Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Distribution of molecular pairs in $Z' = 2$ structures

Anna Collins

Chemistry, University of Oxford, Mansfield Road, Oxford OX2 3TA, England

Correspondence e-mail: anna.collins@chem.ox.ac.uk

Received 24 March 2006
Accepted 29 June 2006

The positions of pairs of independent molecules in $Z' = 2$ structures have been surveyed for six of the most populated space groups for that class of structure. These results have been compared with the $Z' = 1$ situation to reveal whether there are any fundamental differences in the construction of the asymmetric units in these cases. The results indicate that, broadly speaking, the packing of the molecular pairs in $Z' = 2$ structures resembles that of single molecules in structures with $Z' = 1$; this similarity may be chiefly attributed to the constraints imposed by the symmetry operators of the space group. However, there are key differences, which are particularly marked in the space groups with higher symmetry, that indicate that the asymmetric units in $Z' = 1$ and $Z' = 2$ structures are not directly comparable. In those cases where the positions of the pair centroids in $Z' = 2$ structures are similar to the positions of molecular centroids for $Z' = 1$ structures, it follows that the molecular centroids in $Z' = 2$ structures must follow a different distribution. A different pattern is produced if the independent molecules in $Z' = 2$ structures behave like the individual molecules in $Z' = 1$ structures. These two scenarios combine to form the observed distributions of pair centroid positions.

1. Introduction

Bond lengths, angles and to some extent even torsion angles in organic compounds can now be predicted with reasonable accuracy, except in the case of really exotic materials (Gavezzotti & Flack, 2005; Gillespie *et al.*, 1998). However, the laws governing the packing of molecules in a crystal structure are still not completely understood. Although advances have been made in the calculation of packing energies, the limited success of such calculations for structure prediction under blind test conditions (*e.g.* Motherwell *et al.*, 2002) indicates that additional insight is needed. Most structure modelling concerns only structures where $Z' = 1$; molecular structures with $Z' > 1$ result in additional degrees of complexity (Day *et al.*, 2005). However, multiple instances of the same molecule in slightly different environments in the same crystal may throw light onto the relationship between molecular conformation and molecular packing. Structures with pseudosymmetry (Lehmler *et al.*, 2004) or frustrated symmetry (Heine & Price, 1985; Koutentis *et al.*, 2001) may be a particularly rich source of packing information.

2. Crystal packing

Kitaigorodskii produced one of the earliest reviews of patterns in crystal packing (Kitaigorodskii, 1961; Wilson, 1993), but since then such studies have become legion (*e.g.* Brock, 1996;

Table 1
Frequencies for the ten most commonly occurring space groups of $Z' = 1$ and $Z' = 2$ structures in the CSD that contain one chemical residue.

Rank	Space-group number	Space-group name	Frequency	% frequency	Space-group number	Space-group name	Frequency	% frequency
1	14	$P2_1/c$	28 480	38.43	14	$P2_1/c$	2747	26.68
2	19	$P2_12_12_1$	13 489	18.20	2	$P\overline{1}$	2559	24.85
3	2	$P\overline{1}$	10 989	14.83	4	$P2_1$	1978	19.2
4	4	$P2_1$	7152	9.65	19	$P2_12_12_1$	925	6.98
5	61	*Pbca*	3658	4.94	1	*P*1	566	5.50
6	15	*C*2/*c*	3083	1.89	29	$Pca2_1$	251	2.44
7	33	$Pna2_1$	1397	1.06	33	$Pna2_1$	225	2.18
8	9	*Cc*	785	0.97	15	*C*2/*c*	207	2.02
9	5	*C*2	722	0.89	61	*Pbca*	183	1.78
10	29	$Pca2_1$	642	0.41	9	*Cc*	146	1.42

Brock & Dunitz, 1994; Dalhus & Görbitz, 2000; Filippini & Gavezzotti, 1992; Karthe *et al.*, 1993; Pidcock & Motherwell, 2004). Recent studies have put greater emphasis on structures with $Z' > 1$ [see Steed (2003) for a recent review]. Trying to rationalize molecular packing led Motherwell (1997) to survey the molecular positions for $Z' = 1$ structures in the most common space groups in the Cambridge Structural Database (CSD; Allen, 2002): $P\overline{1}$, $P2_1$, $P2_1/c$, $C2/c$, $P2_12_12_1$ and *Pbca*. Structure prediction computations generally yield a number of solutions with approximately the same energy and usually include the known polymorphs. It is widely accepted that the observed forms are selected from these equivalent-energy structures by moderating effects of entropy and kinetic processes. It was hoped that by studying packing arrangements in crystals, general rules could be derived which would initially provide an empirical discriminator for the degenerate predictions, and eventually lead to a proper understanding of the underlying phenomena.

The positions of local (non-crystallographic) centres of inversion in space groups $Pca2_1$ and $Pna2_1$ have also been investigated for structures with $Z' = 2$ (Marsh *et al.*, 1998). In this study, we have investigated the arrangement of pairs of molecules in the most common space groups for structures with $Z' = 2$, and thus there is overlap with those two papers: with the methodology of the former and with the results of the latter. However, the packing of pairs of independent molecules in structures with $Z' = 2$ cannot be assumed to be the same as for single molecules in $Z' = 1$ structures. In addition, we have not restricted the analysis to structures that possess local inversion centres (or indeed of any other specific type of non-crystallographic operator) or placed any restriction on the conformational similarity of the independent molecules.

2.1. Space-group frequencies

The space-group frequency distribution for $Z' = 2$ structures differs from that in the CSD for $Z' = 1$ structures (Table 1). The six most common space groups for $Z' = 2$, excluding *P*1, will be analysed here. The space group *P*1 (and other non-standard settings) is much more common for $Z' = 2$ structures, but unfortunately this space group can provide no information about patterns in crystal packing because there are no restrictions on the choice of origin. No restrictions are imposed on the distribution of molecules within the cell owing to the lack of symmetry elements other than the identity.

2.2. Packing patterns

Although in the final analysis individual atom–atom interactions will play a significant role in determining crystal packing, packing patterns can be visualized by representing whole molecules by a surface derived from their inertial tensor. Further simplification can be achieved by replacing the whole molecule by a point at the average of the atomic coordinates of its component atoms, which approximates to its centre of mass.

Unless a molecule contains a symmetry operator that is also used as a space-group symmetry element, the atoms in a molecular compound cannot lie on a special position. As a direct consequence, in $Z' = 1$ structures, molecular centres of gravity tend to cluster at points removed from symmetry operators (Motherwell, 1997). Similarly for $Z' = 2$ structures; each individual molecule will try to avoid symmetry operators, but because molecules do not interpenetrate, each independent molecule must also avoid the other independent molecule, posing additional restrictions on the positions of molecules in the asymmetric unit of $Z' = 2$ structures.

2.3. Definition of terms

The molecular centroid is here defined as the unweighted average of all the atomic coordinates in the molecule and thus must lie within the envelope of that molecule. For $Z' = 2$ structures, pair centroids are calculated as the average of all the atomic coordinates of a pair of molecules that are crystallographically independent. The position of the pair centroid may lie between the molecules or within the envelope of one of the molecules. The latter case is most likely to occur when the crystal is composed of molecules that are highly elongated or discoidal and where the long axes are not approximately parallel. This situation will also be dependent on space group.

The molecular centroid represents where molecules are located. The pair centroids represent where molecules are not located. It may therefore seem counterintuitive to attempt to make any comparisons of the distributions arising from these positions; however, this process allows the behaviour of asymmetric unit contents in the two classes of structure to be compared.

In a $Z' = 2$ structure in *P*1, the coordinates of the molecular centroids of the independent molecules, molecule 1 and molecule 2, are given by x, y, z and X, Y, Z, respectively. Within a given unit cell, there is only one choice for the pair of independent molecules. In space groups of higher symmetry, the positions of molecular centroids are those for all the general positions in the space group, and so the number of possible molecular centroids is equivalent to the site multiplicity of the space group, M. All possible pair centroids are obtained by combining these positions. The number of pair

centroids is thus given by M^2. These combinations are given in Appendix *A* for $P2_1/c$; possible pair centroids for all other space groups considered in this paper have been deposited as supplementary information.[1]

For all space groups, the pair centroid can be positioned within a range of 0–$\frac{1}{2}$. This arises simply from the translational symmetry of crystals. In $P1$, the molecular centroid in $Z' = 1$ structures must lie in the range 0–1, although in $Z' = 2$ structures the pair centroid in $P1$ can always be located in the 0–$\frac{1}{2}$ range; however, the position of the origin is totally arbitrary.

