Acta Cryst. (1969). B25, 2614

On the Use of the Tangent Formula to Extend the Resolution of Protein Phases

BY GEORGE N. REEKE, JR AND WILLIAM N. LIPSCOMB

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138. *U.S.A.*

(*Received* 10 *March* 1969)

The tangent formula has been applied to data from carboxypeptidase *A* (CPA) to determine its effectiveness at various resolutions for the refinement of multiple isomorphous replacement (MIR) phases and for the extension of MIR phases to reflections at higher resolution for which native data but not heavy-atom derivative data have been measured. Results were judged by comparing electron-density maps made from various phase sets with the known 2·0 Å structure of CPA and by examining the average differences between the MIR and tangent formula phases. The tangent formula was most effective at the lowest resolution tested (6 Å) and produced increasingly poorer results as the resolution was increased to 2·0 Å. At 6 Å, the tangent formula phases were slightly better than the MIR phases. The Karle & Karle '*R* factor' was shown to be a good internal indicator of the accuracy of the tangent formula phases.

Introduction

The popularity of the so-called 'direct methods' of obtaining phases in X-ray crystallography has increased in recent years as their utility has been demonstrated by a succession of published structure solutions. Early contributions to the theory of direct methods were made by a number of workers, including Harker & Kasper (1948), Karle & Hauptman (1950), Sayre (1952), and Zachariasen (1952). The 'tangent formula' (Karle & Hauptman, 1956) for noncentrosymmetric crystals has been used succesfully both for developing phases *ab initio* (Karle & Karle, 1966*a*, *b*) and for refining phases obtained from trial structures (Karle, 1968) . The protein crystallographer has viewed these developments with great interest, in spite of the obvious theoretical and computational difficulties, because any alternative or supplement to the laborious multiple isomorphous replacement (MIR)* method would be welcome. A number of possible applications of the tangent formula in protein work at once suggest themselves: heavy metal atoms could be located in metal–protein complexes; the correct phases could be chosen from the pairs of possible phases determined by a single isomorphous replacement (SIR); MIR phases could be refined in analogy to the refinement of phases from a trial structure; a few known phases from centrosymmetric zones or favorable reflections in SIR could be used to develop a complete set of phases; a set of MIR phases could be extended to include additional reflections, most likely at higher resolution; finally, perhaps phases could even be obtained *ab initio*. The location of heavy metal atoms by phasing structure-factor differences has been discussed by Steitz (1968) for the case of centrosymmetric projections. The choosing of the correct phases from SIR pairs was tested at low resolution by Weinzierl, Eisenberg & Dickerson (1969), who also refined MIR phases at low resolution. Earlier, Perutz & Scatturin (1954) had some success deriving phases for a few strong centrosymmetric reflections, and Coulter (1965) performed a successful model calculation based on his suggestion that a complete set of phases could be developed from a few known phases, particularly from reflections where the phase ambiguity in the SIR method was less than about 30°. Much the same idea was proposed ten years earlier† in a paper by Cochran (1955), who gives credit for the idea to F.H.Crick. The relation between isomorphous replacement and the tangent formula was further discussed by Karle (1966). The present paper proceeds to consider the use of MIR phases in the tangent formula to phase reflections for which heavy atom replacement data are not available and thus also discusses the tangent formula refinement of MIR phases at high resolution.

The tangent formula, as given by Karle & Karle (1966*a*), is

$$\tan \varphi_{\mathbf{H}} \simeq \frac{\sum_{\mathbf{K}} |E_{\mathbf{K}} E_{\mathbf{H}-\mathbf{K}}| \sin(\varphi_{\mathbf{K}} + \varphi_{\mathbf{H}-\mathbf{K}})}{\sum_{\mathbf{K}} |E_{\mathbf{K}} E_{\mathbf{H}-\mathbf{K}}| \cos(\varphi_{\mathbf{K}} + \varphi_{\mathbf{H}-\mathbf{K}})},$$

where

$\varphi_{\mathbf{K}}$ = phase of reflection **K**,
$E_{\mathbf{K}}$ = normalized structure factor of reflection **K**.

It is clear that this formula (without feedback of output phases) is limited to doubling the range of indices included in a set of phases, since contributions to a given phase **H** are from pairs of reflections, **K** and **K′**, where **H**=**K**+**K′**. In the calculations reported here,

* Abbreviations used in this paper are: CPA, carboxypeptidase *A*; MIR, multiple isomorphous replacement; SIR, single isomorphous replacement; TAN, tangent formula.

† The tangent formula not yet having been discovered, Cochran discussed the use of the formula $\varphi_{\mathbf{H}} \simeq \varphi_{\mathbf{K}} + \varphi_{\mathbf{H}-\mathbf{K}}$.

extension of the range of indices covered was limited to a factor of about 1·4 in each index.

It is difficult to estimate theoretically the accuracy to be expected from the tangent formula in phasing protein data, particularly because of the large experimental errors in real data and because of the incompleteness (relative to the limiting sphere) of the data. It has been stated by some, *e.g.* Cochran (1955), that data sufficient to resolve individual atoms are required for the success of relations between structure factors. On the other hand, careful study has not been made of the possible correctness of the tangent formula when the conditions of its usual derivation are not satisfied. Therefore, it appeared that it would be useful at this early stage to attempt a partial evaluation of tangent formula methods by calculations with actual data from a known protein structure. Carboxypeptidase *A* (CPA) forms an ideal system for these tests because an excellent MIR map is available at 2·0 Å resolution (Reeke, Hartsuck, Ludwig, Quiocho, Steitz & Lipscomb, 1967), the structure has been solved, and a structure factor calculation has been performed (Lipscomb, Reeke, Hartsuck, Quiocho & Bethge, 1969). Also required for these calculations is a computer program which is able to handle very large numbers of terms in the tangent formula in a reasonable amount of time. Such a program has been written by one of us (G.N.R.) and has been used succesfully in the solution of several structures, both centrosymetric and noncentrosymmetric.*

With this program and the carboxypeptidase *A* data a number of calculations were performed with various sets of starting phases at various resolutions in an effort to determine under what conditions the tangent formula could be expected to improve or extend a set of MIR phases. The major criterion of success was the examination of electron-density maps computed from measured native protein intensities and phases from the various test calculations, and comparison of these maps with the known carboxypeptidase structure.

