LARSON, A. C. (1967). *Acta Cryst.* **23**, 664–665.

MEUNIER, B. & PROUT, K. (1979). *Acta Cryst.* **B35**, 2558–2562.

NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* **A24**, 351–359.

PROUT, K., CAMERON, T. S., FORDER, R. A., CRITCHLEY, S. R., DENTON, B. & REES, G. V. (1974). *Acta Cryst.* **B30**, 2290–2304.

WHEATLEY, P. J. (1967). *Perspectives in Structural Chemistry*, edited by J. D. DUNITZ & J. A. IBERS, Vol. 1, p. 1. London: John Wiley.

# The Importance of Weak Reflections in Resolving the Centrosymmetric–Noncentrosymmetric Ambiguity: A Cautionary Tale

BY RICHARD E. MARSH

*Arthur Amos Noyes Laboratory of Chemical Physics,\* California Institute of Technology, Pasadena, California 91125, USA*

## Abstract

The weak diffraction intensities are crucially important in determining whether a crystal structure has a real or only an approximate center of symmetry; if these intensities are deleted from the data set, an intelligent conclusion may be impossible. In addition, statistical distribution tests, if based only on the stronger intensities, may be strongly biased toward a non-centrosymmetric indication. In one recently published structure determination [Cotton & Fanwick (1980), *Acta Cryst.* B36, 457–459] a distribution test led to the assignment of the noncentrosymmetric space group *Cc* to a structure which can be better described and refined in the centrosymmetric space group *C2/c*.

In many, perhaps most, X-ray diffraction laboratories it seems to have become customary to ignore low-intensity reflections, either by dropping them from the data set if they fall below some arbitrary signal-to-noise threshold or by programming the diffractometer to skip them if the counting rate is low. The rationale for this procedure is that the weak reflections, besides being of little importance in Patterson maps and most direct phasing techniques, contribute less than their fair share to the least-squares process since, if there is appreciable background, they must be assigned relatively low weights. A side benefit of ignoring the weak reflections

is the cosmetic effect (Hirshfeld & Rabinovich, 1973) of a lower *R* index, which may make the final product more acceptable to some journals.

Hirshfeld & Rabinovich (1973) have pointed out that rejecting, or in any other way biasing, the low-intensity observations introduces systematic errors into the structural parameters, particularly the *B*'s and the scale factor. However, the effect appears to be small, and they remark that 'our limited experience indicates that in real situations the effect of biased data on the structurally interesting parameters is rarely large enough to matter.' I wish to point out that there is one aspect of diffraction crystallography – the determination of the presence or absence of a center of symmetry – where the weak reflections are crucially important, and neglecting them may well lead to an incorrect space-group assignment and severely distorted molecular geometry. I shall cite one example from the recent literature where such an event has almost surely taken place.

There are two areas of a typical crystal-structure analysis where the weak reflections show themselves as vitally important in resolving the centrosymmetric-noncentrosymmetric ambiguity. The more obvious area is during the final least-squares refinement cycles, where it has become popular to carry out a 'significance test' (Hamilton, 1965) on the statistical validity of removing the center and, thereby, increasing the number of structural parameters. Without going into the many pitfalls of such a test as it is usually performed, I wish to emphasize, as has been done in the past (Schomaker & Marsh, 1979), that it is only the weak reflections that can have any importance in distinguishing between a true and an approximate center; tests made on the stronger reflections alone are

0567-7408/81/111985-04$01.00

almost surely invalid. The reason for this is that the noncentrosymmetric distortion of an approximately centrosymmetric structure shows up only as an out-of-phase component to the structure factor – a small (if the noncentrosymmetric distortion is small) scattering vector at right angles to the 'real' vector due to the primarily centrosymmetric nature of the structure. Only if this centrosymmetric component of $F$ is itself small will the out-of-phase component be able to cause a meaningful change in the magnitude of $F$.

The second – and, potentially, more serious – area where weak reflections are of crucial importance is in the evaluation of intensity distributions often used to make an initial decision as to whether a structure is centrosymmetric or not. Prominent among these distribution tests are the $N(z)$ plot (Howells, Phillips & Rogers, 1950) and the evaluation of $E$ statistics (Karle, Dragonette & Brenner, 1965). I wish to point out that deleting the weak reflections before making such tests may severely bias the distributions and hence lead to an incorrect space-group assignment.

