PHILLIPS, G. N., LATTMAN, E. E., CUMMINS, P., LEE, K. Y. & COHEN, C. (1979). *Nature (London)*, **278**, 413-417.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.
STUBBS, G. (1972). *Structural Studies of Crystalline Proteins.* Thesis, Univ. of Oxford.
STUBBS, G. & DIAMOND, R. (1975). *Acta Cryst.* A**31**, 709-718.
STUBBS, G. & MAKOWSKI, L. (1982). *Acta Cryst.* A**38**, 417-425.
STUBBS, G. & STAUFFACHER, C. (1981). *J. Mol. Biol.* **152**, 387-396.
STUBBS, G., WARREN, S. & HOLMES, K. (1977). *Nature (London)*, **267**, 216-221.
STUBBS, G., WARREN, S. & MANDELKOW, E. (1979). *J. Supramol. Struct.* **12**, 177-183.
WANG, B. C. (1985). In preparation.
WASER, J. (1955). *Acta Cryst.* **8**, 142-150.
WEISSMAN, L., STAUFFACHER, C. & EISENBERG, D. (1985). In preparation.
WITTMANN, H. G. (1962). *Z. Vererbungsl.* **93**, 491-530.

Acta Cryst. (1985). A**41**, 262-268

Search for a Fragment of Known Geometry by Integrated Patterson and Direct Methods

BY ERNST EGERT AND GEORGE M. SHELDRICK

Institut für Anorganische Chemie der Universität Göttingen, Tammannstrasse 4, D-3400 Göttingen, Federal Republic of Germany

(*Received 6 July* 1984; *accepted 26 November* 1984)

Dedicated to Professor Jack D. Dunitz on the occasion of his sixtieth birthday

Abstract

A method is presented that attempts to exploit all the *a priori* available information in order to locate a fragment of known geometry in the unit cell. Whereas the orientation of the search model is determined by a conventional but highly automated real-space Patterson rotation search, its position in the cell is found by maximizing the weighted sum of the cosines of a small number of strong translation-sensitive triple-phase invariants, starting from random positions. A Patterson minimum function based on intermolecular vectors is calculated only for those solutions that do not give rise to intermolecular contacts shorter than a preset minimum. This procedure avoids the time-consuming refinement in Patterson space and should be especially efficient for large structures. Finally, the best solutions are sorted according to a figure of merit based upon the agreement with the Patterson function, the triple-phase consistency and an R index involving E_{obs} and E_{calc}. Tests with about 30 known structures, using search fragments taken from other published structures or from force-field calculations, have indicated that this novel combination of Patterson and direct methods is reliable and widely applicable. A few selected examples demonstrate the power of the computer program *PATSEE*, which is compatible with *SHELX*84 and will be distributed together with it. *PATSEE* is valid and efficient for all space groups and imposes no limits on the number of atoms or data. The orientation search for a single fragment allows one additional degree of torsional freedom, and up to two fragments may be translated simultaneously.

Introduction

The choice of strategy for the solution of a crystal structure at atomic resolution is usually determined by the presence or absence of heavy atoms. Thus it is common practice to solve light-atom structures with direct methods and those containing heavy atoms with Patterson techniques. If this (very often straightforward) strategy fails, it may be advisable to resort to the corresponding alternative method: direct methods may well reveal the positions of heavy atoms, and the Patterson function can be interpreted even for purely light-atom structures, such as those of organic molecules, provided that part of the molecular geometry is known. This so-called Patterson search has been shown by various authors to be a powerful tool for solving difficult crystal structures; its great strength is that it employs chemical information directly, and so can compensate for mediocre precision and resolution of the X-ray data (Egert, 1983, and references cited therein). Nevertheless, it is not nearly as popular as direct methods, which owe part of their success to automation and superior computational efficiency. In this paper we describe an attempt to combine the merits of both methods - in a manner that is generally applicable, efficient, automatic, computer independent and easy to use - and thus to exploit *all* the *a priori* available information in order to solve large problem structures.

