

Distribution of Molecular Centres in Crystallographic Unit Cells

W. D. S. MOTHERWELL

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England. E-mail: motherwell@ccdc.cam.ac.uk

(Received 4 October 1996; accepted 8 April 1997)

Abstract

A survey has been made of the position of molecular centres in unit cells for the seven most populated space groups in the Cambridge Structural Database. Visual inspection of histograms and scattergrams reveals very clear peaks and clusters, showing preferred locations for molecules. These preferred positions are midway between centres of symmetry or screw axes. Some simple calculations of interaction energies of molecules with neighbours in a molecular coordination sphere suggest reasons for these preferred positions. These diagrams show that simplification of molecules to spheres is helpful in visualizing the prevalence of layers of molecules and tendencies to the ideal close packing of spheres.

1. Introduction

The archive of molecular crystal structures in the Cambridge Structural Database (CSD) contains, in some form, the answer to the question 'Can we predict crystal packing?'. It is a reasonable hope that as the size of the CSD increases there will be more data for comparison of similar packing structures, which can lead to the formulation of rules based on probabilities of occurrence of particular packing patterns, given certain molecular attributes.

Many attempts have been made to find methods of prediction by studying the packing arrangements of molecules of similar shape, or with similar chemical functional groups, producing some general correlations in certain limited classes of molecules (reviewed in Desiraju, 1989). However, the frustration arises that once some indication of prediction on the basis of molecular shape or chemical functional group is promising, there is found a case which seems to break the rule by forming a completely different packing arrangement.

It is also now accepted by several independent workers that many plausible close-packed crystal structures can be found by calculation for a given molecular structure, all with packing energies within 5 kJ mol^{-1} of the global minimum (see, for example, Chaka, Zaniewski, Youngs, Tessier & Klopman, 1996). The observed crystal structure is often the one found to coincide with the

global minimum, but not always. This observation has been noted both by workers who have exhaustively explored packing parameter space (Williams, 1969; van Eijck, Mooij & Kroon, 1995) or have used Monte Carlo methods plus simulated annealing to take statistical samples of the total space and proceed to a local minimum (Karfunkel & Gdanitz, 1992).

What is certain is the importance of the principle of close packing. The present study takes the approach of looking at the total available data on packing of molecular crystals to identify certain tendencies in packing and certain commonly occurring patterns. If such patterns exist then these will have implications as regards the methods of prediction, if only on the basis of probability of occurrence.

2. Generation of molecular centres

It was decided to begin by investigating subsets of the CSD (Allen & Kennard, 1993a,b) by selection of molecules within the more common space groups, since a complete survey of all 160,000 structures is out of the question as a first study. In the current CSD (October 1996) 82% of structures are accounted for by the seven most populated space groups. These are $P\bar{1}$ (19%), $P2_1$ (6%), $P2_1/c$ (35%), $C2/c$ (7%), $P2_12_12_1$ (9%), $Pbca$ (4%) and $Pnma$ (2%).

The analysis reported here has been restricted to structures with one molecule per asymmetric unit, with no disorder, no polymeric structure and the CSD status of a perfect match between chemical and crystal connectivities. The database search program *Quest* was used to select structures in a given space-group setting and space-group number, e.g. $P2_1/c$, no. 14, with the number of chemical residues (NRES) equal to 1, and a Z' value of 1.00 (the ratio of the number of formula units per unit cell to the Z value). An exception was made for $Pnma$, where all molecules are found to have a mirror plane coincident with the space-group mirror plane; for this study a selection was made with $Z' = \frac{1}{2}$.

The calculation of the coordinates of the molecular centres proceeded as follows; the centre is defined as the mean value of the x, y, z fractional coordinates for all reported atoms, including hydrogen if present.

This centre is formally different from the centre of mass and was felt to be more appropriate for studying molecular packing in space-filling terms. Since space-group symmetry allows the author of a publication a certain amount of arbitrary positioning of the molecules giving the same observed structure-factor magnitudes, a simple normalization procedure was introduced to allow comparison of structures on a standard location of a reference molecule.

In the space groups surveyed in this paper the given molecule may be shifted by $\frac{1}{2}$ in any axial direction, and in any combination of such shifts, and still give the same R factor. Furthermore, replacement of molecule coordinates by their inverse at $(-x, -y, -z)$ will also have no effect on R . Consider a molecule with a centre at $P1$ in Fig. 1; this molecule could equally well have been chosen to lie at points $P2$, $P3$ or $P4$. Thus, it is permissible to move the coordinates of any given centre $P1$ to the position $P2$ in the sector with axial range $0-\frac{1}{2}$, by shifts of $\frac{1}{2}$, and also to generate the equivalent location $P4$ in the same sector (if $P2$ has coordinates x, y, z , then $P4$ has coordinates $\frac{1}{2} - x, \frac{1}{2} - y, \frac{1}{2} - z$). The true asymmetric unit of the distribution of molecular centres is thus a space $\frac{1}{16}$ of the cell volume, in a range such as $x = 0-\frac{1}{2}, y = 0-\frac{1}{4}, z = 0-\frac{1}{2}$. The distribution has a centre of symmetry at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. For ease of visualization of the distribution related to the space-group operators it was decided to display both points $P2$ and $P4$ in the axial range $0-\frac{1}{2}$. To illustrate this procedure, in Fig. 2(a) we see the raw data positions of centres as given by authors in the range $0-1$ for the space group $P\bar{1}$ and one can see

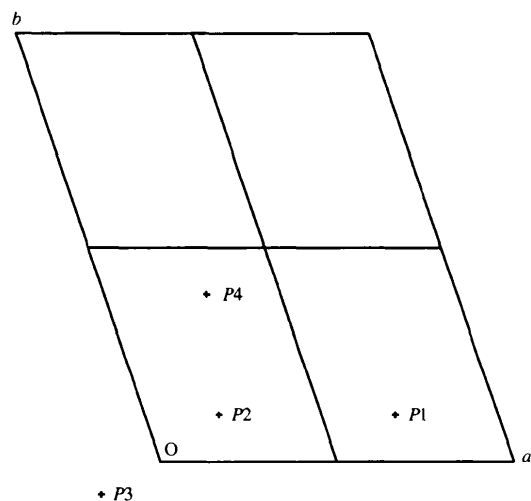


Fig. 1. Schematic diagram of a unit cell, showing some equivalent positions of molecular centres for the space groups named in this paper. $P1$ is a molecular centre position using the published coordinates. $P2$ is this centre moved by shifts of $\frac{1}{2}$ to bring into the sector $0-\frac{1}{2}$ on all axes. $P3$ is the inverse position obtained by the operation $-x, -y, -z$ on $P2$. $P4$ is the position $P3$ moved by shifts of $\frac{1}{2}$ to bring it into the reference sector $0-\frac{1}{2}$. The operator to transform $P2$ to $P4$ is $1/2 - x, 1/2 - y, 1/2 - z$.

the approximate centre of symmetry at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$. In Fig. 2(b) we see the plot of normalized molecular centres in the sector $0-\frac{1}{2}$, projected down each axis.

