$\beta + m$ and $z_3 = \beta + 2m$, where $1 \leq \beta \leq m$, and then performing the summation from 1 to m.

References

ALEXANDER, E., KALMAN, Z. H., MARDIX, S. & STEINBERGER, I. T. (1970). *Philos. Mag.* **21**, 1237–1240.

KIFLAWI, I., MARDIX, S. & KALMAN, Z. H. (1969). *Acta Cryst.* B**25**, 2413–2415.

KIFLAWI, I., MARDIX, S. & STEINBERGER, I. T. (1969). *Acta Cryst.* B**25**, 1581–1586.

MARDIX, S., ALEXANDER, E., BRAFMAN, O. & STEINBERGER, I. T. (1967). *Acta Cryst.* **22**, 808–812.

MARDIX, S., KALMAN, Z. H. & STEINBERGER, I. T. (1970). *Acta Cryst.* A**26**, 599–603.

PÁTEK, K. (1961). *Czech. J. Phys.* B**11**, 686–690.

Acta Cryst. (1983). A**39**, 936–940

Patterson Search – an Alternative to Direct Methods

BY ERNST EGERT

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

(*Received* 10 *March* 1983; *accepted* 12 *July* 1983)

Abstract

On the basis of the solution of six unknown structures by a real-space Patterson search the merits and limitations of the method are discussed. It is shown how chemical information can be used to provide a reliable starting point for direct methods when there is no automatic solution. Two examples demonstrate the potential usefulness of force-field calculations for generating the geometries of appropriate search fragments in the absence of related crystal structures. The further discussion deals with future applications of the vector search method to determine large structures. A Patterson search program, which incorporates all necessary features of a modern program and is compatible with *SHELX*84, is being developed.

Introduction

Nowadays the overwhelming majority of light-atom structures are determined by direct methods, many of them automatically. Although these techniques have proved extremely powerful, they cannot solve all structures and even less complex problems still resist solution, sometimes for reasons not obvious to the investigator. The principal weakness of the method lies in its dependence on a few key reflections during the early stages of phase determination. If some of them are measured incorrectly (*e.g.* weak data from very small crystals, or poor resolution arising from solvent disorder), an inappropriate starting set may be chosen and thus the whole chaining process can go wrong. Current research activities in this field centre on how to improve this situation (Furusaki, 1979; Jia-Xing, 1981; Hull, Viterbo, Woolfson & Shao-Hui, 1981; Schenk, 1983). Furthermore, direct methods are based on a random distribution of atoms, so that structures which deviate considerably from this assumption (*e.g.* those containing planar rings) often present problems.

Difficult structures may eventually be solved by varying one or several parameters that govern the phasing procedure. If this does not work either, the problem reduces to an often frustrating trial-and-error approach in which many phase sets are tested. On the other hand, valuable chemical information, which for most (organic) structures is at least partially present, is not fully used by direct methods. It seems paradoxical, especially to a chemist, that even crystal structures of molecules whose total geometry is well known cannot be determined. Therefore, instead of using chemical knowledge only indirectly, *e.g.* to modify E values or recognize correct electron-density maps, one should try to use it directly in those cases where direct methods fail.

The Patterson function differs from the statistical methods in that all data are used simultaneously and independently of each other; it is therefore less sensitive to a few incorrectly measured or missing reflections. Since the resulting vector map contains in principle the whole structural information, attempts have been made to solve light-atom structures semi-automatically from Patterson syntheses (Mighell & Jacobson, 1963; Gorres & Jacobson, 1964; Simpson, Dobrott & Lipscomb, 1965). However, considering the expected vector peak density it is quite hopeless to unravel more complex problems *ab initio*, *i.e.* without additional information. If part of the molecular geometry is

0108-7673/83/060936-05$01.50

known, one can calculate the resulting vector pattern characteristic of this fragment and varying with its position in the unit cell. Thus it is possible to locate the known fragment by fitting the calculated to the observed vector peak distribution. This is the principal strategy of the so-called Patterson-search methods, which can be subdivided into procedures operating in direct (Huber, 1965, 1969; Nordman, 1966, 1972) and reciprocal space (Rossmann & Blow, 1962; Rossmann, Blow, Harding & Coller, 1964; Tollin & Cochran, 1964; Tollin, 1966; Crowther & Blow, 1967).

