Acta Crystallographica Section B
Structural
Science

ISSN 0108-7681

Topology of molecular packings in organic crystals

E. V. Peresypkina and V. A. Blatov*

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

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

Received 15 May 2000
Accepted 29 August 2000

Using the methods of coordination sequences and of molecular Voronoi–Dirichlet polyhedra, the topological properties of molecular packings and molecular coordination numbers (MCNs) were determined in the crystal structures of 33 575 monosystem organic compounds within the first three coordination spheres. Numerous examples of disagreement between the topology of molecular packing and the system of intermolecular contacts in a crystal structure were found. It is concluded that within the first coordination sphere most of the molecules tend to arrange with MCN = 14, obeying the model of the thinnest covering of space, but molecular packings as a whole tend to be constructed according to one of the close packings.

1. Introduction

During the analysis of molecular packings in crystal structures one usually investigates only the short-range environment of a molecule, analyzes its intermolecular contacts and calculates the molecular coordination numbers (MCNs). Recent studies of the molecular crystal structures of simple substances (Peresypkina & Blatov, 1999) and of more than 23 000 organic compounds (Peresypkina & Blatov, 2000) have shown that a molecule can be correctly represented in a crystal field generally by a nonconvex molecular Voronoi–Dirichlet polyhedron (VDP). A molecular VDP is a union of all the VDPs of atoms in the molecule and its faces correspond to interatomic contacts between different molecules. The interaction strength between two molecules can be evaluated with the $\Omega_{\rm mol}$ parameter

$$\Omega_{\rm mol} = \left(\sum_i \Omega_i\right) / \Omega_\Sigma \times 100\%, \qquad (1)$$

where Ω_i is the solid angle corresponding to the ith intermolecular contact, Ω_Σ is the sum of the solid angles of all the nonvalent bonds formed by a molecule accepted as central.

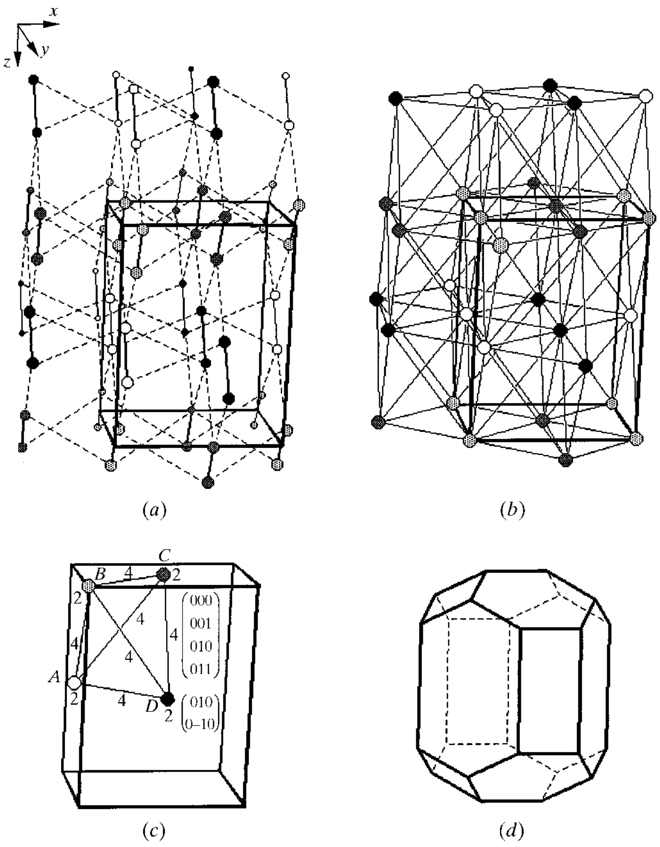
To characterize the topology of the short-range environment of a molecule Peresypkina & Blatov (2000) have proposed the use of so-called *smoothed* and *lattice* VDPs constructed for molecular centroids. A smoothed molecular VDP is constructed with only the molecules of packing, immediately connected to a given molecule, and is used to characterize the local topology of a system of intermolecular contacts or, in other words, the molecular *net*. Therefore, smoothed molecular VDPs do not necessarily form a partition of space, but the number of faces of any such VDP is equal to MCN. The lattice molecular VDP is constructed according to a standard algorithm (Blatov *et al.*, 1999) for a sublattice of molecular centroids, irrespective of existing intermolecular contacts between them. Lattice molecular VDPs characterize the local topology of the spatial *packing* of molecules, not necessarily in contact with each other, and always form a partition of space. However, the number of faces of the lattice molecular VDP is not generally equal to MCN. To emphasize this difference further we shall call the number of faces of lattice molecular VDP the *molecular packing number* (MPN).

Although the method described by Peresypkina & Blatov (1999, 2000) provides some additional resources to study the topology of crystalline molecular compounds, it is based on the investigation of only the first coordination sphere of a molecule. Thus, this study has been performed to extend the method of topological analysis of molecular packing up to any specified depth and to verify it with modern crystal structural data of molecular organic compounds.

2. Method of analysis of global topology for molecular packings

During the topological analysis any crystal structure can be associated with an infinite nonoriented graph (with the *complete* graph of crystal structure, Fig. 1*a*), whose nodes symbolize atoms and edges connecting the nodes imitate the real system of all interatomic contacts typical for a given compound. Although the complete graph shows all the information on interatomic bonds (on the crystal structure *connectedness*), such a model is too complicated for the analysis and usually only enables the investigation of the local topology of a crystal structure. To study the global topological properties the complete graph of a crystal structure is conveniently reduced to a finite *reduced* graph (RG), whose nodes are contained within the unit cell (Blatov, 2000). Keeping the information on the connectedness of the atoms outside the unit cell, all edges corresponding to interatomic contacts, outside the unit cell or on its boundary, should be closed to translationally equivalent nodes conforming to the atoms inside or on the boundary of the unit cell. Such RG edges can be multiple if there are bonds between an atom in the original crystal structure and several translationally identical atoms of a different type, and loops can appear at the nodes if there are bonds between translationally equivalent atoms. The method of reducing contains all the information of the original infinite graph if the nodes and edges of the RG are appropriately labeled (Fig. 1*c*). If there are two or more variations (colors) in nodes, the division of the RG onto subgraphs containing nodes of only one type is possible. The connectedness of a subgraph is kept if the division of an RG is performed by the operation of contracting edges, *i.e.* by sequential removal of all nodes, except those of specified colors, and closing the open ends of edges. The lattice graphs obtained describe the topology for selected atomic sublattices (Blatov, 2000).

