Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Distribution of molecular centres in unit cells with respect to packing patterns

Elna Pidcock* and W. D. Sam Motherwell

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England

Correspondence e-mail: pidcock@ccdc.cam.ac.uk

Received 16 April 2004
Accepted 7 July 2004

Packing patterns, a new description of the limited number of possible arrangements of molecular building blocks in a unit cell, were assigned to many thousands of structures belonging to the space groups $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$ and $C2/c$ [Pidcock & Motherwell (2004). *Cryst. Growth. Des.* **4**, 611–620]. The position of the molecular centre (in fractional coordinates) in the unit cell for these structures has been surveyed, with respect to the space group and the packing pattern. The results clearly show that the position at which the molecular centre is found in the unit cell is correlated with the packing pattern. The relationships between the orientation of the packing pattern in the unit cell and the symmetry operators of the space group are explored. Popular orientations of packing patterns within the unit cell are given.

1. Introduction

Patterns of intermolecular interactions, synthons and motifs can be used to help rationalize molecular crystal structures, but our understanding of crystal packing remains incomplete. Although familiar, polymorphism is not well understood and little can be said *a priori* about whether a molecule will be polymorphic. Even less is known about the many possible polymorphic structures a molecule may exhibit (Bernstein, 2002). Crystal structure prediction trials routinely generate hundreds of seemingly reasonable crystal structures for a given molecule: these structures are close in energy and it appears that there is no single reliable criterion that can be used to separate the 'right' structure from the 'wrong' ones (Lommerse *et al.*, 2000; Motherwell *et al.* 2002). Over the years much work has been undertaken to understand crystal packing; a seminal work being that of Kitaigorodskii (1961) and his examination of how symmetry, crystallographic and molecular, is involved in the close packing of structures. Another large body of work has focused on the role intermolecular interactions play as structure-directing motifs (for examples see Etter, 1990; Desiraju, 1991, 1996, 2002). Efforts have also been made to understand packing from a topological standpoint, examining molecular structures for their similarity to those of packed spheres, *i.e.* body-centred cubic, face-centred cubic and hexagonal close-packed (Peresypkin & Blatov, 2000; Reichling & Huttner, 2000). As part of the continuing effort to demystify crystal packing, we proposed a new model which, for the first time, related unit-cell dimensions to molecular dimensions in a systematic way (Pidcock & Motherwell, 2003, 2004). In the model the unit cell is viewed as a container of molecules and the molecules are arranged in the cell in one of a limited number of packing patterns. For example, the basic structure of a unit cell containing four molecules is either an arrangement of $2 \times 2 \times 1$ molecules (*i.e.*

two molecules high, two molecules wide and one molecule deep) or $1 \times 1 \times 4$ molecules (*i.e.* a stack of four molecules, one on top of the other). For a molecule with the dimensions L, M, S, where $L > M > S$, the unit-cell dimensions are dependent upon which molecular dimension corresponds to which 'direction' of the $2 \times 2 \times 1$ or $1 \times 1 \times 4$ molecular array. To illustrate, a unit cell that is described by $2L \times 2M \times 1S$ ($221S$) is shaped like a book, but a unit cell that is described by $2M \times 2S \times 1L$ ($221L$) has more equal cell dimensions and thus will be more cubic (Fig. 1). A study of thousands of experimental crystal structures demonstrated that the above, simple model provides a viable description of molecular crystal packing (Pidcock & Motherwell, 2004). Molecular dimensions (L, M, S) were determined by 'measuring' the spatial extent of the molecule (difference between the maximum and minimum coordinates including van der Waals radii) along each of the three, perpendicular principal axes of inertia of the molecule. Each molecular dimension was paired with a unit-cell axis by choosing, for two of the molecular dimensions, the smallest angle between the principal axis of inertia describing the molecular dimension and an orthogonalized cell axis. The third pairing of the molecular dimension with a unit-cell axis was assumed. Thus, each unit-cell axis was assigned to a molecular dimension, Fig. 2. Pattern coefficients ($C_{L,\,M\text{ or }S}$) were calculated by dividing each cell axis length by its assigned molecular dimension. Thus, $C_L = \mathrm{Cell}_L/L$, $C_M = \mathrm{Cell}_M/M$ and $C_S = \mathrm{Cell}_S/S$, where $\mathrm{Cell}_{L,\,M\text{ or }S}$ is the unit-cell length to which the molecular dimension (L, M or S) was assigned. Histograms of pattern coefficient *versus* frequency showed very clear peaks at values corresponding to the 1, 2 and 4 of the model above.[1] Each structure was assigned to the packing pattern for which the difference between the calculated pattern coefficients and 'ideal' pattern coefficients was the smallest (for details, see Pidcock & Motherwell, 2004). It was found that not all the packing patterns were populated to an equal extent by experimental crystal structures and those with the most equal unit-cell dimensions and lowest surface area were preferred. Correlations were observed between broadly defined molecular shapes and packing patterns. Molecules with non-cubic shapes were found to be more discriminating in their choice of packing pattern, favouring those patterns that led to the lowest surface-area unit cells to a greater extent than their cubic counterparts. However, no details concerning how the packing patterns were accommodated by space groups were elucidated, but relationships between the packing patterns and cell axes were observed. For example, in $P2_1/c$ it was noted that the '1' of the 221 packing patterns (or '1 direction') was found to be most frequently aligned with the unit-cell a axis.

