Références

BILLIET, Y. (1979). *Acta Cryst.* A**35**, 485–496.

BILLIET, Y. & ROLLEY-LE COZ, M. (1980). *Acta Cryst.* A**36**, 242–248.

BOURBAKI, N. (1952). *Eléments de Mathématique. Algèbre. Groupes et Corps Ordonnés, Modules sur les Anneaux Principaux*, N° 1179. Paris: Hermann et Cie.

HARKER, D. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 5264–5267.

HARTLEY, B. & HAWKES, T. O. (1970). *Rings, Modules and Linear Algebra*. Londres: Chapman and Hall.

ROLLEY-LE COZ, M. & BILLIET, Y. (1980). *Acta Cryst.* A**36**, 785–792.

Acta Cryst. (1982). A**38**, 117–122

Phase Extension and Refinement. II. Application to Metmyoglobin 2·0 Å Data

BY GER J. OLTHOF AND HENK SCHENK

Laboratory for Crystallography, University of Amsterdam, Nieuwe Achtergracht 166, 1018*WV Amsterdam, The Netherlands*

(*Received* 23 *March* 1981; *accepted* 13 *July* 1981)

Abstract

The phase refinement and extension procedure of Olthof, Sint & Schenk [*Acta Cryst.* (1979), A**35**, 941–946] is applied to the protein metmyoglobin. Tests show that the procedure is enantiomorph conserving and can be used to improve the resolution.

Introduction

Convolutional and related equation systems can be used in order to refine and extend phase sets (see, for a recent review article, Sayre, 1980). The most widely accepted is the system based on the tangent formula:

$$\exp(i\varphi_H) = \sum_K E_3 \exp[i(\varphi_K + \varphi_{H-K})] \times \left\{ \left| \sum_K E_3 \exp[i(\varphi_K + \varphi_{H-K})] \right| \right\}^{-1}, \quad (1)$$

with $E_3 = N^{-1/2}|E_H E_K E_{H-K}|$. In noncentrosymmetric polar space groups, however, its repeated use tends to favour centrosymmetric solutions and this may well give rise to both enantiomorphs instead of only one.

Enantiomorph-specific methods have been proposed based on a modified tangent formula

$$\exp(i\varphi_H) = \sum_K E_3 \exp[i(\varphi_K + \varphi_{H-K} - s\Delta_3)] \times \left\{ \left| \sum_K E_3 \exp[i(\varphi_K + \varphi_{H-K} - s\Delta_3)] \right| \right\}^{-1}. \quad (2)$$

$s\Delta_3$ is an estimate of

$$\varphi_3 = -\varphi_H + \varphi_K + \varphi_{H-K}.$$

Such methods have been successfully tested for phase extension and refinement (Sint & Schenk, 1975; Busetta, 1976; Olthof, Sint & Schenk, 1979; Olthof & Schenk, 1981*b*) and for the selection of correct phase sets (Olthof & Schenk, 1981*a*). They are based on the idea, originally proposed by Sint & Schenk (1975), that an estimate Δ_3 can be obtained for the absolute value of the triplet phase sum φ_3 from a plot against E_3 of the values of

$$\delta(E_3) = \langle |\varphi_3| \rangle_{E_3}.$$

Here the averaging is performed over E_3 intervals. Reliable values of Δ_3 are already obtained from a moderately large set of known phases.

Phase refinement can be achieved in an enantiomorph-specific way by selecting the sign s in the modified tangent expression (2) so as to minimize the value of

$$|-\varphi_H + \varphi_K + \varphi_{H-K} - s\Delta_3|.$$

Olthof, Sint & Schenk (1979) showed for two artificial structures in monoclinic polar space groups with about 400 atoms per unit cell that such a refinement is successful. The unmodified tangent formula (1) was found to lead to centrosymmetric phases within a few refinement cycles.

Phase extension can be achieved by use of either (1) or

$$\Phi(\varphi_H) = \sum_K E_3 |-\varphi_H + \varphi_K + \varphi_{H-K} - s\Delta_3|. \quad (3)$$

0567-7394/82/010117-06$01.00

The latter expression selects the phase φ_H which minimizes Φ (Olthof, Sint & Schenk, 1979). It was found that the unmodified expression (1) was most favourable, mainly because the computing effort involved in its application is vastly less than that involved in (3).

