Acta Crystallographica Section B
**Structural
Science**

ISSN 0108-7681

Optimal partitioning of molecular properties into additive contributions: the case of crystal volumes

Sylvain Beaucamp,[a] Didier Mathieu[a]* and Viatcheslav Agafonov[b]

[a]Centre d'Etudes du Ripault, BP 16, 37260 Monts, France, and [b]Laboratoire de Chimie Physique, EA PIMIR 2098, Faculté de Pharmacie, 31 Avenue Monge, F-37200 Tours, France

Correspondence e-mail: didier.mathieu@cea.fr

Received 12 May 2006
Accepted 4 December 2006

A systematic scheme to split the volume of molecular crystals into additive increments is discussed. In contrast to earlier procedures, it relies on the definition of atom types on the basis of their geometrical rather than chemical environment. In addition, the role of the relevant structural features of the compounds is explicitly taken into account. This approach provides insight into the relative influence of chemical bonds, hydrogen bonds and rings on the volume of organic crystals. Compared with group-contribution techniques, it yields very similar results with many fewer empirical parameters. Applied to estimate the densities of 42 880 crystals containing elements up to chlorine and measured at different temperatures, an average absolute deviation from experiment close to 2% is obtained.

1. Introduction

The extensive properties of molecular systems in condensed phases are often estimated using additivity schemes (van Krevelen, 1990; Horvath, 1992). A number of such models may be derived depending on how the property of interest is decomposed into additive contributions. A whole hierarchy of rigorous partitioning schemes of molecular properties into atoms, bonds, groups and components has been defined many years ago (Benson & Buss, 1958). Today, the group-contribution approach has evolved into a versatile technique that proves useful in many different fields (Saeys *et al.*, 2006; Karunanithi *et al.*, 2006; Plyasunov *et al.*, 2006). However, the number of groups proliferates rapidly as the number of compounds to consider increases. As a result, many group-contribution methods rely on loosely defined additive increments more or less associated with functional groups familiar to organic chemists, such as methyl, carbonyl, ether, amine, nitro *etc.* However, the notion of the functional group was introduced to account for chemical reactivity. Therefore, it might not be optimal to describe the physical properties of condensed phases.

The volume of molecular crystals is a simple property for which several additivity schemes have been developed (Horvath, 1992; Stine, 1981; Ammon & Mitchell, 1998; Ammon, 2001; Hofmann, 2002; Beaucamp *et al.*, 2003). True atom-based procedures may be quite easily extended to cover virtually the whole periodic table (Hofmann, 2002). However, modified schemes where atomic contributions depend on the chemical environment of the atom within the molecule allow much more reliable estimates (Ammon & Mitchell, 1998; Ammon, 2001; Beaucamp *et al.*, 2003). Such approaches introduce different atomic volumes according to the func-

doi:10.1107/S0108768106052463

tional group to which the atom belongs and/or to neighboring bonding patterns. As a result, they exhibit several drawbacks.

First, they cannot be applied to compounds bearing functional groups beyond the scope of the parametrization. For instance, the original model of Ammon & Mitchell (1998), hereafter denoted as the AM model, is restricted to crystals made of neutral molecules and having no unusual groups.

Second, whenever an extension of the parametrization to new functional groups is needed, the question of the optimal partioning of the volume arises. In the extension of the AM method to salts (Beaucamp *et al.*, 2003), new volumes are introduced for charged groups, irrespective of the molecular backbone. For instance, a standard volume is attributed to the $-NH_3^+$ moiety, notwithstanding the overall structure of the ammonium cation. In contrast, in the original parametrization of the AM method for neutral compounds, a group such as *e.g.* $-NH_2$ may be assigned different volumes depending on the hybridization of the neighboring atom. Such distinctions might be difficult to rationalize on physical grounds. Furthermore, the arbitrary definition of standard atomic volumes raises some questions. For instance, the original Ammon model yields an average absolute deviation (a.a.d.) from experiment close to 2% for neutral compounds, while its extension encompassing charged groups yields a significantly higher a.a.d. of 2.4%. It is unclear whether this loss of accuracy stems from an inadequate partitioning of the volume of ionic crystals. Alternatively, the representation of the crystal volume in terms of additive contributions might not be as good an approximation for charged systems.

Any improvement in the reliability of density prediction methods is of immediate interest in the field of energetic materials. For instance, current thermochemical models yield detonation velocities with r.m.s. errors close to 2–3% (McGee *et al.*, 1998). However, considering the density dependence of this performance criterion (Kamlet & Jacob, 1968), it is clear that even a 2% error on the input density – for target compounds not yet synthesized – has a significant impact on the reliability of calculated velocities.

From a more fundamental viewpoint, group-contribution methods provide only limited insight into how crystal densities depend on molecular features. While ring structures and hydrogen bonds are likely to yield increased compacities, their effect is absorbed into the many empirical group volumes and therefore cannot be quantified.

2. Present approach

2.1. Partitioning scheme

In this paper, a systematic partitioning of the crystal volume into additive contributions is introduced on the basis of simple physical considerations. Since the packing coefficient of crystals is roughly constant (Dunitz *et al.*, 2000), the volume occupied by an atom k in a molecular crystal appears to be determined by its contribution V_W^k to the van der Waals volume V_W of the compound. This contribution V_W^k depends mostly on:

(i) the van der Waals radius r_k of atom k;

(ii) the number of neighboring atoms q overlapping atom k;

(iii) the van der Waals radii r_q of the neighboring atoms.

