Acta Cryst. (1978). A**34**, 860–863

Positional Errors in (Partial) Structure Models; Their Influence on the Residual R_2

By W. Van de Mieroop and A. T. H. Lenstra

University of Antwerp (UIA), Department of Chemistry, Universiteitsplein 1, *B*-2610 *Wilrijk, Belgium*

(*Received* 7 *March* 1978; *accepted* 2 *May* 1978)

Algebraic expressions are evaluated, from which the influence of small discrepancies between atomic positions in a structure model and the corresponding atomic sites in the observed structure on R_2 can be calculated. Theory and experiment are in good agreement. It is demonstrated that because of the effect of small positional errors, R_2 as a function of the size of a structure model does not necessarily show a monotonic decrease. The path of R_2 may even contain additional minima.

Introduction

The residual R_2 is defined as:

$$R_2 = [\sum (I_o - I_c)^2]/\sum I_o^2. \qquad (1)$$

Function values of R_2 can be enumerated prior to an actual structure analysis using intensity statistics (Lenstra, 1974). Wilson (1976) and Parthasarathy & Parthasarathi (1972, 1974) have demonstrated that, *inter alia*, the influence of anomalous dispersion, space group symmetry and even local symmetry additional to that of the space group can be taken into account.

The impact of small discrepancies between atomic positions in the structure itself and the corresponding positions in the tentative structure model will be studied in this paper.

Let the exact position be $\mathbf{r}_j$ and the position found from a tentative electron density function be $\mathbf{r}_j + \Delta\mathbf{r}_j$. In the original nomenclature of Wilson (1969) the atomic site is incorrect if $\langle \cos 2\pi\, \mathbf{H}.\Delta\mathbf{r}_j \rangle = 0$, in which the angle brackets represent an averaging over all *hkl* values. If an atomic site is correct this average is 1. Intermediate values will be found in the presence of small errors in atomic coordinates.

In this paper the influence of one slightly misplaced atom is calculated for (in)complete structure models. For brevity we confine ourselves to the space groups $P1$ and $P\bar{1}$.

Also, experimental values are enumerated using the experimental intensity data of ammonium hydrogen malate (Versichel, Van de Mieroop & Lenstra, 1978). A good fit between theory and experiment is found. The data also give evidence that the above used definition of an incorrect atomic position is necessary, but not sufficient.

So far, theoretical calculations – taking into account the presence of unequal atoms in the structure, pseudo-symmetry, *etc.* – have indicated R_2 to have only one single minimum, *viz* for a complete and correct structure model. Wilson (1977) states that the probability of, for example, false minima increases with the size of fluctuations in F, be it F_o or F_c. In the last section we confine ourselves to fluctuations in F_c values only. It is demonstrated that positional errors (of the magnitude of the resolution of the Fourier function) can produce additional minima in the curvature of R_2 *versus* the size of the structure model.

R_2 index as a function of one slightly misplaced atom

To calculate R_2 (1) is rewritten as:

$$R_2 = \langle (I_o - I_c)^2 \rangle / \langle I_o^2 \rangle \qquad (2)$$

in which the angle brackets represent the averaging over the complete set of available intensity data.

(*a*) *The space group P*1

The correct, observed structure factor F_o can be written as:

$$F_o = F_n \exp i\alpha + f_{n+1} \exp i\theta + F_u \exp i\chi \qquad (3)$$

in which F_n is the structure amplitude of the known n-atom structure fragment, f_{n+1} is the scattering power of the tentative atom added to this model. In an N-atom structure F_u is the structure amplitude of the remaining, unknown $(N - n - 1)$ atoms.

