Acta Cryst. (1972). A**28**, 412

The Relation between Space Group Symmetry and the Properties of Direct Solutions Based on the Σ_2 Formula. A Method for Using Centrosymmetric Σ_2 Solutions for Solving Non-centrosymmetric Structures

By H. Schenk

Laboratory for Crystallography, University of Amsterdam, Nieuwe Prinsengracht 126, *Amsterdam, The Netherlands*

(*Received* 8 *March* 1972)

The refinement criteria (consistency criteria) of the $\sum_2$ and tangent formulae are discussed. From these criteria the properties of $\sum_2$ solutions in various space groups are evaluated. Structure determinations can then be divided into five categories according to the properties of the most consistent $\sum_2$ solution. In three of these, trivial phases are obtained. For solving non-centrosymmetric structures in which the $\sum_2$ formula produces centrosymmetric phases (such as structures in $P2_1$), a new method ($\sum_2 P$ method) is proposed which makes use of both $\sum_2$ and Patterson techniques. In structures for which the most consistent $\sum_2$ solution corresponds to the trivial set of phases $\psi_{hkl}=0$, Harker–Kasper inequalities or other methods have to be used in order to select the most reliable solution from the different sets of phases built up by means of the $\sum_2$ relation. In an appendix, new weighting schemes for the tangent refinement are proposed.

Introduction

In the direct determination of phases of non-centrosymmetric structures the $\sum_2$ formulae

$$\psi_{\mathbf{h}}=\langle\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}}\rangle_{\mathbf{k}}$$

and

$$\psi_{\mathbf{h}}=\frac{\sum_{\mathbf{k}}|E_{\mathbf{k}}E_{\mathbf{h-k}}|\,(\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}})}{\sum_{\mathbf{k}}|E_{\mathbf{k}}E_{\mathbf{h-k}}|}$$

and the tangent formula

$$\tan\psi_{\mathbf{h}}=\frac{\sum_{\mathbf{k}}|E_{\mathbf{k}}E_{\mathbf{h-k}}|\sin(\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}})}{\sum_{\mathbf{k}}|E_{\mathbf{k}}E_{\mathbf{h-k}}|\cos(\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}})}$$

are used.*

Methods based on these expressions (Karle & Karle, 1966; Germain, Main & Woolfson, 1968, 1970, 1971; Hall, 1970; Schenk, 1971; and others: see Ahmed, 1970) try to select those sets of phases which are most consistent with one of the above formulae. However in some cases it has been observed that the $\sum_2$ formula is best fulfilled by apparently incorrect phases. For instance, in the space groups $P2_1$ and $P4_1$, E maps of consistent phase sets show a pseudo mirror plane perpendicular to the polar axis (see, for example, the discussion in Hall, 1970, pp. 69–70, and in Coulter & Dewar, 1971, p. 1736), and in fact these solutions are centrosymmetric. Hall (1970) states that this is due to an incorrect definition of the enantiomorph. Other workers try to overcome this difficulty by holding constant initial phases which are considered the most reliable.

The first purpose of this paper is to show that in many space groups the $\sum_2$ formula gives solutions of incorrect (higher) symmetry. Moreover, very trivial solutions are sometimes obtained. The second purpose is to give a new method for these space groups which makes use of both the $\sum_2$ and the Patterson techniques. In the Appendix a weighting scheme for the tangent refinement is discussed. A preliminary publication has appeared in *Proc. Roy. Neth. Acad. Sci.* (Schenk, 1972*a*).

Consistency criterion for the $\sum_2$ relation

The most useful form of the $\sum_2$ relation is

$$\psi_{\mathbf{h}}=\frac{\sum_{\mathbf{k}}|E_{\mathbf{k}}E_{\mathbf{h-k}}|\,(\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}})}{\sum_{\mathbf{k}}|E_{\mathbf{k}}E_{\mathbf{h-k}}|}\,. \tag{1}$$

If a phase calculated by means of equation (1) is different from its previous value $\psi_{\mathbf{h}}^{\mathrm{old}}$ then:

$$R_2(\psi_{\mathbf{h}}^{\mathrm{old}})>R_2(\psi_h) \tag{2}$$

in which

$$R_2(\psi_{\mathbf{h}})=\sum_{\mathbf{k}}|E_{\mathbf{k}}E_{\mathbf{h-k}}|\,(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}})^2$$

with

$$-\pi<(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}})\le\pi\,.$$

Therefore refinements based on equation (1) consist in minimizing

$$CC_2=\sum_{\mathbf{h}}|E_{\mathbf{h}}|R_2(\psi_{\mathbf{h}})$$
$$=\sum_{\mathbf{h},\mathbf{k}}|E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h-k}}|\,(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}})^2\,. \tag{2a}$$

---

* See end of paper for a list of symbols.

If equation (1) is valid, the phase set corresponding to the true structure should have the lowest value of CC_2. Moreover, if it is assumed that the distribution of $(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})$ is symmetric, then

$$R(\psi_{\mathbf{h}}^{\text{old}}) \geq R(\psi_{\mathbf{h}}) \tag{3}$$

in which

$$R(\psi_{\mathbf{h}}) = \sum_{\mathbf{k}} |E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\,|-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}}|$$

with

$$-\pi < (-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}}) \leq \pi\,.$$

Then refinements based on equation (1) also result in minimizing

$$\begin{aligned} CC &= \sum_{\mathbf{h}} |E_{\mathbf{h}}| R(\psi_{\mathbf{h}}) \\ &= \sum_{\mathbf{h},\mathbf{k}} |E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\,|-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}}|\,. \end{aligned} \tag{3a}$$

In centrosymmetric structures $-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}}$ has only two possible values 0 or π so that $CC_2=\pi CC$. The minimization of both is equivalent to that of

$$\sum_{\mathbf{h}} \sum_{\mathbf{k}} |E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\,[1-\text{sign}\,(\mathbf{h})\ \text{sign}\,(\mathbf{k})\ \text{sign}\,(\mathbf{h}-\mathbf{k})]\,. \tag{4}$$

This criterion is used successfully in automated centrosymmetric symbolic addition.

Consistency criterion for the tangent refinement

Phase refinements based on the tangent formula

$$\tan \psi_{\mathbf{h}} = \frac{S}{C} = \frac{\sum_{\mathbf{k}} |E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}| \sin(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}} |E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}| \cos(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})} \tag{5}$$

are used frequently in direct phasing procedures. However, equation (5) is not employed in its strict form, because then phases can only be calculated in the interval $0 \leq \psi_{\mathbf{h}} < \pi$. With the help of the signs of S and C, a phase $\psi_{\mathbf{h}}$ is uniquely determined in the interval $0 \leq \psi_{\mathbf{h}} < 2\pi$. Thus, instead of equation (5), in practice phases are calculated with

$$\sin \psi_{\mathbf{h}} = \frac{S}{\sum_{\mathbf{k}} |E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|}$$

and

$$\cos \psi_{\mathbf{h}} = \frac{C}{\sum_{\mathbf{k}} |E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|}$$

or with the equivalent expression:

$$\exp(i\psi_{\mathbf{h}}) = \frac{\sum_{\mathbf{k}} |E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}| \exp[i(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})]}{\sum_{\mathbf{k}} |E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|}\,. \tag{6}$$

If a phase $\psi_{\mathbf{h}}$ calculated with equation (6) is different from its previous value $\psi_{\mathbf{h}}^{\text{old}}$ then

$$R_2(\psi_{\mathbf{h}}^{\text{old}}) > R_2(\psi_{\mathbf{h}}) \tag{7}$$

in which

$$R_2(\psi_{\mathbf{h}}) = \sum_{\mathbf{k}} |E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\,|\exp[i(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})]-\exp(i\psi_{\mathbf{h}})|^2\,.$$

Thus refinements based on equation (6) minimize

$$\begin{aligned} CCT_2 &= \sum_{\mathbf{h}} |E_{\mathbf{h}}| R_2(\psi_{\mathbf{h}}) \\ &= \sum_{\mathbf{h},\mathbf{k}} E_{\mathbf{h}}\,|E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}| \exp[i(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})]-\exp(i\psi_{\mathbf{h}})|^2\,. \end{aligned} \tag{8}$$

If equation (6) is valid, the phase set corresponding to the true structure should have the lowest value in equation (8).