Although these methods have been quite successful in solving difficult small-molecule structures (for more recent examples see Craven, 1976; Go, Kartha & Chen, 1980; Shieh, Hoard & Nordman, 1981; Admiraal & Vos, 1983), they are rarely used except in protein crystallography (Rossmann, 1972). The main reason for this is probably that the existing computer programs are not as automatic as the highly optimized direct-methods program packages such as *MULTAN* and *SHELX* but require extensive user intervention; this prevents them from being used in spite of their considerable effectiveness and flexibility. In order to demonstrate how powerful a tool Patterson search can be, we have determined six unknown structures by a real-space vector search using a program written by Hornstra (Braun, Hornstra & Leenhouts, 1969). Despite intensive efforts, none of them had been solved by direct methods* and most of them had resisted experienced crystallographers for years. The structure determinations, the results of which are or will be published elsewhere, are described here as examples of the conditions under which the method is likely to be successful and also of its scope and limitations.

Solution of six unknown structures

Relevant data for the six crystal structures depicted in Fig. 1 are listed in Table 1. For each a sharpened Patterson function with coefficients $|EF|$ was calculated and stored in a four-valued code with two bits of computer memory representing one Patterson value. The search program used is not as sophisticated as that of Nordman (1966, 1980) but the main features are common to both. It operates in three stages: rotation search, translation search, and final simultaneous optimization of the orientation and position of the fragment, without changing its internal geometry. In all three, possible solutions, *i.e.* those without too close contacts, are sorted according to a figure of merit which is based on the assumption that the model consists of equal atoms.

Structure (I), a steroidal lactone with 47 non-hydrogen atoms, presented the ideal case not only because the data were good, but also because the crystal structure of a diastereoisomer provided an accurate model (Egert, Cruse & Kennard, 1983). A Patterson search with 25 atoms (the maximum number allowed) clearly revealed the correct solution which had by far the best figure of merit at all stages. The structure model was easily completed by a Karle tangent expansion. Similarly, excellent data together with a well-defined search model taken from a closely related molecule (Restivo, Bryan & Kupchan, 1973) made the determination of the 65-atom structure of the tetracyclic triterpene derivative datiscoside C (II) straightforward (Sasamori *et al.*, 1983). In retrospect, there is no obvious explanation for the failure of direct methods with (I) and (II) where a promising fragment was never apparent in the various E maps.

The difficulty with the steroid (III) was the relative paucity of observed reflections mainly resulting from the small crystal size. Also the weak intensities were not available, so that the very useful figures of merit based upon them could not be applied to judge phase sets obtained from direct methods. A characteristic and accurate search model consisting of half the non-hydrogen atoms was available from a previously determined diastereoisomeric structure (Nassimbeni, Sheldrick & Kennard, 1974). Again the result of a Patterson search was convincing, with the best solution corresponding to the correct location of the fragment.

The chances of a successful application of vector search to the determination of structure (IV) (76 atoms) were less promising. The known fragment, a glucopyranose moiety (Jones, Sheldrick, Clegg, Kirby & Glenn, 1982), though very well defined, constituted only a small fraction of the contents of the asymmetric unit, so that interatomic distance tests were rather ineffective. It was therefore surprising when the obviously best solution could be rapidly expanded to the complete structure. Table 1 shows that only the third stage of the search procedure led to the correct answer, while the location of the second independent molecule, whose conformation also agrees well with the search model, was not found among the thirty best solutions. After refinement it was noticed that the intensities of six strong reflections were totally wrong (misplaced beam stop!) and thus had led to problems in direct phasing in spite of the otherwise good-quality data. Correcting these intensities resulted in a successful application of direct methods.