In particular, the fact that H atoms are much smaller than heavier elements indicates that V_W^k depends not only on the coordination number of atom k, but also on its neighbors q being H atoms. For simplicity, all the heavier elements are assumed to have the same influence on V_W^k. This suggests a systematic definition of standard atomic volumes $V(Z, n, n_H)$ that depend only on the atomic number Z of the atom, the total number n of its neighbors and the number n_H of its hydrogen neighbors. Due to the dependence on n_H, atomic volumes cannot be defined unambiguously for H atoms. Therefore, the contribution of H atoms to the crystal volume is absorbed into the atomic volumes of non-H atoms. Furthermore, taking advantage of a previous observation (Beaucamp *et al.*, 2003), a special volume V_{O^-} is introduced for O atoms with a formal atomic charge of -1 attached to C or N through a single bond.

On the other hand, instead of introducing different atomic volumes for atoms involved in rings or hydrogen bonds, structural corrections are used. This approach makes explicit the dependence of the crystal volume on such structural features. In addition, it requires many fewer empirical parameters. In the present scheme, only nine structural volume increments are introduced. To account for the influence of hydrogen bonds, a parameter $h = N_A N_H / N$ is defined, with N_A the number of proton acceptors, N_H the number of labile protons and N the total number of atoms in the chemical crystal unit. Only the most significant hydrogen bonds are considered, following the definition of N_A and N_H adopted by Charlton *et al.* (1995). The correction to the crystal volume for hydrogen bonds is then defined as hV_h, where V_h is a volume increment.

Six additional increments V_s (s = 3–7) are introduced for s-membered rings, while V_8^+ is the correction for any ring with more than seven atoms. In addition, planar five- or six-membered rings are quite common. Since they may not have exactly the same influence on the volume as non-planar rings, two planarity corrections ΔV_5 and ΔV_6 are introduced.

Since it relies on parameters associated respectively with atoms, rings and hydrogen bonding, the present approach is hereafter referred to as ARH. To make easier the comparison of this model with earlier schemes, its reliability is assessed through the relative difference δ between the density ρ derived from the ARH volume and the value deduced from the observed unit-cell volume. The average absolute value of δ, hereafter denoted $\overline{\delta}$, is the parameter most commonly used to compare the performance of various predictive schemes for the density of molecular crystals (Piacenza *et al.*, 1996; Ammon & Mitchell, 1998; Ammon, 2001; Beaucamp *et al.*, 2003). Whatever the method used, values of $\delta > 20\%$ are obtained for crystals that are not closely packed. However, the most accurate methods so far provide crystal densities with $\overline{\delta} \simeq 2\%$.

Table 1
Standard atomic contributions $V(Z, n, n_\mathrm{H})$ of elements Z = C, H, N, O, F to crystal volumes, as a function of the atom coordination numbers n and number of hydrogen neighbors n_H.

The volumes V and their standard deviations σ are reported in Å^3. In addition, atomic densities ρ, defined as the ratio of the atom mass to the corresponding volume increment V, are reported in g cm^{-3}. N_obs is the number of crystals used in the determination of the atomic volume V.

Z	n	n_H	V	σ	ρ	N_obs
C	4	3	33.468	0.008	0.746	12 629
C	4	2	24.409	0.003	0.954	12 583
C	4	1	18.130	0.009	1.192	10 968
C	4	0	10.286	0.017	1.939	8350
C	3	2	29.236	0.031	0.797	873
C	3	1	21.205	0.003	1.019	13 537
C	3	0	12.119	0.008	1.646	16 503
C	2	1	26.442	0.086	0.818	95
C	2	0	16.752	0.019	1.191	1294
C	1	0	20.388	0.167	0.978	15
N	4	4	38.534	0.152	0.777	33
N	4	3	25.857	0.052	1.094	359
N	4	2	21.297	0.076	1.249	133
N	4	1	11.857	0.098	2.103	94
N	4	0	5.392	0.097	4.314	73
N	3	2	24.643	0.038	1.080	907
N	3	1	18.148	0.016	1.374	3479
N	3	0	9.944	0.014	2.339	5406
N	2	1	16.721	0.161	1.491	24
N	2	0	13.324	0.011	1.746	3239
N	1	0	15.300	0.026	1.520	1037
O	2	2	26.118	0.042	1.145	595
O	2	1	15.764	0.013	1.792	6061
O	2	0	12.512	0.005	2.123	10 487
O	1	0	13.935	0.009	1.907	12 516
O^-	–	–	11.818	0.031	2.248	1251
F	1	0	14.760	0.012	2.137	597
F	0	0	28.181	0.385	1.119	4

Table 2
Same data as in Table 1 for the most common coordination numbers of elements Li, B, Na, Mg, Al, Si, P, S and Cl.

The column headings are as defined in Table 1.