Let the n-atom model be exactly correct; only a positional error is present in the location of the $(n + 1)$th atom. Then we find:

$$F_c = F_n \exp i\alpha + f_{n+1} \exp i(\theta + \Delta), \qquad (4)$$

where the phase angle Δ is related to the positional error. Because $I = FF^*$, we obtain

$$\begin{aligned}(I_o - I_c) = {} & F_u^2 \\ & + 2f_{n+1} F_n[\cos(\alpha - \theta) - \cos(\alpha - \theta - \Delta)] \\ & + 2f_{n+1} F_u \cos(\theta - \chi) + 2F_n F_u \cos(\alpha - \chi).\end{aligned}$$

In order to obtain R_2 the H-averaged value of $(I_o - I_c)^2$ has to be calculated by averaging each of the factors in the expression for $(I_o - I_c)^2$. Since

$$\langle 2\sin^2 \Delta/2\rangle = \langle[\cos(\alpha-\theta) - \cos(\alpha-\theta-\Delta)]^2\rangle$$

we find

$$R_2 = (\langle F_u^4\rangle + 8\langle f_{n+1}^2\rangle\langle F_n^2\rangle\langle \sin^2 \Delta/2\rangle + 2\langle f_{n+1}^2\rangle\langle F_u^2\rangle + 2\langle F_n^2\rangle\langle F_u^2\rangle)/\langle F_o^4\rangle. \quad (5)$$

If the tentative atom is located at its proper position $\langle\sin^2(\Delta/2)\rangle = 0$ then R_2 reduces to:

$$R_2 = (\langle F_u^4\rangle + 2\langle f_{n+1}^2\rangle\langle F_u^2\rangle + 2\langle F_n^2\rangle\langle F_u^2\rangle)/\langle F_o^4\rangle.$$

For an atom to be badly misplaced the phase angle $\Delta/2$ has to change by at least π radians [since the periodicity of $\sin^2(\Delta/2)$ is π radians; this is equivalent to Wilson's definition of an incorrect atomic site]. In this situation $\langle\sin^2(\Delta/2)\rangle = \frac{1}{2}$ and (5) becomes:

$$R_2 = (\langle F_u^4\rangle + 4\langle f_{n+1}^2\rangle\langle F_n^2\rangle + 2\langle f_{n+1}^2\rangle\langle F_u^2\rangle + 2\langle F_n^2\rangle\langle F_u^2\rangle)/\langle F_o^4\rangle.$$

The two given extremes of (5) are exactly equal to those derived by Lenstra (1974).

(*b*) *The space group* $P\bar{1}$

A small error in the position of one atom now also occurs in that of a second which is related to the trial atom by inversion symmetry. Analogous to the basic equations (3) and (4) we have

$$\mathbf{F}_o = \mathbf{F}_n + 2f_{n+1}\cos\theta + \mathbf{F}_u$$
$$\mathbf{F}_c = \mathbf{F}_n + 2f_{n+1}\cos(\theta + \Delta).$$

It is easily found that the numerator of the R_2 index is given by:

$$\begin{aligned}\langle(I_o - I_c)^2\rangle = {} & 8\langle f_{n+1}^4\rangle\langle\sin^2\Delta\rangle \\ & + 32\langle f_{n+1}^2\rangle\langle F_n^2\rangle\langle\sin^2\Delta/2\rangle \\ & + \langle F_u^4\rangle + 8\langle f_{n+1}^2\rangle\langle F_u^2\rangle \\ & + 4\langle F_n^2\rangle\langle F_u^2\rangle. \qquad (6)\end{aligned}$$

In the case $\langle\sin^2\Delta\rangle = \langle\sin^2(\Delta/2)\rangle = \frac{1}{2}$ the added, symmetry-related atoms are badly misplaced and R_2 is given by:

$$R_2 = (4\langle f_{n+1}^4\rangle + 16\langle f_{n+1}^2\rangle\langle F_n^2\rangle + \langle F_u^4\rangle + 8\langle f_{n+1}^2\rangle\langle F_u^2\rangle + 4\langle F_n^2\rangle\langle F_u^2\rangle)/\langle F_o^4\rangle.$$

When the two trial atoms are located at the correct positions ($\langle\sin^2\Delta\rangle = \langle\sin^2(\Delta/2)\rangle = 0$) the residual R_2 reduces to:

$$R_2 = (\langle F_u^4\rangle + 8\langle f_{n+1}^2\rangle\langle F_u^2\rangle + 4\langle F_n^2\rangle\langle F_u^2\rangle)/\langle F_o^4\rangle.$$

Again the two extremes of (6) are exactly equal to the equations of Lenstra (1974).

