

quired. Needless to say, practical testing of both methods indicated in the present paper is needed. In our laboratory, a computing program for the least-squares approach to triclinic space groups is under preparation.

References

ANZENHOFER, K. & HOPPE, W. (1962). *Phys. Verh.* **13**, 119.

DIDEBERG, O. (1977). *Acta Cryst.* **A33**, 520–521.

HOPPE, W. (1962). *Naturwissenschaften*, **49**, 536–537.

HOPPE, W. (1963). *Acta Cryst.* **16**, 1056–1057.

KARLE, J. (1972). *Acta Cryst.* **B28**, 3362–3369.

MAIN, P. & WOOLFSON, M. M. (1962). *Phys. Verh.* **13**, 118.

MAIN, P. & WOOLFSON, M. M. (1963). *Acta Cryst.* **16**, 1046–1051.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Acta Cryst. (1979). **A35**, 220–224

A New Monte Carlo Method for Phase Determination

BY AKIO FURUSAKI

Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060, Japan

(Received 17 July 1978; accepted 6 September 1978)

Abstract

An extended version of the multi-solution method has been devised by introducing a Monte Carlo technique. This Monte Carlo direct method differs from the ordinary multi-solution procedure in two respects: (1) The starting set usually consists of as many as 10–50 phases; (2) Tentative phase values assigned to the members of the starting set are derived from successively generated random numbers. The application of the new method to several unknown structures has shown that it can be used as an effective means of phase determination.

Introduction

The direct methods of phase determination have now made it possible to determine unknown structures of complicated organic compounds without converting them into heavy-atom derivatives. These excellent results may give one a false impression that the phase problem would have been completely solved. However, especially in noncentrosymmetric cases, there seem to be a good many structures whose analyses end in failure. It has been pointed out that the probable cause of such a failure is that some phase relationships produce grave errors early in the phase determination (Karle, Karle & Estlin, 1967; Karle, Gibson & Karle, 1969).* If this is the case, it is very hard to overcome the difficulty, because it is practically impossible to find

the serious trouble-makers among the many phase relationships at the early stage of the analysis. Although a promising way of doing this is to add some additional phases to the starting set, it causes a great increase in the number of phase sets to be treated. A method which has been developed to surmount this new obstacle is the magic-integer technique (White & Woolfson, 1975; Declercq, Germain & Woolfson, 1975).

When applying the multi-solution method (Germain & Woolfson, 1968), one sometimes finds that two or more phase sets appearing in succession lead to the correct solution. In this case, generally speaking, one will be able to reach the correct solution earlier by making the phase sets appear in completely random order rather than in systematic order, because the dispersion of the successive, essentially correct phase sets remarkably decreases the average number of trials necessary for finding the correct solution. The larger the starting set, the more frequently such a case must be met. Accordingly, the introduction of a Monte Carlo technique seems to afford a new method for solving the difficulty in phase determination.

Although the use of a Monte Carlo method for X-ray structure analysis was proposed by Vand, Niggli & Pepinsky (1960), the idea now appears to have been given up. One of the main causes of the failure is probably that the optimal-shift method (Niggli, Vand & Pepinsky, 1960) cannot refine random structures sufficiently well. On the other hand, the tangent formula (Karle & Hauptman, 1956) often shows remarkable ability in the refinement of phases. This suggests that it may be more promising to use generated random

* It has recently been reported that there exist structures which the tangent formula itself is inadequate to solve (Lessinger, 1976).

numbers as phase angles rather than as atomic coordinates.

In view of the foregoing circumstances, the author has now attempted to develop a new phase-determining method, hereafter referred to as the Monte Carlo direct method.

