

The Application of Direct Methods to Non-Centrosymmetric Structures Containing Heavy Atoms

BY TH. E. M. VAN DEN HARK, PETER PRICK AND PAUL T. BEURSKENS

Crystallography Laboratory, Toernooiveld, Nijmegen, The Netherlands

(Received 26 January 1976; accepted 1 April 1976)

Direct methods are applied to the difference structure factors for a structure containing one or more heavy atoms in known positions. The present procedure is initiated by subtracting the known heavy-atom contribution from the observed structure factor (assuming that the observed and calculated structure factors have the same phase) to obtain the magnitude and phase of the light-atom contribution. The Σ_2 phase relationship (tangent formula) is used to recalculate the phases of the light-atom contributions, and – consequently – to recalculate the magnitude of the light-atom contribution. An iterative procedure is used to optimize the phases and amplitudes before a difference Fourier map is calculated. The method is applicable also for the solution of partially known structures.

Introduction

In paper I (Beurskens & Noordik, 1971) and in paper II (Gould, van den Hark & Beurskens, 1975) direct methods were used to solve the phase problem, or to speed up the solution, for centrosymmetric structures containing one or more heavy atoms on known positions. The present paper deals with the non-centrosymmetric case. A procedure to handle the special case in which the known heavy atoms do not completely fix the origin and/or enantiomorph is under investigation. The general case* is discussed in this paper.

The positions of the known (heavy) atoms determine the structure; the phase problem is solved in principle: the positions of the remaining (light) atoms can be found by standard Patterson or Fourier techniques. The time and effort necessary for finding the light-atom structure is reduced by the present procedure, especially when the heavy atoms are only marginally sufficient to solve the phase problem.

Define for a reflexion h :

- $|F_{\text{obs}}|$ observed structure factor amplitude, on absolute scale.
- F_H calculated contribution of the known part of the structure ('heavy' atoms).
- φ_H phase of F_H .
- F_L contribution of the remaining part of the structure ('light' atoms), or: the most probable estimate for this contribution.
- φ_L phase of F_L .
- F_{obs} a phased value for the observed structure amplitude, defined by:

$$F_{\text{obs}} = F_L + F_H \quad (1)$$

Possible solutions to this equation, for given $|F_{\text{obs}}|$, φ_L and F_H values will be discussed below.

* A Fortran program for the execution of this procedure, *DIRDIF.D*, is available on request.

In conventional procedures, the difference Fourier coefficients,

$$\Delta F_1 = (|F_{\text{obs}}| - |F_H|) \exp i\varphi_H, \quad (2)$$

are calculated and accepted as an estimate for F_L . Only in very favourable circumstances can the complete light-atom structure be unambiguously deduced from the Fourier synthesis based on these coefficients. In the present procedure, a Σ_2 (tangent formula) refinement procedure is used to convert input ΔF_1 values to more probable F_L values. This procedure depends on a probability estimate for ΔF_1 , relative to the extreme opposite possibility, ΔF_2 ,

$$\Delta F_2 = (-|F_{\text{obs}}| - |F_H|) \exp i\varphi_H, \quad (3)$$

where now F_{obs} is completely out of phase with F_H . Note, that

$$|\Delta F_1| \leq |F_L| \leq |\Delta F_2|.$$

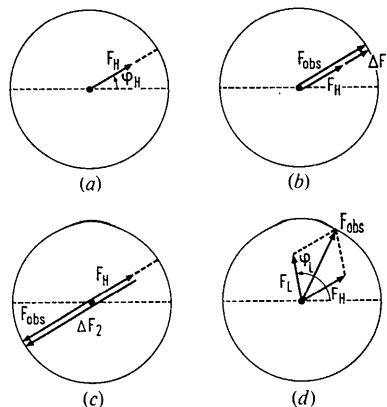


Fig. 1. Definition of ΔF_1 and ΔF_2 , and construction of F_L for case 1: $|F_H| < |F_{\text{obs}}|$. (a) $|F_{\text{obs}}|$ circle, with calculated F_H . (b) Definition of ΔF_1 ; F_H and F_{obs} are in phase. (c) Definition of ΔF_2 ; F_H and F_{obs} out of phase by 180° . (d) General case; φ_L is assumed to be known.

