Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Parameterization of the close packing of molecules in the unit cell

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 12 July 2004
Accepted 7 September 2004

The box model of crystal packing describes unit cells in terms of a limited number of arrangements, or packing patterns, of molecular building blocks. Cell dimensions have been shown to relate to molecular dimensions in a systematic way. The distributions of pattern coefficients (cell length/molecular dimension) for thousands of structures belonging to $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$ and $C2/c$ are presented and are shown to be entirely consistent with the box model of crystal packing. Contributions to the form of the histograms from molecular orientation and molecular overlap are discussed. Gaussian fitting of the histograms has led to the parameterization of close packing within the unit cell and it is shown that molecular crystal structures are very similar to one another at a fundamental level.

1. Introduction

Molecular crystal structures are close packed. The notion of close packing in crystal structures originates from the work of Kitaigorodskii (1961). From a few hundred crystal structures, Kitaigorodskii proposed that the basic rule of crystal packing was the efficient filling of space. Thus, popular space groups are those with symmetry operators, for example, screw axes and inversion centres, which allow 'bumps-into-hollows' type packing. Conversely, mirror planes, unless occupied, do not lead to efficient molecular packing and space groups containing mirror planes were proposed to be 'unpopular' with molecules lacking mirror symmetry. Over the years, Kitaigorodskii's theory of close packing has survived and the hundreds of thousands of available crystal structures have served to validate his proposals (Brock & Dunitz, 1994; Filippini & Gavezzotti, 1992). However, despite the fundamental nature of close packing in crystal structures, it appears to be largely absent from strategies for engineering crystal structure through molecular design. For example, hydrogen-bonding interactions are considered to be 'structure-directing' (Etter, 1990; Desiraju, 2002), but molecular shape is generally ignored. The introduction of a new model of crystal packing, has led to some interesting observations about the consistency of the parametrics of molecular crystal structures (Pidcock & Motherwell, 2003, 2004*a*). The box model of crystal packing describes the possible arrangements for a given number of boxes of three unequal dimensions, stacked with faces touching and edges aligned. There are a limited number of arrangements of a fixed number of boxes and these arrangements are termed packing patterns. The possible packing patterns for four boxes are those of the 221 family (for example, an array of four boxes, two boxes wide, two boxes high and one box deep) and the 114 family (a stack of four boxes, one on top of another). The packing patterns for a

given number of boxes have the same total volume, but are different in terms of total surface area. Applying the box model to crystal structures, a unit cell containing four molecules can be described in terms of an array of $2 \times 2 \times 1$ molecules or $1 \times 1 \times 4$ molecules and the cell lengths are given by multiples (or pattern coefficients) of molecular dimensions. Thus, the dimensions of the unit cell are dependent on the pairings of the pattern coefficients with molecular dimensions. To illustrate, for a molecule with the dimensions L, M and S, where $L > M > S$, a unit cell described by $1M \times 1S \times 4L$ ($114L$) is long and thin, whereas a unit cell described by $2M \times 2S \times 1L$ ($221L$) has more equal cell dimensions and the cell is closer to cubic (Fig. 1). From an analysis of structures belonging to the space groups $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $C2/c$ and $P2_1$, these packing patterns have been shown to be a viable description of molecular crystal structures (Pidcock & Motherwell, 2004*a*). It was found that packing patterns characterized by low surface area were populated to a greater extent by experimental crystal structures. Further work (Pidcock & Motherwell, 2004*b*) showed that the position of the molecule in the unit cell is correlated to the packing pattern that describes the arrangement of molecules in the cell. It was found, for example, that in structures belonging to $P2_12_12_1$ and assigned to a 114 packing pattern, the position of the molecular centre on the cell axis aligned with the '4-direction' of the packing pattern was 1/8: from a starting position of 1/8, symmetry-generated positions occur at 3/8, 5/8 and 7/8 along the same axis, *i.e.* four positions evenly spaced in accordance with the requirements of the 114 packing pattern.

The box model presents a simple view of crystal packing which has yielded some interesting observations. The preference for unit cells with a low surface area suggests that molecular shape is of primary importance. Relationships between molecular dimensions and unit-cell lengths have been established which indicate that, irrespective of the space group and symmetry operators, there is a fundamental description of crystal packing. In this paper we present further evidence that close-packing is expressed in the unit cell of crystal structures.

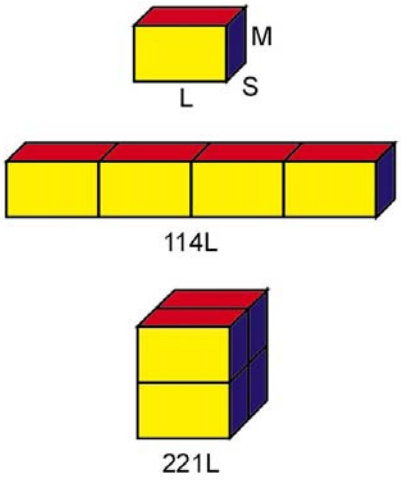

Figure 1
Using a box of the dimensions L, M and S, where $L > M > S$, to represent a molecule, the packing patterns $114L$ ($1M \times 1S \times 4L$) and $221L$ ($2M \times 2S \times 1L$) are shown. The two packing patterns represent unit cells (containing four molecules) of quite different dimensions. The unique 'direction' of the packing pattern is retained in the name, thus an array with the dimensions $1M \times 1S \times 4L$ is named $114L$: $4L$ is the unique direction and the other two molecular dimensions are multiplied by 1. In the case of the packing pattern $221L$, $1L$ is the unique direction and the other two molecular dimensions are multiplied by 2.