There are of course other arbitrary choices affecting the molecular centre position, notably the choice of cell axes, e.g. a, b and c in space group $P\bar{1}$ or a and c in $P2_1$. No attempt has been made in this paper to transform the coordinates to the standard reduced cell, which is planned in future work. Therefore, in these plots there is still a certain amount of randomness due to the authors' choice of axes, which one would expect not to interfere with any overall patterns observed.

It was decided to further simplify the selection by accepting only those structures reported in the normal settings, e.g. $P2_1/c$ rather than $P2_1/a, P2_1/n$ etc. In the case of some space groups it was further necessary to keep the number of database structures within a reasonable limit of 1000 entries for the *Vista* program by selecting from a reduced database. The order of the entries in the CSD is certainly random as regards structural type and no obvious visual difference was found in the analysis of histograms and scattergrams for several samples of 1000 entries selected by taking different orders of reference codes or starting at different points in the database. In order to allow reproducibility of the results all the samples here were made from the October 1996 CSD, using the first 1000 encountered in a search on the normal database. A small number of structures were removed from some samples because, although there was only one chemical residue, there were in fact two independent molecules in the asymmetric unit, each sharing molecular symmetry with a space-group centre of symmetry. A small number of other entries were removed because of some coding error in the CSD, e.g. polymeric structures not correctly flagged.

Some thought was given to the method of presentation of the distributions. Three-dimensional stereoplots of the coordinates were made, which certainly enable computer interactive viewing of clusters of points, but are rather less informative on the printed page in a single orientation. For this paper it was felt better to present projection scattergrams. At this stage of the analysis the purpose is to identify important overall features of distributions and to look for explanations in terms of underlying physical principles.

The computation was carried out by modification of the standard CSD *Pluto* program (Allen & Kennard, 1993b), to output tables of the coordinates, molecular dimensions, cell dimensions and orientation of the principal axes of inertia of the reference molecule, which were then used as input parameters to the statistical display program *Vista*.

2.1. $P\bar{1}$

The sample size was 967 reference codes. The normalized distribution, Fig. 2(b), shows a distinct pattern

with the clustering of centres in planes defined by $x_c = \frac{1}{4}$, $y_c = \frac{1}{4}$ and $z_c = \frac{1}{4}$. The significance of the distribution can be judged by the well defined peak at $\frac{1}{4}$ in the histograms for each axis, Fig. 2(c). As expected, in this simple space group the plots are very similar for each equivalent axial projection. There are maxima at positions $(x_c, y_c) = (\frac{1}{4}, \frac{1}{4})$ and similarly for (x_c, z_c) and (y_c, z_c) . There is significantly low density in the population in the regions $(x_c, y_c) = (0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$ and similarly for

(x_c, z_c) and (y_c, z_c) . These areas of low population correspond to the positions of the only symmetry elements in this space group, namely the centres of symmetry. Thus, these scattergrams are a graphic illustration of the principle of closest packing, since we have chosen molecules which have no centre of symmetry so they cannot occupy the centre of symmetry sites in the crystal. The reason why molecules will tend to avoid the centres is obvious when one considers locating a molecule

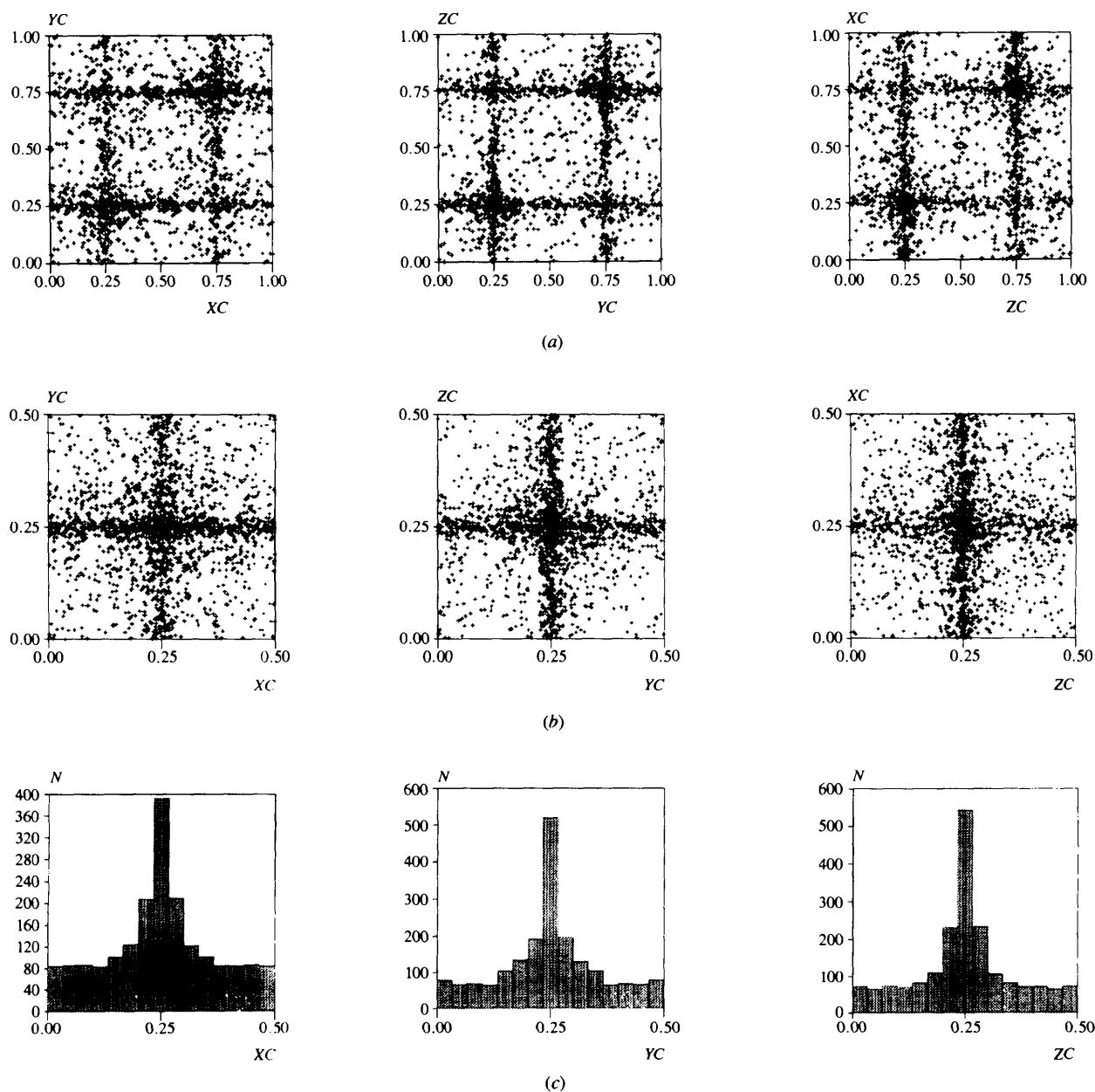


Fig. 2. Distributions for the sample in $P\bar{1}$. Distributions of molecular centres (x_c, y_c, z_c) as fractional coordinates. (a) Scattergrams for the unaltered published coordinates in the range 0-1; (b) scattergrams for x_c, y_c, z_c adjusted to bring the reference molecule into the range 0- $\frac{1}{2}$ on all three axes (see text and Fig. 1); (c) histograms for the adjusted x_c, y_c, z_c .

between any two centres of symmetry separated by a distance D , imagining for the present that it is replaced by just a simple sphere of diameter D . In order for such spheres to be touching the symmetry-generated spheres on either side, we conclude that the best location is halfway between the centres of symmetry, forming a linear arrangement. However, the situation with real molecules is that molecular shapes are far from spherical and molecules have chemical groups which give specific directional interactions, notably hydrogen bonding. The surprise is that there seems to be some principle of balance which causes the molecules to tend to behave as if they were simple spheres, preferring the midpoints between the centres of symmetry at coordinates of $1/4$, on one or more of the axes.