3. Results and discussion

Structures were selected from the CSD on the following search criteria: $Z' = 2$, one chemical unit (*i.e.* homomolecular structures only), three-dimensional coordinates determined, *R* factor < 0.1, not disordered, no errors, not polymeric, no ions, no powder structures and only organic materials. No checks were made for voids in the structures, which may have indicated the omission of solvent, although structures were checked to ensure that both molecules were present.

There are a large number of possible choices for the pair of molecules chosen as the asymmetric unit by the publishing author, but there is no guarantee that the choice of the asymmetric unit is 'best' (*e.g.* it may not illustrate a pseudo-symmetry relationship between molecules clearly). Thus, all the possible pair centroids were calculated and used in analysis.

Using a combination of scatterplots, three-dimensional frequency plots and histograms a complete picture of the distribution of pair centroids can be built up. Each graphical form has its advantages; the scatterplots are very good at showing which areas are sparsely populated by the centroid positions, but are poor at distinguishing between areas that are well populated. This is the advantage of the three-dimensional frequency plots, as they show which are the highly populated regions very well. In general, the projections allow for easy correlation between the positions of centroids and the positions of the symmetry elements of the space group. For a given space group, all possible centroid positions are shown in both the scatterplots and the three-dimensional frequency plots, *i.e.* each crystal structure in $P2_1$ will contribute four points on the scatterplot and a frequency of four to the three-dimensional frequency plot. The effect will be to enhance the areas that are symmetry-related. The histograms are useful for illuminating axis-related trends.

Each possible pair centroid has been used to generate separate data series, which are plotted in histograms that represent projections on to a single axis. This approach will help demonstrate whether there are any differences between the different possible centroid positions.

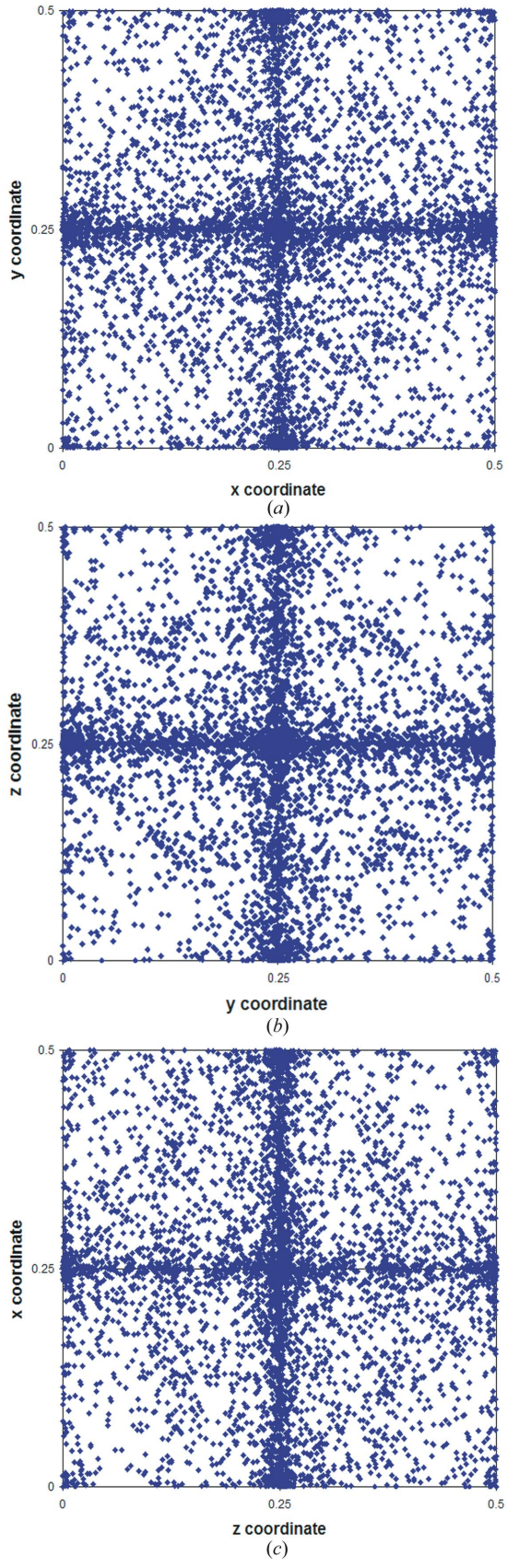


Figure 1
Scatterplots for the positions of pair centroids in $P\bar{1}$. (*a*) *xy* projection; (*b*) *yz* projection; (*c*) *zx* projection.

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: BS5032). Services for accessing these data are described at the back of the journal.

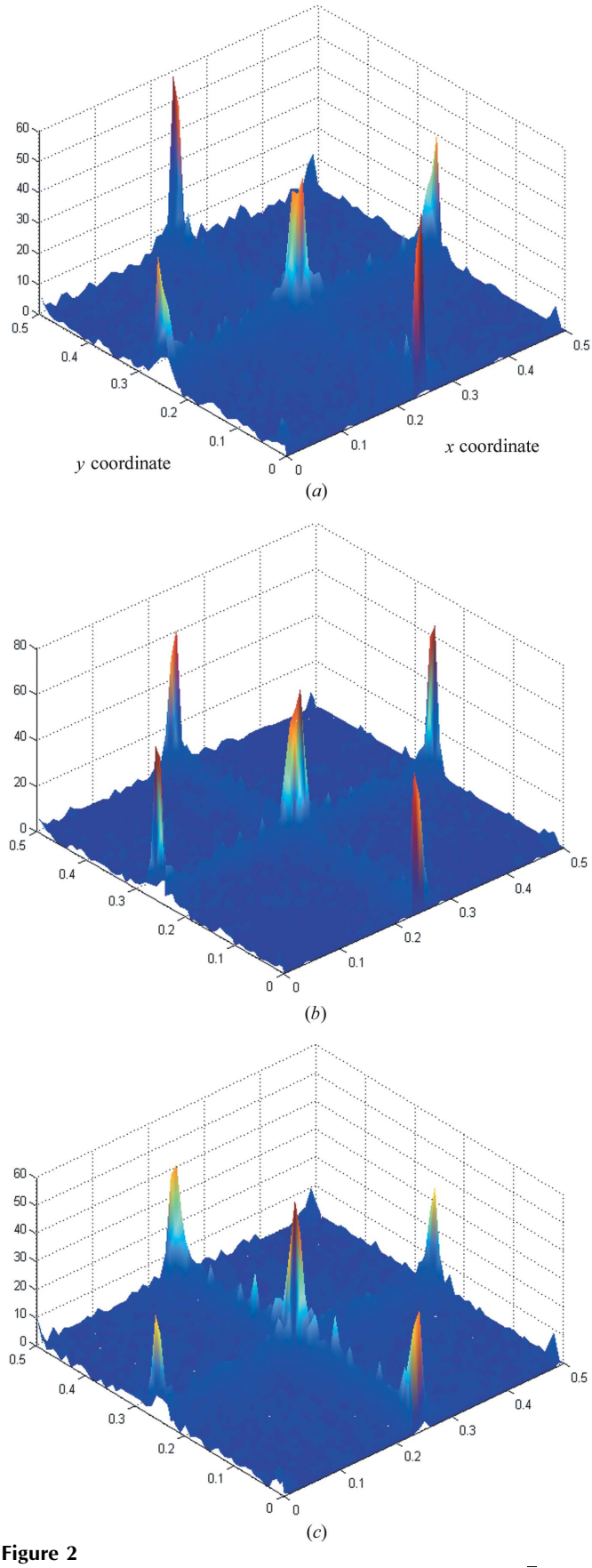


Figure 2
Three-dimensional plots for the positions of pair centroids in $P\overline{1}$. (*a*) *xy* projection; (*b*) *yz* projection; (*c*) *zx* projection.

The reader is referred to Motherwell's (1997) paper for consideration of molecular centroids for space groups $P\overline{1}$, $P2_1$, $P2_1/c$ and $P2_12_12_1$. Space groups $Pca2_1$ and $Pna2_1$ are relatively more common for $Z' = 2$ structures and were not included in that paper. For these space groups, results are therefore also presented for $Z' = 1$ structures to allow for comparison.

Note that it is possible to locate the pair centroids within the range 0–$\frac{1}{4}$ by applying origin shifts to a site of identical symmetry within the space group (Marsh *et al.*, 1998); however, keeping the display within the range 0–$\frac{1}{2}$ facilitates comparison with Motherwell's results. Displaying a quadrant of the unit cell makes the relationships with the positions of symmetry operators clearer, particularly in projection.

