Acta Cryst. (1971). B27, 1550

Phase Determination by Least-Squares Analysis of Structure Invariants: Discussion of This Method as Applied to Two Androstane Derivatives*

BY HERBERT HAUPTMAN† AND JANET FISHER

U.S. Naval Research Laboratory, Washington, D.C. 20390, *U.S.A.*

AND CHARLES M. WEEKS

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

(*Received* 16 *February* 1970 *and in revised form* 29 *September* 1970)

The structures of 5α-androstan-3β-ol-17-one ($C_{19}H_{30}O_2$) and 5β-androstane-3α,17β-diol ($C_{19}H_{32}O_2$) have been solved through the use of the structure invariants $\cos(\varphi_1+\varphi_2+\varphi_3)$. Both these substances crystallize in space group $P2_1$ with two molecules in the unit cell. The computational procedure whereby phases were derived from the invariants is discussed in detail, and a method of calculating invariants which is substantially faster is proposed and shown not to result in any loss of accuracy. The computed invariants were compared with the observed values, and it was found that invariants for which the true value is relatively large are computed more accurately than invariants having smaller values.

Introduction

The most difficult part of solving acentric crystal structures by probability methods lies in the determination of a basic set of phases. Depending on the space group, one to three phases may be arbitrarily specified to select the origin (Hauptman & Karle, 1956; Karle & Hauptman, 1961), and a few additional phases are usually determined by $\sum_1$ relationships (Hauptman & Karle, 1953). After approximately 50 phases have been found, the tangent formula (Karle & Hauptman, 1956) may be employed to evaluate the remaining phases needed to solve the structure.

It is the aim of this paper to show that the structure invariants, $\cos(\varphi_{\mathbf{h}_1}+\varphi_{\mathbf{h}_2}+\varphi_{\mathbf{h}_3})$, may be used to compute additional phases when only a few initial phases are known. These invariants were first used experimentally to solve the structure of the female sex hormone estriol (Hauptman, Fisher, Hancock & Norton, 1969; Cooper, Norton & Hauptman, 1969; Hauptman, 1970). Subsequently this method has been used successfully to solve the structure of 17β-trimethylsiloxy-4-androsten-3-one (Weeks, Hauptman & Norton, to be published) as well as the structures of 5α-androstan-3β-ol-17-one (epiandrosterone, $C_{19}H_{30}O_2$) and 5β-androstane-3α,17β-diol ($C_{19}H_{32}O_2$). These last two structures will be described in the following paper (Weeks, Cooper, Norton, Hauptman & Fisher, 1971). The purposes of this paper are to describe the process of phase determination by least-squares analysis of structure invariants (with reference being made to the phase-determination procedure for epiandrosterone), to point out where difficulties are likely to be encountered, and to compare the merits of various methods of computing the structure invariants based on the results for epiandrosterone and 5β-androstane-3α,17β-diol.

Structure analysis

In space group $P2_1$, the phases of reflections of the type $h0l$, where both h and l are even, are structure invariants whose values depend only on the arrangement of atoms within the unit cell, but the values of all other phases depend on the location of the origin of the unit cell as well as the atomic positions. The location of the origin may be specified by arbitrarily assigning three phases that are linearly independent (Hauptman & Karle, 1956). The linear dependence of phases is a function of space-group symmetry, and for space group $P2_1$, the origin may be uniquely determined by assigning phases to one reflection with $k=1$, and to two reflections $a0b$ and $c0d$ which satisfy the following conditions: (1) a, b not both even, (2) c, d not both even, and (3) the sums $a+c$ and $b+d$ not both even. Phases were assigned to $40\bar{1}$, $50\bar{3}$, and $11\bar{1}$ to specify the origin for epiandrosterone. This set was chosen from among the many possible sets of linearly independent phases because, in each case, both the observed structure-factor amplitude, $|F_{\rm obs}|$, and the normalized structure-factor amplitude, $|E_{\rm obs}|$, were large, and each vector was found to occur in many vector triples $(\mathbf{h}_1,\mathbf{h}_2,\mathbf{h}_3)$* having the property that

$$\mathbf{h}_1+\mathbf{h}_2+\mathbf{h}_3=0\,, \qquad (1)$$

and which also satisfy the condition that $|E_1E_2E_3|$ be large. It is demonstrated below that the probability that the structure invariants, $\cos(\varphi_1+\varphi_2+\varphi_3)$, can be

* Presented at the VIII International Congress of Crystallography, Stony Brook, New York, August 1969.

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

* The abbreviations $\varphi_1=\varphi_{\mathbf{h}_1}$, $\varphi_2=\varphi_{\mathbf{h}_2}$, $\varphi_3=\varphi_{\mathbf{h}_3}$, $E_1=E_{\mathbf{h}_1}$, $E_2=E_{\mathbf{h}_2}$, and $E_3=E_{\mathbf{h}_3}$ will be used throughout the remainder of this communication.

accurately computed increases as the product $|E_1E_2E_3|$ increases. Therefore, reflections with large $|E|$ are introduced into the set of reflections with known phases whenever possible.

In space group $P2_1$, the $\sum_1$ formula (Hauptman & Karle, 1953; Hauptman, 1970)

$$E_{2h,0,2l} \simeq \frac{\sigma_2^{3/2}}{\sigma_3} \langle (-1)^k (|E_{hkl}|^2 - 1) \rangle_k \qquad (2)$$

may be used to obtain, with calculable probability, phases that are structure invariants; σ_n is defined by the relationship

$$\sigma_n = \sum_{j=1}^{N} Z_j^n \qquad (3)$$

where N is the number of atoms in the unit cell, and Z_j is the atomic number of the jth atom. The probability that the normalized structure factor is positive (phase = 0) is given by the relationship (Cochran & Woolfson, 1955)

$$P_+(E_{2h,0,2l}) \simeq \tfrac{1}{2} + \tfrac{1}{2} \tanh \left[\frac{\sigma_3}{\sigma_2^{3/2}} |E_{2h,0,2l}| \sum_k (-1)^k (E_{hkl}^2 - 1) \right] . \qquad (4)$$

The results of the application of the $\sum_1$ formula to those reflections in the epiandrosterone data for which $|E_{obs}|$ was greater than unity are presented in Table 1. A sign was considered to be determined if the probability that it had the indicated value was greater than 95%, and using this criterion, the reflections $20\bar{4}$, $60\bar{4}$, and $40\bar{2}$ were added to the set of vectors with known phases.

Table 1. $\sum_1$ *results for epiandrosterone*

Reflection	$\lvert E_o \rvert$	E_c (by $\sum_1$)	Probability phase positive	True sign
$20\bar{6}$	1·06	0·56	0·58	+
$20\bar{4}$	1·23	−4·54	0·03	−
$60\bar{4}$	1·75	−2·92	0·05	−
$80\bar{4}$	1·42	0·79	0·64	+
$40\bar{2}$	1·83	2·81	0·95	+
$80\bar{2}$	1·05	0·76	0·60	+

The phases of reflections whose indices satisfy equation (1) are linearly dependent, and it is possible to relate all reflections to the origin-determining-reflections and those determined by $\sum_1$ by means of this equation. This use of formulas relating the phases of such linearly dependent reflections allows the set of known phases to be expanded to such an extent that calculation of a Fourier synthesis may reveal the crystal structure. The simplest of the mathematical formulas of this type is the $\sum_2$ type relationship

$$E_{-\mathbf{h}} = E^*_{\mathbf{h}} \simeq \frac{\sigma_2^{3/2}}{\sigma_3} \langle E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}} \rangle_{\mathbf{k}} \qquad (5)$$

(Sayre, 1952; Hughes, 1953) which is valid for all space groups. In this equation $\mathbf{h}$ is a fixed vector, and $\mathbf{k}$ ranges over all vectors for which $-\mathbf{h}-\mathbf{k}$ exists. It is obvious that the indices of the triple $(\mathbf{h}, \mathbf{k}, -\mathbf{h}-\mathbf{k})$ sum to zero, and consequently the equivalence

$$\mathbf{h} = \mathbf{h}_1, \ \mathbf{k} = \mathbf{h}_2, \ -\mathbf{h}-\mathbf{k} = \mathbf{h}_3 \qquad (6)$$

exists for individual terms contributing to the average in equation (5). In the case of noncentrosymmetric space groups, other phase relations which have found considerable use are

$$\varphi_{\mathbf{h}} \simeq -\langle (\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \rangle_{\mathbf{k}} \qquad (7)$$

(Cochran & Woolfson, 1955) and the tangent formula,

$$\tan \varphi_{\mathbf{h}} = \frac{-\sum_k |E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| \sin (\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})}{\sum_k |E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| \cos (\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}})} \qquad (8)$$

(Karle & Hauptman, 1956). The tangent formula is an extremely powerful tool. It cannot be used effectively, however, until several triples exist involving vector $\mathbf{h}$ in conjunction with pairs $(\mathbf{k}, -\mathbf{h}-\mathbf{k})$ for which $\varphi_{\mathbf{k}}$ and $\varphi_{-\mathbf{h}-\mathbf{k}}$ are known.