Previously, Motherwell (1997) conducted a study of the positions of the molecular centres in unit cells of the most populated space groups. For space groups such as $P\bar{1}$, $P2_1$ and $P2_1/c$, with one molecule per asymmetric unit, the most populated regions for molecular centres within the cell were, in general, found to be those which fell mid-way between symmetry operators. These results hinted at some very general rules of crystal packing, because the structures belonging to the datasets were a very diverse mixture of organic and organometallic compounds, and included those with and without hydrogen bonding and electrostatic interactions. However, in the absence of information on the dimensions of unit-cell axes, the position of a molecular centre in fractional coordinates is of limited use. In this paper we survey the distribution of molecular centres in space groups with respect to packing patterns, and examine the relationships between unit-cell axes and the orientation of packing patterns. From molecular dimensions it is possible to estimate the most likely

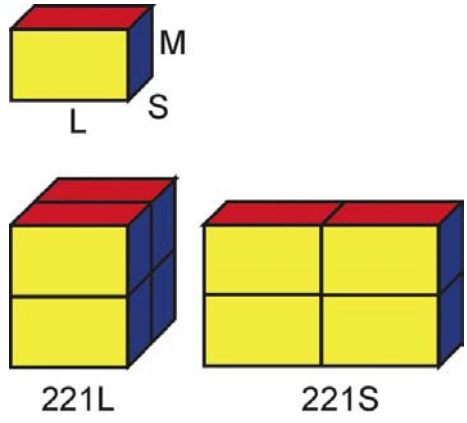

Figure 1
Using boxes of dimensions $L > M > S$ to represent molecules, the packing patterns $221S$ (right) resulting in a 'book-shaped' array and $221L$ (left) resulting in a more cubic array are shown. The coloured faces of the box highlight the different total surface areas of the two packing patterns.

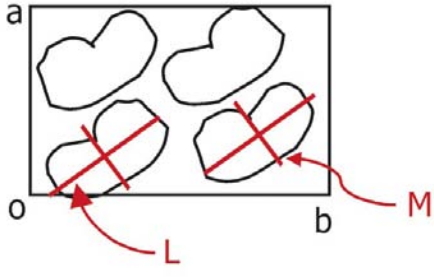

Figure 2
Illustration in two dimensions of the relationships between the molecular dimensions, principal axes of inertia and cell lengths. The cell contains four molecules. Two molecules are shown with their principal axes of inertia marked in red. The extent of the molecule along the principal axes gives the molecular dimensions L and M. Consideration of the angles between the principal axes of inertia and the cell axes shows that L is more closely aligned with the b cell axis and M is more closely aligned with the a cell axis. Thus, the pattern coefficients are $C_L = b/L$ and $C_M = a/M$.

[1] Pattern coefficients were calculated by dividing the cell axis length by the assigned molecular dimension. In a previous paper the average pattern coefficients were presented. The calculated pattern coefficients did not have integer values. Approximately '1' in the box model corresponds to a value of 0.8–0.9 in crystal structures; '2' of the box model corresponds to $\sim$ 1.6 in crystal structures *etc.* See Pidcock & Motherwell (2004) for details.

unit-cell dimensions, within certain confidence limits, with the aid of the most common packing patterns. Supplementing this knowledge with the probable position of the molecule within the unit cell represents another step on the path to understanding crystal structures.

2. Calculation details

Datasets of structures for the five most populated space groups $P\bar{1}$, $P2_1/c$, $P2_12_12_1$, $P2_1$ and $C2/c$ were generated from searches of the Cambridge Structural Database (hereafter known as the CSD; Allen, 2002; Bruno *et al.*, 2002). The datasets were composed of structures with the space-group settings as above and with a maximum of one molecule in the asymmetric unit. No restrictions were placed on the type of molecule included in the datasets, hence organic and organometallic compounds are represented. Dataset sizes were typically several thousand molecules, *e.g.* 12 426 in the case of $P2_1/c$ and 4543 for $P\bar{1}$. These datasets were used in the previous work of calculating pattern coefficients (Pidcock & Motherwell, 2004) and thus each structure is assigned to a packing pattern and the orientation of the molecule in the cell is known. *RPluto* (Motherwell *et al.* 1999) was used to process the structures and output the coordinates of the molecular centre for each structure in each dataset. The method of calculation of the coordinates for the molecular centres was the same as that used by Motherwell (1997): the centre of the molecule was defined as the mean value of all the x, y, z coordinates for all the reported atoms, including H atoms if present. During the processing of the structures by *RPluto* an algorithm was applied to the given molecular centre to move it to be as close to (0,0,0) as possible through the application of allowed shifts of 1/2 and the inversion operation. It is the coordinates of these 'standardized' molecular centres, x_c, y_c and z_c that are shown in the scatterplots below.