3.1. $P\overline{1}$

1380 $Z' = 2$ structures were used in the analysis (Figs. 1–3). The pair centroids would be expected to cluster around ($\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$), as this is the point furthest away from all inversion centres. Any pair centroids that lie near the true centres probably correspond to errors in the original structure determinations owing to missed centres of symmetry.

The histograms for the $Z' = 2$ structures show that the distributions are largely invariant for the different choices of centroid (Fig. 3).

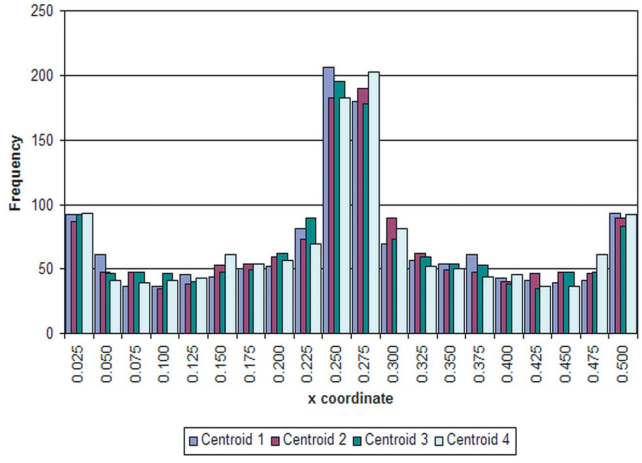
As in the $Z' = 1$ case, there are clusters of centroids at the planes perpendicular to the cell axes at heights $x = \frac{1}{4}$, $y = \frac{1}{4}$ and $z = \frac{1}{4}$, and the three projections down each of the cell axes give very similar plots (Fig. 1). These maxima at $\frac{1}{4}$ in each direction are clear in the frequency charts, again as for the $Z' = 1$ case. However, as can be seen in the three-dimensional frequency plots, there are also definite peaks at the positions (x, y) = (0, $\frac{1}{4}$), ($\frac{1}{2}$, $\frac{1}{4}$), ($\frac{1}{4}$, 0) and ($\frac{1}{4}$, $\frac{1}{2}$), and similarly for (y, z) and (z, x), which do not occur for $Z' = 1$ structures (Fig. 2). These peaks are clearly seen by combining the three axes onto one three-dimensional scatterplot (Fig. 4). This also shows that there is a strong tendency for pair centroids to be located on planes that lie as far as possible from all inversion operators in the space group.

A three-dimensional scatterplot of the molecular centroid positions in $P\overline{1}$ for $Z' = 2$ structures (Fig. 5) shows a similar distribution to that for the pair centroids (Fig. 4), although here the planes are much less sharply defined. Exclusion zones are seen to apply to single molecules as well as to molecular pairs, although these zones extend over a smaller area.

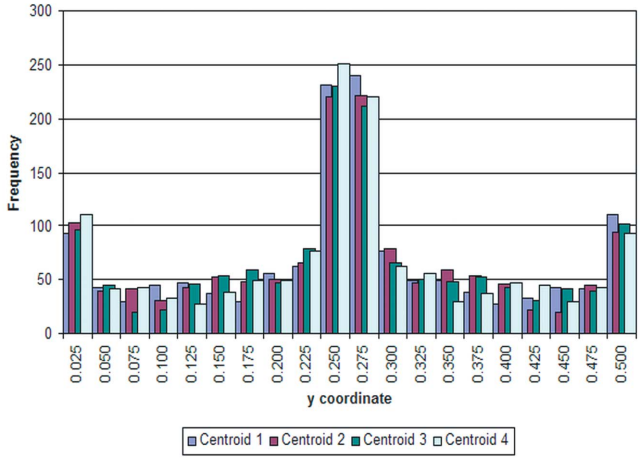
It is worth noting that the choices of cell axes in $P\overline{1}$ are quite arbitrary and therefore it would be possible to interchange the axes; this process will reinforce the observed trends.

If we assume that the most favoured position for a molecule in the homomolecular $Z' = 1$ structure is as far from the existing symmetry operators as possible (excluding $Z' = 1$ structures consisting of two half-molecules placed on centres of symmetry) then, in $P\overline{1}$, the molecular centroid will be at ($\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{4}$). In a $Z' = 2$ structure, if the independent molecules together act in a similar way to the molecule in a $Z' = 1$ structure then there will be a peak at this point; this is

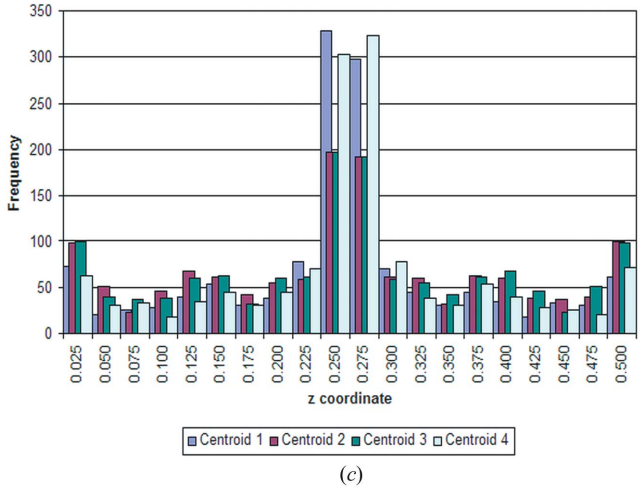
observed. If the independent molecules each behave like the single molecule in $Z' = 1$ structures then if one molecule is centred at $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$, the other will be centred at $(\frac{1}{4}, \frac{1}{4}, \frac{3}{4})$ (or symmetry-related permutations of these), giving rise to peaks for the pair centroids at $(0, \frac{1}{4}, \frac{1}{4})$, $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, $(\frac{1}{4}, 0, \frac{1}{4})$ *etc.* (Pidcock, 2006). These peaks are also observed.



(*a*)



(*b*)



(*c*)

Figure 3
Histograms for the positions of pair centroids by axis in $P\bar{1}$. (*a*) *x* coordinate; (*b*) *y* coordinate; (*c*) *z* coordinate.

The lines along values of $\frac{1}{4}$ seen in projection arise because when the molecules move away from the symmetry-favoured positions, they have greater freedom to move in planes as far from the symmetry operators as possible.

3.2. $P2_1$

1179 $Z' = 2$ structures in $P2_1$ were used in the analysis (Figs. 6–8). Because the origin is floating in *y*, the results are shown in the *xz* projection only. The 2_1 screw is located parallel to the *y* axis at $(0, y, 0)$, $(0, y, \frac{1}{2})$, $(\frac{1}{2}, y, 0)$ and $(\frac{1}{2}, y, \frac{1}{2})$ by convention. Pair centroids would be expected to avoid these areas, and so the *xz* projection should resemble the projections seen for $P\bar{1}$; this is indeed the case (Fig. 6). Although a molecule could lie on the screw axis, this would place restrictions on the features of the upper and lower faces of the

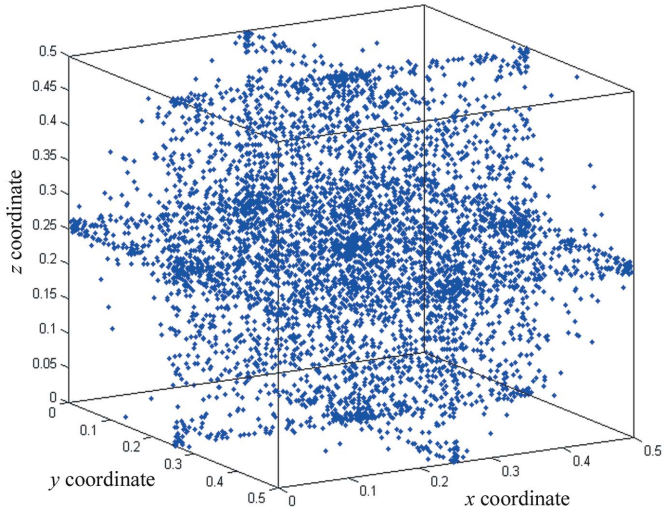


Figure 4
Three-dimensional scatterplot of the pair centroid positions in $P\bar{1}$ for $Z' = 2$ structures.

