Acta Cryst. (1972). B28, 1857

Structure Determination *via* the Two-Dimensional Structure Invariants

BY WILLIAM L. DUAX, CHARLES M. WEEKS AND HERBERT HAUPTMAN

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

(*Received* 12 *April* 1971 *and in revised form* 17 *November* 1971)

Although the least-squares analysis of structure invariants has been successfully employed to solve a number of crystal structures, it is known that the values of the invariants, as predicted by existing probability formulas, are often not quite accurate. Since the cosine invariants for triples involving only two-dimensional reflections may have only the discrete values ± 1 in the space group $P2_12_12_1$, whereas general invariants may have any value within the allowed range of the cosine function, it should be relatively easy to predict the correct values of two-dimensional invariants. Consequently, these invariants were computed for an estradiol–urea (1:1) complex and for 6α-fluorocortisol, and they were used to determine basic sets of phases which were used as input to the tangent formula. Both structures were solved by this method and, subsequently, the relative accuracy of the triple product and *MDKS* relationships for predicting the values of this type of invariant was determined using the data for these two structures.

Introduction

It is well established that a sufficiently accurate and large set of phases used as input to the tangent formula (Karle & Hauptman, 1956) usually results in extension of the set of known phases. For moderately complex structures (20 to 40 atoms) which have appreciable overlap in the Patterson synthesis, it is desirable to determine 40 to 60 phases correctly before applying the tangent formula. In the structures of estriol (Hauptman, Fisher, Hancock & Norton, 1969) and two androstane derivatives (Hauptman, Weeks & Fisher, 1971), the correct phasing of this basic set was achieved through the least-squares analysis of structure invariants $\cos(\varphi_1+\varphi_2+\varphi_3)$ (cosine invariants), which were evaluated by a modification of the triple-product formula (Karle & Hauptman, 1957; Hauptman, 1964).

In the case of the two androstane derivatives, statistical analysis, comparing the cosine invariants calculated by the modified triple-product formula with the true values calculated from the solved structure, indicated that the evaluation by this formula was imperfect. Consequently, it was felt that an initial phasing procedure, based on cosine invariants for triples involving only vectors in the centrosymmetric projections, would be less difficult because these 'two-dimensional' (2D) invariants have only two possible values (± 1),* whereas the general three-dimensional (3D) invariants can have any value in the range -1 to $+1$. The phasing problem is therefore reduced to one of selecting which invariants are $+1$.

For many space groups of orthorhombic or higher symmetry, correct phasing of a large set of two-dimensional reflections can be successfully extended to a three-dimensional solution through the tangent formula. This method was applied to the structures of estradiol.urea (1:1) and 6α-fluorocortisol,† both of which crystallize in the space group $P2_12_12_1$. In the case of the estradiol.urea complex, it was not possible to identify unambiguously those invariants that were $+1$, on the basis of the results of the triple-product formula. The intensity statistics for estradiol.urea indicated greater overlap in the Patterson synthesis than is generally observed; indeed, it was found that the fraction of invariants whose value was -1 was much greater than that predicted by theory in which such overlap was assumed not to exist. Application of a more recently developed formula (*MDKS*) (Hauptman, 1972; Fisher, Hancock & Hauptman, 1970*a*,*b*) did afford a better evaluation of the cosine invariants for this structure. The structure of 6α-fluorocortisol did not have extensive overlap in the Patterson synthesis, and the modified triple-product evaluation of the invariants was adequate.

Following the solutions of the structures, the observed values of the cosine invariants for the solved structures were compared with the triple product and *MDKS* predicted values in an attempt to define the optimal procedure for the use of these formulas in calculating the values of the two-dimensional invariants.

Diffraction measurements

All crystallographic measurements were made on a General Electric single-crystal orienter, and the inten-

* In space group $P2_12_12_1$, there are two-dimensional triples of the type $(h\bar{k}0,\ \bar{h}0l,\ 0k\bar{l})$, where $h+k+l=2n+1$, for which the cosine value must be 0. These triples are of no use in the phasing procedure described here, because any combination of the permitted values of the three phases is consistent with the zero value of the cosine.

† Essentially the same technique has been more recently applied with success to the structure determination of methylphenyl glyoxylate.

sity data were collected by the stationary-counter, stationary-crystal technique. The systematic absences in the diffraction patterns were consistent with the orthorhombic space group $P2_12_12_1$ for both the estradiol.urea complex ($C_{18}H_{24}O_2.CON_2H_4$) and 6α-fluorocortisol ($C_{21}H_{29}O_5F$); the cell constants were $a=24{\cdot}631$, $b=7{\cdot}951$, and $c=9{\cdot}302$ Å for the former and $a=13{\cdot}568$, $b=11{\cdot}447$, and $c=12{\cdot}247$ Å for the latter. The final R values [$R=\sum(||F_o|-|F_c||)/\sum|F_o|$] for the observed data were 6% for the estradiol.urea complex and 7% for 6α-fluorocortisol.