2. Calculation details

The three pattern coefficients (cell axis/molecular dimension) for each structure were calculated from datasets obtained from the Cambridge Structural Database (henceforth CSD; Allen, 2002), as described in Pidcock & Motherwell (2004*a*). The datasets contain structures belonging to the five most populated space groups: $P2_1/c$, $P\bar{1}$, $P2_12_12_1$, $P2_1$ and $C2/c$. A very large proportion (79%) of structures within the CSD belong to one of these five space groups. The datasets were retrieved from the Cambridge Structural Database (November 2002 release) using *Conquest* (Bruno *et al.*, 2002). Structures containing molecules of more than one chemical type and structures containing more than one molecule in the asymmetric unit were excluded and no alternative settings of the space groups were allowed, but no further restrictions were applied to the searches. The number of structures belonging to each dataset is as follows: $P2_1/c$ 15 882, $P\bar{1}$ 4857, $P2_12_12_1$ 8494, $C2/c$ 7782, $P2_1$ 6638. Briefly, the three perpendicular principal axes of inertia (PAI) for each molecule were established (using *RPluto*; Motherwell *et al.*, 1999) and the difference between the maximum and minimum coordinate on each axis (including van der Waals radii) was taken as a molecular dimension. Therefore, each molecule was described by three dimensions, L, M and S, where $L > M > S$. The angle between each vector that described the principal axes of inertia and each cell axis (in an orthogonalized cell) was determined. The orientation of the molecule was established by selecting the smallest angle between the PAI vector and cell axis for two of the PAI's. The third pairing of molecular dimension with cell axis was assumed. Thus, the molecular dimensions are known and each molecular dimension is paired with a cell axis. The three pattern coefficients for each structure were calculated by taking the ratio of the observed cell length to its paired molecular dimension.

3. The distribution of pattern coefficients

In the box model of crystal packing two arrangements of four boxes are possible; the 221 and 114 packing patterns (see §1). Thus, if a histogram was plotted of the distribution of pattern coefficients for a 221 arrangement of boxes (pattern coefficients 2, 2, 1), a peak at 1 and a peak at 2 (with twice the height) would be observed. A 114 arrangement of boxes would contribute a peak at 1 that was twice as high as a peak at 4. For each crystal structure the three pattern coefficients are calculated as above and we construct a histogram including all pattern coefficients. The histograms of pattern coefficient *versus* observations for $Z = 2$, $Z = 4$ and $Z = 8$ structures are