The difficulty with using any of the equations (5), (7), and (8) when only a few phases are known is that they all involve summations, but only a very few terms in these summations can be computed, and the calculated $\varphi_{\mathbf{h}}$ may be in error as a result. The probability that a single term will accurately yield the unknown phase increases as the magnitude of $|E_1E_2E_3|$ increases. Equation (7) has found use in connection with the symbolic addition procedure (Karle & Karle, 1966), in which symbols are assigned to a few phases and equation (7) applied to triples having two known phases and a large value of $|E_1E_2E_3|$. The third phase is then either known or else it can be related to one of the assigned symbols, and it is placed in the set of known phases. Considerable care must be exercised in the use of equation (7) when more than one triple contributes to a new phase determination because of the ambiguity arising from the multiple-valued nature of the phases. It is possible to replace φ with $(\varphi + 2\pi k)$ where k is any integer, and there is nothing in equation (7) to indicate which of these is the proper choice. One method of circumventing this problem is to introduce arbitrary values for the phases of a few (usually one to four) reflections having large $|E|$, and to vary the values of phases systematically and use each set of phases so obtained as input for the tangent formula. Often one or two of the sets of phases output by the tangent formula can be selected as more likely to be correct based on such criteria as the α index (Karle & Karle, 1966),

$$\alpha_{\mathbf{h}} = \frac{2\sigma_3}{\sigma_2^{3/2}} |E_{\mathbf{h}}| \left[\left\{ \sum_k |E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| \sin (\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \right\}^2 + \left\{ \sum_k |E_{\mathbf{k}} E_{-\mathbf{h}-\mathbf{k}}| \cos (\varphi_{\mathbf{k}} + \varphi_{-\mathbf{h}-\mathbf{k}}) \right\}^2 \right]^{1/2} \qquad (9)$$

which is relatively large for the correct solution. However, it may be necessary to compute Fourier syntheses from several sets of phases before the structure is sol-

ved. The obvious drawback to this approach is that even if only two arbitrary phases have been introduced, it may be necessary to assign several values to each before the correct solution is found. Even if only 4 values are assigned to each of 2 arbitrary phases, it may still be necessary to compute 16 Fourier syntheses, and in unfavorable cases, a considerably larger number may be required.

The need for a procedure to eliminate the introduction of deliberate ambiguities during the early stages of phase determination has led to a renewed investigation of the structure invariants $\cos(\varphi_1+\varphi_2+\varphi_3)$. A formula for computing these invariants which requires knowledge only of the normalized structure-factor *amplitudes* was first proposed in 1957 [Karle & Hauptman (1957) equation (2.2)]. Subsequent theoretical considerations have shown that this formula is not valid if the structure contains a substantial amount of overlap among Patterson peaks (Hauptman, 1964). Equation (10),

$$\cos(\varphi_1+\varphi_2+\varphi_3)\simeq \frac{K\psi}{|E_1E_2E_3|}+\frac{R_3}{|E_1E_2E_3|}, \qquad (10)$$

where

$$\psi=\langle(|E_1|^p-\overline{|E|^p})(|E_{\mathbf{h}_1+\mathbf{l}}|^p - \overline{|E|^p})(|E_{-\mathbf{h}_3+\mathbf{l}}|^p-\overline{|E|^p})\rangle_{\mathbf{l}} \qquad (11)$$

and

$$\overline{|E|^p}=\langle|E_1|^p\rangle_{\mathbf{l}}, \qquad (12)$$

is a variant of this formula and, in the form where $\mathbf{l}$ ranges over all vectors in reciprocal space and the exponent p is equal to $\frac{1}{2}$, was first used successfully to solve the structure of estriol (Hauptman, Fisher, Hancock & Norton, 1969). R_3 is a term which depends only on the normalized structure factor amplitudes $|E_1|$, $|E_2|$, and $|E_3|$. When $p=\frac{1}{2}$, it takes the form

$$R_3=\frac{\sigma_3}{4\sigma_2^{3/2}}\left[\tfrac{3}{2}(|E_1E_2|^2+|E_2E_3|^2+|E_3E_1|^2) + |E_1|^2+|E_2|^2+|E_3|^2-\tfrac{7}{2}\right], \qquad (13)$$

and, when $p=2$,

$$R_3=\frac{\sigma_3}{\sigma_2^{3/2}}(|E_1|^2+|E_2|^2+|E_3|^2-2). \qquad (14)$$

K is a scale factor which is expected to be a function of A (Hauptman, Fisher, Hancock & Norton, 1969) where

$$A=\frac{2\sigma_3}{\sigma_2^{3/2}}|E_1E_2E_3|. \qquad (15)$$

The distribution of the structure invariants $\cos(\varphi_1+\varphi_2+\varphi_3)$ is a function of A only, and the scale factors are chosen to make the distribution of the calculated invariants agree as closely as possible with the theoretical distribution. The theoretical distributions of the invariants for several values of A are illustrated in Fig. 1. The most striking feature about these distributions is that, for large A, most invariants are positive, and the proportion of invariants whose value is approximately unity is large.

If two or more structure invariants, $\cos(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})$, involving a common vector $\mathbf{h}$ have been computed, they will probably not yield exactly the same value of $\varphi_{\mathbf{h}}$ for two reasons. First, there are always some experimental errors in the measurement of the normalized structure-factor amplitudes, and, in addition, equation (10) is known not to be exactly valid for real structures having overlapping Patterson peaks. Consequently, the best value of $\varphi_{\mathbf{h}}$ may be found by minimizing the function (Hauptman, Fisher, Hancock & Norton, 1969)

$$\Phi=\frac{\sum_{\mathbf{k}} w_{\mathbf{k}}[\cos(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}})-c_{\mathbf{k}}]^2}{\sum_{\mathbf{k}} w_{\mathbf{k}}} \qquad (16)$$

where several structure invariants,

$$c_{\mathbf{k}}=\cos(\varphi_{\mathbf{h}}+\varphi_{\mathbf{k}}+\varphi_{-\mathbf{h}-\mathbf{k}}), \qquad (17)$$

involving a given vector $\mathbf{h}$, have been determined by means of equation (10), and each invariant has been assigned a weight

$$w_{\mathbf{k}}=|E_1E_2E_3|/n \qquad (18)$$

where n is the number of contributors to the average in equation (11). The minima in the function Φ may be