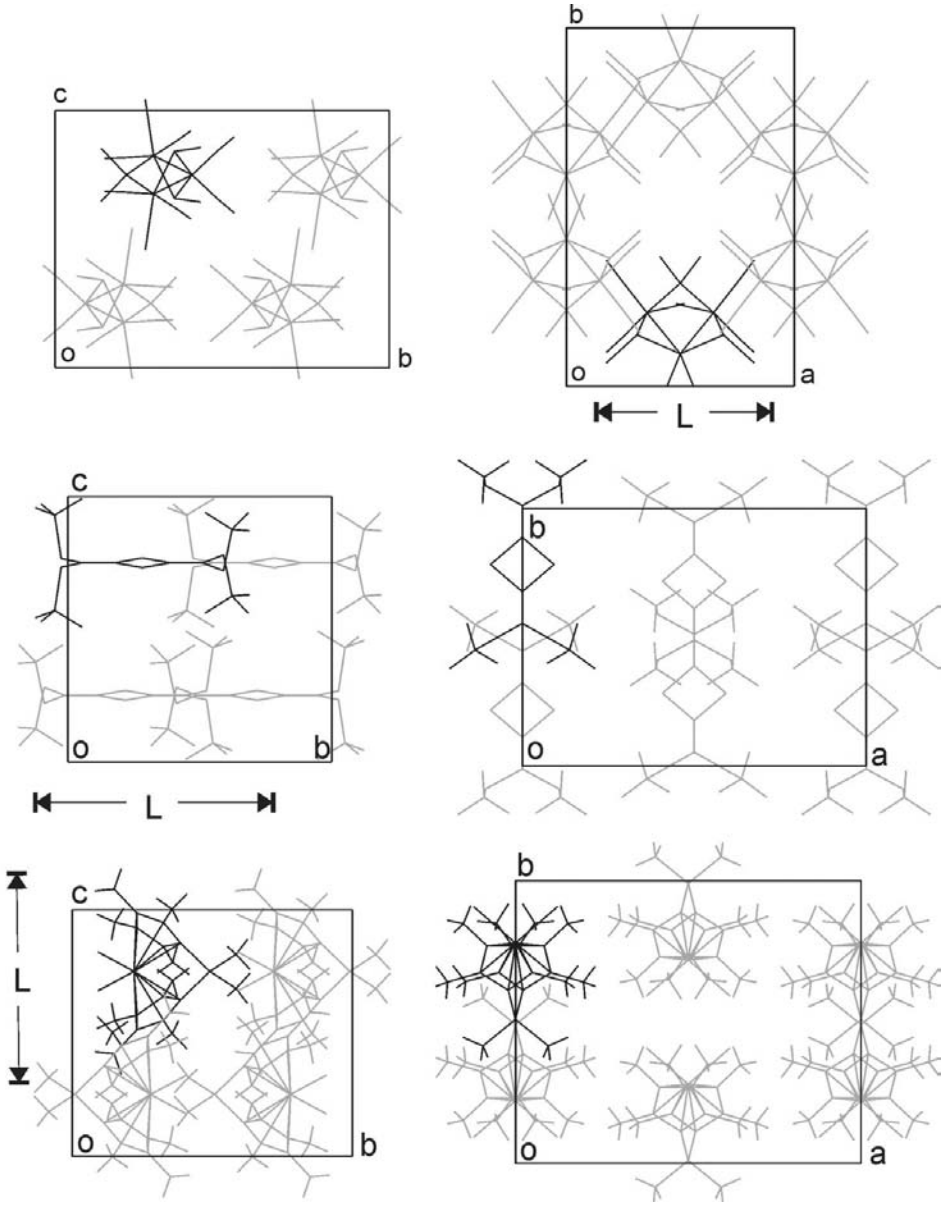

Figure 3
Three $Z = 4$ structures in $C2/c$. The '1 direction' of the 221 packing pattern is coincident with the *a* axis (top), CSD Refcode ALLCOB (Keller & Vahrenkamp, 1981), *b* axis (middle), CSD Refcode PIPNEH (Boese *et al.*, 1994), and *c* axis (bottom), refcode WIYJOD (Koch *et al.*, 2000). A marker has been added to each structure to highlight the orientation of the longest molecular dimension (L).

3. Distribution of molecular centres

To illustrate the interplay between the packing pattern and the space group, three molecules with approximately the same dimensions each belonging to a structure of the space group $C2/c$ have been chosen. Further, each molecule has C_2 (2) symmetry and is found residing on the twofold rotation axis of the space group at 0, y, $\frac{1}{4}$. Each structure is assigned to the 221L packing pattern family and hence the unit-cell dimensions are given

Table 1
Molecular and cell dimensions of three structures belonging to $C2/c$, where $Z = 4$.

Bold typeface indicates the $1L$ cell axis. Pattern coefficients were calculated from cell length/molecular dimension.

Refcode	Molecular dimensions (L, M, S) in Å	Cell dimensions (a, b, c) in Å	Pattern coefficients, C_L, C_M, C_S
ALLCOB	11.5, 9.5, 8.7	**10.2**, 15.8, 12.4	0.9, 1.7, 1.4
PIPNEH	11.7, 9.0, 8.1	15.0, **10.8**, 11.3	0.9, 1.7, 1.4
WIYJOD	10.7, 9.6, 9.0	15.5, 12.2, **11.1**	1.1, 1.6, 1.4

by approximately $1.6M$, $1.6S$ and $0.8L$.[2] In the three structures the '1 direction' of the 221 packing pattern, the direction in which only a single molecular dimension is accommodated by a unit-cell axis, coincides with the a, b or c unit-cell axis. The '2 direction' of the packing pattern is the direction in which two molecules, related by a symmetry operator, are accommodated along a unit-cell axis. As demonstrated by Fig. 3 and the unit-cell dimensions given in Table 1, each axis in $C2/c$ is capable of accommodating either one or two molecules. All axes in $C2/c$ are symmetrically non-equivalent: the symmetry operators that act on molecules in the directions of the unit-cell axes are different, for example the twofold axes are parallel to **b** and the glide planes are situated in the ac plane. Despite the non-equivalence of the unit-cell axes, in the structures chosen, molecules of approximately the same size, which belong to the same packing pattern and which occupy the same position in a unit cell, have unit cells with similar dimensions. However, which particular cell axis is the longest and which particular cell axis is the shortest depends on how the packing pattern is oriented with respect to the cell axes. All combinations of packing pattern and unit-cell axes are possible, but there are preferred orientations. For example, of the structures belonging to the 221 packing patterns in $C2/c$ with $Z = 4$, it is most commonly observed (66% of structures, 2336/3554; Pidcock & Motherwell, 2004) that the '1 direction' of the packing pattern is aligned with the b axis. Therefore, the interaction between space group and packing pattern is potentially very informative, particularly in the field of crystal structure prediction. The top five space groups (in popularity) have been examined in terms of packing pattern, cell axes and the position of the molecules within the unit cell.