Assuming that the distribution of $\exp[i(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})]$ is symmetric it can also be shown that

$$\begin{aligned} CCT &= \sum |E_{\mathbf{h}}| R(\psi_{\mathbf{h}}) \\ &= \sum_{\mathbf{h},\mathbf{k}} |E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\,|\exp[i(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})-\exp(i\psi_{\mathbf{h}})| \end{aligned} \tag{8a}$$

is minimized.

Now

$$|\exp[i(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})-\exp(i\psi_{\mathbf{h}})| = 2|\sin \tfrac{1}{2}(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})|$$

so that the refinement criteria can be written as

$$CCT_2 = \sum_{\mathbf{h},\mathbf{k}} |E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}| \sin^2 \tfrac{1}{2}(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}}) \tag{9}$$

and

$$CCT = \sum_{\mathbf{h},\mathbf{k}} |E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\,|\sin \tfrac{1}{2}(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})|\,. \tag{10}$$

In the case of a centrosymmetric structure for which $-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}}$ equals zero or π, equations (9) and (10) are exactly the same and their minimization is equivalent to that of equation (4).

The effect of symmetry on the consistency criteria

For convenience we will discuss only the properties of equation (3*a*). The results apply equally to equations (2*a*), (9) and (10).

In equation (3*a*) the indices **h** refer to points of the asymmetric part of the reciprocal lattice, and **k** and **h**−**k** to points of the reciprocal lattice as a whole. Now only the phases of the asymmetric part have to be determined, all other phases following from the space-group symmetry relations. If **k** is in the reciprocal lattice as a whole and **k*** is the corresponding reflexion in the asymmetric part, then a symmetry relation can be expressed as:

$$\psi_{\mathbf{k}} = s_{\mathbf{k}}\psi_{\mathbf{k}^*} + c_{\mathbf{k}} \tag{11}$$

in which $s_{\mathbf{k}}=+1$ or -1 and $c_{\mathbf{k}}=0, \frac{1}{2}\pi$, *etc.*, and equation (3*a*) can then be written as

$$CC=\sum_{\mathbf{h},\mathbf{k}} |E_{\mathbf{h}}E_{\mathbf{k}}E_{\mathbf{h}-\mathbf{k}}|\,|-\psi_{\mathbf{h}}+s_{\mathbf{k}}\psi_{\mathbf{k}^*} + c_{\mathbf{k}}+s_{(\mathbf{h}-\mathbf{k})}\psi_{(\mathbf{h}-\mathbf{k})^*}+c_{(\mathbf{h}-\mathbf{k})}| \quad (12)$$

where $\mathbf{h}$, $\mathbf{k}^*$, and $(\mathbf{h}-\mathbf{k})^*$ are now all in the asymmetric part of the reciprocal lattice.

Using the shorthand notations $E_3=|E_{\mathbf{h}}E_{\mathbf{k}}E_{(\mathbf{h}-\mathbf{k})}|$ and $C=c_{\mathbf{k}}+c_{\mathbf{h}-\mathbf{k}}$ the consistency of a set of phases can be given as:

$$CC=\sum_{\mathbf{h},\mathbf{k}} E_3|-\psi_{\mathbf{h}}+s_{\mathbf{k}}\psi_{\mathbf{k}^*}+s_{(\mathbf{h}-\mathbf{k})}\psi_{(\mathbf{h}-\mathbf{k})^*}+C|\,. \quad (13)$$

Although it follows from the validity of equation (1) that the lower the minimum in *CC* the higher the probability of the set phases being correct, in many space groups trivial sets of phases have *CC* values smaller than those based on the correct phases. In this section the properties of sets of phases corresponding to minima in equation (13) will be explored for some space groups, using the symmetry relations as given in *International Tables for X-ray Crystallography* (1952). It will be shown that the triviality of a set of phases is dependent on the values of $s_{\mathbf{k}},s_{(\mathbf{h}-\mathbf{k})}$ and C and on the presence of groups of reflexions, which are restricted to discrete values ($0,\pi,\frac{1}{2}\pi$ *etc.*).

Then structures can be divided into five categories according to the properties of the phase set corresponding to the smallest *CC* value (CC_s):

Category 1: centrosymmetric structures, for which CC_s corresponds to a set of trivial phases.
Category 2: centrosymmetric structures, in which CC_s may correspond to the correct structure.
Category 3: non-centrosymmetric structures, in which CC_s may correspond to the correct structure.
Category 4: non-centrosymmetric structures in which CC_s corresponds to a centrosymmetric (thus incorrect) solution; the centric projection reflexions may have correct phases.
Category 5: non-centrosymmetric structures for which CC_s corresponds to a centrosymmetric trivial set of phases.

Each category will be illustrated by an example.

Example of category 1

In this category all C in equation (13) equal zero. Then the trivial solution with all phases $\psi_{\mathbf{h}}=0$ is the most consistent set of phases, as is well known for space group $P\bar{1}$. The phase relation in $P\bar{1}$ is

$$\psi_{hkl}=\psi_{\overline{hkl}}=\text{zop}\,.$$

Thus $s_{\mathbf{k}}=s_{(\mathbf{h}-\mathbf{k})}=1$ and $C=0$. In this way equation (13) reduces to

$$CC=\sum_{\mathbf{h},\mathbf{k}} E_3|\psi_{\mathbf{h}}+\psi_{\mathbf{k}^*}+\psi_{(\mathbf{h}-\mathbf{k})^*}|\,. \quad (14)$$

This function equals zero for all $\psi_{\mathbf{h}}=0$. Thus in $P\bar{1}$ the lowest *CC* minimum corresponds to a Patterson-related *E* map, and consequently the correct structure must have a higher *CC* criterion. One of the other minima of equation (14) may correspond to the correct structure, but this need not be the solution with the next lowest *CC*.

All centrosymmetric space groups containing positions with point symmetry equal to the point group symmetry of the space group, such as *P*2/*m*, *C*2/*m*, *Pmmm*, *Cmmm*, belong to this category.

Example of category 2

In this category the correct phases can be derived, as will be shown for space group $P2_1/m$. The phase relations of $P2_1/m$ are

$$k=2n \qquad \psi_{hkl}=\psi_{\overline{hkl}}=\psi_{h\bar{k}l}=\psi_{\bar{h}k\bar{l}}=\text{zop}$$

$$k=2n+1 \begin{cases} \psi_{hkl}=\psi_{\overline{hkl}}=\text{zop} \\ \psi_{h\bar{k}l}=\psi_{\bar{h}k\bar{l}}=\psi_{hkl}+\pi\,. \end{cases}$$

Then in (13) s_k, $s_{(h-k)}=+1$ and $C=\text{zop}$. Thus (13) reduces to

$$CC=\sum_{\mathbf{h},\mathbf{k}} E_3|\psi_{\mathbf{h}}+\psi_{\mathbf{k}^*}+\psi_{(\mathbf{h}-\mathbf{k})^*}+\text{zop}|\,. \quad (15)$$

Because $C=\text{zop}$, the trivial solution $\psi_{\mathbf{h}}=0$ is no longer the one with the lowest *CC*. The process of finding the correct phases $\psi_{\mathbf{h}}$ is now equivalent to the location of the lowest minimum of *CC*.