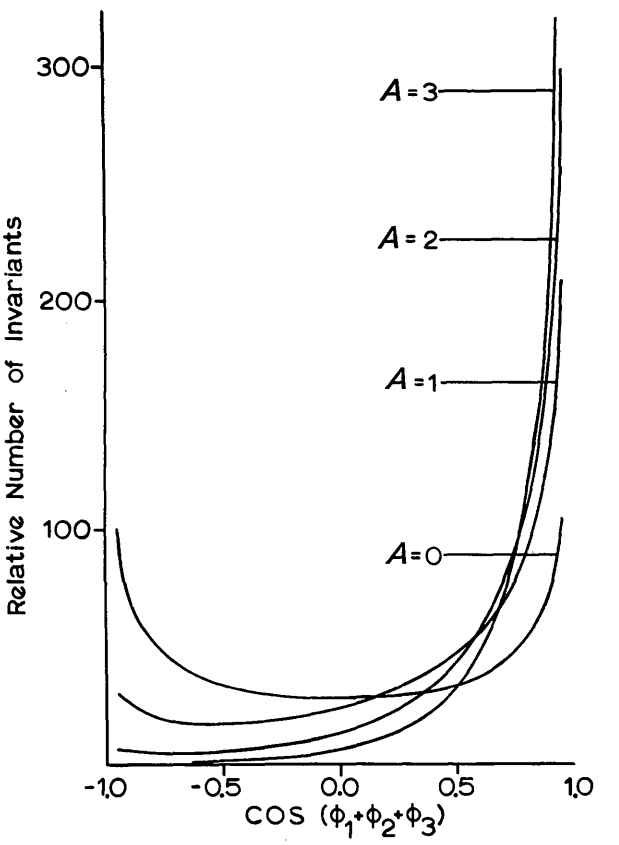


Fig. 1. The theoretical distribution of the structure invariants $\cos(\varphi_1+\varphi_2+\varphi_3)$ as a function of A. The units on the ordinate are arbitrary.

located most readily by varying $\varphi_{\mathbf{h}}$ from 0 to 2π in small increments (*e.g.* 0·01 radians) and evaluating the function at each point. The first time an invariant is used which differs significantly from ± 1, there will be two equal minima in the function Φ, and either of the corresponding values of $\varphi_{\mathbf{h}}$ may be chosen as the correct phase since the choice of one of these minima results in the selection of one of the enantiomorphs. If the structure invariants are internally consistent, and if the enantiomorph determining phase is involved directly or indirectly in all subsequent invariants whose values are not ± 1, only one minimum will be in the function Φ for each additional phase. In practice, however, because the invariants are not computed with great accuracy and are not perfectly consistent with each other, there will be many cases where there are two minima. If these minima are very nearly equal, it is necessary to carry both phases through the remainder of the computations terminating in the calculation of two Fourier syntheses unless one of the phases can, at some intermediate stage, be shown to be less likely to be correct. Thus, even with the use of the structure invariants, it may not be possible to eliminate all the phase ambiguities. The α index [equation (9)], which is one of the quantities used to indicate the reliability of the phasing in the tangent formula, was discussed above. Similar criteria have been used in an attempt to judge the accuracy of phases determined from the structure invariants through equation (16) (Hauptman, 1970). One such quantity, the individual residual for vector **h**,

$$R_{\mathbf{h}} = \Phi_{min}^{1/2} \tag{19}$$

is simply the square root of a weighted minimum of the function Φ. The cycle residual,

$$R_{\text{cycle}} = \left[\frac{\sum_{\mathbf{h}} R_{\mathbf{h}}^2 \left(\sum_{\mathbf{k}} A_{\mathbf{k}} \right)}{\sum_{\mathbf{h}} \left(\sum_{\mathbf{k}} A_{\mathbf{k}} \right)} \right]^{1/2}, \tag{20}$$

is a weighted average of the individual residuals for all vectors **h** for which phases were calculated during the cycle. The centrosymmetric residual,

$$R_{\text{centro}} = \left[\frac{\sum_{\mathbf{h}} d_{\mathbf{h}}^2 \left(\sum_{\mathbf{k}} A_{\mathbf{k}} \right)}{\sum_{\mathbf{h}} \left(\sum_{\mathbf{k}} A_{\mathbf{k}} \right)} \right]^{1/2} \tag{21}$$

(where $d_{\mathbf{h}}$ is the deviation of the phase of a purely real or purely imaginary structure factor from the nearer of its two allowed values, or in the case of phases determined by fixation of the origin or through use of the $\sum_1$ formula, the deviation from the known value), is also a measure of the accuracy of phases computed during a single cycle. These residuals will presumably be small when the use of the structure invariants leads to correct phases.

In the case of epiandrosterone, the phases of six independent spectra and the phases of their eight symmetry-related reflections were initially known. Using this set of reflections as vectors **k** and $-\mathbf{h}-\mathbf{k}$ all possible vectors **h** were generated, and for each of these vectors, the summation, $\sum_{\mathbf{k}} A_{\mathbf{k}}$, over all triples (**h**, **k**, $-\mathbf{h}-\mathbf{k}$) was constructed. The reflection $51\bar{3}$, for which this summation was the largest ($\sum A = 5{\cdot}98$), was selected as the first reflection whose phase was to be determined from the structure invariants by minimizing the function Φ defined in equation (16). This reflection and its three symmetry-related reflections were added to the set of vectors with known phases, additional triples were generated, summations were incremented, and the 3 vectors ($11\bar{2}$, $51\bar{2}$, and $80\bar{3}$) with the largest summations were added to the set with known phases. This process was continued for 21 additional cycles in which 5, 7, 9, 11, ... vectors were successively selected until the order in which the phases of all 533 reflections with $|E| > 1$ were to be found, was specified. Twelve cycles of this procedure were sufficient to select 150 reflections among which there were 1971 independent vector triples whose indices were related by equation (1). For each of these triples, the average in equation (11) was computed using an exponent $p = \frac{1}{2}$ and allowing **l** to vary over all measurable reflections, and the corresponding R_3 terms were computed by means of equation (13).

To find the scaling parameter K, the triples with their associated averages and R_3 terms were sorted on increasing value of A and divided into 17 groups with 120 triples in each of the first 15 groups and 86 and 85 triples in the last 2 groups; the average A was found for each group. Within each group, the triples were sorted so that the terms $\psi/|E_1E_2E_3|$ [equation (11)] were in decreasing order, and inspection of equation (10) shows that this amounts to sorting according to the value of the invariant since A and R_3 are approximately constant within a group.

Some intermediate calculations which were performed to find the value of K for the second group of triples (which had $\langle A \rangle = 0{\cdot}53$ and $R_3/|E_1E_2E_3| \simeq 0{\cdot}23$) are presented in Table 2. An extensive table of the conditional probability that $\cos(\varphi_1+\varphi_2+\varphi_3)$ is greater than X for several values of A and X has been published (Hauptman, 1970, Table V), and the probability that a given invariant with $A \simeq 0{\cdot}53$ is greater than $X = 0{\cdot}071$, 0·170, 0·267, ..., 0·995 was read from this table. The value of $\psi/|E_1E_2E_3|$ for that invariant which should be closest to X in value was then found. For example, the probability that $\cos(\varphi_1+\varphi_2+\varphi_3) > 0{\cdot}622$ is 42% if $A = 0{\cdot}53$, and 42% of the invariants in the experimental group will be greater than 0·622 if the invariant for the 50th term in the sorted list, which had $\psi/|E_1E_2E_3| = 0{\cdot}000371$ and $R_3/|E_1E_2E_3| = 0{\cdot}217$, is set equal to 0·622. This will be the case if $K = 1092$.*

* From equation (10),

$$K = \frac{0{\cdot}622 - R_3/|E_1E_2E_3|}{\psi/|E_1E_2E_3|}$$

and

$$K = (0{\cdot}622 - 0{\cdot}217)/0.000371 = 1092\,.$$

Table 2. *K values for the group of* 120 *epiandrosterone triples having an average* $A = 0{\cdot}53$

The term $R_3/|E_1E_2E_3|$ is approximately constant for these triples and is in the range 0·215–0·255.

