OVERHAUSER, A. W. (1971). *Phys. Rev. B*, **3**, 3173–3182.

PEREZ-MATO, J. M. & MADARIAGA, G. (1990). *Geometry and Thermodynamics*, edited by J.-C. TOLEDANO, pp. 405–415. New York: Plenum Press.

PETRICEK, V., MALY, K., COPPENS, P., BU, X., CISAROVA, I. & FROST-JENSEN, A. (1991). *Acta Cryst.* A**47**, 210–216.

WANG, H. H., BENO, M. A., CARLSON, K. D., THORUP, N., MURRAY, A., PORTER, L. C., WILLIAMS, J. M., MALY, K., BU, X., PETRICEK, V., CISAROVA, I., COPPENS, P., JUNG, D., WHANGBO, M.-H., SCHIRBER, J. E. & OVERMYER, D. L. (1991). *Chem. Mater.* **3**, 508–513.

YAMAMOTO, A. (1982). *Acta Cryst.* A**38**, 87–92.

Patterson-Oriented Automatic Structure Determination. Deconvolution Techniques in Space Group P1

BY F. PAVELČÍK

Department of Inorganic Chemistry, Faculty of Natural Sciences, Comenius University, 842 15 Bratislava, Slovakia

Abstract

Several methods for the automatic determination of heavy-atom structures have been developed and extensively tested. The methods are based on a combination of vector superposition in space group $P1$ with a symmetry minimum function. The peaks of the symmetry minimum function are used as trial origin shifts in the translational search for the cell origin. Three or two Patterson shift vectors, all belonging to a single image, can be obtained for vector superposition by a procedure called cross-vector superposition. The superposition map may be refined by an automatic Fourier recycling in space group $P1$ before the translational search is started.

Introduction

Ab initio Patterson deconvolution techniques can be divided into two main groups depending on the utilization of symmetry:

(i) The symmetry is used from the very beginning and some atomic positions are suggested from the analysis of Harker regions (Harker, 1936) or more automatically using the multiple implication function (Simpson, Dobrott & Lipscomb, 1965). For more sophisticated techniques also using symmetry-related cross vectors, see Borisov (1964), Kuz'min, Golovachev & Belov (1970), Luger & Fuchs (1986), Pavelčík (1988) and Pavelčík, Kuchta & Sivý (1992).

(ii) The Patterson function is deconvoluted by a (weighted) vector minimum superposition (Buerger, 1959; Jacobson & Guggenberger, 1966) in space group $P1$ based on a carefully selected single Patterson vector or on several vectors all belonging to

the same image. The symmetry is introduced after the structure has essentially been solved. Automatic structure determination based on vector minimum superposition in space group $P1$ has been reviewed by Simonov (1982). The general approach consists of several steps:

(1) Calculation of a (sharpened) Patterson function.

(2) Search for an atomic fragment using Patterson peaks or selection of a suitable Patterson peak for the superposition.

(3) Minimum-vector superposition.

(4) Inverse Fourier transform of modified minimum superposition map.

(5) Fourier recycling in space group $P1$.

(6) Search for a standard cell origin consistent with the space-group symmetry by some sort of translation function.

(7) Shift of the origin and electron-density averaging based on the symmetry.

(8) Fourier recycling using only peaks of the asymmetric part of the unit cell.

Even simplified algorithms based only on steps (1), (3) and (6) and the single vector superposition proved to be very successful in solving heavy-atom structures because of more sophisticated translation searching for multiple images (Richardson & Jacobson, 1987) or because of combination with the cross-vector table (Sheldrick, 1991; Sheldrick, Dauter, Wilson, Hope & Sieker, 1993).

The major sources of difficulty in the Simonov scheme are steps (2) and (6). The problem of selection of Patterson peaks may be overcome by a special superposition suggested by Iljukhin, Kuz'min & Belov (1981), which is called cross-vector super-

position. This idea is further developed in this paper and applied for the purpose of automatic structure determination.

The search for the cell origin consistent with space-group symmetry was solved by Simonov (1982), Richardson & Jacobson (1987) and Sheldrick (1991). In this paper, a new approach is suggested in which peaks of the symmetry minimum functions (SMF, or similar multiple implication function) are used as trial origin shifts. This reduces the number of tested trial shifts considerably and also uses information contained in Harker regions.

Various structure-determination schemes have been developed and tested on structures of different size, complexity and symmetry.

Theory and procedures

Cross-vector superposition

Let us select two Patterson vectors (peaks) $\mathbf{u}_1 = \mathbf{r}_{12}$ and $\mathbf{u}_2 = \mathbf{r}_{34}$. The cross-vector superposition can be formulated by

$$S(\mathbf{r}) = \min [P(\mathbf{r}), P(\mathbf{r} - \mathbf{u}_1), P(\mathbf{r} - \mathbf{u}_2), P(\mathbf{r} - \mathbf{u}_1 - \mathbf{u}_2)] \tag{1}$$

or

$$S(\mathbf{r}) = \min [P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12}), P(\mathbf{r} - \mathbf{r}_{34}), P(\mathbf{r} - \mathbf{r}_{12} - \mathbf{r}_{34})]$$

$$= \min \{\min [P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12})], \min [P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{34})],$$

$$\min [P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12} - \mathbf{r}_{34})]\}. \tag{2}$$

