

SHORT COMMUNICATION

Contributions intended for publication under this heading should be expressly so marked; they should not exceed about 1000 words; they should be forwarded in the usual way to the appropriate Co-editor; they will be published as speedily as possible.

Acta Cryst. (1993). **B49**, 383–386

Estimate of a relationship between the number of atoms and the volumes of the unit cells of organic compounds. By PASCUAL ROMÁN,* CARMEN GUZMÁN-MIRALLES and ANTONIO LUQUE, *Departamento de Química Inorgánica, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain*

(Received 9 March 1992; accepted 1 July 1992)

Abstract

A systematic analysis concerning the relationship between the cell volumes and the numbers of atoms in the unit cells of organic molecules with the general formula $C_aH_bF_cN_dO_e$ (where $c, d, e \geq 0$) has been carried out using the Cambridge Structural Database as a source of data and search tool. The results suggest that the H atoms make a significant contribution to the total occupied volume of the unit cell. Although that influence is clearly not uniform and differs from one compound to another, it is possible to obtain two new relationships which take into account this participation. It is also clear that the inter- and intramolecular interactions, such as hydrogen-bonded contacts and π - π interactions, produce a decrease in the non-occupied volume of the cell and a lower average volume per atom.

Introduction

A knowledge of Z , the number of formula units in the crystallographic unit cell, is important at the start of a crystal structure determination. This value can be calculated from the experimental density of the compound, ρ (g cm^{-3}), using the expression:

$$Z = (V \times 0.602)/(M/\rho), \quad (1)$$

where V (\AA^3) is the unit-cell volume and M (g mol^{-1}) is the formula weight. However, in many cases the determination of the density can be time consuming, difficult or even impossible. Because of this, it is necessary to find an expression which permits estimation of the Z value from the volume of the unit cell and the number of atoms alone. In a short communication, Kempster & Lipson (1972) defined an empirical relationship between the volume of the unit cell and Z , for organic compounds containing carbon, hydrogen, nitrogen and oxygen. They took into account 40 crystal structures arbitrarily chosen and established that the Z value is estimable by the following equation:

$$Z_{\text{est}} = (V/18)/N_p, \quad (2)$$

where N_p is the number of non-H atoms per formula unit. This expression implies an approximate proportionality between M/ρ and N_p , the number of 'average' atoms in a

given structure. For 20 years, this relationship has been used by many crystallographers in the crystal structure determination of organic and organometallic compounds. However, it leads to some important differences between Z and Z_{est} for a great number of compounds. This work was undertaken to obtain a more accurate expression, which can be applied to a larger number of compounds. A total of 402 crystal structures of compounds with the general formula $C_aH_bF_cN_dO_e$ (where $c, d, e \geq 0$) were taken as the starting point for this study.

Data selection

Experimental information for this study was taken from the Cambridge Structural Database (CSD), version 4.5 of July 1991, containing 90 296 entries (Allen, Kennard & Taylor, 1983). In order to obtain a representative number of entries for each group of compounds, the following criteria were used:

- (1) crystal structures determined by X-ray and neutron diffraction from 1970 to 1991 were selected;
- (2) between 40 and 90 entries were selected for each group of compounds;
- (3) the R factor is less than 5%, except for the $C_aH_bF_c$ and $C_aH_bF_cN_d$ compounds, for which the limit on R was increased to 7% due to the low number of crystal structures reported for these compounds;
- (4) crystal structures determined at low temperatures were included in the survey.

The original list of hits was edited by hand to remove the duplicate entries. A set of local Fortran programs was used to get a more systematic and rapid treatment of the retrieved information. CSD reference codes for the 402 entries used in this work are given in the supplementary material.†

Results and discussion

Our study of 402 crystal structures shows that the expression given by Kempster & Lipson (1972) (2) is not optimal:

† A list of the CSD reference codes of the 402 crystal structures used in this work classified by $C_aH_bF_cN_dO_e$ families and the error histograms for all expressions have been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 55438 (14 pp.). Copies may be obtained through The Technical Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

* To whom correspondence should be addressed.

there are only 118 entries (29.4%) in the complete data set for which the relative error value (E) between Z (taken from the CSD) and Z_{est} is less than 5%. Fig. 1 shows the error histogram for this expression. A boundary error

value of 5% has been selected as a value generally used in many chemical studies. However, (2) yields Z_{est} to an accuracy of about 10%, hence, the number of entries with $E \leq 10\%$ has been included in the tables that summarize our work.

