COMMUNICATION

ChemComm
www.rsc.org/chemcomm

A new model of crystal packing†

Elna Pidcock* and W. D. Sam Motherwell
Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, UK.
E-mail: pidcock@ccdc.cam.ac.uk; Fax: 44 1223 336033; Tel: 44 1223 762531

Received (in Cambridge, UK) 5th September 2003, Accepted 9th October 2003
First published as an Advance Article on the web 3rd November 2003

A new, conceptually simple model of crystal packing is proposed which uses "packing patterns" to describe unit cells in terms of molecular building blocks.

The question of why molecules pack to form particular crystal structures is an area of great scientific interest. The correct prediction of a crystal structure given only a molecular building block‡ is not achieved routinely or with any high degree of success.[1,2] In general hundreds of equi-energetic, densely-packed structures are generated and the problem is selecting the "right" one. Here we present a model that offers a new description of crystal structures and which it is hoped will contribute positively to the field of crystal structure prediction. The model allows estimation of unit cell size from molecular dimensions and thus during crystal structure prediction trials, the 3-dimensional search space is dramatically reduced. In addition, unit cell surface area is introduced as a parameter with predictive properties.

The following describes the stacking of a discrete number of boxes and the resulting dimensions of the rectangular container that encloses them. The identical boxes, with three unequal sides, $l > m > s$, are stacked with faces touching and edges aligned and hence are described as close-packed. There are six ways to close-pack four boxes and these are shown in Fig. 1. It can be seen that there are only two categories of packing pattern, namely 221 and 114 (where the integers correspond to the box-dimension multipliers henceforth pattern coefficients). The three permutations of the 221 pattern give pattern types 221(*l*), 221(*m*) and 221(*s*): note that the box dimension given in the pattern name corresponds to the dimension that is multiplied by the unique pattern coefficient. Similarly the three permutations of the 114 pattern give the 114(*l*), 114(*m*) and 114(*s*) pattern types (Fig. 1, bottom). A consequence of using a box with 3 unequal dimensions in the model (as opposed to a cube or a sphere) is that the resultant packing patterns can be distinguished in terms of surface area of the container. For example, it can be seen from Fig. 1 that the pattern 221(*l*) yields container dimensions that are most similar and for which the surface area of the container is minimised. The container with the largest surface area belonging to the 221 pattern category is that of 221(*s*). In the case of the 114 packing patterns it is clear that 114(*s*) gives a container with the smallest surface area and 114(*l*) has the largest surface area. The volume of the container is constant for all packing patterns.

Similar arguments can be applied to the stacking of 2 or 8 boxes in a close-packed manner. One pattern category exists for the packing of 2 boxes, namely 112 and the three pattern types 112(*l*), 112(*m*) and 112(*s*) give containers with different surface areas, 221(*s*) yielding a container with the smallest surface area. Three packing categories exist for stacking 8 boxes: 222, 421 and 118.

The top 5 space groups in the Cambridge Structural Database (CSD)[3,4] in terms of frequency of occurrence are $P2_1/c$ (35.5%), $P\bar{1}$ (21.6%), $P2_12_12_1$ (8.6%) $C2/c$ (7.7%), and $P2_1$ (5.6%). Allowing a maximum of one molecule per asymmetric unit ($Z' \leq 1$), the possible number of molecules per unit cell, Z, of these five space groups is 1, 2, 4 or 8. By reducing the description of a molecule to 3 dimensions (*cf.* a box) and likening the unit cell to the container, the relationships between molecular dimensions l_m, m_m, s_m and unit cell dimensions in $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$ and $C2/c$ have been examined.

The CSD (Nov 2002) was searched for crystal structures that belonged to a particular space group (no alternative settings were allowed), which contained a total of $Z = 2$, 4 or 8 molecules in the unit cell (depending on the space group) and which contained only molecules of a single chemical type. The hits from the search were then processed using RPLUTO[5] to determine the molecular dimensions in terms of the principal axes of inertia. The difference between the maximum and minimum atomic coordinates (including the van der Waals radii) on each of the three perpendicular axes was taken as a molecular dimension. The three dimensions were sorted in descending order, thus $l_m > m_m > s_m$. All permutations of the cell axes (D_{cell}) with molecular axes (D_{mol}) were taken to give the calculated pattern coefficients, $c_{l,m,s}$ (1).

