\\ i\neq j,\ i\neq k,\ j\neq k}}^{N}\max_{o=1}^{O}\max_{e=1}^{E}\max_{p=1}^{O}\max_{f=1}^{E}[I^{eo}(\mathbf{r}_{ij})$$
$$+I^{fp}(\mathbf{r}_{ik})+I^{oefp}(\mathbf{r}_{kj})]. \qquad (5)$$

The sum in square brackets is accepted only if all SMI are greater than P_{lim}. This search is much more powerful in removing false peaks of SMF than $Q^2(\mathbf{r}_i)$ but also more time consuming [for SOSMF the time is proportional to $N(N-1)$ and for TOSMF to $N(N-1)(N-2)$]. A tandem approach is in hand. First SOSMF is calculated and then the number of peaks is reduced, say by a factor of two, and TOSMF is calculated. The TOSMF is effective only in cases with three and more heavy atoms in the asymmetric part of the unit cell.

(iii) *Convergence function, CF*

The convergence routine searches for the set of well cross-linked peaks of the SMF (or a similar function), without specifying their relative origin and enantiomorph. For N peaks of the SMF a table of maximal SMI (I^m_{ij}) is calculated,

$$I^m_{ij}=\max_{o=1}^{O}\max_{e=1}^{E} I^{oe}(\mathbf{r}_{ij})\quad (\text{fixed } P_{\text{lim}}). \qquad (6)$$

I^m_{ij} is taken as the maximal SMI among all possible positions of atom j (atom i is fixed). Because of P_{lim} some SMIs are zero. Each peak of the SMF is characterized by

$$\text{ZZ}_i=\sum_{j=1}^{N} I^m_{ij} \qquad (7)$$

and by the number of nonzero I^m_{ij} (n_i) in which atom i is involved. The peak with minimal value of ZZ_i is removed from the peak list (unless $n_i\geq N-2$ for $N\leq 6$), N is reduced to $N-1$ and n_i and ZZ_i are recalculated. The process is repeated until the well cross-linked set of peaks is found. This set can be used as a starting set for multiple Patterson solution by the cross-vector function.

(iv) *Cross-vector function, CVF*

The final deconvolution of the Patterson function means finding the relative positions of the peaks of the SMF (first, second or third order) or convergence function in respect of origin shift and enantiomorph and deletion of false peaks. The multi-solution approach developed here is based on cross vectors. The top list peaks are successively used as a pivot atom. The position of the pivot atom is fixed and for all other maxima of SMF all equivalent origins and enantiomorphs are considered and SMIs are found in the stored Patterson function. A new extended list of atoms (with shifted coordinates) for which the SMIs are nonzero is formed. One peak of SMF can in this way give several potential atomic positions. This is particularly true for non-centrosymmetric groups where both enantiomorphs are contained in the extended list and for the case with the pivot atom in a special (or quasi-special) position. For this new list a full table of SMIs is established (see Table 1). An atom with the highest sum of SMIs [$\text{ZZ}_i=\sum_j I(\mathbf{r}_{ij})$] is selected as the second atom of the solution. This is the most critical point of the algorithm. All atoms with $I_{2j}<P_{\text{lim}}$ are deleted from the atom list and also all atoms which have the same origin in the SMF as the second atom of the solution (the enantiomorph is usually fixed at this point). Then ZZ_i are recalculated and another atom is accepted for solution. The process is repeated until the atom list is exhausted. The solution for the given pivot atom is characterized by $\text{ZZM}=\sum_i\sum_j I_{ij}$. This criterion reflects both the heights of SMIs and the number of atoms in the fragment. The most probable solution is that with the highest ZZM.

(v) *Minimum translational function, MTF*

The higher-order SMF, convergence and cross vector functions are difficult to apply to space groups with an infinite number of equivalent origins in one or two dimensions. Also a $P\bar{1}$ structure with a single not too heavy atom (say Cl in an organic molecule) may not give the heavy atom high enough on the peak list to be easy identifiable. Cross vectors or double-

Table 1. *Table of SMIs for BAVO*

*P*1, *P*5 and *P*8 are the peaks of the SMF having high cross vectors with pivot atom *P*2. Z_i(Patt.) are their heights in atomic number units. P_{lim} for the V–V vector is 18 (on the scale used in the table). In the process of Patterson interpretation the *P*5 (no. 3) peak is selected as the second atom of the solution. *P*1 is eliminated and the enantiomorph is fixed by *P*8 because of the special coordinates of *P*5. Atomic types are assigned according to the final ZZ_i.