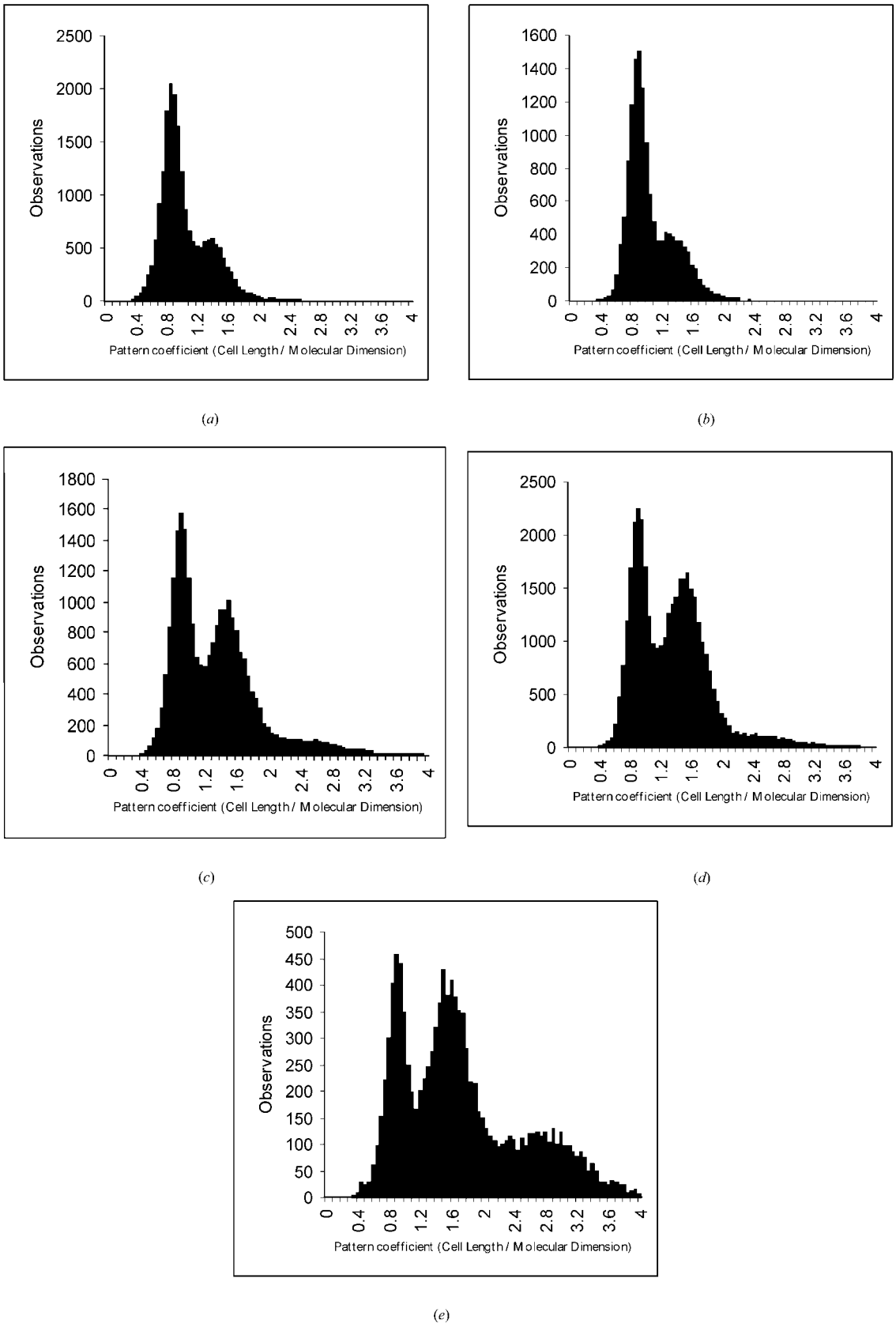

Figure 2
Histograms of pattern coefficients calculated for structures belonging to (*a*) $P2_1$, $Z = 2$; (*b*) $P\bar{1}$, $Z = 2$; (*c*) $P2_12_12_1$, $Z = 4$; (*d*) $P2_1/c$, $Z = 4$; (*e*) $C2/c$, $Z = 8$. The pattern coefficients were calculated by dividing each cell length by the molecular dimension most closely aligned with the cell axis (see Pidcock & Motherwell, 2004*a* for details), thus the orientation of the molecule is taken into consideration.

shown in Figs. 2(*a*)–(*e*). It can be seen that the histograms are structured with distinct peaks. The pronounced structure of these histograms and the similarity they exhibit to one another is an indication that the spatial arrangement of the contents of unit cells is ordered and is ordered in a way that is independent of the space group and symmetry operators. The histograms belonging to structures with the same Z in Fig. 2, *e.g.* for $P2_12_12_1$ and $P2_1/c$, look very similar. There are common features between all the histograms: the first peak, Peak 1, at the lowest value of the pattern coefficient is very consistent in position and shape. The peaks appear to broaden and the maximum height decreases as the pattern coefficient increases. To test the validity of the packing model, pattern coefficients were calculated for structures belonging to the $P2_12_12_1$ dataset, where a randomly selected cell axis was divided by a molecular dimension. A histogram of the results is presented in Fig. 3. As can be seen there is a broad peak centred at a value of approximately 1.2, which tails away slowly to a value of approximately 4 and there is little evidence of any secondary peaks.

It can be seen from Fig. 2 that the peaks of the histograms do not occur at the integer values expected from the idealized box model and as the pattern coefficient increases the differences between the 'ideal' box model values and the observed values increase. In order to understand the similarities and differences shown in the histograms a detailed analysis of the data has been performed and is presented in the next section.

3.1. Analysis of pattern coefficient data

3.1.1. Peak 1. The position of the first peak (Peak 1) is at approximately 0.9 and the width of the peak at half-maximum is approximately 0.3 for all histograms. This peak, according to the box model, represents the 1-direction, *i.e.* the situation where the unit-cell length is approximately equal to a molecular dimension and where molecules along this axis are equivalent and are related to one another by unit-cell translations.[1] In the box model the pattern coefficients are integers since the boxes are stacked with faces touching and edges aligned. Thus, if a molecule in a unit cell were aligned parallel to the cell axis and touched, but did not overlap, the neighbouring molecules, a pattern coefficient of 1 would be returned. However, molecules do not in general align with cell axes and there is often overlap with neighbouring molecules in the form of 'bumps-into-hollows'-type packing. Both of these factors lead to non-integer pattern coefficients. In order to understand the distribution of pattern coefficients observed for Peak 1, the orientation of the molecule with respect to the cell axes for structures belonging to $P2_12_12_1$ has been examined.

[1] In this paper two molecules are said to be 'translationally related' when the molecules reside in adjacent unit cells and are related only by a translation of an entire unit cell. Molecules are described as 'symmetry-related' or 'related by symmetry' when the molecules are related by a symmetry operator of the space group (listed in *International Tables*, Vol. A). For example, two molecules related by a glide plane or a screw axis will be referred to as symmetry related, although the symmetry operator, a glide plane or a screw axis, contains a translational component.

A subset of structures assigned to the packing patterns 221L and 114S were chosen where the 1L of the packing pattern was aligned with the a axis. For each of these structures $L\cos\theta$ where θ is the angle between the vector describing the L direction, and the unit-cell axis was calculated. The molecular dimension projected on the relevant cell axis, $L\cos\theta$, can be thought of as the 'expected' cell length, Fig. 4. The difference between the observed cell length and the expected cell length,

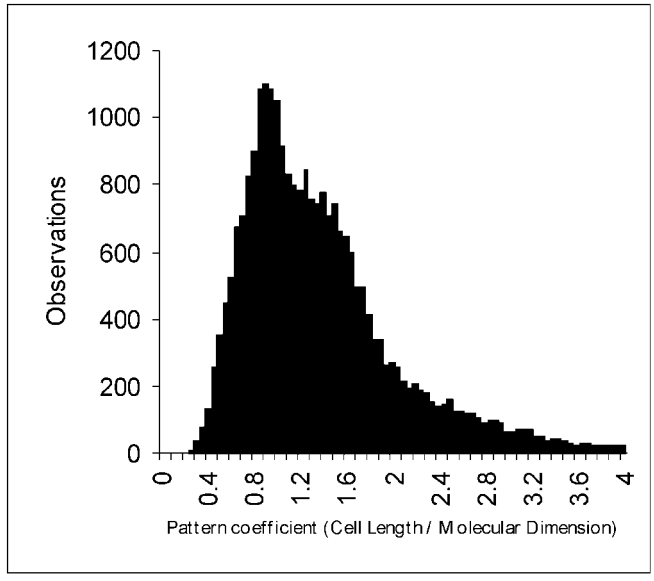

Figure 3
Random pattern coefficients calculated for structures belonging to $P2_12_12_1$. The pattern coefficients (cell length/molecular dimension) were calculated by randomly selecting which cell length of the structure was to be divided by a molecular dimensions L, M or S.

