conventional methods of computation, it offers a handicap to the use of anisotropic molecules like the uranyl ion to exploit the maximum effects at the absorption edges for solving the phase problem. Thus from a pessimistic point of view it is a setback. We adopt the opposite view: where there is a complication there is the opportunity of sharper, more penetrating methods for extracting information from diffraction experiments. There is much to look for in the exploration of this new region of diffraction optics.

Mrs Helena W. Ruben grew some of the crystals we used. We thank her and Dr Allan Zalkin for cooperation in the structure determinations needed for this work and Dr Paul R. Phizackerley and Professor Keith O. Hodgson for assistance with the apparatus at SSRL. This work was supported by the National Science Foundation under Grant No. CHE-7919532. It used facilities of SSRL with support of the NSF under Contract DMR77-27489 (in cooperation with the Department of Energy) and the Lawrence Berkeley Laboratory, supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Chemical Sciences Division of the US Department of Energy under Contract No. W-7405-Eng-48.

References

Barclay, G. A., Sabine, T. M. & Taylor, J. C. (1965). *Acta Cryst.* **19**, 205–209.

Denning, R. G., Snellgrove, T. R. & Woodwark, D. R. (1979). *Mol. Phys.* **37**, 1109–1143.

Fankuchen, I. (1935). *Z. Kristallogr.* **91**, 473–479.

Hoard, J. & Stroupe, J. (1949). *Nat. Nucl. En. Series Div.* 3, Vol. 2, ch. 2, pp. 15–35.

Krause, M. O. & Oliver, J. H. (1979). *J. Phys. Chem. Ref. Data*, **8**, 329–338.

Phillips, J. C., Cerino, J. A. & Hodgson, K. O. (1979). *J. Appl. Cryst.* **12**, 592–600.

Shubnikov, A. V. (1960). *Principles of Optical Crystallography*, pp. 48, 181. New York: Consultants Bureau.

Templeton, D. H. & Templeton, L. K. (1980). *Acta Cryst.* A**36**, 237–241.

Templeton, D. H., Templeton, L. K., Phillips, J. C. & Hodgson, K. O. (1980). *Acta Cryst.* A**36**, 436–442.

Templeton, L. K. & Templeton, D. H. (1978). *Acta Cryst.* A**34**, 368–371.

Teo, B.-K. & Lee, P. A. (1979). *J. Am. Chem. Soc.* **101**, 2815–2832.

Yang, C. Y., Johnson, K. H. & Horsley, J. A. (1978). *J. Chem. Phys.* **68**, 1001–1005.

Zachariasen, W. H. & Plettinger, H. A. (1959). *Acta Cryst.* **12**, 526–530.

Acta Cryst. (1982). A**38**, 67–70

The Influence of the Elimination of Low-Intensity Data on the Applicability Range of R_2 in Automated Structure Evaluations

By G. H. Petit and A. T. H. Lenstra

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

(*Received* 27 *April* 1981; *accepted* 4 *June* 1981)

Abstract

A description is given of the effect on the residual R_2 caused by a misplacement of a fraction of the atoms in a tentative structure model. New expressions are derived for the space groups $P1$ and $P\bar{1}$ for moments as functions of the threshold a, below which intensity data are omitted. It turns out that the range in which R_2 acts as a discriminator between correct and incorrect models is drastically limited even by low threshold values. Theory and experiment are shown to be in excellent agreement.

Introduction

Automation of a crystal structure analysis requires criteria which discriminate between a correct and an incorrect set of atomic positions. If one decides to use mathematical functions for this purpose, then residual functions are an obvious choice (Lenstra, 1974). In this article we will discuss some properties of R_2, which is defined as

$$R_2 \equiv \frac{\sum\limits_H (E_N^2 - E_n^2\,\sigma_1^2)^2}{\sum\limits_H E_N^4}, \qquad (1)$$

where E_N^2 corresponds to the observed normalized intensities and E_n^2 to the normalized intensities related to the tentative fragment of the structure; σ_1^2 is given by $\sum_{j=1}^n f_j^2/\sum_{j=1}^N f_j^2$.