Z	n	n_H	V	σ	ρ	N_obs
Li	0	0	6.414	0.172	1.797	25
Li	4	0	42.978	0.123	0.268	21
B	2	0	29.688	0.708	0.605	2
B	3	0	16.690	0.035	1.076	210
B	3	1	26.720	0.452	0.734	5
B	4	0	11.251	0.044	1.595	359
B	4	1	18.907	0.219	1.038	16
B	4	2	33.202	0.102	0.641	39
B	4	3	40.587	0.101	0.566	56
B	5	0	15.382	0.360	1.167	10
B	5	1	21.814	0.106	0.900	35
B	5	2	21.309	0.262	0.999	8
B	5	3	30.684	0.253	0.749	4
B	6	0	14.845	0.096	1.209	13
B	6	1	17.607	0.034	1.115	43
B	6	2	24.543	0.077	0.868	25
B	6	3	24.291	0.343	0.946	9
F	1	1	16.692	0.333	1.990	3
Na	0	0	13.000	0.072	2.937	95
Mg	4	0	42.155	0.345	0.957	9
Mg	6	0	7.175	0.333	5.625	9
Al	3	0	18.118	0.416	2.473	6
Al	4	0	29.892	0.056	1.499	138
Al	4	2	38.190	0.200	1.261	8
Al	5	0	22.097	0.243	2.028	17
Al	6	0	11.275	0.267	3.974	14
Si	4	0	29.200	0.011	1.597	1091
Si	4	1	34.907	0.065	1.384	42
Si	4	2	45.499	0.152	1.099	5
Si	5	0	24.972	0.117	1.868	62
Si	6	0	28.589	0.302	1.631	11
P	2	0	27.706	0.092	1.856	80
P	3	0	26.910	0.022	1.911	436
P	3	1	31.840	0.163	1.668	13
P	4	0	22.161	0.017	2.321	1642
P	4	1	29.909	0.132	1.776	44
P	5	0	20.768	0.072	2.477	149
P	5	1	27.277	0.289	1.947	12
P	6	0	13.649	0.096	3.768	60
S	1	0	29.521	0.021	1.804	1236
S	2	0	26.152	0.007	2.036	3773
S	2	1	36.980	0.134	1.485	35
S	3	0	21.336	0.030	2.496	772
S	4	0	17.531	0.017	3.037	2232
S	6	0	7.002	0.289	7.605	9
Cl	0	0	29.852	0.030	1.972	856
Cl	1	0	29.084	0.006	2.024	3133
Cl	1	1	30.129	0.302	2.010	5
Cl	4	0	17.845	0.070	3.299	147

2.2. Database

The crystals considered in the present work were retrieved from the Cambridge Structural Database (CSD, Version 5.23; Allen, 2002) obtained from the Cambridge Crystallographic Data Centre. They were selected according to the following constraints: three-dimensional coordinates determined, crystallographic R factor < 0.075, not disordered, no errors, not polymeric, no powder structures. Each entry was exported from the database as a CIF file and a molecular .mol2 file in the Tripos exchange format. The bonding information encoded in the .mol2 files was used to obtain the relevant atomic and structural increments for each crystal. Entries for which the elementary composition from the CIF file was inconsistent with that of the .mol2 file were filtered out. Although this procedure is likely to discard some correct structures with atoms lying on symmetry elements of the crystal, it was used to remove crystals for which the hydrogen positions have not been unambiguously determined. Indeed, in such cases H atoms were missing in the .mol2 file.

Whenever several entries satisfying all the preceding conditions are available for a given compound, the data associated with the lowest R factor are retained. In a first step, only data measured between 273 and 323 K are used to fit the volumic increments, in view of estimating densities at ambient temperature. Then, data obtained at lower temperatures are used to gain insight into temperature effects and derive an average coefficient of thermal expansion.

3. Regression

The increments of the model were fitted to reproduce the experimental crystal volumes, using the singular value decomposition procedure outlined in Beaucamp *et al.* (2003). The regression was carried out in three steps, considering only the 17 359 neutral CHNOF compounds in a first step, including 1519 ionic CHNOF compounds in a second step and

Table 3
Structural increments V_s to the volume of molecular crystals and associated standard deviations σ_s (Å^3).

$N_{\rm obs}$ is the number of crystals used in the determination of V_s and σ_s for the increment s under consideration.

Increment s	V_s	σ_s	$N_{\rm obs}$
V_3	−0.843	0.027	1268
V_4	−2.380	0.033	774
V_5	−4.103	0.017	8921
V_6	−6.639	0.017	14 876
V_7	−7.893	0.030	1065
V_8^+	−9.037	0.036	924
ΔV_5	0.622	0.020	3261
ΔV_6	0.223	0.019	11 380
V_h	−7.341	0.056	9009

13 459 crystals with other elements (Li, B, Na, Mg, Al, Si, P, S and Cl) in the last step.

To assess the predictive value of such models involving empirical parameters, they are usually applied to systems not used for the fitting procedure. This approach has been applied to the present ARH model. More specifically, 500 crystals were randomly discarded from the CHNOF dataset used to fit the parameters. Using the values thus obtained, the densities of the 500 crystals not used in the regression were predicted and associated values of $\overline{\delta}$ calculated. The latter prove to depend only slightly on the crystals removed. For CHNOF compounds, the values of $\overline{\delta}$ calculated from *predicted* densities fall within 0.04% of the value derived from the densities *fitted* using the whole dataset. This indicates a good transferability of the model parameters. Since they do not depend on an arbitrary splitting of the dataset into calibration and validation panels, only data derived from the regression on the whole dataset are reported in the sequel. This allows us to take full advantage of the available experimental data to determine the values of the empirical increments.