3.1. $P\bar{1}$

In the space group $P\bar{1}$ the a, b and c cell axes are equivalent with respect to the symmetry operators. In the work of Motherwell (1997) it was noted that x_c, y_c and z_c values of $\frac{1}{4}$ were very common. The scatterplots indicated when one fractional coordinate was $\frac{1}{4}$, then the remaining two coordinates could take any value. In a $Z = 2$ structure, such as those belonging to $P\bar{1}$ (when the crystallographic inversion centre is not occupied), there are three packing patterns possible: $112L$, $112M$ and $112S$, where the repeated molecular dimension is either L, M or S. A subset of 112 structures was chosen where the '2 direction' of the packing pattern coincided with the unit-cell c axis and it was found that, irrespective of which molecular dimension was aligned with **c**, z_c had a value of $\frac{1}{4}$ (mean $z_c = 0.240$, $\sigma = 0.059$, 1158 observations). The position of the molecule on the a and b axes appeared to be unrestrained (Fig. 4). As expected, equivalent results were observed when the '2 direction' of the packing pattern was coincident with the a or b cell axes. Thus, in $P\bar{1}$ one fractional coordinate of the centre of the molecule is found at $\frac{1}{4}$ and the axis on which this occurs is the axis that corresponds to the '2 direction' of the 112 packing pattern. In $P\bar{1}$ all axes are equivalent and therefore an axis can be chosen arbitrarily to accommodate the '2 direction' and the centre of the molecule placed at $\frac{1}{4}$ on that axis. Coupling this result with the finding that 61% of structures in $P\bar{1}$ belong to the $112S$ packing pattern (Pidcock & Motherwell, 2004) yields a reasonable starting point for the prediction of a $Z = 2$ crystal structure.

3.2. $P2_1$

In $P2_1$ there is no distinction between the a and c axes and the screw axis of the space group is parallel to the unique b axis. The analysis of Motherwell has shown that the centres of molecules are commonly found with a value of $\frac{1}{4}$ on the a or c axis (the position of the molecule on the b axis is of course arbitrary and is not examined further). As above, for a $Z = 2$ structure there is one packing pattern family available, that of 112. Examples of structures belonging to $P2_1$ were found where the '2 direction' coincides with each of the unit-cell axes. A subset of molecules belonging to the $112S$ packing pattern was chosen for which the '2 direction' coincided with the a/c axis. For these structures x_c/z_c was clustered at $\frac{1}{4}$ (mean $x_c/z_c = 0.238$, $\sigma = 0.061$, 1404 observations) and z_c/x_c can take many values, although there is a slight clustering around $x_c/z_c = \frac{1}{4}$ and $z_c/x_c = 0$. When the '2 direction' is along the b axis there is no discernable structure to a scatterplot of x against z (Fig. 5). There does not appear to be any clear preference for the orientation of the packing pattern in the unit cell; structures are divided in the ratio of roughly 1/3:2/3 between those where the '2 direction' coincides with b or with a/c. The screw-axis symmetry operator of $P2_1$ is given by $1 - x, y + \frac{1}{2}, 1 - z$. Thus, the action of the symmetry-operator changes the coordinates of the molecule on all axes and from the above results

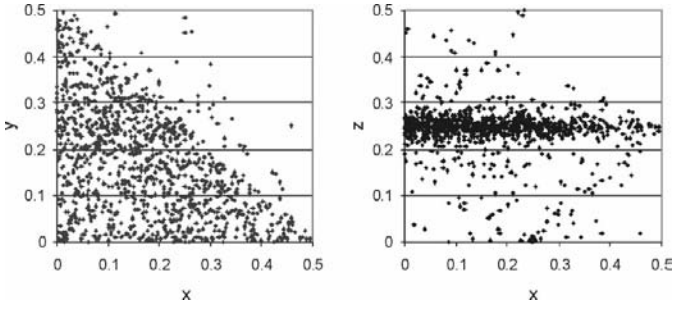

Figure 4
Distribution of molecular centres in $P\bar{1}$ where the '2 direction' of the 112 packing pattern is coincident with the unit-cell c axis. Left, projection onto the xy plane and right, onto the xz plane.

[2] See footnote **1**.

it appears that a reflection on an axis (*i.e.* $1-x$ and $1-z$) and a shift of $\frac{1}{2}$ ($y+\frac{1}{2}$) are equally efficacious at accommodating the '2 direction'. In order to further probe the interaction of symmetry operators with packing patterns a brief examination of structures belonging to *Pc* was performed. For $Z = 2$ structures belonging to *Pc*, it was observed that generally either the *b* or the *c* axis accommodated two molecules, but not the *a* axis. An inspection of the symmetry operator reveals why this may be so. The only symmetry operator of *Pc*, the glide plane, is given by $x, 1-y, z+\frac{1}{2}$. Transformations of the molecular coordinates occur on the *b* and *c* axes due to the action of the glide plane, but not the *a* axis. Therefore, there is no method available to translate the molecule along the *a* axis within the unit cell and hence the unit-cell *a* axis will always host a '1 direction' of the 112 packing pattern.