The idealized Patterson function in a point approximation (vector set) can be described with the help of the Dirac δ function,

$$P(\mathbf{r}) = \sum_i \sum_j Z_i Z_j \delta[\mathbf{r} - (\mathbf{r}_i - \mathbf{r}_j)] = \sum_i \sum_j Z_{ij} \delta(\mathbf{r} - \mathbf{r}_{ji}), \tag{3}$$

where $Z_{ij} = Z_{ji} = Z_i Z_j$, $\mathbf{r}_{ji} = \mathbf{r}_i - \mathbf{r}_j = -\mathbf{r}_{ij}$, $\mathbf{r}_i$ is an atomic positional vector and Z_i is an atomic number. The minimum of shifted vector sets, in the δ-function formulation, can easily be analysed. The first term of (2) is

$$\min [P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12})]$$

$$= \min \left[\sum_i \sum_j Z_{ij} \delta(\mathbf{r} - \mathbf{r}_{ji}), \sum_k \sum_m Z_{km} \delta(\mathbf{r} - \mathbf{r}_{mk} - \mathbf{r}_{12}) \right]$$

$$= \sum_k \min (Z_{1k}, Z_{k2}) \delta(\mathbf{r} - \mathbf{r}_{1k})$$

$$+ \sum_m \min (Z_{m2}, Z_{1m}) \delta(\mathbf{r} - \mathbf{r}_{m2}). \tag{4}$$

The first term of (2) is a sum of two images $\{\mathbf{r}_{1k} = \mathbf{r}_k - \mathbf{r}_1\}$ and $\{\mathbf{r}_{m2} = \mathbf{r}_2 - \mathbf{r}_m\}$, $k, m = 1,2,\ldots,N$, N being the number of atoms in the unit cell. Analogously, the second term leads to $\{\mathbf{r}_{3x}\} + \{\mathbf{r}_{x4}\}$ images. The

third term

$$\min [P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12} - \mathbf{r}_{34})]$$

$$= \min \left[\sum_i \sum_j Z_{ij} \delta(\mathbf{r} - \mathbf{r}_{ij}), \sum_k \sum_m Z_{km} \delta(\mathbf{r} - \mathbf{r}_{km} - \mathbf{r}_{12} - \mathbf{r}_{34}) \right] \tag{5}$$

and, because

$$\mathbf{r}_{12} + \mathbf{r}_{34} = \mathbf{r}_2 - \mathbf{r}_1 + \mathbf{r}_4 - \mathbf{r}_3 = \mathbf{r}_4 - \mathbf{r}_1 + \mathbf{r}_2 - \mathbf{r}_3 = \mathbf{r}_{14} + \mathbf{r}_{32}, \tag{6}$$

the nonzero values are obtained only for $km = 21$, 43, 41 and 23. The third term is reduced to four points $\mathbf{r}_{34}$, $\mathbf{r}_{12}$, $\mathbf{r}_{32}$ and $\mathbf{r}_{14}$. The minimum from all three terms gives only $\mathbf{r}_{14}$ and $\mathbf{r}_{32}$ vectors, which represent cross vectors between $\mathbf{r}_{12}$ and $\mathbf{r}_{34}$. Points belonging to these vectors are of equal weight (height) in the cross-vector superposition. The cross-vector superposition is shown in Fig. 1. Four atom fragments can be formed for subsequent superposition:

(a) $0 = \mathbf{r}_{11}$, $\mathbf{r}_{12}$, $\mathbf{r}_{14}$, $\mathbf{r}_{13} = \mathbf{r}_{14} - \mathbf{r}_{34}$;

(b) $0 = \mathbf{r}_{22}$, $\mathbf{r}_{21} = -\mathbf{r}_{12}$, $\mathbf{r}_{23} = -\mathbf{r}_{32}$, $\mathbf{r}_{24} = \mathbf{r}_{14} - \mathbf{r}_{12}$;

(c) $0 = \mathbf{r}_{33}$, $\mathbf{r}_{34}$, $\mathbf{r}_{32}$, $\mathbf{r}_{31} = \mathbf{r}_{32} - \mathbf{r}_{12}$;

(d) $0 = \mathbf{r}_{44}$, $\mathbf{r}_{41} = -\mathbf{r}_{14}$, $\mathbf{r}_{42} = \mathbf{r}_{32} - \mathbf{r}_{34}$, $\mathbf{r}_{43} = -\mathbf{r}_{34}$.

Although the images are geometrically equivalent, they generally have (if the atoms 1, 2, 3 and 4 do not have equal atomic numbers) different weights.

In practical realizations of this method, the following special cases must also to be considered because Patterson peaks are arbitrarily selected:

(a) the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ belong to different images;

(b) the vectors $\mathbf{u}_1$ and $-\mathbf{u}_2$ belong to different images;

(c) the vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ belong to the same image;

(d) the vectors $\mathbf{u}_1$ and $-\mathbf{u}_2$ belong to the same image;

(e) the vectors $\mathbf{u}_1$, $\mathbf{u}_2$ and $\mathbf{u}_1 + \mathbf{u}_2$ belong to the same image (often the case in centrosymmetric structures with two or more heavy atoms, important for space group $P\bar{1}$).

Cross-vector superposition calculated with a real Patterson function gives more than two (theoretical) peaks. The pairs of top peaks of the cross-vector function are used to form atomic fragments. Each possible solution is assigned a figure of merit, FOM $= H_i \Sigma P(\mathbf{u}_j)$, where H_i is a peak height in the cross-vector superposition, the $P(\mathbf{u}_j)$ are values of the Patterson function at vector positions belonging to the image selected (e.g. at $\mathbf{r}_{12}$, $\mathbf{r}_{13}$ and $\mathbf{r}_{14}$ for a four-atom fragment; $\mathbf{r}_{12}$, $\mathbf{r}_{13}$ and $\mathbf{r}_{23}$ for a three-atom fragment).

There is a formal restriction on the method. Both Patterson peaks in the cross-vector superposition

should be of the H–H type (H being the heavy or intermediate heavy atom). The method could be difficult to apply to, for example, space groups $P\bar{1}$ or $P2_1$ with only one heavy atom in the asymmetric unit and to space groups of higher order and with the single heavy atom in a special position.

The four- or three-atom superposition based on the fragment selected can be calculated, which, in an ideal case, could reveal a single image.

