Acta Crystallographica Section B
**Structural Science**
ISSN 0108-7681

# Sizes of molecules in organic crystals: the Voronoi–Dirichlet approach

**Igor A. Baburin and Vladislav A. Blatov***

Samara State University, Ac. Pavlov St. 1, Samara 443011, Russia

Correspondence e-mail: blatov@ssu.samara.ru

Received 29 March 2004
Accepted 25 May 2004

The sizes of more than 100 000 molecules in organic crystals have been assessed as the volumes of molecular Voronoi–Dirichlet polyhedra. The average molecular volumes for all crystals are shown to be nearly equal to the corresponding values in homomolecular (consisting of identical molecules) crystals. The validity of the Voronoi–Dirichlet approach in determining molecular sizes is substantiated and the reasons for the variations in the molecular volumes are discussed. It is shown that a molecule increases its volume if it is surrounded by a good deal of high-row (*i.e.* an element with more than ten protons) atoms or if there is disorder in the crystal structure.

## 1. Introduction

Assessing the sizes of molecules or functional groups is an important problem in chemistry arising from the study of interatomic or intermolecular contacts, crystal packing analysis, and the prediction of the physico-chemical properties of crystals such as density, expansion coefficient, melting point *etc.* (Zefirov & Zorky, 1995). To solve this problem one should divide crystal space into the regions associated with separate molecules (*molecular domains*). Strictly speaking, such a division could be performed according to different criteria, the advantages and failures of which are discussed below.

The most valid physical concept that follows from the structural experiment describes crystal space as a continuous electron density distribution, $\rho(r)$. Bader (1990) has given a clear interpretation of the $\rho(r)$ function by analysing its gradient-vector field. Such an approach allows the determination of the *atomic domains* in a crystal. An atomic domain is defined as a spatial region bounded by the surfaces through which there is a zero-flux in the gradient-vector field of the electron density. However, when considering a molecular crystal the domain being dealt with belongs to a molecule rather than to an atom. According to Bader's model, a molecular domain is the union of atomic domains. Therefore, the volume and shape of the molecular domain should be taken as a measure of the size and shape of a molecule in a crystal field.

It is apparent that Bader's method provides a physical background for the intuitive notions of molecular sizes and shapes. However, the $\rho(r)$ function can only be obtained for rather simple molecules and crystals because of computational restrictions. Thus, it is reasonable to appeal to some other approximate models of molecular domains in crystal space.

This widespread method uses the values of the van der Waals radii (Bondi, 1964; Kitaigorodskii, 1973; Zefirov & Zorky, 1995). In this approach a molecule is represented by a set of overlapping van der Waals spheres circumscribed

© 2004 International Union of Crystallography
Printed in Great Britain – all rights reserved

**Table 1**
Parameters of the VDP volume distributions for the molecules that occur most frequently.

$V_m$: the mean value and confidence interval [$V_{min}$; $V_{max}$] are given for each molecule; $V_h$: volume of molecule in homomolecular crystal; $\chi^2$: skewness and kurtosis estimates together with their standard errors are given for the samples of $n < 100$. Skewness and kurtosis estimates together with their standard errors are given for the samples of $n < 100$.