This led to a reliable phase extension and refinement system, which required a moderate computing effort. As the computing time would remain reasonable for larger structures it was announced that presently a small protein would be tried. For this purpose we selected metmyoglobin (Takano, 1977) from the data available in the Protein Data Bank, because it crystallizes in space goup $P2_1$, and also because data were available up to a resolution of 2·0 Å. Moreover Coulter & Dewar (1971) used the isomorphous myoglobin data in tangent refinement (1) experiments, with rather disappointing results. The results of extensive tests of our procedure on metmyoglobin are presented in this paper.

Normalization and triplet search

Data for metmyoglobin, $C_{817}H_{1535}FeN_{221}O_{308}S_4$, refined to 2·0 Å resolution, were obtained from the Protein Data Bank (Bernstein *et al.*, 1977). Metmyoglobin has one molecule in the asymmetric unit of space group $P2_1$, cell constants $a = 64{\cdot}56$, $b = 30{\cdot}97$ and $c = 34{\cdot}86$ Å and $\beta = 105{\cdot}86°$. From 10 006 independent reflections 989 were regarded unreliable (Takano, 1977). The remaining 9017 were used to calculate $|E|$ values by the K-curve procedure from the XRAY system (Stewart, 1976). The $|E|$ statistics are given in Table 1.

For the tests five triplet files were created with the program *TRIQUA* from the program system *SIMPEL* (Overbeek & Schenk, 1978), see Table 2.

Tests of the tangent formulae

In order to verify the stability of the phase sets under refinement with the unmodified and the modified tangent formula tests were carried out with all five triplet files of Table 2. The Protein Data Bank supplies phases for 3030 out of the 3261 reflections with $|E| \geq 1{\cdot}0$, the remaining 231 phases are indicated to be unreliable. In all tests five cycles of phase refinement were applied to this set of 3030 phases. It appeared necessary to accept phases calculated by (1) or (2) only if the number of terms exceeded a threshold value of about ten. In each cycle i a value

$$\mathrm{CONS}_i = \sum_H \sum_K E_3 |-\varphi_H + \varphi_K + \varphi_{H-K} - s\Delta_3|$$

was calculated. From the second cycle on convergence was monitored by

$$S_i = |(\mathrm{CONS}_{i-1} - \mathrm{CONS}_i)/\mathrm{CONS}_{i-1}|. \quad (4)$$

(*a*) *Test of the unmodified tangent formula* (1)

The unmodified tangent refinement procedure was carried out with all five files of Table 2, the results of which are given in Table 3, part I. As can be seen from the average phase deviations with respect to the Data Bank phases for structure and enantiomorph, which for all experiments are nearly equal, the sets of phases are approximately centrosymmetric and do not reflect the original phases. An E map was not calculated because the phases most probably map the Fe atom in a centrosymmetric surrounding as described by Coulter & Dewar (1971).

From the column headed $S5$ in Table 3 it can be seen that convergence is not complete for any of the runs and also that an appreciable number of reflections cannot be phased. It is also obvious that, except perhaps for the smallest set A, the enantiomorph is very rapidly lost; this results from the larger errors in the additional triplets of the larger files.

(*b*) *Test of the modified tangent formula* (2)

The results of refinement tests with (2) are summarized in Table 3, part II. As can be seen, the average phase deviation $\langle|\Delta\varphi|\rangle$ relative to the true structure decreases with increasing number of triplets. In all cases the phases remained stable and convergence was reached. No details of the enantiomorph creep in, judging from the nearly random difference between refined phases and the phases of the enantiomorph constructed on the basis of the Data Bank phases.