X	Probability $\cos(\varphi_1+\varphi_2+\varphi_3) > X$	$\frac{\psi}{\lvert E_1E_2E_3\rvert} \times 10^4$	K
0·071	0·64	1·21	−1277
0·170	0·61	1·56	−386
0·267	0·57	1·98	233
0·362	0·54	2·45	535
0·454	0·50	3·07	742
0·540	0·46	3·34	864
0·622	0·42	3·71	1092
0·698	0·38	3·94	1142
0·765	0·34	4·12	1309
0·825	0·29	4·61	1290
0·878	0·25	5·30	1229
0·921	0·20	6·18	1110
0·955	0·15	7·44	946
0·980	0·10	8·27	884
0·995	0·05	9·53	801

Values of K were also computed in analogous fashion for the other values of X, and they are listed in Table 2. If this sample of averages were large and random, and if the accuracy of averages computed by equation (11) were independent of the true value of the invariant, all the values for K so obtained should be nearly equal, but it is observed that this is not the case. K cannot be negative because, if this were true, invariants for which the averages were smallest (negative) would give the largest values of $\cos(\varphi_1+\varphi_2+\varphi_3)$. The instability seen in the values of K results from division by a number which is very close to zero (*i.e.* $\psi/|E_1E_2E_3|$), and if, because of experimental error, a term with a positive average is used when a negative one should be, K will have the wrong sign. Consequently, it is necessary to find that region in which the calculated values of K are positive and reasonably constant, and to find the average K in this region. In the case of these data, K was found to be reasonably stable for all groups of triples in the range $X = 0{\cdot}622$ to 0·955, regardless of the average value of A for that group. The values of K for the various ranges of A are plotted in Fig. 2 to show the dependence of K on A, and it was found from a least-squares analysis that K equals $1015 + 267\,A$ for the epiandrosterone data.

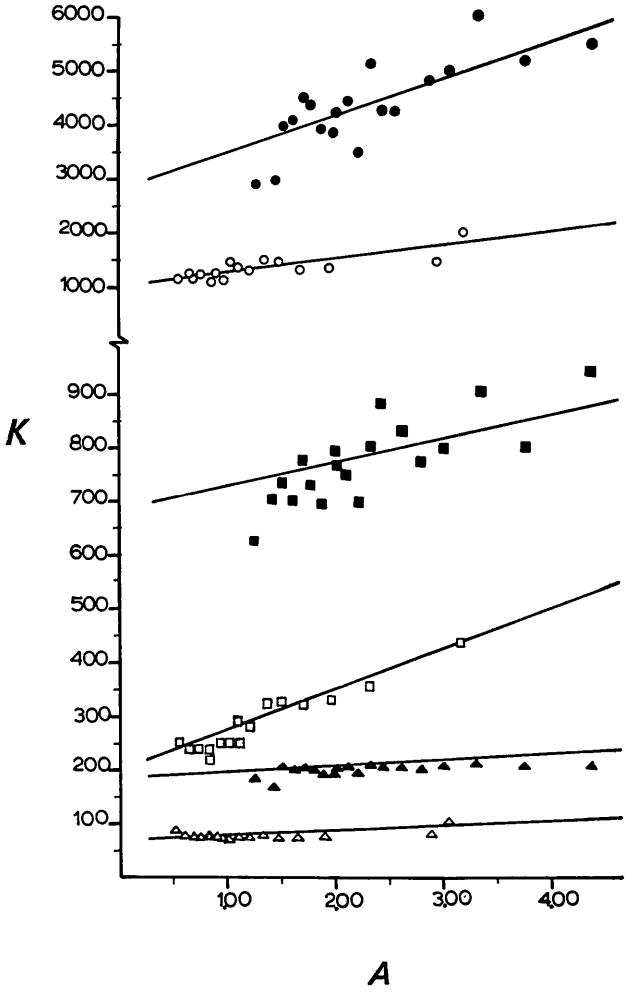
The 1971 structure invariants, $\cos(\varphi_1+\varphi_2+\varphi_3)$, involving the first 150 reflections were then calculated,

Table 3. *Some structure invariants used during the least-squares calculations for epiandrosterone*

	$\mathbf{h}_1$	$\mathbf{h}_2$	$\mathbf{h}_3$	A	Predicted $\cos(\varphi_1+\varphi_2+\varphi_3)$*	Observed $\cos(\varphi_1+\varphi_2+\varphi_3)$
Cycle 1	1 1 $\bar{1}$	$\bar{6}$ 0 4	5 $\bar{1}$ $\bar{3}$	2·92	0·46	1·00
	1 1 $\bar{1}$	4 0 $\bar{2}$	$\bar{5}$ $\bar{1}$ 3	3·06	1·09	1·00
	4 0 $\bar{1}$	$\bar{5}$ $\bar{1}$ 3	1 1 $\bar{2}$	4·45	0·81	0·50
Cycle 2	4 0 $\bar{1}$	1 1 $\bar{1}$	$\bar{5}$ $\bar{1}$ 2	3·64	0·92	0·99
	4 0 $\bar{1}$	4 0 $\bar{2}$	$\bar{8}$ 0 3	3·76	1·32	1·0
	6 0 $\bar{4}$	$\bar{1}$ $\bar{1}$ 2	$\bar{5}$ 1 2	2·44	0·39	0·41
	4 0 $\bar{1}$	$\bar{4}$ 0 2	0 0 1	3·50	1·32	1·0
	5 1 $\bar{3}$	$\bar{5}$ $\bar{1}$ 2	0 0 1	3·60	0·97	1·0
	1 1 $\bar{1}$	$\bar{1}$ $\bar{1}$ 2	0 0 1	3·70	0·81	0·56
	4 0 $\bar{1}$	$\bar{5}$ 0 3	1 0 $\bar{2}$	3·51	0·95	1·0
	4 0 $\bar{1}$	$\bar{6}$ 0 4	2 0 $\bar{3}$	1·73	0·48	1·0
Cycle 3	1 1 $\bar{1}$	1 $\bar{1}$ $\bar{2}$	$\bar{2}$ 0 3	1·92	0·24	0·56
	4 0 $\bar{2}$	$\bar{1}$ $\bar{1}$ 2	$\bar{3}$ 1 0	1·40	1·01	0·33
	4 0 $\bar{1}$	$\bar{1}$ $\bar{1}$ 1	$\bar{3}$ 1 0	2·00	0·98	0·97
	5 1 $\bar{3}$	$\bar{8}$ 0 3	3 $\bar{1}$ 0	2·12	1·07	0·98
	1 1 $\bar{2}$	5 1 $\bar{2}$	$\bar{6}$ $\bar{2}$ 4	2·78	1·08	0·71
	1 1 $\bar{1}$	5 1 $\bar{3}$	$\bar{6}$ $\bar{2}$ 4	3·33	1·21	0·99
	2 0 $\bar{4}$	$\bar{1}$ $\bar{1}$ 0	$\bar{1}$ 1 4	0·40	−0·03	0·90
	2 0 $\bar{4}$	$\bar{1}$ $\bar{2}$ 0	$\bar{1}$ 2 4	0·43	−0·23	−0·32
	2 0 $\bar{4}$	$\bar{6}$ $\bar{1}$ 5	4 1 $\bar{1}$	0·64	−0·19	0·24
	2 0 $\bar{4}$	3 2 2	$\bar{5}$ $\bar{2}$ 2	0·71	0·18	0·99
	2 0 $\bar{4}$	$\bar{8}$ 1 2	6 $\bar{1}$ 2	0·72	0·37	0·76
	2 0 $\bar{4}$	$\bar{3}$ $\bar{2}$ $\bar{1}$	1 2 5	0·73	0·19	0·89
	2 0 $\bar{4}$	1 1 $\bar{3}$	$\bar{3}$ $\bar{1}$ 7	0·79	0·09	0·23
Invariants involving $20\bar{4}$	2 0 $\bar{4}$	1 0 $\bar{2}$	$\bar{3}$ 0 6	0·79	0·47	1·00
	2 0 $\bar{4}$	$\bar{1}$ $\bar{1}$ 3	$\bar{1}$ 1 1	0·84	−0·07	0·38
	2 0 $\bar{4}$	$\bar{4}$ $\bar{1}$ 0	2 1 4	0·88	0·19	0·98
	2 0 $\bar{4}$	$\bar{5}$ $\bar{1}$ 3	3 1 1	0·96	−0·18	0·45
	2 0 $\bar{4}$	$\bar{1}$ $\bar{5}$ $\bar{1}$	$\bar{1}$ 5 5	0·96	0·38	0·67
	2 0 $\bar{4}$	$\bar{6}$ $\bar{6}$ 1	4 6 3	0·97	0·12	0·94
	2 0 $\bar{4}$	$\bar{4}$ 0 1	2 0 3	1·00	0·37	1·00
	2 0 $\bar{4}$	$\bar{2}$ 0 3	0 0 1	1·10	−0·06	1·00