Other space groups in this category are $P2_1/c$, $C2_1/m$, *C*2/*c*, *Pnma etc.* In practice, structures in this category appear to offer little trouble.

Example of category 3

There are some non-centrosymmetric space groups, for instance $P222_1$, for which the phases may be calculated correctly by means of (1). In $P222_1$ the symmetry relations are:

$$\psi_{hk0}=\text{zop},\ \psi_{0kl}=\text{zop}\,,$$

$$1=2n \begin{cases} \psi_{hkl}=-\psi_{\overline{hkl}},\ \psi_{\bar{h}kl}=-\psi_{hkl},\ \psi_{h\bar{k}l}=-\psi_{hkl}, \\ \psi_{hk\bar{l}}=-\psi_{hkl},\ \psi_{h0l}=\text{zop}\,. \end{cases}$$

$$1=2n+1 \begin{cases} \psi_{hkl}=-\psi_{\overline{hkl}},\ \psi_{\bar{h}kl}=-\psi_{hkl}\,, \\ \psi_{h\bar{k}l}=-\psi_{hkl}+\pi,\ \psi_{hk\bar{l}}=-\psi_{hkl}+\pi\,, \\ \psi_{h0l}=\text{htp}\,. \end{cases}$$

Thus $s_{\mathbf{k}}, s_{\mathbf{h}-\mathbf{k}}=+1$ or -1 and $C=\text{zop}$. The consistency criterion (13) reduces to:

$$CC=\sum_{\mathbf{h},\mathbf{k}} E_3|-\psi_{\mathbf{h}}\pm\psi_{\mathbf{k}^*}\pm\psi_{(\mathbf{h}-\mathbf{k})^*}+\text{zop}|\,. \quad (16)$$

A minimum in (16) corresponds to a set of true non-centrosymmetric phases. This can be seen from a reflexion forming the terms of (16) with two projection reflexions of different projections.

Two terms of (16) could be

$$E_3|-\psi_{111}+\psi_{010}+\psi_{101}+(\mathrm{zop})| \qquad (17)$$

$$E_3|-\psi_{111}+\psi_{011}+\psi_{100}+(\mathrm{zop})| \qquad (18)$$

and thus

$$E_3|-\psi_{111}+(\mathrm{zop})+(\mathrm{htp})+(\mathrm{zop})| \qquad (19)$$

$$E_3|-\psi_{111}+(\mathrm{zop})+(\mathrm{zop})+(\mathrm{zop})|\,. \qquad (20)$$

In order to simultaneously satisfy relations (19) and (20) as nearly as possible, ψ_{111} will have to assume a true non-centrosymmetric value. Thus the smallest minimum in equation (16) corresponds to a set of non-centrosymmetric phases and if equation (1) is valid, this set may well define the correct structure.

All space groups having an origin fixed by symmetry and at least one screw axis, such as $P222_1$, $P2_12_12$, $P2_12_12_1$, belong to this category.

Example of category 4

In this category the phases of the centric projection of the non-centrosymmetric structure may be found correctly, but all other phases are calculated centrosymmetric as well. This is the case in $P2_1$ for instance, for which the symmetry relations are:

$$k=2n \quad \begin{cases} \psi_{\bar{h}k\bar{l}}=\psi_{hkl} \\ \psi_{\bar{h}\bar{k}\bar{l}}=\psi_{h\bar{k}l}=-\psi_{hkl} \end{cases}$$

$$k=2n+1 \quad \begin{cases} \psi_{\bar{h}\bar{k}\bar{l}}=-\psi_{hkl} \\ \psi_{\bar{h}k\bar{l}}=\pi+\psi_{hkl} \\ \psi_{h\bar{k}l}=\pi-\psi_{hkl} \\ \psi_{h0l}=\mathrm{zop}\,. \end{cases}$$

Because we choose the asymmetric part of the reciprocal lattice such that $k\geq 0$, equation (13) reduces to:

$$CC=\sum_{\mathbf{h},\mathbf{k}} E_3|-\psi_{\mathbf{h}}+\mathrm{sign}\,(k')\psi_{\mathbf{k}*} + \mathrm{sign}\,(k-k')\psi_{(\mathbf{h}-\mathbf{k})*}+\mathrm{zop}|\,. \qquad (21)$$

The phases of the projection reflexions $h0l$ can be calculated with equation (1):

$$\psi_{h0l}=\frac{\sum\limits_{h'k'l'}|E_{h'k'l'}E_{h-h',\bar{k}',l-l'}|\,(\psi_{h'k'l'}+\psi_{h-h',\bar{k}',l-l'})}{\sum\limits_{h',k',l'}|E_{h'k'l'}E_{h-h',\bar{k}',l-l'}|}\,. \qquad (22)$$

Thus in equation (21) a considerable number of terms are of the type

$$-\psi_{h0l}+\psi_{h_2kl_2}-\psi_{h_3kl_3}+\mathrm{zop}=0 \qquad (23)$$

where, as $\psi_{h0l}=\mathrm{zop}$, the phases of the two hkl reflexions are related by

$$\psi_{h_2kl_2}=\psi_{h_3kl_3}+\mathrm{zop}\,. \qquad (24)$$

From equation (24) it can be seen that the phases of all reflexions with the same value of k are found equal modulo π. As can be easily seen from equation (21) the solutions of the $\sum_2$ relation consist of sets of phases $\psi_{hkl}=kq+\mathrm{zop}$ (in which q is a constant within the set) which, apart from an origin shift, are identical with the centrosymmetrical solution $\psi_{hkl}=\mathrm{zop}$. Therefore the solution with the lowest CC does not represent the correct structure, though the phases of the $h0l$ projection may be correct. Since the correct phases are not restricted to zop, the terms of equation (21) for the correct phases must deviate from 0, which results in a higher consistency index CC for the correct structure.

For the known structure of L-alanyl-glycine (Koch & Germain, 1970), space group $P2_1$, the above statements have been verified. Tangent refinements (see Appendix) have been carried out for the 136 reflexions with the largest $E_{\mathbf{h}}$, using 1085 triples $\mathbf{h},\mathbf{k},\mathbf{h}-\mathbf{k}$, all with $E_3\geq 2$. The first tangent refinement was done with the true phases $\psi_{\mathbf{h}}^t$ of the strongest 41 reflexions as input. These phases were kept constant during 5 cycles, then all 136 reflexions were allowed to change during 25 additional cycles. For each cycle the consistency, equation (3*a*), is given in Table 1. The phases are converging to a centrosymmetric solution with phases $\alpha_{\mathbf{h}}$. At the end of the 30th cycle the mean deviation of the actual values $\psi_{\mathbf{h}}$ of the 136 phases from their centrosymmetric values $\alpha_{\mathbf{h}}$ is 24°, whereas the true structure has a mean deviation from centrosymmetry of 45°.

Table 1. *Tangent refinement of* 136 *reflexions of* L-*alanyl-glycine using the true phases of the* 41 *strongest* *E's as input*

These phases were held constant for five cycles. In column 1 the cycle number is given, in column 2 the consistency criterion CC and in column 3 the total sum of E_3 which is included in the calculation of CC.

