Acta Cryst. (1969). B**25**, 1201

Ambiguity in the results of least-squares analysis - another cautionary tale.

By R. Parthasarathy, *Center for Crystallographic Research, Roswell Park Memorial Institute, Buffalo, N.Y.* 14203, *U.S.A.* and J. G. Sime and J. C. Speakman, *Chemistry Department, The University, Glasgow, W.*2, *Scotland*

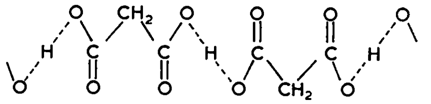
(*Received* 15 *July* 1968 *and in revised form* 29 *November* 1968)

The crystal structure of potassium hydrogen malonate has been studied by the independent use of each of the two space groups *C*2 and *C*2/*m*. The different space groups led to structures which differed crystallographically and chemically, though both were supported by apparently valid three-dimensional analyses from diffractometer data. When the block-diagonal least-squares approximation is used the results of a careful refinement, with good intensity data, can be dependent on the starting structure. Full-matrix refinement avoids this ambiguity, but may not work smoothly in such a situation owing to inverse overlap.

Potassium hydrogen malonate, $KH(C_3H_2O_4)$, crystallizes in the monoclinic system with diffraction symbol *C**/* (Donnay & Kennard, 1964). Its crystal structure has been studied independently at Roswell Park Memorial Institute (RP) and at Glasgow (G). At one stage the two studies led to structures which were different chemically and crystallographically, though both were supported by apparently valid three-dimensional analyses from diffractometer data. We believe our experience is worth reporting. Somewhat similar experiences have been reported recently by Donohue & Goodman (1967) and by Nilsson, Liminga & Olovsson (1968).

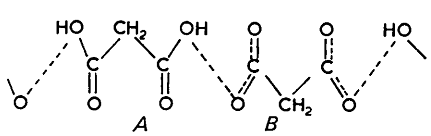
A preliminary account of the G work has been published (Ferguson, Sime, Speakman & Young, 1968). For non-crystallographic reasons *C*2/*m* was chosen as space group. There being four molecules per cell, this symmetry implies that the asymmetric unit is one-half of the above formula. Full-matrix least-squares refinement, based on 480 low-order data and 39 parameters converged with $R = 0{\cdot}058$. The structure, represented by formulation (I), contained infinite chains of hydrogen-malonate residues linked by symmetrical hydrogen bonds. The two carboxyl groups of each residue are crystallographically related by a twofold axis. (Further refinement with more extensive data is in progress. The results given in the preliminary report are confirmed.)

(I)

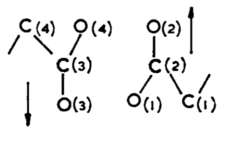


*C*2/*m*

(II)



*C*2



Numbering of Atoms (RP)

At RP the space group was assumed to be *C*2. No specific assumption as to the structure was made; the structure analysis was carried out by the heavy-atom method and iterative Fourier-series and least-squares analysis. The apparently reasonable assumption was made that the structure would converge to the more symmetrical *C*2/*m* if this were indeed the true space group. Refinement was by the block-diagonal approximation, with 860 reflexions and individual anisotropic thermal parameters for non-hydrogen atoms. It led to an *R* of 0·070. The structure found and briefly reported (Parthasarathy, 1968) was of the unsymmetrical type involving discrete malonic acid molecules, $HO_2C.CH_2.CO_2H$, and bivalent anions, $^{-}O_2C.CH_2.CO_2^{-}$, and represented by formulation (II). With one or two exceptions, the molecular dimensions were consistent with this structure.

At this stage, we each became aware of the other's work and exchanged information. At RP the 860 data were used in a block-diagonal refinement, starting from a *C*2/*m* structure with coordinates virtually those found at G. Refinement converged (with $R = 0{\cdot}073$ for the reduced number of parameters) to essentially the same structure. One coordinate [*z* for O(1)] differed by 0·017 Å. For all other coordinates the root mean square difference was only 0·0030 Å, with 0·0046 Å as maximum. According to Hamilton's (1965) tables however, the hypothesis that the structure is centrosymmetric can be rejected by the use of the *R* value ratio* at the 1% significance level, which implies that *C*2/*m* is not a valid choice. We believe that this conclusion, based as it was on the Hamilton test as we applied it, may be false for there are strong non-crystallographic reasons for preferring structure (I). Some of the bond-lengths and angles found on the basis of (II) are chemically implausible; the infrared spectrum of this compound does not resemble what (II) would lead us to expect, *viz.* a superposition of the spectra of free acid and its neutral salt.

At G the 73 parameters, corresponding to *C*2, obtained at RP were refined, with the 480 G data, by the full-matrix program. Generally, the coordinates moved towards the values previously found in the *C*2/*m* refinement: most of them moved into substantial coincidence, but one or two went too far or not far enough, and two coordinates of O(4) moved in the 'wrong' direction. More significantly there was ample evidence of instability in the refinement: standard deviations were five times greater than in *C*2/*m*, though the excess of data over parameter had fallen only from 441 to 407. The shifts were erratic and remained relatively large even after six cycles. The situation resembles that briefly mentioned by Choi & Boutin (1968) for lead azide when the space group was taken to be $Pc2_1n$ instead of *Pcmn*. The mean positions found at G for pairs of atoms which would be symmetry-related in *C*2/*m* do not differ

---

* By *R* we mean, throughout this paper, the quantity commonly used in structure analysis, *viz.* $\Sigma|\Delta|/(\Sigma|F_o|)$. This is not identical with the R'' used by Hamilton. We believed, however, that the *ratio* of two corresponding *R* values would not be greatly affected, since *R* and R'' are quantities of the same order.