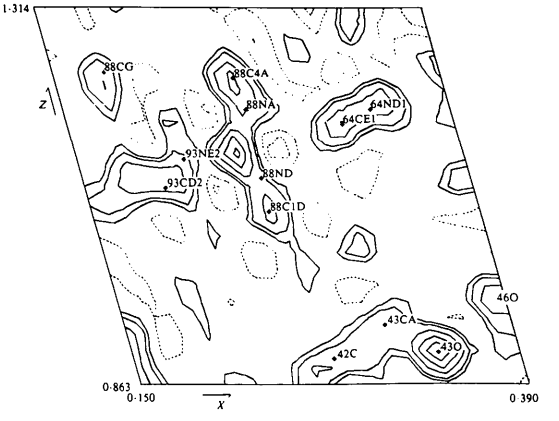

Fig. 1. Section at $y = 0{\cdot}9388$ of the E map of metmyoglobin calculated from the phases of the 3030 reflections with $E \geq 1{\cdot}0$, given by the Protein Data Bank. The atom positions as given by Takano (1977) are marked.

E maps were calculated for the 3030 phases from the Protein Data Bank (the reference map), for the corresponding phases refined by means of triplet file E, and finally for all 3237 refined phases, including most of the 231 phases not found by the protein methods. A typical section of each E map is presented in Figs. 1, 2 and 3 respectively. By comparing Figs. 1 and 2 it can be seen that in the refinement no details of the structure were lost. At a few positions the map even looks a little better. Fig. 3 shows that the additional phases must be approximately correct since the overall quality of the map is the same and again small improvements can be noticed.

Table 1. $|E|$ *statistics and* $|E|$ *distribution for metmyoglobin*

	Metmyoglobin	Theoretical Centric	Theoretical Acentric
Average $\|E\|$	0·895	0·798	0·886
Average $\|E\|^2$	1·004	1·000	1·000
Average $\|E^2-1\|$	0·730	0·968	0·736
Average $\|E^2-1\|^2$	1·023	2·000	1·000
Average $\|E^2-1\|^3$	2·697	8·000	2·000
$\|E\|>3·0$	0·01	0·27	0·01
$\|E\|>2·5$	0·24	1·24	0·19
$\|E\|>2·0$	1·97	4·55	1·83
$\|E\|>1·8$	4·09	7·19	3·92
$\|E\|>1·6$	7·79	10·96	7·73
$\|E\|>1·4$	13·75	16·15	14·09
$\|E\|>1·2$	23·15	23·01	23·69
$\|E\|>1·0$	36·33	31·73	36·79
$\|E\|>0·0$	100·00	100·00	100·00

Table 2. *Data on the test files*

Triplet file	Threshold value for E_3	Number of triplets (non-redundant)
A	0·105	35459
B	0·098	50302
C	0·088	71853
D	0·082	102741
E	0·055	376282

Table 3. *The number of phases refined in cycle i using the tangent formula;* I: *unmodified;* II: *modified*

For the last cycle the average phase deviations $\langle|\Delta\varphi|\rangle$ (in mcycles) are given with respect to the structure and the enantiomorph, and *S*5 gives the convergence monitor (4).

	File	Cycle i = 1	2	3	4	5	$\langle\|\Delta\varphi\|\rangle$ True structure	$\langle\|\Delta\varphi\|\rangle$ Enantiomorph	*S*5
I	*A*	1774	1702	1705	1725	1749	196	223	0·152
	B	2177	2114	2114	2170	2174	206	218	0·258
	C	2666	2604	2609	2663	2658	208	218	0·299
	D	3020	2980	2993	3045	3046	211	219	0·312
	E	3179	3145	3193	3209	3209	211	216	0·153
II	*A*	1810	1816	1816	1816	1816	56	222	0·017
	B	2216	2227	2228	2228	2228	56	224	0·015
	C	2710	2712	2712	2711	2711	55	223	0·012
	D	3080	3094	3094	3094	3094	52	224	0·011
	E	3238	3238	3238	3238	3237	29	223	0·001

(*c*) *Comparison*

The differences in behaviour of refinement based on (1) and (2) are large. Whereas refinement (2) gives approximately correct phases, refinement (1) results in false ones. This is because on the whole $s\Delta_3$ is a reasonable estimate for φ_3 in (2).

