Acta Cryst. (1967). **23**, 339

New Developments in Structure Determination by the Convolution Molecule Method

BY W. HOPPE AND E. F. PAULUS

Abteilung für Röntgenstrukturforschung am Max-Planck-Institut für Eiweiss- und Lederforschung, München, Germany and Physikalisch-Chemisches Institut der Technischen Hochschule München, Abteilung für Strukturforschung, München, Germany

(*Received* 10 *May* 1966 *and in revised form* 30 *December* 1966)

The convolution molecule method interprets a Patterson structure by using the convolution products of molecules or molecular fragments. Essential in the procedure is a separation of the orientational and translational variables, making use of the symmetry relations in and between the convolution molecules. An extension of the method is described which is convenient if the asymmetric unit contains several molecules or molecular fragments of known steric structures. In the new procedure, the determination of the structure of the asymmetric unit ('supermolecule' structure) is separated from the determination of the structure of the unit cell. Examples are given. The influence of constraints is discussed. Possible applications in protein crystal structure determination are mentioned.

Introduction

The convolution molecule method (Hoppe, 1957) was first successfully used in two-dimensional projections with graphically constructed convolution molecules (Hoppe & Will, 1960; Hoppe & Rauch, 1961). The method was later programmed for three-dimensional application in computers. The program scans the position of the convolution molecules in the Patterson function (Huber, 1965). Successful applications of this version have been reported (Huber & Hoppe, 1965; Deuschl, 1965; Hoppe, 1966; Hoppe, Gieren, Hädicke, Brodherr, Röhrl & Huder, 1967).

An interesting extension of the method is possible when the asymmetric unit contains several unsymmetrically distributed molecules or molecular fragments with known steric conformation. The structure determination can then be divided into two steps: The first step concerns the determination of the orientational and translational parameters of these molecules in the asymmetric unit, the second step concerns the determination of the translational parameter of the asymmetric unit.

The procedure described in this paper was first used in the structure analysis of the metal-organic compound μ-cyclooctatetraen-biscyclopentadienylcobalt(I) $C_8H_8(CoC_5H_5)_2$ (Paulus, 1965). The unit cell ($P\bar{1}$) contains three molecules in the asymmetric unit. Neglecting the light atoms, the structure is formed from 2 Co atoms in the geometric configuration of a diatomic molecule (Fig. 1). It was possible to estimate the Co–Co distance from the supposed double sandwich structure of this compound as 3·5–4 Å. An inspection of the inner part of the three-dimensional Patterson function showed high peaks in this region, thus confirming the estimate. The structure determination of this compound did therefore start with a convolution molecule determination of the orientation and positions of the six 'Co_2-dumb-bells' in the unit cell. The positions of the light atoms could then be found by the usual Fourier refinement.

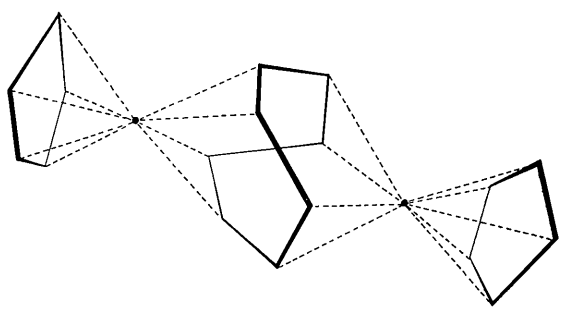

Fig. 1. Molecular structure of $C_8H_8(CoC_5H_5)_2$. The black circles represent the two Co atoms in the molecule.