During the study of molecular packing topology it became clear that the method of binding molecules is firstly important and that their interior structure can be disregarded. Therefore, it is useful to analyze a *graph of molecular net* (Fig. 1*b*), whose nodes coincide with molecular centroids and each of its edges corresponds to a set of contacts between a pair of molecules. Such a graph may be constructed with the complete graph of a crystal structure by contracting all the nodes, symbolizing atoms of a given molecule to a molecular centroid. For instance, in the packing of Cl_2 molecules (Figs. 1*a* and *b*), Cl atoms of a molecule should be contracted to a molecular centroid. At the same time the edges of the complete graph connecting different Cl atoms of a pair of adjacent molecules should be replaced by a single edge between their centroids. This edge should be labeled by the number of intermolecular contacts between a given pair of molecules (to the multiplicity



(*a*) (*b*)

(*c*) (*d*)

Figure 1
(*a*) A fragment of an infinite complete graph of atomic net for the crystal structure of chlorine (Stevens, 1979; solid lines mark the edges of the graph conforming to valent contacts between Cl atoms, *i.e.* between nodes of the graph; dashed lines show the edges corresponding to intermolecular contacts); its sequential transformations into (*b*) the graph of molecular net by contracting atoms of molecules to their centroids (solid lines show the edges of the graph; each edge corresponds to a set of contacts between a pair of molecules; the labels corresponding to edge multiplicities and to the numbers of loops at nodes are omitted); and then to (*c*) the reduced graph of molecular net (RGMN) with labels indicating edge multiplicities and the numbers of loops at the nodes. The labels indicating the translation codes of the reduced centroids are shown as an example only for one node and one edge (see also Blatov, 2000). Translationally nonequivalent molecules (*a*), and the corresponding nodes of graphs (*b*) and (*c*) are shaded differently. In particular, each *A* node of RGMN is connected with two *A* nodes, translationally equivalent to it (the node multiplicity is equal to 2), and also with four nodes of each of the *B*, *C* and *D* types (the multiplicity of each edge connecting node *A* with *B*, *C* and *D* nodes is equal to 4). The molecular lattice VDP (*d*) and the molecular smoothed VDP are combinatorially equal to the Fedorov cuboctahedron.

Table 1
Coordination sequences for some lattices.

Lattice	Coordination sequence, $\{N_k\}$ (k = 1–3)	Analytical expressions (O'Keeffe, 1995)
Body-centered cubic	14,50,110	$N_k = 12k^2 + 2$
Face-centered cubic	12,42,92	$N_k = 10k^2 + 2$
Hexagonal close	12,44,96	$N_k = 10.5k^2 + 2$, k is even $N_k = 10.5k^2 + 1.5$, k is odd

of an edge, Fig. 1*b*). If we ignore the interior structure of a molecule and the number of contacts between adjacent molecules, we should omit the set of labels associated with the edges and nodes of the graph of the molecular net after contracting all atoms of a molecule to its centroid. As a result, the information on the number of intermolecular contacts bonding the molecules is lost, because each edge of the graph of molecular net symbolizes a set of contacts between a pair of molecules. However, with respect to the topology of the molecular net, the information lost is not essential, since the important fact is the presence or absence of intermolecular contacts. Thus, the labels only appear at the nodes and edges of the step where reducing the infinite graph of molecular net to a *reduced graph of molecular net* occurs (RGMN; Fig. 1*c*), which restricts any information on the connectedness of the crystal structure to the intermolecular level. For polysystem molecular crystals, a separate analysis of subnets is possible for different types of molecules that correspond to the decomposition of RGMNs to *lattice* graphs of the molecular net. In this paper only monosystem molecular crystals are considered, therefore, no additional procedure of decomposition is required for RGMNs.

If only the spatial arrangement (packing) of molecules is of interest, but not the real system of intermolecular contacts in a crystal, the connectedness of the net of molecular centroids can be determined by partitioning the crystal space by lattice VDPs (Fig. 1*d*). In this case an edge between two nodes of the graph of molecular net exists only if there is a common face of their lattice VDPs. Thus, the presence of intermolecular contacts or 'tangency points' between molecules is not essential. By reducing the graph of molecular net it is possible to obtain a *reduced graph of molecular lattice* (RGML), whose edges determine the topology of a *lattice* of molecular centroids. It should be noted that, in general, RGMN and RGML are not isomorphic to each other, just as in the case of smoothed and lattice molecular VDPs. However, such isomorphism is highly probable for crystal structures of compounds with molecules whose shape is close to spherical.

The isomorphism of molecular RGs can be investigated by the method of coordination sequences proposed by Brunner & Laves (1971), widely applied in the investigation of the topology of atomic nets in inorganic crystals, especially zeolites (Grosse-Kunstleve *et al.*, 1996), and also for the analysis of RGs (Blatov, 2000). This method is based on the calculation of the composition of long-range coordination spheres for nodes of a given graph of crystal structure. Note that the *k*th coordination sphere (the *k*th layer of nodes), corresponding to a given node of the graph, includes all the nodes which are directly connected to the nodes of the $(k-1)$th coordination sphere, but not included in the first $(k-1)$ coordination spheres. It is the set of integers $\{N_k\}$ $(k = 1 - n)$, each of which is equal to the number of nodes in the *k*th coordination sphere, that is the *coordination sequence* for a given node. If the sets $\{N_k\}$ for a pair of RGs are not equal, the graphs are strictly nonisomorphic, but if their sets $\{N_k\}$ are the same one can assume that the topological properties of the graphs compared are identical only within the first *n* coordination spheres. Thus, equal $\{N_k\}$ sets do not necessarily indicate strict isomorphism of the graphs compared, although in this case the greater the *k* value the more similar the topologies of the RGs compared. In this study the topological properties of molecular nets and lattices were investigated within the first three coordination spheres which, according to Blatov (2000), allows one to assume an essential isomorphism of molecular RGs.

The topological properties of many lattices and their corresponding coordination sequences are well known and obey the analytical equations, with a quadratic dependence of the number of coordination spheres *k* (O'Keeffe, 1995; Conway & Sloane, 1997; Table 1). Further, if any node of a net or a lattice is characterized by the set $\{N_k\}$ fitting one of the equations given in Table 1 within the first three coordination spheres, we shall assume that their topologies are similar to

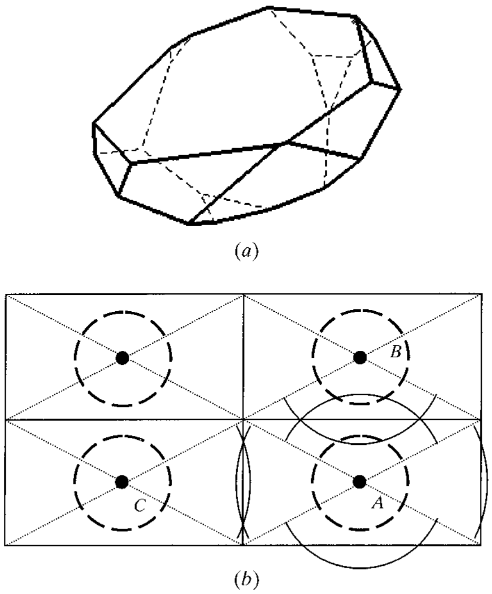


Figure 2
(*a*) VDP of the C(17) atom in the crystal structure of androst-5-ene-3,17-dione (ANDEDP10), associated with an atomic domain. (*b*) An example of the identification of interatomic contacts by the method of spherical sectors in a two-dimensional lattice. Bold lines confine VDPs; dashed lines show the boundaries of pyramids (triangles in this case) based on VDP faces corresponding to the 'basic' interatomic contacts. Dashed circles of radius r_s conform to the maxima of the valent electronic density of atoms; solid arcs of radius $r_{\rm sec}$ confine spherical sectors and show atomic boundaries in a crystal field. Atoms *A* and *B* form a valent contact, to which the triple overlap $r_{\rm sec}(A) - r_s(B)$, $r_s(A) - r_{\rm sec}(B)$ and $r_{\rm sec}(A) - r_{\rm sec}(B)$ corresponds; the contact between *A* and *C* atoms is nonvalent, because the single overlap $r_{\rm sec}(A) - r_s(B)$ corresponds to it.