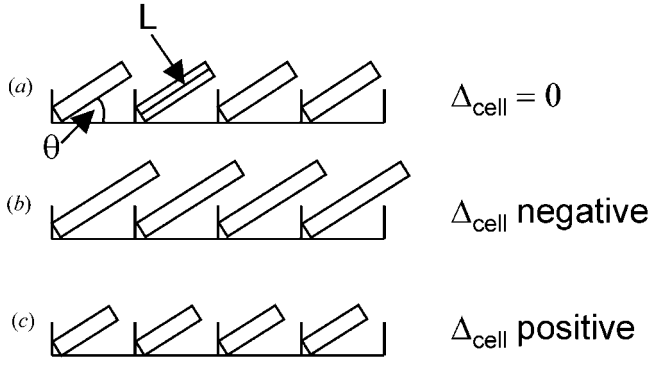

Figure 4
A schematic showing a view of four unit-cell axes which correspond to the 1-direction of the packing patterns where the molecules, represented by bars, are related by translation. The angle θ between the molecular dimension, in this case L, and the cell axis is shown. (*a*) $\Delta_{cell} = 0$, hence the cell length is in good agreement with $L\cos\theta$. The translationally related molecules do not overlap nor are there gaps between them. (*b*) Δ_{cell} is negative. The translationally related molecules overlap and share 'linear space' along the cell axis. (*c*) Δ_{cell} is positive. The unit-cell length is 'too long' for the molecule and hence there are gaps between translationally related molecules.

Δ_{cell}, is calculated for each structure. We now consider three categories of values of Δ_{cell}.

(i) Δ_{cell} approximately 0: When Δ_{cell} is approximately 0 then there is good agreement between the expected cell length ($L\cos\theta$) and the observed cell length. Therefore, irrespective of the orientation of the molecule within the cell, the molecule is touching, but is not overlapping with the space occupied by equivalent molecules in adjacent unit cells, *i.e.* the space occupied by its translationally related neighbours (see Figs. 4 and 5, top). On a scatterplot of C_L *versus* θ (Fig. 6*a*) the points shown in light blue are those for which Δ_{cell} is $\pm$ 0.5. It can be seen, not surprisingly, that these points follow the $\cos\theta$ line closely; for this set of data C_L (a/L) is equal to $\cos\theta$. The range of angles observed for these data fall between approximately 0 and 50°, and this equates to values of C_L between 1.0 and 0.6. From the scatterplot it can be seen that most of the data points lie between 0.6 and 1.0. Thus, the spread of values observed for the pattern coefficients of Peak 1 is due in part to the orientation of the molecule in the unit cell.

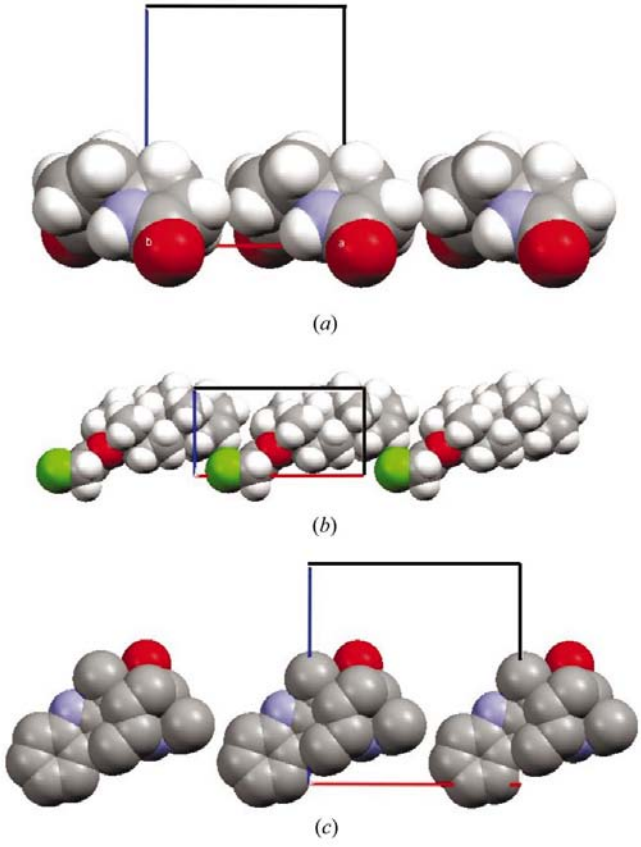

Figure 5
The view down the *b* axis of the structures belonging to $P2_12_12_1$, where the molecular dimension *L* is aligned with the *a* axis. The *a* axis is aligned with the 1-direction of the packing patterns and hence the molecules shown are related by translation. (*a*) AZNPOD (Czugler *et al.*, 1976), 221*L*, Δ_{cell} is approximately 0, hence translationally related molecules are just touching. (*b*) ANYCLA (Duax *et al.*, 1976), 114*S*, Δ_{cell} is negative hence the molecules are overlapped. (*c*) ARISTO (Gopalakrishna *et al.*, 1978) 221*L*, Δ_{cell} is positive and there are gaps between the molecules in the translation direction.