$$c_{l,m,s} = D_{cell}/D_{mol} \quad (1)$$

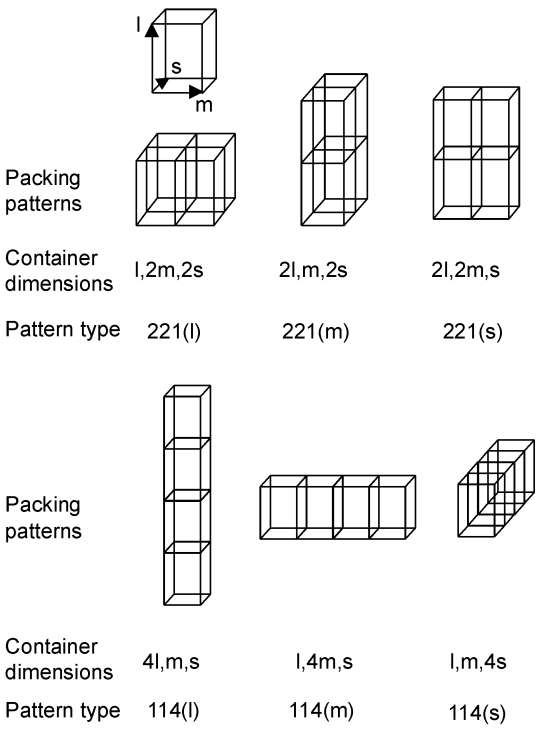

Fig. 1 Illustration of the model used to describe crystal packing. Top: four boxes, each with dimensions $l > m > s$ are stacked in 3 ways that are categorised as 221 patterns. Bottom: The three ways of stacking of 4 identical boxes with dimensions $l > m > s$ in the 114 pattern category.

† Electronic supplementary information (ESI) available: equation used to calculate goodness of fit, mean goodness of fit values calculated for $Z = 4$ datasets, results of pattern assignment for $Z = 2$ and $Z = 8$ structures, examples of unit cell dimension calculations. See http://www.rsc.org/suppdata/cc/b3/b310873b/

DOI: 10.1039/b310873b

For each permutation, a goodness of fit to the "target" pattern coefficients was determined, using the Euclidean distance metric (see ESI). Initially integer target pattern coefficients (from the box model) were used but a better agreement with experimental data was found when pattern coefficients derived from the ideal close-packing of spheres were used. In crystal structures as in close-packed spheres, there is an interpenetration of the layers of molecules not accounted for in the box model. Thus pattern coefficient 1 in the box model equates to 0.82, (the separation of two layers of close-packed spheres[6]), 2 equates to 1.64 *etc*. In the case of 4 molecules in a unit cell a goodness of fit was calculated for patterns 221 and 114 (GoF_{221} and GoF_{114}). GoF_{221} and GoF_{114} were compared and the lower value was used to assign the structure to the corresponding pattern, the results are presented in Tables 1 and 2. Results for structures with $Z = 2$ and 8 are given in the Supplementary Information. The mean GoF for each pattern type was found to be in the range 10–15, where 0 represents a perfect match between the target and calculated pattern coefficients and 100 indicates the difference between the target and calculated pattern coefficients is equal to the target pattern coefficients.§ *Thus the experimental data are in good agreement with the proposed model: packing patterns are applicable to crystal structures and unit cell dimensions are related to molecular dimensions*. For example 42% of structures in $P2_12_12_1$ belong to pattern type 221(l_m) and therefore cell axis $a \approx 0.82l_m$, $b \approx 1.64m_m$ and $c \approx 1.64s_m$. Examples of structures assigned to packing patterns 221(l_m) and 114(s_m) are shown in Fig. 2.

By comparing the number of observations in the two tables it can be seen that the 221 pattern category describes 20708/26091 or 79% of the unit cells with $Z = 4$ in these three space groups. The pattern category 114 is assigned to structures in 21% of cases. Closer inspection of the populations in the 221 pattern category shows that a large proportion (13784/20708, 67%) of unit cells fall within the 221(l_m) pattern type. A similarly uneven distribution in the population of patterns is found in the 114 pattern category, where the pattern type 114(s_m) accounts for 81% of structures. *Thus the most populated packing patterns are those identified by the box model as having the minimum surface area for a given volume.* Packing patterns that generate containers with larger surface areas, such as 221(s_m) and 114(l_m) are observed in a minority of structures.