Table 2
The distribution of molecular VDPs on the combinatorial–topological types for the sample with 33 575 monosystem organic compounds.

The combinatorial–topological VDP types are given which compose no less than 1% of the sample. The combinatorial–topological VDP type is written as $f/v - k$, where f and v are the numbers of VDP faces and vertices, respectively, and k is an ordinal number of the VDP type with given f and v.

Smoothed molecular VDPs			Lattice molecular VDPs		
	Relative part (%)			Relative part (%)	
VDP type	(Peresypkina & Blatov, 2000)	This study	VDP type	(Peresypkina & Blatov, 2000)	This study
14/24-1	13.6	13.9	14/24-1	38.2	37.7
14/24-2	7.1	7.0	12/18-1	12.0	12.0
14/24-3	4.6	4.4	14/24-4	11.7	11.9
14/24-4	3.2	3.4	14/24-2	10.5	10.6
12/20-1	2.6	2.5	16/28-2	3.5	3.7
14/24-5	2.1	2.0	14/24-7	3.2	3.1
14/24-6	1.8	1.9	14/24-8	3.2	3.1
14/24-7	1.5	1.4	16/28-4	1.9	2.2
14/22-1	1.0	1.3	16/28-3	2.1	1.9
14/24-8	1.1	1.1	18/32-1	1.0	1.1
16/28-1	1.0	1.1			

those of the corresponding standard net or lattice, and that the corresponding RGs are essentially isomorphic. Apparently, essential isomorphism represents strict isomorphism for the different types of lattices, given in Table 1, because counter-examples are unknown for three-dimensional space (see also O'Keeffe, 1995; Blatov, 2000).

The method of coordination sequences enables the topological relationship for mono- and polysystem molecular packings to be found by searching for the isomorphism of molecular RGs up to the *n*th layer of nodes, and to classify molecular packings by comparing their coordination sequences with those of known lattices. Unlike the method of analysis of the combinatorial or geometrical characteristics of molecular VDPs (Peresypkina & Blatov, 1999, 2000), which depends on the degree of sphericity of molecules, the method proposed here enables the identification of molecular packing by topology with any degree of distortion.

3. Method of constructing a complete graph of crystal structure

One of the main problems with analysis of the topology of molecular packings is the absence of unambiguous crystal chemical criteria to search for intermolecular contacts. To identify nonvalent bonds a crystal chemist usually compares interatomic distances with the sum of the corresponding van der Waals radii of atoms. However, there is no conventional approach to estimate their values and there are some systems where the radius values can differ from each other by almost 1 Å for the same atom (Zefirov, 1997). A comparison of efficiency of different methods of MCN calculation has shown (Peresypkina & Blatov, 1999) that the use of van der Waals radii sometimes results in MCNs which are too small or even ambiguous. Besides, when analyzing the large number of molecular packings the problem of automatic identification of valent contacts appears, which cannot always be solved unambiguously using only the values of interatomic distances and atomic radii.

For the determination of chemical bonds of any nature Serezhkin *et al.* (1997) proposed the method of intersecting spheres, with which the presence of contacts between atoms is revealed by analyzing the number overlapping pairs of interior and exterior spheres circumscribed around each atom of a given pair. The interior sphere usually has a radius equal to Slater's radius (r_s) of the corresponding atom, and the exterior one has a radius equal to that of a spherical domain (R_{sd}), *i.e.* the radius of a sphere whose volume coincides with the volume of atomic VDP. If there is more than one simultaneously overlapping pair of such spheres, this contact should be considered as a chemical bond and should be taken into account during the determination of the coordination number of this atom. The efficiency of this method has been proved by analyzing several dozens of compounds of a different chemical nature (Serezhkin *et al.*, 1997). In fact, the method of intersecting spheres assumes that the shape of an atomic domain in the crystal structure remains approximately spherical. For the compounds studied here this is a doubtful assumption because atomic domains are usually nonspherical in organic molecules (Fig. 2*a*).

To take into account the nonsphericity of an atomic domain we propose an improved method of intersecting spheres called the method of spherical sectors, which is implemented in the program *AutoCN* integrated into the program package for crystal chemical analysis *TOPOS* (Blatov *et al.*, 1999). According to this method the sphere of R_{sd} radius is substituted by a set of spherical sectors corresponding to separate interatomic contacts, whose number is equal to the number of 'basic' VDP faces of an atom. Note that a basic VDP face is formed by a 'direct' neighbour, *i.e.* by an atom for which a segment connecting it with the central atom intersects this face. The radius (r_{sec}) of each sector may be calculated as follows

$$V_p = 1/3 \cdot \Omega \cdot r_{\rm sec}^3, \qquad (2)$$

where V_p and Ω are the volume and the solid angle of a pyramid based on a VDP face corresponding to an interatomic contact, with the central atom on top. Within the framework of this approach the interior Slater's sphere corresponds to the maximum electron density in a valent shell of an atom and the sector of the sphere confines the atom within the direction of an interatomic contact in the crystal field. Note that within the method of spherical sectors, unlike the method of intersecting spheres, at least one $r_{\rm sec}(A) - r_{\rm sec}(B)$ overlap conforming to the interaction of domains of adjacent atoms A and B always occurs for any interatomic contact between 'direct' neighbours. At the same time, as in the method of intersecting spheres, a necessary condition of existence of a chemical bond is the presence of at least one other overlap $r_{\rm sec}(A) - r_s(B)$, $r_s(A) - r_{\rm sec}(B)$ or $r_s(A) - r_s(B)$, corresponding to an interaction of atomic valent shells (Fig. 2*b*). Owing to errors in atomic coordinates the construction of atomic VDPs can

Table 3
The distribution of molecular net types for the sample with 33 575 monosystem organic compounds.

The molecular net types are given which compose no less than 1% of the sample. The coordination sequence is written as $\{N_k\}$, where N is the number of nodes in the kth coordination sphere (k = 1–3).

All faces of molecular VDPs are considered			The VDP faces with $\Omega_{\rm mol}$ >1% only are considered		
N_{1-3}	Number of nets	Relative part (%)	N_{1-3}	Number of nets	Relative part (%)
14,50,110	7458	22.2	14,50,110	6495	19.3
14,52,116	3084	9.2	12,42,92	3469	10.0
14,54,124	2807	8.4	14,52,116	2842	8.5
14,52,114	2483	7.4	14,54,124	2574	7.7
12,42,92	1311	3.9	14,52,114	1871	5.6
16,58,128	1028	3.1	12,44,96	1426	4.2
14,54,126	828	2.5	13,48,106	1278	3.8
16,60,134	730	2.2	13,47,103	1096	3.3
14,53,120	720	2.1	14,54,126	689	2.1
13,48,106	621	1.9	14,53,120	541	1.6
14,53,122	602	1.8	16,58,128	473	1.4
13,47,103	573	1.7	13,48,109	439	1.3
12,44,96	538	1.6	14,53,122	423	1.3
16,60,132	381	1.1	13,47,104	364	1.1
14,53,123	339	1.0	16,60,134	343	1.0

Table 4
The distribution of molecular lattice types for the sample with 33 575 monosystem organic compounds.