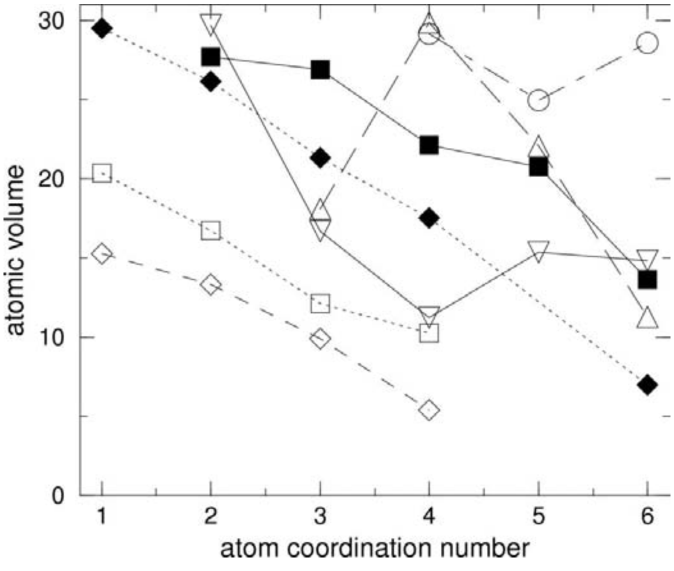

Figure 1
Volumes (in Å^3) of atoms not bonded to any hydrogen as a function of their coordination number: boron (triangles down), carbon (open squares), nitrogen (open diamonds), aluminium (triangles up), silicon (circles), phosphorus (closed squares) and sulfur (closed diamonds).

3.1. Atomic increments

The atomic volume increments (V) and associated standard deviations (σ) obtained are listed in Table 1 for CHNOF crystals, and in Table 2 for other elements. In addition, the corresponding atomic densities are also listed, since they may be useful in practical applications when density is a design criterion (*e.g.* in the field of energetic materials). In molecules, the volume occupied by a given atom is expected to decrease as it becomes more crowded, because of the increasing overlap between van der Waals spheres. This is mostly the case, especially for elements C, N or S, as shown on Fig. 1 for atoms with no hydrogen neighbors. This trend does not emerge so clearly for elements that may be involved in a rich variety of bonding schemes, such as B or P, and it is even less true for metals. Cations tend to occupy a reduced volume owing to smaller ionic radii. For instance, the volume of Li^+ cations is about seven times smaller than the volume of Li atoms surrounded by four ligands. In contrast, F^- appears much larger than fluorine atoms in molecules. The volume of oxygen in *N*-oxides is an exception: despite its formal atomic charge of one electron, it is smaller than oxygen in carbonyl groups, presumably because of a significant overlap with neighboring protons through hydrogen-bonding interactions. It appears difficult to draw definite conclusions concerning the volume of some of the heavier atoms. Since they are often involved in coordination complexes with bulky ligands that do not lend themselves to close-packed structures, their volumes are not so well defined. In addition, after the application of the filtering procedure described in §2.2, only few entries remain available to derive some increments, such as V(B,2,0) or V(B,5,3) (Table 2).

Therefore, we now focus on the volumes of the most common elements. The second derivative of the volume $V(Z, n, 0)$ with respect to the coordination number n may be expected to be negative. Indeed, as more (non-hydrogen) neighbors are added, they overlap not only with the atom k considered, but also with its neighbors already present. This overlap between non-H atoms l, l' bonded to atom k is not explicitly accounted for by the present partitioning scheme, as long as l and l' are not bonded together. Thus, it may only be reflected in the volume of the central atom k. Nevertheless, the stretching of chemical bonds k–l as atom k becomes more crowded is likely to have the opposite effect. In fact, no systematic variation is observed in Fig. 1: while the second derivative of $V(N, n, 0)$ with respect to n is clearly negative, the second derivative of $V({\rm C}, n, 0)$ is rather positive.

On the other hand, given the coordination number n of atom k, its volume increases with $n_{\rm H}$ since it includes the volume of the neighboring H atoms. Finally, given a chemical environment $(n, n_{\rm H})$, atoms with larger van der Waals radii usually exhibit larger volume increments, as expected. Accordingly, V decreases on going from C through N to O. A similar trend is observed for second-row atoms Si–S. However, halogen atoms do not follow this trend. For instance, $V(\mathrm{F}, 1, 0) > V(\mathrm{O}, 1, 0)$ despite the fact that $r_{\mathrm{F}} < r_{\mathrm{O}}$.

3.2. Structural increments

The values (V_s) and standard deviations (σ_s) of the structural increments are reported in Table 3. As expected, V_h as well as all V_s increments are negative. The small value of the standard deviation σ_h is of special interest. It suggests that h is indeed well suited to characterize the influence of hydrogen bonding.