| | | Molecular VDP volume (Å$^3$) | | | |
|---|---|---|---|---|---|
| Molecule | Number of molecules ($n$) | $V_m$ | $V_h$ | $\chi^2_{obs}/\alpha(S_\alpha)$ | $\chi^2/\varepsilon(S_\varepsilon)$ |
| Benzene | 426 | 130.3 [113.7; 146.8] | 122.9 | 8.3 | 11.1 |
| Methanol | 417 | 58.6 [50.7; 66.5] | 54.4 | 19.0 | 11.1 |
| Dichloromethane | 416 | 78.7 [68.2; 89.2] | 82.1 | 6.6 | 9.5 |
| Acetonitrile | 404 | 71.15 [61.0; 81.3] | 66.5 | 13.0 | 11.1 |
| Chloroform | 395 | 94.3 [81.7; 106.8] | 103.8 | 7.8 | 9.5 |
| Acetone | 257 | 104.7 [92.5; 116.8] | 91.6 | 5.9 | 9.5 |
| Dimethylsulfoxide | 200 | 108.5 [96.5; 120.4] | 105.5 | 6.2 | 7.8 |
| Ethanol | 172 | 86.4 [75.2; 97.6] | 74.6 | 2.90 | 6.0 |
| *N*,*N*-Dimethylformamide | 139 | 119.7 [105.6; 133.8] | 102.1 | 2.2 | 7.8 |
| Urea | 139 | 71.3 [64.6; 77.9] | 73.3 | 3.1 | 7.8 |
| Tetrahydrofuran | 127 | 126.0 [107.2; 145.6] | 103.9 | 16.2 | 6.0 |
| 1,4-Dioxane | 121 | 133.1 [115.8; 150.5] | 111.5 | 4.9 | 7.8 |
| 18-Crown-6 | 116 | 372.8 [342.7; 401.9] | 342.3 | 2.1 | 6.0 |
| Toluene | 105 | 157.5 [138.3; 175.4] | 145.1 | 5.0 | 7.8 |
| Thiourea | 100 | 85.9 [77.9; 93.9] | 87.5 | 5.8 | 11.1 |
| Acetic acid | 70 | 82.8 [72.4; 93.1] | 74.3 | 0.5 (3) | 0.6 (6) |
| Diethyl ether | 61 | 146.5 [129.3; 163.7] | 129.3 | 0.0 (3) | 0.0 (6) |
| Pyridine | 54 | 118.3 [106.6; 130.0] | 111.5 | 0.2 (3) | −0.9 (6) |

around its constituent atoms. Thus, a molecular domain is a spatial region bounded by the surfaces of such overlapping spheres. Despite great popularity this approach has essential disadvantages. First, the representation of an atom by a sphere does not account for the electron shell distortions in a crystal field. Besides, no universal set of van der Waals radii has been proposed so far (the radii values for the same atoms vary considerably from each other in different sets) and many metals have no assigned van der Waals radii (Zefirov & Zorky, 1995; Zefirov, 1997). Therefore, an alternative geometrical model of molecular crystals needs to be found.

Such a model has been developed in the last quarter of the past century and rests on the concept of atomic and molecular Voronoi–Dirichlet polyhedra (VDPs). Recently, Blatov & Serezhkin (2000) showed the atomic VDP to be a rough approximation of Bader's atomic domain. Moreover, the average volumes of the monoatomic anions are shown to be independent of anion charge and nearly equal to the volumes of the neutral atoms in the crystal structures of elementary substances. This means that the VDP approach often allows the estimation of not only the relative but also the real sizes of atoms. Thus, a molecular VDP is defined as the union of VDPs of the atoms comprising the molecule and can be interpreted as an image of a molecule in a crystal field. There are several ways to construct molecular VDPs proposed in the literature (Fischer & Koch, 1979; Peresypkina & Blatov, 1999).

According to the concept introduced by Fischer & Koch (1979), molecular VDPs are built up from the VDPs constructed by means of radical planes for the spheres of covalent radii circumscribed around the atoms in a molecule. The authors applied this concept to analyse molecular packings and also to determine the sizes of atoms. They showed (Koch & Fischer, 1980) that the sizes of atoms of a given kind are nearly constant in a wide range of organic compounds. Let us emphasize that in this concept the representation of an atom by a sphere is implicit, and the problem of what sets of atomic radii should be preferred remains.

Another approach to the construction and applications of molecular VDPs was developed by Peresypkina & Blatov (1999). In this model VDPs for the atoms in a molecule are constructed according to the VDP definition, *i.e.* polyhedra are formed by intersecting planes that bisect perpendicularly the lines joining atomic centers.

Several geometrical characteristics of molecular VDPs are shown to have physical meaning, namely:

(i) the volume ($V_{VDP}$) that assesses the molecular size in a crystal field;

(ii) the spherical domain radius ($R_{sd}$), *i.e.* the radius of the sphere of a molecular VDP volume;

(iii) the second normalized moment ($G_3$) that implicitly estimates the shape of a molecule;

(iv) the surface area ($S_{VDP}$) that is proportional to the strength of all intermolecular interactions, in which a given molecule is involved, since the VDP faces correspond to the contacts with the atoms of neighbouring molecules.

