

References

- BERMAN, H. M. & KIM, S. H. (1968). *Acta Cryst.* **B24**, 897.
- BROWN, G. M. & LEVY, H. A. (1965). *Science*, **147**, 1038.
- BUSING, W. R., MARTIN, K. O. & LEVY, H. A. (1962). *ORFLS*. Report ORNL-TM-305, Oak Ridge National Laboratory, Oak Ridge, Tennessee.
- CHU, S. C. & JEFFREY, G. A. (1968). *Acta Cryst.* **B24**, 830.
- CLARK, E. S. & GEERING, G. T. (1970). Papers presented at the ACA Summer Meeting, Aug. 16–22, Carleton Univ., Ottawa, Canada.
- HACKERT, M. L. & JACOBSON, R. A. (1971). *Acta Cryst.* **B27**, 203.
- HORDVIK, A. (1961). *Acta Chem. Scand.* **15**, 16.
- HORDVIK, A. (1966). *Acta Chem. Scand.* **20**, 1943.
- International Tables for X-ray Crystallography* (1952). Vol. I. Birmingham: Kynoch Press.
- NEUMAN, M. A. (1970). *Trans. ACA*, **6**, 111.
- WARSHEL, A. & LIFSON, S. (1970). *J. Chem. Phys.* **52**, 582.
- WILLIAMS, D. E. (1965). *Science*, **147**, 605.
- WILLIAMS, D. E. (1969). *Acta Cryst.* **A25**, 464.
- WILLIAMS, D. E. (1970). *Trans. ACA*, **6**, 21.
- WOBSE, G. & BLASENBREY, S. (1970). *Kolloid Z. Z. Polym.* **241**, 985.

Acta Cryst. (1972). **B28**, 3166

Modified Tangent Procedures*

BY HERBERT HAUPTMAN AND CHARLES M. WEEKS

The Medical Foundation of Buffalo, 73 High Street, Buffalo, New York 14203, U.S.A.

(Received 19 April 1972)

Two procedures for extending a basic set of known phases are described which, unlike the tangent formula, depend on the conditional probability distribution of $\cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})$ as well as the distribution of $\sin(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})$. One of these modified tangent procedures has been applied to one unsolved and three solved crystal structures, and it has been found, in all instances, to be superior to the simple tangent formula when the initial set of phases is small. In addition, the modified tangent procedure, but not the simple tangent formula, is capable of making an enantiomorph selection in space group $P2_1$ when the input set consists only of the origin-fixing phases and those determined by means of Σ_1 .

1. Introduction

One of the most useful relationships which may be derived from the conditional probability distribution of the pair of phases $\varphi_{\mathbf{k}}$ and $\varphi_{-\mathbf{h}-\mathbf{k}}$, where \mathbf{h} is a fixed vector having an unknown phase and \mathbf{k} ranges over all vectors in reciprocal space such that the magnitudes of $|E_{\mathbf{k}}|$ and $|E_{-\mathbf{h}-\mathbf{k}}|$ have fixed values, is the tangent formula (Karle & Hauptman, 1956). Not only is the tangent formula a widely used technique for the solution of equal atom problems in noncentrosymmetric space groups, but it is also useful for refining and extrapolating approximate phases derived in any way, e.g. when phases are calculated from a known fragment of a crystal structure (Karle, 1968).

The simple tangent formula depends on the conditional probability distribution of $\sin Y$, where $Y = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}$, but not on the distribution of $\cos Y$. The purpose of the present paper is to discuss tangent procedures which make more effective use of the probability distribution of the pair of phases $\varphi_{\mathbf{k}}$ and $\varphi_{-\mathbf{h}-\mathbf{k}}$ by forcing two of its parameters, the average values of

$\sin Y$ and $\cos Y$ to agree as closely as possible with their theoretical expected values.

2. Derivation of the modified tangent procedures

Suppose that the space group is $P1$ and that there are N identical atoms in the unit cell. Define $A_{\mathbf{k}}$ by

$$A_{\mathbf{k}} = \frac{2}{N^{1/2}} |E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{h}+\mathbf{k}}|. \quad (2.1)$$

Assume that the reciprocal vector \mathbf{h} is fixed and that \mathbf{k} , the primitive random variable, ranges uniformly over that region of reciprocal space for which $A_{\mathbf{k}}$ has a fixed, specified value. Under these circumstances the conditional probability distribution of the pair of phases $\varphi_{\mathbf{k}}$, $\varphi_{-\mathbf{h}-\mathbf{k}}$, given $A_{\mathbf{k}}$, is known to be (Cochran, 1955; Hauptman, 1972a)

$$P(\Phi_0, \Phi_1 | A_{\mathbf{k}}) \simeq \frac{1}{4\pi^2 I_0(A_{\mathbf{k}})} \times \exp \{A_{\mathbf{k}} \cos(\varphi_{\mathbf{h}} + \Phi_0 + \Phi_1)\} \quad (2.2)$$

where I_0 is the modified Bessel function. From (2.2) it follows readily that (Hauptman, 1966, 1972a; Karle & Karle, 1966)

* Portions of this paper were presented at the meetings of the American Crystallographic Association at Columbia, South Carolina, February 1971, and Ames, Iowa, August 1971.