A comparison of the observed distributions of the normalized structure factor magnitudes, $|E|$, for these structures with the theoretical distribution for noncentrosymmetric structures is presented in Table 1. The extremely large values of the averages $\langle(|E|^2-1)^2\rangle$ and $\langle(|E|^2-1)^3\rangle$ in the case of the estradiol.urea complex were the first indications that extensive overlap in the Patterson function existed, and that the statistics of this structure would be exceptional.

Table 1. *Intensity statistics*

	Theoretical	Calculated estradiol.urea	Calculated 6α-fluorocortisol
Noncentric (3D) reflections			
$\langle(\|E\|^2-1)^2\rangle$	1·00	1·44	1·05
$\langle(\|E\|^2-1)^3\rangle$	2·00	4·86	3·92
Centric (2D) reflections			
$\langle(\|E\|^2-1)^2\rangle$	2·00	4·12	2·32
$\langle(\|E\|^2-1)^3\rangle$	8·00	35·85	17·29

Calculation of structure invariants

The two-dimensional cosine invariants for the two structures were evaluated using the triple-product formula:

$$\cos(\varphi_1+\varphi_2+\varphi_3)\simeq \frac{K\psi}{|E_1E_2E_3|}+\frac{R_3}{|E_1E_2E_3|} \qquad (1)$$

(Hauptman, Fisher, Hancock & Norton, 1969; Hauptman, Fisher & Weeks, 1971; Hauptman, Weeks & Norton, 1969; for a related procedure see also Karle, 1970).* The quantity

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

is a function only of the normalized structure-factor magnitudes of the reflections forming the $\sum_2$ triple. The function

$$\psi=\langle(|E_k|^{1/2}-\overline{|E|^{1/2}})\,(|E_{h_1+k}|^{1/2}-\overline{|E|^{1/2}})\times(|E_{-h_3+k}|^{1/2}-\overline{|E|^{1/2}})\;|\;|E_k|>t\rangle_k\,, \qquad (3)$$

is a restricted average over all reflections k, such that $|E_k|$ is greater than some threshold t, and $\overline{|E|^{1/2}}$ is the average value of the square root of the normalized structure-factor magnitudes for all reflections in reciprocal space. Analysis of the evaluation of cosines by the triple-product formula for two androstane structures (Hauptman, Fisher & Weeks, 1971) indicated that 2·0 was a suitable value for the constant t. The K values of 305·2 for estradiol.urea and 954·7 for 6α-fluorocortisol were chosen in such a way as to make the empirical distribution of predicted invariants agree, as closely as possible, with the theoretical distribution (Hauptman, 1970*a*; for a related distribution, see Cochran, 1955).

The same cosine invariants were evaluated using the $(D-S)/S$ formula (Hauptman, 1970*b*),

$$\cos(\varphi_1+\varphi_2+\varphi_3)\simeq(D-S)/S\,, \qquad (4)$$

where

$$D=\langle(|E_{-h_3+k}|^2-1)\;|\;|E_k|>t,|E_{h_1+k}|>t\rangle_k \qquad (5)$$

and

$$S=\langle(|E_{-h_3+k}|^2-1)\;|\;|E_k|>t\rangle_k+\langle(|E_{-h_3+k}|^2-1)\;|\;|E_{h_1+k}|>t\rangle_k\,. \qquad (6)$$

The threshold t is an arbitrary fixed number exceeding unity, and D, for example, is the average of $(|E_{-h_3+k}|^2-1)$ taken over all reflections k such that $|E_k|>t$ and $|E_{-h_1+k}|>t$. The optimum value of t is unknown. On the one hand, it is desirable that it be large, because if t is large, then S is large and errors that occur in the numerator of (4) are not unduly exaggerated by division by S. However, if t is large, then both the numbers of contributors to the averages in (5) and (6) are small and errors arising from the finite sampling are large. If, on the other hand, t is chosen to be small so as to increase the numbers of contributors to these averages and thus to reduce the errors arising from the finite sampling, then S is small and whatever errors occur in the numerator of (4) are exaggerated by division by the small number S. As a compromise, the D and S terms were computed using $t=1{\cdot}0$ and $t=1{\cdot}3$. For structures of this complexity (20–30 nonhydrogen atoms in the asymmetric unit), it has been found that the predicted cosine values are erratic if there are fewer than 200 contributors to D. A threshold value of $t=1{\cdot}3$ results in a minimum of approximately 400 contributors to D. As the complexity of the structure increases, the minimum number of contributors to D increases, but t remains approximately constant.