It was demonstrated that this approach gives reliable results in calculating molecular coordination numbers (Peresypkina & Blatov, 1999, 2000*a*), in topological analysis of molecular packings (Peresypkina & Blatov, 2000*b*) and in assessing molecular shapes (Peresypkina & Blatov, 2003). This method was found to be correct for comparative analysis, in particular, for predicting organic substrates that can occupy the receptor cavity (Virovets *et al.*, 2004) or for comparing both atomic and

**Table 2**
Environment of molecules with minimum and maximum volumes.

| Molecule | Minimum volume value | $S_\Sigma$ (%) | Maximum volume value | $S_\Sigma$ (%) |
|---|---|---|---|---|
| Benzene | 109.1 | 0 | 166.4 | 68.6 (Cl) |
| Methanol | 49.3 | 0 | 74.64 | 19.7 (Cl) |
| Dichloromethane | 63.2 | 0 | 98.7 | 30.3 (Br) |
| Acetonitrile | 58.7 | 0 | 92.6 | 2.5 (Cl) |
| Chloroform | 78.7 | 5.7 (I) | 121.5 | 81.5(Cl)† |
| Acetone | 86.3 | 0 | 125.5 | 0 |
| Dimethylsulfoxide | 94.5 | 0 | 135.2 | 38.6 (I) |
| 1,4-Dioxane | 111.5 | 0 | 167.5 | 58 (I) |
| Ethanol | 69.3 | 0 | 109.1 | 7.7 (Br)† |
| *N*,*N*-Dimethylformamide | 99.6 | 0 | 140.6 | 5.4 (S)† |
| Urea | 64.8 | 0 | 80.0 | 0 |
| Tetrahydrofuran | 100.7 | 0 | 162.7 | 5.9 (Br)† |
| 18-Crown-6 | 341.3 | 0 | 433.0 | 0† |
| Toluene | 142.4 | 15.5 (Cl) | 202.1 | 89.2 (P, Cl) |
| Thiourea | 77.9 | 0 | 96.2 | 0 |
| Acetic acid | 73.3 | 0 | 99.8 | 0 |
| Diethyl ether | 128.4 | 0 | 168.8 | 0† |
| Pyridine | 108.4 | 0 | 132.3 | 0 |

† The molecule is disordered or surrounded by disordered molecules.

molecular volumes in different crystal structures (Blatov & Serezhkin, 2000; Blatova *et al.*, 2001). Let us also emphasize that no atomic radii sets are required and the sizes and shapes of the molecules can be estimated without using any information about the $\rho(r)$ spatial distribution.

Comparing the aforementioned concepts of molecular domains, notice the following common feature: molecular domains are derived from atomic domains and, hence, molecular volume is the sum of the atomic contributions. That is why the determination of the sizes of atoms or polyatomic groups has been the subject of much research work. At the same time, insufficient attention has been paid to the volume of a molecule as a whole. Using the molecular VDP approach, Blatova *et al.* (2001) recently found volumes of organic $\pi$-ligands to be independent of their coordination type in the crystal structures of rare-earth $\pi$-complexes.

Thus, the goal of this study is to assess molecular sizes as the molecular VDP volumes in a wide range of organic crystals and to compare the results with those observed for homomolecular crystal structures.