3.3. $P2_12_12_1$

In the space group $P2_12_12_1$, as in $P\bar{1}$, all the cell axes are equivalent. For $Z = 4$ structures there are two packing-pattern families available, namely 221 and 114. The 221 packing family dominates with 77% of $Z = 4$ structures in $P2_12_12_1$ assigned to one of these patterns. The 221*L* structures were chosen for further analysis and the scatterplots are shown in Fig. 6. The scatterplots clearly show that the placement of the molecule in the cell for pattern types 221 is not random. The scatterplot of the fractional coordinates of the molecular centre on the axes which correspond to the '2 direction' of the 221 pattern shows a very distinct clustering. There is a clear avoidance of the symmetry operators, particularly the screw axis that runs parallel to the '1 direction' (in this case, at $\frac{1}{4}$ on the ordinate). For the cell axis that coincides with the '1 direction' of the 221 pattern it appears that the fractional coordinate for the molecular centre can take any value and is reasonably evenly distributed across all values. Since a repeated molecular dimension is not required the regions 'occupied' by symmetry operators are not avoided. The results are quite different when structures assigned to the 114 packing patterns are examined. For the axis that accommodates four molecules the molecular centre on that axis was found to be at 1/8 (or 3/8 – a symmetry-related molecule will reside at 1/8 in these cases). Given a *z* coordinate of 1/8, the symmetry operators of the space group $P2_12_12_1$ generate new *z* coordinates at $z+\frac{1}{2} = 5/8$, $\frac{1}{2}-z = 3/8$ and $1-z = 7/8$, a 'stack' of molecules evenly spaced along the *z* axis. The remaining two coordinates, which correspond to a '1 direction' of the packing pattern, are observed to take any value, as above. Therefore, the scatterplots presented by Motherwell (1997) have been resolved into two distinct sets of distributions, dependent on the packing-pattern families to which the structures belong. Thus, a physical basis for the packing patterns is starting to emerge: the different patterns are due to, at least in part, the position the molecular centre adopts with respect to the symmetry operators.

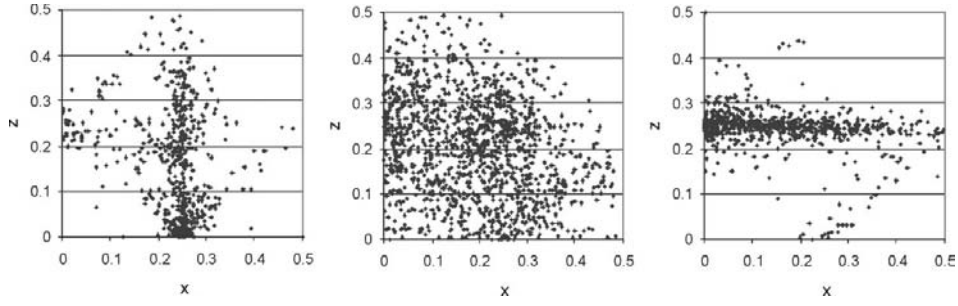

Figure 5
Distribution of molecular centres in $P2_1$. The '2 direction' of the 112 packing pattern is coincident with the *a* (left), *b* (middle) and *c* (right) unit-cell axes.

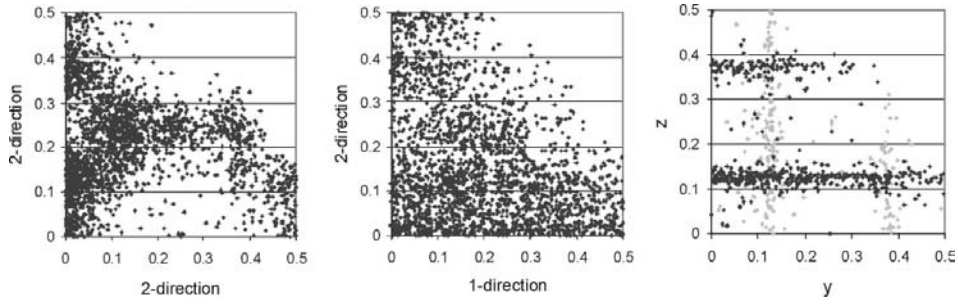

Figure 6
Distribution of molecular centres in $P2_12_12_1$. Left, view down the unit-cell axis that is coincident with the '1 direction' of the 221 packing pattern. Middle, view down a unit-cell axis that accommodates a '2 direction' of the 221 packing pattern. Right, distribution of molecular centres for structures belonging to 114*S*. The '4 direction' is coincident with the *b* (grey) and *c* (black) unit-cell axes.