Table 4. *Results of the phase extension and refinement procedure* (*a*) *from* 2·8 *to* 2·0 Å *and* (*b*) *from* 3·0 *to* 2·0 Å *resolution, based on the triplet files of Table 2 with different thresholds imposed on the number of terms*

For each cycle the number of new phases and the total number of refined phases are given. The average phase error $\langle|\Delta\varphi|\rangle$ (in mcycles) with respect to the structure and enantiomorph refers to the phases of the last cycle.

	File	Threshold	Cycle 1 New	Cycle 1 Total	2 New	2 Total	3 New	3 Total	4 New	4 Total	$\langle\|\Delta\varphi\|\rangle$ True structure	$\langle\|\Delta\varphi\|\rangle$ Enantiomorph
(*a*)	*A*	9	349	1069	505	1586	72	1658	–	–	128	227
	B	9	479	1349	679	2031	37	2068	–	–	129	224
	C	9	655	1699	805	2508	22	2530	–	–	130	228
	D	9	931	2091	854	2945	7	2951	–	–	126	229
	E	9	1936	3115	27	3144	–	–	–	–	108	229
	C	14	464	1360	693	2064	73	2137	–	–	124	226
	D	20	464	1430	790	2234	89	2323	–	–	116	224
	E	76	475	1631	1074	2717	313	3015	–	–	102	228
(*b*)	*A*	9	300	885	569	1481	152	1633	–	–	140	226
	B	9	421	1132	818	1962	85	2047	–	–	146	229
	C	9	587	1452	984	2442	64	2506	–	–	149	230
	D	9	820	1781	1127	2911	27	2938	–	–	146	232
	E	9	2009	2982	148	3125	1	3125	–	–	131	232
	C	14	380	1110	823	1957	162	2119	–	–	141	228
	D	20	381	1156	878	2077	214	2291	–	–	136	229
	E	76	312	1201	837	2113	813	2917	87	2999	125	231

From the experiments described above it can be concluded that a refinement of protein phases based on the modified tangent formula conserves the enantiomorph. If, at a certain resolution, some phases cannot be found by other techniques, these phases can be calculated reliably by the same expression, provided that the number of triplets for these phases is sufficiently large. By doing so the image of the protein is improved.

Phase extension and refinement from 2·8 to 2·0 Å

Because the original isomorphous phases at 2·8 Å resolution were not available to us, the starting point for a test of the combined phase extension based on (1) and refinement based on (2) was the set of 1183 Data Bank phases corresponding to the reflections of 2·8 Å resolution with $|E| \geq 1{\cdot}0$. The procedure was carried out in a number of cycles, each consisting of an extension part, in which as many new phases were determined as possible with (1), and of a refinement part, in which all phases were refined with (2). In case the convergence monitor (4) was larger than some threshold value another cycle was carried out.

To start with, the procedure was run for the five triplet files listed in Table 2, with a threshold value of 9. The results of these experiments are summarized in Table 4. The main result is that by means of file E 3144 phases can be found with an average error with respect to the Data Bank phases of 108 mcycles or 39°. The error with respect to the phases of the enantiomorph is nearly 230 mcycles indicating that the enantiomorph is absent in the final phases.

To decrease the errors in the final phases further runs were made for some models at thresholds which depended on the number of triplets in the different files.

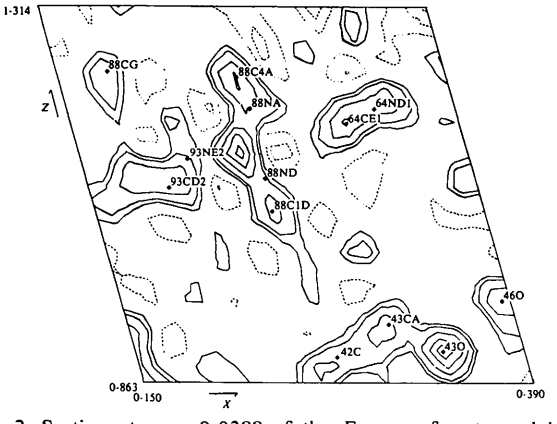

Fig. 2. Section at $y = 0{\cdot}9388$ of the E map of metmyoglobin calculated with all refined phases, corresponding to the 3030 reflections with $E \geq 1{\cdot}0$, used for the calculation of Fig. 1.