Let the structure looked for contain N equal atoms and let the tentative structure model contain n atoms ($n \le N$), of which g atoms are correctly located and f atoms are badly misplaced ($g + f = n$). This model is denoted by (g,f). We will derive expressions for R_2 in the general case (g,f) when a threshold a is applied to the data, *i.e.* when all $E_N^2 < a$ are eliminated. The results will be verified against simulated experiments.

Previously (Petit, Lenstra & Van Loock, 1981) we described the behaviour of R_2 as a function of the threshold a for the extreme cases $(n,0)$ and $(0,n)$. It was demonstrated that for situations $(n,0)$ the elimination of up to 70% of the available data hardly influences the expectation value of R_2. At the same time we noted* that $\sigma(R_2)$ did not vary much when E_N^2 values up to $a = 2$ were discarded.

Since the time needed to compute experimental values of R_2 increases linearly with the number of reflections involved, the constancy of R_2 and $\sigma(R_2)$ made it look surprisingly profitable to omit low-intensity reflections in the testing of the reliability of tentative structure models. The new expressions for $R_2[a,(g,f)]$ made it possible to show that the introduction of a threshold limits the range in which our formulation can be applied to find what correct and incorrect models are. So a large amount of computing time can only be saved if one is willing to give up a substantial part of the region in which R_2 operates as a good discriminator function.

Expressions for $R_2(a)$ for the general case (g,f)

R_2 as a function of the threshold a, which will be applied to the observed data only, is given by

$$R_2(a) = \frac{\langle E_N^4\rangle_a + \sigma_1^4\langle E_n^4\rangle_a - 2\sigma_1^2\langle E_N^2E_n^2\rangle_a}{\langle E_N^4\rangle_a}. \qquad (2)$$

The angular brackets indicate averages over a large number of reflections and the subscript a is linked to the threshold value.

The evaluation of the moments indicated in (2) is simple, for a model (g,f), once the relevant moments for $(n,0)$ and $(0,n)$ are known. We will demonstrate this for $\langle E_n^2\rangle_a$.

The normalized intensity E_n^2 is generally given by

$$E_n^2 = \frac{1}{n}\left[\left(\sum_{j=1}^{n} \cos 2\pi\mathbf{H}.\mathbf{r}_j\right)^2 + \left(\sum_{j=1}^{n} \sin 2\pi\mathbf{H}.\mathbf{r}_j\right)^2\right]. \qquad (3)$$

For a subset of atoms we rigorously apply the normalization condition, that is we take

$$E_g^2 = \frac{1}{g}\left[\left(\sum_{j=1}^{g} \cos 2\pi\mathbf{H}.\mathbf{r}_j\right)^2 + \left(\sum_{j=1}^{g} \sin 2\pi\mathbf{H}.\mathbf{r}_j\right)^2\right]. \qquad (4)$$

Consequently we have $\langle E_n^2\rangle = \langle E_f^2\rangle = \langle E_g^2\rangle = 1$, implying that in the point-atom approximation we use a scattering power depending upon the size of the model.

In general we can now write

$$E_n^2 = \frac{g}{n}E_g^2 + \frac{f}{n}E_f^2. \qquad (5)$$

Taking the average value of E_n^2, we get

$$\langle E_n^2\rangle = \frac{g}{n}\langle E_g^2\rangle + \frac{f}{n}\langle E_f^2\rangle. \qquad (6)$$

Similarly, we obtain

$$\langle E_N^2E_n^2\rangle = \frac{g}{n}\langle E_N^2E_g^2\rangle + \frac{f}{n}\langle E_N^2\rangle\langle E_f^2\rangle \qquad (7)$$

$$\langle E_n^4\rangle = \frac{g^2}{n^2}\langle E_g^4\rangle + \frac{f^2}{n^2}\langle E_f^4\rangle + \frac{\alpha gf}{n^2}\langle E_g^2\rangle\langle E_f^2\rangle, \qquad (8)$$

with $\alpha = 4$ or 6 for the space groups $P1$ or $P\bar{1}$, respectively. Analogous formulae were first derived by Parthasarathi & Parthasarathy (1975) for a crystal containing a few heavy atoms in the asymmetric unit.