Although conventionally ΔF_1 is used as difference Fourier coefficient, ΔF_2 will be the more probable value for relatively small terms (see below).

In the present procedure use is made of those reflexions where ΔF_1 is far more probable than ΔF_2 , and

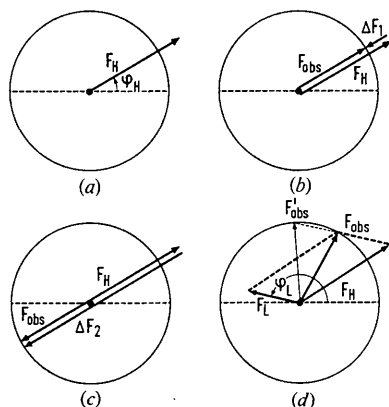


Fig. 2. Definition of ΔF_1 and ΔF_2 , and construction of F_L for case 2: $|F_H| > |F_{obs}|$. (a)–(d) as in Fig. 1.

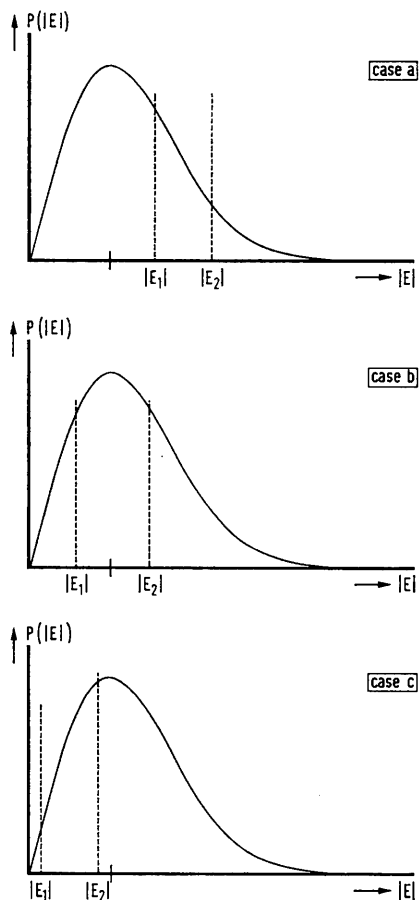


Fig. 3. Classification of reflexions depending on $|E_1|$ and $|E_2|$ pairs. Case (a): $|E_2| > |E_1| > 0.7$. Case (b): $|E_1| < 0.7 < |E_2|$. Case (c): $|E_1| < |E_2| < 0.7$.

ΔF_1 is used as a first estimate of F_L . The application of direct methods then leads to new phases φ_L , and the magnitude $|F_L|$ also is to be recalculated, using (1), which may be written as

$$|F_{obs}| = |F_L| \exp i\varphi_L + F_H \quad (4)$$

(see Figs. 1 and 2, to be discussed later).

Normalization of the difference-structure factors

A modified, two-dimensional Wilson plot, as described in papers I and II, is used to obtain the scale factor K , the overall temperature factor of the heavy atoms B_H , and the overall temperature factor of the light atoms B_L . K is used to bring $|F_{obs}|$ onto an absolute scale. B_H is used in the calculation of the heavy-atom contributions F_H . B_L is used to calculate the normalizing function $g(\mathbf{h})$:

$$g = (\varepsilon_{\mathbf{h}} \sum_{j=1}^{n_L} f_j^2)^{1/2} \exp -B_L \sin^2 \theta / \lambda^2 \quad (5)$$

where n_L is the number of unknown (light) atoms in the unit cell and $\varepsilon_{\mathbf{h}}$ is the usual factor to correct for symmetry enhancement for the reflexion \mathbf{h} . The normalized difference structure factor is defined as:

$$E_L = (F_{obs} - F_H) / g \quad (6)$$

(for a given reflexion \mathbf{h} , if the phase of F_{obs} is known). The above-defined ΔF_1 and ΔF_2 values are brought onto the same scale by defining:

$$E_1 = \Delta F_1 / g \quad \text{and} \quad E_2 = \Delta F_2 / g. \quad (7)$$

The (initially unknown) E_L values correspond to the normalized structure factors of a structure consisting of only the unknown (light) atoms. Conventional direct methods may, in principle, be applied on E_L values, in cases where they are known. For reflexions where E_1 is far more probable than E_2 , we use E_1 as an initial estimate for E_L and use these values to initiate a tangent-formula refinement procedure.

Note: the original $|E_1|$ values do not form a normalized set of structure factors; the average of the squares of $|E_1|$ is less than one. During the refinement of phases, the magnitudes of $|E_L|$ are being increased, on average, and the distribution of $|E_L|$ values becomes more normalized.

Probability considerations

The distribution function for non-centrosymmetric reflexions is given by

$$P(|E|) = 2|E| \exp -|E|^2. \quad (8)$$

This function has a maximum at $|E| = \frac{1}{2}\sqrt{2} \approx 0.7$. The occurrence of this maximum forced us to consider three distinct cases (Fig. 3). [*Note:* special reflexions, having a centrosymmetric distribution, will be treated accordingly; see paper II; we now limit our discussion to general (non-centrosymmetric) reflexions].

For case (a): $|E_1| > 0.7$. Many reflexions will have both $|E_1|$ and $|E_2|$ greater than 0.7. The number of reflexions that belong to this category depends on the known fraction of the scattering power; usually it is about half the total number of reflexions, or fewer.

This is the most important case, as $|E_1|$ (as well as any possible value for $|E_2|$) is large enough to be of importance in a Fourier synthesis and to be useful in the application of the tangent formula. For these reflexions $|E_1|$ is more probable than $|E_2|$, and E_1 may be selected and used as a first estimate for E_L in cases where the probability for $|E_1|$ is significantly greater than the probability for $|E_2|$. Analogous to the centrosymmetric procedure (Paper II) we now use P_1 as a measure for the relative probability of the phase of E_1 :

$$P_1 = P(|E_1|) / [P(|E_1|) + P(|E_2|)], \quad (9)$$

and the weight for this reflexion is chosen to be

$$W_1 = (2P_1 - 1)^2. \quad (10)$$

In principle this is not correct; the weight should be based upon the standard deviation σ_1 for the phase of E_1 , which can be calculated from equation (8) as

$$\sigma_1^2 = \int_1^2 P(|E|) [d\varphi]^2 d|E| / \int_1^2 P(|E|) d|E| \quad (11)$$

where the integration limits are $|E_1|$ and $|E_2|$, and:

$$\cos \Delta\varphi = (|E_1| \cdot |E_2| - |E|^2) / (|E_2| \cdot |E| - |E_1| \cdot |E|).$$

Numerical results are given in Table 1. (These results should also be used in the treatment of the weaker reflexions.)

In practice, however, equation (9) gives acceptable results for the first input phases for the tangent refinement procedure, and the weights for the following cycles of the refinement will be largely determined by the results of the tangent formula.

For case (b) $|E_1| < 0.7 < |E_2|$, and for case (c) $|E_2| < 0.7$, the value of $|E_1|$ is *not* the most probable value. The tangent formula cannot affect the phase of such a reflexion; nevertheless E_1 or E_2 of the reflexion may be used as Fourier coefficient if this is justified by its relative probability (9) or, better, its standard deviation (11). In cases where $|E_1|$ is very small (equal to or less than its estimated error) the reflexion is excluded from further calculations.

Tangent refinement procedure

The case-(a) reflexions may enter into the tangent refinement procedure. As input to the tangent formula we use those reflexions where $|E_1|$ exceeds a given minimum value greater than 0.7 (say $|E_1| > 1.2$). We then use this formula to calculate phases for all reflexions where $|E_1|$ exceeds another minimum greater than 0.7 (say: $|E_1| > 0.9$).