Intuitively, it seems obvious that the empirical expression (2) can be improved by expanding the N_p value to a form:

$$N_p = N_a + N_b + N_c + \dots$$

for different element types a, b, c, \dots etc., just as the formula weight (M) is built up of contributions from the individual weights of those different elements. In this work, organic molecules containing carbon, hydrogen, nitrogen, oxygen and fluorine have been chosen. The non-H atoms in the selected compounds have roughly the same size, so that we may assign to them the same contribution factor. The relationship expressed in (2) does not include the participation of H atoms, which is not negligible, since they are always (or nearly always) on the molecular periphery and their volume increments, although smaller than for non-H atoms, are still important. Therefore, (2) can be modified by the addition of a

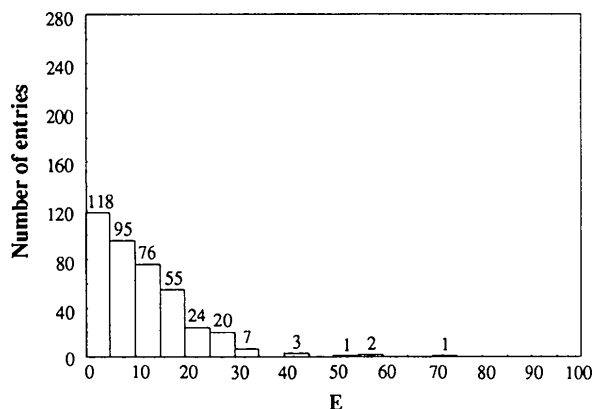


Fig. 1. Error histogram for (2). $E = 100|Z_{\text{est}} - Z_{\text{real}}|/Z_{\text{real}}$.

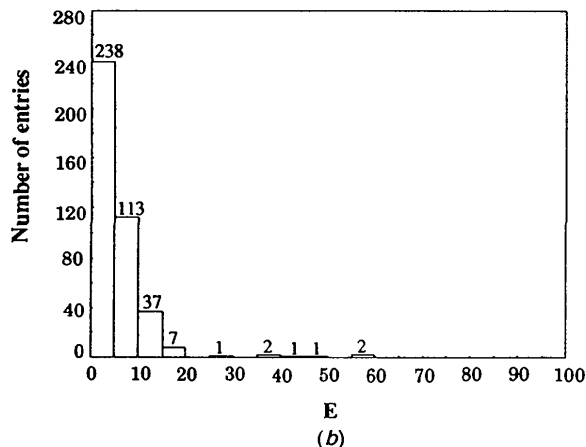
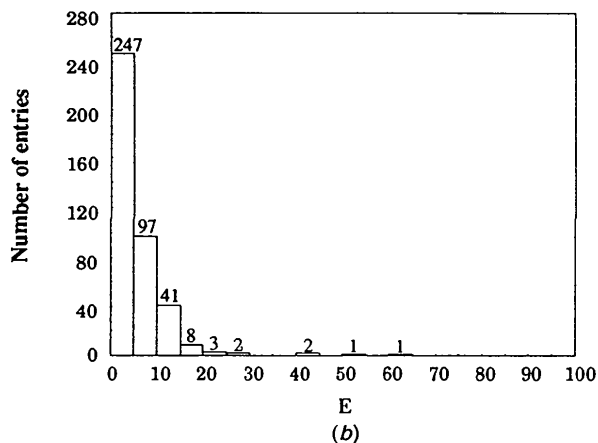
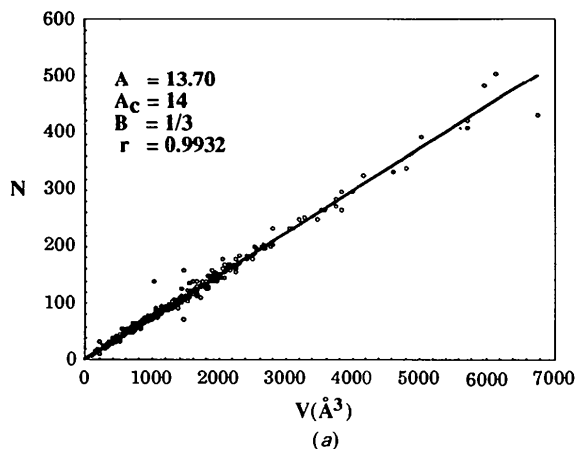
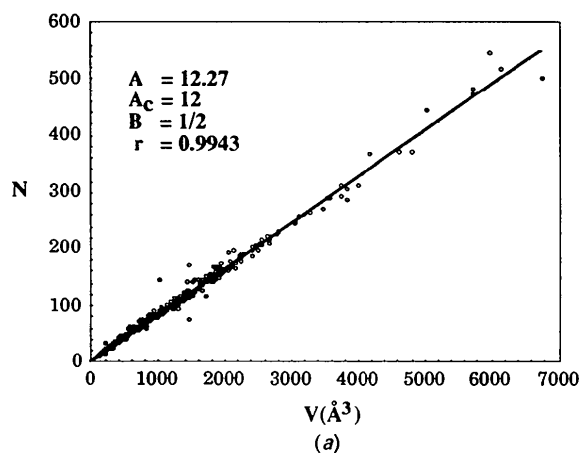


Fig. 2. (a) Scattergram of total number of atoms (including H atoms as one half of non-H atoms) versus the volume of the unit cell. (b) Error histogram for the $F_{1/2}$ expression.