$$\langle \sin(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) | A_{\mathbf{k}} \rangle_{\mathbf{k}} = 0, \quad (2.3)$$

$$\langle \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) | A_{\mathbf{k}} \rangle_{\mathbf{k}} = \frac{I_1(A_{\mathbf{k}})}{I_0(A_{\mathbf{k}})}, \quad (2.4)$$

in which the averages are extended over all vectors \mathbf{k} in reciprocal space for which $A_{\mathbf{k}}$ has a fixed, specified value. An immediate consequence of (2.3) is the simple tangent formula,

$$\tan \varphi_{\mathbf{h}} = \frac{\langle |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \sin(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) \rangle_{\mathbf{k}}}{\langle |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \cos(\varphi_{\mathbf{k}} + \varphi_{\mathbf{h}-\mathbf{k}}) \rangle_{\mathbf{k}}} \quad (2.5)$$

in which the averages are extended over the same, arbitrary set of vectors \mathbf{k} in reciprocal space. Thus (2.5) depends only on (2.3) which in turn is a consequence of the property of (2.2) that it is an even function of $\varphi_{\mathbf{h}} + \varphi_{\mathbf{0}} + \varphi_{\mathbf{1}}$, a relatively weak property of the distribution function. No use is made of (2.4) and the ratio of Bessel functions does not appear in (2.5). If one attempts to make stronger use of the probability distribution (2.2) by employing (2.4) as well as (2.3), one is led to the modified tangent procedures.

If the random variable Y is defined by

$$Y = \varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}} \quad (2.6)$$

and appropriate weights (Hauptman, 1972a)

$$w_{s\mathbf{k}} = [\text{Variance } \sin Y]^{-1/2} = \left[\frac{I_1(A_{\mathbf{k}})}{A_{\mathbf{k}} I_0(A_{\mathbf{k}})} \right]^{-1/2} \quad (2.7)$$

and

$$w_{c\mathbf{k}} = [\text{Variance } \cos Y]^{-1/2} \\ = \left[1 - \frac{I_1(A_{\mathbf{k}})}{A_{\mathbf{k}} I_0(A_{\mathbf{k}})} - \frac{I_1^2(A_{\mathbf{k}})}{I_0^2(A_{\mathbf{k}})} \right]^{-1/2} \quad (2.8)$$

are assigned, one may then attempt to determine the unknown phase $\varphi_{\mathbf{h}}$ by minimizing, in accordance with the principle of least squares, either

$$\Phi = \left\langle w_{s\mathbf{k}}^p \sin^2 Y + w_{c\mathbf{k}}^p \left[\cos Y - \frac{I_1(A_{\mathbf{k}})}{I_0(A_{\mathbf{k}})} \right]^2 \right\rangle_{\mathbf{k}} \quad (2.9)$$

or

$$\Phi' = \langle w_{s\mathbf{k}}^p \sin Y \rangle_{\mathbf{k}}^2 + \left\langle w_{c\mathbf{k}}^p \left[\cos Y - \frac{I_1(A_{\mathbf{k}})}{I_0(A_{\mathbf{k}})} \right]^2 \right\rangle_{\mathbf{k}} \quad (2.10)$$

in which the averages are taken over an arbitrary set of vectors \mathbf{k} in reciprocal space and $p = 1$ or 2 . These modified tangent procedures are substitutes for the tangent formula and, since they make stronger use of the probability distribution of the phases, they presumably constitute an improvement over the tangent formula. Neither (2.9) nor (2.10) may be solved explicitly for the unknown, $\varphi_{\mathbf{h}}$; consequently, it is necessary to evaluate the function at a number of points and to select the value of $\varphi_{\mathbf{h}}$ corresponding to the minimum in the curve. The tangent formula itself gives a unique value for each phase, but introduction of (2.4) may result in the presence of two minima in the function Φ . Equation (2.9) may be rewritten in the form

$$\Phi = \frac{1}{2} C_2 \cos 2\varphi_{\mathbf{h}} - \frac{1}{2} S_2 \sin 2\varphi_{\mathbf{h}} - 2C_1 \cos \varphi_{\mathbf{h}} + 2S_1 \sin \varphi_{\mathbf{h}} + C, \quad (2.11)$$

where

$$C_2 = \langle (w_{c\mathbf{k}} - w_{s\mathbf{k}}) \cos 2(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \rangle_{\mathbf{k}} \quad (2.12)$$

$$S_2 = \langle (w_{c\mathbf{k}} - w_{s\mathbf{k}}) \sin 2(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \rangle_{\mathbf{k}} \quad (2.13)$$

$$C_1 = \left\langle w_{c\mathbf{k}} \frac{I_1(A_{\mathbf{k}})}{I_0(A_{\mathbf{k}})} \cos(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \right\rangle_{\mathbf{k}} \quad (2.14)$$

$$S_1 = \left\langle w_{c\mathbf{k}} \frac{I_1(A_{\mathbf{k}})}{I_0(A_{\mathbf{k}})} \sin(\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \right\rangle_{\mathbf{k}} \quad (2.15)$$

and

$$C = \left\langle w_{c\mathbf{k}} \left[1 + \left(\frac{I_1(A_{\mathbf{k}})}{I_0(A_{\mathbf{k}})} \right)^2 \right] + w_{s\mathbf{k}} \right\rangle_{\mathbf{k}}, \quad (2.16)$$

and evaluated more easily than (2.10) on a digital computer. In the applications which have been made to date, equation (2.11), which will be referred to as the modified tangent formula, has been used and Φ has been evaluated at 100 equal intervals. The formulas used in this least-squares procedure [(2.11)–(2.16)] are similar to those used in the method of the least-squares analysis of structure invariants (Hauptman, Fisher, Hancock & Norton, 1969, equations (4.4)–(4.9); Hauptman, 1970; Hauptman, Fisher & Weeks, 1971), but here it is the sum of the squares of the differences between the cosine invariants $X = \cos(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})$ and their expected values (rather than their calculated values as was done in the earlier work) which, together with the contribution from the squares of $\sin(\varphi_{\mathbf{h}} + \varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})$, is minimized.