Because of the extensive overlap in the Patterson function of real crystals, the $(D-S)/S$ formula yields a

---

* The following abbreviations are employed in this paper: $|E_i|=|E_{h_i}|$, $\varphi_i=\varphi_{h_i}$, $i=1,2,3$ in which it is assumed that $h_1+h_2+h_3=0$, so that $\cos(\varphi_1+\varphi_2+\varphi_3)$ is a structure invariant. A is defined as:

$$A=\frac{2\sigma_3}{\sigma_2^{3/2}}|E_1E_2E_3|\,,$$

where $\sigma_n=\sum_{j=1}^{N}Z_j^n$, Z_j is the atomic number of the atom labeled j, and there are N atoms in the unit cell.

greater percentage of negative cosine invariants than is actually observed, when the structure is solved and the true values of the invariants are computed. Therefore, equation (4) has recently been altered, by the introduction of scaling parameters, K and M, to give the relationship:

$$\cos(\varphi_1+\varphi_2+\varphi_3) \simeq M(D-KS)\,, \qquad (7)$$

which is referred to as the *MDKS* formula. The constant K of the *MDKS* formula is evaluated so that within each group of triples having approximately the same value of A, the proper proportion of invariants, as predicted by theory, will be negative. The value for M is then found such that the empirical distribution of cosine invariants agrees with the theoretical distribution (*cf.* Hauptman, 1970*a*). In these structures, M and K were found to be independent of A. The actual values of M and K are presented in Table 2.

Table 2. *Values of the scaling constants M and K used in the MDKS formula* [*equation* (7)]

	Threshold ($t=1\cdot0$)		Threshold ($t=1\cdot3$)	
Structure	M	K	M	K
Estradiol.urea	8·72	0·78	3·30	0·65
6α-Fluorocortisol	14·57	0·64	6·04	0·58

Phasing the basic set

Initial attempts to phase the estradiol.urea complex, using two-dimensional cosine invariants computed by

Table 3. 2D *triples for estradiol.urea having* $A>1\cdot0$, *for which the calculated value of the quantity* $(D-S)/S$ *was positive*

The terms D and S are defined in equations 5 and 6, respectively.