The molecular lattice types are given which compose no less than 1% of the sample. See remarks to Table 3.

All faces of lattice molecular VDPs are considered			The lattice molecular VDP faces with Ω >1% are only considered		
N_{1-3}	Number of lattices	Relative part (%)	N_{1-3}	Number of lattices	Relative part (%)
14,50,110	12 756	38.0	12,42,92	9637	28.7
12,42,92	4046	12.0	14,50,110	3798	11.3
14,52,114	3881	11.6	12,44,96	3681	11.0
14,52,116	3611	10.8	10,34,74	2972	8.9
16,60,134	1411	4.2	8,26,56	1023	3.1
16,58,128	1341	4.0	13,47,103	1010	3.0
14,53,120	936	2.8	11,39,85	938	2.8
14,53,122	630	1.9	14,52,116	926	2.8
14,53,123	385	1.2	13,48,106	906	2.7
18,66,146	357	1.1	10,36,78	884	2.6
			14,52,114	826	2.5
			13,47,104	591	1.8
			11,40,88	549	1.6
			11,39,86	354	1.1
			8,28,60	353	1.1

become ambiguous, in particular, small VDP faces can appear or disappear. To avoid such problems in the method proposed the VDP faces with $\Omega \leq 1.5\%$ of the total solid angle 4π sr (steradian; which corresponds approximately to the triple Ω error) are omitted.

The efficiency of the method of spherical sectors for organic compounds over the method of intersecting spheres can be estimated using the following facts. During MCN calculation Peresypkina & Blatov (2000) used the method of intersecting spheres, where atomic nonsphericity was not taken into account. As a result lengthened valent contacts, such as C—C, C—H, Si—H *etc*., have been considered as nonvalent, resulting in a molecule 'decomposing' on two or more pseudomolecular fragments. For example, within the method of intersecting spheres the molecules $C_{14}H_{17}Cl_2N_3O$ in the crystal structure of (2*R*,3*S*/2*S*,3*R*)-2-(2,4-dichlorophenyl)-3-(1*H*-1,2,4-triazol-1-yl)-2-hexanol (FEWLIC) 'decompose' on C_2H_5 and $C_{12}H_{12}Cl_2N_3O$ pseudomolecular fragments because the C(10)—C(11) contact of 1.606 Å is considered as nonvalent. Hereinafter, the refcodes of compounds in the Cambridge Structural Database (1999) are given in brackets. Although such a long C—C bond is apparently caused by errors in atomic coordinates, the method of spherical sectors enables the connectedness within a molecule to be restored accurately, taking into account the deformation of atomic domains also caused by the presence of the lengthened contacts. The radii of the spherical sectors of C(10) and C(11), corresponding to a contact between these atoms, are equal to 0.916 and 0.908 Å, respectively, which, using $r_s(\mathrm{C}) = 0.70$ Å, creates a triple overlap {one $r_{\rm sec}[\mathrm{C(10)}] - r_{\rm sec}[\mathrm{C(11)}]$ and two $r_{\rm sec}[\mathrm{C(10)}] - r_s[\mathrm{C(11)}]$ and $r_s[\mathrm{C(10)}] - r_{\rm sec}[\mathrm{C(11)}]$ overlaps} and indicates the existence of a chemical bond.

As a whole, the molecules in 328 of the 28 214 organic compounds studied by Peresypkina & Blatov (2000) were decomposed on 'fragments' while constructing a complete graph of crystal structure, whereas in this study the corresponding numbers are 178 and 37 561, respectively. For 217 of these 328 compounds the method of spherical sectors allows the correct construction of a complete graph of the crystal structure, which also indicates its advantages in comparison with the method of intersecting spheres. However, the method of spherical sectors does not always enable the system of intramolecular bonds to be assigned correctly if there are contacts which are too long, usually caused by experimental errors.

4. Experimental

Molecular organic compounds with completely determined crystal structures without metal or statistically disordered atoms were investigated from the Cambridge Structural Database (1999). Crystal structures with errors in the experimental data revealed with the program package *TOPOS*, and also compounds comprising chemically nonequivalent molecules, were excluded from the sample. Using the method of spherical sectors for each of 37 561 molecular organic compounds selected according to specified criteria, a complete graph of the crystal structure and corresponding RG were constructed. Then with the program *ADS* (Blatov *et al*., 1999) the RG of each crystal structure was transformed into an RGMN and thus information on molecular binding was drawn. It was found that 3986 compounds contained crystallographically nonequivalent molecules and 178 of them comprised the 'decomposed' molecules. All these compounds were also excluded from consideration. For the remaining 33 575 monosystem molecular compounds the coordination sequences $\{N_k\}(k = 1\text{–}3)$ for the nets and lattices of molecular centroids were calculated using the program *IsoTest* (Blatov, 2000). To study the influence of weak intermolecular contacts on the topological properties of crystal structures, the calculation of coordination sequences for RGMN and RGML was performed in two different ways. For the first variant all intermolecular contacts were taken into account, but during the second one only intermolecular contacts with $\Omega > 1\%$ which were strong enough were considered. Note that in the case of RGMN $\Omega = \Omega_{\rm mol}$, but in the case of RGML Ω is equal to the solid angle of a lattice VDP face (Peresypkina & Blatov, 2000). Molecular smoothed and lattice VDPs were also constructed for the first coordination sphere of molecules ($k = 1$) with all intermolecular contacts taken into account.

5. Results

5.1. Topology of the short-range environment of molecules

According to the results obtained, 33 575 molecular smoothed VDPs are divided into 6381 different combinatorial–topological VDP types, and only for 11 of them do the frequencies exceed 1% of the sample size: for nine tetradecahedra, one dodecahedron and one hexadecahedron (Table 2). The Fedorov cuboctahedron (Archimedean truncated octahedron; type 14/24-1) is the most frequent; its part in the sample is almost twice that of a 14/24-2 tetradecahedron occupying the second place. The corresponding distribution of 33 575 molecular lattice VDPs into 322 different combinatorial–topological VDP types is shown in Table 2 by the ten most frequent types. As well as in the case of smoothed VDPs, the Fedorov cuboctahedron is the most frequent lattice VDP; its part is much more than that of any other VDP. However, the second place is occupied by a dodecahedron in the form of, in general, a distorted dodecahedron with four hexagonal faces.

It should be mentioned that the distributions of smoothed and lattice VDPs given in Table 2 correspond to the distributions obtained by Peresypkina & Blatov (2000) with a smaller sample.