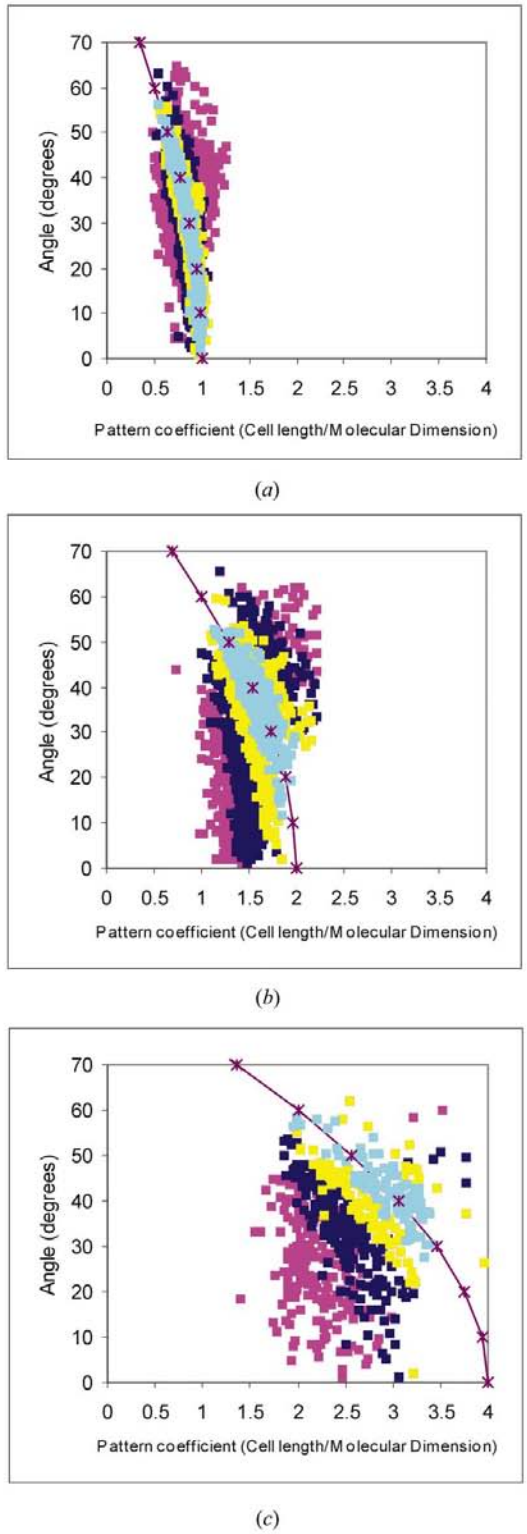

Figure 6
Scatterplots of selected subsets of the pattern coefficient *versus* θ for structures belonging to $P2_12_12_1$. The data points are coloured according to their value of Δ_{cell}: light blue, Δ_{cell} between $\pm 0.5n$; yellow, Δ_{cell} between $\pm 1n$; dark blue, Δ_{cell} between $\pm 2n$; pink, all data. (*a*) C_L *versus* θ for structures where 1*L* is aligned with the cell axis *a*, $n = 1$. (*b*) C_S *versus* θ for structures where 2*S* is aligned with the cell axis *b*, $n = 2$. (*c*) *CS versus* θ for structures where 4*S* is aligned with the cell axis *c*, $n = 4$. A curve calculated with $n(\cos\theta)$ is included on each scatterplot (purple line).

(ii) Δ_{cell} less than 0: When Δ_{cell} is less than 0 the expected cell length, calculated from $L\cos\theta$ is greater than the observed cell length (Fig. 4, middle). This represents the situation where the observed cell is 'too short' for the molecule and therefore there must be some overlap of the molecule with equivalent molecules in adjacent unit cells (see Fig. 5*b*). These data fall below the $\cos\theta$ line shown in the scatterplot of Fig. 6(*a*).

(iii) Δ_{cell} greater than 0: When Δ_{cell} is greater than 0 the expected cell length, calculated from $L\cos\theta$, is less than the observed cell length. Therefore, the observed cell length is 'too long' for the molecule and there is a gap between the molecule and its translationally related neighbours (Fig. 5*c*). These points on the scatterplot of C_L (a/L) lie above the $\cos\theta$ line.

It can be seen from the scatterplot (Fig. 6*a*) that the degree of overlap the molecule has with its translationally related neighbours causes a spread of the data above and below the $\cos\theta$ line. Thus, there are two factors that contribute to the spread of values observed for the pattern coefficients of the first peak: the orientation of the molecule and the degree of overlap with translationally related neighbours.

3.1.2. Peak 2. The second peak of the histogram in Fig. 2(*c*) represents pattern coefficients calculated for a 2-direction of the 221 packing pattern family for structures belonging to $P2_12_12_1$. In a 2-direction two molecular centres are found along the cell axis before the unit cell repeats and these molecules are related by the symmetry operators of the space group. The second peak of the histograms in Fig. 2(*c*) is significantly broader than Peak 1. As discussed for Peak 1, the factors of orientation and overlap are expected to contribute to the spread of values in the calculated pattern coefficients. In the case of Peak 1 (the 1-direction of the packing pattern) the expected cell length was given by $D_{mol}\cos\theta$, where D_{mol} is the molecular dimension, thus for a '2-direction' the expected cell length is given by $2(D_{mol}\cos\theta)$ because there are two molecular centres in this direction. Δ_{cell}, the difference between the observed cell length and $2(D_{mol}\cos\theta)$, was calculated for data where 2*S* was oriented with the *b* axis. The points of the scatterplot of the pattern coefficient, C_S *versus* θ, are coloured according to the value of Δ_{cell} (Fig. 6*b*). The reason for the broadening of Peak 2 is clear. For a range of orientations of the molecule with respect to the cell axis, for example, 0–50°, the expected values for the cell length are given by $2(D_{mol}\cos\theta)$ and hence lie between $2D_{mol}$ (0°) and $1.3D_{mol}$ (50°), twice the breadth of Peak 1. Thus, the range of orientations of the molecule in the cell coupled with the presence of two molecules accounts for the wide range of values observed for the pattern coefficient, C_S. However, as in the case of Peak 1, the degree of overlap between neighbouring molecules also affects the value of the pattern coefficient. It can be seen from the scatterplot of Fig. 6(*b*) that the bulk of the data points (64%) lie below the $2(\cos\theta)$ line, in the region of overlapping molecules. This is an increase in the proportion of data that is found below the $\cos\theta$ line compared with Peak 1 where 59% of the data was found in the region of molecular overlap.