The most obvious reason why rings in molecules decrease the crystal volume lies in the enhanced overlap between atoms that become bonded to each other as a new ring is created. However, the negative values of the ring increments V_s are associated with interatomic overlap between non-bonded atoms, since overlap between bonded atoms is already taken into account by the definition of the ARH atom types. Using space-filling models of cycloalkanes, it clearly appears that the constraints associated with an s-membered ring enhance the overlap between non-bonded atoms involved in the ring, even for s as large as 12. This observation explains the negative sign of the V_s parameters, but not their steady variation on going from $s = 3$ to $s = 8$.

Nevertheless, it proves unnecessary to resort to packing effects to account for the decrease of V_s as the ring size increases. In fact, the latter is required to account for the variation of the van der Waals volume. To illustrate this, the crystal volume per CH_2 unit $V(CH_2)$ of cycloalkane crystals has been estimated using two different approaches:

(i) using the present ARH additive scheme;

(ii) from calculated van der Waals volumes, assuming a constant packing coefficient $C_K = 0.72$.

The latter calculations were carried out using molecular geometries optimized with the Merck molecular force field (MMFF; Halgren, 1996) implemented in the *Spartan* program (Hehre, 2003) and standard van der Waals atomic radii. The results are summarized in Fig. 2. They indicate that $V(CH_2)$ goes through a minimum for $s = 6$. For $s < 7$, the ring correction needed becomes more significant as more atoms are forced to overlap at the ring center. This provides an explanation for the fact that V_s becomes more negative as s increases from 3 to 8. For $s > 7$, the ring influence should become less significant as s increases, since local constraints may be more easily relaxed. This explains why $V(CH_2)$ becomes less negative.

Finally, ΔV_5 and ΔV_6 are slightly positive, indicating that the volume contraction is somewhat less significant for planar rings. In other words, the planarity of rings tends to increase the crystal volume. This might arise either because of an increase of the van der Waals volume of the compounds, or because of packing effects. To investigate the role of the planarity of rings, the van der Waals volume of five- and six-membered rings of C atoms, either planar or not, was calculated. These rings have been taken from the MMFF structures of cyclopentadiene, pentane, benzene and cyclohexane. It turns out that V_W decreases by 6–7% on going from a non-planar to a planar ring. This indicates that the positive values of the planarity corrections ΔV_5 and ΔV_6 arise because of packing effects, while the influence of planarity on the van der Waals volume contributes negatively to these correction terms. This observation that planar rings pack less efficiently is somewhat unexpected. It is presumably due to the relative rigidity of planar rings and the occurrence of exceptionally low compacity crystals for polycyclic compounds such as YAFHAO (Ammon & Mitchell, 1998).

4. Calculated crystal densities

4.1. CHNOF neutral molecular crystals

In order to make easier the comparison with the AM method (Ammon & Mitchell, 1998) only results obtained for the 17 359 crystals made of neutral CHNOF compounds are considered in a first step. For these systems, $\overline{\delta} = 2.02\%$, which is close to the value of 2.04% obtained with the AM method. The distribution of the δ values obtained with the ARH scheme is plotted in Fig. 3 for comparison with the histogram presented by Ammon & Mitchell (1998). In view of the close similarity of the results, it is of special interest to compare the number of empirical parameters needed: 29 for the present approach *versus* 78 for the AM method. The most significant errors concern the densities of CEWHIV (+24%) and YAFHAO (+22%). CEWHIV was not included in the AM dataset owing to a crystallographic R factor > 0.07. The most significant density overestimation using the AM method was for YAFHAO (+23%). Both crystals are exceptions to the close-packing principle, owing to unfavorable shapes. In fact, both structures exhibit empty channels that were almost certainly occupied by disordered solvent, although the latter

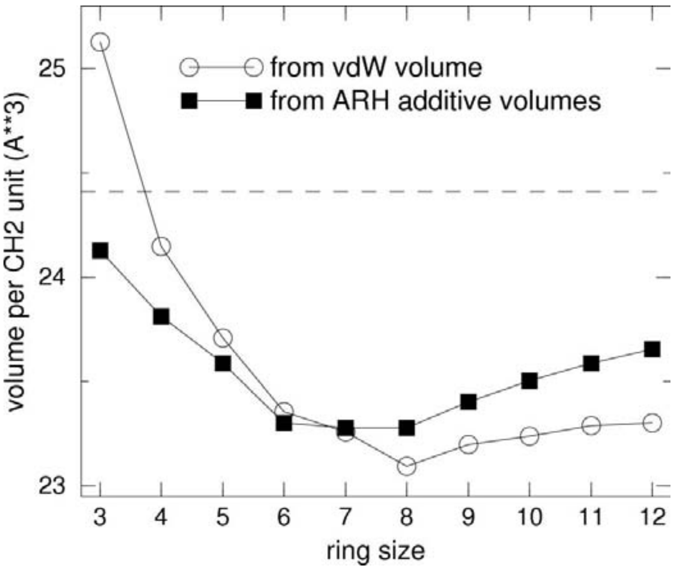

Figure 2
Volume per CH_2 unit for cycloalkane crystals (Å^3) calculated either from the van der Waals volume (assuming a constant packing coefficient C_K = 0.72) or from the ARH additive volumes. The dashed line represents the value derived from the ARH model if ring corrections are omitted.

went undetected in the course of the structure determination. Their densities cannot be predicted from additivity schemes. In contrast, significant underestimations of crystal densities are more puzzling, as fair results might be expected for closely packed crystals. In fact, the AM density of FIJTUN is underestimated by −20%, while it is overestimated by only +5% using the present procedure. The most negative δ values obtained using the present method are for REQZAO (−11%), ISCOFF and PERYTO (−10%). Thus, the present scheme is somewhat more consistent with expectations than the AM model.

