Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Packing and molecular shape: layers of centred hexagons as a guiding principle

Nils Braun and Gottfried Huttner*

Anorganisch-Chemisches Institut der Universität Heidelberg, Im Neuenheimer Feld 270, 69120 Heidelberg, Germany

Correspondence e-mail: g.huttner@aci.uni-heidelberg.de

Received 29 August 04
Accepted 5 January 05

In memory of D. Sellmann and L. Zsolnai

A set of 112 993 structures was analysed for a specific type of packing: Molecular structures for which the centres of molecules form layers of centred hexagons exceed a percentage of 30%. This result has been obtained by a newly developed algorithm, allowing the extraction of these structures automatically and showing correlations between molecular shape and the type of packing.

1. Introduction

A wealth of structural data for molecular compounds has been described by chemists with their main interest focused at the structure, *e.g.* the connectivity and geometry of individual molecules. All relevant information is stored in the Cambridge Structural Database (CSD; Allen, 2002), which is now routinely used by chemists to extract information on molecular structures. Only rarely is it used to extract information on the type of packing of the molecular items within the solid state.

With interest in molecular chemistry steadily growing, chemists are increasingly aware of the importance of the type of arrangement of the molecular building blocks within a crystal. The logical principles for analysing the type of arrangement of molecules within a crystal were laid down decades ago by Kitaigorodskii (1945). His fundamental ideas are still the basis of any attempt to rationalize the patterns of molecular packing.

There are two lines of research originating from his seminal work: one relies upon modelling the patterns by force-field methods (Kitaigorodskii, 1970; Pertsin & Kitaigorodskii, 1987; Lommerse *et al.*, 2000; Motherwell *et al.*, 2002). The other one tries to make intelligent use of data mining procedures with the aim of extracting some of the rules which govern packing and which are hidden in the data (Kuleshova & Antipin, 1999; Motherwell, 1999, 2001; Brock, 1996). This approach had already been used by Kitaigorodskii (1961), who, to give just one example, inspired by the fact that so many crystals belong to the space group $P2_1/c$, was able to correlate space-group symmetry with characteristics of the shape of real molecules (Kitaigorodskii, 1979). Henceforth, crystal symmetry is a commonly used tool in the analysis of molecular packing (Brock & Dunitz, 1994; Cole *et al.*, 2001; Yao *et al.*, 2002; Pidcock *et al.*, 2003). An overview over the systematic study of crystal packing is given by Brock (1999).

A different approach involves analysing the polyhedra (Wells, 1983) formed by the centres of molecules within a crystal. A search for specific polyhedra, *e.g.* octahedra, acts as a filter to extract patterns with that specific polyhedral arrangement of molecular structures. The connectivity between individual polyhedra may then by analysed for the specific idealized type of pattern, *e.g.* cubic close-packed

(f.c.c.), hexagonal close-packed (h.c.p.) or body-centred cubic (b.c.c.); see Reichling & Huttner (2000). An attempt can then be made to correlate the observed specific idealized patterns of packing with the shape of the molecules. A very reduced and approximate description of the shape of a molecule is its ellipsoid of second moments (ellipsoid of moments of inertia). Not unexpectedly, using this shape descriptor as a classifier, it is found that molecules, the shapes of which are described by ellipsoids with rotation symmetry, tend to be packed in one of the idealized close-packing arrangements (Reichling & Huttner, 2000).

The basic idea of this approach is at the foundation of the present work. However, while the latter work made use of pattern analysis by neural networks, the present work concentrates on the search for layers, built from centred hexagons. Structures, the packing patterns of which may be described by parallel stacking of such layers, are extracted by this type of filter. The type of sequence of such layers relative to the normal of the layer planes serves as an additional classifier. A correlation between packing and shape is obvious for some shapes.

An algorithm has been developed which automatically screens crystal structures with respect to packing. A program for performing this analysis and allowing visualization of the results is available for download (Braun & Huttner, 2004). The present article describes the application of the algorithm to approximately 110 000 structures retrieved from the CSD.

2. Methodology

2.1. Dataset

The data were taken from the CSD Version 5.24 with supplements through July 2003. From a total of 296 427 entries a subset of 112 993 served as the basis of the analysis. This subset resulted after discarding structures with the following features:

(i) a conventional R factor greater than 0.10;
(ii) no three-dimensional coordinates;
(iii) disorder;
(iv) polymeric;
(v) inconsistent information;
(vi) more than one residue per asymmetric unit.[1]

The last criterion automatically excludes ionic structures.

The remaining dataset consisted of 113 038 structures. From this total, a further 45 structures had to be excluded since the quantum chemical program used for calculating the molecular volume (see §2.7) could not automatically deal with some of the atoms present in the structure. The analysis is hence based on 112 993 structures.

2.2. Search for centred hexagons

The packing of the centres of the molecules was chosen as the basis of the packing analysis. The centre was calculated as the centre of geometry of all the atoms within the molecular structure (Motherwell, 1997). Starting from one selected molecule within the cell, centres of up to 24 nearest neighbours inside a sphere were determined.

Table 1
Upper limits for the search of hexagons.

Criterion	λ	σ_d	σ_γ	σ_φ	$d_{\min}$	ω
Limit	0.15	0.075	0.15	0.15	0.15	1

All possible combinations of seven points of this ensemble were checked to determine if they form a hexagon that is centered, reasonably regular and approximately planar. The check included the following criteria (see Fig. 1):

(i) coplanarity as determined from the eigenvalues λ of the matrix of second moments;

(ii) the relative standard deviation σ_d of the 12 distances, *i.e.* the six radial (*a*, Fig. 1) and the six tangential (*b*, Fig. 1) distances;

(iii) the relative standard deviation σ_φ of the six central angles from the ideal value of 60°;

(iv) the relative standard deviation σ_γ of the six inner angles between adjacent edges from the ideal value of 120°;

(v) the distance $d_{\min}$ of the central point from the geometrical centre of the seven-point arrangement, normalized by the next shortest distance of a point to the geometrical centre (this kind of normalization is more stringent than normalization by the average distance).

The limits given in Table 1 were found to be appropriate.

As an additional restriction, a combination of all five of the above-named criteria in the form of a weighted sum, $\omega = 4\lambda + \sigma_d + \sigma_\varphi + 4\sigma_\gamma + d_{\min}$, was used. This weighting scheme puts additional weight on planarity (λ). It excludes cases with, for instance, just one edge of a planar centred hexagon being displaced from the position it would occupy in a regular hexagon. In this case σ_d might still be small, while σ_γ reflects this irregularity.

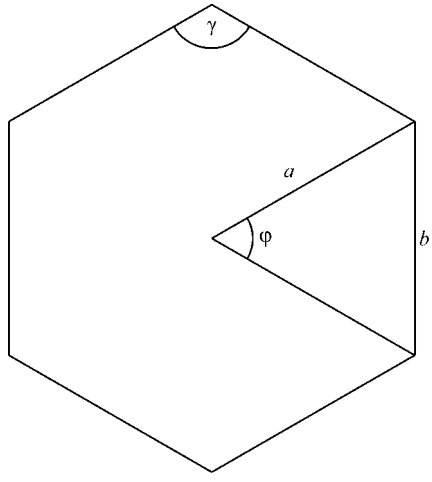

Figure 1
Illustration and definition of some of the symbols pertaining to the data analysed in the search for centred hexagons.

[1] The term 'residue' is defined as a set of atoms where each atom within the set may be reached from any other atom within the set by walking along chemical bonds.