1	2	3
1	5636	90
2	90215	822
3	89923	931
4	88014	931
5	87707	931
6	87628	931
7	83816	931
8	82050	931
9	81140	931
10	80413	931
11	79973	931
12	79593	931
13	79304	931
14	79044	931
15	78860	931
16	78641	931
17	78486	931
18	78285	931
19	78165	931
20	77999	931
22	77760	931
24	77576	931
26	77425	931
28	77296	931
30	77151	931

For the second tangent refinement, the 41 strongest E's were given the pure centrosymmetric phases $\alpha_{\mathbf{h}}$, as deduced from the first tangent refinement. Again these phases were held constant during the five first cycles. Then all 136 phases were refined for another five cycles, the results of which are given in Table 2.

By comparing Table 1 and Table 2 it can be seen that:

(1) The internal consistency of the starting group of 41 reflexions is much better for the $\alpha_{\mathbf{h}}$'s than for the $\psi_{\mathbf{h}}^t$'s (2185 and 5636 respectively).

(2) After five cycles the $\alpha_{\mathbf{h}}$-refinement has a smaller consistency criterion than the corresponding $\psi_{\mathbf{h}}$ refinement (68695 and 87707 respectively).

(3) The final consistency of the set $\alpha_{\mathbf{h}}$ is even better than that of the $\psi_{\mathbf{h}}$ after 30 refinement cycles (67420 and 77151 respectively).

Table 2. *Tangent refinement of the* 136 *phases of* L-*analyl-glycine using as input data the* $\alpha_{\mathbf{h}}$ *values of the strongest* 41 *E's, as calculated from the last cycle of the refinement described in Table* 1

These 41 phases were kept constant for the first five cycles. See Table 1 for the meaning of the columns.

1	2	3
1	2185	90
2	68305	822
3	70635	931
4	68695	931
5	68695	931
6	67935	931
7	67935	931
8	67420	931
9	67420	931
10	67420	931

Other examples of space groups belonging to category 4 are *Pc*, *C*2$_1$, *P*2$_1$*nb*, *C*2*ca*, *P*4 *etc.* These have in common an origin that is not fixed and at least one screw axis or glide plane.

Example of category 5

In this group the solutions of the $\sum_2$ formula are also centrosymmetric, but now the smallest *CC* value corresponds to a trivial solution. An example is *P*1.

The phase relation in *P*1 is:

$$\psi_{hkl} = -\psi_{\bar{h}\bar{k}\bar{l}} \,. \tag{25}$$

Thus with $s_{\mathbf{k}}$, $s_{(\mathbf{h}-\mathbf{k})} = -1$ or $+1$ and $C=0$, equation (13) is changed to

$$CC = \sum_{\mathbf{h},\mathbf{k}} E_3 | -\psi_{\mathbf{h}} \pm \psi_{\mathbf{k}^*} \pm \psi_{(\mathbf{h}-\mathbf{k})^*} | \,. \tag{26}$$

The Patterson related trivial solution with all $\psi_i = 0$ gives a consistency criterion $CC=0$. Moreover all sets of phases

$$\psi_{hkl} = hp + kq + lr \tag{27}$$

in which p, q and r are constants within a set, are fully consistent ($CC=0$) and correspond to the same trivial solution.

An *E* map calculated with the phases from equation (27) is always equivalent to an *E* map based on $\psi_{hkl}=0$, in spite of the fact that, at first sight, the phases appear non-centrosymmetric. The only difference is a shift of origin. The other minima in equation (26) also correspond to centrosymmetric phase sets.

All space groups containing positions with point symmetry equal to the point group symmetry of the space group, such as *P*2, *Pm*, *C*2, *Cm*, *P*222, *P*2*mm*, belong to this category.

Practical consequences

Because of the properties of the $\sum_2$ formula, equation (1), and the tangent formula, equation (6), in various space groups, the strategy of tackling a structure by means of these formulae must be different for each category mentioned in the preceeding section. We will briefly outline the approaches for these categories:

Category 1

The fully consistent $\sum_2$ solution has phases $\psi_{\mathbf{h}}=0$. In our opinion the consistency indices cannot be employed to find the correct phases, because in many cases the next most consistent phase sets do not correspond to the correct structure either, but to superpositions of a number of cell contents at wrong positions. Other methods (for instance Harker–Kasper inequalities, producing negative signs) have to be used in order to select the correct phase set from all the sets built up by means of the $\sum_2$ relation.

An example of this procedure is the structure determination of a derivative of cyclopropane with the formula $C_6H_9O_3N$, space group $P\bar{1}$, $Z=2$, very recently carried out in our laboratory. Automatic symbolic addition (Schenk, 1969) was carried out with three symbols. Using Harker–Kasper inequalities the reliability of the fifth and sixth solution in order of increasing consistency *CC* was higher than that of the others. An *E* map of the fifth solution proved to be correct. Details of the structure will be published in due course (de Jong & Schenk, 1972).

Category 2

In this category the consistency criterion is indeed a measure of the reliability of a phase set. A sensible procedure here is to calculate *E* maps in order of increasing consistency criterion *CC*.

Category 3

For the non-centrosymmetric structures in this category the situation is similar to that of category 2. Thus the symbolic addition method (Karle & Karle, 1966), multisolution tangent refinement (Germain & Woolfson, 1968), fast symbol determination method (Schenk, 1972*a*,*b*) and other methods (see Ahmed, 1970), which are all methods based on the $\sum_2$ formula only, may produce the correct structure.

Category 4

In these non-centrosymmetric space groups the $\sum_2$ relation produces centrosymmetric phase sets. Thus at most it can be expected that in one or more phase sets the phases of the centrosymmetric projection are correct. If this is the case and the complete centrosymmetric phase set does contain information about the true structure, it can only be that both the true structure and its enantiomorph are present in the *E* map. The assumption that the centrosymmetric $\sum_2$ solution of smallest *CC* criterion may correspond to such a superposition is plausible. It is borne out by the experimental results of several investigators.

In the next section a new method based on this principle will be proposed for solving structures in categories 4 and 5. At the moment three structures have been solved using this method.

Category 5

In this category of non-centrosymmetric structures, where $\sum_2$ produces only centrosymmetric phases, the consistency criterion is no indicator of the relation of the centrosymmetric E maps and the true structure for the same reason as for group 1. Other methods have to be used to estimate the reliability of sets of phases obtained by means of $\sum_2$ or tangent formulae. Harker–Kasper inequalities may be effective, but obviously not for $P1$. To find the true structure from a reliable centrosymmetric E map thus obtained, the method to be described in the next section can be used.

The Σ_2Patterson method (Σ_2P method)

This method will be effective in non-centrosymmetric structure determinations of categories 4 and 5. Since the $\sum_2$ relation produces centrosymmetric phases anyway, the most sensible procedure is to tackle these structures centrosymmetrically. Then in the last stage the asymmetric unit of the true structure will be found from Patterson superpositions.

The method will be outlined for the frequently occurring space group $P2_1$, but with very small changes it is useful for all non-centrosymmetric space groups in which the $\sum_2$ or tangent formula produces sets of centrosymmetric phases.