Deconvolution of the Patterson function by double (multiple) Patterson peak

This is a special case of the cross-vector super-position in which $r_{12} = r_{34}$. The cross-vector superposition is

$$S(\mathbf{r}) = \min\left[P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12}), P(\mathbf{r} - 2\mathbf{r}_{12})\right]$$

$$= \min\{\min\left[P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12})\right]$$

$$+ \min\left[P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{34})\right], \min\left[P(\mathbf{r}), P(\mathbf{r} - 2\mathbf{r}_{12})\right]$$

$$+ \min\left[P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12} - \mathbf{r}_{34})\right] + \min\left[P(\mathbf{r} - 2\mathbf{r}_{34})\right]\}.$$

$$(7)$$

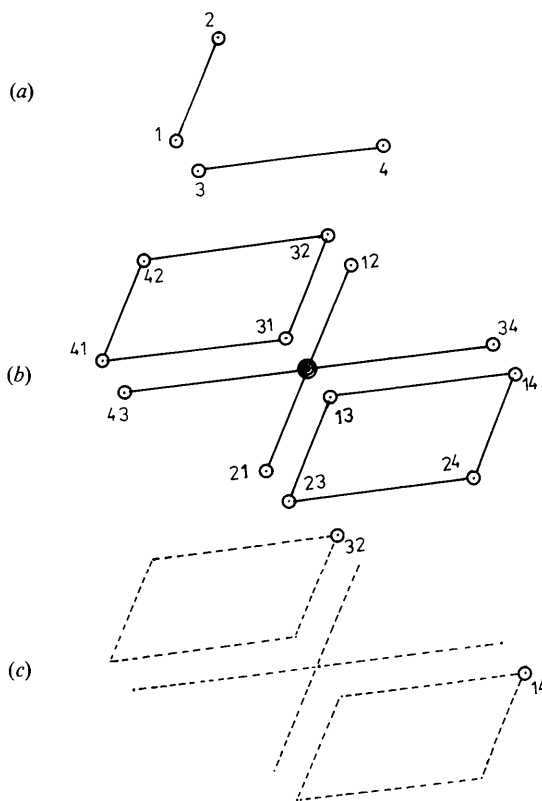The nonzero values can be found only for points r_{12}

Fig. 1. Cross-vector superposition. (a) Fundamental set consisting of four atoms. (b) Vector set. (c) Result of cross-vector superposition: $S(\mathbf{r}) = \min[P(\mathbf{r}), P(\mathbf{r} - \mathbf{r}_{12}), P(\mathbf{r} - \mathbf{r}_{34}), P(\mathbf{r} - \mathbf{r}_{12} - \mathbf{r}_{34})]$.

$= r_{34}$, r_{14} and r_{32}. All three points lie on a line. A four-atom parallelogram (centrosymmetric) can be formed, e.g. $r_{11} = 0$, $r_{12} = r_{34}$, $r_{13} = r_{14} - r_{34}$, r_{14}.

In the automated method, top peaks of the cross-vector function are used to form the four-atom parallelograms. The solution regarded as the best is the one with maximal value of $FOM = H\{P(\mathbf{r}_{13}) + P(\mathbf{r}_{14})\}$. This criterion is important if the overlapped vector is of different-atom type (e.g. Cu–Cl), otherwise second (and third) peaks of the cross-vector superposition can be used directly to form the vector parallelogram.

The overlapped Patterson peaks are usually at the top of the peak list and are then suitable for an automatic structure determination. Because of centrosymmetry of the fragment, the vector-superposition function based on this fragment is also centrosymmetric and is composed of two images of the structure (unless the centre of the parallelogram coincides with the centre of symmetry of the space group). Any peak of this superposition function can be selected and subsequent superposition theo-retically gives a single image.

Fourier recycling in the space group P1

The vector-superposition map generally contains many false peaks because of some degree of overlap of Patterson vectors and because peak coordinates are only approximate vector coordinates. Some false peaks belong to shifted images and some are the result of splitting of peaks due to inaccuracy. The superposition map may be improved by a repeated Fourier calculation in space group $P1$ (phase refinement by the electron-density-modification method). Instead of an inverse Fourier transform, a very simple peak approach [analogous to the Simonov (1975) method] was used. All peaks of the superposition map higher than a prescribed value are regarded as heavy atoms. The site occupation factors, g_i, are calculated from $g_i = H_i/H_1$, where H_i is the peak height. The total scattering power of the structure is

$$\sigma_N = \sum_{i=1}^{N} Z_i^2,$$

$$(8)$$

where N is the number of atoms in the unit cell and Z_i is an atomic number. The actual scattering power of n input peaks is

$$\sigma_P = \sum_{i=1}^{n} (Z_i g_i)^2.$$

$$(9)$$

If $\sigma_P > \sigma_N$, then the g_i are further modified:

$$g_i = (H_i/H_1)\{[\sigma_N - 36(N-n)]/\sigma_P\}^{1/2} \quad \text{if } N > n$$

or

$$(10)$$

$$g_i = (H_i/H_1)(\sigma_N/\sigma_P)^{1/2} \quad \text{if } N < n.$$

For the purpose of the translation search (see below), the coordinates of the peak nearest to the origin are reset to (0,0,0) and g_i is artificially increased by a factor of 1.33 in the first cycle (otherwise by a factor of 1.005) to ensure that the heavy atom is at the origin. The structure factors are calculated from input peaks and Fourier synthesis is calculated as a weighted $2F_o - F_c$ synthesis. The process is repeated until convergence is reached. In each third cycle, the atomic types are assigned to Fourier peaks on the basis of sequence numbers in the ordered peak list and the known contents of the unit cell.