3.1.3. Peak 4. The scatterplot of CS *versus* θ for Peak 4 (the third peak of the histogram) is given in Fig. 6(*c*) for data where 4*S* was aligned with the *c* axis in structures belonging to $P2_12_12_1$. Peak 4 of the histogram corresponds to the 4-direction of the packing pattern, where four molecules, related by the symmetry operators of the space group, are evenly spaced along a single axis of the unit cell. As above, the expected cell length is calculated from $4(D_{mol}\cos\theta)$ and the points on the scatterplot are coloured according to the value of Δ_{cell} [observed cell length $-$ $4(D_{mol}\cos\theta)$]. Again, it can be seen that both orientation and overlap contribute to the spread of values and the spread of values, as expected, is a factor of 4 greater than observed for Peak 1. A large proportion of the data fall below the $4(\cos\theta)$ line (89%) into the region of molecular overlap.

An understanding of the form of the histograms of Fig. 2 is emerging. The broadening of the peaks as the pattern coefficient increases can, to a first approximation, be explained by the relationship

$$\text{pattern coefficient} = D_{cell}/D_{mol} \simeq n\cos\theta, \qquad (1)$$

where n is the number of molecular centres found along the cell axis, *i.e.* 1, 2 or 4, and θ takes values between approximately 0 and 50°.

3.2. Position of peaks

It was noted above that the position of the maximum of Peak 1 in the histograms of Fig. 2 is remarkably consistent at a value of approximately 0.9. It may be expected from the above discussion that the position of Peak 2 is found at 2(Peak 1) or at 2×0.9 and the position of Peak 4 is at 4×0.9, since the breadths of the peaks have been shown to scale with a number of molecular centres. However, it is clear that this is not the case. In general, the peak positions for Peak 2 and Peak 4 are found at the pattern coefficient values substantially less than the 'expected' values. Therefore, the assumption (implicit in expectation that the peak positions of Peak 2 and Peak 4 scale directly the number of molecular centres) that molecular packing mediated by translations of an entire unit cell is as efficient as the molecular packing mediated by symmetry operators is wrong. As shown in the scatterplots of Fig. 6, the proportion of data found below the $n\cos\theta$ line, the region of molecular overlap, increases with increasing n (where n is the number of molecular centres). It seems reasonable to assume that the shift in the positions of Peak 2 and Peak 4 to below the 'expected' values is due to efficient close packing of the molecules (molecular overlap) mediated by symmetry operators when compared with the molecular packing achieved by the translational symmetry that relates neighbouring unit cells. In other words, the 'linear space' required for n molecules related by symmetry operators is less than the space required for n molecules related by translation.

Table 1
Parameters returned from the least-squares fitting of Gaussian curves to the distributions of pattern coefficient *versus* frequency for structures belonging to $P2_1$, $P\bar{1}$, $P2_12_12_1$, $P2_1/c$ and $C2/c$.

It should be noted that the positions of the peaks given below represent the pattern coefficients for the dataset. The goodness-of-fit is measured by the correlation coefficient, R^2, and is given in the last column of the table.

Space group	Height (Peak 1)	Position (Peak 1) $\bar{C}_1$	σ_1 (Peak 1)	Position (Peak 2) $\bar{C}_2$	Scale factor (k_2)	Position (Peak 4) $\bar{C}_4$	Scale factor (k_4)	Fit
$P2_1$	1942	0.86	0.13	1.37	0.5	–	–	0.993
$P\bar{1}$	1443	0.88	0.12	1.36	0.5	–	–	0.997
$P2_12_12_1$	1495	0.89	0.13	1.47	1.23	2.40	0.28	0.997
$P2_1/c$	2140	0.89	0.13	1.50	1.48	2.40	0.20	0.998
$C2/c$	432	0.89	0.13	1.57	1.8	2.74	1	0.987

4. Fitting of the pattern coefficient distributions

Now that the relationships between the peaks of the histogram are understood in terms of the model, it is possible to fit the distributions of pattern coefficient *versus* frequency (Fig. 2) using Gaussian curves and a least-squares fitting procedure (Kevin Raner Software, 2002). The first peak of all the histograms is fit with three parameters corresponding to the height (frequency of observations), the position [$\bar{C}_1$] and the breadth (σ_1) of the peak [see (2)]. The second peak of the histogram, Peak 2, for the same number of contributing values as Peak 1 would be expected, from the above analysis, to be described by height/2 and breadth $2\sigma_1$ to maintain an area equivalent to that of Peak 1. However, the packing patterns do not have equal occurrences of the different pattern coefficients. For example, in the 221 packing pattern family the area of Peak 2 would be expected to be twice that of Peak 1 since there are two occurrences of pattern coefficient '2' for every occurrence of pattern coefficient '1'. For $Z = 2$ structures, where the 112 packing pattern describes all structures, the parameters of Peak 2 are $2\sigma_1$ and 0.5(height/2) as there are half as many contributions from the 2-direction as the 1-direction in the packing pattern. It can be seen from the histograms of Fig. 2 that the height of Peak 2 is roughly 1/4 of that of Peak 1, in accordance with the model. For $Z = 4$ structures, the situation is further complicated by the presence of 2 packing patterns, those of the 221 family and the 114 family. The scaling of the area of Peak 2 and Peak 4 is unknown (the number of members of each packing pattern family is not known) and thus the scaling parameters (k_2 and k_4) of the height of Peak 2 and Peak 4 are included in the fitting of the distributions. The same situation arises for $Z = 8$ structures: two packing patterns are possible, that of the 421 family and the 222 family. Scaling of the area of Peak 2 (k_2) is included in the fit, but the area of Peak 4 is expected to be the same as the area of Peak 1. From the analysis above it has been shown that the breadths (σ) of Peak 2 and Peak 4 are $2\sigma_1$ and $4\sigma_1$, respectively, where σ_1 is the breadth of Peak 1. The parameters, σ_1, $2\sigma_1$ and $4\sigma_1$ are therefore used in the fitting of the distributions of pattern coefficient *versus* frequency for Peak 1, Peak 2 and Peak 4 where appropriate. The positions of Peak 2 and Peak 4 [$\bar{C}_2$ and $\bar{C}_4$, respectively] for all space groups are included in the fit as the positions do not scale with the number of molecular centres.