2.3. Search for layers

Once a centred hexagon is found, the next step is to show that this hexagon is part of a planar layer. Let us define q as the average of the six radial (a, Fig. 1) and the six tangential (b, Fig. 1) distances. Then, within a radius of $4q$ around the centre of the selected hexagon, all centres which are coplanar with it are chosen. Coplanarity with the central hexagon is checked by calculating the deviation in the direction of the normal of the plane. The normal itself is recalculated for the growing ensemble of points after the addition of every ten newly added points. The maximum distance allowed for the deviation of an individual point from the plane is taken as 30% of the shortest distance of the central point to any point not belonging to the layer (*i.e.* to the closest neighbour of an adjacent layer; the 30% criterion was found to be adequate as it guarantees an appropriate tolerance, especially for short interlayer distances). Points above and below the selected layer were sorted in their relevant layers so as to cover at least a total of five parallel layers and a mean of 20 over the whole dataset. The number of parallel layers covered depends on the ratio of the interlayer distance to the mean distance within the layer (q, see above).

2.4. Test for hexagonal tiling of layers

Within the central and the two adjacent layers, a test for an extended hexagonal pattern was performed. The 12 points around the selected central point as shown in Fig. 2 were chosen as the centres of the hexagons. An analogous procedure was applied to the adjacent layers above and below the basal hexagon checking 13 hexagons in each layer. If all the hexagons were found to be congruent, the pattern was ranked as one built of layers of centred hexagons.

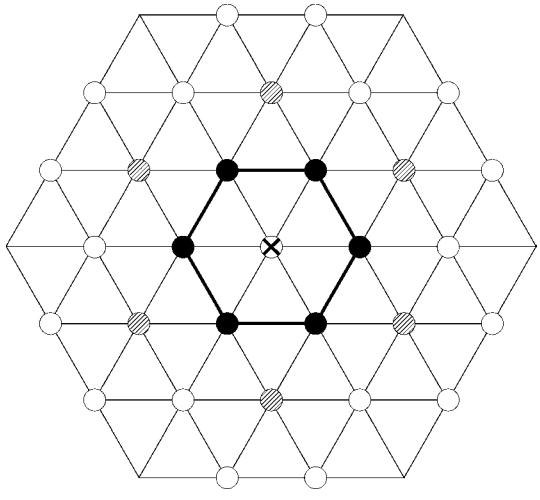

Figure 2
Illustration of the section of a layer of centred hexagons as it was used in a preselection algorithm (see text). Central point: cross; basic hexagon: filled black; centres of additional hexagons: shaded.

2.5. Stacking sequence

In close-packed structures, the sequence of close-packed hexagonal layers, *e.g.* *AB* or *ABC*, is an important discriminator. In general, if the structure is considered to be built from layers, the pattern of shifts of these layers determines the kind of repetitive pattern in the dimension perpendicular to the plane. In order to quantify this shift, the centre of the central hexagon in the basal layer is taken as the origin of an axis perpendicular to the layer plane. The displacements of consecutive layers are measured relative to this axis.

To this end, the three centres which are closest to this axis in the adjacent layer are determined (Fig. 3). Since the centred hexagons are not necessarily fully symmetric, one of the three centres just determined has to be selected as the most appropriate base to measure the shift. This selection is made using the following criteria:

(i) The six-membered cycle around a chosen centre should be related to the basic six-membered cycle by translation alone.

(ii) If this condition is not met by any of the three six-membered cycles defined by the three centres chosen, the one which is related to the basic six-membered cycle by translation and a rotation by 180° around the plane normal should be selected.

(iii) If neither condition (i) nor condition (ii) are met, no stacking analysis should be performed.

If more than one of the three centres meets condition (i) or (ii), a situation schematically shown in Fig. 3 may result. Of the possible alternatives, the one chosen is that for which the projection of its distance vector $\mathbf{r}_0$ onto the basal plane is closest to being perpendicular to an edge of the basal hexagon. In order to standardize the individual deviations from this

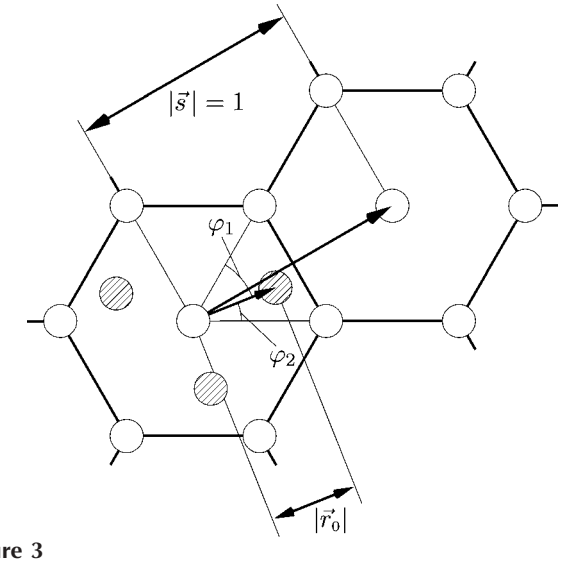

Figure 3
Illustration of the definition of variables and norms used in categorizing the shift of layers. Open circles: centres in layer 0 (basal layer); shaded circles: centres in adjacent layer (layer 1); $|\mathbf{r}_0|$: shift of adjacent layer *versus* basal layer; φ_1, φ_2: angles defining the direction of the shift vector $\mathbf{r}_0$; $|\mathbf{s}|$: distance between the centres of two adjacent hexagons in the basal layer.

ideal pattern (distance vector $\mathbf{r}_0$ pointing to the centre of the edge, see Fig. 3), the deviation ζ is defined as

$$\zeta = \left|\frac{\frac{\gamma}{2} - \varphi_1}{\frac{\gamma}{2}}\right| = \left|\frac{\frac{\gamma}{2} - \varphi_2}{\frac{\gamma}{2}}\right| \quad \text{with} \quad \gamma = \varphi_1 + \varphi_2.$$

With this definition, the deviation is scaled to $0 \leq \zeta \leq 1$, since if $\varphi_1 = \varphi_2$ then $\zeta = 0$ and if $\varphi_1 = 60^\circ$, $\varphi_2 = 0^\circ$ then $\zeta = 1$.

In order to scale the length of the shift, the distance $|\mathbf{s}|$ between two relevant neighbouring centred six-membered cycles (Fig. 3) is taken as the unit distance. This means *e.g.* that for an ideal hexagonal or cubic close-packed structure (h.c.p. or f.c.c.) the shift amounts to $|\mathbf{r}_0| = \frac{1}{3}$, while for an ideal b.c.c.-type packing the shift amounts to $|\mathbf{r}_0| = \frac{1}{2}$. For the specific situation of an *AAA* stacking sequence, *i.e.* a situation in which the centres of the hexagons of adjacent layers lay exactly on top of each other with respect to the plane normal (*i.e.* $|\mathbf{r}_0| = 0$), ζ is set to 0.

The stacking pattern of layers of centred hexagons was analysed in order to find all structures with equal or alternating inter-layer distances. This subset was analysed for the length of the shift vector $|\mathbf{r}_n|$ being an integer multiple of $|\mathbf{r}_0|$ (Fig. 3), the shift vector determined for the shift of the basal layer *versus* the adjacent layer (see above; n counts the layers: $n = 0$: basal layer; $n > 0$: layers above; $n < 0$: layers below).

Two categories were found according to

$$\mathbf{r}_n = (-1)^{|n|} \cdot \mathbf{r}_0$$

(alternating direction of shifts, stacking sequence *AB*) and

$$\mathbf{r}_n = \begin{cases} (n+1) \cdot \mathbf{r}_0 : & n \geq 0 \\ n \cdot \mathbf{r}_0 : & n < 0 \end{cases}$$

(constant direction of shifts, stacking sequences *AA*, *AB*, *ABC*, *ABCD*,...).

The first category comprises structures with a packing pattern characteristic of hexagonal close packing. The second category contains f.c.c.-type packings as a subset ($|\mathbf{r}_0| = \frac{1}{3}$). B.c.c.-type packings are sorted into both categories at the same time: the repetitive pattern is *AB*, taking into account that a shift of $|\mathbf{r}| = 1, 2, \ldots$, reached each second layer, is equivalent to a shift of $|\mathbf{r}| = 0$.