Properly calculated, an $N(z)$ distribution must be based on intensity data for the entire set of reciprocal-lattice points (excluding systematic absences). Howells, Phillips & Rogers (1950) make this point adequately clear: 'those [reflections] accidentally absent must be retained'. Rogers, Stanley & Wilson (1955) went even further in evaluating the effect on the $N(z)$ test of setting the intensities of unobserved (by film techniques) reflections to zero. If the weak reflections are omitted altogether – that is, if the distribution is calculated for a truncated data set that does not contain all accessible intensities – it is not a true $N(z)$ distribution. It is this latter distribution, which I shall call (at a referee's suggestion) $M(z)$, that may well give false information concerning the presence or absence of a center of symmetry.

As an example of the bias that can be introduced into intensity statistics by ignoring weak reflections, I have chosen intensity data from a representative crystal which, because of systematic absences, is known to be centrosymmetric (space group $P2_1/n$). The total data set comprises 2416 independent reflections out to $\sin \theta/\lambda = 0.48$ Å$^{-1}$ (Cu $K\alpha$ radiation), of which 2285 had net intensities greater than zero. The empirical contents of a unit cell are $Cu_8C_{76}H_{99.4}N_{24}O_{13.7}B_4Si_2F_{28}$ (Gagné, McCool & Marsh, 1980). Intensity distributions were evaluated for three different subsets of the data: (1) the entire set of 2416 inensities, unmodified in any way except that $|E|$ (but not $E^2$) was set equal to 0·0 for the 131 reflections with negative net intensities; (2) the 1890 intensities greater than 5 e.s.d.'s above background; (3) the 1561 intensities greater than 12 e.s.d.'s above background. (If the crystal had been smaller, the X-ray source less intense or the scan rate faster, these 1561 reflections might have had, say, a $3\sigma$ threshold rather than $12\sigma$.) For each of these subsets,
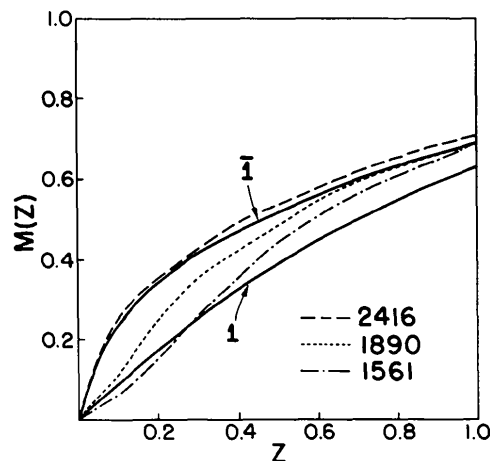


Fig. 1. Intensity distribution curves (Howells et al., 1950) for the three subsets of data, and those expected for centrosymmetric and noncentrosymmetric structures. The ordinate $M(z)$ (see text) is the fraction of reflections having $E^2$ values less than or equal to $z$.

Table 1. *Some E statistics for three subsets of data, and theoretical values for centrosymmetric and noncentrosymmetric arrays (Karle et al., 1965)*

| Number of reflections | $\langle |E| \rangle$ | $\langle E^2 \rangle$ | $\langle E^2 - 1 \rangle$ | $2B$ (Å$^2$) | $K^*$ |
|---|---|---|---|---|---|
| 2416 | 0·792 | 1·004 | 0·999 | 7·8 (5) | 1·15 (5) |
| 1890 | 0·848 | 1·003 | 0·877 | 5·8 (5) | 1·19 (4) |
| 1561 | 0·878 | 1·002 | 0·790 | 4·2 (5) | 1·21 (4) |
| $\bar{1}$ | 0·798 | 1·000 | 0·968 | | |
| 1 | 0·886 | 1·000 | 0·736 | | |

* To be applied to $F_{obs}$.

an isotropic $B$ and a scale factor were evaluated by a least-squares application of Wilson's (1942) method;* the resulting $E^2$ values were then used to calculate $M(z)$ distributions and some $E$ statistics (Karle et al., 1965). The results are shown in Fig. 1 and Table 1.

The object lesson is apparent. Whereas the entire data set (2416 reflections) shows distributions essentially identical to those expected for this centrosymmetric structure, the distributions based on only the stronger intensities are heavily biased towards noncentrosymmetric indications; indeed, when only the strongest 65% of the data are included (1561 intensities), the distributions seem to point unequivocally toward a noncentrosymmetric structure. (Note that the scale and, especially, the $B$ factors are also biased when

the weak intensities are omitted, since the deleted reflections tend to lie at high scattering angles. The resulting bias in derived $E^2$ values might well lead to difficulties in direct-methods phasing.)

