*Acta Cryst.* (1979). A**35**, 497–499

# On the Significance of 'Non-Significant' Reflexions

BY LARS ARNBERG, SVEN HOVMÖLLER AND SVEN WESTMAN

*Departments of Inorganic and Structural Chemistry, Arrhenius Laboratory, University of Stockholm, S-*106 91, *Stockholm, Sweden*

(*Received* 7 *August* 1978; *accepted* 18 *January* 1979)

## Abstract

The information content of so-called 'unobserved', or 'non-significant' reflexions obtained by diffractometer measurements has been experimentally studied, using data and input parameters for three previously determined crystal structures. It can be concluded that even those reflexions having the largest $\sigma(I)/I$ values contain significant information about both thermal and positional atomic parameters. It has been possible to carry out an acceptable refinement of one structure on the basis of only the set of structure factors discarded by the original investigator. Inclusion of the weakest reflexions in the refinements lowers the standard deviations, and so should be standard practice.

## Introduction

A common habit among crystallographers is to discard low-intensity – 'unobserved', or 'non-significant' – X-ray reflexions from crystal structure refinement on the sole criterion that the relative uncertainties, $\sigma(I)/I$, in the measured intensities are larger than some specified value. This practice, transferred from film to diffractometer measurements, lowers the conventional reliability index, $R$, but is theoretically unjustified, as pointed out earlier by several workers (*vide infra*).

It can be argued that, in order for all measurements to have equal importance in a least-squares refinement of a set of structural parameters, $p_i$, it would be desirable that $\mathrm{d}F = \sum (\delta F/\delta p_i)\,\mathrm{d}p_i \simeq F_o - F_c$ be known with constant experimental accuracy. This implies, of course, that $\sigma(F)$, rather than $\sigma(F)/F$ or $\sigma(I)/I$, should be considered the proper criterion for judging data quality; which is indeed recognized by rational weighting schemes for diffractometer data (see *e.g.* discussion by Killean & Lawrence, 1969).

With proper weighting, then, it should be of advantage to include in a refinement all reflexions measured by diffractometry, whether they be strong or almost vanishing in the background. In particular, it has been demonstrated by Hirshfeld & Rabinovich (1973) that selective rejection of low-intensity reflexions introduces a bias into the data, eventually leading to underestimation of thermal parameter and scale factor (in a one-atom structure).

We are not aware, however, of any experimental studies on the influence of very weak reflexions upon the determination of atomic positional parameters and their estimated standard deviations. We have therefore studied this kind of significance of weak reflexions systematically, following the observation that inclusion of 36 originally rejected reflexions, with $\sigma(I)/I > 0{\cdot}16$, in the refinement of the $Cu_9Al_4$ structure actually resulted in a decrease of several positional-parameter e.s.d.'s after the terminal cycle using 114 'reliable' independent reflexions.

The present investigation concerns the effects of using various subsets of available structure factor data for three different substances, namely:

(*a*) $Cu_9Al_4$, which is an alloy with cubic symmetry and, hence, rather few structural parameters, yielding few independent reflexions (Arnberg & Westman, 1978);

(*b*) $C_6H_8O_4Se$, an organic compound with a structure dominated by a heavy atom in a special position (Dahlén & Lindgren, 1979);

(*c*) $C_{30}H_{18}O_2$, an organic substance without heavy-atom influence (Becker, Karlsson & Pilotti, 1976).

## Experimental

Crystal data for the three compounds investigated are given in Table 1. In all cases constant-time counting was carried out on a PW 1100 automatic diffractometer, with monochromatized Cu $K\alpha$ radiation. The least-squares refinement programs *LALS* and *UPALS* were employed, with the structure-factor amplitudes and scattering-factor tables provided by the original investigators. The programs mentioned, which are those easily available to us, base the structure refinement on $|F|$ values, rather than $|F|^2$ which would be theoretically more satisfying (Hirshfeld & Rabinovich, 1973). We have, however, in this investigation obtained extremely few $|F_o|^2 < 0$; this practical difficulty pertaining to the inclusion of all measured structure factors has thus been negligible.