Preparation of the search

There are a number of different methods of performing a Patterson search, but they all fall into one of

0108-7673/85/030262-07$01.50

two general categories depending upon whether they operate in direct (vector) or reciprocal space. Although the latter method is less transparent and flexible (for instance, the molecular boundary has to be a simple geometrical shape), it was chosen by most of the early authors probably because it does not require the storage of the Patterson function (Blow, 1976). Nowadays, the existing main-frame computers permit fast access to sufficient memory, so that there is now no reason to perform a fragment search in reciprocal space (at least as far as 'small' molecules are concerned). Thus our Patterson search procedure, like the methods of Nordman (1966, 1980) and Hornstra (1970), operates in real space and is based on the magnitudes of the Patterson function at positions corresponding to interatomic vectors.

Generally, a Patterson search in vector space consists of the following stages: (1) definition of a search model; (2) calculation and storage of the Patterson function; (3) rotation search; and (4) translation search. It is a serial technique, with the last two stages crucially dependent on the accuracy of all preceding ones. Thus the first step is by no means trivial; this is especially true for a procedure like ours where the search fragments are taken as rigid and no model refinement is attempted (with the exception of one torsional degree of freedom between two rigid groups). Also, the models are not represented by a smooth electron density distribution, but are regarded as an ensemble of point atoms with weights ($Z \times$ s.o.f.)* so that, generally speaking, a small well-defined search model is more appropriate than a larger one containing several incorrect atoms. The model is defined by atomic coordinates in a given coordinate system; these will normally be either fractional (taken from a related crystal structure, very conveniently obtained from the Cambridge Crystallographic Database) or Cartesian (*e.g.* from a force-field calculation). Temperature factors are completely ignored, but the weight of an atom may be modified by changing its site occupation factor. Atoms that are expected to have close contacts (hydrogen bonds *etc.*) or occupy special positions should be marked in order to avoid short distances between them being interpreted as physically unreasonable.

The asymmetric unit of the Patterson function is generated by the *SHELX*84 Fourier program (Sheldrick, 1985), and sampled at $51 \times 51 \times N$ grid points (N arbitrary). Since the order of Fourier summation is chosen so as to make the sampling as uniform as possible, the distance between neighbouring grid points is normally small enough to avoid time-consuming interpolation. For almost all purposes, we recommend using $E \times F$ as coefficients; these lead to a sharper map than F^2 but generate fewer ripples than E^2. However, we do not remove the origin as suggested by Nordman (1966). In order to keep the whole Patterson map in core (a prerequisite for an efficient search procedure) it must be densely packed. A compromise is required between the accuracy of the stored Patterson values and the storage limitations. We decided to represent each grid value by a digit between 0 and 7 so that it can be stored in three bits of computer memory; *cf.* the quite successful two-bit representation of Braun, Hornstra & Leenhouts (1969). The Patterson values are encoded according to seven test levels and care should be taken that the distribution is reasonably smooth (with the exception of the zero values, which include all regions where no vector density is found). If the user does not supply his own test levels, they are set such that the second one equals the median of the cumulative Patterson distribution, and the difference between two successive levels is about half the expected height of the highest single vector calculated from the height of the origin peak (arbitrarily set to 999). Several Patterson values encoded in this way are combined into one (octal) number, which occupies one real word of computer memory. The number of points that can be combined depends on the word length and the unpacking (shifting and masking) operations required for retrieval of a specified digit within this number; for example, seven is optimum for most 32-bit machines.

* s.o.f. = site occupation factor.

Rotation search

The region around the origin of the Patterson function is dominated by intramolecular vectors, which depend on the orientation but not on the position of the fragment. Thus the full six-dimensional search can be split into two three-dimensional searches, a rotation and a translation search (depending on the space group, the latter may be of even lower dimensionality).