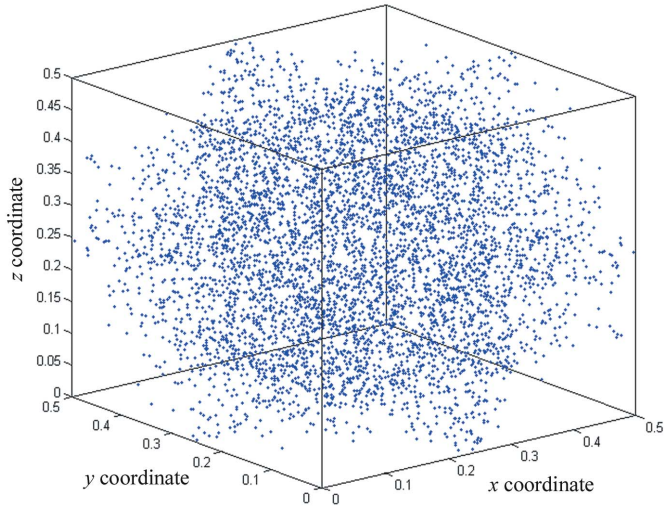


Figure 5
Three-dimensional scatterplot of the molecular centroid positions in $P\bar{1}$ for $Z' = 2$ structures.

molecule in the b direction. In $Z' = 1$ structures, the molecular centroids tend to avoid the screw axes. However, in a $Z' = 2$ structure, if the individual molecules avoid the screw axes, it becomes possible for the pair centroids to occupy this position and there are more pair centroids located in these regions; this is suggestive of the structure approaching a higher symmetry (*e.g.* $P2_1/c$, $P2$).

The populated areas occur in peaks at $(x, z) = (0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{4}, \frac{1}{4})$ (Fig. 7), with lines of low population at $\frac{1}{4}$ for both x and z for $Z' = 2$ structures (Figs. 7 and 8). In the $Z' = 1$ case, these lines at $x = \frac{1}{4}$ and $z = \frac{1}{4}$ are also present, but the peaks for areas of high population are correspondingly less pronounced.

The histograms (Fig. 8) show no marked differences between the choices of centroids.

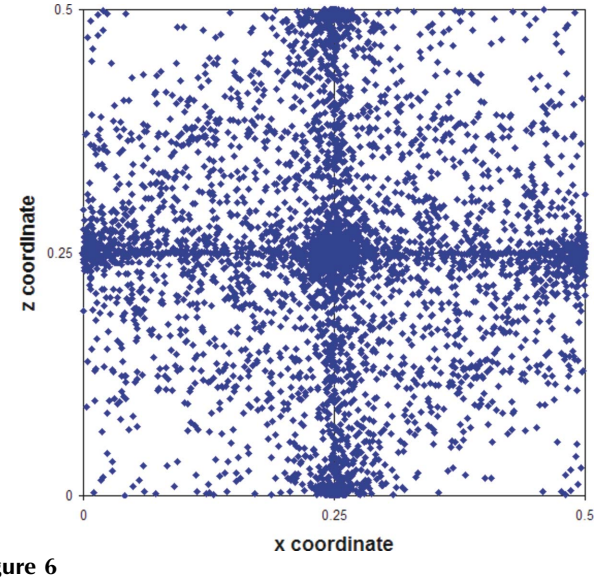


Figure 6
Scatterplot for the positions of pair centroids in the xz projection in $P2_1$ for $Z' = 2$ structures.

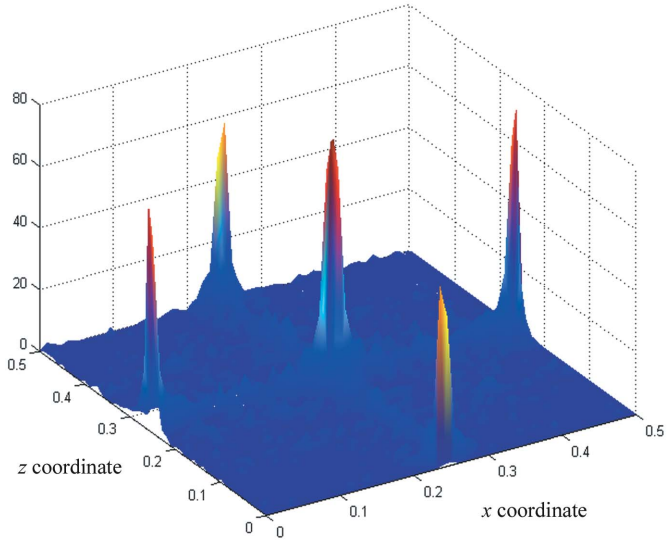


Figure 7
Three-dimensional frequency plot for the positions of pair centroids in the xz projection in $P2_1$ for $Z' = 2$ structures.

3.3. $P2_1/c$

There are 769 $Z' = 2$ structures in $P2_1/c$ (Figs. 9–11). This number is perhaps surprisingly low, but it includes only the structures in the standard setting of space group 14 and no alternative settings (*e.g.* $P2_1/a$, $P112_1/a$). Different settings would result in different centroid distributions, a result of differences in positions of the symmetry elements relative to the cell axes.

The origin is fixed by the centres of symmetry at $(0, 0, 0)$ and $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$. The glide plane would be expected to give rise to unoccupied planes at height $y = \frac{1}{4}$ and $\frac{3}{4}$, while the 2_1 axis at $(0, y, \frac{1}{4})$ should give a largely unoccupied line, for the space group in the standard setting.

Again, the trends for the $Z' = 2$ structures are similar to those observed in the $Z' = 1$ case. The x distribution histograms (Fig. 11*a*) are similar for both types of structure, with the only strong feature being a peak at $x = \frac{1}{4}$. In the y histogram (Fig. 11*b*), there are strong peaks at $y = n/8$ with the $Z' = 2$ structures, while for $Z' = 1$ structures the y distribution is more featureless. The z-coordinate histogram (Fig. 11*c*) also shows more defined peaks in the $Z' = 2$ than the $Z' = 1$ case, although

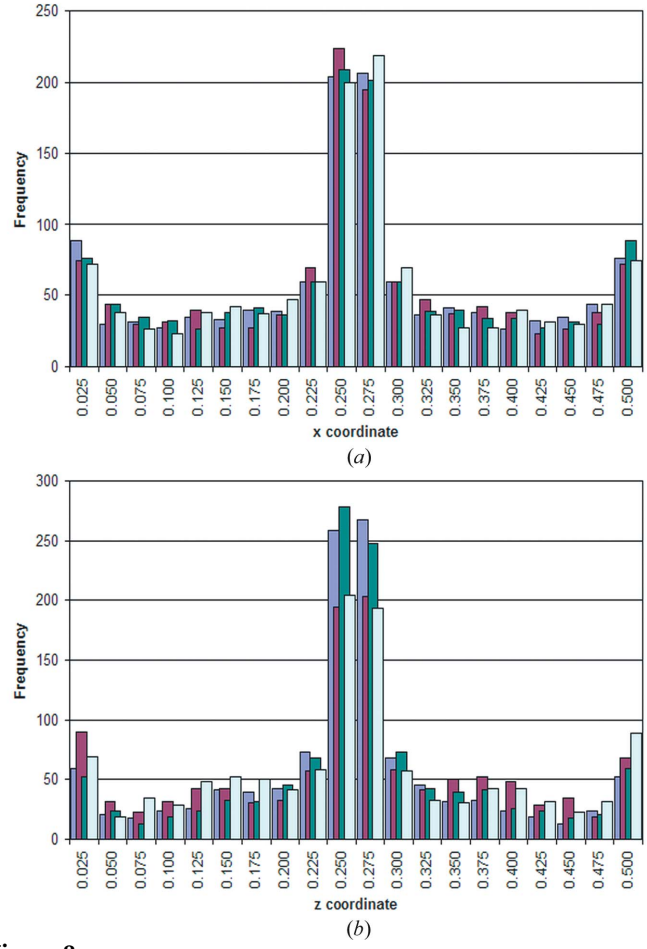


Figure 8
Histograms of the positions of pair centroids by axis in $P2_1$. (*a*) x coordinate; (*b*) z coordinate.