Application of the threshold to the observed data means that $\langle E_f^2\rangle$ is independent of a. Thus, (6) to (8) yield

$$\langle E_n^2\rangle_a = \frac{g}{n}\langle E_g^2\rangle_a + \frac{f}{n} \qquad (9)$$

$$\langle E_N^2E_n^2\rangle_a = \frac{g}{n}\langle E_N^2E_g^2\rangle_a + \frac{f}{n}\langle E_N^2\rangle_a \qquad (10)$$

$$\langle E_n^4\rangle_a = \frac{g^2}{n^2}\langle E_g^4\rangle_a + \frac{f^2}{n^2}\langle E_f^4\rangle + \frac{\alpha gf}{n^2}\langle E_g^2\rangle_a. \qquad (11)$$

Realizing that moments containing E_f correspond to $(0,f)$ and moments containing E_g correspond to $(g,0)$, we can now calculate all terms in (9) to (11) with the help of the previously (Petit & Lenstra, 1979; Petit, Lenstra & Van Loock, 1981) derived moments for the extreme cases (see Table 1). It should be noted that σ_1^2 of Table 1 should now be taken as g/N, whereas σ_1^2 in (2) always remains n/N.

The present R_2 description was verified against simulated experiments. Some typical values are summarized in Table 2. 'Experimental' $\langle R_2\rangle$ values and their spread s ($s^2 = \langle R_2^2\rangle - \langle R_2\rangle^2$) were calculated as averages over a series of 200 structures. Each single structure contained 100 atoms per unit cell, while the corresponding data set was confined to 2000 reflections. Structures, related (correct) and unrelated (incorrect) models were generated by computer simulations.

To check that 200 structures are sufficient to give stable converged values of $\langle R_2\rangle$ [and of $s(R_2)$] we tested the case (30,30) with 700 structures in the

* A theoretical proof for the observation has been presented by Van Havere & Lenstra (1980).

Table 1. *Relevant moments for* $P1$ *and* $P\bar{1}$ *as functions of the threshold* a *for the two extreme situations* $(n,0)$ *and* $(0,n)$

σ_1^2 is given by n/N and $Q = \sqrt{2a/\pi}\, e^{-a/2}/\mathrm{erfc}(\sqrt{a/2})$. n and N are supposed to be large. If f is small $\langle E_f^4\rangle$ is $(2f^2 - f)/f^2$ in $P1$ and $(3f^2 - 3f)/f^2$ in $P\bar{1}$ (Wilson, 1969).

	$P1$		$P\bar{1}$	
Model	$(n,0)$	$(0,n)$	$(n,0)$	$(0,n)$
Moments				
$\langle E_n^2\rangle_a$	$1 + \sigma_1^2 a$	1	$1 + \sigma_1^2 Q$	1
$\langle E_n^4\rangle_a$	$\sigma_1^4 a^2 + 2\sigma_1^2(2 - \sigma_1^2)a + 2$	2	$3 + \sigma_1^2[6 + \sigma_1^2(a - 3)]Q$	3
$\langle E_N^2 E_n^2\rangle_a$	$\sigma_1^2 a^2 + (1 + \sigma_1^2)a + 1 + \sigma_1^2$	$1 + a$	$1 + 2\sigma_1^2 + [1 + \sigma_1^2(2 + a)]Q$	$1 + Q$
$\langle E_N^2\rangle_a$	$1 + a$		$1 + Q$	
$\langle E_N^4\rangle_a$	$a^2 + 2a + 2$		$3 + (3 + a)Q$	

Table 2. *Comparison of 'experimental'* $\langle R_2\rangle$ *values with the theoretical ones as a function of the threshold* a *for non-centrosymmetric and centrosymmetric structures*

Standard deviations of the experimental $\langle R_2\rangle$ values are shown in parentheses.