Since it is most inconvenient to perform the rotation search using crystal coordinates, the known fragment is first placed in a Cartesian coordinate system with its centre located at the origin. The next step is to set up the intramolecular vector set to be used for the search, *i.e.* to express the model geometry (which should always be checked thoroughly) in the form of discrete vectors with associated weights. Of the $N(N-1)/2$ intramolecular vectors, the short ($d <$ 2 Å) and long (say $d > 6$ Å) ones are immediately eliminated. Since the inner sphere around the Patterson origin shows some vector density everywhere, the short vectors provide little angular discrimination and are normally not very useful for determining the orientation of the fragment. However, they may be important for molecules (*e.g.* those consisting of fused aromatic rings) that are characterized by a few short vectors with high weights. An upper limit for the vector length is also advisable because very long

vectors, though quite characteristic of the search model, suffer most from uncertainties in the geometry and could easily miss the corresponding maximum in the Patterson map. All pairs in the remaining vector list with ends closer than 0·4 Å* are replaced by a weighted average vector with the combined weight. In order to save computing time, vectors with weights lower than 10% of the highest may be omitted; experience has shown that a small number of vectors with high weight is often more suitable than a large number of low-weight vectors (Nordman & Schilling, 1970). At this stage, a comparison between the high-weight vectors and the origin-peak distances for the most prominent Patterson maxima should indicate whether or not the gross features of the search model are correct.

Any orientation of a rigid fragment relative to a fixed coordinate system can be described by three angles corresponding to successive rotations about properly chosen axes.† The asymmetric unit of angular space depends on both the Laue group and the model symmetry (Tollin, Main & Rossmann, 1966). In contrast to other Patterson search programs, which scan the respective range of angles by specifying rotation increments, we have chosen to generate random orientations without taking into account problem-specific equivalences between angle triplets; this is later taken care of by the 'equivalence test'. Apart from computational advantages, this strategy automatically recognizes all angular symmetry relations (including non-linear ones) and makes it unnecessary to use a specific initial orientation of a symmetric model. The optimum number of orientations to be tested depends on the size and the shape of the search fragment, the Laue group and the Patterson grid intervals. We usually generate 10 000–60 000 angle triplets, which corresponds to mean rotation increments of about 7°; this is normally sufficient for the coarse location of the maxima.

For each orientation, the correlation between the rotated intramolecular vector set and the Patterson function is measured by a modified sum function; this seems to be the most appropriate criterion for distinguishing correct from false solutions (Hornstra, 1970). Each vector is thus transformed to the asymmetric unit of Patterson space, whereupon its weight w_i is compared with the nearest Patterson grid value P_i. As a figure of merit, we take the average of the, say, 30% worst-fitting vectors, *i.e.* those with lowest P_i/w_i:

$$\mathrm{RFOM} = \frac{1}{n}\sum_{i=1}^{n}\frac{P_i}{w_i} \quad (n \simeq 0{\cdot}3 \times n_{\mathrm{total}}).$$

Table 1. *Rotation search procedure*

1. Compilation of intramolecular vector set
2. Generation of random orientations
3. Retrieval of Patterson grid values and calculation of RFOM
4. Overlap test
5. Equivalence test
6. Sorting of solutions according to RFOM
7. Refinement of best solutions

RFOM is related to the 'weighted minimum average', $\sum_{i=1}^{n} P_i/\sum_{i=1}^{n} w_i$, used by Nordman (1980) and, depending on n, resembles either the sum or the minimum function.

Before an orientation with promising RFOM is sorted into a short list of best solutions, it has to pass two tests. The 'overlap test' ensures that no close interatomic contacts (say $d < 2$ Å) arise from the application of the lattice translations present (except for distances between atoms previously marked) and the 'equivalence test' compares the orientation in question with those already stored. The latter test does not make use of angular relations but is instead based upon distances between equivalent atoms. All symmetry elements of the Laue group that correspond to proper rotations are successively applied to the solutions in the list; if the search model is racemic improper ones may also have to be considered. Two orientations are regarded as similar when *all* pairs of equivalent atoms are closer than, say, 0·6 Å; in that case only the better one is kept. Since these two tests are necessary only for orientations with a larger RFOM than the worst one in the list of 'best' solutions, they are less often applied after a number of good solutions have already been found. Thus the computing time for the rotation search increases only slowly with the number of orientations tested.