Calculation of normalized structure factors

Carboxypeptidase *A*, a Zn-containing enzyme of M.W. 34,600, crystallizes in the monoclinic space group $P2_1$, with unit-cell dimensions $a=51{\cdot}41$, $b=59{\cdot}89$, $c=47{\cdot}19$ Å, and $\beta=97{\cdot}58°$. (This space group is somewhat unfavorable for direct methods because each known reflection gives rise to only three additional reflections by symmetry.) The collection of data on native CPA and four heavy-atom derivatives has been described (Reeke *et al.*, 1967). In order to obtain the parameters needed to calculate normalized structure factors (*E*'s), the CPA intensity data were subjected to a modified Wilson plot. Since the exact numbers of the various kinds of atoms (particularly solvent atoms) in the unit cell were not known, that part of the intensity fall-off with increasing scattering angle due to atomic form factors was absorbed in the overall temperature parameter by plotting the average intensity, $\bar{I}$, rather than the usual $\bar{I}/\Sigma f_i^2$ (where f_i is an atomic form factor), against $\sin^2\theta/\lambda^2$. This procedure can be justified if one considers the relatively large thermal motions or disorders in the protein crystals, which dominate the fall-off, and the small range of $\sin^2\theta/\lambda^2$ covered, over which the behavior of the form factors may be taken as simply exponential. The $0k0$ reflections were divided by two to correct for intensity enhancement by the twofold screw axis. A number of small, poorly measured reflections were omitted from the Wilson plot, giving an apparant thermal parameter smaller than the 'true' value and thus giving smaller calculated *E*'s to the high-angle reflections. This effect serves to give a probably desirable reduced weight to these reflections in the tangent formula calculations.

The Wilson plot was similar in appearance to plots obtained for several other proteins (Rossmann, Jeffery, Main & Warren, 1967). It showed a minimum at $d=6{\cdot}7$ Å, a broad hump near $d=4$ Å, and a fairly uniform fall-off at higher resolution. A straight-line fit was made to the Wilson plot, and the overall thermal parameter $B=16{\cdot}0$ Å^2 was used to calculate the normalized structure factors. The peak near 4 Å is sufficiently broad that with the '*K* curve' procedure of Karle & Hauptman (1953) division by a smooth curve drawn through the points would flatten out the local variations. The use of the straight-line fit would tend to preserve in the *E*'s any structural information contained in the largeness of the reflections near 4 Å, and would tend to give higher weight to these reflections, which are the best measured. The set of *E*'s resulting from these calculations gave a distribution closely similar to the ideal distribution for a noncentrosymmetric structure, and therefore the *E*'s were accepted for use in further calculations.

Tangent formula calculations

The tangent formula calculations performed are summarized in Table 1. The first part of the Table describes the data sets used as input phases to the tangent formula and the data sets phased from these starting sets. The starting sets were chosen on the basis of resolution, large value of *E*, and sometimes large value of the figure of merit, the cosine of the error estimated in MIR phasing. Usually the MIR phase was assigned as the starting value, but in some runs the output phases from a previous tangent formula calculation were used. The sets to be phased were chosen on the basis of resolution and large value of *E* and intensity only. In all cases, phases determined during a particular iteration of the phasing formula were not used for the determination of further phases during the same cycle. This is the iteration method called 'block cycling' by Wein-

* The present version of this program, which is called *FAZE*, is written in machine language for the IBM 7094 computer and will operate on no other machine, unless emulation is employed.

Table 1. *Carboxypeptidase tangent formula calculations*

Run	Starting set					Data phased[a]			
	Å	No. of reflections	FOM[b] min	E min	Phases	Å	No. of reflections	E min	Notes
1	6·0	310	–	0·9	MIR	6·0	310	0·9	*k*
2	6·0	306	0·4	0·9	MIR	6·0	310	0·9	
3*A*	6·0	306	0·4	0·9	MIR	4·5	894	0·9	
3*B*	6·0	306	0·4	0·9	MIR	4·5	894	0·9	
4	6·0	281	–	0·9	TAN	4·5	894	0·9	*l*
5	6·0	~300	0·4	0·9	MIR & TAN	4·5	894	0·9	*m*
6	4·7	643	–	1·0	MIR	4·7	643	1·0	
7*A*	4·7	621	0·4	1·0	MIR	3·5	2085	1·0	
7*B*	4·7	621	0·4	1·0	MIR	3·5	2085	1·0	
8	4·7	603	–	1·0	TAN	3·5	2085	1·0	*o*
9	3·5	1330	0·3	1·0	MIR	2·8	1155	1·0	*p, q*
							1330		*s*
10	3·5	2085	–	1·0	MIR	2·8	2485	1·0	*p*
11	2·8	1366	0·5	1·2	MIR	2·0	3301	1·0	*p, q*
							1366	1·0	*s*
12	2·8	2246	0·4	1·0	MIR	2·0	2421	1·0	*p, q*
							2246	1·0	*s*
13	2·0	~4600	–	1·05	MIR & TAN	2·0	4667	1·0	*p, t, u*

zierl *et al.* (1969). When only one cycle was contemplated for a particular run, the program saved time by searching only for terms in the tangent formula involving the preassigned starting phases. Otherwise, all possible terms were found.

The second part of Table 1 indicates the average numbers of terms used to phase each reflection, the criteria used for acceptance of phase determinations for inclusion in Fourier syntheses, some refinement statistics accumulated by the program for accepted determinations, and some statistics for all of the reflections. Also, references are made to the illustrations of the Fourier syntheses made from the various phase sets (see below). In some runs starting phases were allowed to take on their calculated values after the first cycle; in other runs, they were held fixed at the MIR values. These treatments are indicated by one of the letters '*V*' or '*F*' respectively. In some multicycle runs, determinations were accepted only when the calculated phase did not change by more than 90° from one cycle to the next. This 'jump test' is indicated by the letter '*J*'. In all runs, phase determinations were accepted only when the number of terms was greater than a certain number (5, 10, or 20) and the value of 'α' was greater than a certain minimum value, $\alpha_{\min}$, given in Table 1. The quantity 'α', a measure of the reliability of phase determination, is defined by Karle & Karle (1966*a*), using slightly different notation, to be

$$\alpha_{\mathbf{H}} = 2\kappa E_{\mathbf{H}} \{[\sum_{\mathbf{K}} E_{\mathbf{K}} E_{\mathbf{H}-\mathbf{K}} \cos(\varphi_{\mathbf{K}} + \varphi_{\mathbf{H}-\mathbf{K}})]^2 + [\sum_{\mathbf{K}} E_{\mathbf{K}} E_{\mathbf{H}-\mathbf{K}} \sin(\varphi_{\mathbf{K}} + \varphi_{\mathbf{H}-\mathbf{K}})]^2\}^{1/2},$$

where $\kappa = (\Sigma f_i^3)/(\Sigma f_i^2)^{3/2}$. The probability multiplier, κ was estimated to be 0·013 for CPA. It is clear from the Table that the choice of $\alpha_{\min}$ was not always a judicious one and it would have been better to select $\alpha_{\min}$ after each run. However, the accuracy of phasing is not here a sensitive function of α (see below), so the conclusions are not affected.

