approach is capable of handling this predictable rational dependence. We intend to extend these investigations to multivariate p.d.f.'s of $|E|$, to be used in the field of direct methods.

References

ABRAMOWITZ, M. & STEGUN, I. A. (1972). *Handbook of Mathematical Functions.* New York: Dover.
COLLIN, R. L. (1955). *Acta Cryst.* **8**, 499–502.
FOSTER, F. & HARGREAVES, A. (1963). *Acta Cryst.* **16**, 1124–1133.
HARGREAVES, A. (1956). *Acta Cryst.* **9**, 191–192.
HAUPTMAN, H. & KARLE, J. (1953). *Acta Cryst.* **6**, 136–141.

ILYUKHIN, V. V. & NIKITIN, A. V. (1963). *Sov. Phys. Crystallogr.* **8**, 3–6.
International Tables for X-ray Crystallography (1965). Vol. I, 2nd ed., edited by N. F. M. HENRY & K. LONSDALE. Birmingham: Kynoch Press.
KARLE, J. & HAUPTMAN, H. (1953). *Acta Cryst.* **6**, 131–135.
PRADHAN, D., GHOSH, S. & NIGAM, G. D. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 43–51. Guilderland: Adenine Press.
SHMUELI, U. (1982). *Acta Cryst.* A**38**, 362–371.
SHMUELI, U. & WEISS, G. H. (1985). In *Structure and Statistics in Crystallography*, edited by A. J. C. WILSON, pp. 53–66. Guilderland: Adenine Press.
SHMUELI, U. & WEISS, G. H. (1987). *Acta Cryst.* A**43**, 93–98.
SHMUELI, U., WEISS, G. H. & KIEFER, J. E. (1985). *Acta Cryst.* A**41**, 55–59.
SHMUELI, U., WEISS, G. H., KIEFER, J. E. & WILSON, A. J. C. (1984). *Acta Cryst.* A**40**, 651–660.
SIM, G. (1958). *Acta Cryst.* **11**, 123–124.
SRINIVASAN, R. & PARTHASARATHY, S. (1976). *Some Statistical Applications in X-ray Crystallography.* Oxford: Pergamon Press.
WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

POSIT – a Method for Structure Determination of Small Partially Known Molecules from Powder Diffraction Data. Structure of 6-Methyl-1,2,3,4-tetrahydropyrimidine-2,4-dione (6-Methyluracil)

BY G. RECK, R.-G. KRETSCHMER AND L. KUTSCHABSKY

Zentralinstitut für Molekularbiologie der Akademie der Wissenschaften der DDR, Robert-Rössle-Strasse 10, Berlin DDR-1115, German Democratic Republic

AND W. PRITZKOW

Zentralinstitut für Physikalische Chemie der Akademie der Wissenschaften der DDR, Rudower Chaussee 5, Berlin DDR-1199, German Democratic Republic

Abstract

POSIT is a computer program package for the determination of the approximate structure of small organic molecules using known molecular fragments. The orientation and the translation vector of this fragment are determined and refined by a special R-value analysis. In contrast to other molecular replacement methods, only a few low-order strong reflections are needed to start the structure analysis. Therefore, this method is suitable for the determination of crystal structures of small rigid or semi-rigid organic molecules without heavy atoms from powder diffraction data. $C_5H_6N_2O_2$, $M_r = 126 \cdot 11$, monoclinic, $P2_1/c$, $a = 4 \cdot 5137\,(7)$, $b = 10 \cdot 9888\,(12)$, $c = 11 \cdot 7224\,(9)$ Å, $\beta = 97 \cdot 554\,(8)°$, $V = 576 \cdot 39\,(7)$ Å^3, $Z = 4$, $D_x = 1 \cdot 453$ Mg m^{-3}, λ (Cu $K\alpha$) $= 1 \cdot 54178$ Å, $\mu = 0 \cdot 987$ mm^{-1}, $F(000) = 264$, $T = 297$ K, final $R = 0 \cdot 04$ for 165 approximate structure factors obtained from a powder diffraction measurement. The molecules were found in the diketo form. They are linked by a two-dimensional network of hydrogen bonds forming layers parallel to the yz plane.