Fig. 3. (a) Scattergram of total number of atoms (including H atoms as one third of non-H atoms) versus the volume of the unit cell. (b) Error histogram for the $F_{1/3}$ expression.

Table 1. Values of A , A_c and B and the number of entries ($E \leq 5\%$ and $E \leq 10\%$)

Expression	F_1	$F_{1/2}$	$F_{1/3}$	$F_{1/4}$	$F_{1/5}$	$F_{1/6}$	$F_{1/7}$	$F_{1/8}$	$F_{1/9}$	$F_{1/10}$	F_0
B	1	1/2	1/3	1/4	1/5	1/6	1/7	1/8	1/9	1/10	0
r	0.9836	0.9943	0.9932	0.9907	0.9883	0.9863	0.9847	0.9833	0.9822	0.9812	0.9697
A	9.63	12.27	13.70	14.22	14.69	15.01	15.25	15.45	15.60	15.72	16.90
A_c	10	12	14	14	15	15	15	15	16	16	17*
Entries (A)											
5%	132	247	238	220	197	192	179	170	172	168	131
10%	244	344	351	333	320	317	301	293	291	290	237
Entries (A_c)											
5%	125	226	221	210	194	191	162	149	163	169	130
10%	231	350	341	330	308	318	304	292	281	277	237

* Equation (2): $A_c = 18$; entries (5%) = 118; entries (10%) = 213.

Table 2. Values of A_c and B and the number of entries for the selected $C_aH_bF_cN_dO_e$ families

Expression	Equation (2)		F_0		$F_{1/3}$		$F_{1/2}$		F_1		Total
	B	A_c	Entries	%	Entries	%	Entries	%	Entries	%	
B	0		0		0		1		1		
A_c		18	17		14		12		10		
C_aH_b											
5%	24	40.6	28	46.7	45	75.0	39	65.0	21	35.0	
10%	42	70.0	39	65.0	57	95.0	52	86.7	38	63.3	60
$C_aH_bN_d$											
5%	23	27.4	32	38.1	46	54.8	37	44.0	29	34.5	
10%	53	63.1	59	70.2	76	90.5	64	76.2	53	63.1	84
$C_aH_bN_dO_e$											
5%	16	32.0	15	30.0	25	50.0	28	56.0	9	18.0	
10%	26	52.0	31	62.0	38	76.0	45	90.0	24	48.0	50
$C_aH_bO_e$											
5%	20	40.8	14	28.6	24	49.0	27	55.1	5	10.2	
10%	31	63.3	34	69.4	34	69.4	42	85.7	16	32.7	49
$C_aH_bF_c$											
5%	5	11.1	9	20.0	22	48.9	30	66.7	23	51.1	
10%	8	17.8	19	42.2	44	97.8	42	93.3	33	73.3	45
$C_aH_bF_cN_d$											
5%	10	20.0	13	26.0	27	54.0	23	46.0	18	36.0	
10%	18	36.0	22	44.0	46	92.0	44	88.0	32	64.0	50
$C_aH_bF_cO_e$											
5%	20	31.3	20	31.3	32	50.0	42	65.6	20	31.1	
10%	35	54.7	33	51.6	46	71.9	61	95.3	35	54.7	64
Whole data											
5%	118	29.4	131	32.6	221	55.0	226	56.2	125	31.1	
10%	213	53.0	237	59.0	341	84.8	350	87.1	231	57.5	402

term which represents the volume occupied by the H atoms:

$$Z = (V/A)/N_p + BN_H, \quad (3)$$

where N_H is the number of H atoms per formula; B is the 'contribution factor' of the H atoms to the unit-cell volume, *i.e.* the unknown ratio between the average volumes of the H atoms and the non-H atoms; $1/A$ is the slope of the straight line obtained by plotting the number of atoms in the unit cell [$N = Z(N_p + BN_H)$] versus the cell volume. The physical meaning of A may be described as the average volume occupied by one atom in the unit cell. If one assumes that the volume of non-H atoms is much greater than that of H atoms (*i.e.* $B \rightarrow 0$), the relation obtained (F_0 in Table 1) is similar to (2), although the value of A is now 17 rather than 18. This decrease in A produces only a slight improvement on the original expression (131 entries with $E \leq 5\%$). Table 1 shows the results obtained for different expressions with B running from 0 to 1.