In order to improve the performance of the subsequent translation search, it is worthwhile refining the best solutions by a restricted rotation search. The maximum within each promising region of angular space is found by testing up to 1000 additional random points, which corresponds to a mean rotation increment of less than 2°.

If the search model has one torsional degree of freedom the whole procedure outlined in Table 1 is repeated for each distinct geometry (conveniently defined by a range of possible torsion angles and an appropriate increment), in which case a merged list of best solutions is set up. At the end of the rotation search, a small number of promising orientations are passed over to the translation search. It is our experience that the correct one is usually present among the best two or three for reasonably sized fragments ($p^2 \simeq 0{\cdot}2$).

Translation search

In procedures to position a fragment of known geometry in the unit cell, the translation search has

* This value is increased for low-resolution Patterson maps.

† There are various definitions of the Eulerian angles. For computational reasons, we prefer successive rotations about the **A**, **B** and **C** axes, in that order.

usually proved to be less reliable than the rotation search. We restrict ourselves to Patterson space, this is because the 'cross' vectors used to locate a fragment with respect to the origin suffer from errors in both the model geometry and orientation amplified by the symmetry elements; in addition, model vectors with very high weight are less likely than in the rotation search. In order to obtain atomic positions accurate enough for the subsequent structure expansion or refinement, either a fine search grid or some optimization of promising solutions is necessary. Thus, if a sum or minimum function is used, the time for the translation search rises rapidly with the complexity of the structure. Furthermore, time-consuming interpolation procedures can hardly be avoided.

The calculated phases, in contrast to the way in which the Patterson function must normally be stored, are a *continuous* function of the atomic coordinates. When a fragment is moved through the unit cell keeping its orientation fixed:

$$F_{\mathbf{h}} = F^0_{\mathbf{h}} \exp 2\pi i \mathbf{h} \,.\, \Delta\mathbf{r}$$

since all atomic displacements $\Delta\mathbf{r}$ are the same. So the scattering contributions from the atoms of the search model have to be summed only once for each orientation and reflection to yield a structure factor $F^0_{\mathbf{h}}$ for the starting position; subsequently, the structure factor $F_{\mathbf{h}}$ for any position is readily obtained by multiplication with a simple phase factor. For the true structure, the individual phases of the strongest reflections are linked by various statistical phase relations; amongst these, the three-phase structure invariants have proved to be especially useful:

$$\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}} \simeq 0.$$

The search fragment is usually incomplete and may also be not very accurate. Nevertheless, if its scattering power is significant, the triple-phase relations should hold at least approximately for the correct solution, in the sense that the distribution of the phase sums is far from being random.

These considerations led us to the development of a novel strategy for a Patterson translation search, which, as far as we know, for the first time fully exploits in an integrated fashion the information contained in the sharpened Patterson function, the three-phase structure invariants and allowed intermolecular distances (Table 2). In short, we have chosen the optimization of a weighted sum of cosine invariants as our refinement procedure, with the Patterson correlation and an R index as additional figures of merit, and the minimum intermolecular distance as a possible rejection criterion. This method is computationally efficient, especially for larger structures, because the refinement is based on phase relations derived from a relatively small number of large E magnitudes. Only when an acceptable solution has been found by this 'direct search' is it necessary to calculate the time-consuming Patterson correlation. A less obvious advantage is that the dependence of the phase angles on small coordinate shifts is relatively linear, *i.e.* the sum of cosines is an approximately quadratic function, which results in efficient refinement.