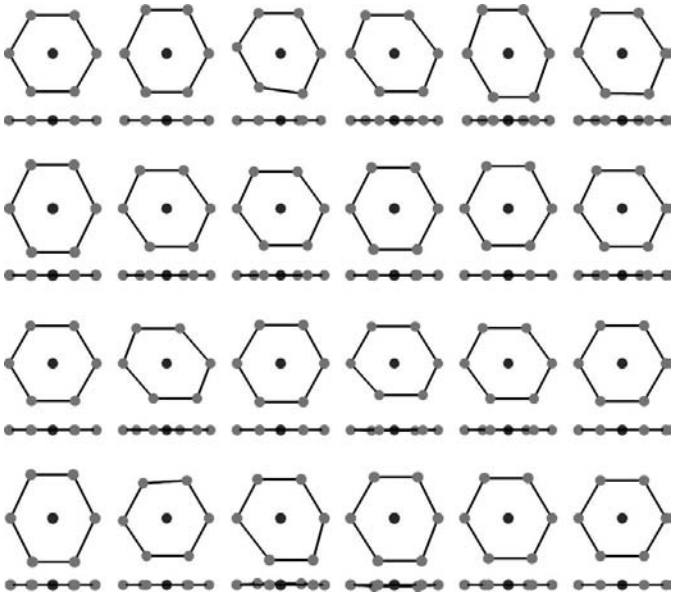

Figure 4
Graphic representation of a random selection of centred hexagons as present in the 36 678 structures which were categorized as being made up from layers of centred hexagons. The graphs are scaled to approximately equal size in order to make them comparable. Each hexagon is shown in two projections, one onto its plane (top) and one perpendicular to it (bottom). It is evident that planarity (each second line in the graph) was used as a stringent criterion in the filter as applied.

2.6. Molecular shape descriptor

It is the ultimate goal of a packing analysis to find rules which allow the correlation of the observed packing pattern to molecular properties. One of the most relevant properties of a molecule is its shape. For an analysis of the influence of molecular shape on the kind of packing of molecules within a crystal, it is necessary to describe the characteristics of the shape in a general way. A very approximate way to do so is to use the ellipsoid of second moments as a descriptor. Two numbers referring to the two aspect ratios m/l and s/m (with l, m and s corresponding to the roots of the eigenvalues in descending order, that is the lengths of the eigenaxes) suffice to characterize the shape this way. This kind of shape descriptor is used throughout this article.

In order to visualize properties of the dataset with respect to molecular shape, the frequency of occurrence of this property with respect to molecular shape has to be analysed. To smooth stochastic variances, the frequency was calculated for a unit area around each point in an m/l *versus* s/m diagram. The unit area was chosen as a circle of radius t around the point considered and the hits within this area were weighted according to their distance d from this point. The following truncated parabolic weight function $f(d)$ was chosen

$$f(d) = \begin{cases} 1 - (\frac{1}{t}d)^2 : & d < t \\ 0 : & d \geq t \end{cases}.$$

For the plots in this paper, t was chosen as 0.02.

2.7. Density of packing

To calculate the volume of individual molecules, the algorithm implemented in the quantum chemical program *GAUSSIAN* (Frisch *et al.*, 2003) was used. Volumes were calculated based on SCF calculations (the standard basis sets sto-3g or if not available for an atom within the molecule, the basis set Lanl2DZ was applied) using the c.p.c.m. model (Barone & Cossi, 1998) with krypton as the 'solvent' in the solvent-excluding surface technique. Bondi's atomic radii (Bondi, 1964) as implemented in *GAUSSIAN* were used.[2] Molecules with an odd count of electrons were calculated in the doublet state.

[2] The standard notation for the volume calculation with *GAUSSIAN* hence implies: "\#P RHF/sto-3g scrf = (solvent = krypton,cpcm,read)" with "surface = ses" and "radii = bondi".

3. Results

3.1. Layers of centred hexagons in molecular packing

The most surprising result of the analysis of the dataset is that of a total of 112 993 structures, a set comprising 36 678 structures [32.5% of the total, subset named *HL* (hexagon layers) hereafter] may be described as composed of parallel layers made up from planar centred hexagons. This result is surprising in so far as the filter used to extract layers of centred hexagons is quite terse. Fig. 4 shows a random selection of seven point arrangements which were ranked by the program as planar centred hexagons using the criteria given in §2.2. It is seen at the same time that planarity was used as an especially important criterion.

If the subset of structures made up by stacking parallel layers with each layer made up by centred hexagons is analysed for space-group distribution, it is evident that compounds crystallizing in some of the less common space groups (right hand side of Fig. 5) have a high tendency to form layers of centred hexagons (up to 80% in $P1$, $C2/m$, *Cmca*, yellow bars, scale on the right). Amongst the more common space groups, $P\bar{1}$ and $C2/c$ present a chance of around 50%. In the most common space group $P2_1/c$, almost 30% of the structures fit to the selected pattern (see Fig. 5).

It is worth while to note that the rather common space group $P2_12_12_1$, for which Kitaigorodskii (1979) had derived a high probability of close packing by analysis of space-group symmetry, has a rather low probability of below 5% to form such layers. The absolute number of structures conforming to the selection pattern has of course to account for the absolute number of structures per space group (Fig. 5, blue bars, scale on the left).

For the subset of structures which may be described as composed of layers of centred hexagons with equal or alternating inter-layer distances and shift vectors (19 751 structures; this subset, named *HLsub* hereafter, comprises 17.5% of the complete dataset and 53.8% of the dataset *HL* of layered structures; see §2.5), it is found that the direction of the shift (see Fig. 3) is most probably towards the centre of the adjacent hexagon ($\zeta = 0$, Fig. 3).

A graph illustrating the distribution of the lengths and directions of the shift vectors of subset *HLsub* is given in Fig. 6. The figure shows this distribution in a diagram with $r \cdot \sin\varphi$ and $r \cdot \cos\varphi$ as the horizontal and perpendicular axes. φ itself is calculated from ζ (see §2.5) by linearly scaling $\zeta[0; 1]$ to $\varphi[0^\circ; 30^\circ]$. The diagram is necessarily symmetric about $y = 0$ since a stack of layers may be seen as well from the top as from the bottom with an associated change of the sign of φ. The frequency with which the values of r_0 and φ occur was eval-