The statistical results for each run include the average value of α for all reflections, the number of reflections with α greater than 1·0 (where this was available), the number of reflections considered determined, and the '*R* factor' of Karle & Karle (1966*a*) for these reflections. This '*R* factor' is obtained from the formula

$$R = \frac{\sum_{\mathbf{H}} |E_{\text{obs}} - E_{\text{calc}}|}{\sum_{\mathbf{H}} E_{\text{obs}}},$$

where E_{calc} is the value of E calculated from the tangent formula. The quantities α and R are here being tested as possible indicators of the accuracy of phase determination, since they are independent of the known MIR phases of the reflections being phased. The quantity $\Delta\varphi = |\varphi_{\text{TAN}} - \varphi_{\text{MIR}}|$ or $360° - |\varphi_{\text{TAN}} - \varphi_{\text{MIR}}|$, whichever is smaller, is used for comparison of the TAN and MIR phases. The average value of $\Delta\varphi$ is given for all accepted reflections, for accepted reflections with $E \geq 1{\cdot}25$, for all reflections, and for the half of the reflections with largest α. As a control, $\Delta\varphi$ was plotted also as a function of the figure of merit of the MIR phases, but as expected no particular trend was discernible.

At each resolution studied, Fourier electron density maps were made from the MIR starting phases and from various combinations of the MIR starting phases with the phases resulting from the tangent formula calculations. The magnitudes of the Fourier coefficients were the measured native CPA structure factors for all maps involving the tangent formula phases, and the structure factors weighted by

Table 1 (*cont.*)

Run	Conditions			Determinations			All reflections					Fig.
	Terms/ reflection	*VFJ*[c]	α[d] min	Cycle	No determined	*R*[e]	Avg.[f] $\Delta\varphi$	Avg.[g] $\Delta\varphi$	Avg. α	No. reflections $\alpha \geq 1{\cdot}0$	Avg.[i] $\Delta\varphi$	Avg.[j] $\Delta\varphi$	
1	112	*V*	0·3	2	281	46	48	40	1·1	142	51	39	3
				3	292	58	60	60	1·4	197	62	52	
2	112	*V*	1·0	2	111	43	40	37	1·0	111	57	40	
3*A*	360	*FJ*	1·0	3	628	37	63	65	2·0	630	70	56	4
3*B*	360	*VJ*	1·0	3	775	44	70	69	2·7	781	73	60	
4	360	*FJ*	1·0	3	843	44	77	78	>3	849	78	70	
	360	*VJ*	1·0	3	857	58	78	79	>3	868	79	77	
5	345[n]	*FJ*	1·0	2	681	44	72	72	2·1	681	76	67	
				3	839	44	76	77	>3	839	77	70	
6	222	*V*	0·3	2	603	53	65	63	1·5	464	66	56	
7*A*	910	*FJ*	0·3	2	1655	45	69	70	>2	1970	71	–	6
				3	1965	48	70	70	>2	2025	71	–	
7*B*	910	*VJ*	0·3	2	1566	57	74	74	>2	2041	76	–	
				3	1980	112	79	79	>2	2074	81	–	
8	910	*FJ*	0·3	1	1800	59	82	81	>2	1715	83	–	
9	313	*F*	0·0*r*	1	1155	52	76	73	1·2	–	76	–	
				1	1330	–	67	66	1·7	–	67	–	
10	800	*FJ*	0·3	3	2405	41	57	53	1·7	2229	58	56	
11	153	*F*	0·3	1	2913	85	78	81	1·0	–	94	–	
	251			1	1302	–	65	64	1·8	–	66	–	
12	325	*F*	0·3	1	2252	77	74	76	1·2	1363	78	73	7
	662			1	2180	–	63	61	>2	1754	64	54	
13	868	*F*	0·3	1′	2118	–	80	80	3·0	1637	80	79	

a. In all runs, only data with observed structure factor greater than ~15 were used.
b. Minimum figure of merit accepted for starting phases.
c. *V*, starting set was allowed to vary; *F*, starting set was held fixed; *J*, only phases were accepted which had changed less than 90° from the previous cycle.
d. $\alpha_{\mathbf{H}} = 2\kappa E_{\mathbf{H}}\{[\sum_{\mathbf{K}} E_{\mathbf{K}}E_{\mathbf{H}-\mathbf{K}} \cos(\varphi_{\mathbf{K}}+\varphi_{\mathbf{H}-\mathbf{K}})]^2 + [\sum_{\mathbf{K}} E_{\mathbf{K}}E_{\mathbf{K}-\mathbf{K}} \sin(\varphi_{\mathbf{K}}+\varphi_{\mathbf{H}-\mathbf{K}})]^2\}^{1/2}$; $\alpha_{\min}$ is minimum value of $\alpha_{\mathbf{H}}$ accepted for phase determinations.
e. $R = \dfrac{\Sigma|E_{\text{obs}} - E_{\text{calc}}|}{\Sigma E_{\text{obs}}}$ The value tabulated is for the accepted reflections only.
f. $\Delta\varphi = |\varphi_{\text{TAN}} - \varphi_{\text{MIR}}|$ or $360° - |\varphi_{\text{TAN}} - \varphi_{\text{MIR}}|$, whichever is smaller. The average is given for all accepted reflections.
g. Average value of $\Delta\varphi$ for all accepted reflections with $E \geq 1{\cdot}25$.
i. Average value of $\Delta\varphi$ for all reflections.
j. Average value of $\Delta\varphi$ for approximately half the reflections with largest values of α.
k. Run 1 utilizes an early version of the program in which the *h*0*l* reflections were incorrectly allowed to take on general phases.
l. Starting set for run 4 was output of run 1 – no MIR phases.
m. Starting set for run 5 was output of run 1 plus MIR phases for the remaining 6 Å reflections.
n. The program used for run 5 eliminated some duplication of terms which occurred in runs 3 and 4.
o. Starting set for run 8 was output of run 6 – no MIR phases.
p. In these runs an overall thermal parameter of 11 was used in the preparation of normalized structure factors.
q. Fringe data plus inner reflections not in starting set (1119 in run 11, 239 in run 12).
r. All determinations were accepted in this run.
s. Starting set reflections only.
t. Starting set for run 13 was 2·8 Å MIR phases plus 2·8–2·0 Å TAN phases from run 12.
u. Data given for fringe reflections only.

the figure of merit for the pure MIR maps. For all maps the same portion of the unit cell was calculated on an undistorted grid, at $-0{\cdot}30 \leq x \leq +0{\cdot}10$, $+0{\cdot}40 \leq y \leq +0{\cdot}50$, and $-0{\cdot}22 \leq z \leq +0{\cdot}10$. Sections are perpendicular to y. This volume, 1/80 the volume of the CPA unit cell, encloses part of the important active site region. The maps are illustrated here as pairs of two-dimensional composites of several sections. The maps have been contoured at equal but arbitrary intervals, except that the highest contours at Zn are omitted. The heavier contours are closer to the reader; the dotted contours are negative. For each map the first composite includes sections $0{\cdot}46 \leq y \leq 0{\cdot}50$ [Figs. 1(*a*), 2(*a*), *etc.*] showing the top of the Zn atom (center top), the portion of extended chain 269–272 including the catalytically important residue Glu 270 (upper left), part of Arg 272 (left edge), chain 197-198 (below Glu 270), and side chain Phe (or His) 279 (bottom center). The second composite includes sections $0{\cdot}40 \leq y \leq 0{\cdot}44$ [Figs. 1(*b*), 2(*b*), *etc.*], showing the Zn atom (X), the Zn ligands His 69 (to right of Zn), Glu 72 (below Zn), and Lys or Glx 196* (left of Zn), and several other smaller portions of the molecule, such as the side chain of

* *Note added in proof:* Residue 196 has now been identified as His in the chemical sequence.