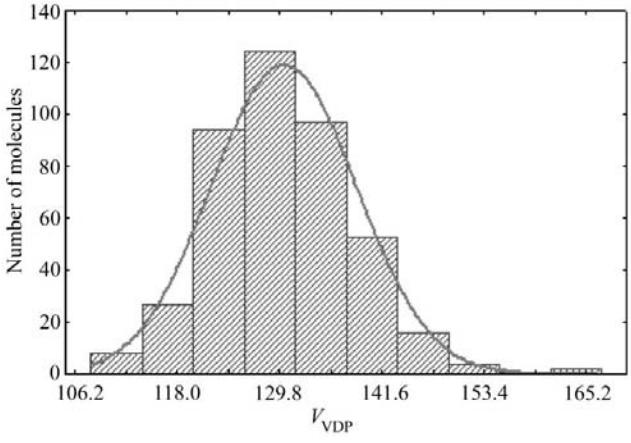


**Figure 1**
Volume distribution for the sample of benzene molecules and the corresponding normal curve (solid line).

## 2. Experimental

Crystallographic data on 69 011 compounds containing non-coordinated molecules (chemically isolated molecules, *i.e.* molecules not involved in any valence bonds with other atoms or molecules) were selected from the Cambridge Structural Database (release 5.24) by means of the program package *TOPOS* (Blatov *et al.*, 2000). All the compounds are completely refined structures without metal atoms. To construct molecular VDPs we used the method by Peresypkina & Blatov (1999) implemented in the *ADS* program (part of the *TOPOS* package). Molecular volumes in heteromolecular crystals were estimated as molecular VDP volumes, whereas those in homomolecular crystals were calculated as the ratio of the unit-cell volume to the number of molecules per unit cell. As a result, 100 632 molecules (both neutral and ions) were detected and volume distributions were constructed for the most frequent molecules (Table 1; Fig. 1).

The assumption that the molecular volume data are normally distributed was verified by performing the chi-square test at the confidence level $p = 0.95$, with a sample size ($n$) of not less than 100, or by calculating the skewness ($\alpha$) and kurtosis ($\varepsilon$) estimates together with their standard errors ($S_\alpha$ and $S_\varepsilon$) if $n < 100$. In the chi-square test we computed the $\chi^2$ observed value ($\chi^2_{obs}$) that assesses the difference between theoretical and empirical distributions, and the $\chi^2$ critical value ($\chi^2_c$) that determines the upper allowable limit for this difference. Thus, the normality hypothesis cannot be rejected if $\chi^2_{obs} < \chi^2_c$ if $n \geq 100$ or $|\alpha| < 3S_\alpha$ and $|\varepsilon| < 5S_\varepsilon$ if $n < 100$.

Additionally, we tested how much the volume of a molecule in a homomolecular crystal deviates from the mean value for all (including heteromolecular) crystals. This test provides statistical evidence about how the environment of a molecule influences its volume. As is known (Brownlee, 1965), for a normally distributed sample $\mathbf{A}$ ($a_1$, $a_2$, ... $a_n$) it is possible to obtain a confidence interval ($a_{min}$; $a_{max}$) for $a_i$ at the specified confidence level $p$ (we assumed $p = 0.95$) using a random Student's variable $\eta$

$$\eta = |\tau|(n-2)^{1/2}/(n-1-\tau^2)^{1/2}, \qquad (1)$$

where $|\tau| = (\bar{a} - a_{min})/\sigma = (a_{max} - \bar{a})/\sigma$; $\bar{a}$ and $\sigma$ are the sample mean and standard deviation, respectively. Solving (1) for both $a_{min}$ and $a_{max}$ gives the confidence interval. Thus, if $a_{min} < a_i < a_{max}$, then $a_i$ does not deviate significantly from the sample mean. In other words, if the volume of a molecule in a

homomolecular crystal lies within the confidence interval, the volumes of the molecule in homomolecular and heteromolecular crystals could be considered to be practically coincident.

To consider what influences the molecular volume, we studied the environment of the molecules with minimum and maximum volume values (Table 2) and assessed the contribution of the contacts between the molecules and the high-row surrounding atoms (other than H, Li—F) as follows