3.4. $P2_1/c$

The analysis of structures belonging to $P2_1/c$ provides similar conclusions. In $P2_1/c$ no unit-cell axes are equivalent and there is more than one type of symmetry operator. The unit-cell *a* axis of $P2_1/c$ has inversion centres at 0 and $\frac{1}{2}$. Screw axes run parallel to the *b* axis and pass through the *ac* plane at 0, $\frac{1}{4}$ and the glide plane in the *ac* plane intercepts *b* at $\frac{1}{4}$. Examination of structures belonging to the 221 packing family, where the '1 direction' of the pattern is aligned with the *a* cell

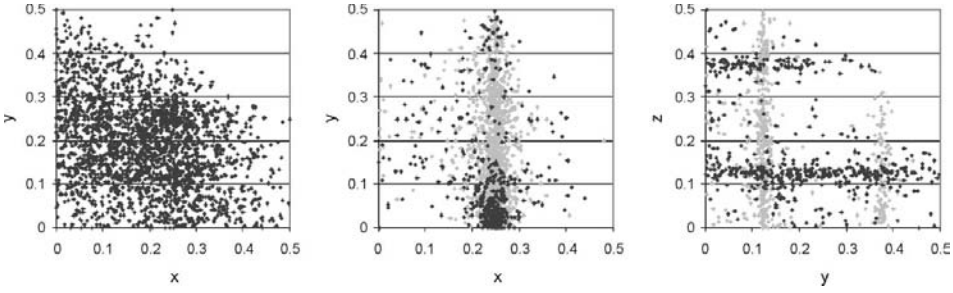

Figure 7
Distribution of the molecular centres of structures in $P2_1/c$. Left, structures assigned to the 221L packing pattern where the '1 direction' is coincident with the a unit-cell axis. Middle, structures belonging to the 221L packing pattern where the '1 direction' is coincident with **b** (grey) and **c** (black). Right, structures assigned to the 114S packing pattern. The '4 direction' is aligned with the b (grey) and c (black) cell axes.

axis, showed that the distribution of molecular centres is reasonably uniform, although there is some avoidance of the region at 0,0,0, the site of a centre of inversion. It appears that the arrangement of symmetry operators in $P2_1/c$ allow the packing pattern 221, where the '1 direction' is aligned with the a unit-cell axis, without placing any particular restraints on the position of the molecular centre. Translations of the molecular centre by $\frac{1}{2}$ are generated by the action of the glide plane and screw axes along the c and b axes, respectively, and thus c and b can accommodate two molecules. However, for structures where the '1 direction' of the packing pattern is aligned with b or c it is observed that the value of the x coordinate of the molecular centre is approximately $\frac{1}{4}$ (mean x_c = 0.241, σ = 0.044, 2186 observations), midway between the inversion centres. When the c axis is coincident with the '1 direction', y_c values tend to cluster around 0 (see Fig. 7). In the space group $P2_1/c$ the symmetry operators either leave x unchanged (the glide plane for example is given by $x, \frac{1}{2} - y, z + \frac{1}{2}$) or reflect x $(1 - x)$. Thus, an x coordinate of $\frac{1}{4}$ leads to x coordinates of $\frac{1}{4}$ and $\frac{3}{4}$ under the action of the symmetry operators, that is, two molecules equally spaced along the a unit-cell axis. Examination of the 114 packing family result in scatterplots similar to those observed for $P2_12_12_1$. When the 'stack' of four molecules is coincident with the b or c cell axes, the distribution of molecular centres clusters in regions where y = 1/8 (or 3/8) or z = 1/8 (or 3/8), respectively. As in the case of $P2_12_12_1$ the symmetry operators of $P2_1/c$ generate coordinates of 3/8, 5/8 and 7/8 from a starting point of 1/8 on the y or z axes. There are few examples of structures belonging to the 114 packing pattern where the four molecules are aligned along **a** (113/1815 structures, 6.2%). As noted above, the x coordinate is either left unchanged or is reflected by the symmetry operators and hence there are only two positions generated for x. Therefore, it seems that the a axis cannot accommodate four molecules and structures that have been assigned to such a packing pattern/cell axis combination have been done so incorrectly. It is likely that these structures are members of the 221 packing pattern family (where x_c = 1/4), but the presence of a molecular dimension < 5 Å results in a large value for the calculated pattern coefficient (cell length/molecular dimension). Since the assignment of a packing pattern to a structure relies on a measure of the goodness-of-fit between the calculated pattern coefficients and 'ideal' pattern coefficients (Pidcock & Motherwell, 2004), the presence of a large value for a pattern coefficient may make the difference between the assignment of the structure to a 114 packing pattern rather than a 221 packing pattern. The proportion of molecules, where the 4-direction is aligned with the cell axis a, with a molecular dimension < 5 Å is 57%, the value calculated for the entire dataset is 8.4% and for the 114 packing patterns the value is 17.2%.

3.5. *C*2/*c*

For Z = 8 structures there are two possible packing pattern families, namely 222 and 421. Structures belonging to the space group $C2/c$, where Z = 8, have been examined and in accordance with the results obtained for $P2_12_12_1$ and $P2_1/c$

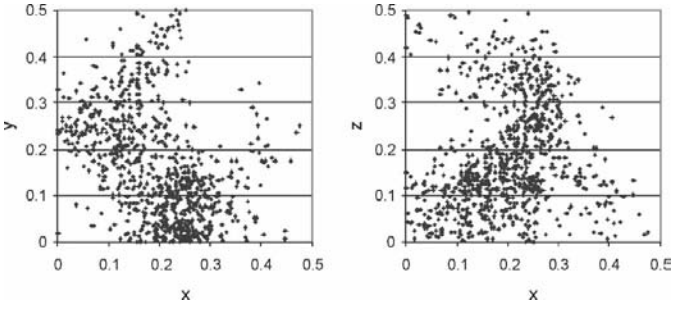

Figure 8
Distribution of molecular centres in the xy plane, left, and xz plane, right, for structures belonging to $C2/c$ and that were assigned to the 222 packing pattern.