Introduction

Normally, powder diffracion data are used for the identification of unknown crystalline samples, for symmetry and lattice-constant determination, for phase analysis and for refinement of approximately known structures. It is very difficult to determine an unknown structure because most reflections in the powder pattern overlap. Therefore, up to now only a few structures, most of them containing heavy atoms, have been solved from powder diffraction

data. One of the first was the structure of copper oxalate (Schmittler, 1968).

Our aim is to determine the structures of organic non-heavy-atom compounds from powder diffraction data. For various reasons the atomic parameters from powder data are less accurate than those obtained from single-crystal data. Very often, however, it is impossible to grow crystals suitable for a single-crystal analysis. Sometimes pharmaceutically important compounds exist as different modifications with different physicochemical and pharmakinetical properties. Interesting modifications are very often thermodynamically less stable than others, and crystals of them are of poor quality. In cases of small rigid or semi-rigid molecules (up to 20 or 25 non-hydrogen atoms) the information on the molecular geometry obtained from a single crystal of a stable modification can be used as starting input for structure determination from powder data. Sometimes this information is also available from the literature. There are some well known molecular replacement methods developed by Hoppe & Paulus (1967), Rossmann & Blow (1962), Tollin, Main & Rossmann (1966), Karle (1970), Nordman (1972) and others. But these methods require a lot of data which are not available in powder diffractometry. In some cases this problem can be solved by *POSIT*.

Apart from the title compound, *POSIT* was used to solve two special problems: first, the structure solution of diaminonaphthoquinone (Schmittler-Fichtner, Reck & Menz, 1988) and, secondly, investigation of order–disorder phenomena in the β' modification of amrinone (Reck, Hagen & Höhne, 1986). In the case of diaminonaphthoquinone, the translation vector of the molecule with respect to the origin was known because the molecule has a centre of symmetry which coincides with that of the space group $P2_1/c$. Thus, only the Eulerian angles defining the orientation were unknown. In the case of amrinone-β', the orientation of a molecular dimer could be derived from the single-crystal structure of the β modification. Therefore, only the translation problem had to be solved by *POSIT*. The title compound is the first example for which *POSIT* was used to determine both the orientation and the translation vector. Furthermore, two single-crystal structures could be solved by *POSIT*, namely the structures of p-toluquinonemonoguanylhydrazone (Reck, Kretschmer & Schulze, 1984) and 3-methoxyestra-1,3,5(10)-triene-14β,17β-diol (Reck, Schubert & Kretschmer, 1985). For these analyses only crystals of poor quality were available.

Method and program

Fig. 1 shows a flowchart of *POSIT*. The first step in the structure determination is the input of a model of the molecule or a known molecular fragment in an arbitrary cell. First the orientation of this fragment is determined. As criterion of fit in the rotation search, the mean R value (Reck, Kutschabsky & Krause, 1975) or other special figures of merit (see below) are used. Normally, only a few low-order strong reflections are needed in this procedure.