Another difficulty was encountered with the structure determination of the antibiotic γ-naphthocyclinone (V), whose molecular shape resembles a roof formed by two planar ring systems (Egert, Noltemeyer, Sheldrick, Saenger, Brand & Zeeck, 1983). The data were poor and no adequate search model could be found in the literature. We therefore calculated the

* It should be mentioned that the new *SHELX*84 program has subsequently solved four of the six structures (G. M. Sheldrick, personal communication).

geometry by the semiempirical force-field program *PIMM* (Lindner, 1974) and used the central part of the molecule as search fragment. The result was not as clear-cut as with structures (I)–(IV) but the figure of merit for the correct solution (ranked as number three) was only slightly worse than for the two best. Later examination revealed that the calculated model deviated considerably from the refined structure mainly

Table 1. *Data on the determination of six crystal structures by Patterson search*

	(I)	(II)	(III)	(IV)	(V)	(VI)
Molecular formula	$C_{37}H_{49}N_3O_7$	$C_{40}H_{58}O_{13}\cdot 2C_6H_6$	$C_{24}H_{28}FN_3O_6\cdot C_3H_6O$	$C_{20}H_{22}N_2O_{14}\cdot 0{\cdot}5C_3H_6O$	$C_{35}H_{30}O_{14}$	$C_{23}H_{20}O_{11}\cdot 0{\cdot}5CH_3OH$
Space group	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$	$P2_1$	$P2_12_12_1$
Z	2	4	2	8	2	8
Number of observed reflections ($F > 3\sigma$)	2650	2904	727	6010*	1422	3003
Model size (fraction of total number of non-hydrogen atoms)	25 atoms (0·53)	25 atoms (0·38)	19 atoms (0·50)	11 atoms (0·14)	25 atoms (0·51)	25 atoms (0·36)
Rank of correct solution after						
(*a*) rotation search	1	1	1	6	2	1
(*b*) translation search	1	1	2	24	21	3
(*c*) vector refinement	1	1	1	1	3	6
Final R value	0·051	0·038	0·058	0·053	0·092	0·079
R.m.s. deviation of model from final coordinates (Å)	0·11	0·12	0·11	0·06	0·51	0·29

* $F > 4\sigma$.