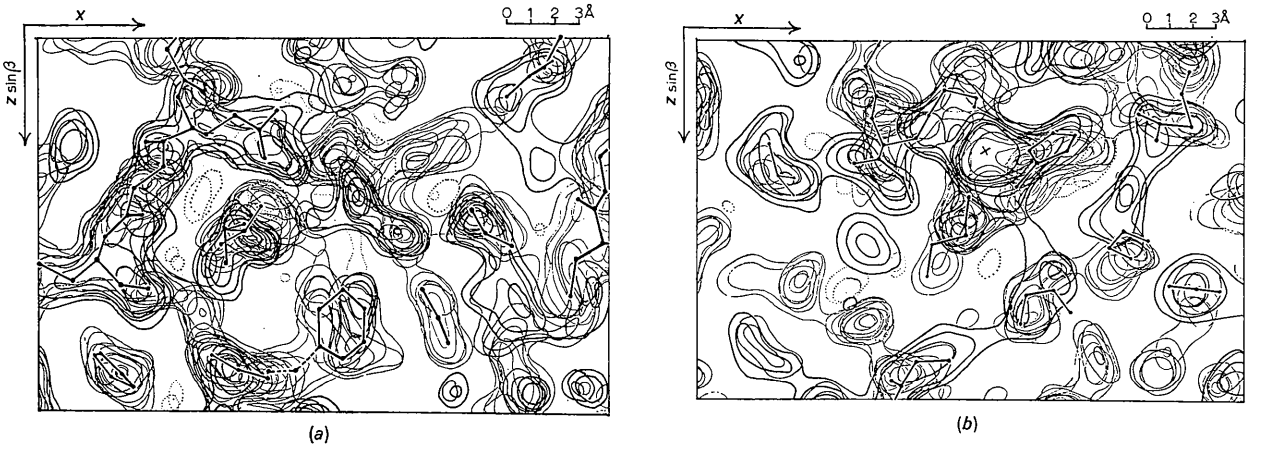


Fig. 1. 2 Å MIR electron density map of CPA, in composites of sections perpendicular to y. Features of the molecule present in these sections are indicated by skeletal model drawings and are described in the text. (*a*) $-0{\cdot}30 \le x \le 0{\cdot}10$, $0{\cdot}46 \le y \le 0{\cdot}50$, $-0{\cdot}22 \le z \le 0{\cdot}10$. (*b*) $-0{\cdot}30 \le x \le 0{\cdot}10$, $0{\cdot}40 \le y \le 0{\cdot}44$, $-0{\cdot}22 \le z \le 0{\cdot}10$. The same portion of the unit cell is shown in all the Figures. The x axis is horizontal.

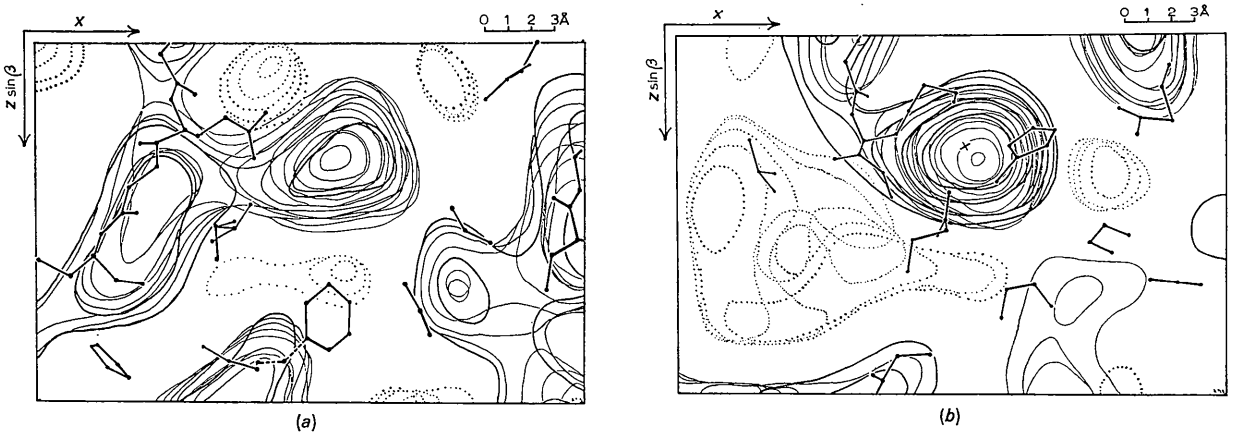


Fig. 2. 6 Å map with MIR phases only.

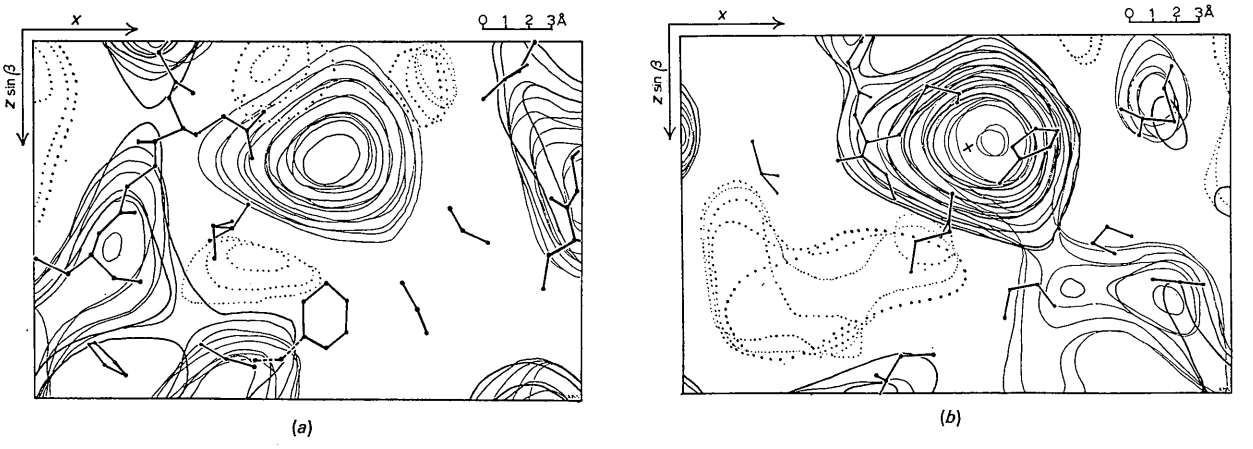


Fig. 3. 6 Å map. 281 TAN phases from run 1, and MIR phases for the remaining reflections.