5.2. Topology of molecular packings

As a result of the first variant of calculation where all the intermolecular contacts were taken into account, 1384 topologically different RGMNs were found within the first three coordination spheres of the molecules. The coordination sequences $\{N_k\}$ for (k = 1–3) of the most frequent RGMNs are given in Table 3. The set $\{N_k\} = \{14,50,110\}$, typical for a b.c.c. lattice, is of great predominance (Table 1). The coordination sequences fitting the close packings ({12,42,92} and {12,44,96} for f.c.c. and a h.c.p. lattice, respectively, Table 1) are much less frequent. At the second variant of calculation, which only takes into account intermolecular contacts which are strong enough, 1320 essentially nonisomorphic RGMNs were found. The data obtained (Table 3) indicate that neglecting weak intermolecular contacts does not influence the number of most frequent molecular nets (15 types), but the number of topologically different nets slightly decreases (from 1384 to 1320) and the topology of some systems of intermolecular contacts changes. As a whole, the number of coordination sequences with large N_1 values decreases and the part of $\{N_k\}$ with small N_1 increases. In particular, it results in a small decrease of nets with the topology of a b.c.c. lattice and an abrupt increase of nets with the topology of an f.c.c. lattice. As a result they take the second place after b.c.c. RGMNs, which keep the leading position. It should be mentioned that the part of the nets whose topology is similar to the topology of the h.c.p. lattice increases insignificantly.

The calculation, which takes into account all intermolecular contacts, gives the distribution of 262 types of topologically different RGMLs. Among them only 10 RGMLs are more frequent than 1% of the sample size (Table 4). The relative number of RGMLs with the topology of a b.c.c. lattice is more than three times greater than that of any other lattice. The f.c.c. lattice takes second place and the h.c.p. lattice trails behind in eleventh place. Another variant of calculation, where only strong intermolecular contacts were considered, gives a distribution of 339 types of essentially nonisomorphic RGMLs. Compared with the first variant of calculation, the list of the most frequent lattices is extended (from 10 to 15 lattices) and their places in this list are redistributed. As well

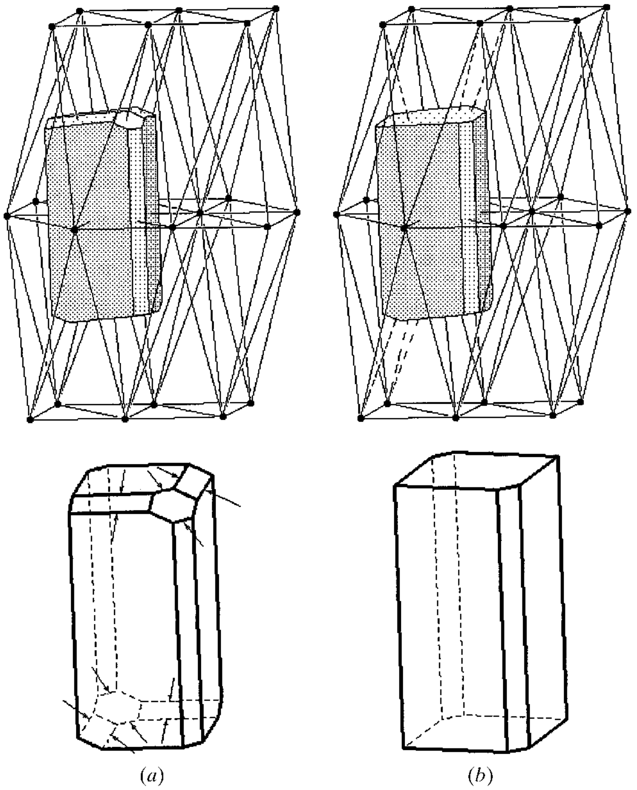

(*a*) (*b*)

Figure 3
The transformation of (*a*) a fragment of the molecular lattice graph with b.c.c. lattice topology into (*b*) a graph of the primitive hexagonal lattice in the crystal structure of 3,4,7,8-dibenzotricyclo(4.2.0.0-2,5-)octa-3,7-diene (BZCOCT) after the omission of six weak intermolecular contacts (by three contacts from each side of the six-coordinated layer; the contacts removed are indicated by dashed lines). During the corresponding transformation of (*a*), a lattice molecular VDP in the form of the Fedorov cuboctahedron, into (*b*) a tetrahedral prism with six faces transforms into edges, as shown by arrows.

as for molecular nets, the number of coordination sequences with a large N_1 value decreases and the number of $\{N_k\}$ with small N_1 increases. In particular, RGMLs with b.c.c. lattice topology drop to second place, decreasing by ~27%, while f.c.c. lattices ascend from second to first place, increasing their part in the sample by more than 16% (Table 4). The h.c.p. lattices are also significantly increased and become comparable with the b.c.c. lattices. Among frequent RGMLs there are RGMLs with the topology related to the close packings and b.c.c. lattice. Let us note the RGML {10,34,74} corresponding to body-centered tetragonal packing and the derivative of b.c.c. or f.c.c. RGMLs, and the RGML {10,36,78} related to b.c.c. RGML, but, unlike body-centered tetragonal packing, also derived from h.c.p. RGML. The RGML {8,26,56} topologically corresponding to primitive hexagonal packing is also concerned with strongly distorted b.c.c. packing (Figs. 3*a* and *b*). Apparently, if weak intermolecular contacts are discounted, the total increase of contributions of the RGMLs considered is caused by the transformation of distorted packings with the topology of a b.c.c. lattice. Let us emphasize that all the aforementioned transformations of packing topology geometrically correspond to arbitrary small motions of molecules, because the lattice VDP faces that disappear due to such motions are small.

6. Discussion

6.1. Local topology of molecular packings

Comparison of the topological properties of the local environment of molecules in a packing characterized by their lattice or smoothed VDPs confirms the conclusion of Peresypkina & Blatov (2000) regarding the predominance of MCN = 14 and the correctness of describing molecular packings within the framework of the model of the thinnest covering of space by deformable spheres (Blatov & Serezhkin, 1997). Let us emphasize that Peresypkina & Blatov (2000) have considered only complete lattice VDPs, taking into account all, even very weak, contacts in a packing. At the same time, comparing data of Tables 3 and 4, one can conclude that the connectedness of molecules in a molecular net is resistant enough to influence the omission of weak intermolecular contacts (see also Peresypkina & Blatov, 2000), but such an influence on molecular domains in a packing gives rather different results. In this case each molecule spatially (without taking into account intermolecular contacts) tends to surround itself by 12 similar molecules (MPN = 12). The geometry of molecular packing is such that the centroids of two molecules out of 14, which gives a high frequency of the Fedorov cuboctahedron (Table 2), are further from the centroid of a given molecule than others, as a rule. At the same time the

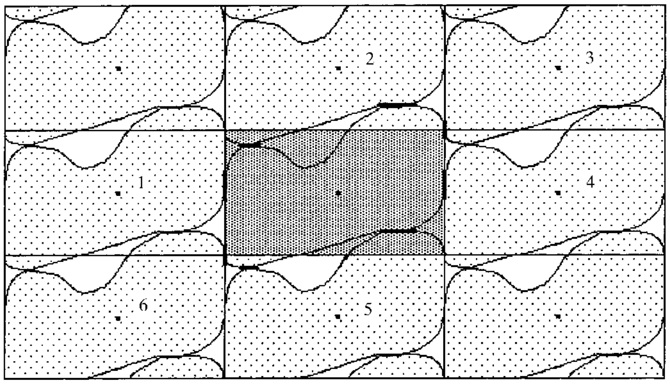


Figure 4
A fragment of hypothetical packing of nonspherical molecules on a plane, where MCN = 6 ≠ MPN = 4. The tangency points of the molecules (bold lines) correspond to the intermolecular contacts; molecular centroids are shown by points inside the molecules and the integers (1–6) enumerate the molecules adjacent to the central one. A lattice molecular VDP of the central molecule is shaded.