0567-7394/79/030497-03$01.00

Table 1. *Crystal data*

| Compound | $Cu_9Al_4$ | $C_6H_8O_4Se$ | $C_{30}H_{18}O_2$ |
|---|---|---|---|
| Space group | $P\bar{4}3m$ | $P4_2/n$ | $Fdd2$ |
| $a$ (Å) | 8·7068 | 8·690 | 29·525 |
| $b$ (Å) | | | 34·118 |
| $c$ (Å) | | 10·084 | 3·971 |
| Atoms per asymmetric unit | 8 | 11 (+8 H) | 16 (+9 H) |
| Crystal size ($\mu m^3$) | $\frac{4}{3}\pi 50^3$ | | 60 × 100 × 450 |

We have utilized unit weights throughout the present work, in order to facilitate comparison of results. It may be argued* that the 'standard deviations' yielded by this procedure are not really standard and may thus not be strictly comparable among the various refinements. The weight analyses obtained show very little trend, however; with all reflexions included $\overline{w\Delta^2}$ maximally varied between 0·9 and 1·2 over the entire $|F|$ range for the (*a*) and (*c*) structures. To a reasonable degree, then, $\sigma(F)$ actually seems to be independent of $|F|$, as indeed it should be in a constant-time experiment, disregarding the influence of Lp and absorption factors. The data for (*b*) $C_6H_8O_4Se$ are of lower quality in this respect ($0{\cdot}4 < \overline{w\Delta^2} < 2{\cdot}5$), but still yield comparable results.

The most obvious subdivision of data would be into sets with high and low values of $|F|$. In order to avoid systematic coupling of $|F|$ and $\theta$ sorting we have instead partitioned our data according to the values of the normalized structure factors, $|E|$. For comparison with refinements based on the data subsets thus constructed, results derived from high-$\theta$ and low-$\theta$ data as well as from the complete reflexion material have been included in Table 2.

Several different calculations were carried out with each data set. Among these was a complete refinement including thermal parameters, based on the total data material, which we have considered to yield 'standard' parameter values, designated $\xi$ for positional coordinates in the following. Since thermal parameters are quite sensitive to selection of data, their standard values have been held fixed in other runs testing the effects of more or less accurate input coordinates. In these tests, inaccuracies were introduced by rounding-off standard values. In order to check for possible bias introduced by the fixed, preselected $B$ values, we ran a few

* As pointed out to us by one of the referees.

Table 2. *Results of the structure refinements*

Standard positional parameter values are designated $\xi$; input and refined positional parameter values, $x_{in}$ and $x_{ref}$, respectively; hydrogen atoms have been included in the refinements but not in the calculations of coordinate deviation and e.s.d. averages.