Basis of the method

As an example, let us consider a set of M two-dimensional and $N - M$ general reflections in space group $P2_12_12_1$. If we assume that every general reflection may arbitrarily take one of the eight discrete phase values $0, \pm\pi/4, \pm\pi/2, \pm3\pi/4$ and π , the total number of possible sets of phase values is 2^{3N-2M} for all reflections. Of these phase sets, sixteen correspond to the correct structure, provided that those sets which can uniquely specify the origin and enantiomorph are included among the N reflections. Therefore, the probability of hitting on any one of the correct phase sets is given by $P = 1/2^{3N-2M-4}$. This probability falls off rapidly as N increases; for example, $P \sim 10^{-5}$ for $N = 10, M = 5$ and $P \sim 10^{-11}$ for $N = 20, M = 10$. Such low values may throw doubt upon the practicability of the Monte Carlo direct method. However, the actual value for the probability of obtaining the correct structure must be considerably higher than P , because, in addition to the sixteen correct phase sets, there will be a considerable number of phase sets which can lead to the correct structure by the aid of the tangent formula. It is probably impossible to derive this actual probability, p , in a theoretical manner. Therefore, the author has initially tried to obtain the values of p for the following known structures empirically: (1) coronafacic acid (hereafter CFA), $C_{12}H_{16}O_3, P2_12_12_1, Z = 4$ (Ichihara, Shiraishi, Sato, Sakamura, Nishiyama, Sakai, Furusaki & Matsumoto, 1977); and (2) grayanotoxin XV (hereafter GXV), $C_{20}H_{30}O_4, P2_12_12_1, Z = 4$ (Hamanaka, Miyakoshi, Furusaki & Matsumoto, 1972).

For the N reflections to which random phase values were assigned, those with the largest E values were chosen in all cases except one; for $N = 7$ in CFA the

necessity of specifying the origin did not allow such a choice. For each of the random phase sets, 8–10 cycles of the tangent iteration were performed using E values >1.30 , and then the resulting phase set was compared with the known correct set. This procedure was repeated until 3–7 correct phase sets appeared. The average number of trials necessary for finding a correct phase set (that is, the value of $1/p$) has thus been obtained. The results are shown in Fig. 1.

First, it should be noted that, in general, the values for $1/p$ are much smaller than could be expected: 15–40 for CFA and 22–66 for GXV. This shows that, compared with the conventional direct methods, the present procedure will not always entail great cost. It is also surprising to see that, over the range of N values up to 40 or 60, the value of $1/p$ increases with N approximately linearly, not exponentially. From these two examples, the slope of a straight line expressing the relationship between N and $1/p$ does not generally seem to be very steep, although it may vary more or less from structure to structure. Consequently, even though one adopts a starting set of several tens of phases, one can continue the phase determination without great disadvantage. This is very important, because the use of such a large starting set may make it possible to surmount various difficulties which the symbolic addition and multi-solution methods (Karle & Karle, 1966; Germain & Woolfson, 1968) have encountered.

Let us examine a little more closely the case when $N = 30, M = 15$ for CFA. In this case, from the value of 23 for $1/p$, it follows that of the 10^{18} possible phase sets no less than 5×10^{16} afford the correct solution. Table 1 gives seven of these essentially correct phase sets.

For the sake of comparison, the phase sets have been adjusted so as to correspond to the same origin and enantiomorph. As can be seen in Table 1, in most of the seven phase sets, even the five strongest reflections are not always assigned phase values close to their correct ones. It is interesting that, as the tangent procedure is repeated, every one of these random phase sets approaches the correct phase set given in Table 1.

Procedure

The Monte Carlo direct method resembles the multi-solution procedure, but differs from the latter in two respects: (1) The starting set is usually made up of as many as 10–50 phases; (2) Tentative phase values assigned to the members of the starting set are derived from successively generated random numbers. Some points to which special attention should be paid in the present procedure are mentioned below.