#	VECTOR TRIPLE									A	COS *
1	20	1	0	-14	5	0	-6	-6	0	7.02	1
2	20	1	0	-20	0	-1	0	-1	1	6.59	1
3	20	1	0	-13	5	0	-7	-6	0	6.16	1
4	0	1	2	0	-4	3	0	3	-5	6.11	1
5	0	1	2	0	4	3	0	-5	-5	5.67	1
6	21	2	0	-14	5	0	-7	-7	0	5.08	1
7	20	1	0	-20	0	-3	0	-1	3	4.96	1
8	0	1	2	0	-3	3	0	2	-5	4.75	1
9	30	1	0	-30	0	-1	0	-1	1	4.71	1
10	0	1	2	0	1	3	0	-2	-5	4.68	1
11	20	1	0	-30	1	0	10	-2	0	4.46	1
12	21	2	0	-13	5	0	-8	-7	0	4.18	1
13	0	4	3	0	-1	4	0	-3	-7	4.01	1
14	0	1	1	0	1	4	0	-2	-5	3.95	1
15	0	1	2	0	3	7	0	-4	-9	3.91	1
16	20	1	0	-10	2	0	-10	-3	0	3.74	1
17	20	1	0	-11	5	0	-9	-6	0	3.69	1
18	0	4	3	0	1	4	0	-5	-7	3.67	1
19	0	1	2	0	-3	5	0	2	-7	3.60	1
20	0	1	2	0	-5	7	0	4	-9	3.58	1
21	20	1	0	-12	6	0	-8	-7	0	3.42	1
22	20	1	0	-13	6	0	-7	-7	0	3.34	1
23	21	2	0	-21	0	-5	0	-2	5	3.26	1
24	0	1	2	0	-4	5	0	3	-7	3.21	1
25	0	1	2	0	-3	7	0	2	-9	3.20	1
26	29	3	0	-14	5	0	-15	-8	0	3.18	1
27	0	1	1	0	3	2	0	-4	-3	3.06	1
28	0	1	1	0	-4	3	0	3	-4	3.00	1
29	1	3	0	14	5	0	-15	-8	0	2.99	1
30	21	1	0	-14	5	0	-7	-6	0	2.99	1
31	8	1	0	7	7	0	-15	-8	0	2.85	1
32	0	1	4	0	3	5	0	-4	-9	2.84	1
33	20	0	1	1	0	4	-21	0	-5	2.83	1
34	20	0	3	1	0	4	-21	0	-7	2.76	1
35	20	1	0	-9	5	0	-11	-6	0	2.66	1
36	14	5	0	-14	0	-5	0	-5	5	2.63	1
37	8	1	0	-14	5	0	6	-6	0	2.61	1
38	21	2	0	-21	0	-7	0	-2	7	2.53	1
39	0	1	2	0	1	5	0	-2	-7	2.53	1
40	21	2	0	-7	3	0	-14	-5	0	2.52	1
41	21	2	0	-10	3	0	-11	-5	0	2.42	1
42	0	3	3	0	-1	4	0	-2	-7	2.41	1
43	0	1	1	0	-3	4	0	2	-5	2.40	1
44	0	1	3	0	1	4	0	-2	-7	2.37	1
45	0	1	3	0	-4	3	0	3	-6	2.36	1
46	0	1	4	0	-3	5	0	2	-9	2.32	1
47	0	1	1	0	2	5	0	-3	-6	2.32	1
48	20	3	0	-20	0	-5	0	-3	5	2.30	1
49	10	3	0	-14	5	0	4	-8	0	2.27	1
50	30	0	1	-13	0	3	-17	0	-4	2.27	1
51	28	0	0	-8	-1	0	-20	1	0	2.23	1
52	7	1	0	8	7	0	-15	-8	0	2.16	1
53	0	1	2	0	-1	4	0	0	-6	2.15	1
54	0	4	1	0	-1	2	0	-3	-3	2.09	1
55	8	1	0	20	1	0	-28	-2	0	2.03	1
56	0	4	3	0	-2	5	0	-2	-8	2.00	1
57	0	3	2	0	-1	3	0	-2	-5	2.00	1
58	7	1	0	-13	5	0	6	-6	0	1.96	1
59	0	1	1	0	2	2	0	-3	-3	1.94	1
60	10	2	0	-6	6	0	-4	-8	0	1.90	1
61	9	1	0	-15	8	0	6	-9	0	1.88	-1
62	21	2	0	-14	4	0	-7	-6	0	1.85	1
63	0	3	2	0	-4	3	0	1	-5	1.84	-1
64	20	0	3	-21	0	7	1	0	-10	1.83	-1
65	18	0	1	-20	0	1	2	0	-2	1.76	1
66	18	0	1	-30	0	1	12	0	-2	1.74	-1
67	21	2	0	-15	4	0	-6	-6	0	1.74	1
68	10	2	0	4	3	0	-14	-5	0	1.74	1
69	4	2	0	11	6	0	-15	-8	0	1.70	1
70	20	3	0	-5	5	0	-15	-8	0	1.69	1
71	4	3	0	11	5	0	-15	-8	0	1.69	1
72	10	3	0	5	5	0	-15	-8	0	1.67	1
73	18	1	0	-14	5	0	-4	-6	0	1.66	1
74	20	0	5	-21	0	5	1	0	-10	1.64	-1
75	20	1	0	0	2	0	-20	-3	0	1.63	1
76	0	1	4	0	1	5	0	-2	-9	1.63	1
77	21	2	0	-4	6	0	-17	-8	0	1.63	1
78	18	0	1	2	0	2	-20	0	-3	1.62	1
79	9	1	0	14	5	0	-23	-6	0	1.61	1
80	20	0	1	-20	0	3	0	0	-4	1.60	1
81	18	1	0	-11	6	0	-7	-7	0	1.59	1
82	30	1	0	-17	4	0	-13	-5	0	1.55	1
83	30	0	1	-17	0	8	-13	0	-9	1.53	1
84	20	0	1	0	0	4	-20	0	-5	1.52	1
85	18	1	0	-11	5	0	-7	-6	0	1.52	1
86	7	1	0	21	2	0	-28	-3	0	1.51	1
87	0	2	2	0	-5	5	0	3	-7	1.51	1
88	5	1	0	-12	6	0	7	-7	0	1.51	1
89	20	0	1	1	0	6	-21	0	-7	1.49	1
90	0	2	2	0	3	5	0	-5	-7	1.49	1
91	0	4	1	0	-1	4	0	-3	-5	1.48	1
92	18	0	1	-1	0	3	-17	0	-4	1.47	1
93	8	1	0	13	1	0	-21	-2	0	1.46	-1
94	30	0	1	-15	0	3	-15	0	-4	1.45	1
95	0	1	3	0	-3	4	0	2	-7	1.44	1
96	18	1	0	-13	5	0	-5	-6	0	1.43	1
97	0	3	2	0	-5	5	0	2	-7	1.43	-1
98	18	0	1	-20	0	3	2	0	-4	1.42	1
99	30	0	1	-22	0	7	-8	0	-8	1.42	1
100	20	0	1	-21	0	5	1	0	-6	1.41	1
101	18	1	0	-15	8	0	-3	-9	0	1.39	1
102	0	3	2	0	1	3	0	-4	-5	1.38	1
103	8	1	0	-13	5	0	5	-6	0	1.38	1
104	0	4	1	0	1	4	0	-5	-5	1.37	1
105	0	4	2	0	-1	4	0	-3	-6	1.36	1
106	22	0	0	-9	-5	0	-13	5	0	1.36	1
107	18	0	1	-19	0	3	1	0	-4	1.34	1
108	18	0	1	2	0	4	-20	0	-5	1.34	1
109	18	1	0	10	2	0	-28	-3	0	1.33	1
110	1	0	3	20	0	5	-21	0	-8	1.32	1
111	9	1	0	-14	5	0	5	-6	0	1.32	1
112	20	0	1	-21	0	2	1	0	-3	1.31	-1
113	0	3	4	0	-5	5	0	2	-9	1.31	-1
114	6	3	0	9	5	0	-15	-8	0	1.31	1
115	0	1	3	0	-3	6	0	2	-9	1.30	1
116	18	0	1	-1	0	4	-17	0	-5	1.30	1
117	7	3	0	-22	5	0	15	-8	0	1.30	1
118	20	1	0	-14	4	0	-6	-5	0	1.29	1
119	21	0	2	-1	0	4	-20	0	-6	1.29	1
120	10	0	0	20	0	1	-30	0	-1	1.28	-1
121	0	4	3	0	0	6	0	-4	-9	1.27	1
122	3	0	2	17	0	4	-20	0	-6	1.26	1
123	3	1	0	-11	6	0	8	-7	0	1.26	1
124	6	0	0	14	0	5	-20	0	-5	1.25	1
125	9	1	0	-13	5	0	4	-6	0	1.24	-1
126	6	2	0	-20	3	0	14	-5	0	1.23	1
127	0	3	4	0	1	5	0	-4	-9	1.21	-1
128	0	1	4	0	-3	4	0	2	-8	1.20	1
129	0	2	2	0	-3	3	0	1	-5	1.17	1
130	0	3	3	0	0	4	0	-3	-7	1.16	1
131	0	4	2	0	-3	3	0	-1	-5	1.13	1
132	4	3	0	-10	3	0	6	-6	0	1.10	1
133	0	0	4	1	0	4	-1	0	-8	1.02	-1
134	0	2	0	0	1	3	0	-3	-3	1.01	1
135	0	2	0	0	3	5	0	-5	-5	1.01	1
136	22	0	0	-2	0	-5	-20	0	5	1.00	1