The reliability of the determination of an individual phase, $\varphi_{\mathbf{h}}$, is indicated by the figure of merit or residual,

$$R_{\mathbf{h}} = \Phi_{\min}^{1/2}, \quad (2.17)$$

where Φ_{\min} is the minimum value of Φ . Clearly, the smaller the value of $R_{\mathbf{h}}$, the better. Next, if a set of phases $\varphi_{\mathbf{h}}$ is determined simultaneously from the same basic set of known phases, then one defines first

$$A_{\mathbf{h}} = \sum_{\mathbf{k}} A_{\mathbf{k}} \quad (2.18)$$

where the sum is taken over all the vectors \mathbf{k} contributing to the value of Φ in (2.9), or (2.11)–(2.16), recalling that, for fixed \mathbf{h} , $A_{\mathbf{k}}$ is a function of \mathbf{k} alone. With this definition for $A_{\mathbf{h}}$, the (cycle) modified tangent figure of merit is defined by

$$R = \frac{\sum_{\mathbf{h}} A_{\mathbf{h}} R_{\mathbf{h}}}{\sum_{\mathbf{h}} A_{\mathbf{h}}} \quad (2.19)$$

which is the weighted average of the minima for all phases in the cycle. Clearly R is small when the phase determination is proceeding correctly. It is found in practice that, for structures of moderate complexity (20–30 independent nonhydrogen atoms) R values in the interval 0.80–0.95 usually indicate correct phasing, and R values in excess of 0.95 are evidence of incorrect

phasing. However, for structures of great complexity (60 or more independent nonhydrogen atoms), larger values for R are to be expected.

3. Application of the modified tangent procedure to three known structures

The efficacy of the modified tangent procedure (2.11), and especially its utility in comparison with the simple tangent formula, (2.5), were first tested on the data for three known steroid crystal structures, each of which has 20–30 nonhydrogen atoms in the asymmetric unit. These structures are epiandrosterone (Weeks, Cooper, Norton, Hauptman & Fisher, 1971), estradiol–urea complex (1:1) (Duax, 1972), and 6 α -fluorocortisol (Duax, to be published). The first of these structures crystallizes in space group $P2_1$, and the other two crystallize in space group $P2_12_12_1$.

In the first test of the modified tangent formula, the input consisted of a large set of correct phases which were refined using both tangent procedures. Since it was desired, in these initial applications, to give approximately equal weights to the contributions of $\sin Y$ and $\cos Y$ to the function minimized by least-squares, and since the variance of $\cos Y$ is much less than that of $\sin Y$ (Fisher, Hancock & Hauptman, 1970), the weights were made proportional to the reciprocals of the corresponding standard deviations rather than the variances [*i.e.* $p=1$ in (2.9)]. Table 1 gives a comparison of the results for four cycles of refinement of the true phases for reflections with $|E| \geq 1.3$, for each of the three structures by means of both the simple and modified tangent formulas. Although large sets of accurate phases are not available at the point at which the tangent procedures are normally applied to an unknown structure, the results of this sort of experiment indicate the maximum accuracy which can be expected. It was found that, for all three structures, the average deviation of the predicted phases from their true values ($\varphi_{\text{true}} - \varphi_{\text{predicted}}$) was approximately zero, and the average absolute value of the deviation was in the range 0.2–0.3 radians for both formulas. Thus, the formulas are about equally good under these circumstances, and the simple tangent formula would be preferred because it requires less computing time.

Table 1. Comparison of simple (top figure) and modified tangent formulas following 4 cycles of refinement of the true phases of all vectors with $|E| \geq 1.3$ using all triples with $A \geq 1.0$

Structure	The deviations are ($\varphi_{\text{true}} - \varphi_{\text{predicted}}$).			
	Number of phases	Number of triples	$\langle \text{Deviation} \rangle$	$\langle \text{Deviation} \rangle$
Epiandrosterone	289	5030	0.01 rad	0.22 rad
Estradiol–urea complex	359	5981	–0.03	0.20
6 α -Fluorocortisol	367	3338	0.02	0.29
			0.03	0.28