Although the present method requires that the starting set should contain reflections appropriate to

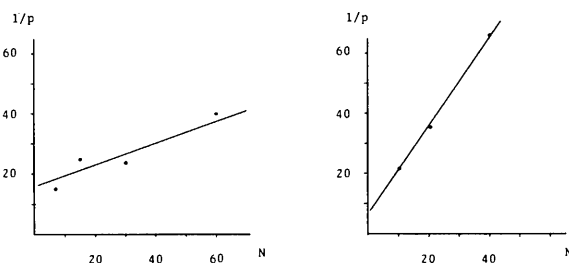


Fig. 1. The average number of trials necessary for finding the correct structure. (a) CFA. (b) GXV.

Table 1. Seven essentially correct phase sets for $N = 30, M = 15$ in CFAPhase values are in $10^{-3} \pi$. Correct phases were calculated with the final atomic parameters.

	h	k	l	$ E $	Correct	No. 10	No. 38	No. 51	No. 116	No. 126	No. 162	No. 163
1	3	14	1	3.54	1022	1000	1500	1000	1000	1500	750	1000
2	7	6	0	3.20	500	1500	500	1500	500	500	500	500
3	0	12	6	3.12	1000	1000	1000	1000	1000	0	0	0
4	7	6	2	3.08	286	1250	250	250	500	500	750	0
5	3	6	1	3.06	1960	0	0	250	1750	1750	0	1000
6	0	1	4	2.91	500	500	500	500	500	500	500	500
7	6	8	1	2.90	1565	1750	1250	1750	1500	1250	1500	0
8	0	13	7	2.89	500	500	500	500	500	500	1500	500
9	4	0	1	2.82	500	500	500	500	1500	500	500	500
10	1	1	8	2.82	483	500	750	1250	1750	750	500	1000
11	5	11	1	2.64	73	750	250	1750	1750	0	250	0
12	0	1	2	2.60	500	500	500	1500	500	1500	500	500
13	10	3	1	2.60	1461	1250	0	1000	1750	1250	1500	1250
14	6	8	0	2.59	0	0	0	1000	1000	0	1000	0
15	5	9	3	2.59	310	1000	750	500	1250	0	750	250
16	8	13	0	2.58	1000	1000	0	0	1000	0	1000	1000
17	1	14	0	2.58	500	500	500	500	500	500	1500	500
18	1	14	1	2.57	972	1000	1250	1250	1250	1750	750	1000
19	5	9	1	2.47	146	0	500	250	0	1750	500	250
20	7	5	0	2.42	1500	1500	1500	1500	1500	500	1500	1500
21	7	2	1	2.41	24	1000	750	250	1750	1750	250	250
22	6	7	4	2.38	577	1000	750	500	0	500	250	500
23	3	16	1	2.38	1087	1250	0	1250	1500	1000	1000	1000
24	7	3	0	2.36	1500	500	1500	500	1500	500	1500	1500
25	7	6	4	2.34	164	1750	1000	0	0	0	500	0
26	10	4	0	2.34	1000	0	1000	0	0	0	0	1000
27	2	15	3	2.33	77	0	500	250	0	1500	0	500
28	0	3	4	2.32	500	500	1500	500	1500	500	500	500
29	7	4	0	2.32	500	500	1500	500	500	1500	500	500
30	8	5	0	2.23	1000	1000	0	0	0	0	1000	0

specify the origin and enantiomorph (Karle & Hauptman, 1961), there is no necessity for assigning definite phase values to these reflections. This is because the origin and enantiomorph can be defined naturally by accidental assignment of random phase values. Since there are usually such origin- and enantiomorph-specifying reflections among the first 10–20 strongest, one may choose the strongest reflections automatically for the starting set.

Random numbers can be easily obtained as values between 0 and 1 by utilizing an available computer program. Each of the projection reflections whose phases may only assume such special values as 0 or π and $\pi/2$ or $3\pi/2$ is assigned either of the two possible values according to the value of a given random number, x ; for example, the value assigned is 0 for $0 \leq x < 0.5$, whereas it is π for $0.5 \leq x < 1$. Although, in this study, the eight discrete values of 0, $\pm\pi/4$, $\pm\pi/2$, $\pm3\pi/4$ and π were used as possible phase values for general reflections, the use of such continuous quantities as $2\pi x$ also probably gives good results.