2.2. $P2_1$

The sample size was 1000 reference codes. The origin choice is arbitrary for the y axis in this space group, so the distribution is simplified to consideration of the (xc, zc) scattergram (Fig. 3) and the distributions along the two equivalent a and c axes are very similar, as can be seen from the histograms. The distribution shows a strong cluster of points at $(xc, zc) = (\frac{1}{4}, \frac{1}{4})$ and, to some extent, there is clustering along the lines at $xc = \frac{1}{4}$ and $zc = \frac{1}{4}$. There are very obvious voids in the regions

$(xc, zc) = (0,0), (1, \frac{1}{2}), (\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$; these are exactly the locations of the screw axes. This suggests a principle that molecules prefer to cluster midway between screw axes.

2.3. $P2_1/c$

The sample size was 994 reference codes after elimination of some cases of two molecules per asymmetric unit. The distribution (Fig. 4) histograms show one strong peak for $xc = \frac{1}{4}$. The yc histogram is more random, but discernable peaks occur at $yc = 0, \frac{1}{8}, \frac{3}{8}$ and $\frac{1}{2}$. The zc histogram has peaks at $\frac{1}{8}, \frac{1}{4}$ and $\frac{3}{8}$. The main feature of the scattergrams is a band of points in the plane defined by $xc = \frac{1}{4}$. It is noticeable in the (xc, zc) scattergram that there are areas of low population at $(0,0), (0, \frac{1}{2}), (\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$, which are the positions of the centres of symmetry. The positions of the screw axes in projection are at $(xc, zc) = (0, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{4})$. These regions show some population on the scattergram, but very much less than the strong bands of points at $xc = \frac{1}{4}$. It is noted that this band is approximately continuous in density and the direction of the glide plane is along the z axis, which would agree with the physical interpretation that when pairs of molecules are in contact due to a glide operator, their position in the direction of the glide has no effect on the intermolecular distances.

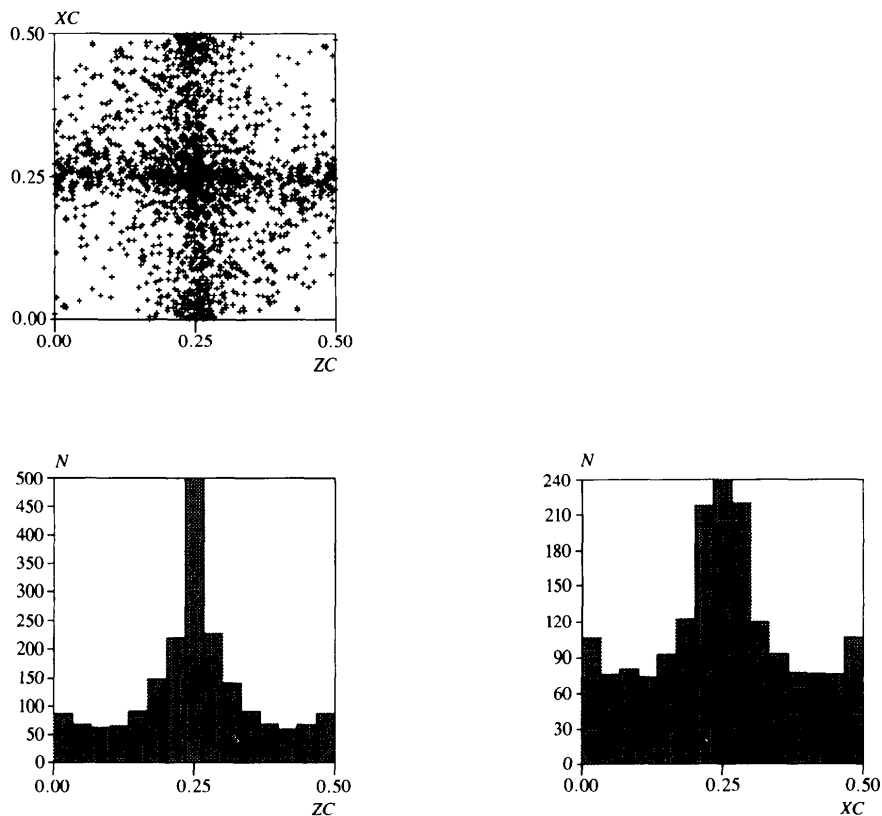


Fig. 3. Histograms and scattergrams for the sample in $P2_1$. Distributions of molecular centres xc, yc, zc .

2.4. $C2/c$

The sample for the setting $C2/c$ consists of 984 reference codes. The distribution histograms show remarkably sharp peaks for xc at $\frac{1}{8}$ and $\frac{3}{8}$ (Fig. 5). The yc distribution is more random, but has sharp peaks at $yc = \frac{1}{4}$. The zc distribution has distinct peaks at positions $\frac{1}{8}$ and $\frac{3}{8}$, similar to the xc histogram. The scattergrams show sharp bands of density in the planes $xc = \frac{1}{8}$ and $\frac{3}{8}$. It may be noted in the (xc, zc) projection that there is visible density extending in the direction of the c glide plane, as would be possible for molecules interacting by the c -glide operator. The scattergram for (xc, zc) shows particularly large voids in the regions $(xc, zc) = (0, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{4})$, which correspond to the positions of the twofold rotation axes. This is an expected result, as molecules cannot pack favourably (bumps into hollows) as they approach such axes. There are low populations in the regions corresponding to the centres of symmetry in projection at $(xc, zc) = (0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{4}, 0)$ and $(\frac{1}{4}, \frac{1}{2})$.

2.5. $P2_12_12_1$

The sample for this setting is 999 reference codes, one structure being eliminated as wrongly processed in CSD. The histograms (Fig. 6) show a much more random distribution than seen in the above space groups.