Table 2. *Translation search procedure*

1. Search for most probable triple-phase relations
2. Calculation of $E^0_{\mathbf{h}}$ for given orientation
3. Selection of suitable phase relations
4. Generation of random positions
5. Preliminary distance test
6. Refinement of fragment positions
7. Final distance test
8. Retrieval of Patterson grid values and calculation of TFOM
9. Sorting of independent solutions according to CFOM

Since, in order to save computing time, relatively few phase relations are employed for the refinement, they have to be selected carefully. Initially the most probable three-phase structure invariants are found by searching a list of, for example, 200 reflections with largest E values generated by *SHELX*84 together with the Patterson map. Normally about 100 phase relations linking an approximately equal number of E magnitudes are sufficient. However, not all of these are used simultaneously because it would not be sensible to employ E magnitudes to which the oriented search model does not contribute significantly. Assuming N^* independent fragments with random positions, their relative contributions to the E magnitudes are given by

$$Q_{\mathbf{h}} = \left(\sum_{i=1}^{N} |E^i_{\mathbf{h},\mathrm{calc}}|^2 \right)^{1/2} / |E_{\mathbf{h},\mathrm{obs}}|.\dagger$$

Accordingly, all triple-phase relations are ignored for which $Q_{\mathbf{h}} Q_{\mathbf{k}} Q_{-\mathbf{h}-\mathbf{k}}$ is less than, say, 80% of its average value for the orientation in question. Thus only the 40–60 most probable *and* translation-sensitive three-phase structure invariants are actually used for a translation search. It is advisable to apply a 2θ limit to the E values before searching for phase relations, since high-order reflections may be influenced considerably by errors in the model. However, if the cut-off is too severe, the accuracy of the phase-refinement procedure suffers because of the occurrence of broad and shallow maxima. It seems that a nominal resolution of about 1 Å is the best compromise; this does not necessarily mean that structures with a lower resolution cannot be solved by this method.

The asymmetric unit of translation search and thus the number and scan range of parameters are uniquely defined by the Cheshire groups (Hirshfeld, 1968).

* $1 \le N \le 3$ for the procedure presented here.

† The average value of $Q_{\mathbf{h}}$ is usually a maximum for the correct orientation but is not very sensitive to errors.

Following these rules, random positions are generated for the rotated search fragment(s). It is our experience that the triple-phase refinement is well able to home in on fragments starting about 0·5 Å away from their true positions. This means that at least one trial per cubic Ångström is necessary in order to have a good chance of locating one search model correctly. Since the number of trials rises as a high power of the number of independent fragments, it is unreasonable to search for more than two simultaneously. However, any number of fixed fragments (obtained from a previous search or a heavy-atom Patterson interpretation, for example) may be added and, in fact, are quite valuable provided they are correct.*

Taking the limited range of the subsequent refinement into account, only those random positions that are fairly close to physically reasonable solutions are worth refining; thus all positions that give rise to very short intermolecular distances (say $d < 1{\cdot}8$ Å) are immediately rejected. The refinement procedure consists of two cycles during which the translation parameters are refined one after another by optimizing

$$\text{TPRSUM} = \frac{\sum E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}} \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}{\sum E_{\mathbf{h}} E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}},$$

where the two sums are taken over all selected three-phase structure invariants. TPRSUM is expected to be large and positive for the correct solution ($-1 \leq \text{TPRSUM} \leq 1$). The fragment is moved stepwise through the cell until a point is found with larger TPRSUM than its neighbours. The coordinates of the maximum are then estimated by parabolic interpolation. If TPRSUM > 0 after the first cycle, the initial step size of, say, one quarter of the nominal resolution of the E values is reduced by $\frac{1}{5}$ (to about 0·05 Å) in order to locate the fragment more accurately. At the end of the second cycle, only positions with TPRSUM $> 0{\cdot}25$ (approximately corresponding to $\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}} < 75°$) are regarded as possible solutions and tested again for short contacts. This time the rejection criterion is more strict (say $d < 2{\cdot}4$ Å) as the atom positions are final (but some tolerance for model errors has to be allowed for).