Once it had been established that the modified tangent formula was at least as accurate as the simple tangent formula in an idealized case, attention was directed to the type of situation encountered in the solution of an unknown structure, and the structures of the estradiol–urea complex and 6 α -fluorocortisol were resolved with the modified tangent procedure using the same basic sets of phases which had previously been used to solve the structures by means of the simple tangent formula. In both cases, after selection of the origin and enantiomorph, additional phases were determined through the two-dimensional structure invariants $X = \cos(\varphi_{\text{h}} + \varphi_{\text{k}} + \varphi_{-\text{h}-\text{k}})$ before beginning the tangent formula application (Duax, Weeks & Hauptman, 1972), and a total of 300 reflections were then phased in the tangent formula. Table 2 summarizes the results of these studies. As in the cases involving a large set of correct input phases, the average deviations of the phases from their correct values are near zero regardless of which formula is used, but the average absolute value of the deviation for the phases computed by the simple tangent formula was 1.5 times as great as the average absolute deviation for the modified tangent phases in the case of 6 α -fluorocortisol and twice as great in the case of the estradiol–urea complex. Although the phases derived by both tangent techniques resulted in intelligible maps for these structures, the more correct modified tangent phases produced cleaner maps as shown by the data in Table 3. Four more atoms were discernible on the 6 α -fluorocortisol E map phased by the modified tangent formula than on the corresponding map phased by the simple tangent formula. Thus, although the modified tangent procedure was superior in both of these applications, the differences were not decisive. A more clear-cut comparison is given in the sequel (§5).

Table 2. Comparison of the simple (top figures) and modified tangent formulas for the phase build-ups used to solve the estradiol–urea complex and 6 α -fluorocortisol

The deviations for the final tangent cycles, in radians, are ($\varphi_{\text{true}} - \varphi_{\text{predicted}}$).

Structure	Estradiol–urea complex (1:1)	6 α -Fluorocortisol
Number of input phases	62	51
Number of incorrect input phases	4	5
$\langle \text{Deviation} \rangle$	0.08	0.06
	0.00	0.03
$\langle \text{Deviation} \rangle$	0.59	0.50
	0.28	0.33

It is instructive to examine some typical examples of the modified tangent minimization function, Φ , and in Figs. 1 and 2, $\Phi^{1/2}$ has been plotted as a function of phase angle for two reflections from the estradiol–urea complex. These curves were obtained by dividing the total

Table 3. Comparison of Fourier maps for the estradiol-urea complex and 6 α -fluorocortisol which were computed using phases calculated by the simple (top figure) and modified tangent formulas

Structure	Estradiol-urea complex	6 α -Fluorocortisol
Number of nonhydrogen atoms in asymmetric unit	24	27
Number of atoms found	22	22
	23	26
Average distance from correct position (Å)	0.23	0.12
	0.15	0.10

possible range of the phases into 100 equal intervals (*i.e.* intervals of 3.6°) and evaluating Φ in each interval. The curves are plotted for each of the seven cycles performed in the modified tangent run used to resolve this structure.

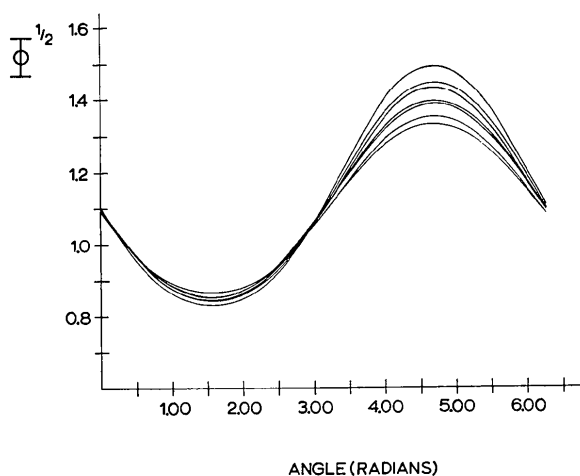


Fig. 1. $\Phi^{1/2}$ [Equation, (2.11)] as a function of ϕ for the well-determined phase ϕ_{011} (estradiol-urea).

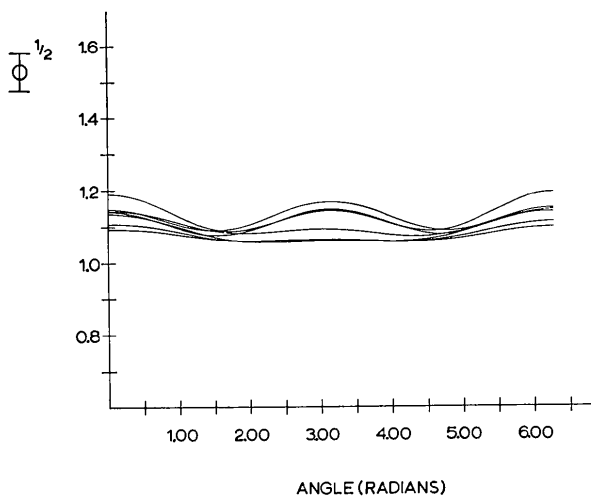


Fig. 2. $\Phi^{1/2}$ [Equation, (2.11)] as a function of ϕ for the poorly determined phase ϕ_{022} (estradiol-urea).

Reflection 011 has a well-determined phase. The curves for this reflection are sinusoidal, and they remain approximately constant throughout all seven tangent cycles. Each curve has a single well-defined minimum occurring at one of the two phase values ($\pi/2$ radians) allowed for this reflection by space group symmetry. The minima are low and give good figures of merit (*i.e.* $R_h \approx 0.85$).

In contrast, reflection 022 has a poorly determined phase. The curves are nearly flat, and they give no clear indication for the value of this phase. Although it is difficult to see, there are actually two minima in each of the curves, these minima occur at widely separated values of the phase, and they have the same poor high figures of merit (R_h in the range 1.06–1.09 for the various cycles). The true value of ϕ_{022} is 0. The simple tangent formula calculates 0 in cycles 1 and 3 and π in all other cycles. The modified tangent formula has minima at $\pi/2$ and $-\pi/2$ in cycle 1, and the positions of the minima move closer to π in subsequent cycles. In cycle 7, the minima corresponded to phase values of -2.03 and 2.08 radians.