Table 1. *Comparison of $C2$ and $C2/m$ refinements*

		Block-diagonal refinement at RP		Full-matrix refinement at G	
Molecule	Bond	$C2$	$C2/m$	$C2$	$C2/m$
A	C(3)–C(4)	1·498 (007) Å	1·531 (007) Å	1·524 (030) Å	1·519 (006) Å
	C(3)–O(3)	1·174 (007)	1·223 (006)	1·200 (030)	1·220 (005)
	C(3)–O(4)	1·306 (007)	1·294 (006)	1·291 (030)	1·293 (005)
B	C(1)–C(2)	1·559 (010)	—	1·514 (030)	—
	C(2)–O(2)	1·292 (007)	—	1·249 (030)	—
	C(2)–O(1)	1·285 (007)	—	1·300 (030)	—

significantly – in relation to the enhanced standard deviations – from those found in the original $C2/m$ refinement. The difference between the $C2$ coordinates found with the full-matrix method at G and those found by the block-diagonal method at RP is large: the r.m.s. difference amounts to 0·030 Å, with a maximum of 0·097 Å.

Inspection of the correlation matrix showed there to be large cross-terms (up to $|0{\cdot}9|$) between corresponding parameters of pairs of atoms, taken to be independent in $C2$ but symmetry-related in $C2/m$. These large correlations in non-centrosymmetric structures with a partial centre of symmetry occur because of 'inverse overlap' (Srinivasan, 1961) between one atom and another related to it by inversion through this centre. When the inverse overlap is present between the ith and jth atoms, then it can be readily shown (Rae & Maslen, 1963) that there is an approximately linear dependence between the ith and jth rows of the least-squares matrix, which is therefore ill-conditioned and would lead to unreliable shifts and large standard deviations. Under these conditions it is not possible to obtain a satisfactory refinement even with full-matrix refinement as may be seen from the discussion by Geller (1961). We believe that the omission of these cross-terms in the block-diagonal refinement at RP was the cause of convergence to a different structure from the one found at G. Table 1 compares typical bond-lengths, resulting from the various refinements. The estimated standard deviations are shown in parentheses, and the numbering of atoms is given in the formulation.

We draw three conclusions:

(1) When one is in doubt whether to use a space group of higher or lower symmetry in a least-squares refinement, the latter is not an automatically safe choice. If the symmetrical structure is correct, the refinement does not necessarily converge to it smoothly.

(2) The choice of the lower symmetry is particularly likely to cause trouble when the block-diagonal approximation is used. The neglect of cross terms, which may be very large between symmetry-related parameters, may lead to a false structure, supported by spuriously low standard deviations.

(3) With full-matrix refinement, the danger is revealed in poor convergence, high standard deviations, and high correlation terms. This does not mean, however, that the lower symmetry is necessarily incorrect, since the inverse overlap also occurs in non-centrosymmetric structures with a partial centre of inversion; strong correlations would still occur between parameters fulfilling the inverse-overlap condition, and these parameters would tend to become indeterminate. To reach a decision on the correct space group, resort to non-crystallographic considerations may be unavoidable after all, as it was in this case.

A fuller joint account of the structure of potassium hydrogen malonate will be published elsewhere. One of us (R.P.) is grateful to Professor David Harker for many discussions and criticisms and to NIH-A-3942 and NSF-GB-4056 for financial support in part. At Glasgow we are happy to acknowledge helpful comments from Professor Cruickshank, whose study of thortveitite (Cruickshank, Lynton & Barclay, 1962) covered the same space group ambiguity. We are particularly indebted to Dr Hamilton, with whom we had valuable discussions.

References

Choi, C. S. & Boutin, H. (1968). *J. Chem. Phys.* **48**, 1397.
Cruickshank, D. W. J., Lynton, H. & Barclay, G. A. (1962). *Acta Cryst.* **15**, 491.
Donnay, J. D. H. & Kennard, O. (1964). *Acta Cryst.* **17**, 1337.
Donohue, J. & Goodman, S. H. (1967). *Acta Cryst.* **22**, 352.
Ferguson, G., Sime, J. G., Speakman, J. C. & Young, R. (1968). *Chem. Comm.* p. 162.
Geller, S. (1961). *Acta Cryst.* **14**, 1026.
Hamilton, W. (1965). *Acta Cryst.* **18**, 502.
Nilsson, A., Liminga, R. & Olovsson, I. (1968). *Acta Chem. Scand.* **22**, 719.
Parthasarathy, R. (1968). *Science.* **161**, 179.
Rae, A. I. M. & Maslen, E. N. (1963). *Acta Cryst.* **16**, 703.
Srinivasan, R. (1961). *Acta Cryst.* **14**, 1163.