For solutions that have survived all these tests, the correlation between the Patterson function and the intermolecular vector set is examined by comparing the weight of each vector with the nearest grid value. The fit is measured by (*cf.* RFOM)

$$\text{TFOM} = \frac{1}{n} \sum_{i=1}^{n} \frac{P_i}{w_i} \quad (n \simeq 0{\cdot}2 \times n_{\text{total}}).$$

The vector fraction to be used for TFOM (or RFOM) should be increased for bad or very small models. A small number of 'best' solutions (according to both TPRSUM and TFOM) are stored provided that they pass various tests for possible equivalence (allowed origin shift or lattice translation). Although the true position of the search fragment is usually recognizable at this stage, an R index based on E magnitudes has proved very useful in distinguishing further between correct and false solutions. It is defined as:

$$R_E = \frac{\sum \{|E_{\text{obs}}| - |E_{\text{calc}}|/p\}}{\sum |E_{\text{obs}}|}$$

since $\overline{E^2_{\text{calc}}} = p^2 \overline{E^2_{\text{obs}}}$ (p^2 = fractional scattering power of the search model). Only positive contributions to the numerator are considered, *i.e.* if $|E_{\text{calc}}|$ is larger than its expected value, complete agreement between fragment position and experiment is assumed. Finally, the solutions are sorted according to a combined figure of merit:

$$\text{CFOM} = \frac{0{\cdot}2}{R_E} \times \text{TFOM} \times \text{TPRSUM}^{1/2}.$$

In this expression, TPRSUM is given lower weight because it was the quantity optimized. If a rotation search preceded the translation search, TFOM is replaced by (RFOM + TFOM)/2. If R_E is less than 0·05 it is reset to 0·05, so that it does not dominate CFOM for very small fragments. For all solutions printed, a Patterson sum function is calculated as a measure of fit/misfit for each individual atom, taking all vectors (intra- and intermolecular) into account; this enables identification of possible wrong atoms and thus model correction.

The procedure described differs from other Patterson translation functions (Nordman, 1966; Hornstra, 1970; Doesburg & Beurskens, 1983) in that the oriented model is placed with respect to *all* symmetry elements of the space group simultaneously. Tests with known structures have indicated that this routine is able to locate very large fragments (of more than 300 atoms), in which case the distance tests already preclude the majority of trial positions, as well as single atoms even when the latter are not very heavy (*e.g.* *P* or *S* in large organic structures). Above all, the variety of different criteria employed to judge solutions should make this combination of Patterson and direct methods a powerful structure-solving strategy, if chemical information is available. One would expect that a position that is in agreement simultaneously with packing criteria (d_{min}), the Patterson function (TFOM), triple-phase relations (TPRSUM) and E values (R_E) is probably correct, and our experience shows that this is indeed the case.

Features of the program *PATSEE*

The procedures outlined have been implemented as a computer program called *PATSEE*, which is valid

* Since known atoms uniquely define the origin of the unit cell it is sometimes more economic *not* to include them, if their size or scattering power is small and the asymmetric unit for the search much smaller (*e.g.* only one or two dimensional) without them.

Table 3. *Data on the solution of five test structures with PATSEE*

[c] refers to the correct, [f] to the highest-ranked wrong solution.

	LAC1	SUOA	TPH	MUNICH1	AZET
Molecular formula	$C_{37}H_{49}N_3O_7$	$C_{28}H_{38}O_{19}$	$C_{24}H_{20}N_2$	$C_{20}H_{16}$	$C_{21}H_{16}ClNO$
Space group	$P2_1$	$P2_12_12_1$	$C222_1$	$C2$	$Pca2_1$
Z	2	4	12	8	8
Model size (atoms)	23	11	13	20	2
Scattering fraction	0·41	0·21	0·30	0·44	0·10
Number of rotation trials	40 000	25 000	10×25 000	40 000	—
RFOM^f/RFOM^c	0·68	0·92	0·97	0·48	—
Number of translation trials	100	500	700	300	500
TPRSUM^f/TPRSUM^c	1·00	1·17	1·17	0·98	1·06
TFOM^f/TFOM^c	0·63	1·05	0·75	0·82	0·93
R_E^c/R_E^f	0·76	0·93	0·80	0·74	0·79
CFOM^f/CFOM^c	0·57	0·90	0·76	0·67	0·67
Computer time (min)*	10	5	24	8	7
Computer time (min) for direct methods*†	15	200	20	60	10