After the selection and storage of a certain number of solutions the best orientations are refined by the method of steepest descents (Bronstein & Semendjajew, 1985). The next step is the determination of the translation vector of the correctly oriented molecular fragment. This is done by the calculation of an R-value map (Kutschabsky & Ziemer, 1973) in a fast procedure. After selection and storage of the positions corresponding to low R values, each one is refined, also by the method of steepest descents. Now the coordinates of the correctly oriented and positioned molecular fragment can be put into a rigid-body least-squares refinement using all well resolved reflections. Missing atoms are located by Fourier syntheses. The final structure refinement can be done by a total-pattern refinement (Rietveld, 1969), by the two-step method using integral intensities (Will, Parrish & Huang, 1983) or by means of approximate observed structure factors. As mentioned above some special problems can be solved by *POSIT* in a straightforward way. For instance, if a heavy atom is part of a known rigid fragment of the molecule, it can first be located. Then the molecule is rotated about the heavy-atom position with normal R values as a criterion of fit. In this way, the crystal structure of a thiogibberellin was solved from single-crystal
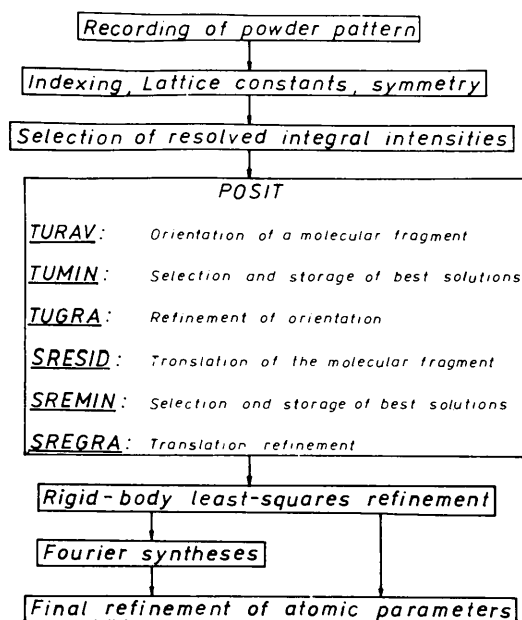

Fig. 1. Flowchart of *POSIT*.

data of very poor quality (Schierhorn, Adam, Kutschabsky & Leibnitz, 1987). If the heavy atom is not attached to the molecule or the structure consists of two molecules or two rigid molecular fragments, *POSIT* is started to determine the arrangement of the first part. Then this can be used as the fixed group and the second part can be determined in the normal way as described above.

In what follows the main procedures of *POSIT*, namely the orientation and translation determination, will be discussed in more detail. Reck, Kutschabsky & Krause (1975) found that the averaged value of an R-value map calculated with strong reflections in order to determine the translation vector of a molecule was significantly lower for a correct orientation than for incorrect ones. This effect can be explained directly by the structure-factor formula. As an example we will use the space group $P2_12_12_1$ having the equipoints x, y, z; $\frac{1}{2}-x, -y, \frac{1}{2}+z$; $\frac{1}{2}+x, \frac{1}{2}-y, -z$ and $-x, \frac{1}{2}+y, \frac{1}{2}-z$. For a shift of the molecule by Δx, Δy, Δz we can write the structure factor as

$$
\begin{aligned}
F(hkl) = &\left\{\sum_j f_j \exp[2\pi i(hx_j + ky_j + lz_j)]\right\} \\
&\times \exp[2\pi i(h\Delta x + k\Delta y + l\Delta z)] \\
&+ \left\{\sum_j f_j \exp[2\pi i(-hx_j - ky_j + lz_j)]\right\} \\
&\times \exp\{2\pi i[h(\tfrac{1}{2}-\Delta x) - k\Delta y + l(\tfrac{1}{2}+\Delta z)]\} \\
&+ \left\{\sum_j f_j \exp[2\pi i(hx_j - ky_j - lz_j)]\right\} \\
&\times \exp\{2\pi i[h(\tfrac{1}{2}+\Delta x) + k(\tfrac{1}{2}-\Delta y) - l\Delta z]\} \\
&+ \left\{\sum_j f_j \exp[2\pi i(-hx_j + ky_j - lz_j)]\right\} \\
&\times \exp\{2\pi i[-h\Delta x + k(\tfrac{1}{2}+\Delta y) + l(\tfrac{1}{2}-\Delta z)]\}.
\end{aligned}
\tag{1}
$$

The first factors of the four products depend only on the orientation of the molecule and may be denoted by $D_m(hkl)$ with $m = 1, 2, \ldots, M$; the second terms contain the information about the translation. For very strong structure factors it is expected that for the correct orientation at least some products will be large. This involves large absolute values of the corresponding orientation terms. The orientation terms referred to the E values are denoted by $D'_m(hkl)$. Fig. 2 shows the distribution of the orientation terms D'_m for strong E_o values corresponding to $|E(hkl)| = 2\cdot0$ for non-centrosymmetric structures. For comparison the distribution of the orientation terms for the same reflections with a spurious orientation of the molecule is shown. Large differences of the distributions make it possible to determine the correct orientation of the molecule by the methods described in this paper.