Examination of the histograms at a finer bin size (not shown here) reveals some structure with small peaks at $xc = n/8$ and $yc = 0, \frac{1}{8}, \frac{3}{8}, \frac{1}{2}$. The zc distribution shows clear peaks at $zc = \frac{1}{8}, \frac{3}{8}$. Since these axes are equivalent in terms of space-group symmetry, it is unexpected that there should be apparently more structure in the zc histogram. A check on whether cells have the standard setting of $a < b < c$ revealed that only ca 40% are correct. However, examination of the reduced cell parameters shows very distinct mean lengths for the three axes (8.3, 12.3 and 22.3 Å), indicating that we could expect differences in the types of intermolecular contacts in the three axial directions. The scattergrams do not show any outstanding pattern, but it is perhaps worth noting a suggestion of lower populations in the regions of the screw axes, e.g. scattergram (xc, yc) has low density at $(xc, yc) = (0, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{4})$, which are the locations of the screw axes parallel to c . This data may well show sharper patterns if reorganized according to the standard reduced cells.

2.6. $Pbca$

The sample consists of 999 reference codes. The histograms (Fig. 7) show very clear maxima on each axis at the positions $\frac{1}{8}$ and $\frac{3}{8}$. The scattergrams show a

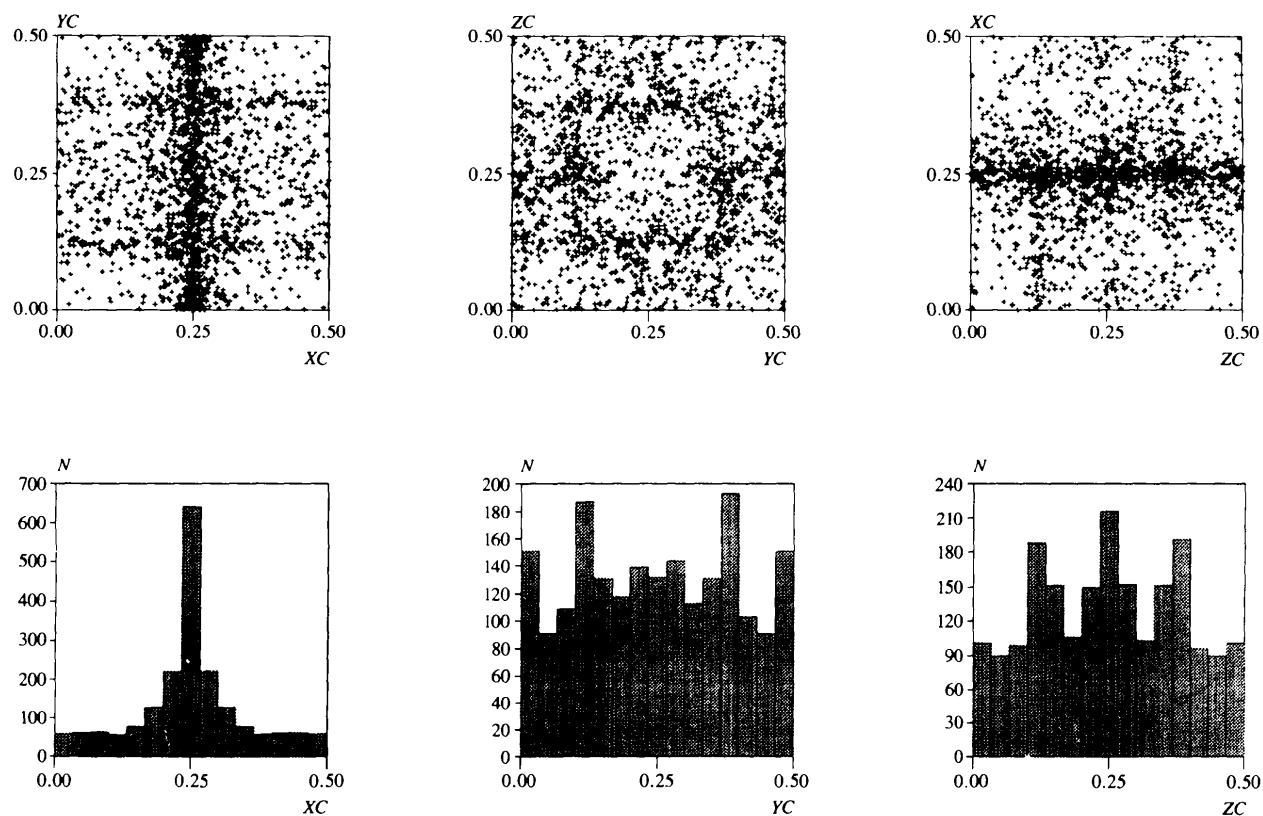


Fig. 4. Histograms and scattergrams for the sample in $P2_1/c$. Distributions of molecular centres xc, yc, zc .

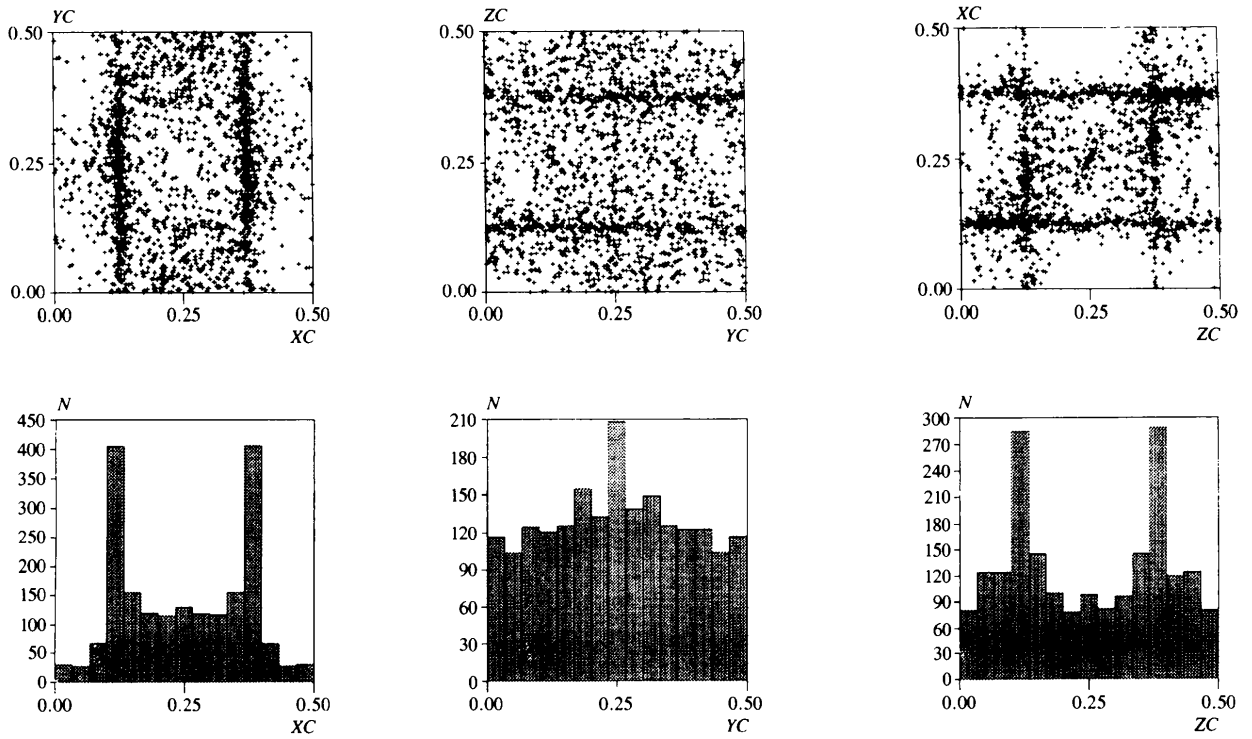


Fig. 5. Histograms and scattergrams for the sample in $C2/c$. Distributions of molecular centres x_c, y_c, z_c .