Rossmann & Blow (1962, 1963, 1964) have published a method which concerns also structures with several molecules in the asymmetric unit. To some extent, this method is also an extension of the convolution molecule method, as it works with convolution molecules, at least in principle (they are named *self vectors* and *cross vectors* of molecular structures by the authors). The basic difference is that the convolution molecules are unknown. The authors use the Patterson structure itself as a replacement of the convolution molecules. This is possible, because every Patterson structure contains all convolution molecules. Every convolution molecule can therefore be brought into coincidence with another convolution molecule of the same Patterson function by rotations and transformations. In the case of several non-identical molecules in the asymmetric unit non-trivial solutions occur. It is an important advantage of this method that nothing of the structure needs to be known. On the other hand the simultaneous coexistence of all convolution molecules during the process of interpretation produces unwanted background noise. Therefore, the scanning with selected convolution molecules is more advantageous if details of the molecular structure are known.

Method of evaluation

Fig. 2 shows schematically three diatomic molecules:

$$\varrho_1(\mathbf{r}_1),\ \varrho_2(\mathbf{r}_2),\ \varrho_3(\mathbf{r}_3)\ .\ \mathbf{r}_1,\ \mathbf{r}_2,\ \mathbf{r}_3\ .$$

The translation vectors of these three molecules with reference to an arbitrarily chosen origin O' are r_1, r_2, r_3. The vector from the origin O of the unit cell to O' is r_0. The symmetry centre in the origin O duplicates

$$\varrho_1, \varrho_2, \varrho_3 \text{ to } \varrho_1^*, \varrho_2^*, \varrho_3^* .$$

Let us now first consider the three molecules $\varrho_1, \varrho_2, \varrho_3$ as a six-atom supermolecule† $\varrho_0(\mathbf{r}_0)$. The Patterson structure can then be represented by the convolution molecule structure:

$$\widehat{\varrho\varrho}^* = 2\widehat{\varrho_0\varrho_0^*} + \widehat{\varrho_0\varrho_0}(2\mathbf{r}_0) + \widehat{\varrho_0^*\varrho_0^*}(-2\mathbf{r}_0) . \qquad (1)$$

The first term in this sum represents the 'identically subscripted convolution molecule' $\varrho_0\varrho_0^*$ with double weight.‡ This convolution molecule is independent of the translation vector $\mathbf{r}_0$. The two other identically subscripted convolution molecules $\varrho_0\varrho_0$ and $\varrho_0^*\varrho_0^*$ are displaced in the Patterson structure by the translational parameter $2\,\mathbf{r}_0$ and $-2\,\mathbf{r}_0$. We now replace the supermolecule ϱ_0 by the three diatomic molecules:

$$\varrho_0 = \varrho_1(\mathbf{r}_1) + \varrho_2(\mathbf{r}_2) + \varrho_3(\mathbf{r}_3) . \qquad (2)$$

$\widehat{\varrho_0\varrho_0^*}$ can now be replaced by a sum of convolution molecules.

$$\begin{aligned}\widehat{\varrho_0\varrho_0^*} = {} & \widehat{\varrho_1\varrho_1^*} + \widehat{\varrho_2\varrho_2^*} + \widehat{\varrho_3\varrho_3^*} \\ & + \widehat{\varrho_1\varrho_2^*}(\mathbf{r}_1-\mathbf{r}_2) + \widehat{\varrho_1\varrho_3^*}(\mathbf{r}_1-\mathbf{r}_3) \\ & + \widehat{\varrho_2\varrho_3^*}(\mathbf{r}_2-\mathbf{r}_3) + \widehat{\varrho_2\varrho_1^*}(\mathbf{r}_2-\mathbf{r}_1) \\ & + \widehat{\varrho_3\varrho_1^*}(\mathbf{r}_3-\mathbf{r}_1) + \widehat{\varrho_3\varrho_2^*}(\mathbf{r}_3-\mathbf{r}_2) . \qquad (3)\end{aligned}$$

The convolution molecules (3) have double weights in the Patterson structure $\widehat{\varrho\varrho}^*$ because they represent the first (double-weighted) term in (1). They can therefore be used to find the structure of the six-atom supermolecule by the convolution molecule technique. The interpretation can therefore be divided into the following steps:

1. Determination of the orientations of the three molecules $\varrho_1, \varrho_2, \varrho_3$ from the first three terms in (3).