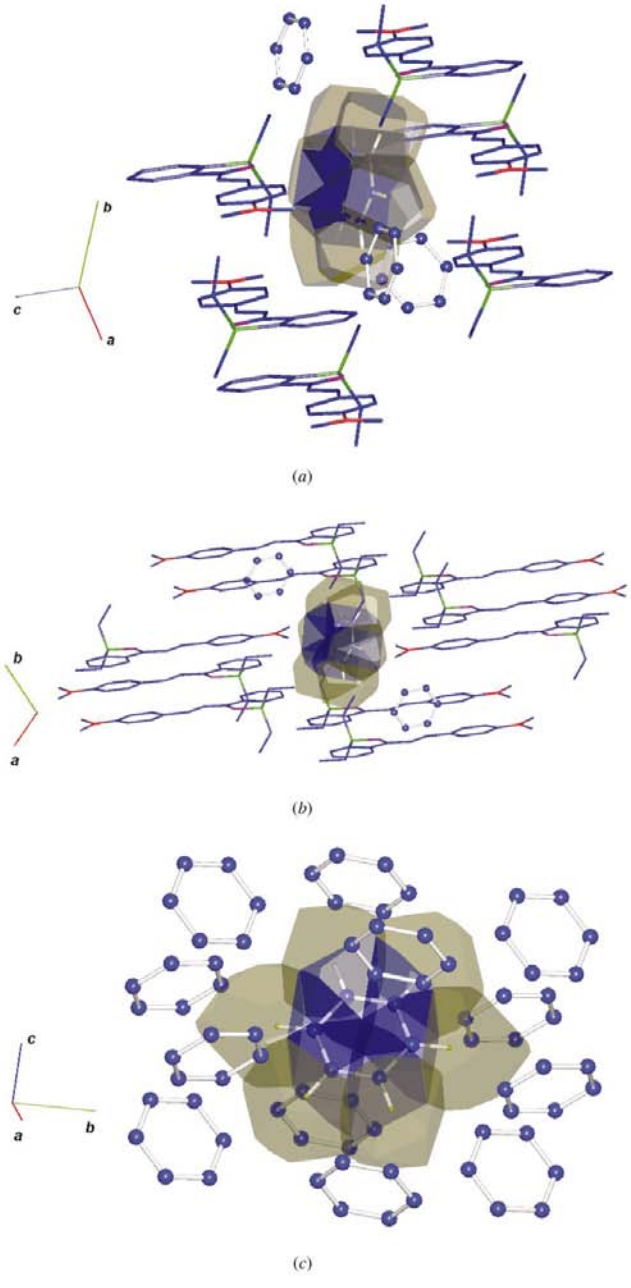


**Figure 2**
Molecular VDPs and the surrounding molecules for the benzene molecules in the crystal structures of (*a*) and (*b*) 2′-diethylboryl-4″-dimethylaminochalcone benzene solvate [HOKLEY]: (*a*) $V$ = 109.1 Å$^3$, (*b*) $V$ = 122.7 Å$^3$ and (*c*) benzene [BENZEN] ($V$ = 122.9 Å$^3$). H atoms of surrounding molecules are hidden for better viewing.

$$S_\Sigma = (S_h/S_{\rm VDP}) \cdot 100\%, \qquad (2)$$

where $S_h$ is the sum of the areas of molecular VDP faces corresponding to the contacts between the molecule and the high-row atoms.

## 3. Results and discussion

The main conclusion from the data obtained (Table 1) is that the molecular volume in a homomolecular crystal is nearly equal to its average value for all the compounds considered. This conclusion follows from the fact that almost all molecular volumes in the homomolecular crystals lie within the confidence intervals, whose half-width is *ca* 10% of the mean value, irrespective of composition. The volume values for acetone, 1,4-dioxane, ethanol, *N*,*N*-dimethylformamide, tetrahydrofuran and 18-crown-6 vary slightly from the lower bounds of the intervals and the difference could be caused by experimental errors. Therefore, the mean volume of a given molecule deviates insignificantly from that in the homomolecular crystal. Recall that molecular volumes in homo- and heteromolecular crystals were calculated separately (§2), therefore the coincidence found confirms that the VDP approach correctly estimates not only the relative, but also the real sizes of molecules. These results bear resemblance to those of Blatov & Serezhkin (2000), who report that electronegative atoms have VDP volumes in organic compounds similar to those in ionic and elementary substances. As a whole, VDPs can provide an appropriate model for both atomic and molecular domains in a crystal field.

Almost all the molecular volume samples can be accurately modelled by a normal distribution (Fig. 1), except the samples for methanol, acetonitrile and tetrahydrofuran molecules. Note that a variable tends to be normally distributed if it is determined by an infinite number of independent random factors and if the influence of each factor is negligible. Therefore, the influence of a crystal field on the size of a molecule is found to be slight and almost equal in homo- and heteromolecular crystals. It should be emphasized that this result indicates that only a common trend that could be violated in particular crystals. Let us consider as an example benzene molecules in different environments.