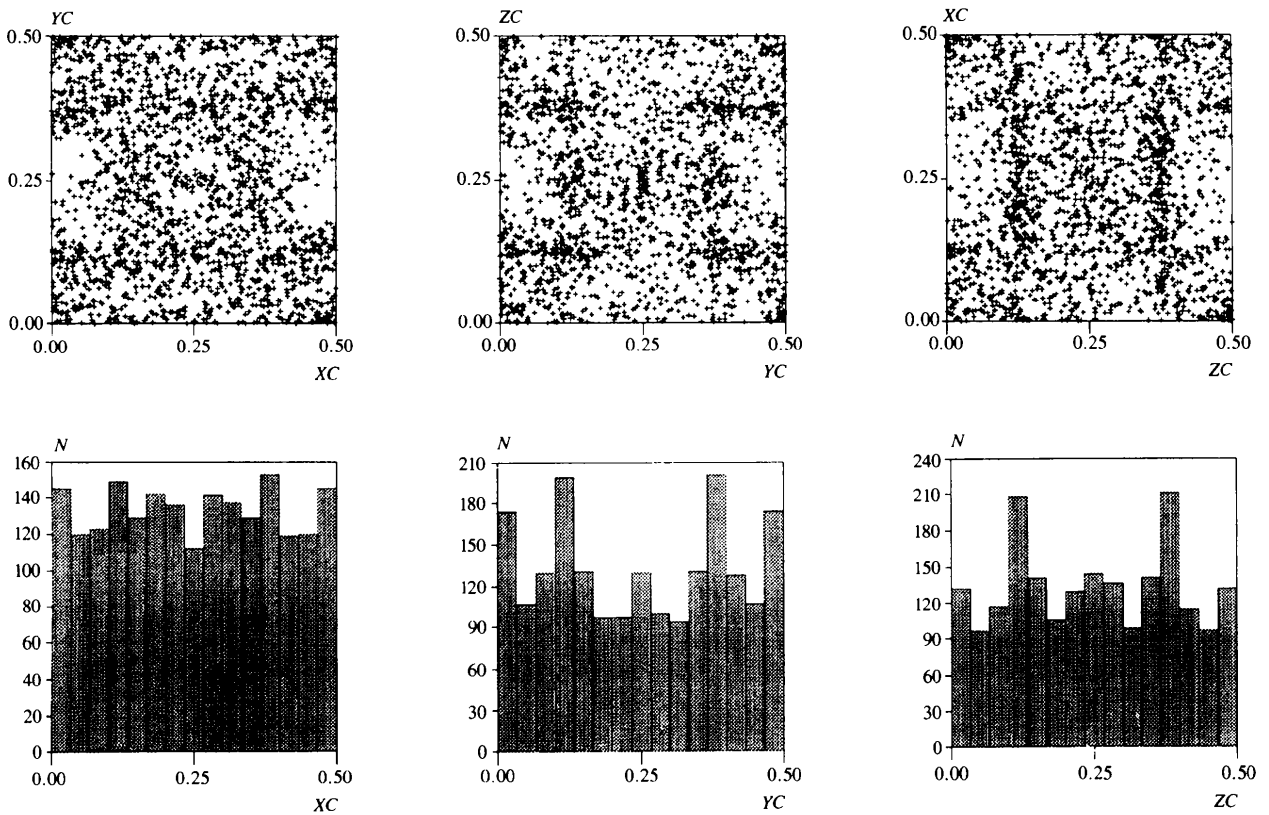


Fig. 6. Histograms and scattergrams for the sample in $P2_12_12_1$. Distributions of molecular centres x_c, y_c, z_c .

pattern of bands of density lying in planes at $\frac{1}{8}$ and $\frac{3}{8}$ on each axis. The lowest population areas correspond to the positions of centres of symmetry in projection at $(0,0)$, $(0,\frac{1}{2})$, $(\frac{1}{2},0)$ and $(\frac{1}{2},\frac{1}{2})$ and the positions of the screw axes, for example, see $(xc,zc) = (\frac{1}{4},0)$ and $(\frac{1}{4},\frac{1}{2})$. As in $P\bar{1}$ and $P2_12_12_1$ the three axial directions should be equivalent.

2.7. *Pnma*

The sample for the setting *Pnma* consists of 731 entries, which represents all the available structures from the CSD. Since all the observed molecules have a mirror plane coincident with the space-group mirror plane at $y = \frac{1}{4}$, the only significant distribution is the (xc,zc) scattergram (Fig. 8). This has clusters in the regions $(xc,zc) = (\frac{1}{4},0)$ and $(\frac{1}{4},\frac{1}{2})$, but very diffuse. Regions of low density are rather better defined; the low populations at $(0,0)$, $(\frac{1}{2},0)$, $(0,\frac{1}{2})$ and $(\frac{1}{2},\frac{1}{2})$ correspond to centres of symmetry. However, the lowest population appears to be in the region $(\frac{1}{4},\frac{1}{4})$, which corresponds to the line of intersection of the glide planes *n* and *a*. The low density near the planes defined by $xc = 0$ and $\frac{1}{2}$ is also very pronounced, but does not correspond to any space-group operator.

3. Discussion

It is emphasized that the data samples cited above are in no way restricted by chemical atom type, so the patterns persist despite the frequent presence of hydrogen bonding, electrostatic interactions and indeed approximately 45% of the molecules are organometallic complexes. A subset of the $P2_1/c$ sample was made for hydrocarbons only (126 refcodes) and no significant difference could be seen in the distribution patterns.

In order to investigate the distribution patterns an extension was written for the *Pluto* program, to display the molecular coordination sphere (Kitaigorodski, 1973; Gavezzotti, 1990) about the given reference molecule in the CSD. A routine was written to calculate approximate packing potential energy, by summing molecular interactions using the atom-pair empirical potentials as collected by Gavezzotti (1994), which do not use any electrostatic terms. In order to elucidate the possible energetic reasons for the patterns, two examples are presented here from $P\bar{1}$ and $P2_1$, without hydrogen-bonding complications.