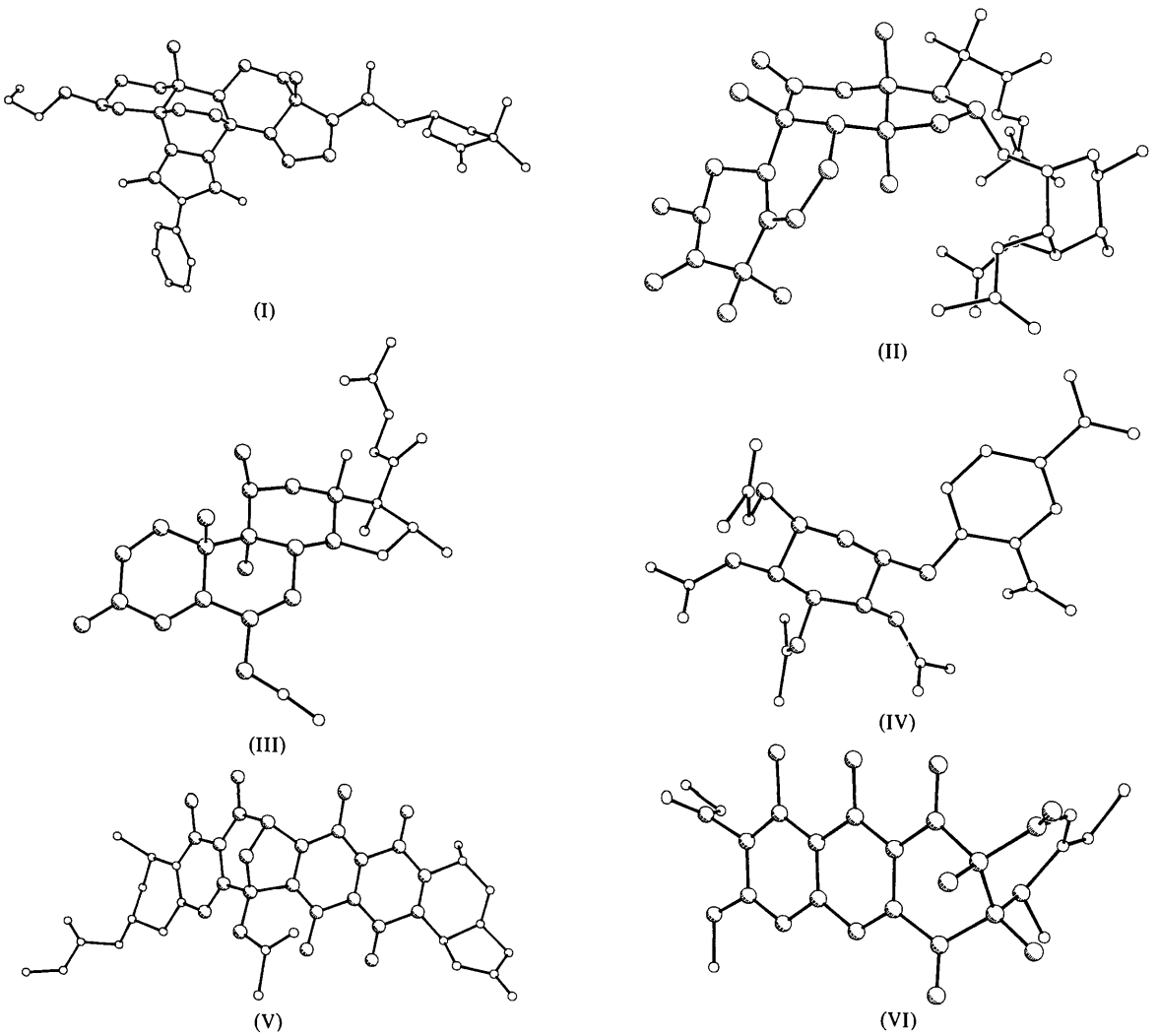

Fig. 1. The molecular conformations observed for (I)–(VI) (H atoms omitted). The atoms defining the respective search model are represented by large shaded circles. For (IV) and (VI), only the molecules successfully located are shown.

because the central fold angle had been overestimated by some 10°. Nevertheless, the vector pattern was obviously characteristic enough to enable the location of the fragment so close to its correct position that the initial phases could be successfully refined.

The limits of the Patterson search program used became evident with the structure solution of the antibiotic tetracenomycin C (VI) which essentially consists of four condensed six-membered rings. Most atoms of the search fragment (again calculated by the program *PIMM*) were coplanar, thus leading to an approximately two-dimensional vector pattern. From the very strong (since highly overlapped) intramolecular vectors within the aromatic rings, the orientation could be determined correctly but the subsequent translation search was not very successful with one correct solution ranking as number six. This may be due to the inability of the Hornstra program to handle several search fragments simultaneously; thus packing criteria were not very effective giving each of the two independent molecules too much space in the unit cell. It was noticed that the best 23 solutions had similar orientations and that [as with structure (IV)] there was no solution for the second molecule among the fifty calculated, although its deviation from the search model (0·25 Å) was even less than for the one correctly located. Direct methods had also yielded a correctly oriented two-dimensional fragment, but attempts to expand it in space group $P1$ were not successful. In cases like this, the best way to find the true location of a misplaced fragment seems to be through the use of a translation function operating in reciprocal space.

Discussion

What can we learn from the above results?

Considering the size of the structures studied and the problems which their solution presented to direct methods, the results of the Patterson searches were strikingly successful when good experimental data were combined with a reliable search model, even if the latter was rather small [as for structure (IV)]. This clearly shows the value of additional chemical information. Chemical knowledge is almost essential in cases like (V) where limited experimental data lead to only few and weak phase relationships. Generally speaking, poor data tend to recommend the Patterson method since the main features of the Patterson map are less sensitive to, for example, limited resolution than the calculated E values and the number of reliable phase relationships. Structures (V) and (VI) demonstrate that even calculated model geometries, which tend to be less accurate than those taken from related crystal structures, can be sufficient for a successful solution. Force-field methods are able to generate reliable geometries of complex molecules very quickly (Osawa & Musso, 1983).