The residual R_2 can be written in terms of a Patterson vector map, namely

$$R_2 = \textstyle\int_V (P_o - P_c)^2 \mathrm{d}V / \int_V P_o^2 \mathrm{d}V.$$

In the expressions of R_2 in the space group $P\bar{1}$ we have in addition to a term containing $\langle\sin^2(\Delta/2n\rangle$ — as in $P1$ –, also a simple term with $\langle\sin^2\Delta\rangle$. In the nomenclature of the Patterson functions the latter term is linked to the two single vectors between the two symmetry-related atoms. If this atomic position is incorrect by $\Delta\mathbf{r}_j$, the single vectors are incorrect by $2\Delta\mathbf{r}_j$. The interatomic vectors between the atoms not related by symmetry are incorrect by $\Delta\mathbf{r}_j$ only.

Experimental

The previous theory is checked using the experimental intensity data and the refined structure of ammonium hydrogen malate ($P2_12_12_1$, final R value = 0·026). Apart from the hydrogen atoms the unit cell contains 4 nitrogen, 16 carbon and 20 oxygen atoms. To simplify the calculations of the corresponding R_2 values we regarded the compound as an equal-atom structure. For a non-centrosymmetric, equal-atom structure R_2 decreases linearly with the number of atoms in the known structure fragment. This and the experimental R_2 values are illustrated in Fig. 1, showing the validity of our simplifying assumption.

Since high-order reflections are more sensitive towards small errors in positional coordinates than the low-order reflections, we calculated R_2 values using two θ ranges, namely $17° < \theta < 21{\cdot}5°$ (set A) and $21{\cdot}5° < \theta < 24{\cdot}9°$ (set B). Both sets contain approximately 200 independent reflections. As additional parameters we chose the size of the known structure fragment and the amount of misplacement of the four symmetry-related atoms added to this fragment. Typical results are summarized in Table 1, in which ΔR_2 is the increase of the R_2 value due to the misplacement of the added atoms.

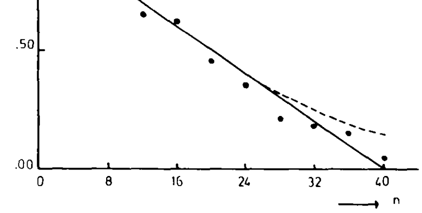

Fig. 1. The straight line represents the theoretical curve of R_2. The dots give the experimental R_2 values for ammonium hydrogen malate. The dashed line represents the general path of experimental R_2 values near the end of a structure analysis.

Table 1. *Theoretical and experimental values of ΔR_2 for sets A and B of ammonium hydrogen malate (see text)*

N is the number of correctly placed atoms in the known structure fragment. The error E in atomic position is given in Å.

	$100 \times \Delta R_2$				
	Set A		Set B		
E	Theory	Experiment	Theory	Experiment	N
0·040	0·05	0·05	0·07	0·25	4
0·076	0·20	0·25	0·29	0·75	4
0·114	0·44	0·50	0·64	1·40	4
0·153	0·76	0·85	1·15	2·05	4
0·190	1·14	1·30	1·79	2·55	4
0·230	1·57	1·70	2·58	3·05	4
0·040	0·10	0·10	0·50	0·30	8
0·076	0·50	0·35	1·50	1·05	8
0·114	1·10	0·75	2·80	2·10	8
0·153	1·70	1·30	4·10	3·25	8
0·190	2·60	1·95	5·10	4·35	8
0·230	3·40	2·65	6·10	5·45	8
0·076	1·00	1·80	1·10	1·80	16
0·153	4·00	5·00	4·00	4·70	16
0·230	8·00	8·60	8·10	7·90	16

The values of ΔR_2 show that the present theory is very satisfactory. Indeed, high-order reflections show larger ΔR_2 values than low-order reflections for the same misplacement.