The first example, CSD reference code ABZNPS (Fig. 9) is deliberately chosen as asymmetric, with a variety of atom types. This is in space group $P\bar{1}$ with only

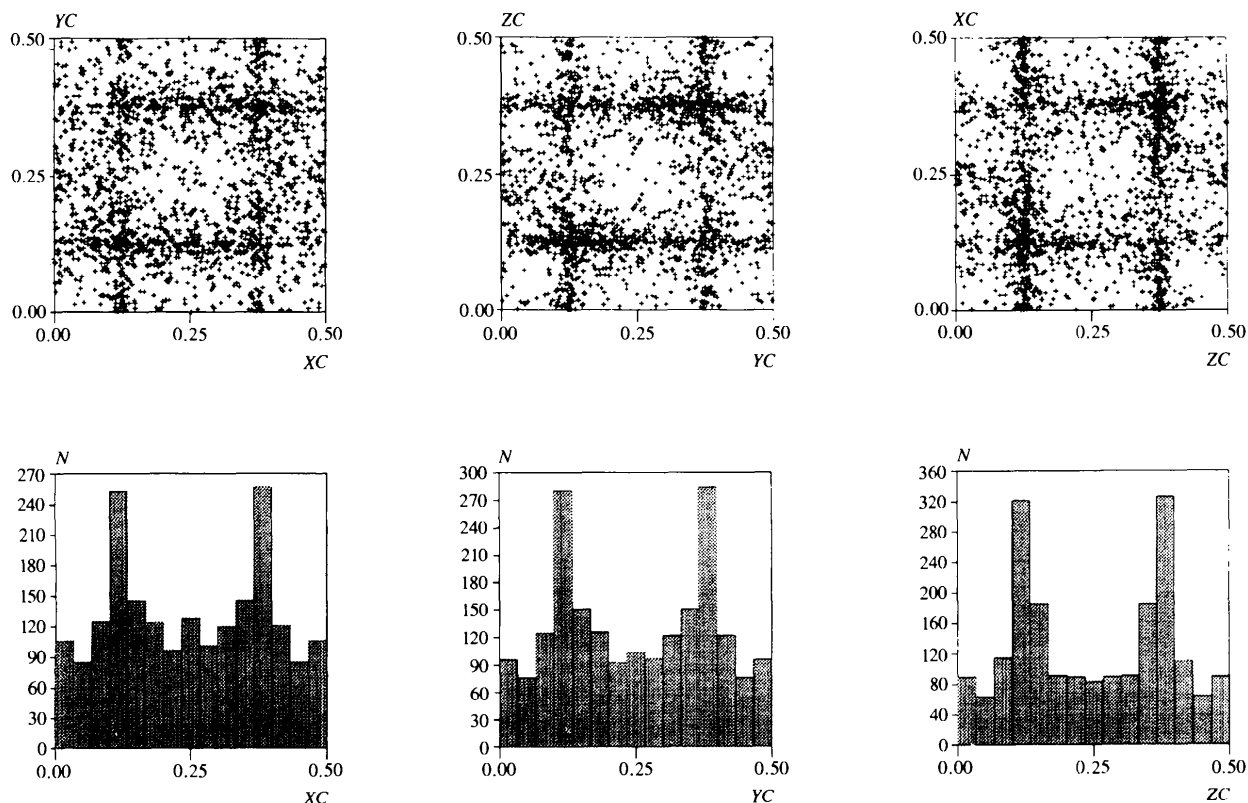


Fig. 7. Histograms and scattergrams for the sample in *Pbc1*. Distributions of molecular centres xc,yc,zc .

centres of symmetry and lattice translation operators available to generate the neighbouring molecules. The molecular coordination sphere is defined on the basis of energy of interactions, including only molecules with interactions stronger than a given threshold. The number of molecules N surrounding the reference molecule in the CSD is typically 12–14 using a default threshold of $-4.18 \text{ kJ mol}^{-1}$. The default coordination sphere for ABZNPS with $N = 13$ is shown in Fig. 9(c), displayed in projection down c . It is difficult to visualize the relative importance of interactions in this sort of diagram, even with on-screen rotation and stereoviews, so a feature was written to allow the display of each molecule as a simple sphere. The sphere model for the same coordination sphere is shown in Fig. 9(d), with slight rotation away from the exact c axis. This is immediately reminiscent of a hexagonal close-packed structure (h.c.p.), albeit with distortions, interactions varying from -13 to -41 kJ mol^{-1} . If we select only the two strongest interactions we find these are due to centres of symmetry and the arrangement is a chain along the abc body diagonal. Note the approximate balancing of strong interaction energy through different centres of symmetry ($-34, -41 \text{ kJ mol}^{-1}$) and how the molecule is thus placed at $(\frac{1}{4}, y, \frac{1}{4})$, halfway between the centres of symmetry.

Looking at the coordination sphere in the projection down a we can see hexagonal packed layers parallel to (010) at $y = \frac{1}{4}$ and also a layer in the plane (01 $\bar{1}$). If we do a simple summation of the energies of interaction in these three layers, E_L , we have for the (110) plane $E_L = -152$ (24, 25, 31, 31, 15, 26 kJ mol^{-1}), the plane (010) $E_L = -122$ (13, 17, 31, 31, 17, 13 kJ mol^{-1}) and the plane (01 $\bar{1}$) $E_L = -159$ (41, 24, 17, 17, 26, 34 kJ mol^{-1}). This shows how the energy of binding within layers is approximately equal, like the ideal h.c.p. structure. In this case it is appropriate to describe the structure according to h.c.p. stacked layers parallel to these planes. Note the symmetry of the layer parallel to (010), where all six molecules are generated by translation alone; the other layers contain both translation and centre-of-symmetry generated molecules. Examination of many hundreds of such coordination spheres has led to the conclusion that many structures seem to form h.c.p. layers. The situation is, however, far from simple and there are structures where molecules are grouped into dimers by one predominantly strong interaction or into columns with a strong interaction. However, a principle is suggested that molecules tend to form a coordination sphere that is as far as possible an approximation to hexagonal or cubic close packing (c.c.p.) of spheres. The