The structure of 2′-diethylboryl-4″-dimethylaminochalcone benzene solvate [HOKLEY][1] contains two non-equivalent benzene molecules, with volumes equal to 109.1 Å$^3$ (the minimum value for this molecule in the sample) and 122.7 Å$^3$. The first molecule is surrounded by six 2′-diethylboryl-4″-dimethylaminochalcone and three benzene molecules (Fig. 2*a*). In other words, its molecular coordination number (Peresypkina & Blatov, 2000*a*), MCN, is 9 and can be written as 6 + 3 to distinguish between the different chemical natures of the molecules. The second benzene molecule is surrounded by ten 2′-diethylboryl-4″-dimethylaminochalcone and two benzene molecules (MCN = 10 + 2), therefore, its environment is similar to a close packing (Fig. 2*b*). For comparison, in

[1] Hereinafter, the Refcodes of compounds in the Cambridge Structural Database are given in square brackets.

the homomolecular crystal [BENZEN] the benzene molecules are arranged according to the face-centred cubic (f.c.c.) topology (Fig. 2*c*) and have MCN = 12 ($V_{VDP}$ = 122.9 Å$^3$). Therefore, the volume of the benzene molecule with MCN = 10 + 2 in [HOKLEY] is close to that in the homomolecular crystal. This example shows that the same molecules, even in a similar environment, may be of different sizes; the difference probably reflects peculiarities of a molecular packing or specific features of electron density distribution in the vicinity of a molecule.

As a rule, the following factors increase the molecular volume:

(i) predominance of high-row atoms in the molecular environment;

(ii) disordering of the central molecule or the surrounding molecules.

The first factor is related to the large sizes and high polarizabilities of the high-row atoms in comparison to hydrogen and the second-row atoms. There can be two reasons for the increase in molecular volumes in this case. First, our method for constructing molecular domains, namely equidistant division of space between small and large atoms, could result in an artificial increase in the sizes of the small atoms and a decrease in the sizes of the large atoms. However, it was the method that allowed Blatov & Serezhkin (2000) to estimate the real sizes of the halogen atoms in organic compounds. Therefore, this 'geometrical' reason may not be sufficient to explain the increase in molecular volume. Second, the electron density could overflow from the large surrounding atoms to a molecular domain that leads to increasing its size. This 'physical' reason has already been considered by Blatov & Serezhkin (2000) to substantiate the use of VDPs for heterogeneous compounds by the effect of electronegativity equalization and the resulting equalization of the sizes of atomic domains. Out of 18 molecules with maximum volumes, three (benzene, 1,4-dioxane and toluene) interact mostly with the high-row atoms and thus have volumes which are larger than the mean values (Table 2). Fig. 3 shows the benzene molecule of maximum volume ($V_{VDP}$ = 166.4 Å$^3$) with *ca* 70% of all the intermolecular contacts with Cl atoms in the structure of perchlorotriphenylmethyl benzene clathrate [FIDVUJ]. Moreover, if a molecule includes the high-row atoms involved in molecular interactions, its volume in a homomolecular crystal is greater than the mean value. Indeed, this effect is observed for dichloromethane, chloroform and, to some extent, thiourea, while dimethylsulfoxide, where sulfur is shielded by other atoms, has practically the same $V_m$ and $V_h$ (Table 1). Such an effect can be explained by the fact that molecules in organic crystals are commonly surrounded by H and second-row atoms, but this is not the case for homomolecular crystals of the molecules with terminal high-row atoms. On the contrary, molecules consisting of light atoms have exceeded the mean volumes compared with those in homomolecular crystals (Table 1).

The second factor is also clear because disordering results in an increase of the number of positions occupied by the molecule in crystal space and, hence, in an increase of its volume. Out of 18 molecules with maximum volumes, four (chloroform, ethanol, tetrahydrofuran and diethyl ether) are disordered in the corresponding crystal structures. Besides, the two molecules (*N*,*N*-dimethylformamide and 18-crown-6) surrounded by disordered molecules also have maximum volumes.