| Compound | Type of reflexions | No. of reflexions | $\overline{\lvert x_{in}-\xi\rvert}$ ($10^{-3}$ Å) | $\overline{\lvert x_{ref}-\xi\rvert}$ ($10^{-3}$ Å) | $\overline{\sigma(x_{ref})}$ ($10^{-3}$ Å) | $R$ (%) | $\bar{B}$ (Å²) | Designation in Fig. 1 |
|---|---|---|---|---|---|---|---|---|
| $Cu_9Al_4$ | All | 150 | 3·6 | 0·0 | 3·7 | 4·6 | 0·88 | |
| | $\sin\theta/\lambda < 0{\cdot}461$ | 77 | 3·6 | 3·5 | 6·1 | 3·4 | 1·12 | |
| | $\sin\theta/\lambda > 0{\cdot}461$ | 73 | 3·6 | 4·0 | 5·2 | 4·9 | 0·61 | |
| | $\lvert E\rvert \le 0{\cdot}56$ | 100 | 0·0 | 4·4 | 6·6 | 9·8 | 1·28 | |
| | | | 8·7 | 6·2 | 6·6 | 10·0 | | |
| | | | 32·0 | 23·8 | 10·6 | 19·5 | | |
| | $\lvert E\rvert > 0{\cdot}29$ | 100 | 0·0 | 3·9 | 4·5 | 2·7 | 0·86 | |
| | | | 8·7 | 3·5 | 3·9 | 2·8 | | |
| | | | 32·0 | 3·5 | 3·9 | 2·8 | | |
| $C_6H_8O_4Se$ | All | 823 | 3·6 | 0·0 | 4·9 | 5·0 | | |
| | $\sin\theta/\lambda < 0{\cdot}465$ | 416 | 25·6 | 3·1 | 8·7 | 5·0 | | |
| | $\sin\theta/\lambda > 0{\cdot}465$ | 407 | 25·6 | 4·3 | 5·8 | 5·2 | | |
| | $\lvert E\rvert \le 0{\cdot}45$ | 414 | 3·6 | 6·3 | 7·2 | 19·2 | | |
| | | | 25·6 | 5·8 | 7·2 | 20·3 | | |
| | | | 211· | 109· | 23·3 | 50· | | |
| | $\lvert E\rvert > 0{\cdot}45$ | 407 | 3·6 | 4·7 | 6·6 | 2·7 | | |
| | | | 25·6 | 4·4 | 6·5 | 2·8 | | |
| | | | 211· | 10·6 | 7·6 | 3·1 | | |
| $C_{30}H_{18}O_2$ | All | 1031 | 9·1 | 0·0 | 9·2 | 10·6 | | *AS* = All, standard |
| | $\lvert E\rvert \le 0{\cdot}75$ | 525 | 12·0 | 9·2 | 12·5 | 20·9 | | |
| | $\lvert E\rvert > 0{\cdot}75$ | 506 | 12·0 | 11·4 | 15·1 | 7·4 | | |
| | $\sigma(I)/I > 0{\cdot}30$ | 338 | 0·0 | 17·8 | 21·8 | 46·0 | | *WS* = Weak, standard |
| | | | 12·0 | 16·6 | 21·6 | 43·6 | | |
| | | | 50·6 | 19·3 | 21·6 | 45·7 | | *WD* = Weak, displaced |
| | $0{\cdot}30 > \sigma(I)/I > 0{\cdot}08$ | 360 | 0·0 | 13·0 | 16·1 | 12·8 | | *MS* = Medium, standard |
| | | | 12·0 | 13·0 | 16·2 | 13·2 | | |
| | | | 50·6 | 13·1 | 16·3 | 12·8 | | *MD* = Medium, displaced |
| | $\sigma(I)/I < 0{\cdot}08$ | 333 | 0·0 | 11·8 | 16·0 | 5·6 | | *SS* = Strong, standard |
| | | | 12·0 | 9·7 | 16·9 | 5·7 | | |
| | | | 50·6 | 11·5 | 15·7 | 5·6 | | *SD* = Strong, displaced |

comparison refinements with variable $B$'s in the cases with data subsets large enough to allow parallel refinement of positional and thermal parameters. The differences between corresponding $x$ values in the two resulting parameter sets were smaller than $\frac{1}{2}$ e.s.d. The e.s.d. values themselves came out about 20% smaller when the $B$ values were held fixed than when they were allowed to vary.

The input and the outcome of the various refinements is summarized in Table 2. Since a greater number of structure factors were available for $C_{30}H_{18}O_2$ than for the other compounds, these diffraction data have been sorted into three further groups, *viz* small, medium and large $\sigma(I)/I$ values. Each of these groups has then been used with standard, and with severely displaced coordinates as input. Fig. 1 gives a survey of these results.

## Conclusions

The standard coordinate values obtained from complete refinements on the basis of all reflexions, measured by diffractometry, have been accepted as the most nearly correct ones. Comparing the other results with this standard, one may make the following observations:

(*a*) Both large- and small-$|E|$ refinements (as well as high- and low-$\theta$) yield positional coordinates differing in each case by less than one e.s.d. from the standard values, when these latter are used as input to the first refinement cycle.

(*b*) Refinements starting with rounded standard coordinates as input converge to the same final results as in (*a*) – the large-$|E|$ refinements most rapidly. Notable exceptions are the small-$|E|$ refinements started with severely rounded-off parameter values; these runs have obviously drifted to false local minima.

(*c*) Both large- and small-$|E|$ refinements (as well as high- and low-$\theta$) yield larger e.s.d. values than the standard refinement, even though a consideration of the $R$ value alone would seem to indicate that the large-$|E|$ refinement is better and the small-$|E|$, worse. This effect is most pronounced with the least well determined structure, $C_{30}H_{18}O_2$.

(*d*) Even though we are here dealing with many-atom structures, the large-$|E|$ refinements yield underestimated and the small-$|E|$, somewhat overestimated thermal parameters, in accordance with the observation by Hirshfeld & Rabinovich (1973).