The influence of the unknown translation terms is suppressed in the calculation of a mean R value because in this way they give a complicated smeared function. This function is smooth enough if the mean R value is calculated from R values at 20 to 30 equidistant points of the translation map.

At the beginning of a structure analysis from powder data sometimes only the Laue group is known. This is sufficient for the orientation determination by mean R-value calculations because all translational components of the equipoints can be transferred into the translation terms. The decision on the space group can be made by subsequent translational R-value maps testing all possible space groups of the Laue group.

The application of the mean R value for the determination of the orientation was most successful for space groups with low symmetry and, therefore, with few D_m terms. The effect of the translational terms and the calculation of mean R values can be avoided if instead of the usual R value a figure of merit is calculated based on the single terms $D_m(hkl)$,

$$
R' = \frac{\sum\limits_{hkl} \sum\limits_{m} \big||F_o(hkl)| - K|D_m(hkl)|\big|}{\sum\limits_{hkl} |F_o(hkl)|}.
\tag{2}
$$

K is a scale factor taking into account the decreased number of atoms. The formula

$$
R'' = \frac{\sum\limits_{hkl} |E_o(hkl)|^{n+1}}{\sum\limits_{hkl} |E_o(hkl)| \sum\limits_{m} |D'_m(hkl)|^n}
\tag{3}
$$

based on the strong $E_o(hkl)$ (with $|E_o(hkl)| > 2\cdot0$) rather than on the $F_o(hkl)$ proved to be useful especially for space groups with three or more symmetrically equivalent molecules. The exponential factor n allows a weighting of the contributing terms by taking into account the effect that for $|E_o(hkl)| > 2$ and for the correct orientation large values of $|D'_m(hkl)|$ are much more frequent than for spurious orientations (Fig. 2).
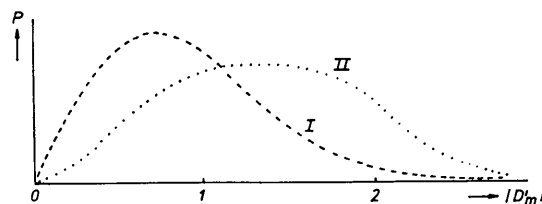

Fig. 2. Distribution P of the orientation terms D'_m for strong E_o values corresponding to $|E(hkl)| = 2\cdot0$ for non-centrosymmetric structures (dotted line) compared with the normal distribution $P(|D'_m|) = 2|D'_m|\exp(-|D'_m|^2)$ (dashed line).

After the determination and refinement of the orientation, the translation vector of the molecular fragment is calculated by an R-value map using the conventional discrepancy factor

$$R = \sum_{hkl} ||F_o(hkl)| - |F_c(hkl)|| / \sum_{hkl} |F_o(hkl)| \qquad (4)$$

as a criterion of fit. If the input molecular fragment contains less than half of the total number of atoms, then only the strongest reflections should be used to calculate the R-value map. But sometimes additional minima appear in the map. Our latest investigation has shown that it is better in these cases to use all low-order reflections and to calculate weighted R values

$$R_w = \sum_{hkl} w|\Delta F(hkl)| / \sum_{hkl} |F_o(hkl)| \qquad (5)$$

with

$$w = A + [B|F_o(hkl)| / |F_o^{max}(hkl)|]. \qquad (6)$$

If all atoms belong to the input molecule, one may set $A = 1$ and $B = 0$. If only half of the atoms are known it is recommended to set $A = 0$ and $B = 1$. It should be mentioned that the calculation of an R-value map is a very fast procedure because if one splits the $F_c(hkl)$ value as in (1), the orientation terms containing all known atoms have to be calculated only once for the starting point.