The interpretation of centric $\sum_2$ solutions of non-centrosymmetric structures

The Fourier summation of a non-centrosymmetric structure can be given by

$$\varrho(r) = \sum_{\mathbf{h}} |F_{\mathbf{h}}| \exp(i\psi_{\mathbf{h}}) \exp(2\pi\, i\mathbf{h}r)$$

and the Fourier summation of its enanthiomorph inverted at the origin by

$$\varrho^1(r) = \sum_{\mathbf{h}} |F_{\mathbf{h}}| \exp(-i\psi_{\mathbf{h}}) \exp(2\pi i\mathbf{h}r)\,.$$

Thus the E map of the sum of structure and enantiomorph is given by

$$\begin{aligned}\varrho_c(r) = \varrho(r) + \varrho^1(r) &= \sum_{\mathbf{h}} |E_{\mathbf{h}}| \left[\exp(i\psi_{\mathbf{h}}) + \exp(-i\psi_{\mathbf{h}})\right] \\ &\quad \times \exp(2\pi i\mathbf{h}r) \\ &= \sum_{\mathbf{h}} |E_{\mathbf{h}}| \cos\psi_{\mathbf{h}} \exp(2\pi\, i\mathbf{h}r) \\ &= \sum s_{\mathbf{h}} |E_{\mathbf{h}} \cos\psi_{\mathbf{h}}| \exp(2\pi\, i\mathbf{h}r)\end{aligned} \tag{30}$$

in which $s_{\mathbf{h}} = \text{sign}(\cos\psi_{\mathbf{h}})$.

For this compound structure the consistency criterion for centrosymmetric structures (4) is:

$$\sum_{\mathbf{h},\mathbf{k}} |E_{\mathbf{h}} \cos\psi_{\mathbf{h}} E_{\mathbf{k}} \cos\psi_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}} \cos\psi_{\mathbf{h}-\mathbf{k}}| \times (1 - s_{\mathbf{h}} s_{\mathbf{k}} s_{\mathbf{h}-\mathbf{k}}) = \text{minimum}\,. \tag{31}$$

In practice only E maps

$$\varrho_c' = \sum_{\mathbf{h}} s_{\mathbf{h}}' |E_{\mathbf{h}}| \exp(2\pi\, i\mathbf{h}r) \tag{32}$$

can be calculated in which the sets of signs $s_{\mathbf{h}}'$ correspond to minima in

$$\sum_{\mathbf{h},\mathbf{k}} |E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| (1 - s_{\mathbf{h}}' s_{\mathbf{k}}' s_{\mathbf{h}-\mathbf{k}}')\,. \tag{33}$$

If in $P2_1$ a phase set $\psi_{\mathbf{h}}^t$ corresponds to the true structure, a shift $q/2\pi$ of the origin along the y axis gives rise to a new phase set

$$\psi_{\mathbf{h}} = \psi_{\mathbf{h}}^t + kq\,.$$

Thus there are an infinite number of phase sets each corresponding to the true structure and each giving rise to a compound E map, equation (30), of symmetry $P2_1/m$ with its own consistency criterion, equation (31). If the set of signs $s_{\mathbf{h}}'$ calculated with equation (33) is approximately equal to the $s_{\mathbf{h}}$ of one of the equations (31) then an E map, equation (32), gives a distorted image of the corresponding compound structure; distorted because instead of $|E_{\mathbf{h}} \cos\psi_{\mathbf{h}}|$ the $|E_{\mathbf{h}}|$ have been used.

Three comments have to be made:

(1) More than one solution of equation (33) may correspond to a correct compound structure, each with its own minimum, equation (31).

(2) Because of the approximations, the lowest minimum in equation (33) does not necessarily contain the structure and its enantiomorph.

(3) In $P2_1$,

$$SE = \sum_{\mathbf{hk}} |E_{\mathbf{h}} E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}} \cos\psi_{\mathbf{h}} \cos\psi_{\mathbf{k}} \cos\psi_{\mathbf{h}-\mathbf{k}}|$$

is a function of the phases $\psi_{\mathbf{h}}$. Minima in equation (33) correspond to those functions of equation (31) in which SE is large. This is the case for those compound structures in which as many atoms of the structure as possible coincide with atoms of the enantiomorph. Since the $\sum_2$ solutions of $P2_1$ are found in $P2_1/m$, at least one atom will be found in the mirror plane. If several atoms have the same y coordinate, these atoms will be found in the mirror plane.

Implementation

The $\sum_2P$ method for solving non-centrosymmetric structures with centrosymmetric $\sum_2$ solutions makes use of both centrosymmetric direct phasing and the Patterson technique. For $P2_1$ the method consists of five steps;

Step 1: centrosymmetric symbolic addition in space group $P2_1/m$. Classification of the various solutions and order of increasing consistency CC.

Step 2: calculation of the compound E map of the most reliable solution.

Step 3: searching for the highest maximum M in the E map. If the compound E map is correct, this peak is the position of coinciding atoms of structure and enantiomorph.

Step 4: calculation of a Patterson superposition map from an E^2–Patterson synthesis on the basis of the coordinates of M and its equivalent position in $P2_1$. If the E map contains a correct compound structure this superposition map should reproduce the E map, and the next step can be carried out. Otherwise, the next most reliable solution is tested starting with step 2.

Step 5: in the E map a second maximum is selected which in all probability belongs to one of the enantiomorphs only, and this is used to calculate another Patterson superposition map. In this map the complete asymmetric unit of the $P2_1$ cell should be present, Difficulties may arise for structures in which the molecule has a pseudo mirror plane perpendicular to the 2_1 axis.

For the other space groups of category 4 a similar procedure can be used. In space groups of category 5, the only modification is in step 1 where the classification of the various solutions had to be carried out using, for instance, Harker–Kasper inequalities.

Results

The known structure of L-alanyl-glycine (Koch & Germain, 1970) space group $P2_1$, has been solved by the $\sum_2 P$ method. Following Koch & Germain the phases of 303, $\bar{2}05$ and $\bar{5}12$ were set at zero in order to define the origin but the additional starting phases of $\bar{1}34$, $\bar{5}21$, 462 and 104 were each restricted to 0 or π. Then, with the 16 different starting sets, tangent refinement was carried out. It appeared that the set $\psi_{\bar{1}34}=0$, $\psi_{\bar{5}21}=0$, $\psi_{462}=0$ and $\psi_{104}=\pi$ resulted in the best consistency. An E map showed peaks at all the positions given by Koch & Germain and at the positions of the corresponding enantiomorph. A fourfold Patterson superposition map revealed all atoms of the correct structure only. It is interesting to note that tangent refinement of the correct phases leads to the centrosymmetric solution described above (see Table 2).

Mr E. Hessling of our laboratory has solved the same structure independently, using an alternative starting set of symbolic phases in the half-automatic symbolic addition (Schenk, 1969). His best solution was similar to that shown in Table 2. The map of the second best solution also represented a correct but different compound structure. Whereas in the first solution the highest maximum in the mirror plane was C(4), in solution 2 it was C(3).

The unknown structure of 1,1,5-trimethyl-6-carboxyl-cyclohexene-5, ($C_{10}H_{16}O_2$), has been tackled succesfully in the same way. The acid crystallizes in the space group $C2ca$ with $Z=8$. Centrosymmetric symbolic addition using automatic programs (Schenk, 1969) was carried out in space group $Cmca$ with 4 symbols. An E map of the most consistent solution, followed by two fourfold Patterson superpositions, revealed all the atoms of the molecule. Full details of the structure will be published elsewhere (Schenk, 1972).

Another structure which has been solved using the $\sum_2 P$ method is that of benzimidazole. This compound crystallizes in space group $P2_1nb$, $Z=4$. Symbolic addition was carried out in space group $Pmnb$. Full details of this structure will be published in due course (Dik, 1972).

Comparison of the $\sum_2 P$ method with other direct methods

For structure determinations in categories 4 and 5, the $\sum_2 P$ method will be compared with two different direct methods: a $\sum_2$ method and the structure invariant method (Hauptman, 1970). As representative of the $\sum_2$ methods the multisolution tangent refinement procedure (Germain & Woolfson, 1968) is arbitrarily chosen.