Similar results are obtained by the use of either the real value of the slope (A) or the nearest integer (A_c), so that A_c

can be used in order to get simpler expressions. It is interesting to note that a lowering of the H-atom contribution factor (B) leads to a compensatory increase in the average volume per atom, ranging from 10 to 17 Å³ for $B = 1$ to 0. F_0 , which does not take into account the H atoms ($B = 0$), and F_1 , which gives the same weight to all atoms ($B = 1$), show the lowest number of correct entries. The equations between these boundary expressions are clearly better, with the best estimates at $B = \frac{1}{2}$ and $B = \frac{1}{3}$. Figs. 2 and 3 show the scattergrams and the error histograms derived from $F_{1/2}$ and $F_{1/3}$, respectively.

In order to distinguish the effects of chemically different atoms, the compounds under study were divided into seven different groups, from simple hydrocarbons, C_aH_b (simplest organic compounds), by adding fluorine, nitrogen and oxygen. Table 2 gives the results for the two best relationships and for the boundary conditions $B = 0$ and $B = 1$.

Given that the ratio between the covalent radius of hydrogen and the average covalent radius of carbon, fluorine, nitrogen and oxygen used in the CSD is 1/3, it is possible to deduce that $F_{1/3}$ may be the most accurate

relationship for compounds containing only these elements. For C_nH_b compounds this expression estimates 75.0% of entries with $E \leq 5\%$, the highest value for all families. However, deviations are marked for aromatic compounds in which there are important π - π interactions between the delocalized π -systems (Hunter & Sanders, 1990). On the other hand, for compounds containing N and/or O atoms, the number of entries with $E \leq 5\%$ decreases, possibly due to the existence of C—H...O and N—H...O contacts in the crystal structures (Desiraju, 1991; Taylor & Kennard, 1984), which produces a more compact crystal packing and a decrease in the average volume per atom. This effect also occurs in compounds containing F atoms in which there are contacts of the types C—H...F and N—H...F (Taylor & Kennard, 1982; Emerson, Román, Luque, Guitérrez-Zorrilla & Martínez-Ripoll, 1991). In this latter case the average volume per atom decreases and the more accurate formula is $F_{1/2}$ in which the average volume (12 \AA^3) is smaller than that in $F_{1/3}$ (14 \AA^3).

This work was supported by Iberdrola SA and UPV/EHU (grant No. 169-310-E180/91). CG-M acknowledges the financial support from the Departamento de Educación del Gobierno Vasco (grant No. BFI90.062 Modalidad BE). We are also indebted to the reviewers for their valuable comments.

References

- ALLEN, F. H., KENNARD, O. & TAYLOR, R. (1983). *Acc. Chem. Res.* **16**, 146–153.
DESIRAJU, G. R. (1991). *Acc. Chem. Res.* **24**, 290–296.
EMERSON, K., ROMÁN, P., LUQUE, A., GUTIÉRREZ-ZORRILLA, J. M. & MARTÍNEZ-RIPOLL, M. (1991). *Inorg. Chem.* **30**, 1878–1881.
HUNTER, C. A. & SANDERS, J. K. M. (1990). *J. Am. Chem. Soc.* **112**, 5525–5534.
KEMPSTER, C. J. E. & LIPSON, H. (1972). *Acta Cryst.* **B28**, 3674.
TAYLOR, R. & KENNARD, O. (1982). *J. Am. Chem. Soc.* **104**, 5063–5070.
TAYLOR, R. & KENNARD, O. (1984). *Acc. Chem. Res.* **17**, 320–326.

International Union of Crystallography

Acta Cryst. (1993). **B49**, 386

Access to CIFs archived in Chester

Since 1 January 1992 all structural papers submitted for publication in Sections B and C of *Acta Crystallographica* have had their numerical content checked in Chester and there is an associated CIF for each paper.

The CIF will contain either (i) numerical information and the whole of the text of the paper if the submission was made electronically in CIF format or (ii) numerical information, the Experimental Section and some of the text if the CIF was generated in Chester from a hard-copy submission.

The CIF for each paper is uniquely identified by the CIF reference number which appears in the deposition footnote; this reference number is used as the basis for storing the file in the archive directory.

An email file server is installed in the Chester office and readers who are interested in retrieving CIFs via email and who are subscribers to (or whose libraries or institutions are subscribers to) Section C of *Acta Crystallographica* (or Section B as appropriate) are invited to contact the Technical Editor (International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England; teched@iucr.ac.uk) who will supply details on the use of the email file server. CIFs can also be supplied on diskette, if preferred.

This service is provided on an experimental basis and there is no charge.