Model	(50,50)		(30,30)		(60,0)		(0,60)	
a	$\langle R_2\rangle$	R_2^{theor}	$\langle R_2\rangle$	R_2^{theor}	$\langle R_2\rangle$	R_2^{theor}	$\langle R_2\rangle$	R_2^{theor}
Space group $P1$								
0·0	0·75 (4)	0·748	0·667 (18)	0·668	0·399 (11)	0·400	0·755 (17)	0·757
1·0	0·49 (2)	0·487	0·595 (17)	0·597	0·378 (11)	0·379	0·661 (13)	0·663
2·0	0·45 (2)	0·449	0·63 (2)	0·629	0·379 (15)	0·381	0·710 (16)	0·711
3·0	0·45 (3)	0·452	0·66 (3)	0·662	0·38 (2)	0·385	0·76 (2)	0·760
4·0	0·46 (5)	0·461	0·68 (4)	0·687	0·39 (4)	0·388	0·80 (3)	0·797
5·0	0·46 (8)	0·471	0·70 (7)	0·706	0·39 (5)	0·391	0·82 (5)	0·825
Space group $P\bar{1}$								
0·0	1·00 (6)	0·995	0·84 (2)	0·837	0·482 (17)	0·480	0·96 (2)	0·954
1·0	0·64 (4)	0·642	0·707 (18)	0·706	0·440 (16)	0·439	0·783 (16)	0·784
2·0	0·58 (3)	0·582	0·71 (2)	0·708	0·429 (18)	0·429	0·796 (16)	0·796
3·0	0·56 (4)	0·559	0·72 (3)	0·720	0·42 (2)	0·424	0·819 (17)	0·818
4·0	0·55 (4)	0·549	0·74 (3)	0·731	0·42 (3)	0·420	0·84 (2)	0·839
5·0	0·54 (5)	0·545	0·75 (4)	0·741	0·42 (3)	0·418	0·86 (2)	0·856

averaging procedure. Comparison of these results with those obtained after 200 cycles showed no differences larger than 0·001. From this we derive that 200 structures are sufficient to tabulate the data significantly to the third digit.

Inspection of Table 2 shows a striking agreement between the theoretical R_2 values and their 'experimental' $\langle R_2\rangle$ counterparts. It is also evident that it is only in situations $(n,0)$ that R_2 does not show large variations in function of the threshold a. In all other cases rather large variations occur. It seems, however, overoptimistic to hope that from an inspection of the path of $R_2(a)$ during an actual structure analysis, one would be able to determine the number of incorrect atoms in a tentative model (*e.g.* in a *MULTAN* solution).

Two more remarks are in order commenting on how realistic the 'experimental' data are. Firstly, in the present analysis any incorrectly placed atom is completely, randomly misplaced. Sometimes, tentative atomic positions are generated (*e.g.* by *MULTAN*) which exhibit systematic errors, for instance a geometrically correct fragment at an incorrect location. In this example only the translation parameter is incorrect. Thus E_N and E_n are related in magnitude, but not in the phase angle. The present theory cannot be applied in this case because E_N and E_n are taken as not interrelated when we deal with incorrect settings. Secondly, the perfect agreement between theory and computer-simulated experiment is certainly favoured by the randomness of our test structures. However, for models of the type $(n,0)$ we also calculated $R_2(a)$ using actual structures as basic information. A detailed description is given elsewhere (Petit, Lenstra & Van Loock, 1981). No contradictions in the behaviour of $R_2(a)$ were found between those 'actual' and the present 'random' enumerations. Therefore, we believe that our present results can be regarded as representative for realistic structure data.

Unfortunately, the present theory does not allow us to predict $\sigma(R_2)$. Nevertheless, an 'experimental' $\sigma(R_2)$ can be obtained as $s^2(R_2) = \langle R_2^2 \rangle - \langle R_2 \rangle^2$. Since $s(R_2)$ is stable to the third digit, *i.e.* independent of the number of the structures in the averaging process, we can conclude that $s(R_2)$ should be practically equal to $\sigma(R_2)$, the relevant parameter of the probability density function $P(R_2)$.