Leu 271 (left of 196). These features can best be seen in Fig. 1, which is the 2·0 Å map made from MIR phases only. Skeletal model drawings of the parts of the CPA molecule depicted are superimposed on all of the composites. The reader is referred to descriptions of the CPA structure which appear elsewhere (Lipscomb *et al.*, 1969).

Results at low resolution

In run 1 (Table 1) an attempt was made to refine the 6 Å MIR phases. After two cycles, the lowest '*R* factor' and best agreement with MIR phases were obtained. Further cycles produced further improvement in a consistency index (not shown in Table 1) at the expense of accuracy of the phases. This phenomenon was also observed by Weinzierl *et al.* (1969). The result of run 1 can be seen by comparing Fig. 2, the 6 Å MIR map (741 reflections), with Fig. 3, the 6 Å 'TAN' map, in which the same set of reflections was used, but in which the tangent formula phases were substituted for the MIR phases for the 281 reflections phased by the tangent formula. The TAN map appears to be somewhat better, in the sense that the electron density more nearly fits the model deduced from the 2 Å MIR map. All measured low angle reflections were included in these maps, and it therefore appears that the tangent formula is better able to phase these reflections than is the MIR method. Another run was made (run 2, not illustrated), in which only the 111 reflections which gave values of α greater than 1·0 were used in the map. This map was intermediate in quality between Fig. 2 and Fig. 3, indicating that reflections down to $\alpha=0{\cdot}3$ included in Fig. 3 were still contributing useful phase information. Run 2 differed from run 1 only in that the starting set was restricted to reflections with figure of merit greater than 0·40. Apparent results of this change were the increase in the average $\Delta\varphi$ from 51° to 57° and the decrease in number of reflections with $\alpha \geq 1{\cdot}0$ from 142 to 111. The larger starting set thus seems better, even though phases of low reliability are included. A similar effect was observed with runs 11 and 12.

The 6 Å MIR phases were used to calculate phases to 4·5 Å resolution (run 3). When the starting phases were allowed to take on their calculated values after the first cycle (run 3*B*) the results were degraded in comparison with 3*A*, in which the reflections within 6 Å were held at the MIR values, as judged by the Karle & Karle '*R* factor' (internal evidence) and by the average error in phase angle compared with the known 4·5 Å MIR phases. Therefore, the results of run 3*A* were used in the preparation of Fig. 4, which may be compared with the best 6 Å results (Fig. 3) and with the 4·5 Å MIR map (Fig. 5), which has been modified by inclusion of the 6 Å TAN phases of run 1. The maps show that the 4·5 Å TAN refinement does provide some additional detail compared with the 6 Å map. However, also apparent is an excessive sharpening of the map, with negative ripples around the Zn. This problem does not seem to affect areas away from the Zn; however, it becomes more severe at higher resolution, as will be seen. When the starting set for 4·5 Å phase determination was the 6 Å TAN phases (run 4), or the 6 Å combined MIR and TAN phases of Fig. 3 (run 5), the results were not satisfactory, and these runs are not illustrated.

4.7 Å MIR phases formed the starting set for the next round of calculations. Self refinement of the 4·7 Å starting set (run 6) gave a higher '*R* factor' and average $\Delta\varphi$ than the 6 Å self refinement. Nevertheless, 3·5 Å phases were calculated from the 4·7 Å MIR set. As was the case with run 3, results were better when the starting phases were not allowed to take on their calculated values (run 7*A*) and results were quite unsatisfactory when 4·7 Å TAN phases were used as a starting set (run 8). In spite of the very poor agreement with known 3·5 Å MIR phases, a map was made from the results of run 7*A* (Fig. 6). This map had severe negative ripples around the Zn, and is probably not even so good as the 4·5 Å MIR map, although some improvement can be seen in Fig. 6(*a*) in the group at upper right.

Results of high resolution

The calculations at higher resolution, where the method had been expected to work best (Weinzierl *et al.*, 1969), were uniformly disappointing. Phases to 2·8 Å were calculated in a single cycle from a 3·5 Å MIR starting set in run 9 (for the high resolution calculations, Table 1 gives the statistical results separately for the starting set and the other reflections). The results were improved in run 10, where a number of cycles were performed in which the starting phases were held fixed, but the outer reflections were allowed to participate in the phasing and to vary. The map made from the results of run 10 (not illustrated) was no better than the 3·5 Å MIR map.

In runs 11 and 12, which utilized only a single cycle each because of the large computer times required, 2·8 Å MIR phases were extended to phase the 2·0 Å data set. Some improvement in the '*R* factor' and average $\Delta\varphi$ was accomplished in run 12 by increasing the number of starting phases from 1366 to 2246. Because the best phases in the low resolution runs were achieved after the second or third cycle, run 13 was then performed – a single cycle of refinement at 2·0 Å in which the starting set was the 2·8 Å MIR phases plus the 2·8 to 2·0 Å TAN output of run 12. The large starting set (~4600 reflections) required the program to enter a special 'tight core' mode of processing, in which the time for the cycle was 3 hr on the IBM 7094. The Karle & Karle '*R* factor' was not available for run 13, but the average $\Delta\varphi$ was larger than for run 12, so a map was made from results of run 12 (Fig. 7). Fig. 7 can be compared with the 2·8 Å MIR map (Fig. 8) and the 2·0 Å MIR map (Fig. 1). The principal differences between the 2·8 and 2·0 Å MIR maps are to be found in the resolution of the oxygen-containing carbonyl and