After a complete set of random phase values has been generated, in order to extend this phase set, about 10 cycles of the tangent procedure are performed using E values >1.20 – 1.50 . During the repetition of the first several cycles of the tangent procedure, the phase

values which have been assigned to the starting reflections are held constant in order that the random phase set may not lose its essential feature.

As in the case of the multi-solution method, here also, the question arises as to how the correct phase set can be distinguished from the others. As will be seen in the next section, the value of $R_K = \sum |E_o| - k|E_c|/\sum |E_o|$ (Karle & Karle, 1966) serves as a good guide for this distinction. After the tangent iteration, the R_K value of the resulting phase set is compared with those which have already been produced. When the R_K value shows a marked decrease or a marked tendency to decrease, several further cycles of the phase refinement are carried out, and then an E map is calculated and examined. Generally speaking, the larger the size of the starting set, the larger the number of cycles of the tangent procedure necessary for catching the correct phase set. If the R_K value is combined with another index, for example the q value defined by (Hoppe, Gassmann & Zechmeister, 1970)

$$q = 1 - \frac{\sum_h |E_h| \left\{ \sum_k |E_k E_{h-k}| \times \exp [i(\phi_k + \phi_{h-k})] \right\}}{\sum_h |E_h|}$$

it may also be useful in space groups, such as $P1$ and

C2, in which it becomes very small when all phase values vanish; for the correct phase set must not only have a low value of R_K but also a moderately high value of q . A concrete example in space group C2 is given in the next section.

Since many reflections are contained in the starting set, it is possible that there may be some phase relationships among them. These phase relationships can be taken into account in the following way. If the phase relationship among three phases φ_h , φ_k and φ_{h-k} is expressed by the equation $\varphi_h = \varphi_k + \varphi_{h-k} + \delta_{h,k}$, then the random variable $\delta_{h,k}$ must have the probability distribution (Karle & Karle, 1966)

$$f(x) = [2\pi I_0(\alpha)]^{-1} \exp(\alpha \cos x),$$

where I_0 is the zero-order modified Bessel function of the first kind, and $\alpha = 2(\sum z_j^2) (\sum z_j^2)^{-3/2} |E_h E_k E_{h-k}|$. In order to derive reasonable values for $\delta_{h,k}$ from uniformly distributed random numbers, it is first necessary to evaluate the following integrals which give approximate estimates for the probability that $\delta_{h,k}$ will assume the value of $n\pi/N$ ($n = 1 - N, 2 - N, \dots, N$):

$$S_n = \int_{(2n-1)\pi/2N}^{(2n+1)\pi/2N} f(x) dx.$$

Here it is assumed that $\delta_{h,k}$ takes any of the $2N$ discrete values of $n\pi/N$. Next, let us define X_m by $X_0 = 0$ and

$$X_m = \sum_{n=1-N}^{m-N} S_n, \quad m = 1, 2, \dots, 2N.$$

Since $X_{2N} = 1$, $2N - 1$ points at $x = X_1, X_2, \dots$,

X_{2N-1} divide the range $0 \leq x \leq 1$ into $2N$ regions proportional to S_{1-N}, \dots, S_{N-1} and S_N respectively. Therefore, if the value of a given random number, x , falls between X_{m-1} and X_m , it is reasonable to assign the value of $(m - N)\pi/N$ to the variable $\delta_{h,k}$. While the values for φ_k and φ_{h-k} are calculated directly from given random numbers in the usual way, that for φ_h is obtained by substituting the $\delta_{h,k}$, φ_k and φ_{h-k} values into the phase relationship $\varphi_h = \varphi_k + \varphi_{h-k} + \delta_{h,k}$.