The multisolution tangent refinement-method (MTR method)

Although the MTR method (Germain & Woolfson, 1968) starts with sets of essentially non-centrosymmetric phases, the final sets are pseudo-centrosymmetric as a result of the properties of the refinement criterion, equation (9). However, at first sight this is not obvious, because, in $P2_1$ for instance, sets of phases $\psi_{\mathbf{h}}=\psi_{hkl}=kq+\alpha_{\mathbf{h}}$, with $\alpha_{\mathbf{h}}=$zop and q a constant within the set, are also centrosymmetric.

The set $\alpha_{\mathbf{h}}$ can be extracted from $\psi_{\mathbf{h}}$ in the following way:

(1) determination of q by finding the linear relation between $\psi_{\mathbf{h}}$ and k,

(2) calculating new phases

$$\psi^1_{\mathbf{h}}=\psi_{\mathbf{h}}-kq\,, \tag{34}$$

(3) forming the centrosymmetric phases $\alpha_{\mathbf{h}}$ such that:

$$\left.\begin{array}{l}\alpha_{\mathbf{h}}=0 \text{ if } -\dfrac{\pi}{2}<\psi^1_{\mathbf{h}}\le\dfrac{\pi}{2}\\ \alpha_{\mathbf{h}}=\pi \text{ if } \dfrac{\pi}{2}<\psi^1_{\mathbf{h}}\le\dfrac{3\pi}{2}\end{array}\right\}. \tag{35}$$

Then the mean deviation from centrosymmetry of the phases $\psi^1_{\mathbf{h}}$ is:

$$DC=\frac{\sum\limits_{\mathbf{h}}|\psi^1_{\mathbf{h}}-\alpha_{\mathbf{h}}|}{\sum\limits_{\mathbf{h}}1}\,. \tag{36}$$

Now, if the values of the starting phases in the *MTR* method were restricted to 0 or π, the phases $\alpha_{\mathbf{h}}$ would be obtained directly. Since in the *MTR* method the values of the starting phases were chosen as $\frac{\pi}{4}$, $\frac{3\pi}{4}$, $\frac{5\pi}{4}$ and $\frac{7\pi}{4}$, many different starting sets must refine to the same centrosymmetric phase set $\alpha_{\mathbf{h}}$.

In their paper on L-alanyl-glycine, Koch & Germain (1970) reported that in *E* maps of the ten most consistent solutions, large parts of the molecule were present in addition to a number of ghost peaks (note that these peaks must belong to the enantiomorph). We repeated the tangent-refinement process of Koch and Germain; for each set of phases $\psi_{\mathbf{h}}$ we determined the corresponding set $\alpha_{\mathbf{h}}$ and the deviation from centrosymmetry, *DC*. The results are given in Table 3. Nine solutions have the same set of centrosymmetric phases $\alpha_{\mathbf{h}}$, the *E* map of which gives the superposition of structure and enantiomorph. A second set of centrosymmetric phases $\alpha'_{\mathbf{h}}$ was the basis of the next 13 solutions. Only 20% of the phases, $\alpha'_{\mathbf{h}}$, were different from the $\alpha_{\mathbf{h}}$, so that possibly part of the molecule will be present in an *E* map based on $\alpha'_{\mathbf{h}}$.

In conclusion, the following objections to the use of the *MTR* method in structures coming in the categories 4 and 5 can be formulated:

(1) Many essentially equal solutions are all refined by the tangent procedure. In order to avoid the calculation of a great number of nearly equal *E* maps, a rather complicated analysis is necessary to find out which solutions have approximately equal phases.

(2) The *E* maps may be difficult to interpret because many ghost peaks arising from the enantiomorph are present. These ghost peaks may even be higher than the corresponding 'correct' peaks.

The structure-invariant method (*SI method*)

With the *SI* method (Hauptman, 1970) numerical values of the structure invariants

$$C_{\mathbf{k}}=\cos\left(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h-k}}\right)$$

are predicted using a complicated expression. The origin defining reflexions are chosen, and, with other relations ($\sum_1$ *etc.*), a small number of additional phases

Table 3. *Multiple tangent-refinement of* L-*analyl-glycine*

The method used for searching the centrosymmetric phase set is given in equations (34), (35) and (36) The phases are given in mc/s=(radians/2π)10^3. The origin defining reflexions are ψ_{303}=0mc/s; $\psi_{\bar{2}05}$=0mc/s and $\psi_{\bar{5}12}$=0mc/s. The 31 different solutions belong to only three different sets of centrosymmetric phases.

Additional starting phases				Refinement criterion on relative scale	Constant *q* from equation (34) in mc/s	Centro- symmetric set of equation (35)	*DC* of equation (36) in mc/s
$\bar{1}34$ $E=2\cdot27$ ψ	$\bar{5}21$ $E=2\cdot55$ ψ	462 $E=2\cdot32$ ψ	104 $E=2\cdot56$ ψ				
125	125	125	500	0·17	+26	α_h	44
125	125	375	500	0·17	+60	α_h	33
625	625	125	500	0·17	−145	α_h	26
625	875	375	500	0·18	−110	α_h	34
625	625	375	500	0·18	−112	α_h	65
125	875	125	500	0·19	+25	α_h	36
625	375	125	500	0·20	+196	α_h	63
625	875	125	500	0·21	−150	α_h	80
625	875	375	500	0·21	+230	α_h	60
125	125	375	0	0·17	+59	α_h'	9
625	625	125	0	0·17	−143	a_h'	27
125	125	125	0	0·17	+25	α_h'	26
625	875	375	0	0·18	−106	α_h'	38
625	375	125	0	0·18	+191	α_h'	49
625	625	375	0	0·18	−105	a_h'	50
625	875	125	0	0·19	−142	α_h'	41
125	875	125	0	0·19	+21	α_h'	39
125	375	125	0	0·19	−311	α_h'	55
625	375	375	0	0·19	+228	α_h'	50
125	375	375	0	0·20	−276	α_h'	57
125	875	375	0	0·23	−396	α_h'	54
875	875	125	0	0·26	+20	α_h'	85
375	875	125	0	0·21	−60	α_h''	16
375	875	375	0	0·21	−27	α_h''	21
875	375	375	0	0·21	+147	α_h''	18
875	125	125	0	0·22	+114	α_h''	22
875	125	375	0	0·22	+148	α_h''	36
375	125	375	0	0·22	−26	α_h''	25
875	625	125	0	0·23	−234	α_h''	41
875	375	125	0	0·23	+113	α_h''	37
375	125	125	0	0·23	−60	α_h''	37

are determined. With the above phases as input all other reflexions are phased. Each phase $\psi_{\mathbf{h}}$ is determined such that

$$\varphi_{\mathbf{h}}=\frac{\sum_{\mathbf{k}} W_{\mathbf{k}}\,[\cos(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})-C_{\mathbf{k}}]^2}{\sum_{\mathbf{k}} W_{\mathbf{k}}} \qquad (37)$$

is a minimum. The weights $W_{\mathbf{k}}$ are proportional to E_3. Because the $C_{\mathbf{k}}$ may have all values between -1 and $+1$ the resulting phase set is non-centrosymmetric, and may be correct. In fact the $\sum_2$ methods use $C_{\mathbf{k}}$ values which are all equal: $C_{\mathbf{k}}^{\Sigma}=1$. If phases are calculated with equation (37) and $C_{\mathbf{k}}^{\Sigma}=1$ then the resulting phases cannot be very different from those calculated by means of equation (1).