* The exponent p [equation (10)] was $\frac{1}{2}$, and **l** ranged over all measurable reflections ($|E_l| > 0$). The scaling parameter used was $K = 1015 + 267\,A$.

and the phases of these reflections were found by 12 cycles of the least-squares procedure outlined above. The invariants used during the first three cycles are listed in Table 3. Several of these invariants were calculated to be greater than 1 and a few less than -1, but before calculation of the function Φ [equation (16)], such invariants were set equal to $+1$ or -1, respectively. In noncentrosymmetric space groups, the enantiomorph is selected the first time a phase is assigned that has substantially different values for the two enantiomorphs. This requires the use of a structure invariant whose value is different from ± 1. The first calculated invariant to be used for epiandrosterone involved the reflections $11\bar{1}$, $60\bar{4}$, and $51\bar{3}$; since the computed value, 0·46, was significantly different from ± 1, the function Φ had two equal minima in the first cycle, and the value 0·73 was arbitrarily chosen for $\varphi_{51\bar{3}}$ to specify the enantiomorph. The results of the least-squares process for the first three cycles are presented in Table 4. In the second cycle, two additional reflections ($11\bar{2}$ and $51\bar{2}$) were encounted each of which had two equal minima in the least-squares function. These phases are ambiguous because the last arbitrary choice of phase was made in specifying the enantiomorph, and these ambiguities arise because there is only one invariant contributing to each phase determination. If a calculated phase is not 0 or π and there is only one contributor, there will always be two solutions because, for all angles except 0 and π, there is a second angle having the same value of the cosine. A second contributor is needed to resolve such ambiguities, and, unfortunately, it is not always possible to avoid them at the beginning of phase determination since there may not be an alternate path for building up a set of phases.



Fig·2. The dependence of K on A. Epiandrosterone, $|E_l|$ greater than: 0·0, ○; 1·0, □; 2·0, △. 5β-androstane-3α,17β-diol, $|E_l|$ greater than: 0·0, ●; 1·0, ■; 2·0, ▲.

In the third cycle, other double minima occur but there are no additional ambiguities of the type encountered with the reflections $11\bar{2}$ and $51\bar{2}$. The two minima for $20\bar{3}$ do not create an ambiguity because the second residual is much larger than the first. While the two residuals for $10\bar{2}$ are equal, this phase is still determined because it is known to be a centrosymmetric phase with a value true of either 0 or π, and both of the calculated minima lie close to one of these values. In no instance were more than two minima observed.

A disturbing feature in the results of the least-squares phase determination procedure, is that the calculated phase for $20\bar{4}$ was 0 (see Table 4) whereas the $\sum_1$ formula had strongly determined this phase to be π. Input phases that have been determined by origin specification or through use of the $\sum_1$ formula are normally used at their original value in all phase-determination cycles regardless of the calculated value in the previous cycle. Consequently, the least-squares procedure was repeated with the alternative phase for $20\bar{4}$; little difference was observed in the two sets of cycle residuals (Table 5) and $\varphi_{20\bar{4}}$ was still calculated to be 0. Three additional runs, in which $\varphi_{20\bar{4}}=0$, and which differed from each other in the initial minimum selected for $11\bar{2}$ and $51\bar{2}$ were then performed. Again, the cycle and centrosymmetric residuals were all about equally good, and it was also impossible to select the set of phases most likely to be correct based on the criterion that there be little fluctuation of calculated values of a single phase in successive cycles. Because of the ambiguities with the reflections $20\bar{4}$, $11\bar{2}$, and $51\bar{2}$, it was necessary to perform 8 runs in the tangent formula. The set of 150 phases resulting from the first least-squares run in which $\varphi_{20\bar{4}}$ equaled 0 was arbitrarily chosen, and 11 cycles in which no phases were forced to the input values, were performed using the tangent formula to determine phases for all 533 independent X-ray spectra with $|E|>1$. All phases were then refined for five additional cycles. The phase of the $20\bar{4}$ reflection fluctuated from 0 to π during the tangent formula cycles, and in the final refinement cycles, it was consistently calculated to be π.

A Fourier map was prepared using all 533 phases resulting from the final tangent formula refinement

cycle, and peaks were found that appeared to fit the expected model of epiandrosterone. Positive electron density was found at the positions of all the nonhydrogen atoms, but the axial carbons, C(18) and C(19), as well as the two oxygen atoms, O(3) and O(17), were not so well resolved as the atoms in the steroid nucleus. Four cycles of least-squares refinement of the positional and isotropic thermal parameters of the 21

Table 4. *Some results of the least-squares phase determination cycles for epiandrosterone*

Cycle	h	Number of invariants contributing to phase determination	1st (lowest) minimum φ_h (rad)	1st (lowest) minimum R_h	2nd minimum φ_h (rad)	2nd minimum R_h
	$20\bar{4}$		3·14			
	$60\bar{4}$		3·14			
Input	$50\bar{3}$		0·00			
(origin and $\sum_1$)	$40\bar{2}$		0·00			
	$40\bar{1}$		0·00			
	$11\bar{1}$		0·00			
1	$51\bar{3}$*	2	0·73	0·27	−0·73	0·27
	$60\bar{4}$	1	3·14	0·29		
	$80\bar{3}$*	1	0·00	0·00		
	$51\bar{3}$	2	0·73	0·27	0·00	0·27
2	$40\bar{2}$	1	0·00	0·25		
	$11\bar{2}$*	1	1·35	0·00	0·11	0·00
	$51\bar{2}$*	1	0·41	0·00	−0·41	0·00
	$11\bar{1}$	2	0·73	0·27	0·00	0·27
	$60\bar{4}$	2	3·14	0·25		
	$62\bar{4}$*	2	1·23	0·13		
	$20\bar{3}$*	2	3·14	0·27	0·00	0·86
	$80\bar{3}$	1	0·00	0·00		
	$51\bar{3}$	3	0·73	0·20		
	$10\bar{2}$*	1	0·32	0·00	−0·32	0·00
3	$40\bar{2}$	2	0·00	0·21		
	$11\bar{2}$	2	1·40	0·09	0·61	0·30
	$51\bar{2}$	2	0·31	0·05		
	$40\bar{1}$	5	0·00	0·00		
	$11\bar{1}$	3	0·00	0·21	0·56	0·27
	310*	3	−2·50	0·17		
	001*	3	0·00	0·40		
4	$20\bar{4}$	1	−2·32	0·00	2·32	0·00
5	$20\bar{4}$	2	0·00	0·63		
6	$20\bar{4}$	2	0·00	0·63		
7	$20\bar{4}$	3	0·00	0·59	−3·10	0·70
8	$20\bar{4}$	4	0·00	0·46	3·14	0·61
9	$20\bar{4}$	5	0·00	0·45	3·14	0·67
10	$20\bar{4}$	5	0·00	0·51	−3·12	0·74
11	$20\bar{4}$	10	0·00	0·68	−3·10	0·76
12	$20\bar{4}$	12	0·00	0·67		

* This phase was first determined in this cycle.

Table 5. *Average residuals for the least-squares phase determination cycles for epiandrosterone*

	Cycle residual			Centrosymmetric residual		
Cycle	No. reflection	$\varphi_{20\bar{4}}=\pi$	$\varphi_{20\bar{4}}=0$	No. reflection	$\varphi_{20\bar{4}}=\pi$	$\varphi_{20\bar{4}}=0$
1	1	0·27	0·27	0	—	—
2	7	0·22	0·22	3	0·00	0·00
3	13	0·21	0·22	6	0·07	0·07
4	22	0·29	0·29	10	0·34	0·30
5	31	0·30	0·30	12	0·46	0·50
6	42	0·34	0·34	12	0·42	0·36
7	55	0·36	0·37	14	0·49	0·44
8	70	0·35	0·36	16	0·45	0·43
9	87	0·37	0·42	16	0·42	0·43
10	106	0·41	0·42	18	0·37	0·33
11	127	0·40	0·39	22	0·43	0·41
12	150	0·37	0·37	23	0·33	0·37

atoms were performed using all observed reflections and a block-diagonal approximation to the normal equations. The R indices in successive cycles were 47·8, 38·3, 36·9, and 36·7%, and the phase of the $20\bar{4}$ reflection, as calculated using the original atomic positions, was equal to 0. Since this refinement appeared to converge after two or three cycles, and since there were no individual isotropic temperature factors which became either very large or very small and the overall geometry of the steroid molecule appeared to be quite satisfactory, it was hypothesized that the entire molecule had been translated into a false minimum as the result of an incorrect phase assignment during the early stages of the phase buildup.

