selection of E's. The more random starts one makes the better the chance of finding the solution!

Another characteristic of phase developing and refining methods, such as the tangent formula or linear equations, is that even when wrong phase sets are obtained, an E map from the phases will reveal a molecular fragment. Particular sets of E's may relate most strongly to particular parts of the structure and if the relative phases of the E's are correctly deduced then the corresponding portion of the molecule will be seen although it may be incorrectly positioned.

This is probably the basis of the success of the symbolic addition method. If one is prepared to make several starts then within a few trials a set of phases giving a fragment may be found. With such a fragment an extremely efficient process of structure development may be initiated (*e.g.* Karle, 1968).

It should not be thought that we are saying that *all* structures are solved by a pseudo random approach. For simple structures the *MULTAN* development process can start from a best set of phases such that as the phase development proceeds so the phases being developed are close to their correct values. Again, if techniques are employed which can give reliable estimates for the values of structure-invariant quantities, then the complexity of structure for which a solution is obtained by a 'non-random' pathway may well be greater.

The question does nevertheless present itself – with large and fast computers available and with hardware such as array processers becoming ever more common – how worthwhile is it to develop intricate and subtle techniques when a bull-at-the-fence process may be enough?

We are grateful to the Computer Science department of the University of York and the Centre de Calcul of the University of Louvain for the generous provision of computing facilities. Financial support by the Science Research Council and by the British Council is also acknowledged.

References

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473–478.
COLENS, A., DECLERCQ, J. P., GERMAIN, G., PUTZEYS, J. P. & VAN MEERSSCHE, M. (1974). *Cryst. Struct. Commun.* **3**, 119–122.
DECLERCQ, J. P., GERMAIN, G. & WOOLFSON, M. M. (1975). *Acta Cryst.* A**31**, 367–372.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.* B**26**, 274–285.
HULL, S. E., LEBAN, I., MAIN, P., WHITE, P. S. & WOOLFSON, M. M. (1976). *Acta Cryst.* B**32**, 2374–2381.
KARLE, J. (1968). *Acta Cryst.* B**24**, 182–186.
KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–868.
MAIN, P. (1978). *Acta Cryst.* A**34**, 31–38.
WHITE, P. S. & WOOLFSON, M. M. (1975). *Acta Cryst.* A**31**, 53–56.
WOOLFSON, M. M. (1977). *Acta Cryst.* A**33**, 219–225.

Acta Cryst. (1978). A**34**, 892–894

The Use of Karle–Hauptman Determinants in Small-Structure Determinations

BY W. J. VERMIN AND R. A. G. DE GRAAFF

X-ray Department, Gorlaeus Laboratories, PO Box 9502, 2300 *RA Leiden, The Netherlands*

(*Received* 1 *February* 1978; *accepted* 4 *May* 1978)

This paper illustrates the potential of Karle–Hauptman determinants for the solution of the phase problem in a 'difficult' small structure. An outline of the procedure followed is given. Suggestions for the future are presented, together with a discussion of the results obtained.

Introduction

Efforts to solve small and medium structures (as opposed to large protein structures) by direct methods have centred on the $\sum_2$ relation. Attempts to include relations among more than three phases have met with some initial success (Gilmore, 1977; Gilmore, Hardy, MacNicol & Wilson, 1977; Blank, Rodrigues, Pletcher & Sax, 1976; Sax, Rodrigues, Blank, Wood & Pletcher, 1976). Quartets and quintets have yielded efficient criteria for discriminating between solutions obtained by $\sum_2$-based processes (Schenk, 1973*a*,*b*). The empirical formulae obtained by Schenk have been supplemented by theory (Hauptman, 1974*a*,*b*, 1975*a*,*b*; Giacovazzo, 1974, 1975; Heinerman, 1975).

A totally different approach to the problem was

much earlier suggested by Karle & Hauptman (1950). The method involves the use of the so called Karle–Hauptman determinants.

An important property of the Karle–Hauptman determinants has been found by Goedkoop (1950). He demonstrated that the value of the determinant decreases with its order, reaching zero if the order of the determinant is greater than the number of independent atoms in the unit cell. Tsoucaris (1970*a*,*b*) published his well-known 'maximum-determinant rule', *i.e.* the most probable set of phases connected with the reflexions in a certain Karle–Hauptman determinant yields the maximum value of this determinant. The meaning of this in terms of $\sum_2$, $\sum_3$ *etc.* relations is: all possible relations between a set of phases of a certain specific group of reflexions must be satisfied as well as possible. The relative weights of these relations decrease with the order of the relation by a factor equal to E_{000}. The rule has been applied successfully to the extension of phases from medium to high resolution in the structure determination of proteins (de Rango, Mauguen & Tsoucaris, 1975).