Structure determination of 6-methyluracil

For the structure determination of this biologically interesting compound only the crystalline powder was available. Neglecting all zwitterionic forms, six tautomers are possible from the theoretical point of view. The structure analysis should decide which form is realized in the crystal. Furthermore, the molecular packing and the hydrogen bonds are of special interest.

The lattice constants were determined from a Guinier photograph by the program *GIKO* (Walther, 1983). Observed systematic absences $h0l$ for l odd and $0k0$ for k odd indicated space group $P2_1/c$, which was confirmed by the results of the subsequent structure anlysis. The diffraction pattern in the range of $10 \leq 2\theta \leq 80°$ in steps of $0.02°$ was recorded with Cu $K\alpha$ radiation on a computer-controlled two-circle HZG-4 diffractometer produced by VEB Freiberger Präzisionsmechanik.

From the powder pattern only 23 resolved integral intensities could be separated by the program *PEAK-FIT* (Mortier & Costenoble, 1973). The five strongest reflections were used in *TURAV* to determine the orientation of an ideal molecular model containing carbon atoms only. The endocyclic and exocyclic bonds were set to 1·39 and 1·32 Å, respectively; all bond angles were assumed to be 120°. In the rotation search, the Eulerian angles were varied in steps of 15°. Only one significant mean R-value minimum was observed at $\psi = 40°$ (rotation about $\mathbf{c}$), $\theta = 15°$ (about $\mathbf{a}'$) and $\varphi = 45°$ (about $\mathbf{b}''$). The *TURAV* results were refined by *TUGRA* to 52, 9 and 39°, respectively.

This orientation of the molecular model was then input to *SRESID*. Using all 23 resolved reflections *SRESID* gave a significantly low R value at $\Delta x = 0.33$, $\Delta y = 0.20$ and $\Delta z = 0.40$. These values were refined to 0·317, 0·192 and 0·407, respectively.

It should be added that *TURAV* and *TUGRA* gave similar results independently, whether $P2_1/c$ or $P2/m$ symmetry was used. $P2/m$ was excluded by *SRESID*. The scaling factor for the $F_o(hkl)$ used in *POSIT* was obtained from the quotient $\sum|F_o(hkl)| / \sum|F_c(hkl)|$ for which the F_c was calculated assuming a structure with a randomly oriented and positioned molecule. An overall isotropic temperature factor U was set to 0·05 Å^2.

A calculation of all interatomic distances with atomic coordinates obtained by the *POSIT* procedure showed no steric hindrances. The two shortest distances of 2·87 and 2·85 Å, respectively, between ring atoms and substituents were established as N···O hydrogen bonds. In this way oxygen, nitrogen and carbon atoms of the molecule could be identified. A subsequent rigid-body least-squares refinement with *SHELX*76 (Sheldrick, 1976) converged with $R = 0.27$ for 23 reflections.

At this stage of the structure analysis three possible tautomeric forms could be excluded. The final structure refinement should decide between the remaining forms. A conventional least-squares refinement using the 23 resolved integral intensities derived from the powder pattern was impossible.