Examination of the unit-cell dimensions, and of the Patterson synthesis, showed that the molecule was located in the unit cell with its long axis almost parallel to the twofold screw axis. One conspicuously large peak, which indicated the position of the center of mass of the molecule, was observed on the Harker section, and it was found that this position was related to the location of the molecule on the Fourier map by a translation perpendicular to the $20\bar{4}$ planes over a distance approximately equal to one and one half times the spacing of these planes. The molecule was translated so that the position of its center of mass coincided with the location of this large peak, and following refinement of the atomic positional and anisotropic thermal parameters the R index fell to a final value of 6·4%. The final phase of the reflection $20\bar{4}$ was π.

To try to understand the behavior of the $20\bar{4}$ reflection, the true invariants, $\cos(\varphi_1+\varphi_2+\varphi_3)$, were computed using the phases for the refined structure and were compared to the predicted invariants. The 15 invariants involving the $20\bar{4}$ reflection which were used in the least-squares phase determining procedure are listed in Table 3. In all but one case, the value of the predicted invariant is less than the observed value for the refined structure. This distinctly nonrandom deviation of the predicted invariants from their true values is not the normal situation as can be seen from Table 6 where the root-mean-square deviation and average deviation of all predicted invariants involving a common vector are listed for several centrosymmetric reflections. Although there are other cases (201 and 603) where the r.m.s. deviation is relatively large, there is no other case where the average deviation is as large as it is in the case of $20\bar{4}$, and a large average deviation is indicative of nonrandom error. The fact that a large nonrandom error occurred in the computation of the invariants involving the $20\bar{4}$ reflection is, of course, consistent with the facts that its phase was wrongly indicated in the least-squares phase calculations, and that it had a comparatively high residual (see Table 4). The 201 reflection is unusual in that it is involved in several invariants whose observed values are small, and the observed errors for such invariants are discussed in greater detail below.

Table 6. *R.m.s. and average deviations of epiandrosterone invariants involving a common reflection*

Reflection	Number of invariants	R.m.s. deviation	Average deviation
$20\bar{6}$	21	0·48	0·19
$30\bar{6}$	22	0·50	0·21
$20\bar{4}$	15	0·64	−0·56
$60\bar{4}$	26	0·55	−0·19
$20\bar{3}$	32	0·53	−0·19
$50\bar{3}$	43	0·32	0·08
$80\bar{3}$	22	0·38	0·16
$10\bar{2}$	45	0·30	−0·20
$40\bar{2}$	44	0·37	0·17
$40\bar{1}$	49	0·34	0·03
100	30	0·42	−0·04
001	48	0·41	0·04
201	26	0·71	0·00
203	23	0·44	−0·02
603	20	0·60	0·31
105	35	0·37	−0·01
205	24	0·37	0·09

* The deviation is the difference between the predicted and the observed values of $\cos(\varphi_1+\varphi_2+\varphi_3)$.

Examination of the observed invariants presented in Table 3 also shows that the invariant involving $11\bar{1}$, $60\bar{4}$, and $51\bar{3}$, which was predicted to be significantly different from ± 1 and which was presumed to cause selection of the enantiomorph, has an observed value of $+1$. Consequently, the enantiomorph was not really specified until $\varphi_{11\bar{2}}$ was selected in the second cycle. This example points out the critical nature of the invariant(s) which actually result in enantiomorph selection. It would be desirable to try several different ways of selecting the origin until a case could be found where there are at least two collaborating invariants not equal to ± 1 that contribute to the phase assignment which specifies the enantiomorph. In the present situation, enantiomorph selection actually took place at the later, second cycle, not the first. As it turned out, one and only one phase had substantially different values for

Table 7. *Comparison of four E maps resulting from different initial choice of phase for* $11\bar{2}$ *and* $51\bar{2}$

Map	Phase(s) forced to second minimum in cycles 2 and 3	Number of real atoms among top 30 peaks	Rank of largest spurious peak	R.m.s. deviation*
1	—	16	1	0·29 Å
2	$11\bar{2}$	20	19	0·13
3	$51\bar{2}$	17	7	0·27
4	$11\bar{2}$, $51\bar{2}$	16	6	0·39

* R.m.s. deviation (in Å) of original map positions from refined atomic positions.

the two enantiomorphs in the second cycle and so the enantiomorph was unambiguously selected, unwittingly, in this cycle. However, even if, unsuspected by the investigator, enantiomorph selection were to occur during a cycle in which two or more phases, one of which alone would suffice for enantiomorph selection, were simultaneously determined, no damage would be done. In such a situation all combinations would yield the same least-squares residuals, one combination would correspond in a consistent way to one choice of enantiomorph, another to the other choice, and all possibilities would normally be considered.

To determine if the epiandrosterone molecule could be found at the correct position on a Fourier map resulting from phases determined by the least-squares analysis of the structure invariants without making any assumptions concerning molecular packing, four additional syntheses were computed. In each case, $\varphi_{20\bar{4}}$ was taken to be π, and the phasing differed on the basis of the initial phase used for $11\bar{2}$ and $51\bar{2}$, two of the reflections for which two equal least-squares minima had been calculated. Also in each case, 150 phases were determined by the least-squares analysis, but only 8 cycles (two of which were refinement cycles) were performed with the tangent formula, and a total of only 330 phases were determined, which is approximately 15 phases per nonhydrogen atom in the asymmetric unit. The epiandrosterone molecule was found in the correct position on each of the resultant four maps although the definition of the molecule was considerably better in one case (map 2) as shown in Table 7 by the values of the average r.m.s. deviation of the atoms from their refined positions, the number of real atoms among the highest thirty peaks on the map, and the rank of the largest spurious peak after the peaks had been sorted according to their relative heights. As expected, choice of the second least-squares phase for $11\bar{2}$ resulted in selection of the second enantiomorph. The large individual deviations on map 4 make it doubtful whether the molecule could have been recognized had the correct position not been known, but nevertheless, it was possible to refine this molecule to the correct solution. Since the deviations for the map of the computed synthesis using the second

Table 8. *R.m.s. deviations of the predicted structure invariants, cos* $(\varphi_1+\varphi_2+\varphi_3)$, *from the observed values for the refined structure*

The column headings are threshold values for $|E_1|$. When $|E_1|>1{\cdot}0$ or 1·5, only the permutation $\mathbf{k}_1=\mathbf{h}_1, \mathbf{k}_3=\mathbf{h}_3$ was used to compute the averages in equation (10). When $|E_1|>2{\cdot}0$, 2·25 or 2·5, the three even permutations of $\mathbf{k}_1$, $\mathbf{k}_2$ and $\mathbf{k}_3$ were used.

(*a*) R.m.s. deviations for the epiandrosterone invariants.