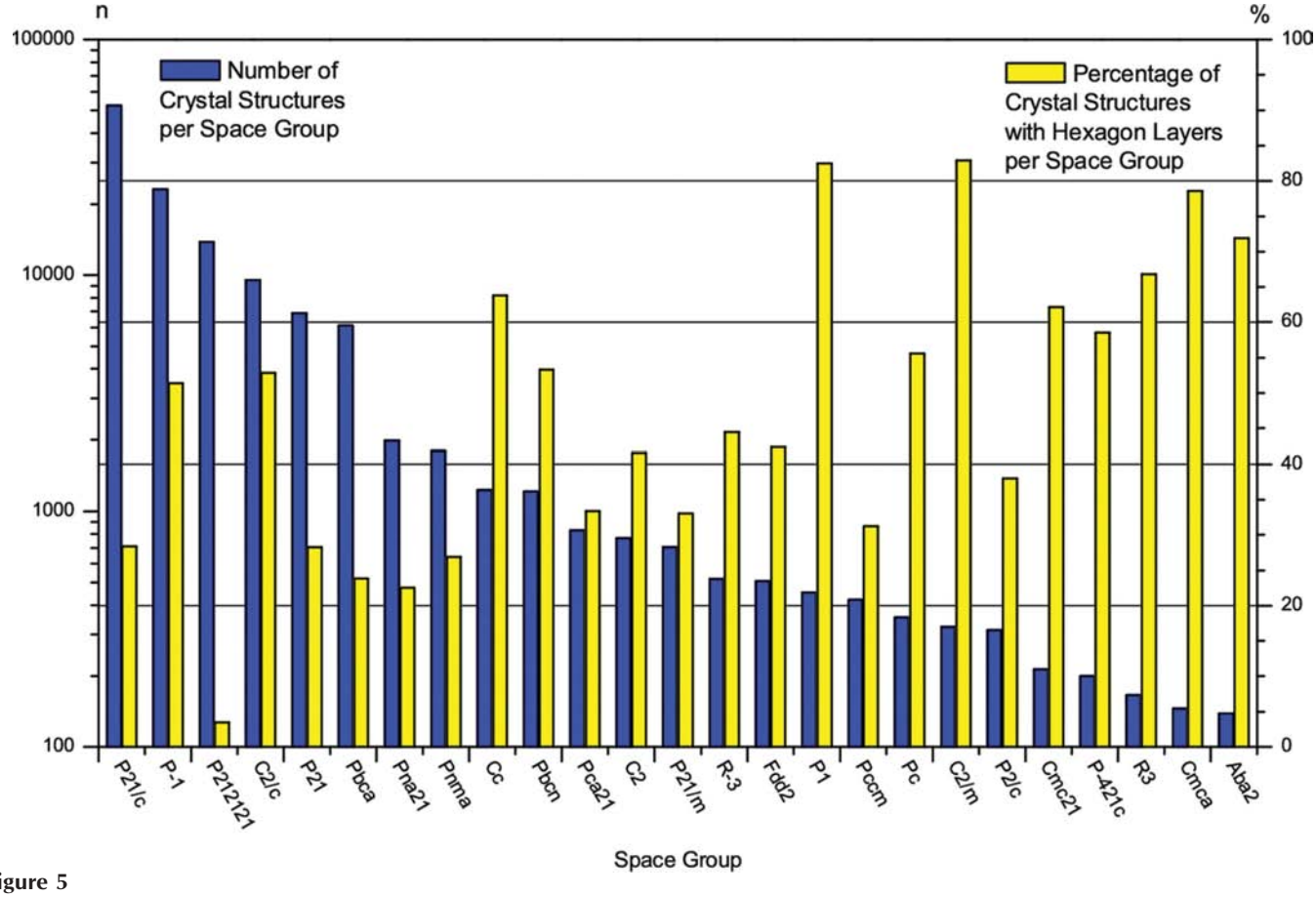

Figure 5
Histographic illustration of the frequency of the occurrence of the 25 most frequent space groups (blue bars, scale on the left). Percentage of structures made up from layers of centred hexagons (*HL*) for each of the 25 space groups (yellow bars, scale on the right).

uated within a rectangular net in an *r versus* φ diagram (100 steps for $0 \leq r \leq 0.52$ and 45 steps for $0 \leq \zeta \leq 1$, see §2.5 and Fig. 3 for the definition of r, φ and ζ). These values are presented in a colour-coded form in Fig. 6 at the corresponding location of $x = r \cdot \cos\varphi$ and $y = r \cdot \sin\varphi$.

An important feature of Fig. 6 is that it shows that a shift direction of $\zeta = 0$ is by far preferred. For this direction of shift, there is a high probability that the length of the shift of $|\mathbf{r}_0|$ is $\frac{1}{2}$, which places the centre of the hexagon of the adjacent layer just above the centre of the edge of the basic hexagon (see Fig. 3). There is also a high probability for a shift of $|\mathbf{r}_0| = \frac{1}{3}$ (see Fig. 6) which, in the ideal case, places the centre of a hexagon of the adjacent layer just on top of the midpoint of the equilateral triangle of the basic hexagon. The distribution of the lengths of the shifts at around $y = 0$ ($\zeta \leq 0.1$) is more easily visualized in Fig. 7, which shows the length of the shift vector *versus* the number of structures in the sector $\zeta \leq 0.1$ of Fig. 6. It is apparent that $|\mathbf{r}_0| = \frac{1}{2}$ is greatly preferred, followed by $|\mathbf{r}_0| = \frac{1}{3}$. The diagram also tends to indicate that shift vectors of lengths 0 and $\frac{1}{4}$ are relatively probable as well. A shift vector of $|\mathbf{r}_0| = \frac{1}{2}$ corresponds to an *AB* sequence of layers, as present in b.c.c.-type lattices.

A shift vector of $|\mathbf{r}_0| = \frac{1}{3}$ is characteristic of the *AB* sequence of a h.c.p.-type lattice as well as for the *ABC* sequence of an

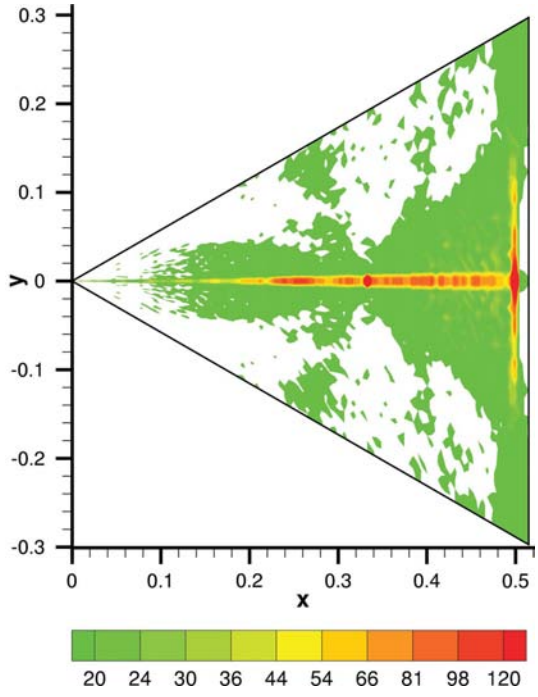

Figure 6
Distribution of shift vectors with respect to their orientation and length (φ and r_0, see Fig. 3). x and y refer to a polar coordinate system with $x = r \cdot \sin\varphi$ and $y = r \cdot \cos\varphi$. Frequency sampling was carried out for a grating with 100 steps for r and 45 steps for $\varphi = 30°$. The number of structures found in each interval area is represented by a colour code (logarithmic scale). For less than two structures per area, the colour was set to white.

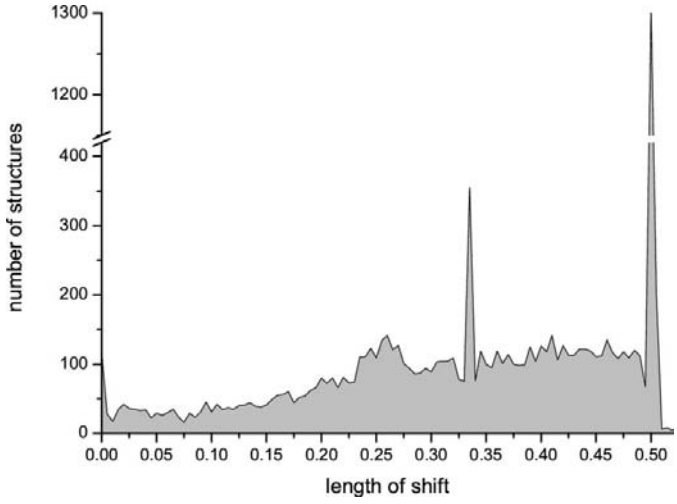

Figure 7
Histographic projection of the sector of Fig. 6 along $y = 0$ ($\zeta \leq 0.1$). The predominance of shift lengths around 0, $\frac{1}{3}$ and $\frac{1}{2}$ is evident. (NB The perpendicular scale is discontinuous.)