In at least one instance in the recent literature there is clear evidence that the misleading distribution based on a partial data set has led to the assignment of an incorrect, noncentrosymmetric space group to a centrosymmetric structure. Tetrachlorotetrakis(diethyl sulfide)dimolybdenum(II), $Mo_2(C_4H_{10}S)_4Cl_4$, (Cotton & Fanwick, 1980; CF) crystallizes with four molecules in a $C$-centered monoclinic cell having dimensions $a = 18\cdot769$ (5), $b = 10\cdot096$ (2), $c = 15\cdot793$ (6) Å, $\beta = 109\cdot94$ (2)°. Of the 2221 reflections out to $2\theta = 45°$ (Mo $K\alpha$ radiation), 1567 – or 71% – had intensities $I$ greater than $3\sigma(I)$; the rest were rejected. A Howells, Phillips & Rogers test 'strongly indicated that the unit cell was noncentrosymmetric'; accordingly, the solution and refinement of the structure were based on the noncentrosymmetric space group $Cc$. In fact, the structure can be described and refined more satis-

factorily in the centrosymmetric space group $C2/c$. Presumably the incorrect indication resulted from the Howells, Phillips & Rogers test being based on the truncated data set; that is, $M(z)$ rather than $N(z)$.

In order to describe the reported structure in $C2/c$, the atomic parameters (Table 1, CF) were incremented by $x = 0\cdot25$, $z = -0\cdot25$ so that the Mo atoms lie on a twofold axis at $x = 0$, $z = \frac{1}{4}$; coordinates of the remaining atoms were symmetrized across this twofold axis and averaged in pairs. Anisotropic $U$'s and isotropic $B$'s (Supplementary Publication No. SUP 34868) were likewise averaged in pairs. Least-squares minimization of the quantity $\sum w(F_o^2 - F_c^2)^2$ was based on $w = 1/F_o^2$ for $F_o \geq 60$ and $w = 1/60F_o$ for $F_o \leq 60$ (Hughes, 1941; Marsh & Schomaker, 1979), since values of $\sigma(F)$ were not included in the supplementary structure factor table.

Progress of the $C2/c$ refinement is shown in Table 2. While the model formally analogous to the $Cc$ structure – model ($b$), with anisotropic heavy atoms and isotropic C atoms – converged at a somewhat higher $R$ ($0\cdot045$ vs $0\cdot041$), the comparison is misleading because the analogy is inappropriate: two isotropic atoms in a slightly noncentrosymmetric array can closely approximate, in |F| values, a pair of anisotropic atoms in a centrosymmetric array. Thus, the better comparison is with model ($c$), with all atoms anisotropic, which attains a slightly better $R$ with considerably fewer parameters than ($a$). [Comparisons among models ($a$), ($b$), and ($c$) point up only one of the problems associated with significance tests, as referred to earlier.]

The final model ($d$), Table 2, includes H atoms in positions calculated to complete tetrahedra about the C atoms, the conformations of the methyl groups having been clearly indicated by difference maps in the planes of the H atoms. The H atoms were not further refined. A final difference map had, as its largest feature, a peak

Table 2. *Agreement indices for various models of* $Mo_2(C_4H_{10}S)_4Cl_4$

|  | ($a$) | ($b$) | ($c$) | ($d$) |
|---|---|---|---|---|
| $R$ | 0·041 | 0·045 | 0·040 | 0·033 |
| Number of parameters, $p$ | 153 | 79 | 119 | 119* |
| Goodness-of-fit† |  | 10·1 | 9·0 | 6·9 |

($a$) $Cc$ (Cotton & Fanwick, 1980). Mo, Cl, S anisotropic, C isotropic, no H atoms.
($b$) $C2/c$. Mo, Cl, S anisotropic, C isotropic, no H atoms.
($c$) $C2/c$. Mo, Cl, S, C anisotropic, no H atoms.
($d$) $C2/c$. Mo, Cl, S, C anisotropic, H atoms included.

\* H atoms from difference maps; not refined.

† $[\sum w(F_o^2 - F_c^2)^2/(n - p)]^{1/2}$; weights on arbitrary scale (see text). The number of reflections $n$ is, in all cases, 1567.