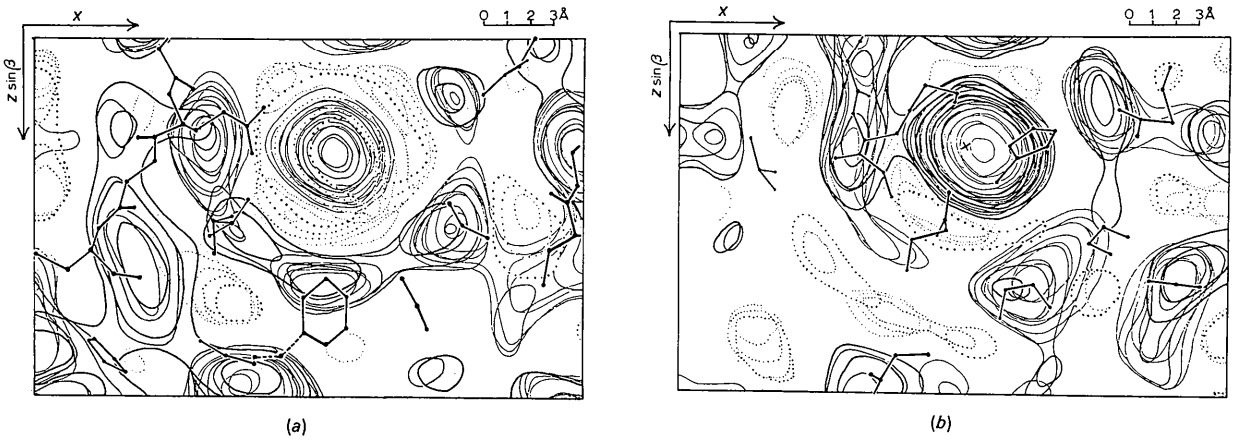


Fig. 4. 4·5 Å map. TAN phases from run 3*A* between 6 and 4·5 Å, and the phases of Fig. 3 within 6 Å.

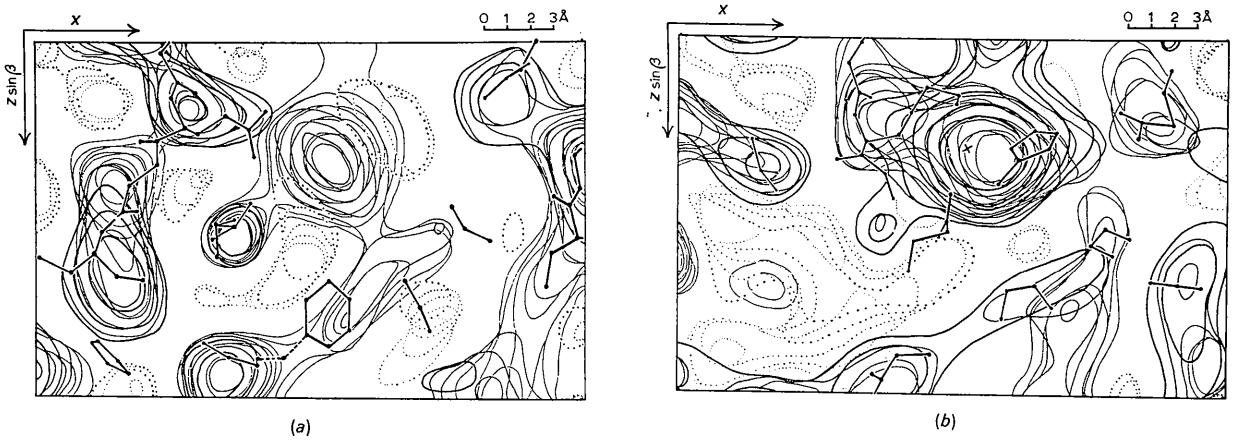


Fig. 5. 4·5 Å map with MIR phases beyond 6 Å, and the phases of Fig. 3 within 6 Å.

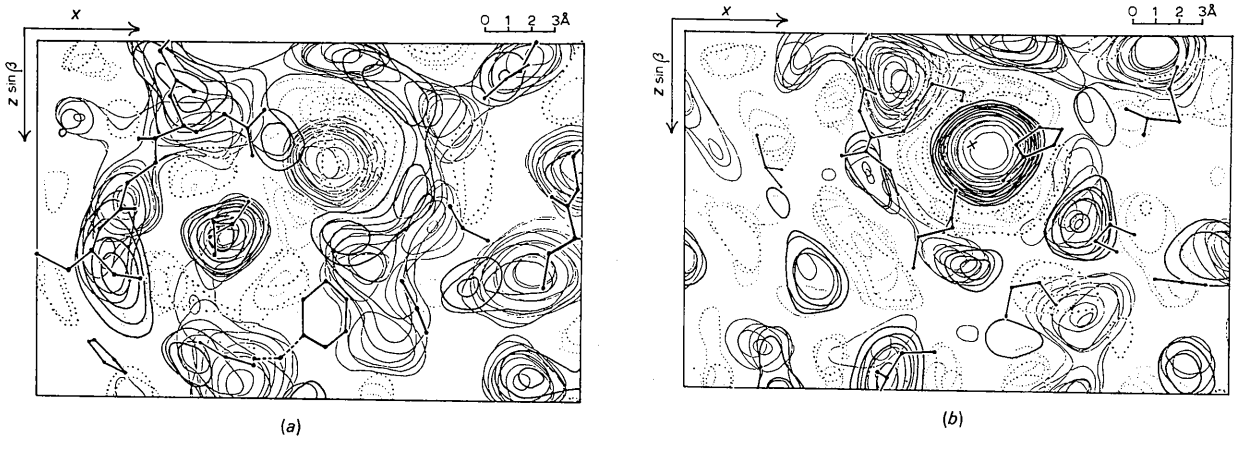


Fig. 6. 3·5 Å map. TAN phases from run 7*A* between 4·7 and 3·5 Å, and MIR phases within 4·7 Å.

carboxyl groups. Four carbonyl groups can be examined in the portion of extended chain at the left in Figs. 1(*a*), 7(*a*), and 8(*a*), and two carboxyl groups are present in these same figures – the carboxyl of Glu 270, top center, and another at top right. The 2·0 Å TAN map, Fig. 7, is little improved in these regions relative to Fig. 8.* It was hypothesized that perhaps Fig. 7 was unable to approach Fig. 1 in quality because only 2250 reflections were included between 2·8 and 2·0 Å, whereas Fig. 1 contains 6000 such reflections. A control map was then made (not illustrated) in which MIR phases were used for all reflections to 2·8 Å and for only those reflections between 2·8 and 2·0 Å which were included in Fig. 7. This control map was intermediate in quality between Fig. 1 and Fig. 8, but was more nearly similar to the 2·0 Å MIR map (Fig. 1). It shows that Fig. 7 could be expected to approach Fig. 1 in quality if the phases were good, and that the smallness of the number of reflections that can be well determined, as judged by large E or α, is not in itself a severe limitation of the tangent formula method. (It also shows, incidentally, that good quality high resolution protein maps can be obtained when only the largest reflections are included. In the case of CPA, only the largest $\frac{1}{2}$ of the reflections between 2·8 and 2·0 Å were measured on the heavy metal derivatives and included in the MIR phasing calculation. It is now apparent that an even smaller fraction could be made to suffice.) A second test was designed to determine whether the phases of run 12 contained any structural information at all. A control

* After a suggestion by one referee that these illustrations do not convincingly demonstrate the conclusions made, the maps were re-examined in three dimensions. While a few groups, notably CO 272, CO_2 270, and CO 269, were somewhat improved in sharpness in Fig. 7 relative to Fig. 8, a number of other, mostly larger, groups were distorted in shape: His 69, His or Phe 279, and His 281. In general, the TAN phases appear to give an increase in resolution accompanied by distortions of some groups. The balance between improvements and distortions is a delicate one, and examination of more cases at this and higher resolutions is suggested.