4.2. CHNOF salt crystals

When extending the AM method to salts through the inclusion of 12 additional volume increments for charged groups, a value of 2.38% was obtained for $\overline{\delta}$ (Beaucamp *et al.*, 2003). At that time, the reason for this increase of $\overline{\delta}$ on going from neutral to charged compounds was not obvious. It could arise either from an inadequate definition of the volume increments or because of problems specific to ionic systems, such as stronger non-additive effects or less reliable experimental data. In this context, the performance of the ARH method for salts is of interest. The ARH densities of the 1519 CHNOF salt crystals considered here have been calculated, using seven atomic volumes in addition to the 29 values needed for neutral CHNOF species. This data yield $\overline{\delta} = 2.41\%$. Since ionic and neutral compounds are treated on the same footing by the ARH approach, it may be concluded that the larger errors observed on extending the AM method to salts did not arise as a result of inadequate increments, but are inherent to salt crystals. They may occur either because X-ray data for salts are more error prone or because the use of additive volumes might not be as good an approximation as for neutral crystals. The much stronger electrostatic forces in salts are more likely to squeeze groups with opposite charges very close to each other, especially if H atoms are involved. Indeed, while the packing coefficients of crystals do not differ much from $C_K = 0.74$ (Dunitz *et al.*, 2000) much larger values are observed for salts (Beaucamp *et al.*, 2005). Thus, the picture of atoms as overlapping hard spheres – which underpins the present model – is more questionable.

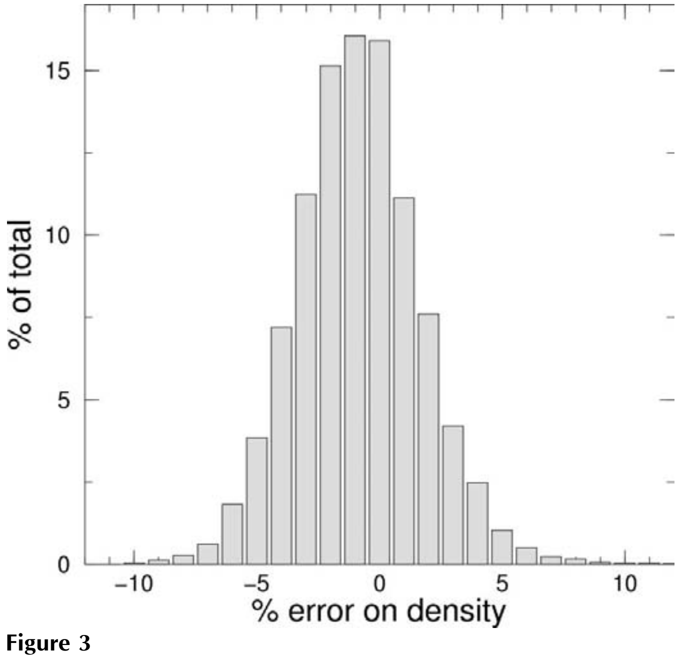

Figure 3
Distribution of the relative errors on ARH crystal densities for CHNOF neutral crystals between 273 and 323 K.

On applying the ARH model to CHNOF salt crystals, it is gratifying to observe that densities are not too much underestimated, with most negative values of δ obtained for JIJCOU (−12%) and SIMRAH (−10%). By far the most significantly overestimated density concerns LETGIA, for which $\delta = 23\%$. Notwithstanding this compound, the largest values of δ are observed for ZOGSOD (+13%), YUWMOS (+12%), BZOFOX and WIKGEC (+11%). For ZOGSOD and YUWMOS, a close packing of the constitutive compounds is hindered by non-parallel rings. BZOFOX is an *N* oxide compound completely having no H atoms. The examination of the unit cell indicates a significant departure from a close-packed structure, presumably because of the repulsive character of atom–atom interactions associated with many lone pairs. LETGIA and WIKGEC are nitrate salts where the NO_3^- anion is in close interaction with $-NH_2$ or $=NH_2^+$ moieties, and for which the magnitude of the hydrogen-bonding correction to the crystal volume appears to be overestimated. Indeed, the significance of hydrogen bonds in a crystal may be characterized by the ratio h/N. This ratio range from 0 (in the lack to hydrogen bonds) to 0.25 for strongly hydrogen-bonded systems such as hydrogen fluoride. In practice, using the present definition of hydrogen bonds, $h/N < 0.1$ in most cases. Among the CHNOF crystals considered in the present study, 10 334 are hydrogen bonded, but $h/N > 0.18$ only in 12 cases, including LETGIA and WIKGEC. For these crystals, the ARH densities turn out to be systematically overestimated by more than 5%. This suggests that the present hydrogen bond correction to the crystal volume is not suitable for extreme values of the ratio h/N. This is understandable since h characterizes the number of hydrogen bonds expected per chemical unit on the basis of the law of mass action, *i.e.* assuming a dilute solution of independent proton donors and acceptors. In contrast, large values of h/N imply a high concentration of hydrogen bonds in the crystal. Moreover, the fact that proton donors and acceptors lie on the same molecules induces significant constraints. In this context, the linear increase of the number of hydrogen bonds with N_A or N_H appears unrealistic for large values of h/N.