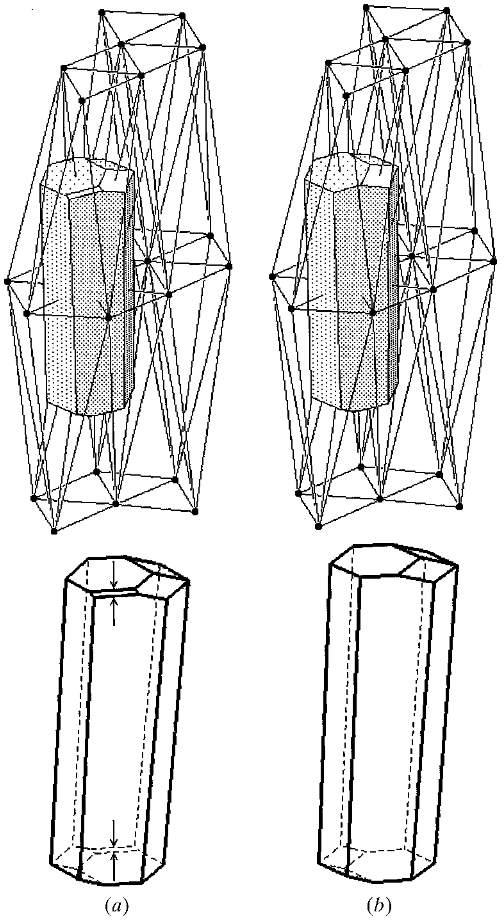


Figure 5
(*a*) A fragment of the graph of molecular net with the topology $\{N_{1-3}\}$ = {14,54,124}, where each node has 14 adjacent nodes (MCN = 14) and (*b*) the graph of molecular lattice with the topology of a f.c.c. lattice, possessing 12 nodes within the first coordination sphere (MPN = 12), for the crystal structure of (20*S*)-ethyl-3-β-(*tert*-butyldimethylsiloxy)-20-methylpregna-5,6-dien-21-oate (GIHLUE). (*a*) This graph is transformed into graph (*b*) by discounting two contacts which correspond to the small quadrangular faces of a smoothed molecular VDP (*a*), which are allocated to the opposite sides of six-coordinated layers. These contacts are not shown in the figure because they are outside the graph fragment. The operation of 'tightening' these faces is shown by arrows and corresponds to a transformation of a smoothed molecular VDP of the type 14/24-3 (*a*) into a lattice VDP of the type 12/18-1 (*b*).

contacts between these molecules are often strong enough so that the topology of smoothed VDPs and MCN = 14 are persistent. An apparent contradiction caused by the disagreement of MCN and MPN can be easily explained if one takes into account that most organic molecules are non-spherical, therefore, smoothed and lattice VDPs also differ from each other (Peresypkina & Blatov, 2000). For instance, in a hypothetical molecular packing on a plane (Fig. 4), a lattice molecular VDP in the form of a rectangle (MPN = 4) is realised, but MCN = 6. Thus, the smoothed molecular VDP has six faces, but in the lattice VDP there are no two small faces corresponding to the tangency points of exterior surfaces of contacting molecules, because their centroids are rather far apart. In real crystal structures such a situation is ordinary (Fig. 5). Let us emphasize that if a molecule has a spherical shape (for example, it consists of a single atom, as in the case of noble gases), MCN and MPN (and also smoothed and lattice VDPs) are equal.

The features of the topology of molecular crystal structures also result in large differences between the total numbers of lattice and smoothed VDP types. Owing to the variable chemical nature of the molecules studied, topologically similar packings are generally characterized by different systems of intermolecular contacts, even within the first coordination sphere. Thus, each of 322 lattice VDP types corresponds at least to several of 6381 smoothed VDP types. For instance, in the crystal structures of (20*S*)-ethyl-3-β-(*tert*-butyldimethylsiloxy)-20-methylpregna-5,6-dien-21-oate (GIHLUE; Figs. 5*a* and *b*) and androst-5-en-3,17-dione (ANDEDP10; Figs. 6*a* and *b*), with identical combinatorial properties of lattice VDPs (type 12/18-1), the topologies of smoothed VDPs differ from each other (types 14/24-3 and 14/24-6, respectively). However, the correspondence between lattice and smoothed VDPs is ambiguous because one type of smoothed VDP can also correspond to various types of lattice VDPs. For instance, in the crystal structures of ANDEDP10 and 1,5-anhydro-2,3,4-tri-*o*-benzoyl-ribitol (ATBRIB10), type 14/24-6 of the smoothed VDP (Fig. 5*a*) corresponds to the lattice VDPs 12/18-1 (Fig. 5*b*) and 14/24-4 (Table 2), respectively. However, let us emphasize that both molecular VDPs are used to describe the topology of the same molecular crystal, therefore, the differences in their combinatorial properties are usually insignificant. Thus, smoothed VDPs can be rather easily transformed into lattice VDPs constructed for the same molecule (Peresypkina & Blatov, 2000; Figs. 5*a* and *b*).

During the analysis of lattice VDPs and their combinatorial stability Peresypkina & Blatov (2000) concluded that the local topology of molecular packings tends to the thinnest covering of space to which the b.c.c. lattice corresponds (Blatov & Serezhkin, 1997). However, one should not assume that packing, especially distorted, could only be identified with the analysis of VDP combinatorial properties. Such an identification has to be only approximate because there is no information on the composition of long-range coordination spheres in this case. Both smoothed and lattice molecular VDPs depend on the geometrical characteristics of concrete molecular packing and their combinatorial properties can vary at metrical distortions of the packing, while the global topology of the crystal structure remains unchanged. There is also a contrary, but rare case when two nets or lattices have different topologies, but combinatorially identical molecular VDPs. In particular, the fact that the amount of topologically different RGMNs (1384 types) and RGMLs (262 types) is much less than number of types of smoothed (6381) and lattice VDPs (322), respectively, complies with these properties of molecular packings. Therefore, the realisation of a lattice or smoothed molecular VDP in the form of the Fedorov cuboctahedron, strictly speaking, allows a conclusion on the similarity of the environments of the molecules and of an atom in a b.c.c. lattice within the first coordination sphere, *i.e.* the similarity of their local, but not global, topologies. In particular, in the crystal structures of compounds with lattice VDPs in the form of the Fedorov cuboctahedron there are both $\{N_{1-3}\} = \{14,50,110\}$ and $\{N_{1-3}\} = \{14,52,116\}$ or $\{14,54,124\}$ lattices, *i.e.* lattices with the topologies of both b.c.c. lattices and also some other lattices. However, the cases are frequent when 'unusual' VDPs are realised for given RGMN or RGML. In particular, the molecular net $\{N_{1-3}\} = \{14,50,110\}$ in the crystal structure of 1,1-diphenyl-2,2-dimethyl-3(2,2-dimesitylvinil)cyclopropane (DOFGIO) is characterized by a smoothed molecular VDP with combinatorial properties of the 14/24-5 type (Table 2), but not the Fedorov cuboctahedron 14/24-1, typical for a b.c.c. lattice. In this case the analysis of VDP combinatorial properties gives inexact results, because it only shows the