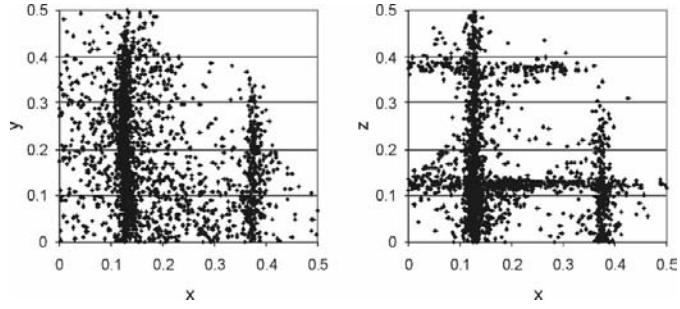

Figure 9
Distribution of molecular centres in the xy plane, left, and xz plane, right, for structures belonging to $C2/c$ and that were assigned to a member of the 124 packing pattern family. x_c or z_c is $\sim$ 1/8 (3/8) when the a or c cell axis accommodates the '4 direction' of the packing pattern.

Table 2
Distribution of packing-pattern orientations within unit cells in $P2_1$, $P2_1/c$ and $C2/c$.

Space group	Packing pattern (observations in italics)	Packing-pattern direction	Coincident cell axis	Number of obs.	%
$P2_1$	112	'2 direction'	*a* or *c*	3485	56
	6228		*b*	2743	44
$P2_1/c$	221	'1 direction'	*a*	5589	53
	10 610		*b*	4156	39
			c	865	8
	114	'4 direction'	*a*	113	6
	1815		*b*	852	47
			c	865	48
$C2/c$	421	'4 direction'	*a*	1527	66
	2320		*b*	32	1
			c	761	33
		'1 direction'	*a*	95	4
			b	2124	92
			c	101	4
		'2 direction'	*a*	698	30
			b	164	7
			c	1458	63

structures belonging to the two packing pattern families can be distinguished in terms of the position of the molecule in the unit cell. For structures belonging to the 222 packing pattern, when x is $\frac{1}{4}$, y tends to be zero and when x is 0, y tends to be $\frac{1}{4}$. The twofold axis running parallel to the b axis ensures that an area around $x = 0$, $z = \frac{1}{4}$ is almost entirely unpopulated (Fig. 8). For a molecular centre at $\frac{1}{4}, 0, \frac{1}{4}$ the symmetry operators of $C2/c$ generate further centres translated by $\frac{1}{2}$ on all axes, *i.e.* two molecules evenly spaced on each cell axis.

For structures belonging to the 421 packing patterns the scatterplots are reminiscent of those obtained for the 114 packing patterns (Fig. 9). It is clear that when an axis accommodates four molecules the molecular centre is found at ~1/8 (or 3/8) on that axis. There are very few structures (less than 2%) for which the stack of four molecules is found coincident with the b axis. Examination of the actions of the symmetry operators on coordinates does not provide an obvious explanation as to why the b axis is an unpopular choice for four molecules. Both the a and b axes appear equivalent in terms of the transformations that are applied to the coordinates. However, the presence of the twofold rotation axes parallel to the b cell axis require that when four molecules are coincident with **b**, the two molecules that constitute each of the four 'layers' are related by a twofold rotation axis. It appears that this is not a mode of crystal packing that is suitable for the majority of molecules. Indeed, Kitaigorodskii postulated that twofold rotation axes are not necessarily conducive to close packing. Once again, however, the packing pattern families for a particular Z value are distinguished by the different distributions of molecular centres.

4. Packing pattern orientation in unit cells

By examining the distribution of packing patterns and unit-cell axis combinations it is clear that there are preferred combinations. Thus, a clearer picture of crystal packing is emerging: popular packing patterns have been identified (allowing the estimation of unit-cell size from molecular dimensions) and within particular space groups the likely sites for molecules belonging to the packing patterns have been presented. All combinations of packing pattern and unit-cell axes are not equally populated. For example, in $P2_1/c$, of the 10 610 structures assigned to 221 packing patterns (Pidcock & Motherwell, 2004) 5589 of them are found with the '1 direction' of the packing pattern aligned with the a unit-cell axis, 4156 align the single molecular dimension with the b axis (and hence $x_c \simeq \frac{1}{4}$) and only 865 (8%) are found where c accommodates the '1 direction' of the packing pattern ($x_c \simeq \frac{1}{4}$ and $y_c \simeq 0$). Of the structures assigned to 114 packing patterns, 48.6% (850/1815) belong to a structure where the '4 direction' of the pattern is aligned with the c unit-cell axis and thus the z coordinate of the molecular centre is approximately 1/8 and 48.9% (852/1815) belong to a structure where the '4 direction' of the packing pattern is aligned with the b unit-cell axis (and hence $y_c = 1/8$).

In $C2/c$, when $Z = 8$, it has been observed that the b axis is most commonly found to accommodate a single molecule for structures belonging to the 421 pattern family. Of the 2320 structures assigned to the 421 pattern-packing family (Pidcock & Motherwell, 2004) it was found that 2124 (91.6%) showed the '1 direction' of the packing pattern to be coincident with the b unit-cell axis. The '4 direction' of the packing pattern was along either the a or c axis in 2288 cases (98.6%) and in 1527 of the 2288 cases (65.8%) the '4 direction' was coincident with **a**. In 61.3% of the $C2/c$ structures assigned to one of the 421 packing pattern family, the '4 direction' is along **a** ($x_c = 1/8$), the '2 direction' is along **c** and the '1 direction' is along **b**. In the case of structures belonging to the 222 packing pattern, the b axis is required to accommodate two molecules and it is found that in these structures the molecular dimension L aligns with **b** in 17.8% of cases and S aligns with **b** in 52.7% of cases.