SMF number	*P*2 1	*P*1 2	*P*5 3	*P*8 4	*P*8 5	*P*8 6	*P*8 7	Starting ZZ_i	Final ZZ_i	Z_i (Patt.)	Type
1	—	129	125	45	45	25	25	394	170	51·4	Ba
2	129	—	0	4	4	4	4	145	0	53·6	—
3	125	0	—	19	19	19	19	201	144	30·6	V(1)
4	45	4	19	—	0	0	0	68	64	23·5	V(2)
5	45	4	19	0	—	0	0	68	0	23·5	—
6	25	4	19	0	0	—	0	48	0	23·5	—
7	25	4	19	0	0	0	—	48	0	23·5	—

weighted cross vectors can be used to resolve the problem. A translation function based on the symmetry minimum function was developed, which is properly weighted if atoms forming non-Harker vectors are not of the same type. Instead of searching for one atom in Harker regions, there is a simultaneous search for two (and eventually more) atoms which have strong Harker and cross vectors.

Let $\mathbf{r}_i^o$, $i=1,2,\ldots,n$ be relative coordinates of a correctly oriented fragment. Their correct position with respect to symmetry elements can be established with the help of a search vector $\mathbf{r}$: $\mathbf{r}_i=\mathbf{r}_i^o+\mathbf{r}$. The search function has the form

$$D(\mathbf{r})=\left\{\min_{k=1}^{n}\min_{l>k}^{n}\min_{i=1}^{S}\min_{j=1}^{S}\frac{Z_1Z_1}{m_{ij}Z_kZ_l}\times P[(R_i\mathbf{r}_k+\mathbf{t}_i)-(R_j\mathbf{r}_l+\mathbf{t}_j)]\right\}^{1/2}. \quad (8)$$

Z_1 is the atomic number of the heaviest atom, the multiplicity m_{ij} for self vectors is the same as the multiplicity in the SMF (Pavelčík, 1986); for the cross vectors $m_{ij}=1$ for non-centro- and $m_{ij}=2$ for centrosymmetric groups. Because of the symmetry of the Patterson function the equation can be simplified to

$$D(\mathbf{r})=\left[\min_{k=1}^{n}\min_{l\geq k}^{n}\min_{j=2}^{S}\frac{Z_1Z_1}{m_{ij}Z_kZ_l}\times P(\mathbf{r}-R_j\mathbf{r}+\mathbf{r}_k^o-R_j\mathbf{r}_l^o)\right]^{1/2}. \quad (9)$$

For the normalized Patterson function $[P(000)=\sum Z_i^2]$ the heaviest atom height in the $D(\mathbf{r})$ function is on the atomic number scale. Though the function was originally developed for low-symmetry space groups and one non-Harker vector as a generalization of the Jacobson & Beckman (1979) approach, it is clear that it is valid for all space groups and more than two atoms. The combined direct method/Patterson translation search of Egert & Sheldrick (1985) would be an efficient alternative to the MTF.

(vi) *A strategy for the automatic structure determination*

The strategy is decided on the basis of the unit-cell contents and symmetry. The calculation is started with a sharpened Patterson followed by the SMF (except for the group *P*1). The default sharpening is $(\frac{1}{6}+s^2)[F(000)/\sum f_i]^2$, $s=(\sin\theta)/\lambda$. For structures with one heavy atom in the symmetric unit and for structures in which two coordinates are not fixed by space-group symmetry (*e.g.* *Pc*) one heavy atom is selected on the basis of the $R(E)$ criterion from the top list peaks. By trial and error the functional form was settled at