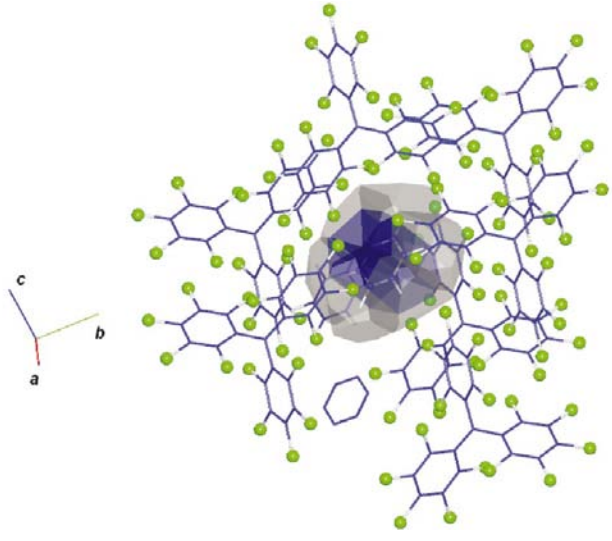


**Figure 3**
Molecular VDPs and the surrounding molecules for the benzene molecule in the crystal structure of perchlorotriphenylmethyl benzene clathrate [FIDVUJ] ($V$ = 166.4 Å$^3$). Cl atoms are shown as balls. The H atoms of the neighbouring benzene molecule are hidden.

## 4. Conclusions

The results show that molecular VDPs can be used to model molecular domains and to reveal factors causing the variations in domain sizes. However, the factors discovered are not the only reasons that the molecular volumes increase (Table 2). Besides, no clear explanation was found for the decreasing molecular volumes (Table 2). To elucidate these facts a detailed topological analysis of the molecular packings should be performed. At the same time, the methods for analysing molecular packings are aimed mainly at homomolecular crystals (Peresypkina & Blatov, 2000*b*), whereas most of the molecules considered are solvates in complex structures. Thus, these methods should be developed to investigate the molecular crystals of any composition and structure complexity.

## References

Bader, R. F. W. (1990). *Atoms in Molecules – a Quantum Theory*. Oxford University Press.
Blatov, V. A. & Serezhkin, V. N. (2000). *Russ. J. Inorg. Chem.* Suppl. S105–S220.
Blatov, V. A., Shevchenko, A. P. & Serezhkin, V. N. (2000). *J. Appl. Cryst.* **33**, 1193.
Blatova, O. A., Blatov, V. A. & Serezhkin, V. N. (2001). *Acta Cryst.* B**57**, 261–270.
Bondi, A. (1964). *J. Phys. Chem.* **68**, 441–451.

Brownlee, K. A. (1965). *Statistical Theory and Methodology in Science and Engineering*. New York: John Wiley and Sons.
Fischer, W. & Koch, E. (1979). *Z. Kristallogr.* **150**, 245–260.
Kitaigorodskii, A. I. (1973). *Molecular Crystals and Molecules*. New York: Academic Press.
Koch, E. & Fischer, W. (1980). *Z. Kristallogr.* **153**, 255–263.
Peresypkina, E. V. & Blatov, V. A. (1999). *J. Mol. Struct. (Theochem.)* **489**, 245–236.
Peresypkina, E. V. & Blatov, V. A. (2000*a*). *Acta Cryst.* B**56**, 501–511.
Peresypkina, E. V. & Blatov, V. A. (2000*b*). *Acta Cryst.* B**56**, 1035–1045.
Peresypkina, E. V. & Blatov, V. A. (2003). *Russ. J. Inorg. Chem.* **48**, 237–245.
Virovets, A. V., Blatov, V. A. & Shevchenko, A. P. (2004). *Acta Cryst.* B**60**, 350–357.
Zefirov, Yu. V. (1997). *Crystallogr. Rep.* **42**, 111–116.
Zefirov, Yu. V. & Zorky, P. M. (1995). *Usp. Khim.* **64**, 446–461.