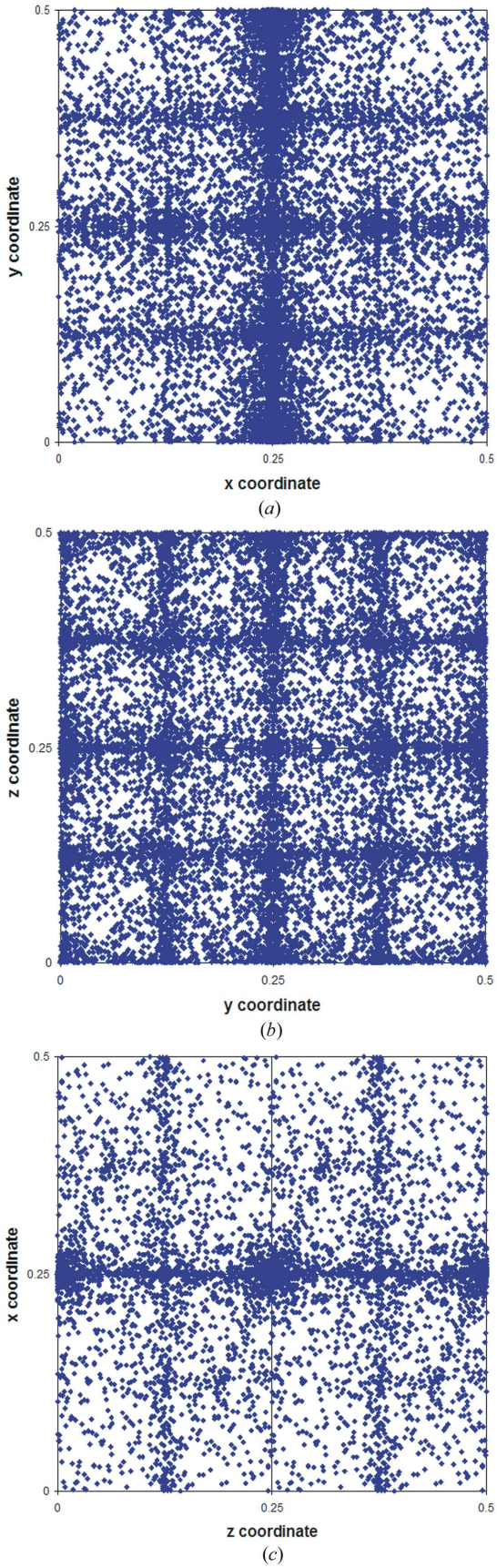


Figure 9
Scatterplots for the positions of pair centroids in $P2_1/c$ for $Z' = 2$ structures. (*a*) *xy* projection; (*b*) *yz* projection; (*c*) *zx* projection.

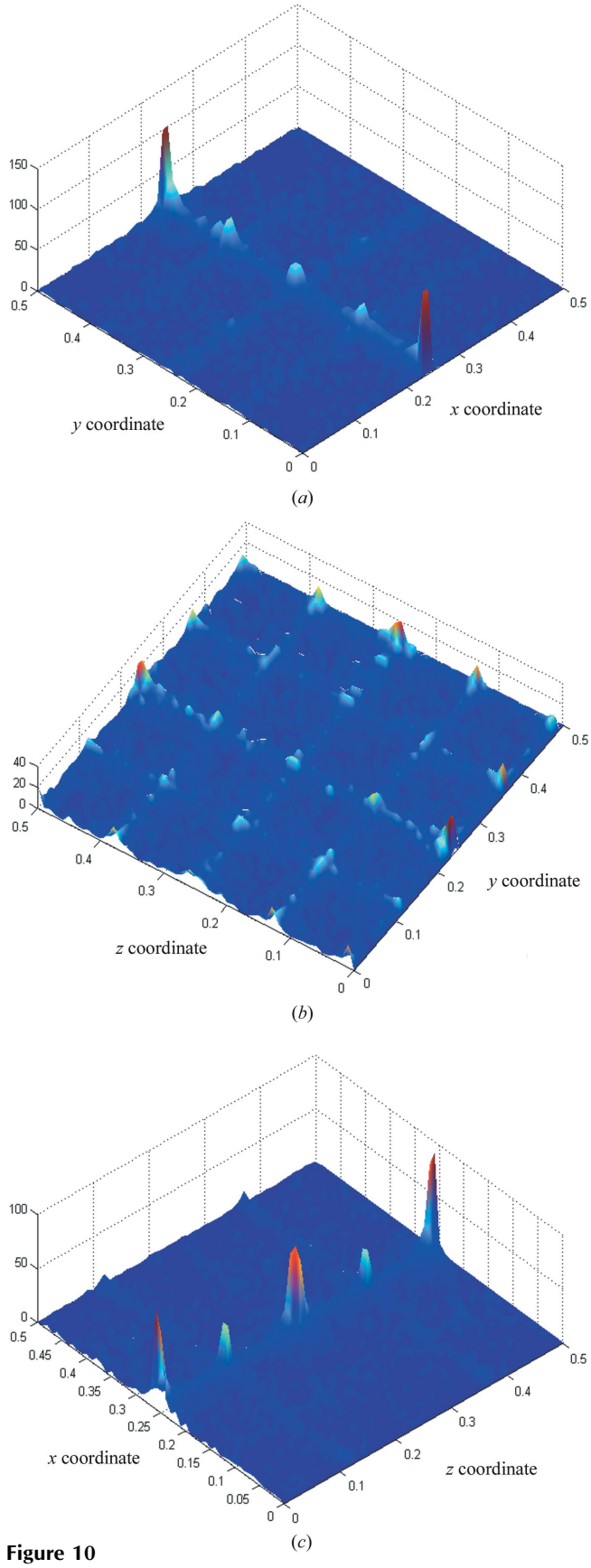


Figure 10
Three-dimensional frequency plots for the positions of pair centroids in $P2_1/c$ for $Z' = 2$ structures. (*a*) *xy* projection; (*b*) *yz* projection; (*c*) *zx* projection.

in both cases these occur at $z = n/8$; the peaks at $z = 0, \frac{1}{4}$ and $\frac{1}{2}$ are particularly strong for $Z' = 2$ structures.

In the xy projection (Fig. 9a), both types of structure exhibit a line of high population at $x = \frac{1}{4}$ and weaker lines at $y = \frac{1}{8}$ and $\frac{3}{8}$. For $Z' = 2$ structures, however, there are stronger peaks at values of $(x, y) = (\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$ (Fig. 10a). There is also a noticeable line of pair centroid location at $x = \frac{1}{4}$ (Fig. 9a); as the screw axis is positioned at values of $x = 0$ and $\frac{1}{2}$, this situation corresponds to pair centres lying as far from the screw axis as possible. The presence of the glide plane, which lies at height $y = \frac{1}{4}$, may explain the position of the weaker lines. However, in the $Z' = 2$ case, there is a marked line of population at $y = \frac{1}{4}$, *i.e.* pair centres are lying on the glide plane. This result is suggestive of a pseudo-glide plane.

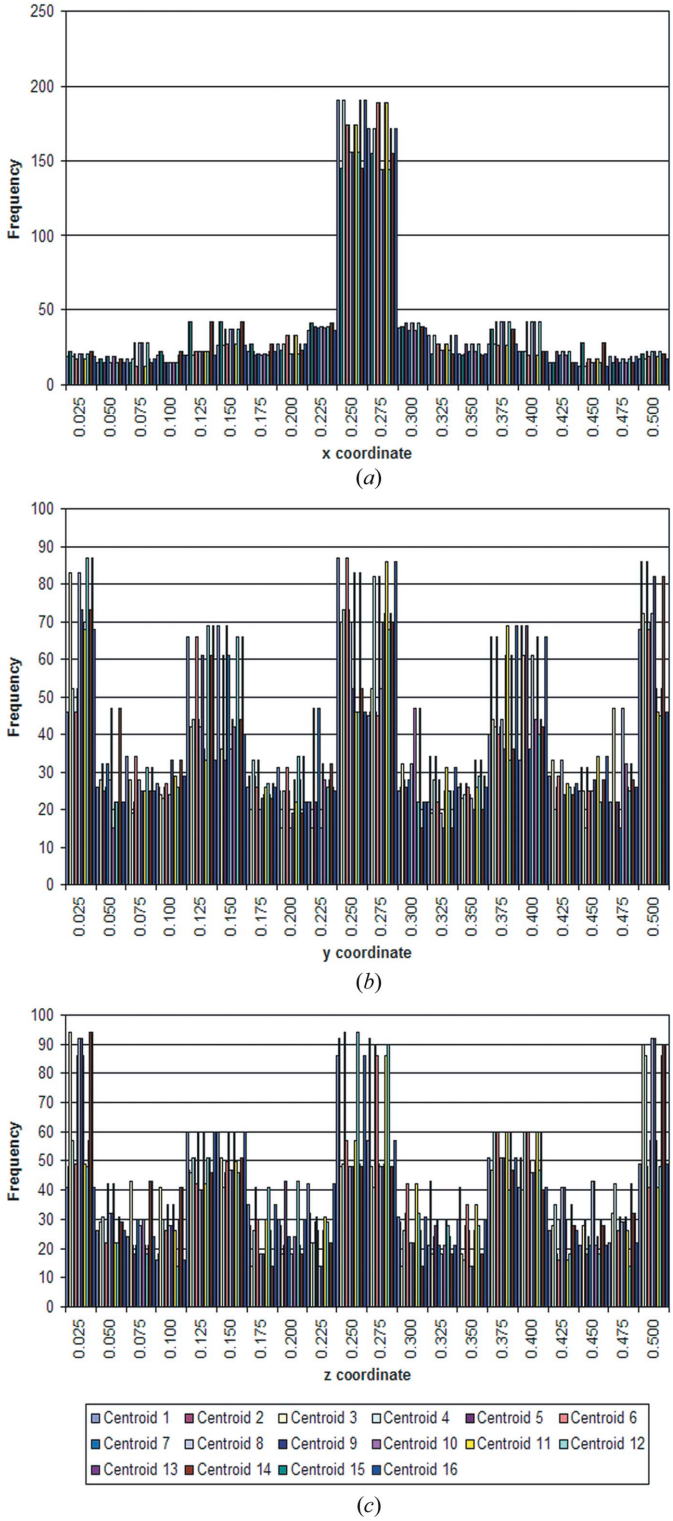


Figure 11
Histograms of the positions of centroids by axis in $P2_1/c$. (*a*) x coordinate; (*b*) y coordinate; (*c*) z coordinate.