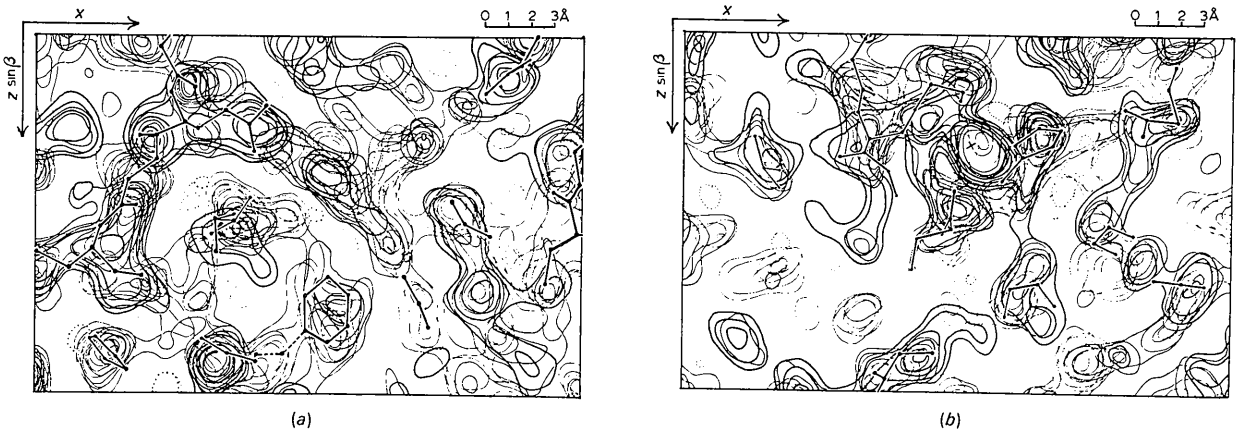


Fig. 7. 2·0 Å map. TAN phases from run 12 between 2·8 and 2·0 Å, and MIR phases within 2·8 Å.

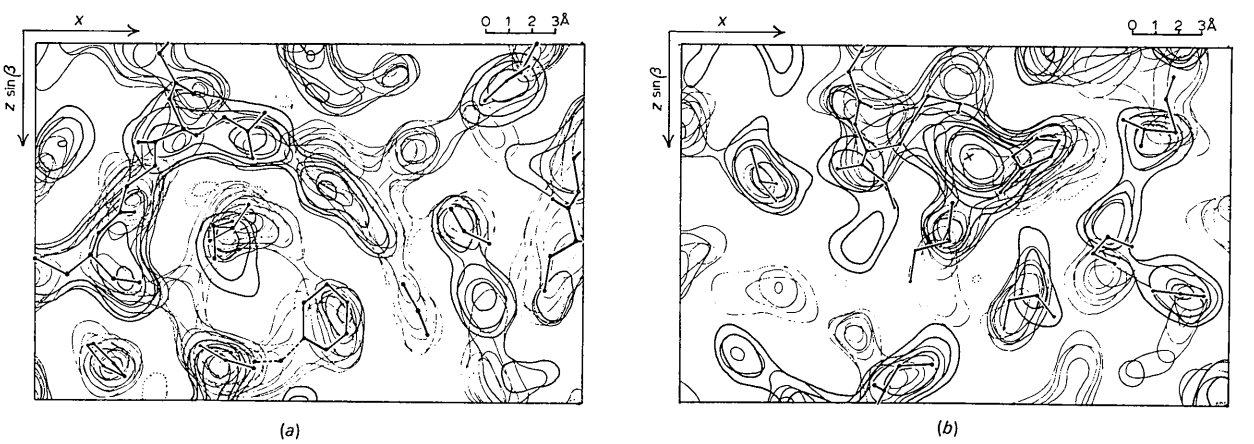


Fig. 8. 2·8 Å MIR map.

map was made from the complete 2·8 Å MIR phases and with phases assigned by a random number generator (uniformly distributed between 0 and 2π) to the reflections between 2·8 and 2·0 Å which were included in Fig. 7. This map (not illustrated) was significantly worse in quality than even Fig. 8, showing that the phases of run 12 were decidedly better than random.

Discussion and conclusions

The findings reported here, if they can be extended to other proteins, show that it is not true that the tangent formula will work better with protein data as the resolution is increased from 6 to 2·0 Å. This result is contrary to theoretical expectation based on inclusion of an assumption of resolvable point atoms in derivations of direct-methods phasing formulas. Perhaps sufficiently high resolution for the success of the tangent formula has not yet been reached – since atoms are not yet resolved at 2·0 Å. The poor results at 2·0 Å can probably also be attributed to the larger errors in the intensity measurements and in the MIR starting phases at the higher resolutions. The usefulness of the tangent formula for extending a set of MIR phases to higher resolution under the conditions used here appears to be limited to data within approximately the 3·5 Å sphere, and even then the method provides only preliminary phase estimates. On the other hand, the method seems to be of definite value in improving low resolution phases, where it can give phases better than those obtainable from MIR for some reflections. These results might be even more striking in a case where the positions of the heavy metal atoms had not yet been refined at high resolution, as had been done with CPA. The improved 6 Å phases were also used in a 2 Å map calculated in the vicinity of residue 196 from the phases of the CPA structure factor calculation, and here also there seems to be some improvement in the resolution of residue 196, which was omitted from the structure factor calculation. Further studies are contemplated on the results obtainable when structure factor calculation phases are refined with the tangent formula.

It is instructive to compare some observations made in these experiments with the experiences of Weinzierl *et al.* for the common features can be expected to occur also with other proteins. In both sets of calculations, the best phases were obtained in the first three or four cycles, and in one cycle at the highest resolutions. Further cycles improved the index of self consistency even as the phases deteriorated. This phenomenon remains unexplained. Contrary to the experiences of Weinzierl *et al.*, however, the '*R* factor' advocated by Karle & Karle was found here to be a fairly good indicator of the overall quality of the tangent formula phases and to be an excellent indicator of the cycle at which convergence was best achieved. There has been less success at estimating the accuracy of individual phase determinations. Both here and in the work of Weinzierl *et al.* the quantity α was only weakly correlated with the correctness of the phases. Table 2 gives the average error, relative to MIR phases, of the phases of run 3*A* and of run 12 as a function of α. The usefulness of α appears to be greatest at the lower resolution, but the error becomes smallest only at very large values of α, where there are too few reflections to be of much use. The average α changes erratically from run to run, so it would be advisable to choose the cut-off value of α after the phasing calculation rather than before. However, in the case of run 12, Table 2 shows that an $\alpha_{\min}$ of 1·0 rather than 0·3 would not have greatly improved the average error. The average error is also not greatly improved by choosing only the largest values of E, as shown by Table 1, unless only the few reflections with very largest E's are considered. Until some better index of correctness is found, reliance must be placed upon pre-selecting reflections with reasonably large E's to be phased, and rejecting those with the lowest α's.