The study of the above structures confirms the common experience that the application of the tangent formula or related procedures is very useful in extending a correctly placed molecular fragment to the complete structure. In other words, direct methods usually yield the correct solution once they have been provided with approximate initial phases. In this respect, Patterson search is one possible method of obtaining a good starting point. Since tangent expansion is able to improve the input geometry as long as the corresponding phase errors are not too large or too systematic, difficult structures can be solved even with less-accurate search models; structure (V), with a r.m.s. deviation between search fragment and refined structure as large as 0·51 Å, was completed without difficulty. Instead of continuing the structure determination with Patterson methods where errors are likely to be accumulated, it is far better to switch to direct methods after the successful location of a search fragment. This approach obviously uses the inherent strengths of both methods and thus leads to a safe and rapid structure solution.

Why does Patterson search deserve a revival?

Although formerly popular, Patterson search has now been completely superseded by direct methods; there are, however, good reasons to keep such a tool available. First of all, every progress in structure solution strategies has stimulated the investigation of still larger problems, so that there will always be structures that cannot be solved routinely. The use of chemical information is certainly a very promising approach, especially with the availability of a large amount of structural data accumulated in the Cambridge Crystallographic Database (Allen *et al.*, 1979), which provides accurate search models for most organic problems. This makes the application of Patterson search much more general than twenty years ago.

Secondly, with the advent of high-intensity X-ray sources with variable wavelengths it will be possible to tackle much larger structural problems at atomic resolution than hitherto, thus closing the gap between small-molecule and protein crystallography. This will be especially important for studies of biologically active compounds; for these, chemical information has already been used for structure determination (Rossmann, 1972; Shakked, Rabinovich, Cruse, Egert, Kennard, Sala, Salisbury & Viswamitra, 1981). A further strategy of considerable potential is provided by anomalous scattering of phosphorus and sulphur in nucleic acids and proteins (Hendrickson & Teeter, 1981). Here Patterson search could be useful to locate

the anomalous scatterers in cases where part of the geometry is known (single and double helices, β sheets *etc.*).

When other methods do not succeed, Patterson search can also be applied to structure determinations of inorganic compounds such as metal clusters or complexes, if the atomic arrangement is known. The additional chemical information is particularly useful in order to eliminate spurious symmetry elements introduced by heavy atoms in non-centrosymmetric space groups. Finally, the Patterson search procedure, especially when it operates in direct space, is clear and easy to understand. The user is thus able to make logical decisions after inspection of the results, which is not always true for statistical methods. Of course, direct methods should generally be used first because they are more automatic and do not depend on the correctness of a search model, which might be based on a wrong chemical assumption.

What are the requirements for a modern Patterson search program?

Notwithstanding the inherent advantages of Patterson search certain conditions must be met by a computer program for it to come into general use. Besides many other desirable features these are:

(1) It should be valid for all space groups.

(2) It should use the full potential of the search model, *i.e.* different atom types ought to be treated differently.

(3) There should be no limit in model size.

(4) It should be able to handle several independent search fragments.

(5) It should be user friendly by employing default values and format-free input.

When we decided to design such a program it was clear that its only chance of becoming widely used was to make it compatible with a widespread program package. Apart from other advantages, tested routines are then available to generate the Patterson function and to expand the fragments. A Patterson search program compatible with the new *SHELX*84 program system is being developed and tested extensively (Egert & Sheldrick, 1984). In view of the encouraging preliminary results we hope that the method will be powerful enough to provide an alternative pathway to the solution of difficult structures.

I thank Professor G. M. Sheldrick for many stimulating discussions, Dr Z. Shakked for directing my attention to the Hornstra program and the Fonds der Chemischen Industrie for a Liebig-Stipendium.

References

ADMIRAAL, G. & VOS, A. (1983). *Acta Cryst.* C**39**, 82–87.