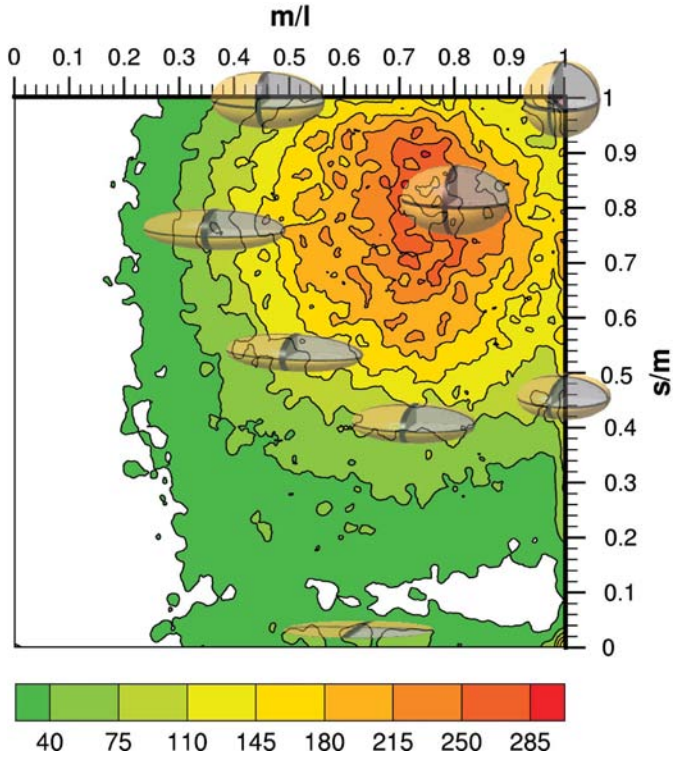

Figure 8
Distribution of molecular shapes for the basis data set (112 993 structures). m/l, s/m: aspect ratios of shape ellipsoid. The frequency of molecules with a specific shape ellipsoid was analysed in a grating of 1000 steps in each direction. The number of structures per point was calulated as a weighted sum of hits in a circular area around each point (see §2.6). The frequency is shown by the colour code, as shown at the bottom of the diagram. For frequencies of less than 10 (unweighted number of hits) the colour was set to white. Iso-frequency lines are shown at the border of individual areas. As an overlay, the shape of the ellipsoids, which characterize each point in the diagram, are shown for selected aspect ratios.

f.c.c.-type lattice. A shift vector of $|\mathbf{r}_0| = \frac{1}{4}$ indicates a stacking sequence of *ABCD*, while a shift vector of 0 represents the sequence *AAA* with the layers directly on top of each other.

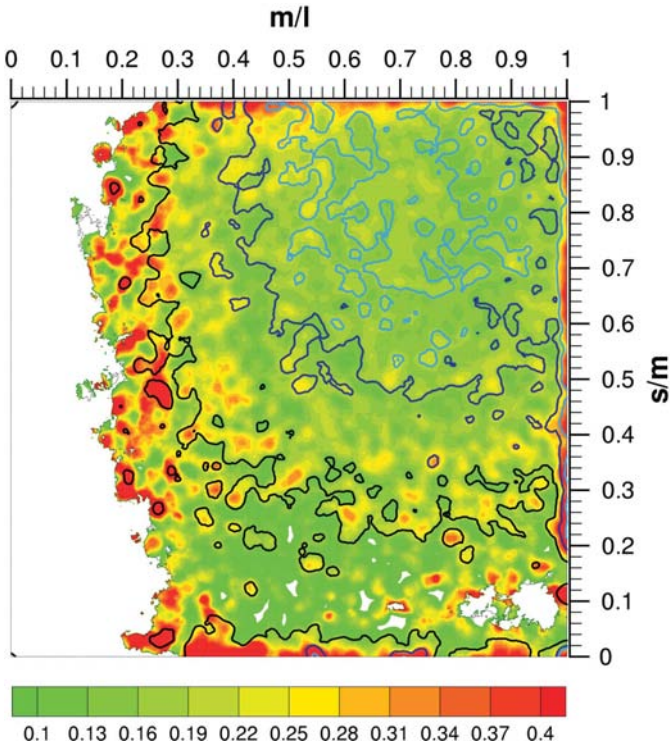

Figure 9
Frequency of structures of dataset *HLsub* with respect to the shapes of their ellipsoids of second moments. The frequency (colour code at the bottom of the diagram) is given as the percentage of structures showing this type of packing with respect to the total number of structures in the data set in each area. Iso-frequency lines are shown for 5 (black), 20 (dark blue) and 35 (light blue) weighted hits per unit area (see §2.6). The colour for areas with less than 5 unweighted hits was set to white.

3.2. Distribution of molecular shapes

The dataset of 112 993 structures, *i.e.* all structures with just one residue per asymmetric unit (*i.e.* just one type of molecule in the crystal, see §2.1), was analysed with respect to the distribution of molecular shapes. The shapes themselves were roughly approximated by ellipsoids of second moments (ellipsoids of moments of inertia with equal weights assigned to all atoms, including H atoms if present in the list of coordinates; see §2.6). The distribution is shown in a diagram scaled by the two aspect ratios of the ellipsoids (Fig. 8). To illustrate how these aspect ratios relate to the shape of the ellipsoids, these shapes are overlaid in a semi-transparent manner on the basic diagram at their respective aspect ratios. The diagram itself shows that the compounds that chemists have prepared and which have been structurally characterized tend to have a preferred shape around $m/l = s/m \simeq 0.8$. The frequency of shapes decreases in an almost radially symmetric way with increasing distance from this preferred aspect ratio. No definitive explanation for this finding can be given. It might be due to the shape properties of the families of compounds in which chemists have so far been interested; it might also mirror the differing propensity of different shapes to allow the formation of high quality crystals.

It is worth noting that a numerical simulation as well as experimental modelling (using candies of ellipsoidal shape as models) have shown that ellipsoids with both aspect ratios around 0.8 allow for a closer packing than all the other ellipsoidal shapes studied (Donev *et al.*, 2004).

For the subset *HLsub*, the iso-frequency lines show a grossly similar distribution of molecular shapes. In the corresponding diagram (Fig. 9), the iso-frequency lines pertaining to this subset of data are overlaid on a colour-coded graph. The colours in this graph indicate the percentage of structures which show the selected pattern of packing relative to the total number of structures at each point. Red colours correspond to a high percentage. It is obvious that the red-coloured areas concentrate in four regions. There are small/narrow red bands along the $m/l \simeq 1$ and $s/m \simeq 1$ axes, respectively, of the diagram. These locations correspond to shape ellipsoids with rotational symmetry. A concentration of structures with the basic packing patterns f.c.c., h.c.p. and b.c.c., which necessarily contain layers of centred hexagons, in these regions of the diagram has already been documented for a small subset of the CSD. The rotational symmetry of the shape ellipsoids has been invoked as the most

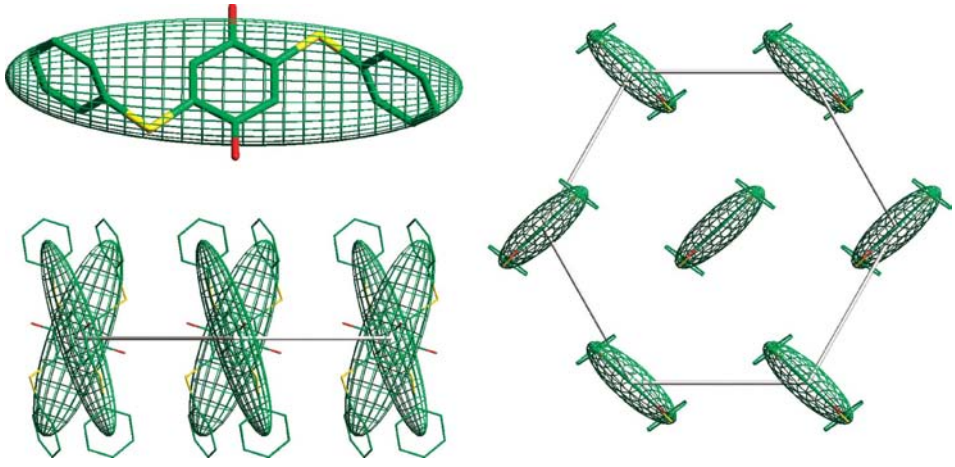

Figure 10
Packing of rod-type shape molecules. *SAFPIY01* (s/m = 0.596, m/l = 0.295) as an example. The centres of the rods form layers of centred hexagons. The rods are tilted relative to the plane normal.

probable explanation of this finding (Reichling & Huttner, 2000). The present analysis shows that the findings made on a small subset are also characteristic of a complete subset of the *CSD*, containing all structures with one 'residue'.