Search for the standard cell origin

The vector-superposition function (based on a single Patterson vector, vector parallelogram, four-atom image) or electron-density function obtained by Fourier recycling in space group $P1$ is calculated in such a way that some heavy atom is at the origin of the unit cell. The search for the standard cell origin can be based on this atom. Peaks of the symmetry minimum function (or similar multiple implication function) are used for generating trial origin shifts. The trial origin shifts are

$$\mathbf{r}_o = \mathbf{R}_s \mathbf{r}_{SMF} + \mathbf{t}_s \tag{11}$$

because each atom of the image of symmetry-related atoms can be at the origin. Each trial origin shift is checked for the presence of all symmetry-related peaks in the superposition or Fourier-refined superposition map. Their coordinates are

$$\mathbf{x}_i = s(\mathbf{R}_i \mathbf{r}_o + \mathbf{t}_i - \mathbf{r}_o), \tag{12}$$

where s is $+1$ for all space groups except 11 enantiomorphous pairs where both $+1$ and -1 have to be checked. The shift is accepted only if all the $\mathbf{x}_i$s, related to the atom residing at the origin, are found among peaks of the superposition map within a specified distance limit. Each accepted origin shift is characterized by criteria based on the presence of as many as possible symmetry-equivalent peaks found in the shifted superposition map.

$$FOM1 = \sum_s \prod H(\mathbf{x}_{is})$$

$$FOM2 = \sum_s \min H(\mathbf{x}_{is})$$

$$FOM3 = \sum_s \sum H(\mathbf{x}_{is}) \tag{13}$$

$$FOM4 = \sum_s \sum 0.16/[0.16 + 4d(\mathbf{x}_{is})^2]$$

$$FOM5 = n.$$

$H(\mathbf{x})$ is the peak height of the superposition peak found within a specified distance of the calculated peak position of a symmetry-equivalent peak, $d(\mathbf{x})$ is the difference between the calculated and actual peak

positions. The calculated position is

$$\mathbf{x}_{is} = \mathbf{R}_s(\mathbf{x}_i + \mathbf{r}_o) + \mathbf{t}_s - \mathbf{r}_o. \tag{14}$$

The sum, product or minimum is added to the FOM only if all $\mathbf{x}_{is}$ peaks are found in the superposition map. The total number of these contributions is n. The accepted origin shifts are ordered according to an empirical combined figure of merit

$$FOM = FOM1 + 0.2(FOM2 + 0.5\,FOM3). \tag{15}$$

The vector-superposition function based on one vector and the subsequent search for the cell origin is a very discriminating and selective 'figure of merit' for peaks of the symmetry minimum function (SMF). It is selective in the sense that only SMF peaks belonging to atoms forming the superposition cross vector can be found as the shift vectors in the search for the cell origin; if a Harker vector is used as the superposition vector only one SMF peak can be found.

Multisolution

The origin shifts found in the translation search having the highest FOMs are used in the multisolution. The criterion that selects the final solution can be either a conventional R factor (on F) or a cross-vector function (CVF) (Pavelčík, 1988). In each particular tested solution, the origin shift is applied to all peaks of the superposition map. Peaks that do not have all symmetry equivalents are deleted from the peak list and from all the symmetry-equivalent peaks only one is retained. The selected peaks go to either CVF or R-factor calculation. The best cross-linked set of heavy atoms from the CVF is used in the atomic minimum superposition. The set of atoms with minimum R factor is used in the calculation of the electron density. Because it is difficult to assign the atomic types to the superposition peaks at this stage of structure determination, all atoms are assigned the atomic type of the heaviest atom and site occupation factors, g_i, are calculated by $g_i = H_i/H_1$.

Automatic structure determination

Automatic structure determination is started with a calculation of the sharpened Patterson map and peak picking. The Patterson peaks are ordered according to their heights. Peak no. 1 is the origin peak. In the fully automated procedure, peak no. 2 is used in vector superposition or peaks nos. 2 and 3 are used in the cross-vector superposition. If these procedures fail to give the correct solution, the Patterson peaks have to be selected manually (if there is not strong chemical evidence for an H–H peak, the simplest way is to try all top-list peaks) and inserted into the automatic procedure. Several deconvolution

schemes for the automatic structure determination have been developed.

Scheme A.

Selection of the Patterson peak (automatically or manually).

Minimum vector superposition based on this Patterson peak (in space group $P1$).

SMF.

Search for the cell origin and ordering trial origin shifts on the basis of the FOM. Only shifts with FOMs greater than 0.50 of the best FOM are considered in the multisolution.

Multisolution. (*a*) CVF multisolution for a maximum of three origin shifts with greatest FOM by cross-vector function followed by the atomic minimum superposition (Pavelčík, 1986). (*b*) R-factor multisolution for the maximum of three origin shifts. The set of atoms with the best R factor is used for electron-density calculation.

Two Fourier cycles.

Scheme B.

Selection of two Patterson peaks.

Cross-vector superposition.

Determination of atomic fragment consisting of four or three atoms all belonging to one structure image.

Minimum vector superposition based on this fragment (structure-determination superposition).

SMF.

Search for the cell origin. Only shifts with FOMs greater than 0.75 of the best FOM are considered in the multisolution.

Multisolution. (*a*) CVF multisolution as in strategy *A*. (*b*) R-factor multisolution as in strategy *A*.

Two Fourier cycles.

Scheme C.

Selection of the overlapped (double or multiple) Patterson peak.

Cross-vector superposition.

Selection of the vector parallelogram.

Minimum-vector superposition based on this parallelogram.

SMF.

Search for the cell origin. Only shifts with FOMs greater than 0.50 of the best FOM are considered in the multisolution.

Multisolution. (*a*) CVF multisolution. (*b*) R-factor multisolution.

Two Fourier cycles.

Scheme D. Like scheme *B*. The structure-determination superposition is followed by Fourier recycling in space group $P1$ until the structure is solved (this step may eventually be followed by the SMF and the search for the cell origin).

The number of SMF peaks was limited to 40, but only peaks with H_i greater than $0.8Z_H$ were used, where Z_H is the atomic number of the heaviest atom.

The result of the electron-density calculation is shown on the computer screen (a graphic procedure was adapted from *XFPS* (Pavelčík, Sivý, Rizzoli & Andreetti, 1992). If only a partial solution can be found, the false peaks are deleted in a graphic mode and the next Fourier synthesis of the electron density is calculated. The process can be repeated several times until the whole structure appears on the screen.