The tangent formula may be given as:

$$\varphi_L(t) = \text{phase of } (E_L)_n \simeq \text{phase of } \Sigma_2 \quad (12)$$

with $\Sigma_2 = \sum_k W_k W_{h-k} E_k E_{h-k}$ and E_k is the most probable E_L value for the reflexion \mathbf{k} . Analogous to the centrosymmetric formula, and to the formulae (9) and (10), we use the following simple expression for the relative probability and the corresponding weight:

$$P_t = \frac{1}{2} + \frac{1}{2} \tanh \sum_j Z_j^3 (\sum_j Z_j^2)^{-3/2} |E_n| \Sigma_2 \quad (13)$$

$$W_t = (2P_t - 1)^2 \quad (14)$$

where Z_j is the number of electrons for the j th atom. [For the weighted tangent formula, see Germain, Main & Woofson (1971).]

In the first cycle we only have E_1 values (with phases φ_1) as first estimates for E_L for the reflexions \mathbf{k} and $(\mathbf{h}-\mathbf{k})$. Whether or not the output phases φ_L are accepted depends on the corresponding weights. If $W_t > W_1$ then the new phase φ_L is accepted with weight W_t . If $W_t < W_1$ then the calculated φ_L value is only partially accepted where $|\varphi_L(t) - \varphi_1|$ is less than 90° :

$$\varphi_L(\text{new}) = \varphi_1 + (W_t/W_1)[\varphi_L(t) - \varphi_1]. \quad (15)$$

In addition, there is a limitation on $|\varphi_L(t) - \varphi_1|$ for reflexions of category (a2), see below, and φ_L may be reset accordingly. The new φ_L value is used to calculate a new value for E_L , equation (4), which may be used as input for the next tangent refinement cycle.

Note: As a consequence of the definitions of W_1 and W_t [given in equations (10) and (14), which on average result in a slightly overestimated W_1 relative to W_t] the original phases φ_1 still play an important role in the second refinement cycle. Using a weighting scheme based upon equation (11) calls for the replacement of

Table 1. Standard deviations ($^\circ$) of φ_1 for pairs of $|E_1|$ and $|E_2|$ values, calculated from equation (11)

$ E_2 =$	0.3	0.6	0.8	1.0	1.2	1.4	1.7	2.0	2.5	3.0	4.0
$ E_1 = 0.3$	104	109	111	111	110	108	104	99	92	88	82
$ E_1 = 0.6$		104	101	102	101	99	94	89	81	75	69
		$ E_1 = 0.8$	104	99	98	96	91	85	76	70	63
			$ E_1 = 1.0$	104	97	95	90	83	73	66	58
				$ E_1 = 1.2$	104	96	90	83	71	63	54
					$ E_1 = 1.4$	104	92	85	71	61	52
						$ E_1 = 1.7$	104	90	74	61	49
							$ E_1 = 2.0$	104	80	63	47
								$ E_1 = 2.5$	104	75	47
									$ E_1 = 3.0$	104	51
										$ E_1 = 4.0$	104

the conventional scheme (14) by a scheme that is based upon the standard deviations associated with the tangent formula.

Description of the procedure

We have assumed that the heavy-atom part of the structure is known, and that its contribution to the structure factors is calculated. The normalization procedure, described above, will then lead to values for $|F_{\text{obs}}|$, F_H , ΔF_1 , ΔF_2 , E_1 , E_2 , P_1 and W_1 .

The distinction between several different cases, as given above, leads to the following categories of reflexions: (a1), (a2), (b1), (b2), (c1) and (c2). A detailed description for each of these categories is given below.

At the end the final E_L values are transformed back to F_L values and used in a weighted Fourier synthesis.