The reason for the disparate behavior of the two phases is clear from a *post mortem* examination of the average values of the cosine seminvariants, $\cos(\phi_h + \phi_k + \phi_{-h-k})$ for the Σ_2 triples which were used in the tangent procedures and which involve these reflections. The average A value for the triples involving 011 was 1.95, and the average cosine value was 0.57 in acceptable agreement with the theoretical expected value of 0.69. On the other hand, the triples involving 022 (for which the average A was 1.33) had an average cosine value of 0.10, in disagreement with the theoretical expected value of 0.55. This low average cosine value for 022 reflects the fact that this phase is, in a sense, 'out of phase' with the majority of estradiol-urea reflections, and this explains why the phase was so poorly determined.

4. Enantiomorph selection by the modified tangent procedure

One of the difficulties in the application of direct methods to structures in a noncentrosymmetric space group like $P2_1$ where all restricted phases have the same two possible values lies in the selection of the enantiomorph. On the other hand, in space group $P2_12_12_1$, enantiomorph selection is simplified because some phases are restricted to values of 0 or π and other phases are restricted to $\pm\pi/2$. In space groups of the $P2_1$ type, enantiomorph discrimination depends on the accurate identification of some cosine seminvariants, $\cos(\phi_h + \phi_k + \phi_{-h-k})$, whose values are far from +1 or -1 and will follow from the least-squares analysis of such seminvariants (Hauptman, Fisher, Hancock & Norton, 1969). Alternatively, this enantiomorph discrimination may be accomplished by the modified tangent procedure. When the first enantiomorph sensitive phase is determined by this procedure, there will be two equal minima in the minimization function Φ . The phase

corresponding to one minimum will be the phase for one enantiomorph whereas the other minimum will occur at the phase for the other enantiomorph. In contrast, in cases where all starting phases have values equal to one of the two cardinal points (0 or π), no values other than these may be calculated for any phase using the simple tangent formula (2.5) except through the accumulation of round-off errors.

The enantiomorph distinguishing ability of the modified tangent procedure may be demonstrated by considering what would have happened had the tangent techniques been applied to the starting set of six phases (three phases determined by the origin specifying procedure and three additional phases determined by the Σ_1 formula) for epiandrosterone. In the actual structure solution (Hauptman, Fisher & Weeks, 1971), least-squares analysis of the cosine seminvariants was used to derive a larger set of phases, including phases sensitive to the enantiomorph, before application of the simple tangent formula. Table 4 shows the progress of the enantiomorph selection by the modified tangent procedure. During the first three cycles, five three-dimensional phases ($51\bar{3}$, $11\bar{2}$, $51\bar{2}$, 310 , and $62\bar{4}$) were determined, but all were calculated to have cardinal point values by both the simple and modified tangent formulas. In the fourth cycle, six additional phases ($41\bar{1}$ through $62\bar{5}$) were determined, and one of these, $\varphi_{41\bar{1}}$, was calculated by the modified tangent formula to have two minima, and these minima did not correspond to cardinal point values for the phase. It may be seen from the second column in Table 4 that, of all the three-dimensional reflections for which phase values had been determined up to this point, reflection $41\bar{1}$ was the most sensitive to the enantiomorph. The absolute value

of the deviation of the true phase of $41\bar{1}$ from the nearest cardinal point was 1.48 radians. The maximum value of this deviation is 1.57 radians, and the closer the value of the deviation to 1.57, the greater the difference between the values of the phase in question for the two enantiomorphs. For example, the difference between the values of $\varphi_{41\bar{1}}$ for the two enantiomorphs is 2.96 radians, but this difference is only 0.12 radians in case of the enantiomorph insensitive reflection $51\bar{3}$.

All phases were calculated, by the simple tangent formula, to have cardinal point values in the fourth cycle, and all phases except $\varphi_{41\bar{1}}$ were calculated to be equal to the cardinal point nearest to the true phase. The same results were obtained for the simple tangent formula in the fifth cycle. However, the phases of reflections which depended on $\varphi_{41\bar{1}}$ were not calculated to have cardinal point values by the modified tangent formula. Only one ($51\bar{3}$), of the seven phases which were calculated off cardinal points by the modified tangent formula in cycle 5 was less accurate than the phase calculated by the simple tangent formula.

5. 9-t-Butyl-9,10-dihydroanthracene

The modified tangent procedure has been shown to result in more accurate phasing than the simple tangent formula when applied to the data for three solved steroid crystal structures. However, the simple tangent formula had been adequate to solve these structures and, based on these results alone, it was questionable whether the modified tangent procedure would allow for a structure solution in a case where the tangent formula itself was unsuccessful. The analysis of 9-t-butyl-9,10-dihydroanthracene ($C_{18}H_{20}$), a hydrocarbon crys-

Table 4. Enantiomorph selection for epiandrosterone (space group $P2_1$) by the modified tangent procedure