Test calculations and discussion

The test structures are given in Table 1. The results of calculations for scheme *A* are given in Table 2. If the second Patterson peak (automatic run) had not led to structure determination, the third peak (inserted in the input data) was tried and so on until the structure was solved. Only the first successful or partially successful solution is usually given in Table 2 (*e.g.* 'FUNG 6' means that Patterson peaks 2–5 failed to solve the structure); more information is presented only for complicated or interesting structures. The results of calculations for scheme *B* are given in Table 3. In this scheme, various combinations of peaks were tried, starting from low peak numbers (peaks 2 and 3 are defaults in an automatic run). In the structure APAPA, only experimental peak positions related to known vector positions (calculated from atomic coordinates) were used for testing purposes.

The results show that both schemes are very powerful for the solution of simple structures with dominating *H–H* vectors. In complicated structures, the correct shift sometimes did not have the best FOM or multisolution criteria (R factor or CVF) failed to select the correct shift. The CVF criterion is in general preferable to the R factor. More elaborate combined FOMs may further increase the power of these methods. This will be considered in the development of a user-oriented computer program. In some structures, Patterson peaks with the lowest sequence numbers are not of the *H–H* type (*e.g.* VULM peak no. 2, AZET peak no. 3, MGHEX peaks nos. 2–6).

Scheme *A* sometimes failed to give the solution if the superposition vector was highly overlapped. This was the case for, for example, FUNG with *H–H* vectors 3 and 4, MORF 3 and 4, BETA 2, AZET 2, STR 2 and 3, APAPA 2, 6 and 8.

Scheme *B* gave slightly more reliable results than scheme *A*. In most cases, cross-vector superposition with *H–H* Patterson peaks nos. 2 and 3 led to solution. The exception is STR. The correct shift vector in the translation search is usually the shift with the highest FOM. Also, the number of accepted

Table 1. *Test structures for Patterson methods*

Code	Formula	Z	Space group	Reference
BAVO	$Ba(VO_3)_2.H_2O$	4	$P2_12_12_1$	(1)
PYOX	$C_{14}H_{12}Cl_4CuN_2O_6$	2	$P\bar{1}$	(1)
FUNG	$C_{17}H_{22}ClN_2O_4PS_2$	8	$Pccn$	(1)
VULM	$C_{18}H_{30}ClNO_2$	4	$P2_12_12_1$	(1)
HAVE	$C_7H_{15}N_2O_7P$	4	$P2_1/c$	(1)
KEID	$C_{14}H_{24}CoKN_4O_{12}$	2	$P\bar{1}$	(1)
CUSA	$C_{17}H_{13}CuKN_2O_3S$	2	$P\bar{1}$	(1)
MORF	$C_7H_{11}Cl_3N_2O_2$	4	$P2_1/c$	(1)
BETA	$C_{18}H_{29}ClN_2O_5$	8	$C2/c$	(2)
CUMOS	$C_{18}H_{33}CuMo_{0.5}NS_3$	8	$C2/c$	(1)
AZET	$C_{42}H_{32}Cl_2N_2O_2$	4	$Pca2_1$	(1)
BOBBY	$C_6H_6CaNNaO_6$	4	$P2_13$	(1)
APAPA	$C_{30}H_{49}N_{15}O_{22}P_2$	8	$P4_12_12$	(1)
MGHEX	$C_{56}H_{80}Cl_2MgN_{16}O_{20}$	3	$P3_1$	(1)
SELENID	$C_{22}H_{28}O_2Se$	2	$P2_1$	(1)
CUIM	$C_{13}H_{18}CuN_3O_3$	4	$P2_1/c$	(1)
DAMO	$C_{15}H_{50}B_4Cu_2F_{16}N_{10}$	8	$Pbca$	(1)
HDAV	$C_3H_9NO_3V$	8	$Pbca$	(3)
MCO	Cr_2MnO_4	8	$Fd\bar{3}m$	(1)
ES67A	$C_{13}H_{19}Cl_6Ta$	2	$P2_1/c$	(1)
FC1014	$C_{44}H_{52}O_5W$	4	$I422$	(1)
FC999	$C_{95}H_{114}O_{10}W$	2	$P4/n$	(1)
FC969	$C_{32}H_{28}O_8W$	4	$C2/c$	(1)
STR	$C_{30}H_{13}Co_2FeMnO_{11}$	4	$P2_1/c$	(1)
PV206	$C_{38}H_{34}N_2Zr$	8	$P2_1/n$	(1)
PV214	$C_{24}H_{22}ClNOZr$	2	$P\bar{1}$	(1)
PV213	$C_{34}H_{50}KN_2O_8$	2	$P\bar{1}$	(1)
JR171	$C_{57}H_{63}N_4O_4V$	2	$P\bar{1}$	(1)
PV220	$C_{29}H_{22}ClCrO_6PZr$	4	$P2_1/c$	(1)
PV216	$C_{29}H_{22}ClCrNO_6Zr$	4	Cc	(1)
WL480	$C_{58}H_{78}ClO_4Ti_2$	2	$P\bar{1}$	(1)
FC955	$C_{62}H_{70}Cl_2O_4W$	4	$P2_1/n$	(1)
ZAGA11	$C_{28}H_{38}Cl_4N_2O_6P_2Pd$	1	$P\bar{1}$	(1)
SGPY	$C_{17}H_{16}CuN_2O_5$	4	$P2_1/n$	(4)
NTAB	$C_8H_{12}BaNO_{12}V$	4	$Pna2_1$	(5)
SOLD	$C_{12}H_{17}CuNO_8$	2	$P2_1$	(6)
MEXI	$C_{11}H_{18}ClNO$	4	$P\bar{1}$	(7)
DATTA5	$C_{11}H_{11}ClN_4$	4	$P\bar{1}$	(8)
PIPE	$C_8H_{31}N_4O_{12}V_3$	8	$Pbca$	(9)
BL238	$C_{44}H_{60}N_8Na_4O_4$	4	$P\bar{1}$	(10)

References: (1) Pavelčík, Kuchta & Sivý (1992); (2) Kettmann (1990); (3) Tyršelová & Pavelčík (1992); (4) Soldánová, Tyršelová & Pavelčík (1993); (5) Kuchta, Sivák & Pavelčík (1993); (6) Krätsmár-Šmogrovič, Pavelčík, Soldánová, Sivý, Seressová & Žemlička (1991); (7) Sivý (1992); (8) Bocelli (1990); (9) Tyršelová, Pavelčík & Kuchta (1993); (10) Pavelčík (1992).

shift vectors was in general smaller in comparison with scheme A (not documented in the tables). Some structures (SELENID, JR171, SOLD) with only one H–H vector in the asymmetric unit of the Patterson function have also been solved by scheme B.