* On a UNIVAC 1100/83.
† Using *SHELX*84.

and efficient for all space groups in all settings. Since it is written in a simple subset of Fortran it may be run without significant alteration on a wide range of computers, provided that sufficient memory (at least 40K words) may be addressed directly and that the word length is at least 32 bits. The program has been designed to be fully automatic, but the default settings may easily be changed by experienced users. The rotation search can find the orientation of a fragment of any size and allows one torsional degree of freedom. The translation search may locate up to two independent search models of any size (including single atoms), taking into account known atoms at fixed positions, if any. Since the program is compatible with *SHELX*84, convenient facilities exist for the generation of the Patterson function before, and the structure expansion after, the fragment search. *PATSEE* is available on request and will be distributed together with *SHELX*84.

Test structures

The program has already been tested on about 30 known structures of different size and complexity, using fragments taken from related structures in the Cambridge Crystallographic Database or calculated by force-field methods. In all cases the best solution on the basis of the combined figure of merit could easily be expanded by tangent expansion and Fourier recycling to give the complete structure. The results for five test structures are summarized in Table 3. They all present difficulties in direct methods (three of them were originally solved by Patterson search) and will be briefly discussed in order to illustrate the range of possible applications for *PATSEE*. Apart from some minor modifications the default parameters were used throughout.

LAC1, a steroidal lactone, was ideally suited to Patterson search since a large and reliable model was available from the crystal structure of a diastereoisomer (Egert, Cruse & Kennard, 1983). Indeed, the solution was straightforward, as indicated by the convincing figures of merit. Also the structure of sucrose octaacetate (SUOA) (Oliver & Strickland, 1984) did not cause any problems, although the small size of the search model taken from the structure of sucrose (Hanson, Sieker & Jensen, 1973) resulted in less clear-cut figures of merit (but, on the other hand, in a very rapid structure determination). The problem with tetraphenylhydrazine (TPH) (Hoekstra, Vos, Braun & Hornstra, 1975) was that the largest rigid fragment consists of only seven planar atoms. Therefore an extended search model Ph-N-Ph comprising one torsional degree of freedom was constructed. By specifying a suitable range (0-90°) of torsion angles and an increment (10°) altogether ten geometries were tested. Since the short intramolecular vectors were expected to be important due to their high weights they were retained in this particular case. Furthermore, the N atom and its nearest neighbours were marked in order to locate the molecule on the twofold axis; otherwise the distance tests would have rejected the correct solution. For the structure determination of 9,10,11,12-dibenzopentacyclo-[6.2.2.$0^{2,6}$.$0^{2,7}$.$0^{3,7}$]dodeca-9,11-diene (MUNICH1) (Szeimies-Seebach, Harnisch, Szeimies, Van Meerssche, Germain & Declercq, 1978) no suitable model was at hand. We therefore calculated the molecular geometry by the force-field program *PIMM* (Lindner, 1974) and used it as input for *PATSEE*. Since there are two independent molecules with *mm*2 symmetry, eight distinct orientations could lead to a correct solution. Indeed there were just eight positions with excellent figures of merit - all of them correct! This example clearly demonstrates how fragment search and force-field methods can be successfully combined (see also Egert, 1983).

If the known fragment consists of a single 'heavy' atom its orientation is of course always correct and a translation search with *PATSEE* should reveal its

position. This strategy was employed for the straightforward structure solution of 3-chloro-1,3,4-triphenylazetidin-2-one (AZET) (Colens, Declercq, Germain, Putzeys & Van Meerssche, 1974) with two Cl atoms at arbitrary starting positions (for example, both placed at the origin). For such a search, a considerable amount of computing time may be saved by specifying a proper value for the allowed intermolecular distance (*e.g.* for metal complexes); otherwise the distance tests could be rather ineffective. It may also be useful to calculate CFOM without TPRSUM (a *PATSEE* option) in order to give increased weight to TFOM, which is very reliable in such cases.