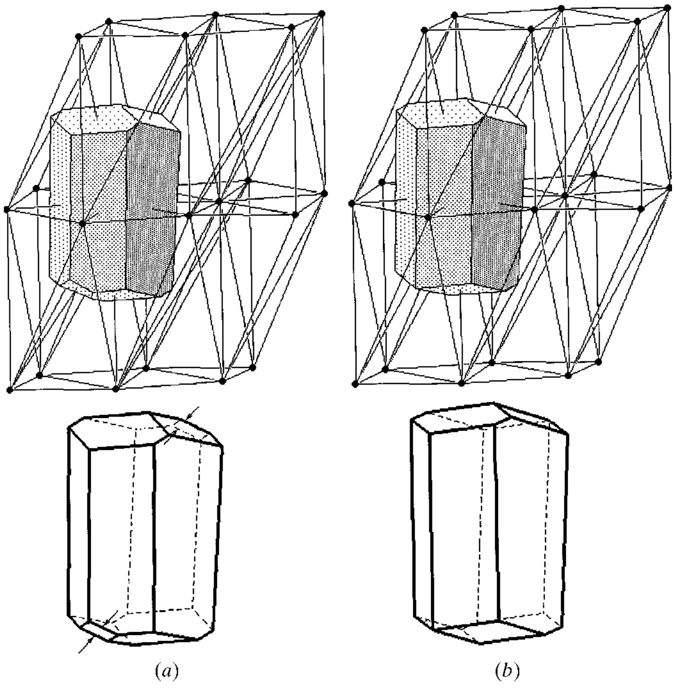

(*a*) (*b*)

Figure 6
(*a*) Fragment of the molecular net graph with b.c.c. lattice topology and (*b*) the molecular lattice graph with f.c.c. lattice topology for the crystal structure of androct-5-ene-3,17-dione (ANDEDP10) with corresponding molecular (*a*) smoothed and (*b*) lattice VDPs. Arrows indicate the direction of the transformation of a tetradecahedron of the type 14/24-6 (*a*) into a dodecahedron of the type 12/18-1 (*b*).

Table 5
The distribution of types of molecular lattices with the topology of the b.c.c. net for the sample with 7458 monosystem organic compounds, considering all the faces of the smoothed molecular VDPs.

See remarks to Table 4.

All faces of lattice molecular VDPs are considered			The lattice molecular VDP faces with Ω >1% are only considered		
N_{1-3}	Number of lattices	Relative part (%)	N_{1-3}	Number of lattices	Relative part (%)
14,50,110	5463	73.3	12,42,92	2391	32.1
16,58,128	443	6.0	14,50,110	2270	30.5
14,52,114	384	5.2	12,44,96	747	10.0
12,42,92	383	5.1	10,34,74	614	8.2
14,52,116	247	3.3	13,47,103	187	2.5
16,60,134	172	2.3	8,26,56	177	2.4
			11,39,85	155	2.1
			10,36,78	136	1.8
			13,47,104	98	1.3
Total number of types is 70			Total number of types is 92		

similarity of a given molecular net to a b.c.c. net, not their strict isomorphism. It is of interest that the RGML of this crystal structure has $\{N_{1-3}\}$ = {14,52,114}, *i.e.* in this case the topologies of packing and the net differ from each other, although MCN and MPN are identical. The ambiguity of the correlation between local and global topology can appear in different ways. Thus, in the crystal structures of 2-amino-3-nytropyridine (AMNTPY) and *syn*-1,6-imino-8,13-methano(14)annylene (BEJPEL) RGMN and RGML are characterized by smoothed and lattice VDPs, whose combinatorial properties coincide with those of molecular VDPs in the crystal structure of DOFGIO. In the crystal structure of AMNTPY RGML topology coincides with DOFGIO RGML topology, however, RGMN is characterized by another coordination sequence ($\{N_{1-3}\}$ = 14,54,124; Fig. 5*a*), while in the case of BEJPEL the RGMN and RGML topologies are the same ($\{N_{1-3}\}$ = 14,52,115) and do not correspond to any coordination sequence, typical for two other compounds.

The examples given confirm the aforementioned conclusion that the molecular packing (net) can be unambiguously topologically determined only if complete isomorphism of its RGML (RGMN) and a lattice with known topology can be found. In this study we have searched only for the essential isomorphism of molecular nets and lattices that, however, allows discussion on their complete topological similarity from the chemical point of view (Blatov, 2000). Owing to these reasons it is more valid to use coordination sequences to analyze and carry out a topological classification of molecular packings, whereas a comparison of combinatorial and geometrical VDP properties allows the degree of metrical distortions in a molecular packing to be discovered.

6.2. Global topology of molecular packings

The disagreement found between local topologies of molecular nets and lattices also indicates variations in their global topological properties, which is shown by comparing their coordination sequences (Tables 3 and 4). At the same time, some correlation between various RGMN and RGML types exists such as smoothed and lattice molecular VDPs. Thus, a large number of topologically different nets symbolizing different ways of binding molecules can conform to any molecular packing with a given RGML. The number of such nets depends on the abilities of real molecules to form intermolecular contacts, *i.e.* mainly on the shape of a molecule and on its chemical composition. In particular, the calculations show that for the sample of the most frequent molecular lattices with the topology of a b.c.c. lattice, the corresponding nets have the same topology in 43% of cases, which indicates slight distortions of the molecular packings. In the remaining 57% of cases RGMNs are subdivided into 606 topologically different types, among which the f.c.c. net (5.5%) and also the nets $\{N_{1-3}\}$ = {14,52,114} (4.8%) and $\{N_{1-3}\}$ = {14,54,124} (4.7%; Fig. 5*a*) are the most frequent. At the same time any net with a given topology (with a given set $\{N_k\}$) and different degrees of geometrical distortion (*i.e.* in various molecular crystal structures) can correspond to several lattices, and these lattices should be topologically similar, *i.e.* they should be easily transposed owing to distortions. In this sense the geometrical distortion of a packing can result topologically in breaking or, contrarily, intermolecular contacts appearing. The crystal structures of the compounds investigated can be divided into groups of topologically identical packings, which are metrically distorted differently owing to the variations in shape and the chemical nature of molecules. In rather large samples practically all the possible variants of distortion of a given packing and, therefore, all topologically similar packings can be realised. The frequency of each RGML type obtained at the metric distortion of a packing corresponds to the probability of this transformation, *i.e.* the degree of its similarity to this distorted variant. For instance, if one takes into account all the contacts of molecules in a packing (Ω > 0) then the b.c.c. RGMN conforms to 70 RGMLs to be topologically different, but similar to each other, among which the molecular packings with the topology of a b.c.c. lattice, and also the packings with $\{N_{1-3}\}$ = {16,58,128}, {14,52,114} and the f.c.c. lattice (Table 5), are frequent enough. Note, that the amount of nonisomorphic RGMNs corresponding to a single RGML type is larger, as a rule, than the amount of nonisomorphic RGMLs available for a given RGMN type. This fact is confirmed by a great difference in the number of topological types of RGMNs (1384) and RGMLs (262), and shows a much greater variety of ways of binding molecules in comparison to the number of different spatial arrangements.