* True value of the cosine invariant.

the triple-product formula proved unsuccessful. The same cosine invariants were then re-evaluated using the $(D-S)/S$ formula [equation (4)], and only those invariants for which $(D-S)/S>0$ were retained for use in determining phases. The resulting set of $\sum_2$ triples was assumed to have cosine invariant values of unity, and is presented in Table 3. The criterion that the calculated value of $(D-S)/S$ be positive, in order for the invariant to be accepted as having a value of $+1$, was a stringent restriction, since it resulted in the selection of many fewer invariants than the theoretical number which should have values $+1$. For example, among the invariants with $A>3$, more than 98% should have values of $+1$, but the $(D-S)/S$ calculation resulted in the acceptance of only 65% of the invariants with A values in this range, and these were invariants for which the $(D-S)/S$ values were highest. Among the groups of invariants having lower A values, slightly less than 50% of the invariants were accepted, although even in the group of invariants with $1{\cdot}0<A<1{\cdot}5$, more than 75% should have values of $+1$. Initial phasing, based on this set of cosine invariants, led to the solution of the estradiol.urea structure, and it is this procedure that is described here in detail. The values of the cosine invariants indicated by the triple-product formula were sufficient to allow solution of the 6α-fluorocortisol structure by a similar phasing procedure. Only those aspects of the phasing of 6α-fluorocortisol that differ materially from the procedure followed for estradiol.urea will be discussed.

The list of triples (Table 3), presumed to have cosine invariant values of $+1$ based on the $(D-S)/S$ calculations, was inspected in order to find a set of reflections that interacted with many other reflections and which would consequently be suitable for selecting an origin and enantiomorph. The reflections 043, 20,1,0, and 104 interacted well and also satisfied the parity restrictions placed on the origin-defining reflections in space group $P2_12_12_1$ (Hauptman & Karle, 1956); reflection 011 proved to be suitable for specifying the enantiomorph (Karle & Hauptman, 1956). These reflections, $|E|$ values, and assigned phases, as well as the reason for the phase assignment, are displayed in Table 4 along with similar information for the other two-dimensional reflections in the basis set.

From invariants Nos. 1 and 37 (Table 3), it is apparent that

$$\varphi_{810}=\varphi_{20,1,0}=0\,.$$

Table 4. *Basis set for the estradiol.urea complex as determined from the two-dimensional invariants predicted by the MDKS formula*

The serial number(s) of the invariant(s) in Table 3 used for each phase assignment are indicated.