$$R(E)=\left(\sum\left||E_0|^2-0{\cdot}65\Sigma_Q/\Sigma_N-|E_c|^2\right|\right)\Big/\left(\sum|E_o|^2\right) \quad (10)$$

where

$$\Sigma_Q=\sum_{i=1}^{Q}Z_i^2,\quad \Sigma_N=\sum_{i=1}^{N}Z_i^2$$

and Q is the number of unknown and N is the total number of atoms in the unit cell. All reflections with $E_o>0{\cdot}3$ are included. For the other structures the SOSMF and CVF are used for determination of the heavy-atom positions. The heavy atoms form the input for atomic minimum superposition. The CVF without equivalent origin shifts is used as the 'filtering' function if superposition was started only with one heavy atom in multiple-heavy-atom structures. This may resolve enantiomorph or pseudosymmetry problems. The procedure is finished with weighted $2F_o-F_c$ synthesis. In space group *P*1 a Patterson vector is used for the minimum superposition and the structure is completed by repeated Fourier syntheses.

If this simple strategy fails the strategy should be decided by the crystallographer, but the single job determination can also be used. An asymmetric part of the unit cell is selected automatically for Patterson, SMF, superposition or Fourier map for groups up to orthorhombic. For higher-symmetry groups the asymmetric part is derived from an appropriate orthorhombic or monoclinic subgroup.

The main features of the *XFPS* computer program are as recently described (Pavelčík, 1986), particularly for Fourier, Patterson and superposition methods. The program has been extended by the procedures described in this paper. Now there is no limit on the number of symmetry operations and the number of scattering-factor curves included in the program has been extended to 80. The program–user interface was improved. The program *XFPS* is available on request.

Test structures

The newly developed methods have been tested on ten known structures of different size and complexity. The most important results will be briefly discussed to illustrate the range of possible applications. Purely organic structures are so far resistant to the methods developed here.

BAVO

$P2_12_12_1$, $BaV_2O_8.H_2O$, $Z = 4$. The standard program test (Ulická, Pavelčík & Huml, 1987). The heavy atoms have quasi-special positions in the SMF: Ba (0·10, 0·02, 0·25), V(1) (0·00, 0·00, 0·25), V(2) (0·25, 0·33, 0·47). The highest peak of the SMF is false because of ambiguity in interpreting Harker vectors and the correct atoms have sequential numbers 2, 5, 8. The SOSMF gave the order as 1, 3, 4. TOSMF and CF left only the correct peaks. The CVF starting from any of the preselection functions led to the correct solution. The MTF based on the Ba–V vector clearly gave the Ba coordinates.

FUNG

Pccn, $C_{17}H_{22}ClN_2O_4PS_2$, $Z = 8$ (Vrábel, Pavelčík, Kellö, Miertuš, Konečný & Lokaj, 1987). The SMF gave the heavy atoms as peaks 1, 2, 28, 38. The SOSMF started from 45 peaks gave correct peaks as 2, 5, 6, 9 and TOSMF started from 22 peaks of SOSMF gave 1, 2, 3, 5. The result depends on the number of starting peaks. Another run with 60 peaks for SOSMF and 30 peaks for TOSMF gave the results 2, 4, 5, 7 and 1, 4, 7, 9, respectively. The CVF gave the correct solution. The results of the convergence function depend critically on the number of peaks used.

HAVE

$P2_1/c$, $C_7H_{15}N_2O_7P$, $Z = 4$ (Pavelčík & Havetta, 1988). The SMF has two solutions for the P atom because of its quasi-special position (0·23, 0·25, 0·01). The $R(E)$ factor test prefers either the correct or a false peak depending on the number of reflections and the formula used. It seems that success of the automatic procedure depends critically on the $R(E)$ factor criterion in complicated single-atom cases, and this aspect is open for new developments. The conditions and exact formula for calculation of the $R(E)$ factor refined on the basis of this structure were found suitable for other test structures. A minimum translational function with the non-Harker vector clearly gave the correct position for the P atom.