It is of interest to determine the influence of weak intermolecular contacts on the topology of molecular nets and lattices. If they were omitted the main tendencies for a certain arrangement of short-range and long-range environments of a molecule could be found and the topological stability of a certain net (lattice) type could be statistically characterized. For the sample studied the omission of weak contacts of both RGMN and RGML considerably influences the graph topology (Tables 3 and 4), however, the greatest changes occur in the case of molecular lattices. These changes come to light in the strong decrease of the b.c.c. lattices in the molecular lattice distribution and in a significant increase of the close packings. Thus, the group of packings with b.c.c. lattice topology at all the intermolecular contacts considered gives 80 topologically different variants of its distortion at the expense of omitting weak contacts. Among them RGMLs with the topology of f.c.c. (32.4%), h.c.p. (10.7%) or tetragonal body centered (9.4%) lattices are the most frequent. Thus, owing to the geometrical distortion of a packing with the topology of the b.c.c. lattice, one of the close packings is realised with high probability. It was already mentioned that this tendency to change lattice topology indicates an inclination of the molecules to a spatial arrangement according to one of the close packings, preferably the f.c.c. packing. The packings derived from close packings (Table 4) are realised due to some steric restrictions preventing the appearance of f.c.c. or h.c.p. packings and are caused by the tendency of the local topology of a molecular environment to a b.c.c. motive.

The short-range environment of a molecule realising the thinnest covering of space and its long-range environment constructed according to a close packing do not exclude each other, because the situation is typical for a crystal structure where RGMN topology corresponds to a b.c.c. lattice and the RGML has the topology of one of the close packings. Thus, out of all the compounds with b.c.c., RGMN and RGML most frequently correspond to f.c.c. or h.c.p. lattices (Table 5; Figs. 6*a* and *b*) discounting weak intermolecular contacts in a packing ($\Omega > 1\%$). In our opinion, such combinations of b.c.c. RGMN and f.c.c. RGML in the same molecular packing are expected, because it is a compromise between the aforementioned antagonistic tendencies in molecular packing formation. This assumption is confirmed by an easy transformation of the b.c.c. net, appropriate for the local arrangement of a molecule, into the f.c.c. lattice, preferably to the global packing topology. Another confirmation is the aforementioned large number of strongly metrically distorted and topologically unstable b.c.c. lattices, in most cases capable of transforming into a f.c.c. lattice with the omission of weak contacts. Apparently, the real arrangement of molecules in a crystal is intermediate between these two lattices, representing either a distorted b.c.c. or distorted f.c.c. lattice, which is corroborated by extremely small numbers of regular Fedorov cuboctahedra and rhombododecahedra among lattice VDPs (Peresypkina & Blatov, 2000). The rarity of the h.c.p. lattice indirectly confirms this conclusion, since the coexistence of the b.c.c. net and a h.c.p. packing of molecules in a crystal structure seems to be less preferable because of the more essential structural reorganization needed in this case. Note that an abnormally high frequency of close packings compared with other types of molecular packings, revealed by studying RGML global topology, was not found by Peresypkina & Blatov (2000). The reasons include the extremely low frequency of cubic and hexagonal rhombododecahedra, typical for ideal close packings, among molecular lattice VDPs (18 of 23067 cases), and the topological similarity of the only frequent dodecahedron type 12/18–1 (Table 2) both to the Fedorov cuboctahedron and to the rhombododecahedron, typical of an f.c.c. lattice.

7. Conclusions

From these results the following main conclusion can be made: during the formation of the structure of molecular crystals the molecular nets prefer the b.c.c. lattice topology, while the molecular packings are topologically similar to close packings. These counteracting tendencies are caused by the different nature of various forces which are responsible for the molecular crystal formation. On the one hand, the first tendency reflects the striving of molecules for the most uniform environment by other molecules, and also for the maximum number of intermolecular contacts to obtain as high an MCN as possible. Obviously this rule is governed by short-range van der Waals interactions, which are capable of influencing the arrangement of molecules only within the first coordination sphere. On the contrary, the second tendency is connected with long-range factors which provide a periodic spatial arrangement of molecules, in particular, with acoustic lattice vibrations and dispersing the long-wave phonons in a crystal. At the same time, a concrete mechanism resulting in the predominance of close and topologically similar to close molecular packings remains unclear. Although this result formally agrees with the conclusions of Kitaigorodskii (1973), it is impossible to consequently use his model for most of the compounds, because it predicts a predominance of MCN = 12. According to the calculations performed each molecule tends to surround itself by not 12, but 14 molecules, which is contrary to Kitaigorodskii's model, therefore, the model of deformable spheres (Blatov & Serezhkin, 1997) should be used. Let us emphasize that Kitaigorodskii's model assumes that a packing as a whole, including its global topology, is determined by the arrangement of adjacent molecules according to the principle 'a bulge to a hollow', *i.e.* by the local topology. However, detailed analysis of the results obtained shows that there is no unambiguous correlation between the local and global topology of the structure of a molecular crystal. Molecular packing is a result of a compromise between the factors indicated and usually is rather geometrically distorted. At the same time even a strong local deformation stipulated by a complex shape of molecules and by intermolecular interactions of different intensity is unable to noticeably change the global topology of a crystal structure, which keeps its general pattern even if the compounds are chemically and stereometrically different.

The authors are indebted to the RFFI for covering the License fee for the use of the Cambridge Structural Database (project 99–07-90133).

References

Blatov, V. A. (2000). *Acta Cryst.* A**56**, 178–188.

Blatov, V. A. & Serezhkin, V. N. (1997). *Acta Cryst.* A**53**, 144–160.

Blatov, V. A., Shevchenko, A. P. & Serezhkin, V. N. (1999). *J. Appl. Cryst.* **32**, 377.

Brunner, G. O. & Laves, F. (1971). *Wiss. Z. Tech. Univ. Dresden*, **20**, 387–390.

Cambridge Structural Database (1999). Version 5.18. Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, England.

Conway, J. H. & Sloane, N. J. A. (1997). *Proc. R. Soc. London*, **453**, 2369–2389.

Grosse-Kunstleve, R. W., Brunner, G. O. & Sloane, N. J. A. (1996). *Acta Cryst.* A**52**, 879–889.

Kitaigorodskii, A. I. (1973). *Molecular Crystals and Molecules.* New York: Academic Press.

O'Keeffe, M. (1995). *Z. Kristallogr.* **210**, 905–908.

Peresypkina, E. V. & Blatov, V. A. (1999). *Theoret. Chem.* **489**, 225–236.

Peresypkina, E. V. & Blatov, V. A. (2000). *Acta Cryst.* B**56**, 501–511.

Serezhkin, V. N., Mikhailov, Yu. N. & Buslaev, Yu. A. (1997). *Russ. J. Inorg. Chem.* **42**, 1871–1910.

Stevens, E. D. (1979). *Mol. Phys.* **37**, 27–45.

Zefirov, Yu. V. (1997). *Crystallogr. Rep.* **42**, 111–116.