	No. of	$p=\frac{1}{2}$						$p=2$					
A range	invariants	0·0	1·0	1·5	2·0	2·25	2·5	0·0	1·0	1·5	2·0	2·25	2·5
0·2–0·4	36	0·78	0·68	0·71	0·67	0·71	0·92	0·84	0·75	0·78	0·79	0·84	1·09
0·4–0·6	254	0·61	0·62	0·61	0·58	0·58	0·64	0·60	0·60	0·60	0·60	0·61	0·75
0·6–0·8	362	0·53	0·58	0·56	0·54	0·54	0·58	0·59	0·60	0·60	0·60	0·61	0·71
0·8–1·0	369	0·50	0·53	0·52	0·51	0·52	0·56	0·57	0·56	0·57	0·60	0·62	0·76
1·0–1·2	244	0·49	0·49	0·48	0·47	0·47	0·51	0·57	0·54	0·55	0·58	0·60	0·81
1·2–1·4	170	0·40	0·38	0·38	0·38	0·38	0·42	0·48	0·45	0·46	0·48	0·50	0·64
1·4–1·6	138	0·42	0·44	0·43	0·43	0·44	0·47	0·52	0·49	0·50	0·54	0·56	0·75
1·6–1·8	119	0·31	0·32	0·32	0·32	0·32	0·35	0·41	0·38	0·39	0·40	0·42	0·57
1·8–2·0	80	0·33	0·32	0·33	0·32	0·33	0·37	0·47	0·42	0·44	0·47	0·51	0·68
2·0–2·2	57	0·35	0·39	0·38	0·36	0·37	0·39	0·46	0·45	0·46	0·49	0·52	0·67
2·2–2·4	37	0·30	0·34	0·33	0·32	0·32	0·33	0·38	0·36	0·37	0·37	0·39	0·47
2·4–2·6	29	0·30	0·34	0·34	0·33	0·33	0·36	0·44	0·43	0·45	0·47	0·49	0·72
2·6–2·8	21	0·25	0·22	0·21	0·21	0·22	0·25	0·40	0·36	0·35	0·36	0·41	0·52
2·8–3·0	8	0·29	0·20	0·21	0·22	0·21	0·23	0·31	0·25	0·25	0·27	0·27	0·32
3·0–5·0	47	0·24	0·28	0·27	0·26	0·26	0·28	0·46	0·45	0·45	0·49	0·55	0·72
0·2–5·0	1971	0·48	0·50	0·49	0·47	0·48	0·53	0·55	0·53	0·54	0·56	0·58	0·73

(*b*) R.m.s. deviations for the 5β-androstane-3α,17β-diol invariants

	No. of	The exponent $p=\frac{1}{2}$					
A range	invariants	0·0	1·0	1·5	2·0	2·25	2·5
1·0–1·2	23	0·47	0·64	0·57	0·57	0·61	0·67
1·2–1·4	115	0·57	0·47	0·45	0·45	0·44	0·48
1·4–1·6	187	0·53	0·43	0·43	0·41	0·41	0·47
1·6–1·8	229	0·54	0·44	0·44	0·42	0·44	0·46
1·8–2·0	239	0·45	0·38	0·38	0·36	0·37	0·41
2·0–2·2	227	0·41	0·37	0·36	0·35	0·36	0·40
2·2–2·4	173	0·38	0·24	0·24	0·26	0·27	0·33
2·4–2·6	150	0·35	0·25	0·25	0·24	0·25	0·30
2·6–2·8	108	0·36	0·30	0·29	0·28	0·28	0·33
2·8–3·0	70	0·37	0·27	0·27	0·26	0·26	0·29
3·0–9·0	382	0·29	0·20	0·19	0·19	0·20	0·24
1·0–9·0	1903	0·43	0·35	0·34	0·33	0·34	0·38

phase for $11\bar{2}$ were markedly smaller than those found on the other maps, and since this map had many fewer spurious peaks with a height comparable to the height of the real atoms, it was expected that the corresponding residuals for the least-squares phase determination cycles might be significantly less than in the other cases, but this was not observed to be true. Furthermore, the observed residuals were about the same as those encountered in the least-squares run which eventually resulted in a Fourier map on which the molecule was translated perpendicular to the $20\bar{4}$ plane.

To determine the overall accuracy of the invariants used to solve the structure, and to see if some groups of invariants are calculated more accurately than others, the true invariants were computed, and the r.m.s. deviation of the predicted invariants from their observed values was used as an indicator of the precision of invariants calculated using equation (10). For purposes of this comparison, invariants predicted to be greater than +1 or less than −1 were not forced to the nearest allowed value of the cosine. If the predicted invariants were distributed at random, then the expected value of the r.m.s. deviations of the invariants from their observed values ranges from 1·15 for very large values of A to 0·82 for very small values of A if all invariants are calculated to be within the allowed range of the cosine. The observed values of 0·48 for 1971 epiandrosterone invariants and 0·43 for 1903 invariants for 5β-androstane-3α,17β-diol are substantially smaller than this. The invariants were then grouped according to the value of A, and the r.m.s. deviation was computed for each range of A values. The results for epiandrosterone are presented in Table 8(*a*) and show that predicted invariants with high values of A (1·5–5·0) are computed much more accurately than are those with low A (0·0–1·0). Similar data for 5β-androstane-3α,17β-diol are given in Table 8(*b*) and confirm the observation that the accuracy of predicted invariants increases as A increases. It should be recalled that the invariants actually used to solve these structures were computed using an exponent p [see equation (10)] equal to $\frac{1}{2}$, and **l** ranged over all measurable reflections (*i.e.* $|E_l|>0$).

It is also a point of interest to determine if the range of **l** can be restricted since the computation of ψ is time consuming even on very fast computers. Consequently, ψ was computed for the epiandrosterone and 5β-androstane-3α,17β-diol invariants by imposing each of the restrictions $|E_l|>1{\cdot}0$, 1·5, 2·0, 2·25, and 2·5 in turn. The calculation of the epiandrosterone invariants was also repeated using an exponent $p=2$ to test experimentally the relative reliability of exponents 2 and $\frac{1}{2}$. The r.m.s. deviations of the invariants computed by each of these methods are shown, as a function of A, in Table 8, and the approximate number of contributors to ψ computed by making the various restrictions on the range of **l**, as well as the amount of time required to calculate 1971 epiandrosterone invariants by each method, are listed in Table 9. These data demonstrate that a substantial amount of time can be saved, without sacrificing accuracy, by restricting the range of **l**. There is little difference in the accuracy of invariants computed with the different restrictions for either structure except for those computed with $|E_l|>2{\cdot}5$ which have slightly larger deviations. However, in view of the small number of contributors to ψ when $|E_l|$ is required to be greater than 2, and that the amount of computing time required is not much greater when $|E_l|>2$, it is suggested that the optimum procedure would be to require that $|E_l|$ be greater than 2. The theoretical justification for this restricted averaging process will be published at a later date. The invariants computed using exponent $p=2$ seem to be less accurate than those computed using $p=\frac{1}{2}$, especially at higher values of A. Since invariants with large A are in general the most accurate and the most useful invariants, it seems to be advisable to use exponent $p=\frac{1}{2}$ rather than an exponent equal to 2.

Table 9. *Number of contributors to individual averages [equation (11)] and* IBM 1130‡ *computing time required to calculate* 1971 *structure invariants for epiandrosterone using various restrictions on the range of* **l**

$\lvert E_l\rvert >$	Number of contributors	Time (hours)
0·0	2000–4000	13·25
1·0*	400–800	4·5
1·5*	150–300	1·75
2·0†	150–300	1·5
2·25†	100–200	1·0
2·5†	50–100	0·5

* Using only the permutation $\mathbf{k}_1=\mathbf{h}_1$ and $\mathbf{k}_3=\mathbf{h}_3$.

† All three even permutations used to generate contributors.

‡ 3·6μs, 8K word core, single-disk storage.

If the three vectors in a given triple are denoted by $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$, three different values of ψ may be computed by permuting the order of the vectors in the triple if the range of **l** depends on $|E_l|$. For example, one value is obtained by making the equivalences $\mathbf{k}_1=\mathbf{h}_1$ and $\mathbf{k}_3=\mathbf{h}_3$ [where $\mathbf{h}_1$ and $\mathbf{h}_3$, as used in equation (11), are any two of the three vectors in the triple], and other averages are formed if $\mathbf{k}_3=\mathbf{h}_1$ and $\mathbf{k}_2=\mathbf{h}_3$ or $\mathbf{k}_2=\mathbf{h}_1$ and $\mathbf{k}_1=\mathbf{h}_3$. If **l** ranges over all measurable reflections (*i.e.* $|E_l|>0$), the three even permutations of vectors in the triple $(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$ will result in the same value for ψ. No new values will be obtained by a permutation of the type $\mathbf{k}_1=\mathbf{h}_3$ and $\mathbf{k}_3=\mathbf{h}_1$, because the only difference in the way in which $\mathbf{h}_1$ and $\mathbf{h}_3$ enter into equation (11) is that $\mathbf{h}_3$ is used in the form $-\mathbf{h}_3$ whereas the sign of $\mathbf{h}_1$ is not changed, and the change of sign on $\mathbf{h}_3$ is compensated by the fact that, for every vector **l** which is used, $-\mathbf{l}$ is also used. When the restrictions $|E_l|>1{\cdot}0$ or 1·5 were made, only the permutation $\mathbf{k}_1=\mathbf{h}_1$ and $\mathbf{k}_3=\mathbf{h}_3$ was used, whereas all three permutations were made and entered into the averages when the range of **l** was more severely restricted. The dependence

of K on A, as a function of the restriction imposed on $\mathbf{l}$, was also studied, and the results are seen in Fig. 2. K appears to become largely independent of A as the restriction on the range of $\mathbf{l}$ becomes more severe, and when $|E_l|$ is required to be greater than 2 (or larger), it seems to be a constant.