Applications

Five new structures have quite recently been elucidated by the application of the Monte Carlo direct method. The crystallographic data for the five compounds and the main points of their structure determinations are summarized in Table 2.

For each of the five crystals, the intensity data with 2θ values up to 140° were collected on an automatic four-circle diffractometer, using Cu $K\alpha$ radiation monochromatized with a LiF crystal. After correction for the Lorentz and polarization factors, these intensities were converted into normalized structure factor magnitudes ($|E|^2$'s). All the phase determinations were carried out according to the procedure described in the preceding section. In every case, the strongest reflections were adopted for the starting set. Although the maximum number of pairs of φ_k and φ_{h-k} used for the evaluation of φ_h in the tangent procedure was limited to 50, this showed no signs of hindering the progress of the phase determination. In this manner, the five structures could be solved with little difficulty. Some of the results are shown in Table 2. With the intention of

Table 2. Crystal data and some key points of the structure determinations

Compound	(I)*	(II)	(III)	(IV)	(V)
Formula	$C_{10}H_{16}N_2O_5 \cdot 3H_2O$	$C_{20}H_{28}O_4$	$C_{20}H_{32}O_5$	$C_{28}H_{26}N_4O_3 \cdot CH_3OH$	$C_{22}H_{34}O_6 \cdot 0.5H_2O$
Space group	$P2_1$	$P2_12_1$	$P2_12_1$	C2	$P2_12_1$
a (Å)	10.499	9.500	10.040	23.487	17.503
b (Å)	12.812	21.524	26.517	7.636	37.933
c (Å)	5.322	8.758	6.672	15.638	6.371
β	$98^\circ 32'$			$116^\circ 43'$	
Z	2	4	4	4	8
Size of the starting set	20	10	20	20	30
E values used	331 (≥ 1.20)	380 (≥ 1.30)	421 (≥ 1.30)	487 (≥ 1.30)	932 (≥ 1.30)
Correct phase set	No. 54	No. 40	No. 3	No. 72	No. 261
Cycles of the tangent procedure	10 (16)	10	10 (12)	12	12 (20)
R_K value (%)	28.2 (18.8)	23.5	24.6 (24.1)	27.4	34.4 (27.6)
Phases obtained	306 (329)	367	406 (407)	458	796 (864)
Atoms found in the E map	20	24	25	33	56
Hydrogen atoms found in the D map	22	28	32	25	68
Data used for the refinement	1377	1871	1908	2532	4180
Final R value (%)	3.1	4.5	3.8	4.7	5.5

* The five compounds are: (I) palythine trihydrate; (II) $5\beta,6\beta$ -(dimethylmethylenedioxy)-15,16,17-trinorgrayan-10(20)-ene-3,14-dione; (III) grayanotoxin II; (IV) AM-2282 methanol solvate; (V) grayanotoxin XVI hemihydrate.

subjecting the present method to severe tests, in all these phase determinations no account was taken of phase relationships among the starting reflections. The phase determinations for (I)–(IV) were accomplished in moderate computing time; for instance, the time required for (IV) was about 8 min on a FACOM 230-75 computer at the Computing Center of Hokkaido University. On the other hand, the phase determination for (V) took about 1 h. However, it seems possible to reduce this computing time considerably by utilizing phase relationships among the starting reflections in the manner given in the preceding section. The details of the structures will be published separately. For (IV) and (V), which possess the most complex crystal structures, a more detailed description is given below.

Since (IV) belongs to space group $C2$, the most important problem in the phase determination was how to discriminate the correct phase set. As previously mentioned, values of R_K and q were used for this discrimination. As the calculation proceeded, it gradually became apparent that the phase sets with R_K values below 30% could be grouped into three classes with respect to their q values: (1) 0%, (2) 15–18% and (3) 36–40%. Of the 72 phase sets obtained, only two belonged to the last class. The 72nd phase set, which was one of the two, turned out to be correct. The resulting E map showed 33 of the 37 independent non-hydrogen atoms.