Suggestions have been made to use the method for small structures (Woolfson, 1977). We have applied the maximum-determinant rule to the structure determination of pyrocalciferol ($P2_1$, $Z = 2$, $N = 86$). The structure has been determined by Patterson search methods (de Kok & Romers, 1975). The reflexions were measured at −170 °C using Mo $K\alpha$ radiation and a graphite monochromator. 4435 intensities were recorded, 3302 of which were significant. The maximum glancing angle was 29°. The final weighted R value based on the significant reflexions only was 4·86%. Attempts to use conventional multisolution programs have failed, magic integers (White & Woolfson, 1975; Declercq, Germain & Woolfson, 1975) gave about 40% of the structure, enough to solve the phase problem. The reason for the difficulties encountered in solving this fairly simple structure is the breaking down of the $\sum_2$ relation for a significant number of very strong triple products. However, if a sufficiently large number, about 20, of strong reflexions are known, tangent refinement does not diverge.

Building the determinant

We have developed a routine to build a Karle–Hauptman determinant using all the available E values. A not significant or unobserved reflexion is assigned the value 0. A reflexion lying outside the sphere of measurement is assigned the value 1. The program attempts to include as many strong reflexions as possible, while at the same time keeping the number of independent reflexions to a minimum. The weights of these conflicting requirements can be varied according to specification. In this way we have built a 20 × 20 determinant containing 79 independent reflexions, 23 of which belong to the group of the 200 strongest E's.

Solving the structure

The origin was fixed by giving three general reflexions their correct phases. Three more reflexions were selected from the determinant to act as symbolic reflexions (exactly as in the multi-solution procedure). The choice of origin-defining and symbolic reflexions was made on the basis of the value of E as well as the number of times the reflexion occurred in the determinant. The three symbolic reflexions were strongly interrelated; moreover, one of the symbols was a special reflexion. This is the reason why only four solutions had to be tried. All other phases were initially set equal to zero.

The determinant is optimized by the following procedure: (1) Origin and symbols are fixed at their current values. (2) All other phases in the left uppermost 7 × 7 determinant are allowed to vary until convergence. (3) A new row and column are added and only the new phases (initially set equal to zero) are allowed to vary until convergence. (4) Next, the full determinant is optimized with respect to the phases; the symbols are allowed to vary as well. The determinants are optimized using a modified Newton–Raphson

Table 1. *The starting reflexions*

Phases are in millicycles (m.c.): $|\Delta\varphi| = |\varphi_{\text{struct.}} - \varphi_{\text{determinant}}|$, $|\Delta\varphi|_{\text{av}} = 80$ m.c.

$h\ k\ l$	$\varphi_{\text{struct.}}$	$\varphi_{\text{determinant}}$	$\varphi_{\text{tgref.}}$	$\lvert\Delta\varphi\rvert$	E
1 1 2*	282	282	280	0	3·62
10,0, $\bar{2}$†	500	500	500	0	3·12
9 $\bar{1}$ $\bar{4}$†	244	228	200	16	2·72
0 2 0†	51	44	897	7	2·34
$\bar{1}$0,2, 2	557	544	526	13	2·11
6 1 7*	832	832	831	0	2·39
$\bar{3}$,2,11*	580	580	579	0	2·67
7 0 9	0	0	0	0	1·89
$\bar{8}$ 2 6	134	125	96	9	1·91
$\bar{7}$ 9 2	924	778	619	146	1·91
$\bar{7}$ 7 2	870	693	761	177	2·14
$\bar{6}$ 8 4	69	986	922	83	2·36
2 6 $\bar{2}$	95	879	16	216	1·90
2 2 $\bar{4}$	525	613	599	88	2·12
5 $\bar{7}$ $\bar{6}$	749	793	9	44	2·46
7 $\bar{5}$ $\bar{2}$	153	375	321	222	2·31
$\bar{4}$ 6 8	256	403	141	147	2·01
10,$\bar{1}$, $\bar{4}$	350	140	464	210	2·45
11,$\bar{2}$, $\bar{2}$	537	317	738	220	1·99
2 $\bar{1}$ 2	958	52	965	94	2·27
3 $\bar{2}$ 4	153	267	131	114	2·18
7 $\bar{8}$ $\bar{4}$	861	872	48	11	1·88
9 0 6	0	73	500	73	1·83

* Origin-defining reflexion.

† Symbol. Starting values of symbols are: 10,0,$\bar{2}$ 500, 9$\bar{1}\bar{4}$ 220 follows from 112 and 10,0,$\bar{2}$, 020 64 follows from 112.

method as supplied by the Nottingham Algorithm Group (Ford, 1977). During optimization the determinant is kept positive-definite.

The derivatives of the determinants with respect to the parameters are evaluated numerically.