Table 3. *Final coordinates* ($\times 10^5$ *for* Mo, $\times 10^4$ *for others*) *and* $U_{ij}$ *values* (Å$^2$) ($\times 10^4$ *for* Mo, $\times 10^3$ *for others*) *for* $Mo_2(C_4H_{10}S)_4Cl_4$, *model* ($d$)

The form of the anisotropic temperature factor is $\exp(-2\pi^2)(U_{11}h^2a^{*2} + \ldots + 2U_{23}klb^*c^*)$.

|  | $x$ | $y$ | $z$ | $U_{11}$ | $U_{22}$ | $U_{33}$ | $U_{12}$ | $U_{13}$ | $U_{23}$ |
|---|---|---|---|---|---|---|---|---|---|
| Mo(1) | 0 | 15004 (8) | 25000 | 376 (6) | 254 (5) | 236 (5) | 0 | 110 (4) | 0 |
| Mo(2) | 0 | 36235 (8) | 25000 | 330 (5) | 246 (4) | 254 (5) | 0 | 138 (4) | 0 |
| Cl(1,2) | 664 (1) | 732 (2) | 4004 (1) | 50 (1) | 41 (1) | 29 (1) | 1 (1) | 9 (1) | 9 (1) |
| Cl(3,4) | -1110 (1) | 4383 (2) | 2794 (1) | 40 (1) | 42 (1) | 42 (1) | 7 (1) | 23 (1) | 1 (1) |
| S(1,2) | 1262 (1) | 1139 (2) | 2288 (1) | 46 (1) | 42 (1) | 32 (1) | 13 (1) | 16 (1) | 3 (1) |
| S(3,4) | -629 (1) | 3990 (2) | 828 (1) | 39 (1) | 36 (1) | 28 (1) | 2 (1) | 14 (1) | 5 (1) |
| C(1,5) | 1176 (4) | 1405 (8) | 1116 (4) | 66 (5) | 50 (4) | 31 (4) | 11 (4) | 23 (4) | 4 (4) |
| C(2,6) | 1935 (5) | 1456 (12) | 1002 (6) | 63 (6) | 127 (8) | 61 (6) | 0 (6) | 37 (5) | 13 (6) |
| C(3,7) | 1498 (7) | -586 (9) | 2449 (7) | 167 (11) | 46 (5) | 68 (6) | 35 (6) | 59 (7) | 12 (5) |
| C(4,8) | 1034 (9) | -1493 (12) | 1885 (8) | 229 (15) | 70 (8) | 87 (9) | -9 (10) | 36 (10) | 12 (7) |
| C(9,13) | -520 (4) | 5738 (7) | 645 (5) | 58 (5) | 43 (4) | 41 (4) | -4 (4) | 19 (4) | 13 (4) |
| C(10,14) | -848 (5) | 6702 (8) | 1153 (6) | 86 (6) | 41 (5) | 66 (6) | 4 (4) | 34 (5) | 5 (4) |
| C(11,15) | -1644 (4) | 3837 (8) | 501 (5) | 43 (4) | 55 (5) | 46 (4) | -5 (4) | 12 (4) | 6 (4) |
| C(12,16) | -2061 (5) | 4148 (11) | -468 (6) | 57 (6) | 122 (9) | 53 (5) | -14 (6) | -5 (4) | 10 (6) |

Table 4. H-*atom coordinates* ($\times 10^3$) *and assumed isotropic B values* ($\text{Å}^2$)

| Bonded to | $x$ | $y$ | $z$ | $B$ |
|---|---|---|---|---|
| C(1,5) | 86 | 67 | 73 | 5 |
|  | 89 | 226 | 89 | 5 |
| C(2,6) | 189 | 166 | 37 | 6 |
|  | 220 | 57 | 117 | 6 |
|  | 226 | 214 | 141 | 6 |
| C(3,7) | 202 | −72 | 240 | 10 |
|  | 157 | −84 | 309 | 10 |
| C(4,8) | 123 | −242 | 200 | 12 |
|  | 94 | −127 | 123 | 12 |
|  | 51 | −148 | 197 | 12 |
| C(9,13) | 4 | 594 | 81 | 5 |
|  | −74 | 594 | −1 | 5 |
| C(10,14) | −141 | 656 | 97 | 6 |
|  | −75 | 764 | 102 | 6 |
|  | −61 | 655 | 182 | 6 |
| C(11,15) | −178 | 290 | 64 | 5 |
|  | −184 | 444 | 89 | 5 |
| C(12,16) | −262 | 405 | −64 | 6 |
|  | −194 | 508 | −62 | 6 |
|  | −187 | 352 | −86 | 6 |