Another region with an accumulation of red areas is apparent in the left-hand side of the diagram around $m/l \simeq 0.25$ over the whole range of s/m. With m/l in this range, the molecules are rather elongated. The fact that the centres of these molecules somehow prefer an arrangement in centred hexagonal layers should be correlated with the rod-type shape of these molecules. The fact that centred hexagonal layers are formed independent of the aspect ratios s/m is explained by the orientation of the molecules relative to each other: ellipsoids of adjacent lines of the layer are rotated relative to each other so as to produce a herring-bone appearance of the projection (see Fig. 10).

The fourth region with a high probability of finding structures conforming to the search pattern is located at the bottom of the diagram at $s/m \simeq 0$ over the whole range of m/l. The molecules are hence close to planar and chemically they are preferentially aromatic molecules or square-planar complexes. For these flat molecules, the formation of layers of centred hexagons is generally achieved in the following way: the molecular planes tend to be parallel to the normal of the layer

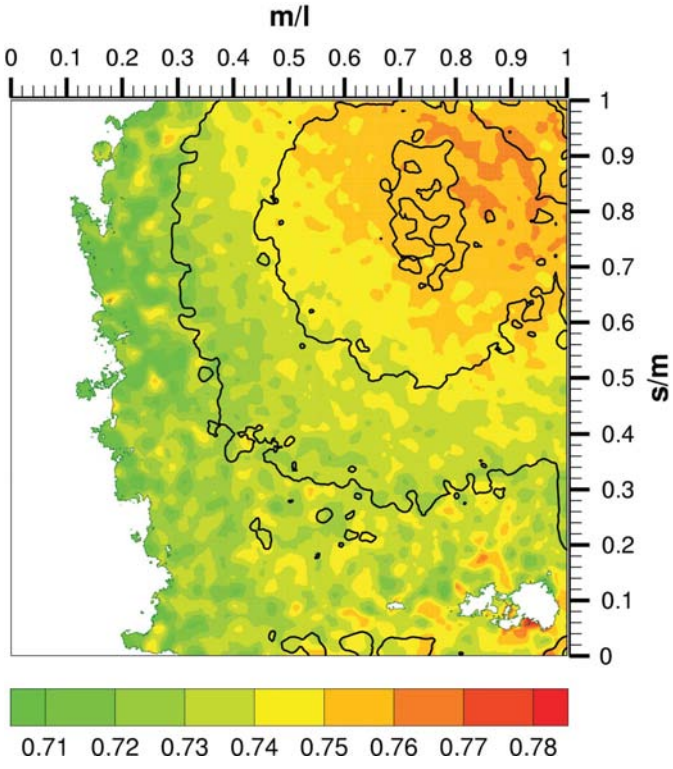

Figure 11
Distribution of packing coefficients with respect to the shape of molecules as characterized by the aspect ratios of their ellipsoid of second moments. The colour code used for a specific range of packing coefficients is shown at the bottom of the diagram. For less than 10 structures per unit area, the colour was set to white. The diagram refers to the basis dataset of 112 993 structures. Iso-frequency lines relate to the weighted number of data per unit area (analogous to Fig. 8, see §2.6). They are given for values of 40, 145 and 250 in an outside in sequence (see also Fig. 8).

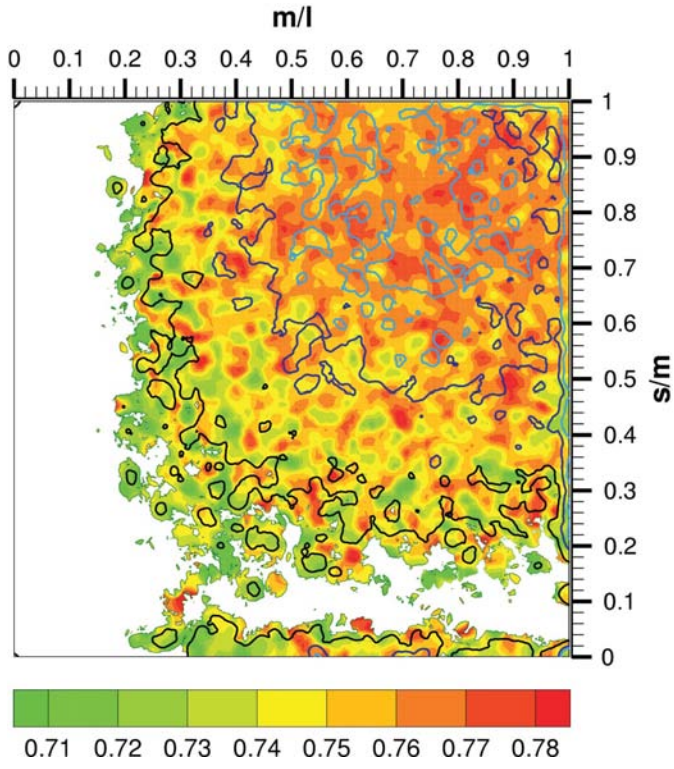

Figure 12
Distribution of packing coefficients for the dataset *HLsub*. The colour code used for the individual ranges of packing coefficients is given at the bottom of the diagram. It is the same colour code as used in Fig. 11. For less than 5 contributors per unit area, the colour was set to white. Iso-frequency lines indicate the weighted number of structures per unit area (see §2.6); black: 5; dark blue: 20; light blue: 35.

with different degrees of tilting allowed. The planes of molecules along neighbouring lines of the pattern of centred hexagons are rotated relative to each other around the normal to the layer plane by angles close to 90° (Kitaigorodskii, 1979).

Even though the shape descriptor used is very approximate, some correlations between the shape and the kind of packing are obvious with this descriptor. It is remarkable that the predictions based on this descriptor [forecasting an arrangement of the centres of molecules in a pattern of centred hexagons with equal or alternating inter-layer distances and constant or alternating shift vectors (dataset *HLsub*)] may well have a probability of above 35% to be true.[3] It is also remarkable that there is an even higher probability (up to more than 90%) that a given molecule will not pack in this type of pattern.

3.3. Density of packing

The volume of all the molecules was determined using the same algorithm (see §2.7). The solvent-excluded volume was used throughout. The distribution of packing coefficients as

[3] In the dataset used, structures represented more than once were not eliminated. Fig. 9 is hence biased by counting these structures several times, but we have established that eliminating such multiplicates in specific prominent areas does not significantly change the appearance of the diagram.

calculated on this basis is shown in Fig. 11 for the whole dataset of 112 993 structures. The colour coding in Fig. 11 represents the average packing coefficient in an area around each point in the diagram. This average was calculated as a weighted average per unit area by the same approach as described in §2.6 for finding the probability of regular stacking of layers of centred hexagons (Fig. 9). The iso-frequency lines indicate the number of structures in a given area.