In the yz projection (Figs. 9b and 10b), both $Z' = 1$ and $Z' = 2$ structures have lines of intensity at $y = \frac{1}{8}$ and $\frac{3}{8}$, and at $z = \frac{1}{8}$ and $\frac{3}{8}$. In $Z' = 2$ structures, these are stronger for certain centroid choices. Examination of the positions of the symmetry operators shows that these are the areas as far as possible from the screw axes and glide planes in that projection. For $Z' = 1$ structures, the areas around $(y, z) = (0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$ show low population; these areas correspond to the positions of the centres of symmetry. The area around $(y, z) = (\frac{1}{4}, \frac{1}{4})$ is also sparsely populated; this corresponds to the intersection of the screw axis and c-glide. The situation is different in the $Z' = 2$ case. The distribution of pair centroids appears more symmetrical, possibly as a consequence of the overlay of the 16 possible centroid positions. There are small peaks at values of $(y, z) = (0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$, as well as at $(\frac{1}{4}, \frac{1}{4})$ (Fig. 10b). Thus in this projection, the pair centroids do not seem to show the same reluctance to occupy the positions of symmetry operators as is seen in the $Z' = 1$ case.

In the zx projection, the line of intense population occurs at $x = \frac{1}{4}$, with weaker lines at $z = \frac{1}{8}$ and $\frac{3}{8}$, which corresponds to avoidance of the screw axes and inversion centres (Figs. 9c and 10c). In the $Z' = 2$ case, the line at $x = \frac{1}{4}$ is weaker and contains more peaks, at $(z, x) = (0, \frac{1}{4})$, $(\frac{1}{4}, 0)$ and $(\frac{1}{2}, \frac{1}{4})$; the lines at $z = \frac{1}{8}$ and $\frac{3}{8}$ are relatively more pronounced.

For $P2_1/c$, the histograms show differences between the centroid positions for the y and z axes, although in the x axis there are no apparent differences between centroid choice (Fig. 11).

3.4. $P2_12_12_1$

375 $Z' = 2$ structures in $P2_12_12_1$ have been used for this analysis (Figs. 12–14).

Analysis of $P2_12_12_1$ is more difficult because the axial directions are chosen by convention rather than being based on symmetry elements (Gruber, 1973). The three principal projections show the same symmetry operators, and scatter plots have been orientated so that the operators lie in equivalent positions on the page, *i.e.* the projected space-group symmetry of the diagrams is the same.

In the $Z' = 1$ case, Motherwell found very little fine structure distinguishable for any projection. For $Z' = 2$ structures, however, there are projections that show a greater degree of clustering. From the projections onto a plane, the xy projection appears the most random (Fig. 12a and 13a). It is also the projection that deviates most markedly from its equivalent $Z' = 1$ projection. There is perhaps a ring, of diameter of

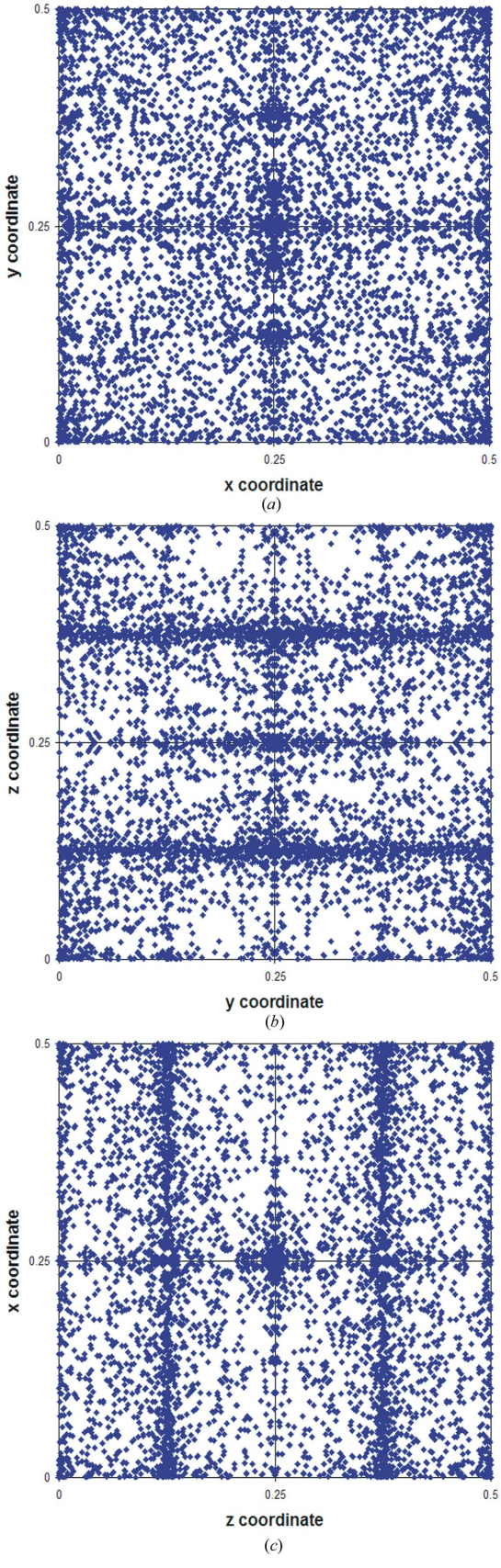


Figure 12
Scatterplots for the positions of centroids in $P2_12_12_1$. (*a*) *xy* projection; (*b*) *yz* projection; (*c*) *zx* projection.

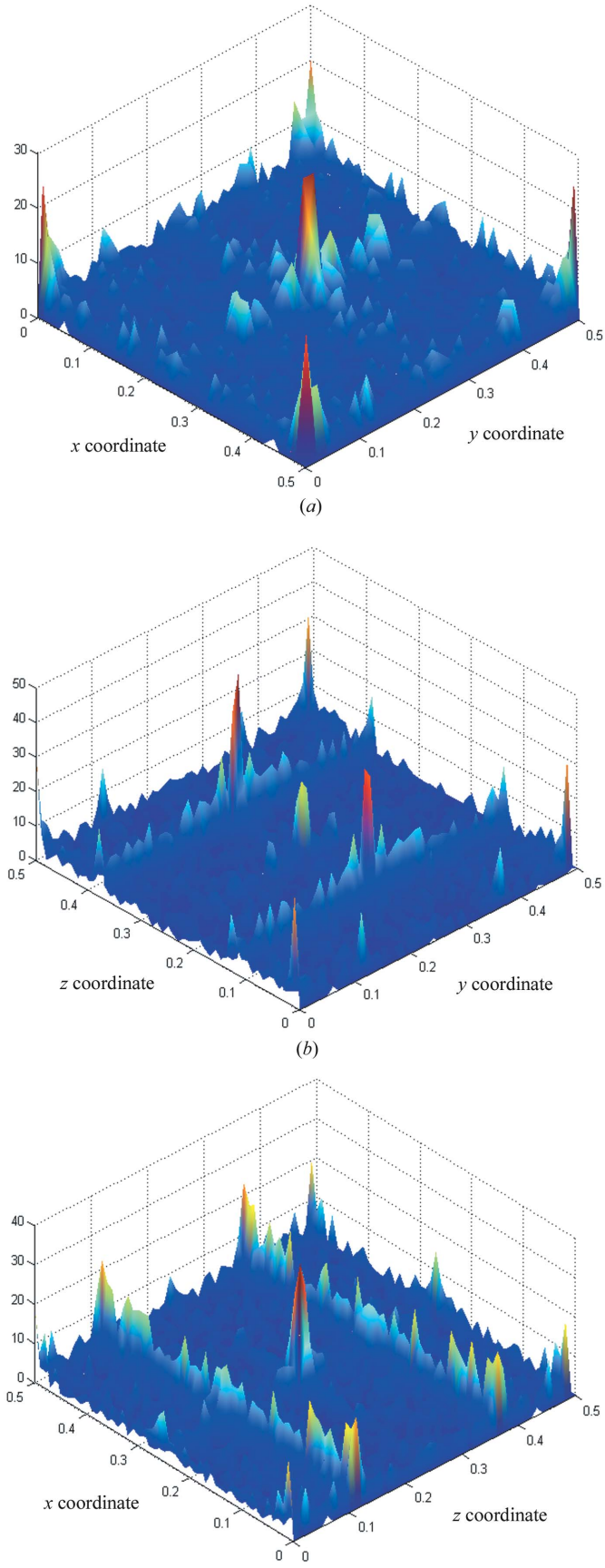


Figure 13
Three-dimensional frequency plots for the positions of centroids in $P2_12_12_1$. (*a*) *xy* projection; (*b*) *yz* projection; (*c*) *zx* projection.