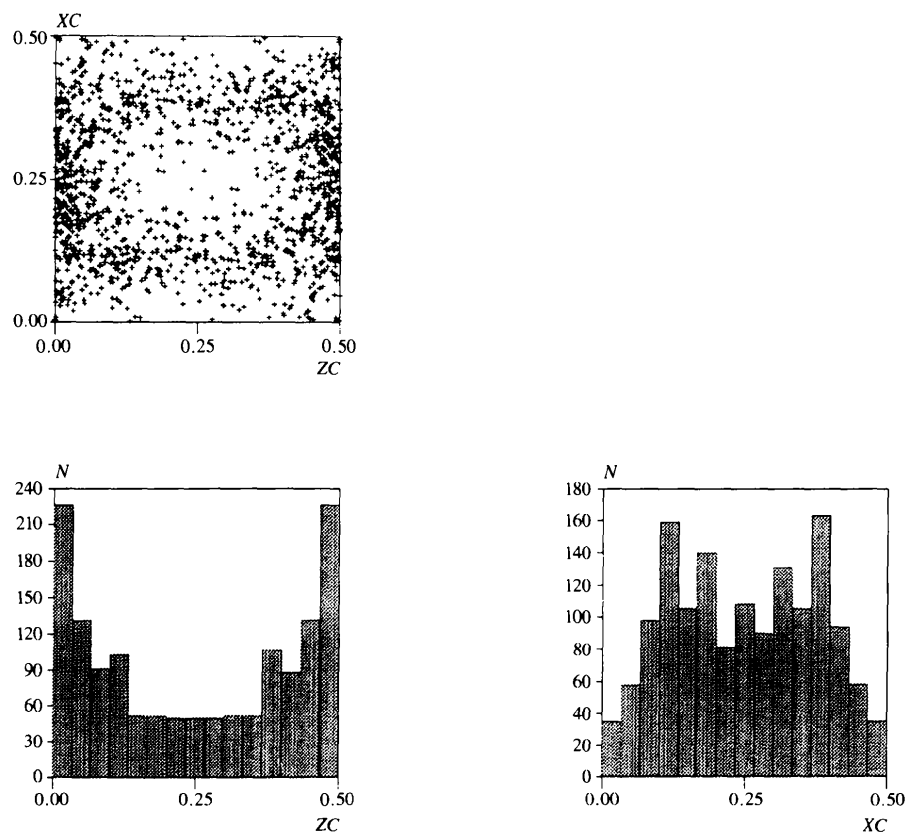


Fig. 8. Scattergrams for the sample in $Pnma$. Distribution of molecular centres x_c, z_c , all molecules lying at $y_c = \frac{1}{4}$.

coordination spheres are found to fit together in a great variety of ways in networks to form the full space-group symmetry.

A second example, ACDXYL10 (Fig. 10), is taken from space group $P2_1$ to show a coordination sphere

involving screw axes. The selection of the top six interactions shows a layer at $z = \frac{1}{2}$ in Fig 10(c), simplified to the sphere model in Fig. 10(d). Note that in order to achieve a balance between the interactions due to the screw axes ($-42, -40 \text{ kJ mol}^{-1}$), the position of the

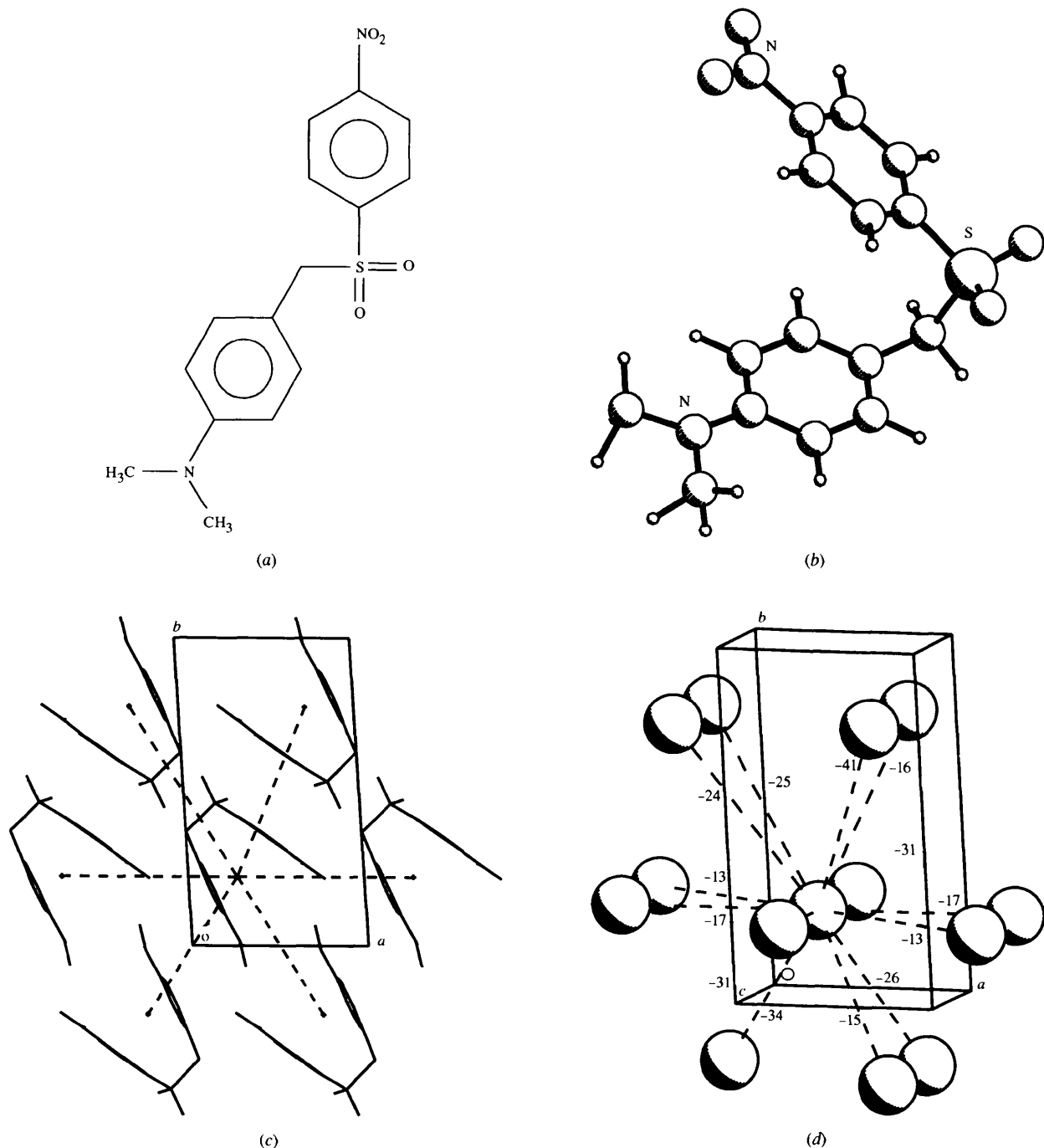


Fig. 9. Example molecule in $P\bar{1}$, CSD reference code ABZNPS. (a) Chemical diagram; (b) molecular structure; (c) molecular coordination sphere for 13 molecules, selected with interaction energy $< -4.18 \text{ kJ mol}^{-1}$, projected down c . Small crosses mark the molecular centres, dashed lines indicate the molecular interactions; (d) another view with the molecules represented as spheres, with interaction energies in kJ mol^{-1} .