At first sight the *SI* method has the advantage over the $\sum_2$ method and the $\sum_2$ Patterson technique that an image of the structure can be produced by direct phasing only. However there are several objections to the method as it used in practice:

(1) The quality of the $C_{\mathbf{k}}$ values. This is the bottleneck of the method because only if the predicted $C_{\mathbf{k}}$ values and the true values $C_{\mathbf{k}}^t$ have small differences can expression (37) give the true phases. The $C_{\mathbf{k}}$ and $C_{\mathbf{k}}^t$ for the 25 largest structure invariants E_3 of estriol (Hauptman, 1970) are listed in Table 4. There are several large deviations $C_{\mathbf{k}}$–$C_{\mathbf{k}}^t$ among these invariants Moreover it can be seen from the table that
$\sum|C_{\mathbf{k}}^t-C_{\mathbf{k}}|\gg\sum|C_{\mathbf{k}}^t-C_{\mathbf{k}}^{\Sigma}|=\sum|C_{\mathbf{k}}^t-1|$.

Thus in estriol in general $C_{\mathbf{k}}^{\Sigma}=1$ is a better estimate of $C_{\mathbf{k}}^t$ than $C_{\mathbf{k}}$. This implies that if a structure can be solved by the *SI* method, it can be solved by the $\sum_2 P$ method as well. It is even possible that when a structure cannot be solved by the *SI* method, the better estimates of $C_{\mathbf{k}}^{\Sigma}=1$ in the $\sum_2 P$ method may give the solution.

(2) The structure invariant method is a single-solution method and thus in order to make progress in finding new phases it is necessary to assume that an invariant of small E_3 is correct. In practice it is found that the larger E_3 the better its predicted value. Multisolution procedures are therefore favoured, because from the beginning they make use of the largest E_3 values as a basis for the phase determination. However, in this case the price to be paid is that one ends up with a considerable number of solutions, but in most cases consistency criteria or supplementary information (Harker–Kasper inequalities) give enough information about the most probable correct solution. A single-solution method may fail as a result of the incorrectness of an invariant with a relatively small E_3. Multisolution methods fail if one of the highest E_3 invariants is wrong.

(3) In the *SI* method the phase determination with expression (37) is followed by tangent refinement. Thus even if the correct phases $\psi_{\mathbf{h}}$ have been found, the tangent refinement changes the phases towards centrosymmetric values.

Table 4. *Values of the structure invariant* $C_{\mathbf{k}}$ *for estriol*(*Hauptman*, 1970)

The predicted values $C_{\mathbf{k}}$ (*SI* method); the true $C_{\mathbf{k}}^t$ values and the predicted $C_{\mathbf{k}}$ values, as used in $\sum_2$ methods are given

Values given by Hauptman (1970)

A ($\sim E_3$)	$C_{\mathbf{k}}$	True $C_{\mathbf{k}}$: $C_{\mathbf{k}}^t$	Difference $\lvert C_{\mathbf{k}}^t-C_{\mathbf{k}}\rvert$	$C_{\mathbf{k}}$ as used in $\sum_2$ methods $C_{\mathbf{k}}^{\Sigma}$	Difference $\lvert C_{\mathbf{k}}^t-C_{\mathbf{k}}^{\Sigma}\rvert$
13·5	0·9458	0·9888	0·0430	1	0·0112
10·5	1·0000	0·9988	0·0012	1	0·0012
10·2	1·0000	0·9991	0·0009	1	0·0009
9·6	0·8424	1·0000	0·1576	1	0·0000
8·7	1·0000	0·9820	0·0180	1	0·0180
8·2	0·6485	0·9553	0·3068	1	0·0447
8·0	1·0000	0·9838	0·0162	1	0·0162
8·0	0·9629	0·9926	0·0297	1	0·0074
6·7	1·0000	0·9976	0·0024	1	0·0024
6·7	1·0000	0·9737	0·0263	1	0·0263
6·5	1·0000	0·9976	0·0024	1	0·0024
6·4	0·9293	0·9888	0·0595	1	0·0112
6·3	0·8970	0·9759	0·0789	1	0·0241
6·2	1·0000	0·9960	0·0040	1	0·0040
6·1	1·0000	0·9992	0·0008	1	0·0008
6·0	0·9358	0·9998	0·0640	1	0·0002
6·0	0·9975	1·0000	0·0025	1	0·0000
6·0	1·0000	0·9717	0·0283	1	0·0283
5·8	0·7416	0·9664	0·2252	1	0·0336
5·7	1·0000	0·9689	0·0311	1	0·0311
5·7	0·6617	0·9606	0·2989	1	0·0394
5·5	1·0000	0·9928	0·0072	1	0·0072
5·5	0·6876	0·7452	0·0576	1	0·2548
5·5	0·8523	0·9988	0·1465	1	0·0012
5·4	0·8866	0·9916	0·1050	1	0·0084
			Sum 1·7140		Sum 0·5750

(4) Since, in practice, equations (1) and (6) have demonstrated their great value, the following expressions:

$$\varphi_\Sigma = \sum_{\mathbf{k}} W_{\mathbf{k}}[|-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}}| - \arccos(C_{\mathbf{k}})]^2$$

and

$$\varphi_t = \sum W_{\mathbf{k}} \{|\sin \tfrac{1}{2}(-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})| - \sin \tfrac{1}{2}[\arccos(C_{\mathbf{k}})]\}^2$$

may be more useful than (37). These expressions are closely related to the refinement criteria of the $\sum_2$ formula and the tangent refinement.

Objections 2, 3 and 4 can be overcome. However, as long as $C_{\mathbf{k}}^{\Sigma}=1$ is a better estimate of the invariant $C_{\mathbf{k}}^{t}$ than the predicted $C_{\mathbf{k}}$ value of the *SI* method, in our opinion the $\sum_2 P$ method is the stronger one of the two. If the predicted $C_{\mathbf{k}}$ values could be improved such that $\sum|C_{\mathbf{k}}^{t}-C_{\mathbf{k}}|$ is very small and $\sum(C_{\mathbf{k}}^{t}-C_{\mathbf{k}})=0$ then the *SI* method would possess considerable potential since it would be able to predict, for instance, whether in a centrosymmetric structure a $C_{\mathbf{k}}$ value should be $+1$ or -1, implying that $-\psi_{\mathbf{h}}+\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}}$ in equation (3*a*) is equal to 0 or π respectively. A structure in $P\overline{1}$ could then be solved directly, a goal far removed as yet, judging from Table 10 of Hauptman, Fisher & Weeks (1971).

The author is greatly indebted to Dr C. H. Stam for many helpful discussions about the project and the manuscript. I wish to thank Professor Dr C. H. MacGillavry for valuable criticism of the manuscript and Mr E. Hessling for assisting in some of the computational work.

APPENDIX

Weights in tangent refinement

Germain, Main & Woolfson (1971) describe a tangent refinement:

$$\tan \psi_{\mathbf{h}} = \frac{\sum_{\mathbf{k}} W_{\mathbf{k}} W_{\mathbf{h}-\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \sin(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})}{\sum_{\mathbf{k}} W_{\mathbf{k}} W_{\mathbf{h}-\mathbf{k}} |E_{\mathbf{k}} E_{\mathbf{h}-\mathbf{k}}| \cos(\psi_{\mathbf{k}}+\psi_{\mathbf{h}-\mathbf{k}})} = \frac{S_{\mathbf{h}}}{C_{\mathbf{h}}} \qquad (38)$$

where

$$W_{\mathbf{h}} = \tanh[\sigma_3 \sigma_2^{-3/2} |E_{\mathbf{h}}| (S_{\mathbf{h}}^2 + C_{\mathbf{h}}^2)^{1/2}]\,. \qquad (39)$$

This allows the use of all reflexions simultaneously; acceptance and rejection criteria are replaced by the weights. Of course a necessary condition is that a weight $W_{\mathbf{h}}$ is proportional to the accuracy of phase $\psi_{\mathbf{h}}$.