approximately $\frac{3}{8}$ centred at $(x, y) = (\frac{1}{4}, \frac{1}{4})$, of lower population, but no exclusion zones at $(x, y) = (0, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{4})$ that are seen in the distribution for $Z' = 1$ structures. However, there are peaks at $(x, y) = (0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{4}, \frac{1}{4})$ (Fig. 13*a*).

The *yz* and *zx* pair centroid scatterplots (Figs. 12*b*, 12*c*, 13*b* and 13*c*) may be regarded as sharpened-up versions of the $Z' = 1$ molecular centroid plots. There are areas of low population in the regions of $(y, z) = (0, \frac{1}{4})$ and $(\frac{1}{2}, \frac{1}{4})$ (Figs. 12*b* and 13*b*), and also, but less strongly, at $(y, z) = (\frac{1}{4}, 0)$ and $(\frac{1}{2}, 0)$. There are now much stronger lines of higher population density at $y = \frac{1}{8}$ and $\frac{3}{8}$ and at $z = \frac{1}{8}$ and $\frac{3}{8}$, compared with the *xy* projections, and these lines in particular are much stronger for $Z' = 2$ structures. The area at $(y, z) = (\frac{1}{4}, \frac{1}{4})$ is also more populated.

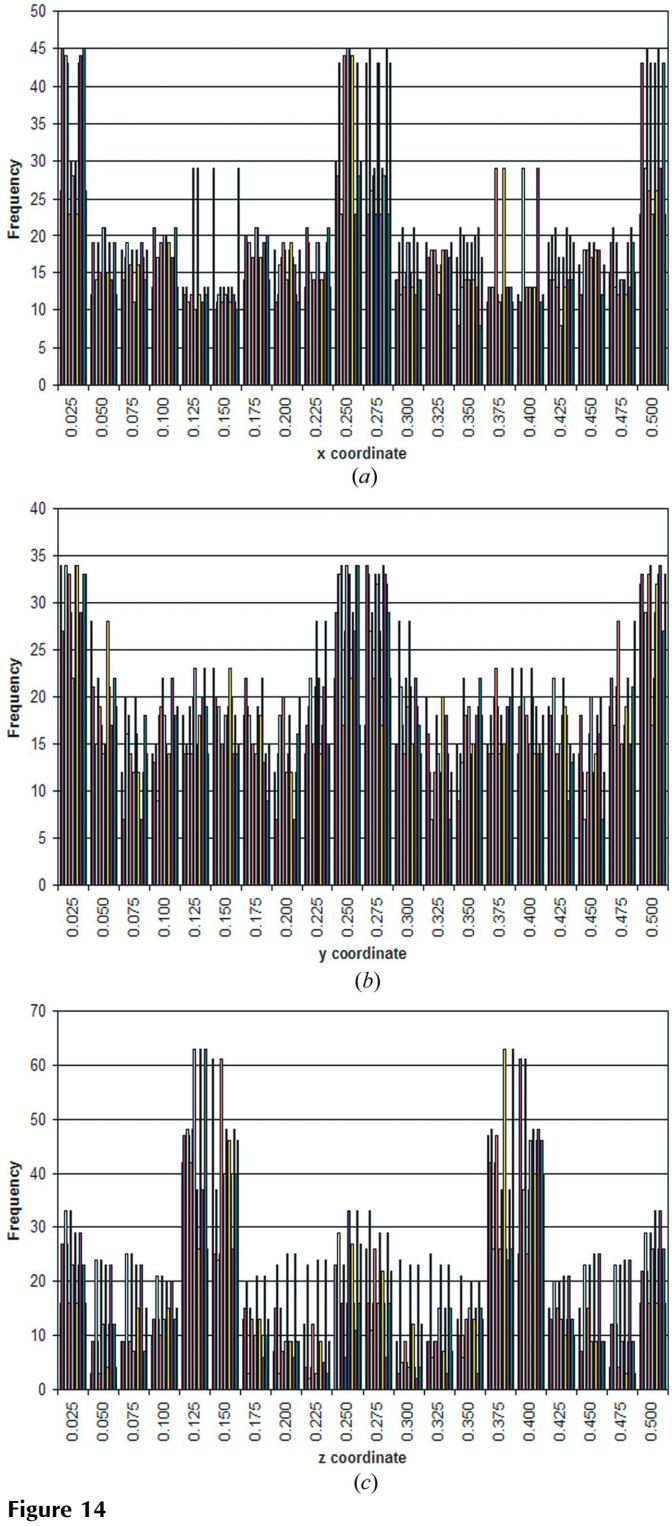


Figure 14
Histograms of the positions of centroids by axis in $P2_12_12_1$. (*a*) *x* coordinate; (*b*) *y* coordinate; (*c*) *z* coordinate.

The *zx* projection (Figs. 12*c* and 13*c*) has very strong lines at $z = \frac{1}{8}$ and $\frac{3}{8}$, as may be expected given the *yz* projection; these lines are present in the distributions for $Z' = 1$ structures but are much more pronounced in the $Z' = 2$ case. There are two exclusion zones, at $(z, x) = (0, \frac{1}{4})$ and at $(\frac{1}{2}, \frac{1}{4})$, which should be considered as one exclusion zone that is split as a result of the choice of cell axis ranges.

The histograms for the $Z' = 2$ structures have more apparent features than for the $Z' = 1$ structures. The *x* coordinate histogram has peaks at $x = 0$, $\frac{1}{4}$ and $\frac{1}{2}$ (Fig. 14*a*); the *y* coordinate histogram also has peaks at values of 0, $\frac{1}{4}$ and $\frac{1}{2}$, although these are less pronounced (Fig. 14*b*). The *z* coordinate histogram shows very strong peaks at $z = \frac{1}{8}$ and $\frac{3}{8}$, with a weak peak at $z = \frac{1}{4}$ (Fig. 14*c*).

The three-dimensional frequency plots for all three projections exhibit a peak at $(\frac{1}{4}, \frac{1}{4})$, which is not observed in the $Z' = 1$ cases (Fig. 13).

Surprisingly, for both $Z' = 1$ and $Z' = 2$ structures, the distribution of molecular centroids and pair centroids, respectively, is not the same in all projections, even though the three axes are equivalent in terms of symmetry elements. In particular, the *z* distribution (Fig. 14*c*) is markedly different from that for the other two axes; this is true of both $Z' = 1$ and $Z' = 2$ structures. Motherwell could offer no explanation of the phenomenon but it was noted that only ∼40% of structures had the standard setting with $a < b < c$. In this work, all the $P2_12_12_1$ $Z' = 2$ structures have cells in the standard settings. Note that the strongest features in Fig. 12 are in the planes perpendicular to the longest axis.

3.5. *Pca*2$_1$

There are 439 structures with $Z' = 1$ (Figs. 15–17) and 110 structures with $Z' = 2$ (Figs. 18–20). The origin is floating in the *z* direction, so only the distributions in *x* and *y* need to be considered.

In both $Z' = 1$ and $Z' = 2$ structures the centroids are very strongly located at values of $y = \frac{1}{4}$, and the *y* coordinate distributions have a very similar form (Figs. 17*b* and 20*b*). This is unsurprising given the positions of the symmetry elements in the space group; the screw axes are located at $(x, y) = (0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$ and $(\frac{1}{2}, \frac{1}{2})$, and the *c*-glide is coincident with the *a* axis, making values of $x = 0$ and $\frac{1}{2}$ unfavourable for positioning molecular centroids. The *a*-glide occurs at $y = \frac{1}{4}$ and $\frac{3}{4}$, which will tend to disfavour centroids at values of $y = \frac{1}{4}$.