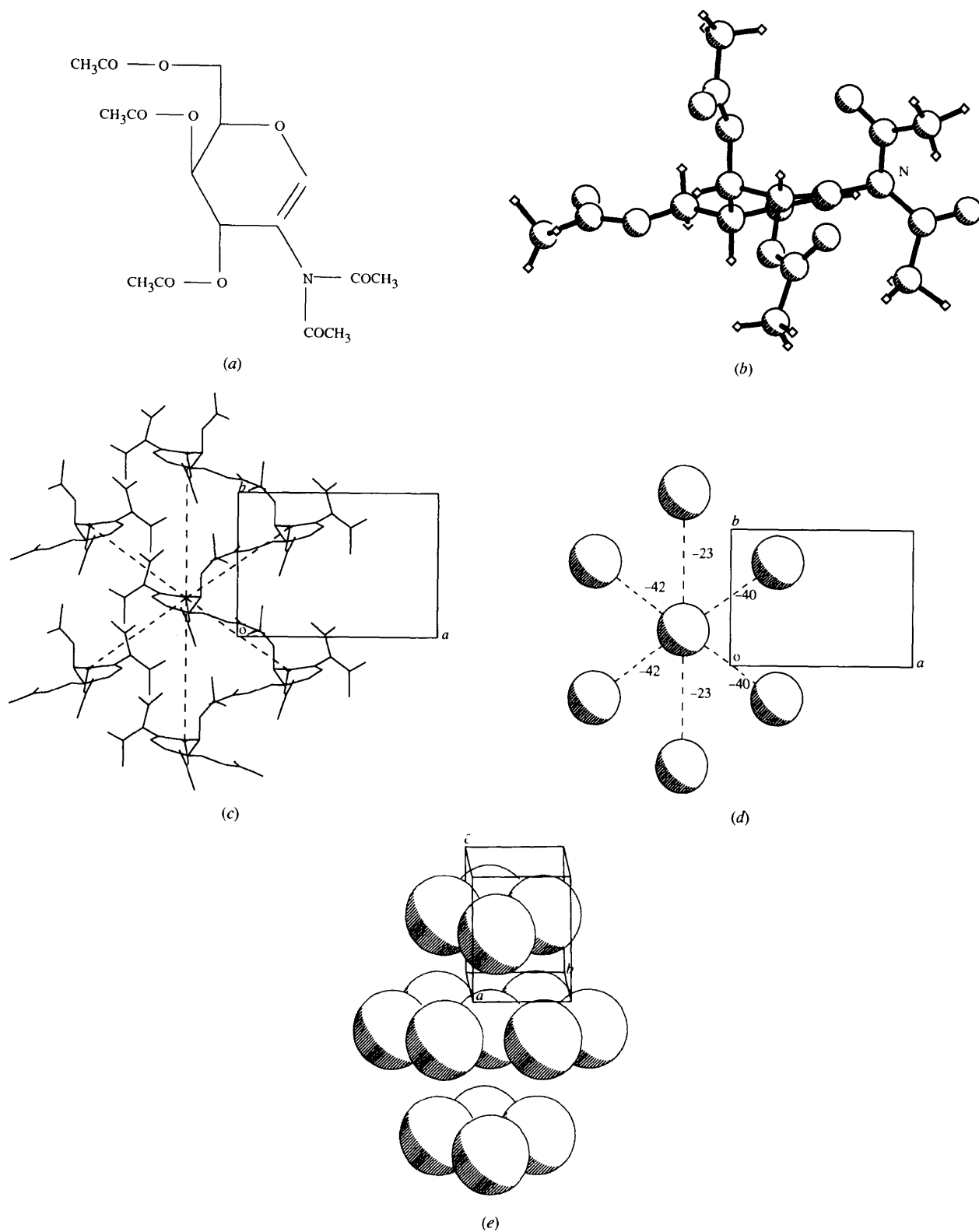


Fig. 10. Example molecule in $P2_1$, CSD reference code ACDXYL10. (a) Chemical diagram; (b) molecular structure; (c) display of the strongest six interactions, lying in the plane at $z = \frac{1}{2}$, energies in kJ mol^{-1} ; (d) the same six molecules simplified to spheres; (e) view of molecular coordination sphere for 14 neighbours with energy $< -4.18 \text{ kJ mol}^{-1}$.

molecular centre is at $xc = \frac{1}{4}$. The layer symmetry is formed by four molecules generated by screw axes and two by translation along b . A view of the $N = 14$ coordination sphere, for all interactions $< -4.18 \text{ kJ mol}^{-1}$ (Fig. 10e), again suggests a way of looking at complex packing structures as approximations to h.c.p or c.c.p. of spheres.

Examination of coordination spheres in $P2_1/c$ shows a wide variety of types of packing. There are many $N = 14$ examples where there is an approximation to ideal h.c.p., but here there are now four types of operator involved (T = translation, I = centre of symmetry, S = screw axis and G = glide plane, following the notation of Gavezzotti, 1991). This can explain the fact that this space group is by far the most commonly observed for organic molecules, as there is a higher probability that a close-packed $N = 14$ sphere can be formed with more operators than with the restriction of only T , I in $P\bar{1}$ or T , S in $P2_1$. Since the position of a molecule on the axis in the direction of the glide operator does not have any effect on the energy of interaction of a pair of molecules related by the glide plane, we have an explanation of the elongation of the heavily populated positions into bands in the c direction (Fig. 4).

Discussion of the other space groups must be shortened in this first study of the topic. However, the occurrence of close-packed layers is frequent in $P2_12_12_1$, $C2/c$, $Pbca$ and $Pnma$. The arrangements in $P2_12_12_1$, where the only operators available for forming a coordination sphere are T and S , seem to be easy to form pseudo-hexagonal close-packed layers regardless of the position of the centre on the cell axes.

It can be envisaged that with further study these distributions, taking into account transformations to standard reduced cell settings, could be used to assist structure prediction programs. In methods based on random exploration of parameter space a certain probability could be attached to the position of the molecular centre, according to the observed frequency from the CSD data,

and thus speed up search times. In cases where calculated packing energies are very similar, for a set of possible polymorphs it may be possible to use these distributions-of-centres plots to rank the structures by likelihood of occurrence.

Finally, a somewhat negative result. In the hope that some obvious correlation might be found with molecular shape or symmetry, an inspection of many molecules selected from the most heavily populated areas was carried out. Other parameters available were the molecular dimensions as an enclosing box based on the cell axes and the principal axes of inertia. No immediate conclusions could be reached and it is hoped that readers may be encouraged to explore such clusters of molecules themselves and detect some similarity of attributes.

The author records his thanks to the many thousands of contributors to the CSD and to Dr R. Gould for helpful discussions.

References

- Allen, F. H. & Kennard, O. (1993a). *Chem. Des. Autom. News*, **8**, 1.
Allen, F. H. & Kennard, O. (1993b). *Chem. Des. Autom. News*, **8**, 31–37.
Chaka, A. M., Zaniwski, R., Youngs, W., Tessier, C. & Klopman, G. (1996). *Acta Cryst.* **B52**, 165–183.
Desiraju, G. (1989). *Crystal Engineering*. Amsterdam: Elsevier.
Eijck, B. P. van, Mooij, W. T. M. & Kroon, J. (1995). **B51**, 99–103.
Gavezzotti, A. (1990). *Acta Cryst.* **B46**, 275–283.
Gavezzotti, A. (1991). *J. Am. Chem. Soc.* **113**, 4622–4629.
Gavezzotti, A. (1994). *Acc. Chem. Res.* **27**, 309–314.
Karfunkel, H. R. & Gdanitz, R. (1992). *J. Comput. Chem.* **13**, 1171–1183.
Kitaigorodski, A. I. (1973). *Molecular Crystals and Molecules*. New York: Academic Press.
Williams, D. E. (1969). *Acta Cryst.* **A25**, 464–470.