The phases of the 23 strongest reflexions obtained from the determinant were used as a starting point for a weighted tangent refinement. An E map based on the fourth solution contained 41 out of 43 atoms of the molecule. The 23 starting phases are listed in Table 1. The mean phase error was 80 millicycles before and 111 millicycles after tangent refinement.

Discussion

Inspection of Table 1 shows that while some phases are calculated with surprising accuracy others are quite wrong. Furthermore it may be argued that by giving the three origin-defining reflexions their correct phases we have provided the process with a better starting point than is generally available in a structure determination. The result obtained nevertheless convincingly illustrates the potential of the Karle–Hauptman determinants.

As Tsoucaris has pointed out, the 'quality of the determinant' is the key factor in the process of structure determination. The meaning of this term is not rigorously defined. Our findings have confirmed Tsoucaris's idea that a determinant of high quality will have a value as close to zero as is practical, while the number of independent phases will be small in relation to the order of the determinant. By improving the algorithm designed to construct the determinant as well as increasing the size of the said determinant the problems outlined in the beginning of this section may well be avoided. The results we have obtained lead us to believe that Karle–Hauptman determinants will play a significant role in the determination of structures eluding solution by $\sum_2$-based processes.

The future

We are planning to replace the numerical calculation of the derivatives by an analytical one, taking advantage of the fact that $\partial \det \mathbf{A}/\partial a_{ij} = b_{ij} \det \mathbf{A}$, where $\mathbf{A}$ is a general $n \times n$ matrix, a_{ij} are the elements of $\mathbf{A}$ and b_{ij} are the elements of the inverse of $\mathbf{A}$. This will result in a speeding up of the optimization process with a factor equal to the number of parameters refined.

We then hope to be able to obtain more information on the essence of the term 'quality of the determinant'. Optimization of a 30×30 determinant will be a matter of minutes instead of hours. We hope this will enable us to tackle structures which have up to now eluded solution by direct methods. Our aim is to construct a determinant for a given structure, which after optimization contains enough correct phase information to calculate an E map directly, without resorting to the tangent formula.

Our gratitude is owed to Professor C. Romers for his continuing stimulating interest, and to Dr J. van der Griend for his helpful suggestions concerning the mathematics of the optimization process. The calculations were performed on the 370/158 IBM computer of the Central Computing Centre of Leiden University and on the CDC 6600 computer of the ECN in Petten.

References

BLANK, G., RODRIGUES, M., PLETCHER, J. & SAX, M. (1976). *Acta Cryst.* B**32**, 2970–2975.
DECLERCQ, J. P., GERMAIN, G. & WOOLFSON, M. M. (1975). *Acta Cryst.* A**31**, 367–372.
FORD, B. (1977). *NAG Fortran Library Manual Mark V*, chs. EO4 and EO4 CDF. Nottingham Algorithm Group.
GIACOVAZZO, C. (1974). *Acta Cryst.* A**30**, 481–487.
GIACOVAZZO, C. (1975). *Acta Cryst.* A**31**, 602–609.
GILMORE, C. J. (1977). *Acta Cryst.* A**33**, 712–716.
GILMORE, C. J., HARDY, A. D. U., MACNICOL, D. D. & WILSON, D. R. (1977). *J. Chem. Soc. Perkin Trans.* 2, pp. 1427–1434.
GOEDKOOP, D. (1950). *Acta Cryst.* **3**, 374–378.
HAUPTMAN, H. (1974*a*). *Acta Cryst.* A**30**, 472–477.
HAUPTMAN, H. (1974*b*). *Acta Cryst.* A**30**, 822–829.
HAUPTMAN, H. (1975*a*). *Acta Cryst.* A**31**, 671–679.
HAUPTMAN, H. (1975*b*). *Acta Cryst.* A**31**, 680–687.
HEINERMAN, J. J. L. (1975). *Acta Cryst.* A**31**, 727–730.
KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181–187.
KOK, A. J. DE & ROMERS, C. (1975). *Acta Cryst.* B**31**, 1535–1542.
RANGO, C. DE, MAUGUEN, Y. & TSOUCARIS, G. (1975). *Acta Cryst.* A**31**, 227–233.
SAX, M., RODRIGUES, M., BLANK, G., WOOD, M. K. & PLETCHER, J. (1976). *Acta Cryst.* B**32**, 1953–1956.
SCHENK, H. (1973*a*). *Acta Cryst.* A**29**, 72–82.
SCHENK, H. (1973*b*). *Acta Cryst.* A**29**, 480–483.
TSOUCARIS, G. (1970*a*). *Acta Cryst.* A**26**, 492–499.
TSOUCARIS, G. (1970*b*). *Acta Cryst.* A**26**, 499–501.
WHITE, P. & WOOLFSON, M. M. (1975). *Acta Cryst.* A**31**, 53–56.
WOOLFSON, M. M. (1977). *Acta Cryst.* A**33**, 219–225.