The values for the figures of merit show that RFOM, TFOM and R_E together are strongly indicative of the correct solution. TPRSUM is often only a local maximum but it enables the rapid location of the search fragment. All test examples (not only the five discussed here) confirm that *PATSEE* is reliable and widely applicable. In terms of computing times, it is also competitive with direct methods; under favourable circumstances (see SUOA and MUNICH1) it can even prove more economical. In any case, *PATSEE* offers a powerful alternative if chemical information is available.

We thank the Fonds der Chemischen Industrie for a Liebig-Stipendium (to EE) and the Deutsche Forschungsgemeinschaft for financial support. One of us (EE) is very much indebted to Professor Jack D. Dunitz for his enthusiastic interest and hospitality during a three-months stay in Zürich.

References

BLOW, D. M. (1976). *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 229-238. Copenhagen: Munksgaard.
BRAUN, P. B., HORNSTRA, J. & LEENHOUTS, J. I. (1969). *Philips Res. Rep.* **42**, 85-118.
COLENS, A., DECLERCQ, J.-P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119-122.
DOESBURG, H. M. & BEURSKENS, P. T. (1983). *Acta Cryst.* A**39**, 368-376.
EGERT, E. (1983). *Acta Cryst.* A**39**, 936-940.
EGERT, E., CRUSE, W. B. T. & KENNARD, O. (1983). *Acta Cryst.* C**39**, 95-99.
HANSON, J. C., SIEKER, L. C. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 797-808.
HIRSHFELD, F. L. (1968). *Acta Cryst.* A**24**, 301-311.
HOEKSTRA, A., VOS, A., BRAUN, P. B. & HORNSTRA, J. (1975). *Acta Cryst.* B**31**, 1708-1715.
HORNSTRA, J. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 103-109. Copenhagen: Munksgaard.
LINDNER, H. J. (1974). *Tetrahedron*, **30**, 1127-1132.
NORDMAN, C. E. (1966). *Trans. Am. Crystallogr. Assoc.* **2**, 29-38.
NORDMAN, C. E. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 5.01-5.13. Bangalore: The Indian Academy of Sciences.
NORDMAN, C. E. & SCHILLING, J. W. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 110-114. Copenhagen: Munksgaard.
OLIVER, J. D. & STRICKLAND, L. C. (1984). *Acta Cryst.* C**40**, 820-824.
SHELDRICK, G. M. (1985). In preparation.
SZEIMIES-SEEBACH, U., HARNISCH, J., SZEIMIES, G., VAN MEERSSCHE, M., GERMAIN, G. & DECLERCQ, J.-P. (1978). *Angew. Chem. Int. Ed. Engl.* **17**, 848-850.
TOLLIN, P., MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **20**, 404-407.

Acta Cryst. (1985). A**41**, 268-273

Statistical Mechanics Approach to the Structure Determination of a Crystal

BY S. V. SEMENOVSKAYA

Institute of Organo-Element Compounds, Academy of Sciences of the USSR, Moscow, USSR

AND K. A. KHACHATURYAN AND A. G. KHACHATURYAN

Institute of Crystallography, Academy of Sciences, Moscow, USSR

(*Received* 22 *June* 1984; *accepted* 7 *December* 1984)

Abstract

The previously formulated new approach to the structure analysis of a crystal based on the profound analogy between the problem of determination of thermodynamic equilibrium in statistical mechanics and the optimization problem for a function of many variables [Khachaturyan, Semenovskaya & Vainshtein (1979). *Sov. Phys. Crystallogr.* **24**, 519-524; (1981). *Acta Cryst.* A**37**, 742-754] is developed. In this approach, a crystal structure is determined by the equilibrium low-temperature state of a model non-ideal gas composed of the atoms within a crystal unit cell, the unit cell and the R factor being regarded as a vessel and an interatomic interaction Hamiltonian, respectively. In contrast to the above cited papers, the low-temperature equilibrium state is found by means of the Monte Carlo sampling scheme usually utilized in statistical mechanics applications. The main advantage of such a treatment is that the

0108-7673/85/030268-06$01.50