2. Formation of the six *m*-convolution molecules† representing the remaining terms in (3) by graphical construction or (better) by Fourier calculations.

3. Determination of the translation vectors $\mathbf{r}_1$—$\mathbf{r}_2$, $\mathbf{r}_1-\mathbf{r}_3$, $\mathbf{r}_2-\mathbf{r}_3$ by the usual convolution molecule procedure using the double weight criterion for the selection of the translation vectors.‡

4. Construction of the supermolecule ϱ_0 and arbitrary choice of an origin O'.

5. Repetition of the convolution molecule operation with convolution molecules of the supermolecules using relation (1). In procedures (1)–(4) not only the structure but also the orientation of the supermolecule ϱ_0 has been found. It is therefore only necessary to determine the translation vector $\mathbf{r}_0$.

Fig. 3 shows the difference Fourier synthesis (light atoms subtracted) in the structure of $C_8H_8(CoC_5H_5)_2$ in order to demonstrate the positions of the three 'Co_2-dumb-bells' which have been determined previously by the procedure described above. It is not necessary for the different 'molecules' (or molecular fragments) contained in the 'supermolecules' to be of the

† The expression *supermolecule* is used in a purely geometrical way. It does not imply bonds between the three molecules.

‡ The expression *convolution molecule* was introduced in our first paper (Hoppe, 1957). It means the convolution product of two molecular structures. The molecular structure can be a molecule (or a molecular fragment) in the physical sense of the word. But it can also be a geometrical configuration of partly unbonded atoms, ions and (or) molecules. Note that *identically subscripted convolution molecules* (which we abbreviate to *i-convolution molecules*) have a structure which is independent of the orientation of the molecules in the unit cell, whereas the structures of mixed subscripted convolution molecules (*m-convolution molecules*) vary with the orientation in the unit cell. The *i*-convolution molecules $\varrho_i\varrho_i^*$ are the Patterson structures of the molecular structures. They are not only independent of the orientation in the unit cell but also of the translations of the molecular structures. They will be used in the convolution molecule method for the determination of the molecular translations.

† Note that only three convolution molecules are different in structure. The important symmetry relations between convolution molecules have been discussed in our earlier papers. General symmetry considerations show for example that a structure containing *n* molecules in the unit cell and one molecule in the asymmetric unit has only *n* different convolution molecules in the Patterson structure. Symmetry relations generate from these *n* convolution molecules a structure of n^2 convolution molecules.

‡ The automated version of the convolution molecule method (Huber, 1965) can be used in the present form because it registers separately the positive and negative differences between Patterson structure and convolution molecules. The double-weighted convolution molecules can therefore only be subtracted with positive differences when they coincide with double-weighted Patterson peaks. It is possible to translate operations with convolution molecules into reciprocal space in a similar way to operations with the image-seeking functions. It should, however, be mentioned that criteria for the fit which can easily be translated (*e.g.* the sum or the product between convolution molecule and the Patterson structure) cannot be used for differentiation between convolution molecules of different weights.

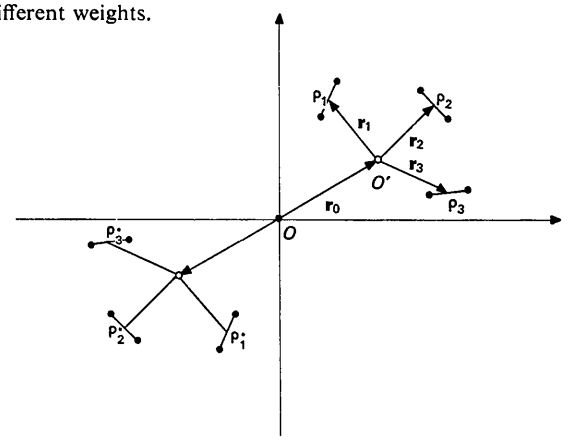

Fig. 2. Schematic representation of a structure containing three diatomic molecules in the asymmetric unit (space group $P\bar{1}$). The three molecules form a 'supermolecule' ϱ_0 with reference to an arbitrarily chosen origin O'.