Category (a1)

For reflexions with $|F_H| < |F_{\text{obs}}|$ we have (see Fig. 1)

$$\varphi_1 = \varphi_H = \varphi_2 + 180^\circ. \quad (16)$$

The strongest of these form the category (a1). This is the most important set of reflexions. In principle all of these reflexions may be treated alike; in practice, however, a considerable amount of computer time is saved by limiting the number of reflexions that enter into the tangent refinement procedure. These reflexions will have, after four to six cycles of tangent refinement, a calculated E_L value. If this calculated value is unreliable, then the original E_1 value will be used according to formula (7).

The remaining reflexions are those with $0.7 < |E_1| < E_{\text{min}}$ (= say 0.9); for these reflexions the E_1 value will be used, with its proper weight W_1 .

Category (a2)

For reflexions with $|F_H| > |F_{\text{obs}}|$ we have (see Fig. 2):

$$\varphi_1 = \varphi_2 = \varphi_H + 180^\circ. \quad (17)$$

As can be seen from the diagram in Fig. (2d), there are two possible $|E_L|$ values for a given phase φ_L ; the smallest $|E_L|$ value is the most probable one and, naturally, this is the value to be used in our procedure. For a given $|F_{\text{obs}}|$ and F_H there is a restriction on the possible phase values φ_L , see Fig. 4.

Let us define:

$$\Delta\varphi = \varphi_L - \varphi_1, \quad (18)$$

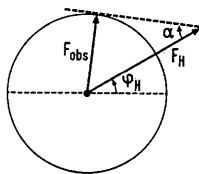


Fig. 4. Definition of the limiting angle for reflexions with $|F_H| > |F_{\text{obs}}|$; $\alpha = \Delta\varphi_{\text{max}}$ [equation (19)].

that is: $\Delta\varphi$ is the correction to the original φ_1 value; we hope to find $\Delta\varphi$ by the weighted tangent procedure.

The maximum value for $|\Delta\varphi|$ is given by

$$\sin |\Delta\varphi_{\text{max}}| = |F_{\text{obs}}|/|F_H|. \quad (19)$$

For $|E_1| > 0.7$ and (most likely) $|E_2| < 4.0$ we have $\Delta\varphi_{\text{max}} = 44.6^\circ$ as the largest possible value for $\Delta\varphi_{\text{max}}$; so the few reflexions in this category have well-determined phases and will therefore be given unit weight ($W_1 = 1$) in all calculations. If the $\Delta\varphi$, calculated by the tangent formula (12), exceeds $\Delta\varphi_{\text{max}}$ for the given reflexion, then $\Delta\varphi_{\text{max}}$ is substituted for $\Delta\varphi$. In the last cycle, however, the calculated $\Delta\varphi$ is accepted to allow for possible experimental errors in $|F_{\text{obs}}|$ and model errors in F_H .

Category (b1)

Because of the low $|E_1|$ value the tangent formula may lead to incorrect results, and a change in phase, leading to larger $|E_L|$ values cannot be trusted; so the tangent formula is not used at all. Although the reflexions in this category have rather low $|E_1|$ values, this category is not unimportant because of the large number of reflexions. E_1 values will be used, with the proper weight W_1 , only in case $W_1 > W_2$. Reflexions with $W_1 < W_2$ are rejected for several reasons [see also category (c1)].

Category (b2)

The reflexions belonging to this category have reliable phases φ_1 [$W_1 = 1$, see category (a2)] and are therefore useful for the Fourier synthesis; φ_L is taken as φ_1 , and the absolute value $|E_L|$ is taken as its expectation value:

$$\int_1^2 |E| P(|E|) d|E| / \int_1^2 P(|E|) d|E|. \quad (20)$$

Category (c1)

For these reflexions E_2 is a more probable estimate for E_L than E_1 (in contrast to the conventional definition of difference Fourier coefficients). It is certainly *not* useful to put these reflexions as E_1 into a Fourier synthesis. To use E_2 as Fourier coefficients on the other hand, may easily lead to an increased Fourier noise level because of the uncertainties in $|F_{\text{obs}}|$ and F_H . At present we reject these reflexions, but we will change our computer program as soon as our experiments show the usefulness of these reflexions.