Our 'experimental' $\sigma(R_2)$ values do agree with the theoretical values obtained by Van Havere & Lenstra (1980).

The dependency of $\sigma(R_2)$ on a in all cases (g,f) still holds out hope that low-intensity data can be neglected and thus computing time saved, without loss of information. In the next section, we will show that the price to be paid is a substantial shrinkage of the region in which R_2 operates as a good discriminator function.

The influence of the threshold *a* on the applicability of R_2

The heavy-atom technique has been successfully automated (Lenstra, 1974; Van de Mieroop, 1979) to handle R_2 as a discriminator function. In this routine the model of the structure is enlarged by adding the atoms one by one to a starting model. If addition of an atom increases R_2 with respect to the previous model, the new atom is regarded as incorrect. This definition of an incorrectly located atom is algebraically given by

$$\frac{R_2(g,1)}{R_2(g,0)} \geq 1 \qquad (12)$$

for the non-centrosymmetric space group $P1$ and by

$$\frac{R_2(g,2)}{R_2(g,0)} \geq 1 \qquad (13)$$

for space group $P\bar{1}$. Condition (13) simply reflects that it is impossible to add one atom to the model without the introduction of its symmetry-related atom.

After substitution of the appropriate values in (2), (9), (10) and (11) and after some manipulation, (12) can be rewritten as

$$4ag^2 + 4Ng - 2N^2 a - 2N^2 + N \geq 0. \qquad (14)$$

Similarly, one obtains from (13):

$$4Qg^2 + 4Ng - \tfrac{4}{3}N^2 Q - \tfrac{4}{3}N^2 + 2N \geq 0 \qquad (15)$$

Table 3. $\varepsilon(a)$ *values for the space groups* $P1$ *and* $P\bar{1}$

a	$\varepsilon(a)$ $P1$	 $P\bar{1}$
0·0	0·500	0·333
1·0	0·618	0·484
2·0	0·651	0·515
3·0	0·667	0·530
4·0	0·675	0·539
5·0	0·681	0·545
6·0	0·685	0·549

with Q defined as in Table 1. Since g and N are large, it is assumed implicitly in the previous sections, the contributions of the smallest terms (N and $2N$) can be neglected. Conditions (14) and (15) can then be reduced to

$$g \geq \varepsilon(a)\, N. \qquad (16)$$

That is to say that our criterion to reject a newly added atom as being badly misplaced will only be a good criterion when the number of already correctly located atoms g is larger than a certain fraction $\varepsilon(a)$ of the total number of atoms N in the cell. This fraction depends upon the space group as well as on the threshold. Some values of $\varepsilon(a)$ are given in Table 3.

The impact of small threshold values is large, while larger thresholds do not lead to a substantial additional decrease in the operational validity range of (12) and (13). This behaviour of $\varepsilon(a)$ contrasts in this aspect with the behaviour of $\sigma(R_2)$ (see also Table 2) near the situations $(g,0)$.

One of us (GPH) thanks the Belgian Institute for Scientific Research in Industry and Agriculture (IWONL) for a predoctoral grant.

We thank Professor Dr H. J. Geise for his critical reading of this manuscript.

References

Lenstra, A. T. H. (1974). *Acta Cryst.* A**30**, 363–369.

Parthasarathi, V. & Parthasarathy, S. (1975). *Acta Cryst.* A**31**, 38–41.

Petit, G. H. & Lenstra, A. T. H. (1979). Abstracts of the Fifth European Crystallographic Meeting, Copenhagen, pp. 344–345.

Petit, G. H., Lenstra, A. T. H. & Van Loock, J. F. (1981). *Acta Cryst.* A**37**, 353–360.

Van de Mieroop, W. (1979). PhD Thesis (in Dutch), Univ. of Antwerp (UIA), Belgium.

Van Havere, W. & Lenstra, A. T. H. (1980). Abstracts of the Sixth European Crystallographic Meeting, Barcelona, p. 156.

Wilson, A. J. C. (1969). *Acta Cryst.* B**25**, 1288–1293.