same kind. In the example, the molecular fragment was a simple heavy atom structure. In such a case it is probable that conventional image seeking procedure would also lead to the correct heavy atom structure in the unit cell. The method is therefore of special interest where the molecular fragments are of complicated structure, *e.g.* where they consist only of light atoms. It may however be mentioned that even in the simple case described above, the convolution molecule method has some distinct advantages compared with the image seeking technique:

1. The known chemical information (formation of three diatomic Co_2 structures with known distance) will be used.

2. The method starts from doubly weighted Patterson peaks, whereas the image seeking method has to start from a singly weighted Patterson peak.

It should be mentioned that in some cases multiple solutions in the first steps of the determination of the supermolecule structure may occur. The ambiguities can be solved in the following steps. The correct solution in our example explains *e.g.* all high peaks in the Patterson function.

Generalization to other space groups

It is quite clear from the description in the preceding chapter that the first step of the method is the interpretation of the Patterson structure of the supermolecule (the geometrical configuration of the molecules or molecular fragments in the asymmetric unit). This structure is independent of the translation vectors which combine the asymmetric units to the unit cell. The convolution molecule structure of every space group must contain the Patterson structures of the supermolecules. Therefore, the determination of the structure of the supermolecule is possible as a first step of structure determination in every space group. It must only be taken into account that in general the origin of the convolution molecule structure is occupied by several Patterson structures of the supermolecules. (There are, *e.g.*, in $P2_12_12_1$ four i-convolution molecules (Patterson structures) related by mirror planes perpendicular to the crystal axes.) This means that the Patterson structure should be scanned simultaneously with the symmetrically related convolution molecules of the submolecules. It is of general interest that the 'double-weight criterion' is not valid in all space groups. It requires that two supermolecules in the unit cell have orientations which are identical or related by a centre of symmetry. The last condition is valid in all centrosymmetric space groups. It is advisable to discuss the symmetry properties of the space group in question in terms of the 'convolution molecule point positions' which we introduced in our first papers (Hoppe, 1957; Hoppe & Will, 1960; Hoppe & Rauch, 1961). Two examples will be given:

1. *Space group* $P2$

$$\varrho = \varrho_1(x_0, y_0, z_0) + \varrho_2(\bar{x}_0, y_0, \bar{z}_0)$$

$$\widehat{\varrho\varrho}^* = \widehat{\varrho_1\varrho_1}^* + \widehat{\varrho_2\varrho_2}^* + \widehat{\varrho_1\varrho_2}^*(2x_0, 0, 2z_0) + \widehat{\varrho_2\varrho_1}^*(2\bar{x}_0, 0, 2\bar{z}_0)\,.$$