h	k	l	E	φ	Invariant number
0	1	1	2·01	$\pi/2$	Enantiomorph
0	4	3	3·31	0	Origin
1	0	4	2·69	0	Origin
20	1	0	4·15	0	Origin
8	1	0	1·72	0	1 and 37
20	0	1	2·74	$-\pi/2$	2
0	3	2	1·87	$-\pi/2$	27
0	3	4	1·80	$-\pi/2$	28
0	2	5	2·72	π	43
0	1	4	3·01	$\pi/2$	14
0	3	7	2·20	$\pi/2$	13
0	5	7	2·01	$\pi/2$	18
0	3	6	1·65	$-\pi/2$	47
0	1	3	2·47	$\pi/2$	45, 57
0	1	2	2·80	$\pi/2$	10
0	3	3	2·45	$\pi/2$	8
0	3	5	2·52	$\pi/2$	4
0	5	5	2·15	$\pi/2$	5
20	0	3	2·50	$-\pi/2$	7
0	4	9	2·05	π	15, 20
0	2	7	1·84	π	19, 42
0	4	5	1·76	0	24
0	2	9	1·69	π	25
21	0	7	2·08	$-\pi/2$	34
21	2	0	3·07	$-\pi/2$	38
0	1	5	1·67	$\pi/2$	39, 76
21	0	5	1·96	$-\pi/2$	23, 33
28	0	0	2·27	π	51
28	2	0	1·48	0	55
0	4	1	1·18	0	54, 91
0	0	6	1·39	π	53
0	2	2	1·66	0	59, 87
0	2	8	1·21	0	56
0	0	4	1·69	0	80
20	0	5	2·34	$-\pi/2$	84
20	3	0	1·93	0	48
1	0	10	1·60	0	64, 74*
0	2	0	1·51	0	75
1	0	6	1·30	0	89, 100
13	1	0	1·54	$-\pi/2$	93*
10	3	0	2·11	π	70 and 72
10	2	0	2·40	π	16
30	1	0	2·24	0	11
11	5	0	1·93	$-\pi/2$	41
9	6	0	2·18	$\pi/2$	16
0	4	2	1·55	π	106
30	0	1	3·47	$-\pi/2$	9
10	0	0	0·97	0	120*
1	0	8	1·78	0	133*
14	5	0	3·28	π	
6	6	0	2·42	0	1
7	7	0	2·30	$-\pi/2$	6
13	6	0	1·60	$-\pi/2$	22
15	8	0	3·73	$-\pi/2$	31
29	3	0	1·21	$-\pi/2$	26
1	3	0	1·30	$\pi/2$	29
14	0	5	1·75	$\pi/2$	36
7	3	0	1·27	$\pi/2$	40
4	8	0	1·67	π	49
4	3	0	1·27	0	68
5	5	0	1·16	$\pi/2$	72
6	0	0	2·26	π	124

* Phased incorrectly, due to undetected negative invariants.

(Table 5 may be consulted to aid in transforming the phases of reflections to correspond to sign changes in the indices). From invariant No. 2, the phases of 011 and 20,1,0 determine $\varphi_{20,0,1}$ to be $-\pi/2$. The systematic analysis of those invariants in which the phases of two reflections were known, thereby determining the phase of the third reflection, yielded a set of 49 phases. Table 4 lists the order in which the phasing proceeded, together with the serial number(s) of the invariant(s) determining the phase. The only conflicts in phasing detected during this process concerned 015 and 055. Decisions made regarding these conflicts, based on A values and calculated cosine values, proved to be correct. It was observed that a block of 13 strongly interacting reflections could be introduced into the basic set by assigning a phase to 14,5,0. When $\varphi_{14,5,0}$ was assumed to be π, three resultant phase assignments (6 6 0; 4 8 0; and 6 0 0) agreed with $\sum_1$ indications, and this phasing was therefore considered to be correct.

Table 5. *Phase transformations corresponding to sign changes of indices* ($\alpha = \pm\pi/2$, $\beta = 0$ *or* π)

	hkl	$h\bar{k}l$	$\bar{h}kl$	$\bar{h}\bar{k}l$
ouu	α	$-\alpha$	α	$-\alpha$
oug	α	$-\alpha$	$-\alpha$	α
gou	α	α	$-\alpha$	$-\alpha$
uou	α	$-\alpha$	$-\alpha$	α
ugo	α	$-\alpha$	α	$-\alpha$
uuo	α	α	$-\alpha$	$-\alpha$
ogu	β	β	$\pi+\beta$	$\pi+\beta$
uog	β	$\pi+\beta$	β	$\pi+\beta$
guo	β	$\pi+\beta$	$\pi+\beta$	β
ggg	β	β	β	β

In Table 3, the triples that have true cosine invariant values of -1 are so indicated and constitute a 10% error in invariant evaluation. Fortunately, they caused incorrect phase assignment to only 4 of the 62 reflections used as tangent-formula input (a 6% error). Starting from this basis set, the tangent formula was used to determine 238 additional phases, and 5 or more contributors were required for each new phase. The 62 input phases were held constant during all cycles of tangent-formula refinement.

During the attempts to determine phases using cosine invariants computed by the triple-product formula, the space group invariants 020 and 022 were involved in numerous conflicts between cosine calculations and $\sum_1$ indications. Furthermore, when these reflections were among those whose phases were input for the tangent formula, the refinement invariably altered the phase assignment by π. When phasing was based on invariants computed by the $(D-S)/S$ formula, the ambiguities regarding the phases of 020 and 022 were removed and both phases were determined to be 0, in agreement with weak $\sum_1$ indications. The tangent formula calculated the phase of 020 to be π through all seven cycles of refinement, and, although $\varphi_{0\,2\,2}$ was calculating 0 in early cycles, it eventually became π also. An E map, constructed from this tangent-formula output, revealed the 24 nonhydrogen atoms of the complex among the highest peaks in the map. The true values of the phases of the 020 and 022 reflections proved to be 0, in agreement with their initially determined values, but in conflict with the tangent-formula calculations of π, which were based on 113 and 134 contributors, respectively. This miscalculation of the tangent formula is a result of the unusual statistical distribution of negative cosine invariants and illustrates clearly how a too early and too heavy dependence on the tangent formula may lead to incorrect phasing.