Table 5. *Bond distances* ($\text{Å}$) *and angles* ($°$), *model* ($d$)

| | | | | | |
|---|---|---|---|---|---|
| Mo(1)–Mo(2) | 2·144 (1) | | | | |
| Mo(1)–Cl(1,2) | 2·404 (2) | | Mo(2)–Cl(3,4) | 2·409 (2) | |
| –S(1,2) | 2·530 (2) | | –S(3,4) | 2·525 (2) | |
| S(1,2)–C(1,5) | 1·822 (8) | | S(3,4)–C(9,13) | 1·811 (8) | |
| –C(3,7) | 1·794 (11) | | –C(11,15) | 1·803 (8) | |
| C(1,5)–C(2,6) | 1·496 (13) | | C(9,13)–C(10,14) | 1·519 (12) | |
| C(3,7)–C(4,8) | 1·364 (18) | | C(11,15)–C(12,16) | 1·495 (13) | |

| | | | | |
|---|---|---|---|---|
| Mo(2)–Mo(1)–Cl(1,2) | 108·8 (1) | Mo(1)–Mo(2)–Cl(3,4) | 108·6 (1) |
| –S(1,2) | 98·3 (1) | –S(3,4) | 98·4 (1) |
| Cl(1,2)–Mo(1)–Cl(1,2)$^i$ | 142·3 (1) | Cl(3,4)–Mo(2)–Cl(3,4)$^i$ | 142·9 (1) |
| –S(1,2) | 83·7 (1) | –S(3,4) | 91·6 (1) |
| –S(1,2)$^i$ | 90·9 (1) | –S(3,4)$^i$ | 83·0 (1) |
| Mo(1)–S(1,2)–C(1,5) | 110·5 (3) | Mo(2)–S(3,4)–C(9,13) | 105·8 (3) |
| –C(3,7) | 108·5 (4) | –C(11,15) | 111·0 (3) |
| S(1,2)–C(1,5)–C(2,6) | 111·6 (6) | S(3,4)–C(9,13)–C(10,14) | 116·9 (6) |
| –C(3,7)–C(4,8) | 119·2 (9) | –C(11,15)–C(12,16) | 114·1 (6) |

(i) at $\bar{x}, y, \frac{1}{2} - z$.

of about 0·9 e $\text{Å}^{-3}$ at a position ($x = 0·18$, $y = -0·15$, $z = 0·20$) which suggests an alternative site for the methyl group C(4,8) with an occupancy factor of 10% or so; the large and anisotropic $U$'s for C(3,7) suggest that this atom might accompany C(4,8) in any such disorder. Final parameters are given in Tables 3 and 4, and bond distances and angles in Table 5.

The only surprises are associated with ethyl group C(3,7)–C(4,8) which, as noted above, seems to be disordered and in any event shows large $U$ values.

Except in this group, the ranges of chemically equivalent distances are well explained by the e.s.d.'s and are considerably smaller than reported for the earlier, *Cc*, refinement (CF). The relatively large range of values reported by CF is typical for refinement of a centrosymmetric structure in a noncentrosymmetric space group (Marsh & Schomaker, 1979), and reflects the problem of near-singularity (Ermer & Dunitz, 1970).

Finally, it cannot be asserted that the structure of this compound is definitely centrosymmetric; there may be small deviations from a centrosymmetric structure which could only be detected by examining the weak reflections, and these were not included in the data set. What can be asserted is that, on the basis of the 1567 reflections making up the final data set, there is no reason to choose the lower-symmetry space group *Cc*. Furthermore, the fact that these 1567 intensities are so well explained by the *C*2/*c* structure is clear proof that the Howells, Phillips & Rogers test was misleading in this instance, and that statistical indications of noncentrosymmetry are not to be trusted if the weak reflections have been discarded.

### References

COTTON, F. A. & FANWICK, P. E. (1980). *Acta Cryst.* B36, 457–459.

ERMER, O. & DUNITZ, J. (1970). *Acta Cryst.* A26, 163.

GAGNÉ, R. R., McCOOL, M. W. & MARSH, R. E. (1980). *Acta Cryst.* B36, 2420–2422.

HAMILTON, W. C. (1965). *Acta Cryst.* 18, 502–510.

HIRSHFELD, F. L. & RABINOVICH, D. (1973). *Acta Cryst.* A29, 510–513.

HOWELLS, E. R., PHILLIPS, D. C. & ROGERS, D. (1950). *Acta Cryst.* 3, 210–214.

HUGHES, E. W. (1941). *J. Am. Chem. Soc.* 63, 1737–1752.

KARLE, I. L., DRAGONETTE, K. S. & BRENNER, S. A. (1965). *Acta Cryst.* 19, 713–716.

MARSH, R. E. & SCHOMAKER, V. (1979). *Inorg. Chem.* 18, 2331–2336.

ROGERS, D., STANLEY, E. & WILSON, A. J. C. (1955). *Acta Cryst.* 8, 383–393.

SCHOMAKER, V. & MARSH, R. E. (1979). *Acta Cryst.* B35, 1933–1934.

WILSON, A. J. C. (1942). *Nature (London),* 150, 151–152.