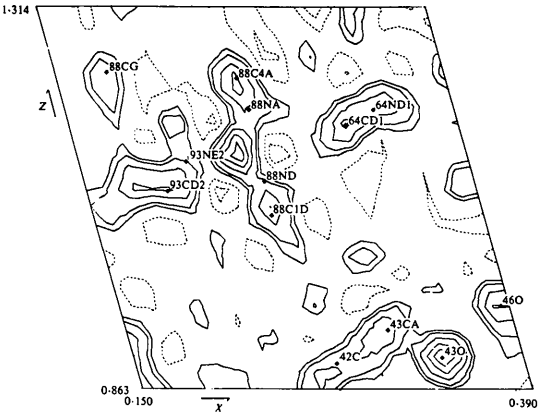

Fig. 3. Section at $y = 0{\cdot}9388$ of the E map of metmyoglobin calculated with 3237 extended and refined phases corresponding to reflections with $E \geq 1{\cdot}0$, starting with 3030 phases given by the Protein Data Bank.

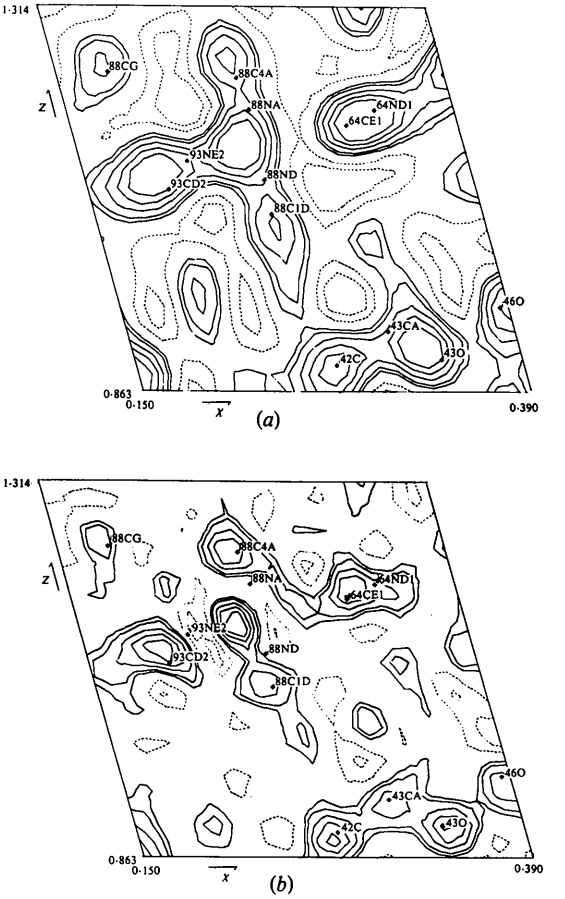

Fig. 4. Fourier sections at $y = 0{\cdot}9388$. (*a*) Calculated from the phases of 1183 reflections, corresponding to the 2·8 Å resolution starting set. (*b*) Calculated from the set of 3105 phases, resulting from the extension and refinement of the 2·8 Å resolution starting set.

Now, using file E, in three cycles 3015 phases could be found with an average error of 102 mcycles or 37°. These and other results are summarized in Table 4. It can be seen that the larger the triplet file the larger the number of phased reflections and the smaller the final average error. From Table 4 it is also evident that a high threshold value for the number of terms in the tangent expressions (1) and (2) is desirable.

E maps were calculated for the 1183 starting phases and the 3015 phases from experiment $E76$, sections of which are given in Figs. 4(a) and (b), respectively. Comparison of Figs. 1 and 4(a), both based on the Data Bank phases (2·0 and 2·8 Å resolution, respectively) shows that the first contains much more detail. The map of the refined phases (Fig. 4b) is of somewhat lower quality than the map based on all Data Bank phases (Fig. 1), but it shows much more detail than the map of the starting phases.