(*e*) It must be particularly emphasized that a refinement of $C_{30}H_{18}O_2$, based on one third of the total structure factor material, with $\sigma(I)/I > 0{\cdot}30$, which was originally discarded, converged to acceptable coordinate values.

Fig. 1 attempts to show how the several results based on more or less 'significant' reflexion sets for $C_{30}H_{18}O_2$ are related, and to illustrate the influence of the weak reflexions on the refinement using the full material.

The convergence of the several refinements is illustrated by the proximity of related results (*cf.* Table 2 for symbols):

$$\overline{|x_{ref}^{S,S} - x_{ref}^{S,D}|} = 0{\cdot}7 \times 10^{-3}\ \text{Å}, \quad \sigma(x_{ref}^{S,}) = 16 \times 10^{-3}\ \text{Å};$$
$$\overline{|x_{ref}^{M,S} - x_{ref}^{M,D}|} = 3{\cdot}6 \times 10^{-3}\ \text{Å}, \quad \sigma(x_{ref}^{M,}) = 16 \times 10^{-3}\ \text{Å};$$
$$\overline{|x_{ref}^{W,S} - x_{ref}^{W,D}|} = 16 \times 10^{-3}\ \text{Å}, \quad \sigma(x_{ref}^{W,}) = 22 \times 10^{-3}\ \text{Å};$$
$$|x_{in}^{D} - x_{ref}^{D,}| \simeq 50 \times 10^{-3}\ \text{Å}, \quad \text{in all cases.}$$

We think it can be concluded from these results that traditionally-discarded reflexions in general do contain significant information not only about thermal, but also about positional atomic parameters. No set of reflexions should therefore be excluded from a least-squares refinement on the pretext of having too large $\sigma(I)/I$ values.

We are much indebted to Dr Birgitta Dahlén and Dr Bengt Karlsson, who provided us with diffraction data for two of the compounds studied. The investigation has been carried out with financial support from the Swedish Natural Science Research Council.

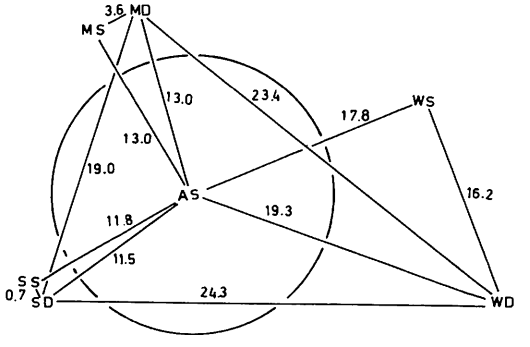


Fig. 1. A two-dimensional picture of the mean deviations, $\overline{|x_{ref} - \xi|}$, obtained from refinements of the $C_{30}H_{18}O_2$ structure, based on data sorted according to $\sigma(I)/I$ (for symbols, see Table 2). The mean deviations $\overline{|x_{ref}^{I} = x_{ref}^{J}|}$ are also represented, approximately to scale. The radius of the circle equals the mean e.s.d. of the standard coordinates, $\xi$. The mean e.s.d. of each $x_{ref}^{I}$ is always larger than $\overline{|x_{ref}^{I} - \xi|}$. It should be observed that $\overline{|x_{ref}^{I} - x_{ref}^{J}|} < \sigma(x_{ref})$ when $i$ and $j$ denote refinements based on the same data subset, with different input parameters but $\overline{|x_{ref}^{I} - x_{ref}^{J}|} > \sigma(x_{ref})$ when the refinements are based on different data subsets. The figure is intended to convey an impression of the fact that an average of $x_{ref}^{S}$, $x_{ref}^{M}$ and $x_{ref}^{W}$ may approximately coincide with $\xi$, but that this is not true for an average of $x_{ref}^{S}$ and $x_{ref}^{M}$.

### References

ARNBERG, L. & WESTMAN, S. (1978). *Acta Cryst.* A**34**, 399–404.

BECKER, H.-D., KARLSSON, B. & PILOTTI, A.-M. (1976). *Chem. Phys. Lett.* **44**, 589–593.

DAHLÉN, B. & LINDGREN, B. (1979). To be published.

HIRSHFELD, F. L. & RABINOVICH, D. (1973). *Acta Cryst.* A**29**, 510–513.

KILLEAN, R. C. G. & LAWRENCE, J. L. (1969). *Acta Cryst.* B**25**, 1750–1752.