4.3. Other systems

On going from CHNOF crystals to other elements, the present ARH approach might prove less reliable. Indeed, as mentioned in §3.1, some increments are derived using only a few observations, often carried out on structures that are not closely packed (*e.g.* coordination complexes). Moreover, as atoms beyond the first row are considered, the occurrence of larger van der Waals radii is likely to affect the validity of the simple assumptions at the basis of the ARH approach, because a significant part of the interatomic overlap may involve non-bonded atoms.

In fact, no significant loss of accuracy is observed on considering lithium, boron and second-row elements in addition to CHNOF compounds. The densities of the 464 crystals containing lithium or boron yield $\overline{\delta} = 2.17\%$. Only one deviation from observed values is $>10\%$, namely for FUZVIF, a crystal made of boron cages. The calculated density for QECNAN is 31% larger than the value of $1.029\ \mathrm{g\ cm^{-3}}$ reported in the CSD. However, the latter proves to be erroneous: the lattice parameters and chemical unit yield a density of $1.372\ \mathrm{g\ cm^{-3}}$, very close to the ARH value of $1.351\ \mathrm{g\ cm^{-3}}$. For the 12 995 crystals with second-row atoms, $\overline{\delta} = 2.01\%$. Again, the application of the ARH model reveals the presence of erroneous densities in the database, not consistent with the associated structural data. For instance, the density of ZUFDAF is estimated to be $1.248\ \mathrm{g\ cm^{-3}}$, while a value of $1.813\ \mathrm{g\ cm^{-3}}$ is reported in the database. The value derived from the unit-cell parameters and chemical unit is $1.208\ \mathrm{g\ cm^{-3}}$. Taking such errors into account, the most negative value of δ (-21%) is obtained for BENHOR, an ammonium salt with countercharges associated with pentacoordinated Si atoms. Notwithstanding PIZDOR, a sulfur compound having no H atoms, $\delta < -12\%$ only for some silicon-containing crystals. A significant number (29) of crystal densities are overestimated by 10% or more (up to 23% for OFUXES). Most of these deviations may be explained by the shape of the molecules, which precludes the formation of closely packed structures.

Clearly, the present ARH method is designed with molecular crystals and organic salts in mind. It cannot be unambiguously applied to arbitrary inorganic crystals, since the definition of chemical bonds is not always obvious, *e.g.* for bonds with hybrid covalent/ionic character. Moreover, even in covalent crystals, the ARH approach meets with difficulties because of the very large number of rings in three-dimensional bond networks. For instance, as it stands, this approach overestimates the density of diamond by 55% because the atomic volume is dramatically decreased by the ring corrections (one per atom). In contrast, the value obtained with ring corrections ignored is underestimated by 45%. Nonetheless, the present scheme appears to provide fair results for two-dimensional solids ($\delta = -0.5\%$ for graphite) or inorganic salts ($\delta = 4\%$ for rocksalt, -1% for ammonium nitrate).

4.4. Temperature effects

In the CSD database, 10 543 entries are consistent with the requirements listed in §2.2, except for the fact they were observed at temperatures $T < 273\mathrm{K}$. The ARH densities of these crystals yield $\delta = 2.90\%$. As expected from the low temperatures, more negative values of δ are observed, especially a value of -15% for KIDBUU and KAPFIQ. However, thermal effects being quite small, large positive values of δ are still obtained (up to +20% for JAVFUH).

In principle, using the volume of the unit cell at low temperature and the corresponding volume at ambient temperature estimated using the ARH approach, and assuming a linear variation of the crystal volume over the temperature range considered, a coefficient of thermal expansion α may be derived for every crystal. In practice, the ARH method is clearly not accurate enough for meaningful values of α to be obtained. In fact, α is often found to be negative, with values ranging from $-2 \times 10^{-3}\ \mathrm{K^{-1}}$ to $+1.7 \times 10^{-3}\ \mathrm{K^{-1}}$ for the 10 543 low-temperature crystals. Nevertheless, the average value $\overline{\alpha} = 1.7 \times 10^{-4}\ \mathrm{K^{-1}}$ is reasonable, although almost twice as large as the value obtained by Hofmann (2002) using volumes independent of the atom environment. Applying this value, the densities of the low-temperature crystals are estimated with the same accuracy as ambient data, with an average deviation $\overline{\delta} = 2.09\%$.