A further study of the errors entering into the calculation of the structure invariants revealed that the magnitude of such errors depends on the observed values of the invariants. The average and r.m.s. deviations for two of the sets of invariants computed using $p=\frac{1}{2}$ and a scaling parameter having a linear dependence on A, are shown in Table 10 for various ranges of the observed values of the invariants. The major conclusion to be drawn from these data is that invariants whose observed values are small are the most poorly computed. Not only is the r.m.s. deviation for these invariants much larger than the deviations found when the observed value is larger, but the average deviation is also very large and indicates that there is a substantial nonrandom error entering into the computation of such invariants which appear, on the average, to be calculated substantially larger than they really are. Furthermore, this trend is independent of A.

Table 10. *R.m.s. and average deviations of the computed structure invariants grouped according to the true values of the invariants*

The deviations are ($\cos_{\text{predicted}}-\cos_{\text{observed}}$).

(*a*) Deviations for the epiandrosterone invariants:

			Observed $\cos(\varphi_1+\varphi_2+\varphi_3)$				
			−1·00 to −0·00	0·00 to 0·25	0·25 to 0·50	0·50 to 0·75	0·75 to 1·00
	No. of invariants		204	68	112	163	474
A 0·0–1·0	$\|E_l\|>0$	avg. dev.	0·64	0·18	0·15	−0·04	−0·16
		r.m.s. dev.	0·62	0·17	0·24	0·22	0·23
	$\|E_l\|>2$	avg. dev.	0·90	0·36	0·26	0·04	−0·16
		r.m.s. dev.	0·95	0·19	0·16	0·12	0·13
	No. of invariants		54	21	50	74	290
A 1·0–1·5	$\|E_l\|>0$	avg. dev.	0·69	0·35	0·25	0·00	−0·13
		r.m.s. dev.	0·60	0·20	0·19	0·16	0·14
	$\|E_l\|>2$	avg. dev.	0·94	0·48	0·32	0·08	−0·10
		r.m.s. dev.	0·96	0·28	0·17	0·07	0·07
	No. of invariants		25	10	18	64	344
A 1·5–5·0	$\|E_l\|>0$	avg. dev.	0·83	0·16	0·23	0·12	−0·04
		r.m.s. dev.	0·79	0·09	0·15	0·07	0·06
	$\|E_l\|>2$	avg. dev.	1·09	0·44	0·36	0·14	−0·07
		r.m.s. dev.	1·31	0·23	0·20	0·04	0·03

(*b*) Deviations for the 5β-androstane-3α,17β-diol invariants

			Observed $\cos(\varphi_1+\varphi_2+\varphi_3)$				
			−1·00 to 0·00	0·00 to 0·25	0·25 to 0·50	0·50 to 0·75	0·75 to 1·00
	No. of invariants		26	12	26	44	122
A 1·0–1·5	$\|E_l\|>0$	avg. dev.	0·51	−0·02	−0·11	0·05	−0·08
		r.m.s. dev.	0·58	0·11	0·36	0·29	0·27
	$\|E_l\|>2$	avg. dev.	0·94	0·22	0·17	0·15	−0·10
		r.m.s. dev.	0·99	0·11	0·14	0·10	0·10
	No. of invariants		76	42	86	213	1256
A 1·5–90	$\|E_l\|>0$	avg. dev.	0·28	0·18	0·06	−0·02	−0·04
		r.m.s. dev.	0·30	0·19	0·17	0·15	0·16
	$\|E_l\|>2$	avg. dev.	0·78	0·50	0·28	0·11	−0·04
		r.m.s. dev.	0·70	0·30	0·14	0·07	0·05

It does seem to be true that invariants whose observed values are near unity are computed more accurately if A is large, but the dependence of accuracy on A is not as striking as the dependence of accuracy on the observed value of the invariant. Consequently, the large r.m.s. deviations seen when all invariants with low A are grouped together result primarily from the fact that invariants whose observed values are relatively small are much more frequent at low A.

Summary

The following conclusions are based on the analysis of the epiandrosterone and 5β-androstane-3α,17β-diol data.

1. If an incorrect phase assignment is made in the early stages of a phase buildup, the molecule may still appear on a resulting Fourier map but be translated perpendicular to the planes having the same indices as the reflection in question.

2. Invariants involving certain reflections may be computed less accurately than a general set of invariants, and they may be subject to a nonrandom error.

3. The residuals calculated during the least-squares phase determining procedure do not distinguish all false minima from the true minimum.

4. An exponent $p=\frac{1}{2}$ in equation (11) seems to be somewhat more reliable than an exponent $p=2{\cdot}0$.

5. Invariants computed after imposing restrictions on $|E_1|$ are, on the average, as accurate as those computed when **l** is allowed to range over all reflections, but computing time is substantially less. The condition $|E_1|>2$ is suggested for general use.

6. The scaling parameter K is largely independent of A when the range of **l** is severely restricted. When the range of **l** is unrestricted, K is a function of A, and in the case of these structures, the dependence was linear.

7. Invariants whose observed values are large are computed more accurately than those which are, in reality, small. Invariants which are small show a relatively large nonrandom error since their predicted values are usually larger than their observed values.

8. Invariants with large A appear to be computed much more accurately than invariants with small A, but this is largely a reflection of lower accuracy of the computation of invariants whose observed values are relatively small because such invariants are more frequent at low A.

The authors wish to express their gratitude to Dr A. Cooper for his invaluable advice throughout this investigation, to Mr Harrison Hancock who wrote a number of the computer programs, and to Dr D. A. Norton for her encouragement and support. We are also indebted to Mr C. T. Lu, Miss Jean Ohrt and Mrs C. DeVine who made the experimental X-ray measurements, to Mr D. Maracle who prepared the drawings, and Mrs L. Kirwan and Miss D. Hefner who typed the manuscript.

This investigation was supported in part by U.S.P.H. Research Grant No. CA 10906-02 from the National Cancer Institute.

References

COCHRAN, W. & WOOLFSON, M. M. (1955). *Acta Cryst.* **8**, 1.

COOPER, A., NORTON, D. A. & HAUPTMAN, H. (1969). *Acta Cryst.* B**25**, 814.

HAUPTMAN, H. (1964). *Acta Cryst.* **17**, 1421.

HAUPTMAN, H. (1970). *Crystallographic Computing.* Proceedings of the 1969 International Summer School on Crystallographic Computing, pp. 45–51 Ed. F. AHMED, S. R. HALL & C. P. HUBER. Copenhagen: Munksgaard.

HAUPTMAN, H., FISHER, J., HANCOCK, H. & NORTON, D. A. (1969). *Acta Cryst.* B**25**, 811.

HAUPTMAN, H. & KARLE, J. (1953). *Solution of the Phase Problem. I. The Centrosymmetric Crystal.* Ann Arbor: Edwards Brothers, Inc.

HAUPTMAN, H. & KARLE, J. (1956). *Acta Cryst.* **9**, 45.

HUGHES, E. W. (1953). *Acta Cryst.* **6**, 871.

KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635.

KARLE, J. & HAUPTMAN, H. (1957). *Acta Cryst.* **10**, 515.

KARLE, J. & HAUPTMAN, H. (1961). *Acta Cryst.* **14**, 217.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60.

WEEKS, C. M., COOPER, A., NORTON, D. A., HAUPTMAN, H. & FISHER, J. (1971). *Acta Cryst.* B**27**, 1562.

WEEKS, C. M., HAUPTMAN, H. & NORTON, D. A. To be published.