Experience showed that schemes A and B are more reliable if the Patterson peaks do not contain special or quasispecial coordinates like (0,0.5,0.25) *etc*. It is recommended that two vectors be selected for scheme B in such a way that there is at least one nonspecial coordinate in both peaks [*e.g.* one vector of the type $(x,0,z)$ and a second one of $(0,y,0.5)$].

Scheme C also showed itself to be fairly powerful in solving crystal structures (Table 4). The main problem is to select a peak consisting of two overlapping H–H vectors. For this reason, only centrosymmetric structures with two or more heavy atoms were tested and a twice overlapped cross vector was

Table 2. *Results of test calculations for scheme A*

PV Patterson peak number used in the Patterson deconvolution. CS sequence number of sorted FOM belonging to correct shift vector(s). CVF multisolution by cross-vector function. RF multisolution by R-factor calculation. S shift vector selected by CVF or R-factor multisolution. R R-factor resulting from automatic structure determination. c number of extra graphic Fourier cycles needed to finish structure determination. A fully automated structure determination, otherwise Patterson peaks have been specified on input. N no solution found.

Code	PV	CS	CVF			RF		
			S	R	c	S	R	c
BAVO	2A	1	1	0.13	0	2	N	–
PYOX	2A	1	1	0.31	0	1	0.34	0
FUNG	6	1,3	1	0.37	1	3	0.33	0
VULM	3	2	1	N	–	2	0.33	0
VULM	4	1	1	0.35	0	1	0.37	2
HAVE	2A	1,3	1	0.56	0	3	0.45	0
KEID	2A	2	1	N	–	2	0.40	0
CUSA	2A	1	1	0.23	0	1	0.22	0
MORF	2A	3	3	0.33	0	2	N	–
MORF	5	2	2	0.31	0	2	0.33	0
BETA	3	2	1	0.55*	–	2	0.44	0
CUMOS	2A	3	3	0.23	0	3	0.20	0
AZET	2A	1	1	N	–	1	0.52	2
AZET	6	3	1	N	–	3	0.43	1
BOBBY	2A	1	1	0.41	1	1	0.25	0
APAPA	16	1	1	0.49	1	1	0.53	3
APAPA	17	1	1	0.49	1	1	0.50	1
MGHEX	7	2	2	0.45	0	2	0.45	0
SELENID	2A	1	1	0.29	0	1	0.29	0
CUIM	2A	1	1	0.29	0	1	0.31	0
DAMO	2A	1,3	1	0.33	0	3	0.35	1
HDAV	2A	1	1	0.48	0	1	0.42	0
MCO	2A	1	1	0.33	0	1	0.19	0
ES67A	2A	1	1	0.30	1	1	0.27	2
FC1014	2A	1	1	0.26	1	1	0.21	1
FC999	2A	1	1	0.24	0	1	0.23	0
FC969	2A	1	1	0.27	0	1	0.26	0
STR	4	1-4	1	0.29	0	3	0.28	0
PV206	2A	1	1	0.32	0	2	N	–
PV214	2A	1	1	0.26	0	1	0.25	0
PV213	2A	1	1	0.36	0	1	0.36	0
JR171	2A	1	1	0.32	0	1	0.32	0
PV220	2A	1	1	0.27	0	1	0.27	0
PV216	2A	1	1	0.20	0	1	0.36	2
WL480	2A	2,3	2	0.36	0	1	N	–
FC955	2A	2	2	0.28	0	1	N	–
ZAGA11	2A	2	2	0.25	0	2	0.25	0
SGPY	2A	1	1	0.29	0	1	0.31	0
NTAB	2A	1	1	0.24	0	1	0.19	0
SOLD	2A	1	1	0.33	2	1	0.29	1
MEXI	2A	2	2	0.28	0	1	N	–
PIPE	2A	1	1	0.25	0	1	0.45	0

* The whole structure was found but it was shifted by the vector (0.0,0.0,0.08).

used for Patterson deconvolution by cross-vector superposition. The method is less powerful if the H–H atoms are of different atomic types like Cu–Cl in PYOX or Cu–K in CUSA, because peak heights in the cross-vector minimum superposition are given by the self vectors Cl–Cl or K–K, which are near to background level. Scheme C could be very useful for automatic structure determination of structures with two or more equal heavy atoms in space group $P\bar{1}$ (WL480, MEXI, DATTA5). The cross vectors dominate on the peak list, whereas the single-weighted H–H vector is more difficult to recognize.

Table 3. *Results of test calculations for scheme B*

For symbols see Table 2.