It is evident that high packing coefficients (red colour) are especially probable around aspect ratios of m/l and $s/m \simeq 0.8$ (Donev *et al.*, 2004). It is well known that the upper limit for the packing coefficient of spheres is 0.74 (Sloane, 1998). The

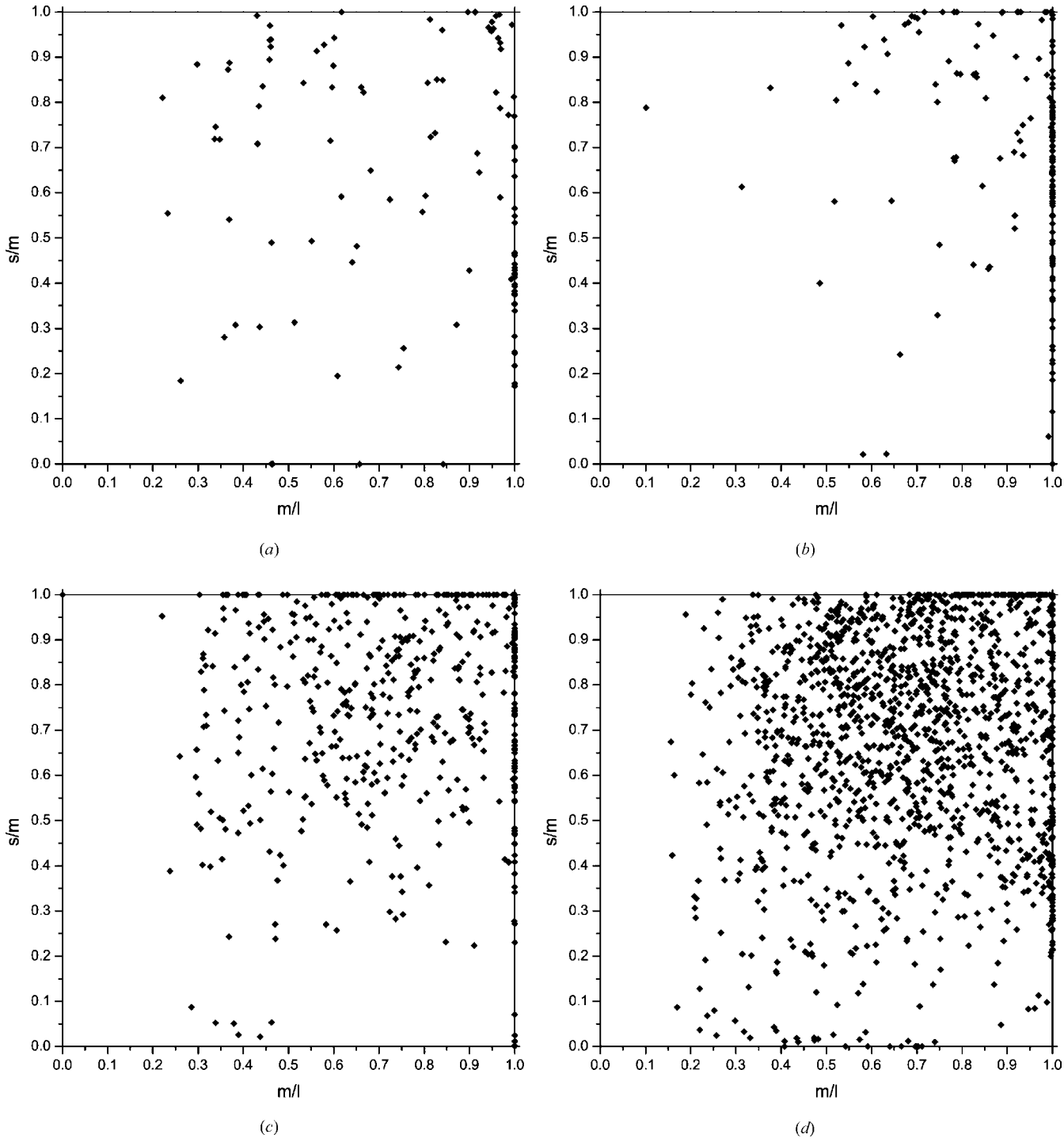

Figure 13
Plots showing the scatter of specific stacking sequences. (*a*) represents the distribution of structures with an *AA* stacking sequence with respect to the aspect ratio of the shape ellipsoids of the respective molecules. (*b*) shows analogous data for the stacking sequence *AB* corresponding to an h.c.p.-type arrangement. (*c*) presents analogous data for an *ABC* stacking sequence (f.c.c.-type arrangement). (*d*) relates to the distribution of an *AB* stacking sequence in b.c.c. mode.

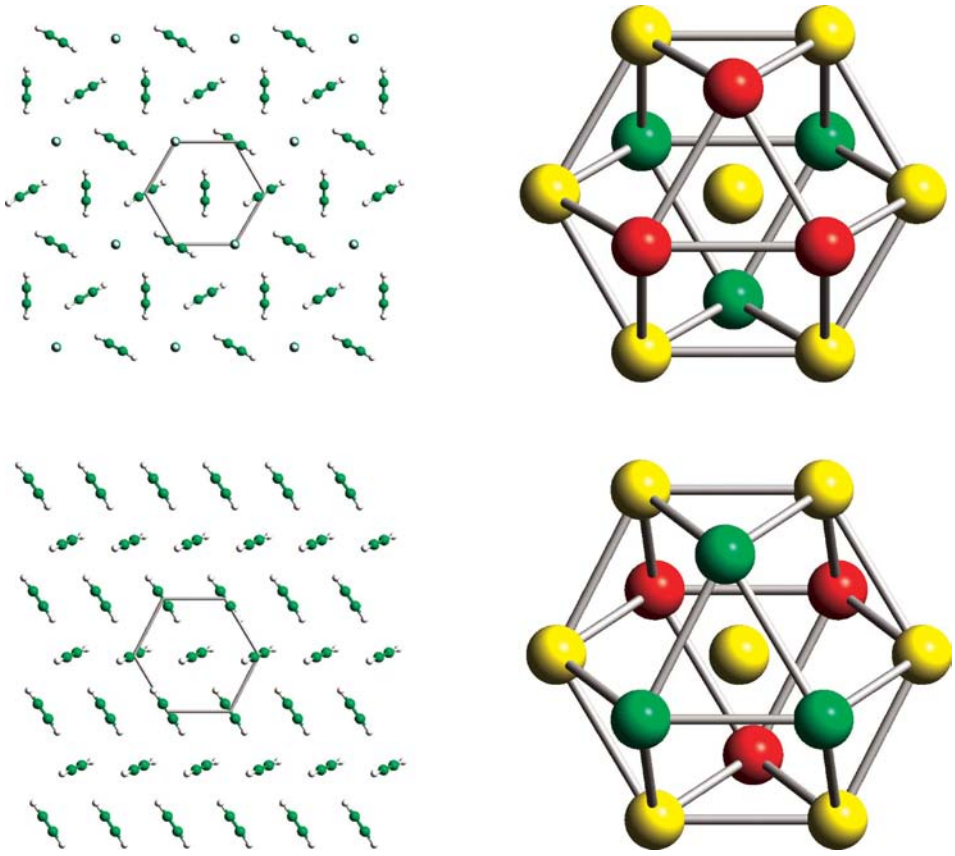

Figure 14
Illustration of two kinds of f.c.c.-type packing of acetylene (ACETYL02 at the top, ACETYL11 at the bottom). The left-hand side of the diagram: projection of the molecules on to the plane of centred hexagons; right-hand side: cuboctahedron illustrating the position of the centres of molecules in adjacent layers.

packing coefficients of real molecules may well be higher (general area of the diagram, see also Fig. 8) and packing of molecules characterized by second-moment ellipsoids without rotational symmetry (at axes m/l and s/m) tend to pack with higher packing coefficients than those corresponding to rotationally symmetric ellipsoids (see Fig. 11).

If the subset of structures for which the regular packing of layers of centerd hexagons was found (*HLsub* with 19 751 structures, see §3.1) is analysed the same way for the distribution of packing coefficients, the diagram shown in Fig. 12 is obtained. In comparing both diagrams (Figs. 11 and 12) it is apparent that the density of packing is higher in the mean for this subset than for the whole set of data. Is is tempting to hypothesize that the high frequency of occurrence of structures containing stacked layers of centred hexagons is correlated with the high packing coefficient induced by this kind of arrangement.