VULM

$P2_12_12_1$, $C_{18}H_{40}ClNO_2$, $Z = 4$ (Kettmann, Pavelčík & Rybár, 1988). The Cl atom was second on the SMF and the $R(E)$ factor was successful in revealing the correct position. There were no problems in the fully automated solution. The superposition based on the Cl atom was slightly more powerful than a weighted Fourier function in revealing light atoms. By hand interpretation 18 and 16 atoms respectively could be picked up from the 50 top peaks with the help of interpeak distances and angles. A degree of sharpening of the Patterson function optimized on this structure was incorporated as default sharpening in the program.

PYOX

$P\bar{1}$, $C_{14}H_{12}ClCuN_2O_6$, $Z = 2$ (Pavelčík, Žemlička, Kettmann & Krätsmár-Šmogrovič, 1987). Peak heights of the SMF and their order in the sorted peak list depend a little on the grid selection and degree of sharpening. This is important here because Cu–Cu and double-weighted Cu–Cl vectors are of the same strength. The correct peaks for the $CuCl_4$ fragment have sequential numbers 3, 32, 34, 60, 66 in the SMF but their coordinates are not very accurate because Cl–Cl vectors are near the background level. The SOSMF gives the order as 3, 7, 10, 32, 36 and TOSMF as 2, 3, 11, 23, 25. A peak (0, 0, 0) from the origin of the Patterson was still the first in the list. This demonstrates lower effectivity of SOSMF and TOSMF in this low-symmetry case. Experience showed that it may be advisable to delete the (0, 0, 0) peak from the list (unless one is expecting to have a heavy atom at the centre of symmetry). After removing this peak the SOSMF and TOSMF gave the correct atoms as 2, 10, 13, 27, 28 and 1, 5, 10, 15, 39. In the automatic structure determination the $R(E)$ factor was powerful enough to find the Cu atom among the highest peaks of SMF. A superposition based on the Cu atom clearly revealed all Cl atoms and subsequent Fourier synthesis showed the complete structure. In the automatic run the input to the program is as simple as this:

```
PYOX. AUTOMATIC.
HKLF
CELL 1·5418 8·807 10·657 11·628 73·43 74·62 68·55
UNIT 2, CU 1, CL 4, O 6, N 2 C 14 H 12
SOLV
```

The MTF based on the double-weighted Cu–Cl vector gave the correct position of the Cu atom and the trivial peak at $(\mathbf{r}_{\rm Cu}-\mathbf{r}_{\rm Cl})/2$ as top peaks.

References

ALDEN, R. A., STOUT, G. H., KRAUT, J. & HIGH, D. F. (1964). *Acta Cryst.* **17**, 109-121.

EGERT, E. & SHELDRICK, G. M. (1985). *Acta Cryst.* **A41**, 262-268.

HIGH, D. F. & KRAUT, J. (1966). *Acta Cryst.* **21**, 88-96.

HUBBARD, C. R., BABICH, M. W. & JACOBSON, R. A. (1977). *A PL/1 Program System for Generalized Patterson Superposition.* Ames Laboratory-ERDA. Iowa State Univ., Ames, USA.

JACOBSON, R. A. & BECKMAN, D. E. (1979). *Acta Cryst.* **A35**, 339-340.

KETTMANN, V., PAVELČÍK, F. & RYBÁR, A. (1987). In preparation.

KIM, S. H., JEFFREY, G. A., ROSENSTEIN, R. D. & CORFIELD, P. W. R. (1967). *Acta Crýst.* **22**, 733-743.

LUGER, P. & FUCHS, J. (1986). *Acta Cryst.* **A42**, 380-386.

PAVELČÍK, F. (1986). *J. Appl. Cryst.* **19**, 488-491.

PAVELČÍK, F. & HAVETTA, K. (1988). In preparation.

PAVELČÍK, F., ŽEMLIČKA, M., KETTMANN, V. & KRÄTSMÁR-ŠMOGROVIČ, J. (1987). *Chem. Pap.*, **41**, 433-440.

RABINOWITZ, I. N. & KRAUT, J. (1964). *Acta Cryst.* **17**, 159-168.

SHELDRICK, G. M. (1985*a*). In *Crystallographic Computing* 3, edited by G. M. SHELDRICK, C. KRUGER & R. GODDARD, pp. 175-189. Oxford Univ. Press.