Vector	Predicted phase		True phase	Cardinal pt nearest true value	Dev. true phase from nearest cardinal pt.		Dev. predicted phase from nearest true phase		Dev. predicted phase from true phase		A-B
	Single	Modified			Simple	Modified	A= Simple	B= Modified			
4 1 $\bar{1}^*\dagger$	0	1.35	1.66	3.14	1.48	3.14	1.79	1.66	0.31	1.35	
5 1 $\bar{3}^\ddagger$	0	-0.18	-0.06	0	0.06	0	0.18	0.06	0.12	-0.06	
1 1 $\bar{2}^\ddagger$	3.14	-2.98	-2.16	3.14	0.98	0	0.16	0.98	0.82	0.16	
5 1 $\bar{2}$	3.14	3.14	2.98	3.14	0.17	0	0	0.17	0.17	0	
3 1 0	0	0	-0.25	0	0.25	0	0	0.25	0.25	0	
6 2 $\bar{4}$	0	0	0.05	0	0.05	0	0	0.05	0.05	0	
4 1 $\bar{1}^*$	0	1.06	1.66	3.14	1.48	3.14	2.08	1.66	0.60	1.06	
2 2 $\bar{3}$	3.14	3.14	-2.08	3.14	1.06	0	0	1.06	1.06	0	
2 2 $\bar{2}$	0	0	0.11	0	0.11	0	0	0.11	0.11	0	
3 1 $\bar{1}$	3.14	3.14	2.27	3.14	0.87	0	0	0.87	0.87	0	
3 1 1^\ddagger	3.14	-2.94	-2.10	3.14	1.04	0	0.20	1.04	0.84	0.20	
6 2 $\bar{5}$	3.14	3.14	-2.52	3.14	0.62	0	0	0.62	0.62	0	
1 1 0	3.14	3.14	1.85	3.14	1.29	0	0	1.29	1.29	0	
1 2 $\bar{1}^\ddagger$	0	0.21	0.80	0	0.80	0	0.21	0.80	0.59	0.21	
2 2 1	3.14	3.14	2.19	3.14	0.95	0	0	0.95	0.95	0	
3 2 0^\ddagger	3.14	-2.47	-2.05	3.14	1.09	0	0.67	1.09	0.42	0.67	
4 1 $\bar{2}^\ddagger$	3.14	3.00	2.73	3.14	0.41	0	0.14	0.41	0.27	0.14	
4 2 $\bar{2}$	3.14	3.14	1.73	3.14	1.41	0	0	1.41	1.41	0	
7 1 $\bar{5}$	3.14	3.14	-3.04	3.14	0.10	0	0	0.10	0.10	0	

* Enantiomorph selecting phase.

† Phase assignment dependent on $\varphi_{41\bar{1}}$.

‡ Calculations for cycle 4. All other rows pertain to cycle 5 of the tangent procedures as described in the text.

tallizing in space group $P2_12_12_1$ (Brennan, Putkey & Sundaralingam, 1971), provided an example where this situation occurred.

A base set of 29 two-dimensional phases was obtained for 9-*t*-butyl-9,10-dihydroanthracene through the use of a variety of techniques, and these starting phases are recorded in Table 5. Phases of $\pi/2$ were assigned to reflections 091, 190, 0,3,14, and 105 in order to specify the origin and enantiomorph, and the \sum_1 and squared-tangent formulas (Hauptman, 1970, 1972*b*) indicated the phases of 0,10,0, 006 and 200 to be π . It should be noted that, while specifying arbitrarily the values of φ_{091} , φ_{190} , $\varphi_{0,3,14}$, φ_{105} does serve to fix the origin and enantiomorph so that all phases are uniquely determined, it is necessary to assign a phase to a reflection hkl where k is not divisible by three in order to make all phases accessible from \sum_2 type relationships alone. In this case, assigning $\varphi_{0,10,0}$ the value determined by \sum_1 served this purpose.

The \sum_2 triples with A values greater than 1.3 which involved only two-dimensional reflections with $|E|$ greater than 1.2 were generated, and the cosine seminvariants, $\cos(\varphi_h + \varphi_k + \varphi_{-h-k})$, for these triples were computed by both the modified triple product and MDKS formulas (Hauptman, Fisher, Hancock & Norton, 1969; Hauptman, 1970, 1972*b*). The triples

were then divided into four groups having approximately constant A values and, within each group, they were first ranked according to the modified triple product computed cosines and then according to the MDKS cosines. Acceptance criteria, as shown in Table 6, were then assigned to each of the A ranges, and a triple was assumed to have a cosine of unity and was used to determine a new phase only if its cosine, as computed by both formulas, ranked among the top $T\%$ for triples in that range. Similarly, triples were assumed to have cosine values of -1 , and were used to determine new phases, if their computed cosines ranked among the bottom $B\%$ for the A range in question.