The structure refinement by the Rietveld method with a modified version of the program *PFSR* (*DBW* 3·2) (Wiles & Young, 1981) gave entirely unacceptable bond lengths and angles. In this program it was not possible to constrain these parameters, and therefore a refinement minimizing

$$||F_o^{ap}(hkl)| - |F_c(hkl)||^2$$

was carried out. $F_o^{ap}(hkl)$ is the structure factor for the hkl reflection calculated from the observed profile by separating the contributing peaks according to the calculated integral intensities. This approximation was described by Rietveld (1969). The following formula was used for the determination of the approximate observed structure factors:

$$F_o^{ap}(hkl) = \frac{\sum\limits_{i=m}^{M} y_i(\text{obs})}{\sum\limits_{i=m}^{M} y_i(\text{calc})} F_c(hkl). \qquad (7)$$

$y_i(\text{obs})$ and $y_i(\text{calc})$ are the observed and calculated intensities, respectively, at equidistant 2θ values

Table 1. *Atomic coordinates* ($\times 10^4$) *and isotropic temperature factors* ($\text{Å}^2 \times 10^3$) *with e.s.d.'s in parentheses*

	x	y	z	U_i
N1	2588 (15)	2703 (6)	2954 (5)	19 (3)
C2	730 (21)	1724 (8)	2994 (9)	13 (3)
N3	1020 (14)	1171 (6)	4076 (6)	17 (2)
C4	2975 (19)	1494 (9)	5082 (9)	16 (3)
C5	4763 (18)	2518 (8)	4909 (8)	18 (3)
C6	4608 (18)	3123 (8)	3893 (8)	18 (3)
O2	-1024 (12)	1349 (5)	2186 (5)	16 (2)
O4	2896 (11)	886 (5)	5928 (5)	17 (2)
C6'	6556 (17)	4228 (8)	3781 (6)	16 (3)

Table 2. *Bond lengths* (Å) *and bond angles* (°) *with e.s.d.'s in parentheses*

C2–N1	1·369 (9)	O2–C2–N1	125·1 (9)
C2–O2	1·224 (8)	N1–C2–N3	112·8 (9)
C2–N3	1·397 (8)	O2–C2–N3	122·0 (9)
C4–N3	1·422 (8)	C2–N3–C4	128·2 (9)
C4–O4	1·200 (8)	N3–C4–O4	117·7 (9)
C4–C5	1·415 (10)	N3–C4–C5	112·5 (9)
C6–C5	1·358 (9)	O4–C4–C5	129·8 (10)
C6–C6'	1·515 (10)	C4–C5–C6	123·4 (9)
C6–N1	1·411 (8)	C5–C6–C6'	120·3 (9)
		C5–C6–N1	118·5 (9)
		C6'–C6–N1	121.2 (8)
		C6–N1–C2	124·5 (8)

between *m* and *M*. The 2θ range $M - m$ is the assumed width of the reflection *hkl*.

For the first calculation of y_i(calc) and $F_c(hkl)$ the atomic coordinates obtained by the rigid-body least-squares refinement and functions for profile, half-width, background and θ corrections were used. 165 $F_o^{ap}(hkl)$ obtained in this way were available to refine
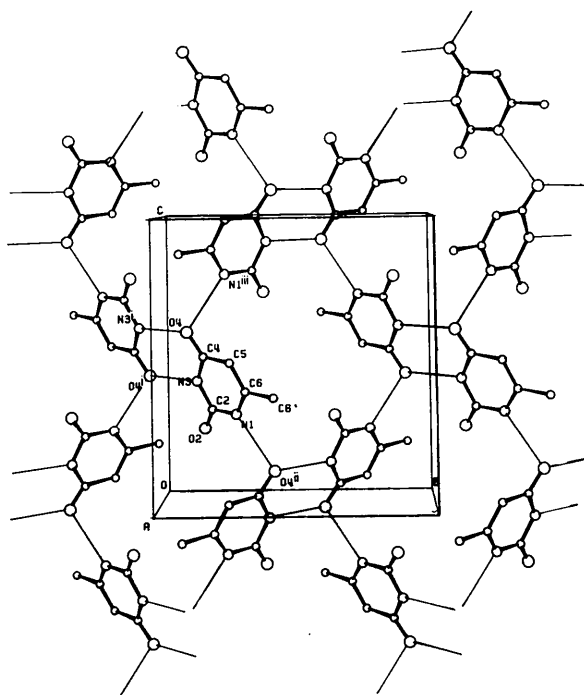
the atomic coordinates of the non-hydrogen atoms and isotropic temperature factors using the *SHELX*76 program. The refined parameters were the input for a subsequent $F_o^{ap}(hkl)$ calculation and the refinement was repeated. After six cycles of this refinement procedure the R value dropped from 0·25 to 0·04.* It should be mentioned that this value depends also on the model. The agreement value R_p for the y_i(obs) and y_i(calc) was 0·19. The resulting coordinates are listed in Table 1.