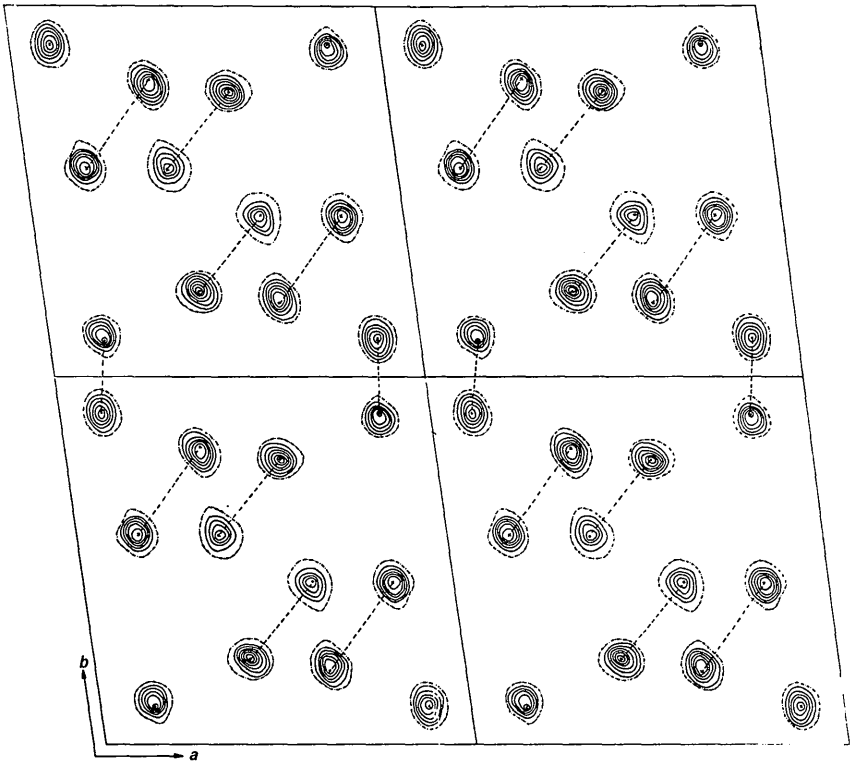

Fig. 3. Two-dimensional difference Fourier synthesis of the structure of $C_8H_8(CoC_5H_5)_2$ (light atoms subtracted), showing the configuration of the three Co_2-dumb-bells (c projection).

The space group has no centre of symmetry. The Patterson structure consists only of single-weighted convolution molecules. The first two terms in the sum represent the translation independent i-convolution molecules (Patterson structures) of the asymmetric unit. After the configuration of this unit has been found, the translation parameter $2x_0$, $2z_0$ will be determined.

2. *Space group* $P2/m$

$$\varrho = \varrho_1(x_0, y_0, z_0) + \varrho_2(\bar{x}_0, y_0, \bar{z}_0) + \varrho_1^*(\bar{x}_0, \bar{y}_0, \bar{z}_0) + \varrho_2^*(x_0, \bar{y}_0, z_0)$$

$$\begin{aligned}\widehat{\varrho\varrho}^* = {} & 2\widehat{\varrho_1\varrho_1^*} + 2\widehat{\varrho_2\varrho_2^*} \\ & + 2\widehat{\varrho_1\varrho_2}(0, 2y_0, 0) + 2\widehat{\varrho_1^*\varrho_2^*}(0, 2\bar{y}_0, 0) \\ & + 2\widehat{\varrho_1\varrho_2^*}(2x_0, 0, 2z_0) + 2\widehat{\varrho_2\varrho_1^*}(2\bar{x}_0, 0, 2\bar{z}_0) \\ & + \widehat{\varrho_1\varrho_1}(2x_0, 2y_0, 2z_0) + \widehat{\varrho_1^*\varrho_1^*}(2\bar{x}_0, 2\bar{y}_0, 2\bar{z}_0) \\ & + \widehat{\varrho_2\varrho_2}(2\bar{x}_0, 2y_0, 2\bar{z}_0) + \widehat{\varrho_2^*\varrho_2^*}(2x_0, 2\bar{y}_0, 2z_0) .\end{aligned}$$

The space group has a centre of symmetry. The first two terms of the sum represent the same two i-convolution molecules (mirror plane perpendicular to y) as in the first example, but with double weight. The example shows that further doubly weighted convolution molecules can occur in special convolution molecule point positions (terms 3–4 in the sum). Therefore, in spite of the double-weight criterion, the differentiation is not as sharp as in the space group $P\bar{1}$.