SHELDRICK, G. M. (1985*b*). *J. Mol. Struct.* **130**, 9-16.

SIMPSON, P. G., DOBROTT, R. D. & LIPSCOMB, W. N. (1965). *Acta Cryst.* **18**, 169-179.

TERWILLIGER, T. C., KIM, S.-H. & EISENBERG, D. (1987). *Acta Cryst.* **A43**, 1-5.

ULICKÁ, L., PAVELČÍK, F. & HUML, K. (1987). *Acta Cryst.* **C43**, 2266-2268.

VRÁBEL, V., PAVELČÍK, F., KELLÖ, E., MIERTUŠ, S., KONEČNÝ, S. & LOKAJ, J. (1987). *Collect. Czech. Chem. Commun.* **52**, 692-706.

Acta Cryst. (1988). A**44**, 729-735

A Priori Estimation of Scale and Overall Anisotropic Temperature Factors from the Patterson Origin Peak

BY ROBERT H. BLESSING AND DAVID A. LANGS

Medical Foundation of Buffalo, 73 *High Street, Buffalo, New York* 14203, *USA*

(*Received* 1 *October* 1987; *accepted* 26 *April* 1988)

Abstract

An idea due to D. Rogers [*Computing Methods in Crystallography* (1965), edited by J. S. Rollett, pp. 117-148. Oxford: Pergamon Press] has been developed and implemented. The method is an advantageous alternative to Wilson plot or *K*-curve scaling of intensity data. On the relative experimental scale the structure factor can be written in matrix notation as $F(\mathbf{h})=k^{-1}\sum_j f_j(\mathbf{h})\exp(2\pi i\mathbf{h}^T\mathbf{x}_j)\exp(-\mathbf{h}^T\mathbf{b}_j\mathbf{h})$; and the squared structure-factor magnitude can be written as $|F(\mathbf{h})|^2=k^{-2}\exp(-2\mathbf{h}^T\mathbf{b}\mathbf{h})\{\sum_j f_j^2+2\sum_j\sum_{k>j}f_jf_k\exp[2\pi i\mathbf{h}^T(\mathbf{x}_j-\mathbf{x}_k)]\}$, if a common, or average, anisotropic temperature factor is factored out of the atomic summations. The f_j^2 summation corresponds to the Patterson origin peak, and the f_jf_k double summation to the off-origin Patterson peaks. A trivariate Gaussian density function, $P(\mathbf{u})-P_{\rm min}=p_0\exp(-\mathbf{u}^T\mathbf{p}\mathbf{u})$, is fitted by least squares to the origin peak from a Patterson synthesis with coefficients $|F|^2_{\rm meas}/\sum_j f_j^2$. Fourier inversion of the fitted Gaussian gives the scale and thermal parameters, $k^2=(\det\mathbf{p})^{1/2}/(\pi^{3/2}V_{\rm cell}P_0)$ and $\mathbf{b}=(\pi^2/2)\mathbf{p}^{-1}$. The fit of the parameter $P_{\rm min}$ is constrained by the condition that $P_{\rm min}=-F(000)^2/(k^2V_{\rm cell}\sum_j Z_j^2)$, and thus only p_0 and the six coefficients p_{ij} $(i<j=1,2,3)$ are independent parameters.

Background

One of the first steps in an X-ray crystal structure analysis is the estimation of the scale and overall temperature factor for the diffraction intensities. Over the years, several methods for doing this have been devised (Rogers, 1965, 1980; Levy, Thiessen & Brown, 1970; Ladd, 1978; Giacovazzo, 1980; Subramanian & Hall, 1982; Hall & Subramanian, 1982).

The Wilson plot

The widely employed method of Wilson (1942) takes advantage of the cosine form of the product of the structure factor with its complex conjugate (Patterson, 1935). In matrix notation,

$$F(\mathbf{h})=\sum_j f_j(s)\exp(-B_js^2)\exp(2\pi i\mathbf{h}^T\mathbf{x}_j), \quad (1)$$

0108-7673/88/050729-07$03.00