For the space group $P2_1$, only the b axis is unique. It was found that the '2 direction' of the 112 packing pattern was reasonably evenly distributed over the three axes; 44.0% of structures (2743/6228) are found with the '2 direction' coincident with the b axis and the remaining 56.0% are found with the '2 direction' along the a/c axes [and mean x_c (z_c) = 0.242, $\sigma = 0.053$]. A summary of these statistics is given in Table 2. In $P2_12_12_1$ and $P\bar{1}$, where all axes are equivalent, there are no preferred orientations of the packing pattern within the cell. Thus, for $P\bar{1}$ the molecule centre will be at $\frac{1}{4}$ on one axis (of 112*S*, mean value = 0.248, $\sigma = 0.064$, 2783 observations) and this axis accommodates two molecules. For $P2_12_12_1$, when the structure belongs to the packing pattern 114, the centre of the molecule will be found at 1/8 on one axis and this axis will accommodate four molecules.

5. Conclusions

With the advent of packing patterns we believe that our understanding of crystal packing has developed. Crystal

structures can be classified into a limited number of packing patterns, depending on the number of molecules in the unit cell. The packing patterns are not populated to an equal extent by experimental crystal structures and the popular packing patterns are those that are 'most cubic', *i.e.* those that are described by the most equal cell dimensions and low surface area. In this paper a physical basis for the previously reported packing patterns has been established: the different packing patterns can be explained, in part, by the position of the molecular centre with respect to the symmetry operators of the space group. There are common features between space groups, for example, in structures where a unit-cell axis accommodates two molecules (112 in $P\bar{1}$ and $P2_1$, 221 in $P2_1/c$ with the '2 direction' along **a**), the molecular centres are generally found clustered at $\frac{1}{4}$ on that axis, irrespective of the symmetry operators available. Likewise, in structures where an axis accommodates four molecules (in $C2/c$, $P2_12_12_1$ and $P2_1/c$), the molecular centre is found at 1/8. When an axis accommodates only one molecule then it appears that the position of the molecular centre on that axis is unrestrained. The correlations between the packing pattern and the position of the molecular centre may help in the assignment of new structures to a packing pattern. Indeed, in the light of the above findings, where the consideration of pattern coefficients alone led to the incorrect assignment of a small percentage of $P2_1/c$ structures, it may be prudent to re-examine the classification of the structures that are outliers or which belong to less popular combinations of packing patterns and unit-cell axes.

At this juncture, although we are not at the stage of being able to predict crystal structures reliably, we feel we are making progress in reducing the required search space. Thus, given a molecule, an assessment of likely space groups can be made on the basis of the molecular symmetry (Pidcock *et al.*, 2003). With a chosen Z, the most likely packing patterns can be identified (using the idea of 'maximum cubicity') and further refined upon consideration of the molecular shape (Pidcock & Motherwell, 2004). The combination of molecular dimensions and the packing pattern means that unit-cell dimensions can be estimated for structures of the nominated space groups. Common orientations of the packing pattern within the unit cell can be explored with restraints placed on the position of the molecule within the unit cell. The reduction of search space is therefore quite substantial.

However, there is a great deal more to be done. For example, it is expected that the study of polymorphic structures, in the light of packing patterns, will be fruitful. In some cases molecules of the same conformation pack to yield different three-dimensional structures. It is likely that these structures will belong to different packing patterns and thus it may be possible to begin to understand why and what choices are available. In addition, the role played by intermolecular interactions in crystal structures has yet to be assessed in terms of packing patterns. However, we believe that we have developed a simple, useful and widely applicable model upon which to build our understanding of molecular crystal structures.

References

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

Bernstein, J. (2002). *Polymorphism in Molecular Crystals.* Oxford: Clarendon Press.

Boese, R., Haas, A., Lieb, M. & Roeske, U. (1994). *Chem. Ber.* **127**, 449–455.

Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. M., McCabe, P., Pearson, P. & Taylor, R. (2002). *Acta Cryst.* B**58**, 389–397.

Desiraju, G. R. (1991). *Acc. Chem. Res.* **24**, 290–296.

Desiraju, G. R. (1996). *Acc. Chem. Res.* **29**, 441–449.

Desiraju, G. R. (2002). *Acc. Chem. Res.* **35**, 565–573.

Etter, M. C. (1990). *Acc. Chem. Res.* **23**, 120–126.

Keller, E. & Vahrenkamp, H. (1981). *Chem. Ber.* **114**, 1111–1123.

Kitaigorodskii, A. I. (1961). *Organic Chemical Crystallography*. New York: Consultants Bureau.

Koch, T., Blaurock, S., Somoza Junior, F. B., Voight, A., Kirmse, R. & Hey-Hawkins, E. (2000). *Organometallics*, **19**, 2556–2563.

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., 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.

Motherwell, W. D. S., Shields, G. P. & Allen, F. H. (1999). *Acta Cryst.* B**55**, 1044–1056.

Peresypkin, E. V. & Blatov, V. A. (2000). *Acta Cryst.* B**56**, 1035–1045.

Pidcock, E. & Motherwell, W. D. S. (2003). *Chem. Commun.* pp. 3028–3029.

Pidcock, E. & Motherwell, W. D. S. (2004). *Cryst. Growth. Des.* **4**, 611–620.

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.* pp. 857–877.