3.4. Stacking sequence

From the set of structures corresponding to an arrangement of equidistant layers of centred hexagons (*HLsub* with 19 751 structures, see §3.1) only a small percentage conforms almost exactly to the special type of stacking characteristic of h.c.p.-, f.c.c.- or b.c.c.-type packings. Even if the distance between individual layers is not used as a classifing criterion and if hence only the shift vector and its repetition pattern (see §2.5) are taken into account, the number of structures conforming to these nevertheless still quite tight criteria is small. There are 116 structures which show stacking with a shift vector of **0**, stacking sequence *AA* (Fig. 13*a*). There are 175 structures which show an *AB* pattern of stacking with a shift vector corresponding to $|\mathbf{r}_0| = \frac{1}{3}$; $\zeta = 0$ ($0.325 \leq |\mathbf{r}_0| \leq 0.345$, $\zeta \leq 0.1$; see §3.1, Fig. 13*b*). An *ABC* pattern is found for 540 structures ($0.33 \leq |\mathbf{r}_0| \leq 0.35$, $\zeta \leq 0.1$; Fig. 13*c*). There are 1534 structures which show a stacking pattern characteristic of b.c.c.-type structures corresponding to a shift vector of $|\mathbf{r}_0| = \frac{1}{2}$; $\zeta = 0$ ($0.495 \leq |\mathbf{r}_0| \leq 0.5$, $\zeta \leq 0.1$; Fig. 13*d*).

A high probability of some of these types of stacking is observed at the borders of the corresponding diagrams, that is at $m/l \simeq 1$ or $s/m \simeq 1$, respectively. The second-moment ellipsoids characterizing the points along these lines have rotational symmetry. The preponderance of b.c.c.-type lattices (see Fig. 13*d*) might be correlated with the fact that in this type of packing there are 14 nearest neighbours instead of only 12 in an f.c.c.- or h.c.p.-type lattice (Peresypkina & Blatov, 1999, 2000).

To illustrate some of the various possibilities of molecules which give rise to one or the other of these special stacking sequences, the layers found in two different modifications of acetylene are shown in Fig. 14. In both cases, the stacking sequence is *ABC* and even the layer distance is close to the ideal value of an f.c.c. lattice, as illustrated by the shape of the corresponding coordination cuboctahedron shown.

4. Conclusion

An algorithm is described which allows for the automatic retrieval of special types of molecular packing. By applying this algorithm to 112 993 structures containing 'one residue'

stored in the CSD it is found that about a third (36 678 structures, dataset *HL*) of these structures show a packing pattern made up from layers of centred hexagons stacked onto each other. If this subset is screened for patterns made up from layers of centred hexagons with the additional restriction of equal or alternating inter-layer distances and shift vectors a subset (dataset *HLsub*) of 19 751 structures is found to meet these additional criteria.[4]

Packing coefficients are found to be somewhat higher on average for the dataset *HLsub* than for the whole set of 112 993 structures. The type of packing characteristic of the dataset *HLsub* appears to be correlated with the molecular shape in a characteristic way: molecules for which the ellipsoid of second moments ('ellipsoid of inertia') has rotational symmetry show a high probability of packing this way, as do rod-type molecules with an aspect ratio of $m/l \simeq 0.25$ and planar molecules ($s/m \simeq 0$). For some combinations of m/l and s/m, the probability that the packing of the respective molecules conforms to the category defined for *HLsub* is higher than 30%. For some other combinations, the probability that the respective molecule will not form centred hexagonal layers is over 90%.

We are grateful to the German Science Foundation (*DFG*) for financial support (HU 151/31-1) and the Graduiertenkolleg 850 'Modellierung von Moleküleigenschaften'.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Barone, V. & Cossi, M. (1998). *J. Phys. Chem. A*, **102**, 1995–2001.
Bondi, A. (1964). *J. Phys. Chem.* **68**, 441–451.
Braun, N. & Huttner, G. (2004). *Crystpack*2.0. www.aci.uni-heidelberg.de.
Brock, C. P. (1996). *J. Res. Natl. Inst. Stand. Technol.* **101**, 321–325.
Brock, C. P. (1999). *NATO Sci. Ser. Ser. E*, **360**, 251–262.
Brock, C. P. & Dunitz, J. D. (1994). *Chem. Mater.* **6**, 1118–1127.
Cole, J. C., Yao, J. W., Shields, G. P., Motherwell, W. D. S., Allen, F. H. & Howard, J. A. K. (2001). *Acta Cryst.* B**57**, 88–94.
Donev, A., Cisse, I., Sachs, D., Variano, E. A., Stillinger, F. H., Connelly, R., Torquato, S. & Chaikin, P. M. (2004). *Science*, **303**, 990–993.
Frisch, M. J. *et al.* (2003). *GAUSSIAN*03, Revision B.03. Gaussian, Inc., Pittsburgh PA, USA.
Kitaigorodskii, A. I. (1945). *J. Phys. Acad. Sci. USSR*, pp. 351–352.
Kitaigorodskii, A. I. (1961). *Organic Chemical Crystallography*. New York: Consultants Bureau.
Kitaigorodskii, A. I. (1970). *Adv. Struct. Res. Diffr. Meth.* **3**, 173–247.
Kitaigorodskii, A. I. (1979). *Molekülkristalle*. Berlin: Akademie.
Kuleshova, L. N. & Antipin, M. Y. (1999). *Russ. Chem. Rev.* **68**, 1–18.
Lommerse, J. P. M., Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Mooij, W. T. M., Price, S. L., Schweizer, B., Schmidt, M. U., Van Eijck, B. P., Verwer, P. & Williams, D. E. (2000). *Acta Cryst.* B**56**, 697–714.
Motherwell, W. D. S. (1997). *Acta Cryst.* B**53**, 726–736.
Motherwell, W. D. S. (1999). *Nova Acta Leopold.* **79**, 89–98.
Motherwell, W. D. S. (2001). *Mol. Cryst. Liquid Crys. Sci. Technol. A*, **356**, 559–567.
Motherwell, W. D. S., Ammon, H. L., Dunitz, J. D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D. W. M., Leusen, F. J. J., Lommerse, J. P. M., Mooij, W. T. M., Price, S. L., Scheraga, H., Schweizer, B., Schmidt, M. U., van Eijck, B. P., Verwer, P. & Williams, D. E. (2002). *Acta Cryst.* B**58**, 647–661.
Peresypkina, E. V. & Blatov, V. A. (1999). *Theochem.* **489**, 225–236.
Peresypkina, E. V. & Blatov, V. A. (2000). *Acta Cryst.* B**56**, 501–511.
Pertsin, J. A. & Kitaigorodskii, A. I. (1987). *The Atom-Atom-Potential Method. Applications to Organic Molecular Solids*, Vol. 43, Springer Series in Chemical Physics. Berlin: Springer.
Pidcock, E., Motherwell, W. D. S. & Cole, J. C. (2003). *Acta Cryst.* B**59**, 634–640.
Reichling, S. & Huttner, G. (2000). *Eur. J. Inorg. Chem.* **5**, 857–877.
Sloane, N. J. A. (1998). *Nature (London)*, **395**, 435–436.
Wells, A. F. (1983). *Structural Inorganic Chemistry*, 5th Ed. Oxford: Clarendon Press.
Yao, J. W., Cole, J. C., Pidcock, E., Allen, F. H., Howard, J. A. K. & Motherwell, W. D. S. (2002). *Acta Cryst.* B**58**, 640–646.

[4] The datasets *HL*, *HLsub* and the set of 112 993 structures that served as a basis of the analysis are available from the IUCr electronic archives (Reference: BM5019). Services for accessing these data are described at the back of the journal.