In our opinion the weights in expression (39) do not fulfil this condition. Take, for instance, the values, $A=2\sigma_3\sigma_2^{-3/2} E_3$, given by Hauptman, Fisher & Weeks (1971) for two steroids with the same space group and about equal molecular weight and cell volume; the one structure contains 382 A values between 3·0 and 9·0 and the other 47 between 3·0 and 5·0. In our tangent refinement procedure we try to use more realistic weights in the following way:

$$B_{\mathbf{h}} \ge g \to W_{\mathbf{h}} = B_{\mathbf{h}} / (\sum_{\mathbf{k}} W_{\mathbf{k}} W_{\mathbf{h}-\mathbf{k}} A)$$

$$B_{\mathbf{h}} < g \to W_{\mathbf{h}} = [B_{\mathbf{h}}/g]^p$$

in which

$$B_{\mathbf{h}} = 2\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}|\,(S_{\mathbf{h}}^2 + C_{\mathbf{h}}^2)^{1/2}$$

$$p = 2\tfrac{1}{2}$$

$$\text{and } g = 0{\cdot}8(\text{maximum of } A)\,.$$

Phases which are below the relative reliability level g get a low weight as a result of the exponent p. Phases above this level get a weight proportional to their relative accuracies in the refinement.

It should also possible to insert a varying scale factor L in expression (39):

$$W_{\mathbf{h}} = \tanh\{L\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}|\,(S_{\mathbf{h}}^2 + C_{\mathbf{h}}^2)^{1/2}\}\,. \qquad (40)$$

where L is chosen such that

$$\tanh[L\sigma_3\sigma_2^{-3/2}|E_{\mathbf{h}}|\,.\,\tfrac{1}{2}(\text{maximum of } A)] = 0{\cdot}98\,.$$

List of symbols

$\mathbf{h} \equiv hkl$ $\mathbf{k} \equiv h'k'l'$ $\mathbf{h}-\mathbf{h}' \equiv h-h',\ k-k',\ l-l'$	Indices of reflexions
$\lvert E_{\mathbf{h}} \rvert$	Normalized structure factor
N	Number of atoms in the unit cell
$\psi_{\mathbf{h}}$	Phase of **h**
$\sigma^n = \sum_{j=1}^{N} Z_j^n$	
Z_j^n	Atomic number of j
zop ≡ either 0 or π	
htp ≡ either $\frac{1}{2}\pi$ or $\frac{3}{2}\pi$	

References

Ahmed, F. R. (1970). Editor, *Crystallographic Computing*. Copenhagen: Munksgaard.

Coulter, C. L. & Dewar, R. B. K. (1971). *Acta Cryst.* B**27**, 1730.

De Jong, J. & Schenk, H. (1972). To be published.

Dik, C. (1972). To be published.

Germain, G. & Woolfson, M. M. (1968). *Acta Cryst.* B**24**, 91.

Germain, G., Main, P. & Woolfson, M. M. (1970). *Acta Cryst.* B**26**, 274.

Germain, G., Main, P. & Woolfson, M. M. (1971). *Acta Cryst.* A**27**, 368.

Hall, S. R. (1970). In *Crystallographic Computing*. Edited by F. R. Ahmed. Copenhagen: Munksgaard.

HAUPTMAN, H. (1970). In *Crystallographic Computing*. Edited by F. R. AHMED. Copenhagen: Munksgaard.
HAUPTMAN, H., FISHER, J. & WEEKS, C. M. (1971). *Acta Cryst.* **B27**, 1550.
International Tables for X-ray Crystallography (1952). Vol. I. Birmingham: Kynoch Press.
KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849.
KOCH, M. H. J. & GERMAIN, G. (1970). *Acta Cryst.* **B26**, 410.
SCHENK, H. (1969). *Automation of Symbolic Addition*. Amsterdam: North Holland.
SCHENK, H. (1971). *Acta Cryst.* **B27**, 2037, 2039, 2040.
SCHENK, H. (1972*a*). *Koninkl. Ned. Akad. Wetenschap. Proc.* **B75**. In the press.
SCHENK, H. (1972*b*). To be published.

Acta Cryst. (1972). A**28**, 422

Derivation of van der Waals Radii from Known Crystal Structures

BY G. GAFNER, F. H. HERBSTEIN* AND C. M. LEE

Chemical Physics Group of the National Physical and Chemical Research Laboratories,
South African Council for Scientific and Industrial Research, P.O. Box 395, Pretoria, South Africa

(*Received* 4 *April* 1972)

Methods which can be used for determining the van der Waals diameter constant in interatomic interaction potentials by using known structures of molecular crystals are discussed. It is shown that the usual lattice energy minimization procedure is invalid due to the presence of molecular strain energy. A method based on the equilibration of nearest-neighbour interactions and the internal pressure is developed. Calculations using all available methods have been made for the three crystalline phases of selenium and results are given.

Introduction

Accurate prior knowledge of the interactions between non-bonded atoms is essential if calculations relating to such matters as the magnitude of packing forces in molecular crystals and the relationship between conformation and energy in sterically hindered molecules are to be meaningful. A further important use for non-bonded interactions has been found (Coulson, 1960) in the field of molecular physics where it has been realized that their influence on the length of chemical bonds may be highly significant. Another consequence of the availability of satisfactory interaction potentials would be that trial structure determinations from packing considerations, such as described by Milledge (1962), could be put on a more quantitative (energy) basis for molecular crystals containing molecules of known geometry.

The purpose of this paper is to investigate various methods which might be used for determining the van der Waals diameter constant in an assumed mathematical form of interaction potential from a knowledge of the geometrical arrangement of atoms in a molecular crystal. We will show that lattice energy calculations cannot be handled exactly as the molecular strain energy contribution cannot be calculated. A method based on internal pressure is developed to circumvent the difficulty.

Methods which have been used for estimating interaction potentials

Two basically different approaches have been used by other workers in attempting to derive interatomic interaction potentials for non-bonded atoms. The first approach involves the deduction from first principles of the attractive term in the interaction expression. Examples of this are due to Slater & Kirkwood (1931), Kirkwood (1932) and London (1930). These have been applied to hexachloroethane by Sasada & Atoji (1953) and result in attractive energies with a spread of some 12% about the mean value. All these interactions involve the reciprocal sixth power of the interatomic distance. The repulsive part of the interaction has been derived by Born & Mayer (1932) using a quantum-mechanical treatment. This, however, was not used by Sasada & Atoji (1953) due to the unavailability of certain constants needed in the expression.

Theory thus indicates that a reasonable general expression for non-bonded interactions would have an exponential repulsive part and an inverse sixth power attractive part (considering only dipole–dipole interactions). This is the basis for the existence of various heuristic expressions for the interatomic interaction: the 6-exp deriving directly from theory and the 6–12 interaction of Lennard-Jones (1937) differing from this in the repulsive part alone. Of these two, the latter has the advantage of containing only two constants and as its ability to describe non-bonded interactions is

---

* Present address: Department of Chemistry, Technion – Israel Institute of Technology, Haifa, Israel.