Category (c2)

This small set of reflexions is treated like category (b2).

Examples

The procedure has been used successfully in the structure analysis of $(\text{HgI}_2)_2\text{CH}_2(\text{S}_2\text{CNC}_4\text{H}_{10})_2$, and of two modifications of heptahelicene, $\text{C}_{30}\text{H}_{18}$. In the former structure one HgI_2 unit was found from the Patterson

synthesis, and we assumed the space group to be Cc ; application of the program *DIRDIF.D* revealed the unexpected presence of a second HgI_2 unit (Beurskens, Enckevort, Menger & Bosman, 1976). The application of *DIRDIF.D* on the heptahelicene structures, where a fragment of the structure was found by direct methods, will be described in the following paper (Beurskens, van den Hark & Beurskens, 1976).

Some numerical results for a test structure will be given in Tables 2-4 for 4,4-dichloro-2a-aza-*A*-homocholestan-3-on, $C_{27}H_{45}NOCl_2$, space group $P2_12_12_1$, $Z=4$ (Mootz & Berking, 1970). The published parameters of the two chlorine atoms were used as heavy atoms; the Fourier coefficients produced by *DIRDIF.D* gave an electron density map that revealed the molecule much better than the normal difference Fourier synthesis. The improvement of the phases is shown in Tables 3 and 4.

Table 2. Number of reflexions of the test structure in each of the six categories

Category	Number of reflexions
(a1)	458 (360 reflexions with $ E_1 > 0.9$)
(a2)	62 (18 reflexions with $ E_1 > 0.9$)
(b1)	613
(b2)	364
(c1)	399
(c2)	45

Conclusions

The procedure described above proved to be useful in routine structure analysis of heavy-atom compounds, as well as for partially known equal-atom structures. The special case, where the origin and/or enantiomorph is not completely fixed by the known part of the structure is presently being studied.

Table 3. Results of the tangent refinement procedure for the test structure, for reflexions of the categories (a1) and (a2) with $|E_1| > 0.9$

Ranges in W_1	Ranges in phase deviation*	0-30°	30-60°	60-90°	90-180°	Average phase deviation	Number of reflexions
Category (a1)							
0.0-0.3	Before†	6	6	9	15	85°	36
	After †	29	4	2	1	20°	
0.3-0.6	Before	6	8	5	6	62°	25
	After	22	2	0	1	21°	
0.6-0.99	Before	23	21	13	4	45°	61
	After	46	12	3	0	19°	
> 0.99	Before	51	39	17	6	39°	113
	After	98	14	1	0	15°	
Category (a2)							
$W_1 = 1.0$	Before	13	3	1	1	28°	18
	After	13	5	0	0	20°	
Projection reflexions							
0.0-1.0	Before	94	-	-	31	-	125
	After	121	-	-	4	-	

* Phases are compared with the true phases φ_{calc} , calculated from the final light-atom structure. Before applying the tangent formula, the 'phase deviation' of a reflexion is defined as $|\varphi_1 - \varphi_{calc}|$. After the tangent refinement the 'phase deviation' is defined as $|\varphi_L - \varphi_{calc}|$.

† Tabulated is the number of reflexions belonging to the specified ranges in W_1 and 'phase deviation', before and after the tangent refinement.

Table 4. Average deviations of refined phases for the test structure

Ranges in W_1	Ranges in phase deviation before tangent refinement*	0-30°	30-60°	60-90°	90-180°
Category (a1)					
0.0-0.3	After†	14° (6)	10° (6)	26° (9)	23° (15)
0.3-0.6		14 (6)	16 (8)	20 (5)	33 (6)
0.6-0.99		12 (23)	22 (21)	29 (13)	6 (4)
> 0.99		15 (51)	17 (39)	12 (17)	12 (6)
Category (a2)					
$W_1 = 1.0$		15 (13)	47 (3)	3 (1)	30 (1)

* Ranges in $|\varphi_1 - \varphi_{calc}|$, see Table 3.