The equation used to fit the pattern coefficient *versus* the frequency distribution of the $P2_12_12_1$ data is given below and is the sum of three Gaussians, one for each peak of the histogram. The same equation is used to fit the pattern coefficient *versus* frequency distribution of $P2_1/c$ and $C2/c$. The sum of only the first two Gaussians is required for histograms of $P\bar{1}$ and $P2_1$.

$$\begin{aligned}\text{Observations} = {} & \text{height}\{\exp -[(C_1 - \bar{C}_1)^2/2\sigma_1^2]\} \\ & + k_2(\text{height}/2)\{\exp[(C_2 - \bar{C}_2)/2(2\sigma_1)^2]\} \\ & + k_4(\text{height}/4)\{\exp[(C_4 - \bar{C}_4)/2(4\sigma_1)^2]\} \quad (2)\end{aligned}$$

The same method is used to fit all of the pattern coefficient *versus* frequency distributions shown in Fig. 2 and the parameters used to fit the distributions are given in Table 1.

5. Discussion

Pattern coefficient *versus* frequency distributions together with fits for $P2_1$, $P2_1/c$ and $C2/c$ are shown in Fig. 7. It can be seen that the agreement between the experimental data and the calculated curves is extremely good. The box model of crystal packing is shown to be a viable representation of molecular crystal structures containing two, four or eight molecules in the unit cell. Peaks of the histogram are found at positions entirely consistent with the box model and the broadening of the peaks is accounted for by the orientation and overlap of the molecules within the cell. Broadly speaking the histograms and the parameters used to describe them are similar to one another. Certainly, the similarity between the histograms for structures of the same Z is striking. Thus, close packing, the primary rule governing molecular crystal packing, has been revealed within the unit cell.

Closer analysis of the histograms has shown that the first peak of all the pattern coefficient *versus* frequency distributions can be fitted with a Gaussian curve of the same parameters. Thus, the position and breadth of the first peak is invariant to the space group and Z of the structures. Peak 1 represents the situation where a cell length is approximately equal to a molecular dimension and hence where the molecule is related by translation to equivalent molecules in adjacent unit cells, along this axis, *i.e.* a 1-direction of the packing pattern. Thus, the consistency in the parameters of this peak

indicates that translation relates molecules in an unchanging way: $D_{cell} = (0.89 \pm 0.13) \times D_{mol}$.

The positions of Peak 2 and Peak 4, the 'symmetry peaks', are less consistent across the space groups than was observed for the 'translation peak', Peak 1. However, the parameters determined from each histogram for structures with the same Z are very similar. Thus, the Peak 2 parameters that describe the 2-direction in $Z = 2$ structures belonging to $P2_1$ and $P\bar{1}$ are the same, but different from the parameters obtained from Peak 2 for $Z = 4$ structures belonging to $P2_12_12_1$ and $P2_1/c$. The observation that the positions of Peak 2 are similar for structures of the same Z but which belong to different space groups indicates that the type of symmetry operator involved in the molecular packing does not have a significant effect on the values of pattern coefficient. Rather, the position of the symmetry peaks (Peak 2 and Peak 4) appears correlated with Z. For structures with $Z = 2$ the peak is centred at the lowest values observed and for structures with $Z = 8$ Peak 2 is positioned at the highest value. Similar behaviour is observed in the position of Peak 4. The position of the peak is very similar in structures of the same Z, for example, $P2_12_12_1$ and $P2_1/c$, but quite different for structures of different Z (for example, $P2_1/c$, $Z = 4$ and $C2/c$, $Z = 8$). Currently we have no explanation for this observed trend. A hypothesis is that as the number of molecules in the unit cell increases it is more difficult to pack the molecules 'ideally' and thus more 'space' per molecule is required. Alternatively, as suggested by a referee, if a molecule is 'awkward' to pack, the number of molecules within the unit cell increases. Clearly this is a topic that will be revisited. The observation that, irrespective of Z, the 'symmetry peaks' lie to lower values than expected from an equivalent number of translationally related molecules indicates that molecular packing mediated by symmetry operators is more efficient than molecular packing described by translation. Thus, in the packing of molecules related by symmetry operators, molecular overlap (bumps-into-hollows) is promoted and the space required by molecules related by symmetry operators is less than the space required by an equivalent number of translationally related molecules.

In a previous paper (Pidcock & Motherwell, 2004*a*) pattern coefficients were presented for structures belonging to particular packing patterns and space groups. These pattern coefficients were calculated by gathering structures assigned to the same packing pattern (on consideration of the goodness-of-fit between the calculated pattern coefficients and 'ideal pattern coefficients') and calculating average pattern coefficients for the dataset. In this paper pattern coefficients for structures belonging to a particular space group have been established but without the need of first assigning structures to a packing pattern. It is now possible to verify the previous assignment of packing patterns to structures. Representative pattern coefficients (and associated standard deviations) have been established and it will be possible to assess how well experimental structures fit to these 'real' (as opposed to 'ideal') pattern coefficients. It should be noted that the values of pattern coefficient presented in the previous work and in this work show good agreement, indicating that the initial packing pattern assignments were reasonable.