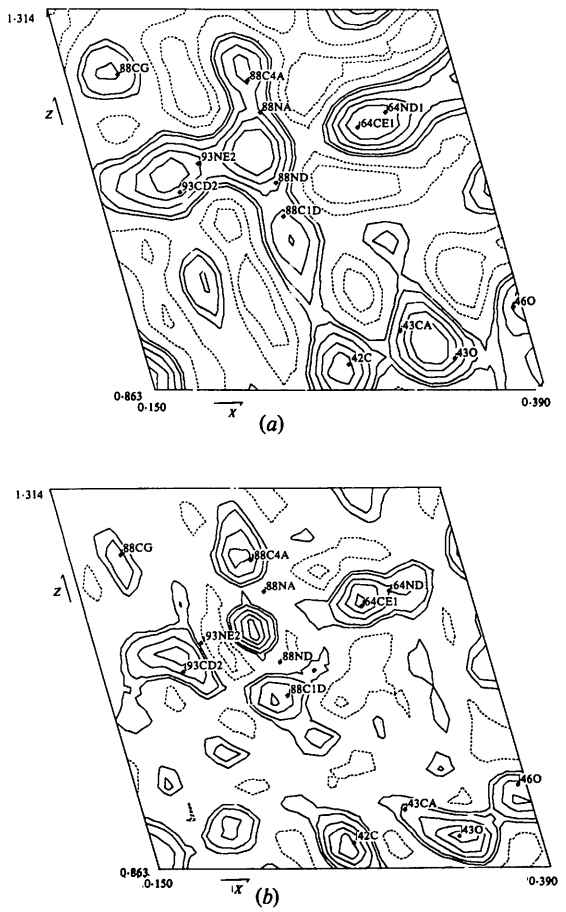

Fig. 5. Fourier sections at y = 0·9388. (a) Calculated from the phases of 977 reflections, corresponding to the 3·0 Å resolution starting set. (b) Calculated from the set of 2999 phases, resulting from the extension and refinement of the 3·0 Å resolution starting set.

For the experiment with file E and a threshold of 76 the computing time involved was 145 s on a CYBER 175/750. It can be noted that the computing time is linearly correlated with the number of triplets used.

Phase extension and refinement from 3·0 to 2·0 Å

In order to study the effect of the size of the set of starting phases on the quality of the final set of extended and refined phases the experiments of the last section were repeated, but now starting from the 977 phases corresponding to a resolution of 3·0 Å. In one experiment it was necessary to apply four cycles of extension and refinement. The results are summarized in Table 4(b). The average phase error in experiment $E76$ was 125 mcycles (=45°) for 2999 refined phases, about 25% larger than in the corresponding experiment from the preceding section.

E maps were calculated with the 977 starting phases and with the 2999 refined phases, respectively, and sections of both maps are reproduced in Fig. 5. Comparison of Figs. 4 and 5 shows that both sections of Fig. 5 show less detail than those of Fig. 4.

It can be concluded that the larger the difference in resolution between starting set and final phases the larger the average phase error will be.

Discussion

A general remark with respect to the E maps is that the peak height of the Fe atom is approximately of the expected order ($\simeq$ three times the height of the others). Coulter & Dewar (1971) had noticed that in their maps the Fe atom tended to be too large, two or three times the expected height. Sayre (1974) reported the opposite effect for the phase extension and refinement of rubredoxin; here the structure was driven to equality of atom heights.

In view of the differences between the results of the 2·8–2·0 and the 3·0–2·0 Å extensions the extension procedure must be applied to limited ranges only. Map interpretation (Diamond, 1980) and successive phase calculation then will have to be followed by application of the procedure described here.

In a preceding paper (Olthof, Sint & Schenk, 1979), it was concluded that new phases could be calculated reliably with the unmodified tangent formula (1). A better though much slower method, with the phase φ_H corresponding to the minimum of (3), was also described there. In view of the fact that the Δ_3 values in the present experiments are more than twice as large as those described for the $N \simeq 400$ test structures, the latter method may lead to a further improvement of the errors given in Table 4. Another possibility of improving the quality of the maps is to include reflections with $|E| < 1{\cdot}0$. The computing time involved will not be influenced dramatically by these changes.