Code	PV	CS	CVF			RF		
			S	R	c	S	R	c
BAVO	2,3A	1	1	0.13	0	1	0.15	0
PYOX	2,3A	1	1	0.33	0	1	0.31	0
FUNG	2,3A	1	1	0.50	1	1	0.41	0
FUNG	3,4	1	1	0.41	0	1	0.29	0
VULM	3,4	1	1	0.35	0	1	0.32	0
HAVE	2,3A	1	1	0.56	0	1	0.35	0
CUSA	2,3A	1	1	0.22	0	1	0.22	0
MORF	2,3A	1	1	0.32	0	1	0.33	0
BETA	2,3A	1	1	0.53	0	2	0.45*	–
CUMOS	2,3A	2	2	0.23	0	2	0.20	0
AZET	2,6	1	1	0.43	1	1	0.43	1
BOBBY	2,3A	1	1	0.41	1	1	0.25	0
APAPA	11,17	1	1	0.49	1	1	0.47	0
SELENID	2,3A	1	1	0.29	0	1	0.28	0
CUIM	2,3A	1	1	0.29	0	1	0.31	0
DAMO	2,3A	1,2	2	0.38	1	2	0.32	0
DAMO	2,4	1	1	0.38	0	1	0.31	0
HDAV	2,3A	1	1	0.48	0	1	0.44	0
MCO	2,3A	1-4	1	0.30	0	2	0.30	0
ES67A	2,3A	1	1	0.30	1	1	0.39	1
STR	2,4	1	1	0.32	0	1	0.31	0
PV206	2,3A	1,2	1	0.32	0	2	0.32	0
JR171	2,3A	1	1	0.32	0	1	0.29	0
PV220	2,3A	1	1	0.27	0	1	0.27	0
PV216	2,3A	1	1	0.23	0	1	0.27	1
WL480	2,3A	1,2	2	0.36	0	1	0.35	0
FC955	2,3A	1	1	0.47	2	1	0.36	1
ZAGA11	2,3A	1	1	0.24	0	1	0.26	0
SGPY	2,3A	1	1	0.29	0	1	0.31	0
NTAB	2,3A	1	1	0.24	0	1	0.16	0
SOLD	2,3A	1	1	0.33	2	1	0.25	0
MEXI	2,3A	1	1	0.29	0	1	0.27	0
PIPE	2,3A	1,2	1	0.26	0	2	0.36	0

* The whole structure was found but it was shifted by the vector (0.0,0.0,0.08).

Scheme D was tested on difficult structures and on several more-or-less arbitrarily selected less-complicated structures for comparison (Table 5). The phase refinement by Fourier recycling in space group $P1$ is a rather slow process, so in many cases it is stable. Nevertheless, it may be regarded as the last choice for otherwise unsolvable structures or for structures with uncertain space group or pseudosymmetry. The structures of DATTA5 and BL238 were solved by this method. In complicated structures, the success of the automated recycling may depend on the fine details of the algorithm. The variables are, for example, the number of input atoms in individual cycles, the method of calculation of site-occupation factors, degree of preference for the origin peak, type of Fourier synthesis ($2F_o - F_c$, F_o, E_o, weighted, unweighted). More research is needed to settle these problems.

The problem structures are now commented on in more detail.

DATTA5. The A, B, C strategies failed to find the correct solution because the SMF peaks belonging to Cl atoms were not among the 40 peaks used for generation of trial origin shift vectors. It seems that peaks preselected by, for example, the second-order

Table 4. *Results of test calculations for scheme C*

For symbols see Table 2.

Code	PV	CS	CVF			RF		
			S	R	c	S	R	c
WL480	2A	1	1	0.35	0	1	0.34	0
MEXI	2A	1	1	0.29	0	1	0.31	0
PYOX	3	3	3	0.32	0	3	0.30	0
PYOX	4	1	1	0.34	0	1	0.31	0
CUMOS	2A	2	2	0.23	0	2	0.21	0
MORF	10	1	1	0.31	0	1	0.31	0
DAMO	12	1-3	3	0.47	2	3	0.42	1
STR	4	1	1	0.28	0	1	0.31	0
CUSA	2A	1	1	0.22	0	1	0.21	0
PV220	4	1	1	0.27	0	1	0.26	0
PV206	2A	1	1	0.55	2	2	N	–
PV206	7	1	1	0.32	0	2	N	–
PV214	3	1	2	N	–	1	0.25	0

Table 5. *Results of test calculations for scheme D*

G represents extra graphics cycles.

Code	PV	R factors (%)
BAVO	2,3A	49,35,25,18,16,15,13,12,10
CUMOS	2,3A	52,36,29,25,22,18,22,19,13
DATTA5	2,4	54,36,41,35,26,27,29,25,24
VULM	3,4	58,38,39,31,30,32,29,28,27, G:28
PYOX	2,3A	52,39,31,21,18,18,19,17,17
BETA	2,3A	57,44,39,34,33,28,25,23,19
FUNG	3,4	47,55,39,44,49,34,40,45,32,39,42,29,36,35,24,23,24,18
STR	3,4	47,36,30,33,31,25,32,29,24, G:23,22,21,23,21
AZET	6,11	71,54,51,40,35,41,35,34,38,35,34,36,34,34,34, G:33,33,21
BL238	3,5	54,55,43,50,55,40,47,53,38,47,52,37,47,52,36,46,51,35,44, 49,34,42,47,33,42,47,33,42,45,32,37,36,25,34,31,20,30, 26,18

symmetry minimum function (Pavelčík, 1988) would be useful and this case will be considered in future. The cross-vector superposition and Fourier recycling easily gave the solution.

BL238. This is a structure in which the heavy atom is Na$^+$. It was supposed that there are clusters of atoms in the shape of a cube. Three multiple over-lapped Na–O vectors forming the edges of the cube were easily recognizable in the Patterson function (because of length 2.3 Å and approximate right angles) as well as Patterson vectors belonging to the square and body diagonals of the cube. Two overlapped Na–O vectors were selected and the cross-vector superposition gave a four-atom fragment. Neither CVF nor R-factor multisolution were able to give the solution. The information from the superposition map based on this small fragment was refined by Fourier recycling and all 232 atoms could be seen on the computer screen (divided into eight fragments with central Na$_4$O$_4$ cubes). There is also strong translational pseudosymmetry relating symmetry-independent cubes. The second Patterson vector (0.5,0.0,0.5) has a height 56% of that of the origin peak.

STR. The failure of scheme B for fully automated structure determination using Patterson peaks nos. 2 and 3 was caused by the fact that both Patterson

peaks are overlapped and their coordinates are (0.51,0.50,0.75) and (0.0,0.0,0.5). Also, the first two peaks of the SMF are false. The correct shift vector was the second one but the FOM was outside the limit specified (0.75 of the best FOM). The structure was solved using Patterson peaks nos. 2 and 4.