† Average in $|\varphi_L - \varphi_{calc}|$, after the tangent refinement for the reflexions in the given ranges (the number of reflexions is given in parentheses).

One of us (Th. E. M. van den Hark) acknowledges support from the Dutch Foundation for Pure Research, ZWO/FOMRE.

References

- BEURSKENS, P. T., ENCKEVORT, W. J. P., MENGER, E. M. & BOSMAN, W. P. (1976). To be published.
- BEURSKENS, P. T., VAN DEN HARK, T. E. M. & BEURSKENS, G. (1976). *Acta Cryst.* A32, 821–822.
- BEURSKENS, P. T. & NOORDIK, J. H. (1971). *Acta Cryst.* A27, 187–188.
- GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1971). *Acta Cryst.* A27, 368–376.
- GOULD, R. O., VAN DEN HARK, T. E. M. & BEURSKENS, P. T. (1975). *Acta Cryst.* A31, 813–817.
- MOOTZ, D. & BERKING, B. (1970). *Acta Cryst.* B26, 1362–1372.

Acta Cryst. (1976). A32, 821

Application of Direct Methods on Difference Fourier Coefficients for the Solution of Partially Known Structures

BY PAUL T. BEURSKENS, TH. E. M. VAN DEN HARK AND GEZINA BEURSKENS

Crystallography Laboratory, University of Nijmegen, Toernooiveld, Nijmegen, The Netherlands

(Received 26 January 1976; accepted 1 April 1976)

A partially known structure, which may be a fragment of a molecule found by direct methods, can be solved completely by applying a weighted tangent refinement procedure on difference structure factors, similar to the authors' procedure for heavy-atom structures. Often the molecular fragment is found to be misplaced with respect to the symmetry elements: then, the space-group symmetry is reduced to $P1$, and the procedure is used to locate the symmetry elements.

Introduction

In previous papers (Gould, van den Hark & Beurskens, 1975; van den Hark, Prick & Beurskens, 1976) we have described a procedure (for centrosymmetric, and non-centrosymmetric structures respectively) for the solution of heavy-atom structures, where the positions of some heavy atoms are known. This procedure is called *DIRDIF*. Naturally, the procedure can also be used for equal-atom structures when a molecular fragment is known.

Application of direct methods for the solution of not too small structures often leads to an electron density map (or E map) from which a molecular fragment can be recognized. Patterson search techniques or tangent-formula recycling techniques may then be used to complete the solution of the structure. Often the molecular fragment is in the correct orientation but shifted with respect to the symmetry elements. The solution of the

structure can then be sought by translation functions, or by reducing the space-group symmetry to $P1$, where any position of the molecular fragment is correct by definition.

For these cases the *DIRDIF* procedure is a very convenient tool, which is illustrated by the following examples.

Examples

Heptahelicene, $C_{30}H_{18}$, (see Fig. 1) crystallizes in two modifications:

(I) Space group $P2_1$; $Z=4$; two independent molecules per unit cell (Beurskens, Beurskens & van den Hark, 1976). We had troubles in solving this structure, as could be expected. We managed to find one heptahelicene molecule, incorrectly placed, and we could solve the structure as described below.

(II) Space group $P2_1/c$; $Z=4$; one molecule per asymmetric unit (van den Hark & Beurskens, 1976). We had troubles in solving this structure, partly because of the relatively small number of reflexions that could be measured. One heptahelicene molecule was found, also incorrectly placed, and this was used as described below.

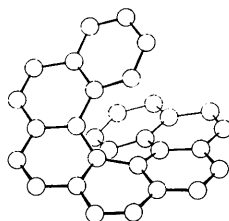


Fig. 1. Carbon skeleton of the heptahelicene molecule.

Description of the procedure

The symmetry is supposed to be $P1$, and the monoclinic reflexion data set is expanded using $|F_{obs}(h\bar{k}l)| =$