Less interaction of the two-dimensional reflections having large $|E|$ values was observed in the structure of 6α-fluorocortisol. Consequently, the working set of triples having $A < 1{\cdot}00$, which could be considered to have values of $+1$ based on triple-product or *MDKS* calculations, was composed of approximately 150 invariants. In addition to the origin and enantiomorph defining reflections, four of the $\sum_1$ predictions that were considered most reliable were required to build, successfully, a basic set of 51 input phases. Three of these phases were for three-dimensional reflections, each of which occurred in invariants with two-dimensional reflections for which A was greater than 2·5 and the computed cosine was nearly 1·0, so that they could be phased with relative certainty. Twenty-two of the nonhydrogen atoms were located in the first E map, and the remaining five atoms of the steroid A ring were located in a Fourier synthesis based on the E map positions. Four of the 51 input phases were determined to be incorrect.

Statistical distribution of invariant values

The true distributions of negative cosine invariants, as a function of A, for triples composed of two-dimensional (centric) reflections in the estradiol.urea and the 6α-fluorocortisol structures, are compared with the theoretical probability distribution of negative invariants in Table 6. At all A values examined, the number of negative invariants for estradiol.urea is greater than that predicted by theory, and the excess of negative invariants increases as A increases. This observation explains the numerous conflicts encountered in the early attempts to solve this structure. In comparison, 6α-fluorocortisol has fewer negative invariants than are theoretically predicted, except in the $2 < A <$-range, and even in this range the number of excess negative invariants is small. This difference in percentages of negative invariants in the two structures is directly related to the difference in intensity statistics seen in Table 1. Structures having large values for the averages $\langle(|E|^2-1)^2\rangle$ and $\langle(|E|^2-1)^3\rangle$ may be expected to have high percentages of negative invariants, especially for large A.

To compare the relative accuracy with which the cosine invariants, $\cos(\varphi_1+\varphi_2+\varphi_3)$, are calculated by

Table 6. *Percentages of negative two-dimensional cosine invariants*

A Range	No. of invariants	% Negative	% Negative (theoretical)*
Estradiol.urea			
1·0–1·5	282	24	22·3
1·5–2·0	87	20	14·8
2·0–3·0	64	11	7·6
3·0–7·0	48	10	< 2
6α-Fluorocortisol:			
1·0–1·5	154	17	22·3
1·5–2·0	34	9	14·8
2·0–3·0	22	9	7·6
3·0–7·0	9	0	< 2

* Percentage negative 2D invariants = $100/[1+\exp(A)]$.

different methods, the sets of invariants calculated by the triple-product formula [equation (1)] and the *MDKS* formula [equation (7)] for estradiol.urea were each arranged in increasing order of their calculated values, after first breaking the triples into groups having similar A values. Each set of invariants was then broken into quarters, and the percentage of actual negative invariants occurring in each quarter was computed and is displayed in Table 7. For example, 30% of the invariants, with A values in the 1·0–1·5 range, whose values as computed by *MDKS* ($t=1{\cdot}0$) ranked in the lowest quarter for all invariants in this A range, actually had values of -1. It is encouraging to note that for both formulas, and at nearly all A values, the greatest percentages of negative invariants are found among the quarter with lowest calculated invariant values. Also, in general, the percentage of negative invariants in each higher ranking quarter is less than the percentage in the next lower quarter. The deviations from this trend probably result from the smallness of the sample size. A comparison of the results for the two formulas shows that MDKS, especially with a threshold value of 1·0, is more successful than the triple-product formula in identifying the negative invariants in the most useful, higher A ranges. That is, the negative invariants occur more frequently in the lower ranking quarters.

Table 7. *Percentages of actual negative invariants for estradiol.urea, occurring in each quarter of the predicted invariants sorted in increasing order*

The 1st quarter consists of those invariants which had the lowest computed values, and the 4th quarter consists of those invariants with the highest computed values.

Formula	*A* range	Quarter 1st	2nd	3rd	4th
	1·0–1·5	30%	35%	23%	7%
MDKS	1·5–2·0	35	26	9	18
($t=1{\cdot}0$)	2·0–3·0	37	6	0	0
	3·0–7·0	38	0	0	0
	1·0–1·5	32	32	21	10
MDKS	1·5–2·0	47	13	9	18
($t=1{\cdot}3$)	2·0–3·0	25	6	12	0
	3·0–7·0	38	0	0	0
	1·0–1·5	36	26	22	10
Triple product	1·5–2·0	43	26	18	0
($t=2{\cdot}0$)	2·0–3·0	19	19	6	0
	3·0–7·0	23	0	8	8