DAMO. The coordinates of two copper atoms Cu1 (0.10,0.24,0.94) and Cu2 (0.07,0.22,0.28) have similar x and y coordinates. The SMF gave these atoms at (0.09,0.25,0.06) and (0.09,0.25,0.19) with the same y coordinate. Patterson vectors nos. 2 and 3 with coordinates (0.0,0.0,0.50) and (0.5,0.0,0.38) used in the cross-vector superposition have undefined y coordinates. The delicate balance in the inaccurate y SMF coordinate gave in some preliminary test calculations inaccurate coordinates of Cu atoms in the CVF and subsequent atomic superposition interchanged the y coordinates. Two Fourier syntheses were not able to correct for this error. The atomic minimum superposition started with only one Cu atom gave better results.

BETA. The whole structure has appeared on the screen in all cases presented in Tables 2 and 3 but in two cases the whole molecule was shifted by a vector (0,0,0.08) with respect to the cell origin (the wrong shift vector was selected in the multisolution).

AZET. The well known test structure distributed with *MULTAN*. The Cl–Cl vectors in the sharpened Patterson function have sequential numbers 2, 4, 6, 8, 9, 11, 42 and 52. This is an example where fine details of the algorithm are crucial for the success of the method. The cross-vector superposition and Fourier recycling in $P1$ did not give the solution for low-order Patterson peaks.

MGHEX. The Patterson peak for the Mg–Mg vector is no. 7. Although the structure is large, the space group is of order 3 and the scattering power of Mg^{2+} is not very high, there were no special problems with the structure determination. The only problem was to find the $H–H$ vector. One had to try each Patterson vector starting from peak no. 2.

APAPA. The P–P vectors can be found in the broad interval from 2 to 123. The top P–P peaks are numbers 2, 11, 16 and 17. Peak no. 2 has coordinates (0.5,0.5,0.5) resulting from averaging because of the overlap of several P–P vectors. The CVF is not able to filter out two P atoms effectively. There are too many atoms (552) in the unit cell, the experiment is not very accurate and the execution times are very long so the structure has not been tested by Fourier recycling in $P1$.

The results show that more than 80% of heavy-atom structures can be solved automatically and almost all structures can be solved with minimal intellectual effort simply by inserting the top Patterson peaks into the automatic procedure. These methods are useful alternatives to previously published methods (Pavelčik, 1988; Pavelčik, Kuchta & Sivý, 1992), particularly for structures with lower symmetry (especially $P\bar{1}$) and uncertain space groups.

References

BOCELLI, G. (1990). Personal communication.
BORISOV, S. V. (1964). *Kristallografiya*, **9**, 603–610.
BUERGER, M. J. (1959). *Vector Space*. New York: Wiley.
HARKER, D. (1936). *Chem. Phys.* **4**, 381–390.
ILJUKHIN, V. V., KUZ'MIN, E. A. & BELOV, N. V. (1981). *New Approaches to the Patterson Function*. Moscow: Nauka. (In Russian.)
JACOBSON, R. A. & GUGGENBERGER, L. J. (1966). *Acta Cryst.* **20**, 592–593.
KETTMANN, V. (1990). Personal communication.
KRÄTSMÁR-ŠMOGROVIČ, J., PAVELČIK, F., SOLDÁNOVÁ, J., SIVÝ, J., SERESSOVÁ, V. & ŽEMLIČKA, M. (1991). *Z. Naturforsch. Teil B*, **46**, 1323–1327.
KUCHTA, Ł., SIVÁK, M. & PAVELČIK, F. (1993). *J. Chem. Res. (S)*, p. 393; *J. Chem. Res. (M)*, pp. 2801–2810.
KUZ'MIN, E. A., GOLOVACHEV, N. V. & BELOV, N. V. (1970). *Dokl. Akad. Nauk SSSR*, **192**, 86–88.
LUGER, P. & FUCHS, J. (1986). *Acta Cryst.* A**42**, 380–386.
PAVELČIK, F. (1986). *J. Appl. Cryst.* **19**, 488–491.
PAVELČIK, F. (1988). *Acta Cryst.* A**44**, 724–729.
PAVELČIK, F. (1992). Unpublished.
PAVELČIK, F., KUCHTA, Ł. & SIVÝ, J. (1992). *Acta Cryst.* A**48**, 791–796.
PAVELČIK, F., SIVÝ, J., RIZZOLI, C. & ANDREETTI, G. D. (1992). *J. Appl. Cryst.* **25**, 328–329.
RICHARDSON, J. W. JR & JACOBSON, R. A. (1987). *Patterson and Pattersons. Fifty Years of the Patterson Function*, edited by J. P. GLUSKER, B. K. PATTERSON & M. ROSSI, pp. 311–317. New York: IUCr/Oxford Univ. Press.
SHELDRICK, G. M. (1991). *Crystallographic Computing* 5, edited by D. MORAS, A. D. PODJARNY & J. C. THIERRY, pp. 143–157. Oxford: IUCr/Oxford Univ. Press.
SHELDRICK, G. M., DAUTER, Z., WILSON, K. S., HOPE, H. & SIEKER, L. C. (1993). *Acta Cryst.* D**49**, 18–23.
SIMONOV, V. I. (1975). International Summer School on Crystallographic Computing, Prague, Czech Republic.
SIMONOV, V. I. (1982). *Computational Crystallography*, edited by D. SAYRE, pp. 150–158. Oxford: Clarendon Press.
SIMPSON, P. G., DOBROTT, R. D. & LIPSCOMB, W. N. (1965). *Acta Cryst.* **18**, 169–179.
SIVÝ, J. (1992). Personal communication.
SOLDÁNOVÁ, J., TYRŠELOVÁ, J. & PAVELČIK, F. (1994). *Chem. Pap.* Submitted.
TYRŠELOVÁ, J. & PAVELČIK, F. (1992). *Acta Cryst.* C**48**, 1207–1209.
TYRŠELOVÁ, J., PAVELČIK, F. & KUCHTA, L. (1993). Unpublished.