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.* B**35**, 2331–2339.

BRAUN, P. B., HORNSTRA, J. & LEENHOUTS, J. I. (1969). *Philips Res. Rep.* **42**, 85–118.

CRAVEN, B. M. (1976). *Nature (London)*, **260**, 727–729.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

EGERT, E., CRUSE, W. B. T. & KENNARD, O. (1983). *Acta Cryst.* C**39**, 95–99.

EGERT, E., NOLTEMEYER, M., SHELDRICK, G. M., SAENGER, W., BRAND, H. & ZEECK, A. (1983). *Liebigs Ann. Chem.* pp. 503–509.

EGERT, E. & SHELDRICK, G. M. (1984). In preparation.

FURUSAKI, A. (1979). *Acta Cryst.* A**35**, 220–224.

GO, K., KARTHA, G. & CHEN, J. P. (1980). *Acta Cryst.* B**36**, 1811–1819.

GORRES, B. T. & JACOBSON, R. A. (1964). *Acta Cryst.* **17**, 1599–1603.

HENDRICKSON, W. A. & TEETER, M. M. (1981). *Nature (London)*, **290**, 107–113.

HUBER, R. (1965). *Acta Cryst.* **19**, 353–356.

HUBER, R. (1969). *Crystallographic Computing*, edited by F. R. AHMED, pp. 96–102. Copenhagen: Munksgaard.

HULL, S. E., VITERBO, D., WOOLFSON, M. M. & SHAO-HUI, Z. (1981). *Acta Cryst.* A**37**, 566–572.

JIA-XING, Y. (1981). *Acta Cryst.* A**37**, 642–644.

JONES, P. G., SHELDRICK, G. M., CLEGG, W., KIRBY, A. J. & GLENN, R. (1982). *Z. Kristallogr.* **160**, 269–274.

LINDNER, H. J. (1974). *Tetrahedron*, **30**, 1127–1132.

MIGHELL, A. D. & JACOBSON, R. A. (1963). *Acta Cryst.* **16**, 443–445.

NASSIMBENI, L. R., SHELDRICK, G. M. & KENNARD, O. (1974). *Acta Cryst.* B**30**, 2401–2406.

NORDMAN, C. E. (1966). *Trans. Am. Crystallogr. Assoc.* **2**, 29–38.

NORDMAN, C. E. (1972). *Acta Cryst.* A**28**, 134–143.

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.

OSAWA, E. & MUSSO, H. (1983). *Angew. Chem.* **95**, 1–12.

RESTIVO, R. J., BRYAN, R. F. & KUPCHAN, S. M. (1973). *J. Chem. Soc. Perkin Trans.* 2, pp. 892–897.

ROSSMANN, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon and Breach.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

ROSSMANN, M. G., BLOW, D. M., HARDING, M. M. & COLLER, E. (1964). *Acta Cryst.* **17**, 338–342.

SASAMORI, H., REDDY, K. S., KIRKUP, M. P., SHABANOWITZ, J., LYNN, D. G., HECHT, S. M., WOODE, K. A., BRYAN, R. F., CAMPBELL, J., LYNN, W. S., EGERT, E. & SHELDRICK, G. M. (1983). *J. Chem. Soc. Perkin Trans.* 1, pp. 1333–1347.

SCHENK, H. (1983). *Recl J. R. Neth. Chem. Soc.* **102**, 1–8.

SHAKKED, Z., RABINOVICH, D., CRUSE, W. B. T., EGERT, E., KENNARD, O., SALA, G., SALISBURY, S. A. & VISWAMITRA, M. A. (1981). *Proc. R. Soc. London Ser. B*, **213**, 479–487.

SHIEH, H.-S., HOARD, L. G. & NORDMAN, C. E. (1981). *Acta Cryst.* B**37**, 1538–1543.

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

TOLLIN, P. (1966). *Acta Cryst.* **21**, 613–614.

TOLLIN, P. & COCHRAN, W. (1964). *Acta Cryst.* **17**, 1322–1324.