The structure determination of (V) involved a risk of getting into trouble, because there were as many as 57 non-hydrogen atoms in the asymmetric unit of this crystal. With the intention of diminishing this risk, a starting set of 30 phases was adopted. After 12 cycles of the tangent procedure, an R_K value for the 261st phase set showed a marked tendency to decrease, although it was still 34.4%. Eight additional cycles of the tangent iteration reduced the R_K value to 27.6%. An E map thus obtained yielded all the non-hydrogen atoms except one, which was later found to occur in two statistical positions.

We have so far confined our remarks to noncentrosymmetric cases. It should be stated here that the Monte Carlo direct method is also applicable to centrosymmetric structures. As examples of such a case, the structure determinations of the following two crystals are cited: (1) *o,p*-dinitrocinnamyl acetate, $C_{11}H_{10}N_2O_6$, $P2_1/c$, $Z = 4$; and (2) 3-(1-ethoxyaminopropylidene)-6-ethyl-2,3-dihydro-4H-pyran-2,4-dione, $C_{12}H_{17}NO_4$, $C2/c$, $Z = 8$. Each of these two phase determinations was based on a starting set of 10 reflections; in the latter case, not all these reflections were the strongest. Almost the same procedure as that for noncentrosymmetric structures was used. The 1st and 14th phase sets gave the respective correct solutions. The details will be given elsewhere. Thus, it has been proved that

the present phase-determining method is widely applicable.

Concluding remarks

If the starting set in the multi-solution procedure is employed in the present method just as it is, then these two methods come to the same, except that the order in which the phase sets are dealt with differs. Hence, the present method may be regarded as an extended version of the multi-solution procedure.

It has now been shown that the Monte Carlo direct method is an effective means of phase determination. If the free use of a large high-speed computer were permitted, even the structures of considerably complicated compounds might be solved by this method. Accordingly, as highly efficient computers become more available in the future, the present method will play a more important role in X-ray structure elucidation.

The author wishes to express his sincere thanks to Professors I. Nitta and T. Matsumoto for their continued support and encouragement. This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education (No. 254140), and in part by a subsidy from the Matsunaga Scientific Foundation.

References

- DECLERCQ, J. P., GERMAIN, G. & WOOLFSON, M. M. (1975). *Acta Cryst.* **A31**, 367–372.
- GERMAIN, G. & WOOLFSON, M. M. (1968). *Acta Cryst.* **B24**, 91–96.
- HAMANAKA, N., MIYAKOSHI, H., FURUSAKI, A. & MATSUMOTO, T. (1972). *Chem. Lett.* pp. 779–782.
- HOPPE, W., GASSMANN, J. & ZECHMEISTER, K. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 26–36. Copenhagen: Munksgaard.
- ICHIHARA, A., SHIRAIISHI, K., SATO, H., SAKAMURA, S., NISHIYAMA, K., SAKAI, R., FURUSAKI, A. & MATSUMOTO, T. (1977). *J. Am. Chem. Soc.* **99**, 636.
- KARLE, I. L., GIBSON, J. W. & KARLE, J. (1969). *Acta Cryst.* **B25**, 2034–2039.
- KARLE, I. L., KARLE, J. & ESTLIN, J. A. (1967). *Acta Cryst.* **23**, 494–500.
- KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.
- KARLE, J. & HAUPTMAN, H. (1961). *Acta Cryst.* **14**, 217–223.
- KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.
- LESSINGER, L. (1976). *Acta Cryst.* **A32**, 538–550.
- NIGGLI, A., VAND, V. & PEPINSKY, R. (1960). *Acta Cryst.* **13**, 1002.
- VAND, V., NIGGLI, A. & PEPINSKY, R. (1960). *Acta Cryst.* **13**, 1001.
- WHITE, P. S. & WOOLFSON, M. M. (1975). *Acta Cryst.* **A31**, 53–56.