Special problems of the structure refinement will be discussed in a separate paper. The resulting bond lengths and angles given in Table 2 establish unambiguously the diketo form of 6-methyluracil. There are two symmetry-independent hydrogen bonds in the crystal structure, namely N3–H$\cdots$O4^i = 2·87 Å [(i): $-x, -y, 1-z$] and N1–H$\cdots$O4^{ii} = 2·85 Å [(ii): x, $\frac{1}{2} - y, z - \frac{1}{2}$]. These and their symmetry-related bonds link molecules into infinite layers as illustrated in Fig. 3.

Fig. 3. Packing of molecules in the crystal structure of 6-methyluracil; the hydrogen bonds are indicated.

References

BRONSTEIN, I. N. & SEMENDJAJEW, K. A. (1985). *Taschenbuch der Mathematik*, edited by G. GROSCHE, V. ZIEGLER & D. ZIEGLER. Leipzig: B. G. Teubner Verlagsgesellschaft.
HOPPE, W. & PAULUS, E. (1967). *Acta Cryst.* 23, 339–343.
KARLE, J. (1970). In *Crystallographic Computing*, edited by F. R. AHMED, S. R. HALL & C. P. HUBER. Copenhagen: Munksgaard.
KUTSCHABSKY, L. & ZIEMER, B. (1973). *Krist. Tech.* 8, 437–441.
MORTIER, W. J. & COSTENOBLE, M. L. (1973). *J. Appl. Cryst.* 6, 488–490.
NORDMAN, C. E. (1972). *Acta Cryst.* A28, 134–143.
RECK, G., HAGEN, V. & HÖHNE, E. (1986). *Pharmazie*, H3, 181–186.
RECK, G., KRETSCHMER, R.-G. & SCHULZE, W. (1984). *Cryst. Res. Technol.* 19, 213–217.
RECK, G., KUTSCHABSKY, L. & KRAUSE, CH. (1975). *Krist. Tech.* 10, 947–950.
RECK, G., SCHUBERT, G. & KRETSCHMER, R.-G. (1985). *Cryst. Res. Technol.* 20, 757–762.
RIETVELD, H. M. (1969). *J. Appl. Cryst.* 2, 65–71.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* 15, 24–31.
SCHIERHORN, A., ADAM, G., KUTSCHABSKY, L. & LEIBNITZ, P. (1987). *J. Chem. Soc. Perkin Trans.* 1, pp. 2111–2118.
SCHMITTLER, H. (1968). *Monatsber. Dtsch. Akad. Wiss. Berlin*, 10, 581–603.
SCHMITTLER-FICHTNER, H., RECK, G. & MENZ, M.-L. (1988). *Cryst. Res. Technol.* Submitted.
SHELDRICK, G. M. (1976). *SHELX*76. Program for crystal structure determination. Univ. of Cambridge, England.
TOLLIN, P., MAIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* 20, 404–407.
WALTHER, G. (1983). *GIKO. A Program for the Determination of Cell Parameters from Diffraction Data by the Least-Squares Method.* Berlin: Zentralinstitut für Physikalische Chemie der AdW der DDR.
WILES, D. B. & YOUNG, R. A. (1981). *J. Appl. Cryst.* 14, 149–151.
WILL, G., PARRISH, W. & HUANG, T. C. (1983). *J. Appl. Cryst.* 16, 611–622.