Table 6. Acceptance criteria for the two-dimensional triples

A Range	Average A	Number of triples in range	$T\%$	$B\%$
1.3- 1.5	1.40	69	50%	5%
1.5- 2.0	1.73	69	60	2
2.0- 3.0	2.38	60	80	2
3.0-10.5	5.00	46	90	0

At A values of 2.0, 12% of the cosine seminvariants for triples of two-dimensional reflections have values

Table 5. Initial two-dimensional phases for 9-*t*-butyl-9,10-dihydroanthracene

Serial	Reflection	$ E $	Phase (radians)	Reason
1	0 9 1	4.02	$\pi/2$	Define origin and enantiomorph $\sum_1 P_+ = 0.00$; squared-tangent $\cos \varphi_{1D}/K = -16\ddagger$ $\sum_1 P_+ = 0.01$; squared-tangent $\cos \varphi_{1D}/K = -35\ddagger$ Squared-tangent $\cos \varphi_{1D}/K = -43\ddagger$
2	1 9 0	3.25	$\pi/2$	
3	0 3 14	2.90	$\pi/2$	
4	1 0 5	3.09	$\pi/2$	
5	0 10 0	4.71	π	
6	0 0 6	4.08	π	
7	2 0 0	1.68	π	Cycle 1
8	0 9 5	2.80	$-\pi/2$	
9	0 7 14	2.66	$-\pi/2$	
10	1 0 1	1.75	$\pi/2$	Cycle 2 (105; 103) pair*
11	0 10 6	1.63	0	
12	0 9 7	1.42	$-\pi/2$	
13	1 0 11	1.66	$-\pi/2$	Cycle 3
14	0 2 13	1.85	0	
15	0 2 15	1.80	0	
16	1 0 3	2.05	$\pi/2$	Cycle 4
17	2 0 2	2.11	0	
18	2 0 8	2.89	π	
19	3 0 3	1.33	$\pi/2$	Cycle 5, $\cos(\varphi_{0,2,13} + \varphi_{0,3,8} + \varphi_{0,5,5}) = -1$
20	0 3 8	1.41	$-\pi/2$	
21	0 1 1	1.21	$\pi/2$	
22	0 8 0	1.30	0	Cycle 6
23	0 5 5	2.77	$-\pi/2$	
24	0 2 3	2.80	π	
25	0 5 11	2.90	$\pi/2$	Cycle 7
26	0 8 3	1.41	π	
27	0 11 4	1.91	$-\pi/2$	
28	0 4 10	1.80	0	
29	0 5 17	2.35	$-\pi/2$	

* Grant, Howells & Rogers (1957), Hauptman (1971, 1972*b*).

† Hauptman (1972*b*), Hauptman & Duax (1972).

of -1 . Using only those triples with A values of two or greater whose cosines were among the top 80% when computed by both formulas, six additional reflections (reflections with serial numbers 8 to 13 in Table 5) could be related to the origin and \sum_1 reflections, and a second cycle of this procedure yielded two more phases. Reflection 103 was then introduced into the known set because it was strongly paired to 105 (Grant, Howells & Rogers, 1957; Hauptman, 1971, 1972b) and the inclusion of this phase allowed four additional phases to be determined from only those triples having $A > 2$. Reflection 011 was included despite its relatively low normalized structure factor amplitude ($|E_{011}| = 1.21$) because it could be obtained from a triple with an A value of 5.2 for which the cosine values calculated by the triple product and MDKS formulas were 1.35 and 1.43 respectively.

It was considered to be important to obtain $\varphi_{0.5.5}$ because knowledge of this phase would make several additional phases accessible, but it was necessary to use triples with lower A values in order to reach this phase. The triple product value for $\cos(\varphi_{011016} + \varphi_{01378} + \varphi_{01714})$ was 1.99, and this triple was used to find φ_{038} even though the A value was only 1.4. Reflection 038 provided a bridge which led to 055 by means of $\cos(\varphi_{01213} + \varphi_{038} + \varphi_{055})$ for which the computed values were -0.27 (triple product) and -0.64 (MDKS). This was the only triple from the bottom $B\%$ in any A range which was actually used to find a new phase. Two more cycles were then based on these two-dimensional triples, and six more phases were obtained.

The 29 phases found by means of this analysis of the two-dimensional triples, which were all later shown to be correct, were used as input to both the simple tangent and modified tangent formulas, and 221 more phases were found during the course of ten tangent cycles. The total number of phases known at the end of each cycle and the modified tangent cycle figures of merit, R cycle (2.19) are given in Table 7. Although the figures of merit for the final two or three cycles seemed to be a little high, the modified tangent phases yielded an excellent E map nevertheless. The highest 18 peaks on this E map had a r.m.s. deviation of 0.10 \AA from the refined positions for the 18 carbon atoms, and the initial R value was 22%. In contrast, the simple tangent phases produced an E map in which 13 of the top 50 peaks were within 0.50 \AA of the carbon atom positions, and formed a recognizable fragment (after the true solution was known) of the enantiomorph of the original molecule referred to a different origin! Table 8 shows the average absolute deviations of the 250 modified and simple tangent calculated phases from their true values for all 16 possible positions of the two enantiomorphs of the molecule. The average absolute deviations of the phases corresponding to the position of the molecule consistent with the original origin and enantiomorph selection was 0.27 radians for the modified tangent formula and 0.98 radians for the simple tangent formula. The average absolute deviation of the simple

tangent phases from the true values for the position at which the enantiomorph fragment appeared was 1.04 radians. If the phases issuing from the tangent program were randomly distributed, these deviations would be 1.57 radians, and it can be seen that the deviations are close to this value for most of the positions of the molecule. It is clear that such a movement of the molecule as was seen on the tangent map is not a phenomenon reproducible in other sets of data, and must be attributed to the fact that the various positions of the molecules have some phases in common. In this case, the tangent formula retained a majority of the phases common to the two positions and altered many of the other phases from the values consistent with the original origin and enantiomorph selections, but it is doubtful that a recognizable molecule would have resulted had the structure been very complex.