This work would not have been possible without the contributions to the carboxypeptidase structure determination of our present and former colleagues J. C. Coppola, J. A. Hartsuck, M. L. Ludwig, H. Muirhead, F. A. Quiocho, J. Searl and T. A. Steitz. Kathleen Seyfarth prepared the figures for this paper. We are also indebted to Professor R. E. Dickerson for making his results available to us prior to their publication. A National Science Foundation Fellowship to G. N. R. is also acknowledged.

References

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473.
COULTER, C. L. (1965). *J. mol. Biol.* **12**, 292.

Table 2. *Error in phase determination as a function of* α

Run 3*A*					
Range of α	≤1·0	1·0–1·5	1·5–2·0	2·0–3·0	>3·0
Avg. $\Delta\varphi$	84	76	66	63	49
No. of reflections	453	158	132	185	156
Run 12					
Range of α	≦0·5	0·5–0·83	0·83–1·33	1·33–2·0	>2·0
Avg. $\Delta\varphi$	85	88	75	75	67
No. of reflections	340	395	687	563	335

HARKER, D. & KASPER, J. S. (1948). *Acta Cryst.* **1**, 70.
KARLE, J. (1966). *Acta Cryst.* **21**, 273.
KARLE, J. (1968). *Acta Cryst.* B**24**, 182.
KARLE, J. & HAUPTMAN, H. (1950). *Acta Cryst.* **3**, 181.
KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 473.
KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.
KARLE, J. & KARLE, I. L. (1966*a*). *Acta Cryst.* **21**, 849.
KARLE, I. L. & KARLE, J. (1966*b*). *Acta Cryst.* **21**, 860.
LIPSCOMB, W. N., REEKE, G. N., HARTSUCK, J. A., QUIOCHO, F. A. & BETHGE, P. H. (1969). *Proc. Roy. Soc.* B. In the press.
PERUTZ, M. F. & SCATTURIN, V. (1954). *Acta Cryst.* **7**, 799.
REEKE, G. N., HARTSUCK, J. A., LUDWIG, M. L., QUIOCHO, F. A., STEITZ, T. A. & LIPSCOMB, W. N. (1967). *Proc. nat. Acad. Sci. Wash.* **58**, 2220.
ROSSMANN, M. G., JEFFERY, B. A., MAIN, P. & WARREN, S. (1967). *Proc. nat. Acad. Sci. Wash.* **57**, 515.
SAYRE, D. (1952). *Acta Cryst.* **5**, 60.
STEITZ, T. A. (1968). *Acta Cryst.* B**24**, 504.
WEINZIERL, J. E., EISENBERG, D. & DICKERSON, R. E. (1969). *Acta Cryst.* B**25**, 380.
ZACHARIASEN, W. H. (1952). *Acta Cryst.* **5**, 68.

Acta Cryst. (1969). B**25**, 2623

The Molecular Structure of *N*-Phenyl-2,4,6-Trimethylpyridinium Perchlorate

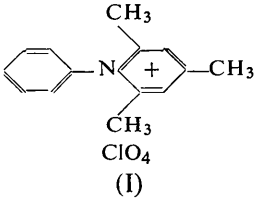
BY ARTHUR CAMERMAN AND L. H. JENSEN
Department of Biological Structure, University of Washington, Seattle, Washington, U.S.A.
AND A. T. BALABAN*
Institute of Atomic Physics, Bucharest, Roumania

(*Received* 27 *August* 1968 *and in revised form* 17 *January* 1969)

N-Phenyl-2,4,6-trimethylpyridinium perchlorate exhibits the nuclear magnetic resonance peak due to the α-methyl protons at higher field than the peak due to the γ-methyl protons, unlike 2,4,6-trimethyl derivatives of pyridine or *N*-alkylpyridinium salts. The compound crystallizes in space group $P2_1/a$ with unit-cell dimensions $a=16{\cdot}619\pm0{\cdot}004$, $b=12{\cdot}110\pm0{\cdot}007$, $c=8{\cdot}340\pm0{\cdot}003$ Å, $\beta=114{\cdot}44\pm0{\cdot}03°$. The structure has been refined to an *R* value of 0·061 using diffractometer data. The planes of the phenyl and pyridinium rings are rotated 83·5° with respect to each other and the C–N bridge bond length after correction for thermal motion is 1·470 Å. These results indicate that the shielding of the α-methyl protons is due to the ring current induced in the *N*-phenyl ring and not to any resonance effect. 14% of the perchlorate ions have disordered oxygen positions and as a result all Cl–O bonds appear shorter than usual.

Introduction

The reaction of pyrylium salts (Balaban, Schroth & Fischer, 1968) with primary amines represents a convenient method for preparing pyridinium salts, especially with *N*-aryl groups, which cannot be obtained by direct quaternization (Brody & Ruby, 1960). Thus 2,4,6-trimethylpyrylium perchlorate and aniline afford in high yield *N*-phenyl-2,4,6-trimethylpyridinium perchlorate (I). Unlike 2,4,6-trimethyl derivatives of pyridine, *N*-alkylpyridinium, or pyrylium salts whose α-methyl groups give rise to n.m.r. peaks at lower fields than the γ-methyl group (Toma & Balaban, 1966; Balaban, Bedford & Katritzky, 1964), compound (I) exhibits the n.m.r. peak due to the six α-methyl protons at higher field than the peak due to the three γ-methyl protons. This supplementary shielding of the α-methyl groups in (I) could be due either to a marked perturbation of the π-electron density in the pyridinium ring, by interaction with the phenyl ring electron cloud, or to an effect of the ring current induced in the phenyl group (Pople, Schneider & Bernstein, 1959) which



(I)

owing to steric hindrance is presumably tilted out of the plane of the pyridinium ring.

In order to test these alternatives, an X-ray diffraction analysis of *N*-phenyl-2,4,6-trimethylpyridinium perchlorate was undertaken. As far as we are aware no other structure determination of an *N*-alkyl or *N*-aryl pyridinium salt has been published.

Experimental

N-Phenyl-2,4,6-trimethylpyridinium perchlorate was prepared (Toma & Balaban, 1966) from 2,4,6-trimethylpyrylium perchlorate and aniline in ethanol. Large yellow prismatic single crystals elongated along ***b*** were grown from ethanol, m.p. 128–129°C. For intensity measurements an approximately cubic crystal

* Present address: IAEA, Kärntnerring 11, Vienna, Austria.