5. Conclusion

Using the ARH model, the standard crystal densities of compounds made of elements up to chlorine may be estimated with an average deviation from experiment close to 2%. Using an average thermal expansion coefficient of $\overline{\alpha} = 1.7 \times 10^{-4}\ \mathrm{K^{-1}}$, the same accuracy is obtained at low temperatures. Only four negative deviations below -12% are observed, for three silicon compounds (BENHOR, HIRPUT and FSINBR) and one sulfur compound (PIZDOR). On the other hand, some positive deviations as large as +23% are observed, either because the crystal does not obey the close-packing principle owing to unfavorable molecular shapes or because of an overestimation of the role of hydrogen bonds, in the few cases when $h/N > 0.18$. We plan to improve the modelling of the volume contraction associated with hydrogen bonds before extending the scope of the model. The present definition of the atomic volumes bears some similarity with the atom code concept indroduced by Ammon (2001), following the systematic procedure defined earlier by Benson & Buss (1958) to split molecular compounds into groups. The latter was found to be inferior to the attribution of atomic volumes according to the functional groups to which they belong (Ammon, 2001). The present procedure overcomes two drawbacks associated with the use of the 1601 Ammon atom codes. First, the number of empirical parameters is much smaller, hence their values are statistically better defined, except for a few unusual bonding environments. Secondly, the structural corrections provide an efficient way to take into account rings and hydrogen bonds. Furthermore, the corrections afford new insight into the magnitude of their influence.

The question whether further improvement in the prediction of crystal densities can be obtained within the framework of additivity schemes remains unclear. Progress has been hampered by the lack of a systematic approach for decomposing the crystal volume into additive contributions. On the basis of the present scheme, introducing some specific correction terms might allow improved predictions within a limited range of compounds for which extensive crystallographic data are available. For instance, more reliable estimates of the densities of energetic materials might be obtained using correction terms for nitramines, nitroaromatic or nitroalkane compounds. However, this approach provides no systematic procedure for increasing the flexibility of the ARH scheme in order to improve its overall reliability.

To avoid the problems associated with the fit of many parameters, it might be interesting to compute atomic volumes directly from the crystallographic data instead of fitting their values against the cell volumes. The most common procedures to split physical properties into atomic contributions, such as the Hirshfeld (1977) or Bader (1990) analyses, are not well suited for this purpose as they would involve time-consuming quantum calculations. However, more straightforward approaches to split the cell volume into atomic contributions may be used, such as the Voronoi procedure (Baburin & Blatov, 2004). Atom types associated with standard Voronoi atomic volumes could be defined on this basis and subsequently used to estimate the volume of new crystals.

A Python script to compute easily the ARH densities from *.xyz* molecular files is available upon request or from the CEA–Le Ripault website (Mathieu, 2006).

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.

Ammon, H. L. (2001). *Struct. Chem.* **12**, 205–212.

Ammon, H. L. & Mitchell, S. (1998). *Propel. Explos. Pyrotechn.* **23**, 260–265.

Baburin, I. A. & Blatov, V. A. (2004). *Acta Cryst.* B**60**, 447–452.

Bader, R. F. W. (1990). *Atoms in Molecules – A Quantum Theory.* Oxford University Press.

Beaucamp, S., Marchet, N., Mathieu, D. & Agafonov, V. (2003). *Acta Cryst.* B**59**, 498–504.

Beaucamp, S., Mathieu, D. & Agafonov, V. (2005). *J. Phys. Chem. B*, **109**, 16469–16473.

Benson, S. W. & Buss, J. H. (1958). *J. Chem. Phys.* **29**, 546.

Charlton, M., Docherty, R. & Hutchings, M. (1995). *J. Chem. Soc. Perkin Trans. II*, **17**, 2023.

Dunitz, J. D., Fillipini, G. & Gavezzotti, A. (2000). *Tetrahedron*, **56**, 6595–6601.

Halgren, T. A. (1996). *J. Comput. Chem.* **17**, 490.

Hehre, W. J. (2003). *A Guide to Molecular Mechanics and Quantum Chemical Calculations.* Wavefunction Inc.

Hirshfeld, F. L. (1977). *Theoret. Chim. Acta*, **44**, 129–138.

Hofmann, D. W. M. (2002). *Acta Cryst.* B**58**, 489–493.

Horvath, A. L. (1992). *Molecular Design.* Amsterdam: Elsevier Science Ltd.

Kamlet, M. L. & Jacob, S. J. (1968). *J Chem. Phys.* **48**, 23–35.

Karunanithi, A. T., Acheniea, L. E. & Gani, R. (2006). *Chem. Eng. Sci.* **61**, 1247–1260.

Krevelen, D. W. van (1990). *Properties of Polymers.* Amsterdam: Elsevier Science Ltd.

Mathieu, D. (2006). *ARH*, Version 1.0. http://www-dam.cea.Fr/statique/sitesDAM/vdc_ripault_accueil.htm or didier.mathieu@cea.fr.

McGee, B. C., Hoobs, M. L. & Baer, M. R. (1998). *Exponential 6 parameterization for the JCZ3-EOS.* Sandia Report SAND98-1191.

Piacenza, G., Legsai, G., Blaive, B. & Gallo, R. (1996). *J. Phys. Org. Chem.* **9**, 427.

Plyasunov, A. V., Plyasunova, N. V. & Shock, E. L. (2006). *J. Chem. Eng. Data*, **51**, 276–290.

Saeys, M., Reyniers, M. F., Speybroeck, V. V., Waroquier, M. & Marin, G. B. (2006). *ChemPhysChem*, **7**, 188–199.

Stine, J. R. (1981). *Predictions of Crystal Densities of Organic Explosives by Group Additivity.* Report LA-8920. Los Alamos National Laboratory, New Mexico, USA.