Wilson's formulation of a badly misplaced atom is correct for one single isolated atom. However, it is insufficient to reveal the consequences of a structure fragment being known already. The presence of such an influence is clearly seen in Table 1. ΔR_2 for a positional error of 0·23 Å increases from 0·016 to 0·080 when the known structure fragment increases from 4 to 16 atoms.

From equation (5) it can be seen that the larger the known fragment, the larger ΔR_2 for a given misplacement will be. This can be easily illustrated in the Patterson function. When an atom is added to an n-atom model $2n$ vectors (not coinciding with the original maximum) must be accounted for. So the larger the known structure fragment is, the more R_2 will react to small misplacements in the added atoms.

This sensitivity of R_2 towards positional errors is met in the behaviour of the experimental R_2 values (see Fig. 1; dashed line). A quantitative explanation cannot be provided, because fluctuations on I_0 and other defects were not taken into account.

Additional information and conclusions

If in the structure model all atoms contain small positional errors, R_2 will be well above its ideal value. Upon addition of an extra slightly misplaced atom one might be confronted with an increase in R_2 rather than with the expected decrease.

A mathematical formulation of this idea is as complex as describing a Patterson vector map in terms of correlated, but incorrect interatomic functions. A simple example will prove our point. Suppose we translate the refined atomic positions in ammonium hydrogen malate by 0·23 Å in the x direction. Such a shift is of the same magnitude as the normal grid distance used in the Fourier calculation in view of $h_{\max}$. Moreover, this displacement is less than an atomic diameter, and so these positions are refinable in terms of least-squares procedures (Lipson & Cochran, 1953). The behaviour of R_2 for the original model and for the shifted one are listed in Table 2 for the first 20 atoms only, underlining the correctness of our previous statement. It even shows the presence of an additional minimum in the path of R_2 *versus* the number of atoms in the structure model.

Table 2. *R_2 for a correct model (A) and a structure model with small errors on the x coordinate only (B)*

All structure factors in the range $0 < \theta < 30°$ are used in the R_2 calculation.

Number of atoms in the structure model	A	B	Error on x coordinate (Å)
4	0·893	0·894	+0·23
8	0·779	0·805	−0·23
12	0·650	0·701	+0·23
16	0·619	0·733	+0·23
20	0·449	0·728	0·23
24	0·354	0·570	+0·23
28	0·215	0·443	+0·23

In automated structure analysis we use as a criterion (Lenstra, 1974) for the correctness of an atomic position: $[R_2(\text{new}) - R_2(\text{original})] < 0$.

Apparently this criterion does not always hold, especially not with high-order reflections. Fortunately, in automated structure analysis the experimental ΔR_2 values are not as sensitive for positional errors as Table 1 might suggest, because the lowest-order reflections cause a serious damping in the calculated ΔR_2.

In view of this an automated structure analysis based on R_2 criteria must use refined peak coordinates rather than the coordinates of the grid point with the highest electron density. In addition, in an iterative process of Fourier calculation the atomic coordinates of the already accepted atoms should be replaced by the coordinates of their refined maxima. So one avoids as much as possible a cumulation of errors in positional coordinates. Thus the chances of wrongly rejecting a slightly misplaced atomic position using the above mentioned criterion are minimized.

References

LENSTRA, A. T. H. (1974). *Acta Cryst.* A**30**, 363–369.
LIPSON, H. & COCHRAN, W. (1953). *The Crystalline State*, Vol. III, 1st ed., p. 250. London: Bell.
PARTHASARATHY, S. & PARTHASARATHI, V. (1972). *Acta Cryst.* A**28**, 426–432.
PARTHASARATHY, S. & PARTHASARATHI, V. (1974). *Acta Cryst.* A**30**, 310–315.
VERSICHEL, W., VAN DE MIEROOP, W. & LENSTRA, A. T. H. (1978). *Acta Cryst.* B**34**, 2643–2645.
WILSON, A. J. C. (1969). *Acta Cryst.* B**25**, 1288–1293.
WILSON, A. J. C. (1976). *Acta Cryst.* A**32**, 53–56.
WILSON, A. J. C. (1977). *Acta Cryst.* A**33**, 523–524.