Constraints in the asymmetric unit

The number of parameters can be reduced if there are constraints in the formation of the supermolecule. An example is shown in Fig. 4. The 'diphenylmethane' molecule contains the two benzene rings ϱ_1 and ϱ_2. It is convenient to use O' as the origin of both benzene rings. In this case a determination of the translation parameters $\mathbf{r}_1$, $\mathbf{r}_2$ is not necessary; both vectors are equal to zero owing to the choice of the origin O'. The determination of the structure of the supermolecule will therefore be restricted to the determination of the orientation of the two benzene rings. The m-convolution molecules of the submolecules can be used to check the orientations found. It can be seen from the Figure that there are also correlations in the orientation of the benzene rings. The simplest way to deal with them is to determine all possible orientations of the two benzene rings and to neglect all combinations which are impossible for chemical reasons. In general, constraints can often be introduced into the procedure in the same way as in the rigid body refinement techniques (Scheringer, 1965). Constraints of a more complicated nature can exist if the convolution molecule method is used for the determination of the quaternary structure of those proteins which are oligomers consisting of several identical subunits. If the structure of a subunit is known (*e.g.* from a crystal structure determination of the monomer), the described molecule convolution procedure can be used for the determination of the structure of the oligomer. Sometimes, some features of the oligomer structure might be known (formation of a ring of monomers, formation of tetrahedral, octahedral or other regular conformations). In these cases, the remaining unknowns can be described with reference to the known structural principle; such unknowns are *e.g.* the radius of the ring, the angular orientations of the ring, the orientation of a monomer in the ring *etc.* It might even often be possible to determine the general structural principle by electron microscopy (images of negatively stained molecules of the oligomer). Therefore, a quite complicated structure in the difference Patterson synthesis might occur. In the interpretation of this structure, the convolution molecule concept can be used for the heavy atom configurations in the monomer; *e.g.* in the case of double derivatives, the difference Patterson structure can be discussed in a way similar (diatomic configuration) to that in our first example.

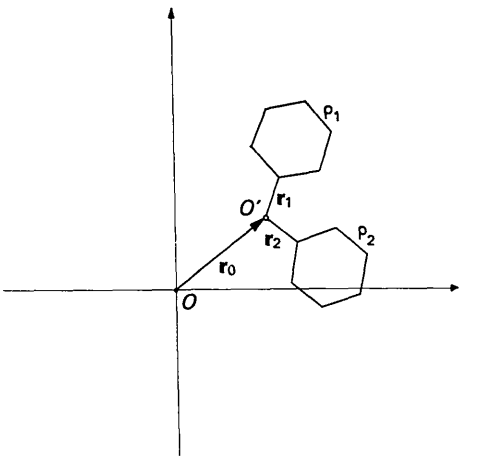

Fig. 4. 'Diphenylmethane'-molecule.

References

DEUSCHL, H. (1965). *Ber. Bunsenges. physikal. Chem.* **69**, 550.

HOPPE, W. (1957). *Z. Elektrochem.* **61**, 1076.

HOPPE, W. (1966). *Angew. Chem.* **78**, 289.

HOPPE, W., GIEREN, A., HÄDICKE, E., BRODHERR, N., RÖHRL, M. & HUBER, R. (1967). Vortrag auf der 9. Diskussionstagung der Sektion für Kristallkunde der Deutschen Mineralogischen Gesellschaft.

HOPPE, W. & RAUCH, R. (1961). *Z. Kristallogr.* **115**, 141.

HOPPE, W. & WILL, G. (1960). *Z. Kristallogr.* **113**, 104.

HUBER, R. (1965). *Acta Cryst.* **19**, 353.

HUBER, R. & HOPPE, W. (1965). *Chem. Ber.* **98**, 2403.

PAULUS, E. F. (1965). Dissertation, Univ. München.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24.

ROSSMANN, M. G. & BLOW, D. M. (1963). *Acta Cryst.* **16**, 39.

ROSSMANN, M. G. & BLOW, D. M. (1964). *Acta Cryst.* **17**, 338.

SCHERINGER, C. (1965). *Acta Cryst.* **19**, 513.