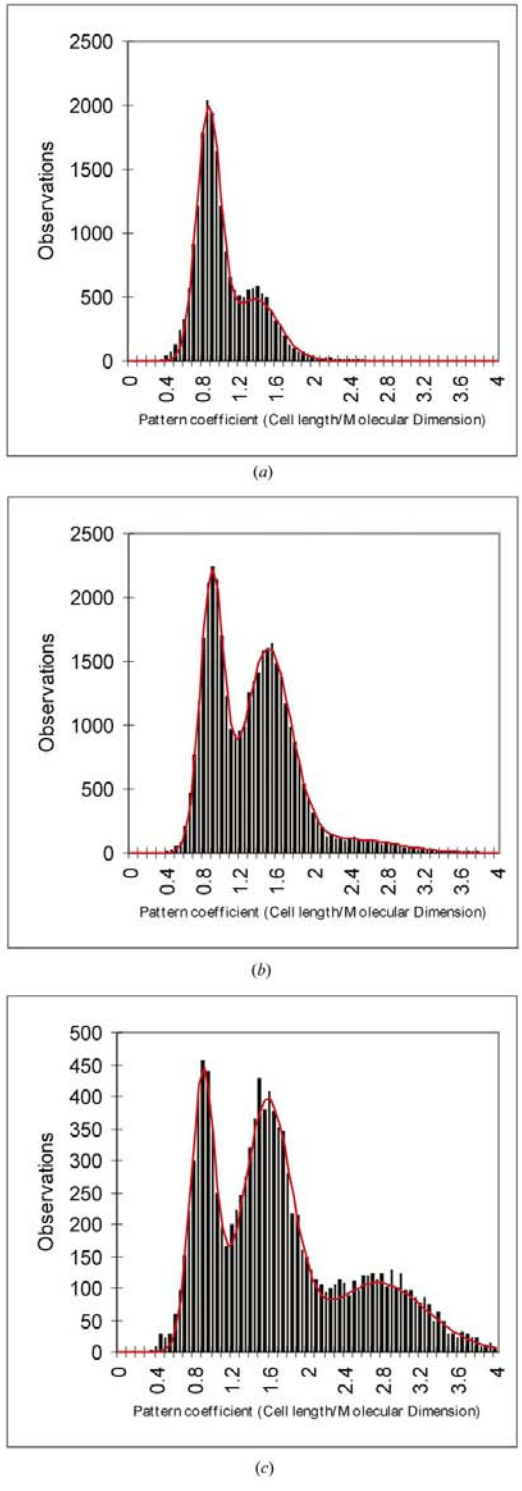

Figure 7
The distribution of pattern coefficients together with fits (red lines) calculated using Gaussian curves and the parameters given in Table 1 for structures belonging to (*a*) $P2_1$, (*b*) $P2_1/c$ and (*c*) $C2/c$.

6. Conclusions

A regular structure present in unit cells has been revealed in the plots of the pattern coefficient (the cell length/molecular dimension ratio), where the alignment of the molecule in the cell is taken into consideration. The histograms show very clear peaks at positions that are entirely consistent with the box model of crystal packing. There are a limited number of packing patterns that describe all $Z = 2$, 4 or 8 crystal structures and the packing patterns, for structures of different Z, have much in common. The uniformity of the histograms of Fig. 2 demonstrates that beneath symmetry operators and space groups there is a fundamental structure to unit cells and hence crystal structures: the concept of close packing has been parameterized.

The observation that, irrespective of Z, the 'symmetry' cell axes (as represented by Peak 2 and Peak 4 of the histograms; Fig. 2) are shorter than the linear space required by an equivalent number of translationally related molecules indicates that molecular packing mediated by symmetry operators is more efficient than molecular packing described by translation between unit cells. Thus, a crystal structure can be viewed as a collection of molecules related by 'tight links' (*i.e.* symmetry operators) and 'loose links' (translation). This is an oversimplification, but serves to illustrate the experimental results. This finding, coupled with the result from previous work that structures characterized by low surface-area packing patterns are most common, leads to an intriguing view of crystal packing. To return to the box model, low surface-packing patterns, *i.e.* 112S, 221L and 221M, for example, are constructed by placing the largest faces of the boxes together and hence the surface area of the overall arrangement is minimized. Therefore, crystal structures that are described by low surface-area packing patterns are those where the largest faces of the molecule interact *via* symmetry operators. Analysis of the histograms of Fig. 2 has shown that the symmetry operators mediate more efficient molecular packing than translation. Thus, for the majority of crystal structures the 'tight links' (symmetry operators) relate to the largest faces of the molecule and the 'loose links' (translation) relate to the smallest faces of the molecule. In other words, for the small faces of the molecule, where energetically there is little to be lost or gained by their interaction, translation, the 'loose link', is employed.

Molecular crystal structures, irrespective of space group, symmetry operators and Z are remarkably similar to one another at a fundamental level and it appears that the molecular shape is of primary importance in determining crystal structures. With these foundations it is hoped we can begin to understand the role played by molecular shape in the engineering of crystal structures.

References

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

Brock, C. P. & Dunitz, J. D. (1994). *Chem. Mater.* **6**, 1118–1127.

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.

Czugler, M., Kalman, A. & Kajtar, M. (1976). *Cryst. Struct. Commun.* **5**, 25.

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

Duax, W. L., Erman, M. G., Griffin, J. F. & Wolff, M. E. (1976). *Cryst. Struct. Commun.* **5**, 775.

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

Filippini, G. & Gavezzotti A. (1992). *Acta Cryst.* B**48**, 230–234.

Gopalakrishna, E. M., Watson, W. H., Silva, M. & Bittner, M. (1978). *Acta Cryst.* B**34**, 3778–3780.

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

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

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

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

Pidcock, E. & Motherwell, W. D. S. (2004*b*). *Acta Cryst.* B**60**, 539–546.

Kevin Raner Software (2002). *WinCurveFit.* Kevin Raner Software, 55 Smyth Street, Mt Waverley, Victoria 3149, Australia.