Table 7. Modified tangent cycle figures of merit, R cycle (equation 2.19)

Cycle	Number of phases determined	Number of triples used	R cycle
8	35	106	0.79
9	45	178	0.81
10	60	324	0.81
11	80	634	0.82
12	105	1008	0.85
13	135	1649	0.88
14	170	2542	0.90
15	210	3818	0.92
16	250	5745	0.93
17	250	5745	0.93

After the structure was solved, several other phase buildups differing in E minima and rate of new phase pick-up were investigated, and in one of these cases both tangent procedures worked, but no case has yet been found where the simple tangent formula yielded the solution and the modified tangent procedure did not. Consequently, it cannot be claimed that it would not have been possible to solve the 9-t-butyl-9,10-dihydroanthracene structure by means of the simple tangent formula alone had other build-ups been used. However, it has been shown that the initial phases which were assigned in the build-up reported here were all logically and correctly assigned, and in this case time was saved by the immediate application of the modified tangent procedure. The computing time required to generate sets of phases by both the simple and modified tangent procedures simultaneously is approximately 1.75–2.0 times as great as the time required by the simple tangent formula alone.* This time require-

* Times for the modified tangent formula alone are not available, but it is expected that they will be only slightly less than the times required for simultaneous application of the two formulas.

Table 8. Average absolute deviations of the 250 modified and simple tangent phases calculated in the 17th cycle from their true values

Origin position	1st enantiomorph		2nd enantiomorph	
	Modified tangent	Simple tangent	Modified tangent	Simple tangent
0 0 0	0.27 rad	0.98 rad	1.55 rad	1.56 rad
$\frac{1}{2}$ 0 0	1.54	1.56	1.49	1.56
0 $\frac{1}{2}$ 0	1.57	1.58	1.29	1.52
0 0 $\frac{1}{2}$	1.50	1.56	1.19	1.28
$\frac{1}{2}$ $\frac{1}{2}$ 0	1.51	1.51	1.23	1.04*
$\frac{1}{2}$ 0 $\frac{1}{2}$	1.52	1.46	1.52	1.55
0 $\frac{1}{2}$ $\frac{1}{2}$	1.54	1.58	1.54	1.52
$\frac{1}{2}$ $\frac{1}{2}$ $\frac{1}{2}$	1.44	1.36	1.54	1.52

* A fragment of the molecule was found at this position as described in the text.

ment must be weighed against the possible costs of investigating a greater number of initial phase combinations if only the simple tangent formula is used, and it is difficult to guess, based on the limited data available, how many additional combinations will be necessary. Thus, definite statements concerning the relative merits of the tangent procedures must await the results of more extensive applications of the modified tangent procedure, both in this laboratory and elsewhere.

6. Concluding remark

A number of attempts to expand a large set of known phases for very complex molecules, mostly protein structures, by means of the simple tangent formula have been made with indifferent success (Coulter & Dewar, 1971). In view of the comparisons between the two techniques made in this paper, it is suggested that similar attempts employing the modified tangent procedure instead of the older tangent formula may prove to be more successful.

The authors are grateful to Professor M. Sundaralingam who supplied the 9-t-butyl-9,10-dihydroanthracene data and to Dr T. Brennan who helped to solve the structure. One of us (HH) acknowledges helpful discussions with Professor Giovanni Andreotti concerning the procedures employed here.

This research was supported in part by U.S.P.H. Research Grant No. CA 10906.

References

- BRENNAN, T., PUTKEY, E. F. & SUNDARALINGAM, M. (1971). *Chem. Commun.* p. 1490.
 COCHRAN, W. (1955). *Acta Cryst.* **8**, 473.
 COULTER, C. & DEWAR, R. (1971). *Acta Cryst.* **B27**, 1730.
 DUAX, W. (1972). *Acta Cryst.* **B28**, 1864.
 DUAX, W., WEEKS, C. & HAUPTMAN, H. (1972). *Acta Cryst.* **B28**, 1857.
 FISHER, J., HANCOCK, H. & HAUPTMAN, H. (1970). Naval Research Laboratory Report 7132.
 GRANT, D. F., HOWELLS, R. G. & ROGERS, D. (1957). *Acta Cryst.* **10**, 489.
 HAUPTMAN, H. (1966). *Acta Cryst.* **21**, A6.
 HAUPTMAN, H. (1970). *Acta Cryst.* **B26**, 531. *Crystallographic Computing*, pp. 41–51. Edited by F. AHMED, S. R. HALL & C. P. HUBER. Copenhagen: Munksgaard.
 HAUPTMAN, H. (1971). Meeting of the ACA, Ames, Abstract E1.
 HAUPTMAN, H. (1972a). *Z. Kristallogr.* **135**, 1.
 HAUPTMAN, H. (1972b). *Crystal Structure Analysis: The Role of the Cosine Seminvariants*. New York: Plenum Press.
 HAUPTMAN, H. & DUAX, W. L. (1972). *Acta Cryst.* **A28**, 000.
 HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. (1969). *Acta Cryst.* **B25**, 811.
 HAUPTMAN, H., FISHER, J. & WEEKS, C. (1971). *Acta Cryst.* **B27**, 1550.
 KARLE, J. (1968). *Acta Cryst.* **B24**, 182.
 KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.
 KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.
 WEEKS, C., COOPER, A., NORTON, D., HAUPTMAN, H. & FISHER, J. (1971). *Acta Cryst.* **B27**, 1562.