In conclusion it can be stated that the enantiomorph-specific refinements and extensions described here are capable of improving the electron-density maps of large molecules. Its limit of applicability, and the possible improvements suggested in this paragraph, will be pursued further.

The authors thank Professor B. O. Loopstra and Dr C. H. Stam for valuable comments on the manuscript. The Protein Data Bank is acknowledged for providing us with the data on metmyoglobin. We are grateful to Mr W. Gaykema of the Laboratory voor Chemische Fysica of the University of Groningen for assisting with the calcualtion of the E maps. One of us (GJO) is indebted to the Netherlands Organization for the Advancement of Pure Research (ZWO) for financial support.

References

BERNSTEIN, F. C., KOEZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F. JR, BRICE, M. D., RODGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BUSETTA, B. (1976). *Acta Cryst.* A**32**, 139–143.

COULTER, C. L. & DEWAR, R. B. K. (1971). *Acta Cryst.* B**27**, 1730–1740.

DIAMOND, R. (1980). In *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN. Bangalore: Indian Academy of Sciences.

OLTHOF, G. J. & SCHENK, H. (1981*a*). *Acta Cryst.* A**37**, 684–689.

OLTHOF, G. J. & SCHENK, H. (1981*b*). *Acta Cryst.* A**37**, 689–691.

OLTHOF, G. J., SINT, L. & SCHENK, H. (1979). *Acta Cryst.* A**35**, 941–946.

OVERBEEK, A. R. & SCHENK, H. (1978). In *Computing in Crystallography*, edited by H. SCHENK, R. OLTHOF-HAZEKAMP, H. VAN KONINGSVELD & G. BASSI, pp. 108–112. Delft Univ. Press.

SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.

SAYRE, D. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER, pp. 271–286. New York: Plenum.

SINT, L. & SCHENK, H. (1975). *Acta Cryst.* A**31**, S22.

STEWART, J. M. (1976). The XRAY system–version of 1976. Tech. Rep. TR-446. Computer Science Center, Univ. of Maryland, College Park, Maryland.

TAKANO, T. (1977). *J. Mol. Biol.* **110**, 537–568.

Acta Cryst. (1982). A**38**, 122–128

The Maximum Determinant Method and the Maximum Entropy Method

BY RAMESH NARAYAN AND RAJARAM NITYANANDA

Raman Research Institute, Bangalore-560080, India

(*Received* 9 *March* 1981; *accepted* 24 *July* 1981)

Abstract

A generalized maximum determinant rule is shown to be equivalent to maximizing the integral of the logarithm of the electron density, *i.e.* equivalent to the 'maximum entropy method' (MEM) of image reconstruction. Relations between the structure factors and the Fourier coefficients of the reciprocal of the electron density follow, leading to new algorithms for phase determination and refinement. Although structures with equal, spherical, resolved atoms automatically satisfy the MEM phase relations, the method really requires only positivity and 'peakiness' of the electron density.

Editorial note: The similarity between this and the following paper, by Britten & Collins [*Acta Cryst.* (1982), A**38**, 129–132], has been recognized and, although they represent completely independent work, they have been published together to facilitate comparison.

1. Introduction

The maximum determinant method (MDM) of crystallography (Lajzerowicz & Lajzerowicz, 1966; Tsoucaris, 1970) has received fair attention as an alternative 'direct method' to the conventional approach of estimating low-order structure invariants and seminvariants (Sayre, 1952; Ladd & Palmer, 1980). Tsoucaris (1980) describes the use of the MDM to rederive low-order relations, determine phases *ab initio* (from medium-sized determinants), and refine and extend phases (from large determinants). In this paper we give a new interpretation of the MDM and use it to draw a number of conclusions of relevance to crystallography.

We introduce in § 2 a theorem relating a certain limit of the Karle–Hauptman (1950) determinant to the integral of the logarithm of the electron density in the

0567-7394/82/010122-07$01.00