An important question which now arises is what restrictions should be placed on the invariants that are to be accepted as forming the basis of a phase determination, so that a maximum number of triples are available but a minimum number of triples with negative cosine invariants are included in the working set. To answer this question, the invariants were again divided into four groups having similar A values (1·0–1·5, 1·5–2·0, 2·0–3·0, and 3·0–7·0), and sorted according to their values as predicted by the triple-product and *MDKS* formulas. Within each group of triples with similar A values, certain percentages of the highest ranking invariants were accepted as having values of unity, and the total number of invariants available for use, as well as the percentage error (*i.e.* the percentage of invariants accepted as having values of $+1$ whose true values were -1) in several cases, are presented in Table 8. If 100% of the invariants which should be positive, based on the MDKS ($t=1{\cdot}0$) calculations, are accepted, then 400 estradiol.urea invariants are available for use, and 17% of these invariants have true cosine values of -1. Since 77% of the invariants with A values in the range 1·0–1·5 should, in theory, have values of $+1$ (as should 85% of the invariants having $1{\cdot}5<A<2{\cdot}0$), then accepting 50% of the invariants which 'should be positive' as calculated by *MDKS* means that the highest ranking 38·5% of the invariants with A values in the range 1·0–1·5 were accepted, as were 42·5% of the highest ranking invariants having A values in the range 1·5–2·0. The results shown in Table 8 demonstrate that, for both structures, if all invariants with $1{\cdot}0<A<7{\cdot}0$ are considered, approximately the same amount of error is made by using the results of either formula. If only 50% of the highest ranking invariants that should be positive are accepted, the error is reduced by 3–5% but the number of available triples is only about half the number available if 100% of the invariants that should be positive are accepted. If it is required that a given triple must pass restrictions placed on its calculated value from both formulas simultaneously, the percentage error is reduced an additional 3–4%, but only in the case where 50% of the invariants that should be positive are accepted.

The data presented in Table 9 show a further breakdown of the percentage of negative invariants for estradiol.urea that are accepted as positive. If $A<2$, the requirement that the computed value be among the upper 50% of the invariants expected to be positive, as predicted by both formulas, results in substantial reduction in error. If $A>2{\cdot}0$, restriction to the highest 75% of the invariants predicted to be positive by MDKS greatly reduces the errors encountered. Further restriction only serves to reduce the number of available triples, and the need for corroboration between the formulas is not evident.

Table 8. *Comparison of criteria for acceptance of predicted invariants*

Formula	% of presumed positive invariants accepted	Estradiol.urea Number of invariants accepted	Estradiol.urea % error	6α-Fluorocortisol Number of invariants accepted	6α-Fluorocortisol % error
	100%	400	17%	179	13%
MDKS ($t=1\cdot0$)	75	303	12	134	13
	50	197	12	89	9
	100	400	16	179	11
Triple product ($t=2\cdot0$)	75	303	16	134	10
	50	197	13	89	8
	100	364	16	161	12
MDKS and triple product	75	247	12	107	10
	50	130	8	66	5

Table 9. *Percentage negative invariants accepted as positive for estradiol.urea*

A range	% of presumed positive invariants accepted	*MDKS* ($t=1\cdot0$)	Triple product ($t=2\cdot0$)	*MDKS* and triple product
	100%	21%	19%	19%
1·0–2·0	75	15	19	16
	50	16	16	10
	100	9	8	8
2·0–7·0	75	1	5	2
	50	0	6	0

Conclusion

The necessity of acquiring a strong base of correctly phased reflections before beginning tangent formula refinement, is demonstrated by the ambiguous behavior of the highly interacting 020 and 022 reflections of estradiol.urea. The simplicity and effectiveness of consistently phasing a basic set of two-dimensional reflections from accurately computed cosines have been demonstrated. While the triple-product formula was satisfactory, and perhaps preferable, for computing cosine invariants for 6α-fluorocortisol, it is apparent that the extensive overlap in the Patterson function of estradiol.urea introduced elements not incorporated in the derivation of the triple-product formula. In the derivation of the *MDKS* formula, the problem of overlap has been considered, and the accuracy of the computed cosines for estradiol.urea illustrates the success of this formula. It is hoped that the unambiguous phasing procedure, and the analysis of the comparison between the predicted and the observed true values of the structure invariants described here will provide useful guidelines for future structure solutions.

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

References

COCHRAN, W. (1955). *Acta Cryst.* **8**, 473.

FISHER, J., HANCOCK, J. & HAUPTMAN, H. (1970*a*) *The Conditional Probability Distribution of Crystal Structure Invariants.* Naval Research Laboratory Report 7132.

FISHER, J., HANCOCK, J. & HAUPTMAN, H. (1970*b*). *Computer Program for the Calculation of the Crystal Structure Invariants Based on Certain Conditional Distributions.* Naval Research Laboratory Report 7157.

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

HAUPTMAN, H. (1970*a*). *The Calculation of Structure Invariants in Crystallographic Computing.* Edited by F. R. AHMED, S. R. HALL & C. P. HUBER, Copenhagen: Munksgaard.

HAUPTMAN, H. (1970*b*). Abstract B8, ACA Meeting, New Orleans.

HAUPTMAN, H. (1972). *Z. Kristallogr.* In the press.

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

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

HAUPTMAN, H., WEEKS, C. & FISHER, J. (1971). *Acta Cryst.*, B**27**, 1550.

HAUPTMAN, H., WEEKS, C. & NORTON, D. (1969). Abstract 1X-34, VIIth International Congress of Crystallography at Stony Brook.

KARLE, J. (1970). *Acta Cryst.* B**26**, 1614.

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

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