Acta Cryst. (1978). A**34**, 863–870

On the Application of Phase Relationships to Complex Structures. XIV.* The Additional Use of Statistical Information in Tangent-Formula Refinement

BY S. E. HULL AND M. J. IRWIN

*Department of Physics, University of York, Heslington, York YO*1 5*DD, England*

(*Received* 22 *February* 1978; *accepted* 9 *May* 1978)

Suitable weighting schemes for use in tangent-formula phase development and refinement are discussed, and a statistically based weighting scheme, which can easily be incorporated in existing computer programs, is proposed. Examples of the use of this method in structure solution and completion are presented for a previously unknown structure and for several structures which had been difficult to solve by other methods. In addition a method is given for subtracting the contribution of heavy atoms from observed $|E|$ values on a statistical basis which is very useful when the presence of such atoms leads to problems in defining the enantiomorph. These methods add greatly to the power of the *MULTAN* computer program.

Introduction

The multisolution method of phase determination using convergence mapping and the tangent formula as embodied in computer programs such as *MULTAN* (Main, 1978) and *SHELX* (Sheldrick, 1976) has been very successful in solving both centrosymmetric and noncentrosymmetric crystal structures containing up to 70 or 80 atoms in the asymmetric unit. Sometimes, however, these programs fail to yield any interpretable molecular fragments. The fact that there exist certain structures which are impossible to solve using the simple tangent formula is well known (see, for example, Lessinger, 1976). A common symptom of this inadequacy in the tangent formula is the tendency for the definition of the enantiomorph to be lost, or indeed never to be defined properly. This may lead either to a trivial solution containing both enantiomorphs or to a 'uranium-type' E map with one large peak and little else. Schenk (1972) has divided the problems caused by enantiomorph definition into two classes. Our experience with *MULTAN* is that such structures are difficult to solve for one of the following reasons.

(1) The enantiomorph is initially defined but upon phase extension or refinement is lost; for example, 3-chloro-1,3,4-triphenylazetidin-2-one (AZET) ($C_{21}H_{16}ClNO$, $Pca2_1$, $Z = 8$) (Colens, Declercq, Germain, Putzeys & Van Meerssche, 1974).

(2) Light-atom structures in space groups such as $C2$, $P2_1$, or $P1$ where it is often difficult to define the enantiomorph; for example, prostaglandin E_2 ($C_{20}H_{32}O_5$, $P1$, $Z = 1$) (Edmonds & Duax, 1974*a*,*b*).

(3) Structures containing one heavy atom where the squaring effect of the tangent formula leads to a trivial solution with associated symmetry about the heavy-atom position; for example, the K salt of alborixin (ALB) ($C_{48}H_{83}O_{14}^-.K^+$, $P2_1$, $Z = 2$) (Alléaume, Busetta, Farges, Gachon, Kergomard & Staron, 1975).

Only in the second of these categories is the enantiomorph not well defined at the start of phase determination: for structures in this group there exist several methods making use of enantiomorph-discriminating triple-phase invariants or quartets (Duax & Hauptman, 1972; Busetta, 1976) or enantiomorph-sensitive quartets (Gilmore, 1977), which have met with limited success.

Groups (1) and (3) consist mainly of structures for which the tangent formula is unstable even when the correct phases are used (Lessinger, 1976). This lack of stability appears to be linked to the fact that in normal tangent refinement no explicit use is made of the cosine invariant information or of Sayre's (1952) equation. The next two sections of this paper outline the development of a simple statistical weighting scheme for

---

* Part XIII: Hull (1978).