Preliminary results have shown that by applying the packing patterns described above and using calculated packing coefficients it is possible to estimate the lengths of unit cell axes (examples are included in ESI). This is an important result to those interested in the field of crystal structure prediction. Instead of searching a unit cell of $30 \times 30 \times 30$ Å, (volume of 27000 Å^3; a typical unit cell volume is ~2500 Å^3) realistic lengths for unit cell axes can be estimated. The search space is dramatically reduced and thus the likelihood of finding the correct structure is significantly increased.

Table 1 Distribution of unit cells over 221 packing patterns for $Z = 4$ structures

		Number of structures assigned to pattern		
Space group	Total obs.	221(l_m)	221(m_m)	221(s_m)
$P2_1/c$	13516	7809	2388	842
$P2_12_12_1$	8494	3612	1781	655
$C2/c$	4081	2363	715	543

Table 2 Distribution of unit cells over 114 packing patterns for $Z = 4$ structures

		Number of structures assigned to pattern		
Space group	Total obs.	114(l_m)	114(m_m)	114(s_m)
$P2_1/c$	13516	28	301	2148
$P2_12_12_1$	8494	52	596	1798
$C2/c$	4081	0	58	402

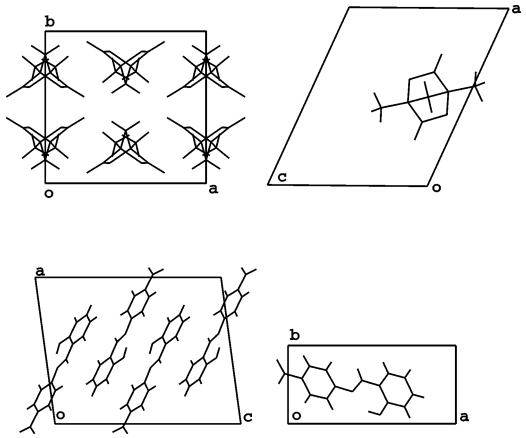

Fig. 2 Crystal structures assigned to pattern types 221(l_m) (top, CSD Refcode ABIVIQ[7]) and 114(s_m) (bottom, CSD Refcode, CAVQOF[8]). The views illustrate the correlations between molecular dimensions and cell axis lengths: with molecular dimensions $l > m > s$, in the 221(l_m) structure (top), approximately, $a \approx 2m$, $b \approx 2s$ and $c \approx l$. Likewise for the 114(s_m) structure (bottom), approximately, $a \approx l$, $b \approx m$ and $c \approx 4s$. For clarity, only one molecule of the unit cell is retained in the structures on the right.

In summary, a novel conceptually simple model of crystal packing has been presented. The model is space group independent and reinforces the principle of closest-packing.[9] It has been shown that the model is a viable representation of experimental crystal structures. The observed preference for the "most cubic" cell has implications for a number of areas including crystal structure prediction. A preliminary examination of polymorphic structures indicates that packing pattern and cell surface area considerations may prove illuminating.

Notes and References

‡ Successes have been reported in the field of inorganic materials. See S. M. Woodley, P. D. Battle, J. D. Gale and C. R. A. Catlow. *Phys. Chem. Chem. Phys*. 1999, **1**, 2535 and references therein.

§ A table of the calculated mean $GoF_{pattern}$ is included in the ESI.

1 J. D. Dunitz, *Chem Commun.*, 2003, **5**, 545.

2 W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, W. T. M. Mooij, S. L. Price, H. Scheraga, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer and D. E. Williams, *Acta. Crystallogr., Sect. B.*, 2002, **58**, 647.

3 F. H. Allen, *Acta. Crystallogr., Sect. B.*, 2002, **58**, 380.

4 I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta. Crystallogr., Sect. B.*, 2002, **58**, 389.

5 W. D. S. Motherwell, G. P. Shields and F. H. Allen, *Acta. Crystallogr., Sect. B.*, 1999, **55**, 1044.

6 A. F. Wells, *Structural Inorganic Chemistry*. Clarendon Press, Oxford. 5th Ed. 1984, p 143.

7 J. Burgess, J. Fawcett, D. R. Russell, S. R. Gilani and V. Palma, *Acta. Crystallogr., Sect. C.*, 1999, **55**, 1707.

8 I. V. Vystorop, K. A. Lyssenko and R. G. Kostyanovsky, *Mendeleev. Commun.*, 2001, **2**, 49.

9 A. I. Kitaigorodskii, *Organic Chemical Crystallography.*, Consultant's Bureau, New York, 1961.