However, the distributions in the *x* coordinate are quite different (Figs. 17*a* and 20*a*). The $Z' = 2$ case has definite peaks at values of $n/8$ with the strongest occurring at $x = \frac{1}{8}$ and $\frac{3}{8}$. The distribution is more uniform in the *x* coordinate for $Z' = 1$ structures, although there is a noticeable minimum at $x = \frac{1}{4}$; this

is in marked contrast to the small but definite peak at $x = \frac{1}{4}$ in the $Z' = 2$ case.

The *xy* scatterplots (Figs. 15 and 18) also show these distinct differences and make the ordering in the *x* coordinate very apparent.

This higher level of ordering may be explained by considering the space group implications of placing an inversion centre at the position $(x, y) = (\frac{1}{8}, \frac{1}{4})$. There is no space group that corresponds to this (even with applying an origin shift to obtain the conventional setting to place the inversion at the origin); one of the problems is that the unit-cell volume would need to be reduced, because of how the other space-group operators must be used in a real space group to generate

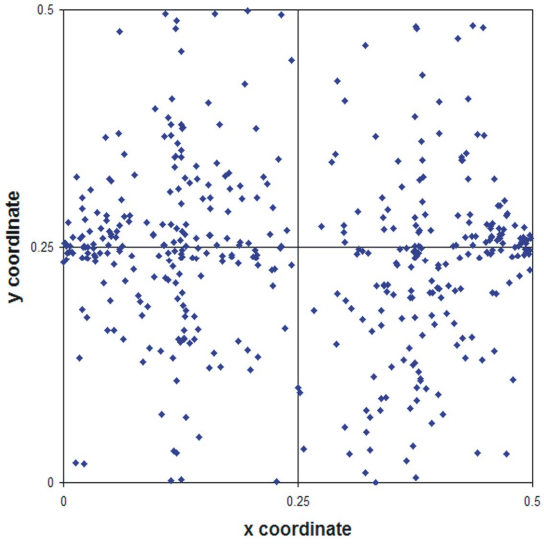


Figure 15
Scatterplot for the positions of molecular centroids in $Z' = 1$ structures in the *xy* projection in $Pca2_1$.

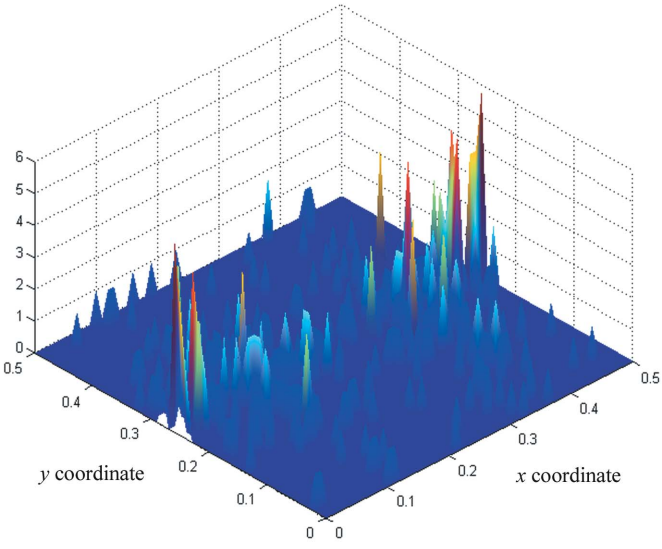


Figure 16
Three-dimensional frequency plot for the positions of molecular centroids for $Z' = 1$ structures in the *xy* projection in $Pca2_1$.

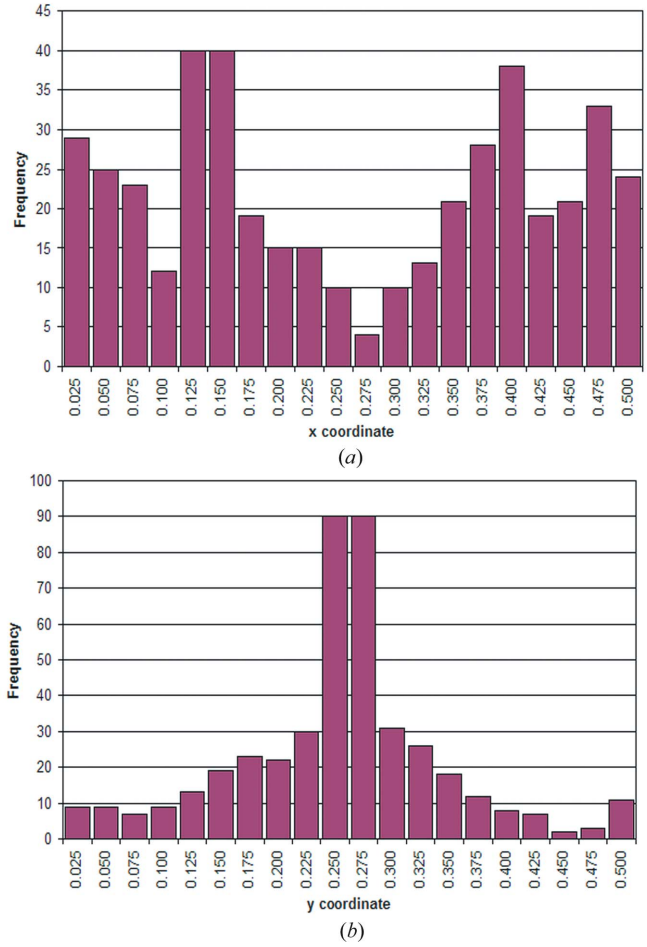


Figure 17
Histogram of the positions of molecular centroids in $Z' = 1$ structures by axis in $Pca2_1$. (*a*) *x* coordinate; (*b*) *y* coordinate.

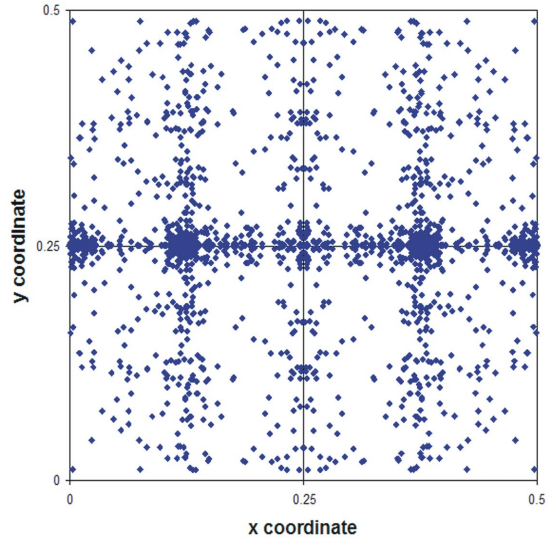


Figure 18
Scatterplot for the positions of pair centroids in $Z' = 2$ structures in the *xy* projection in $Pca2_1$.

further inversion centres. However, the location of the pair centroid at $(x, y) = (\frac{1}{8}, \frac{1}{4})$ is suggestive of how $Z' > 1$ structures may try to evoke a higher symmetry in their crystal packing.

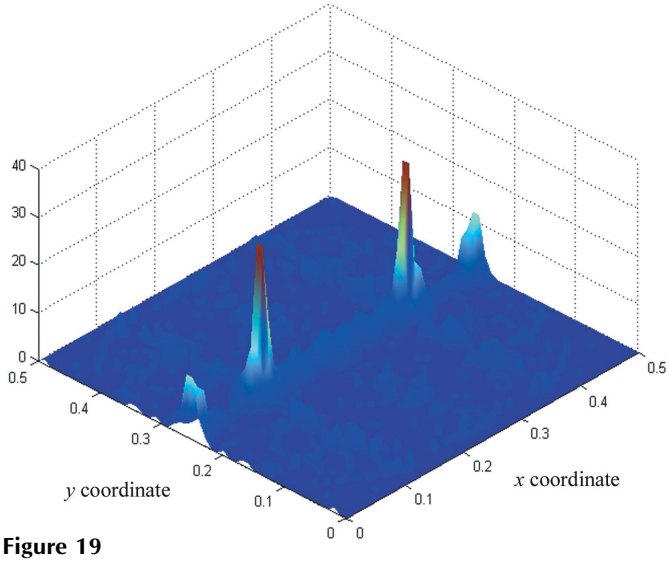
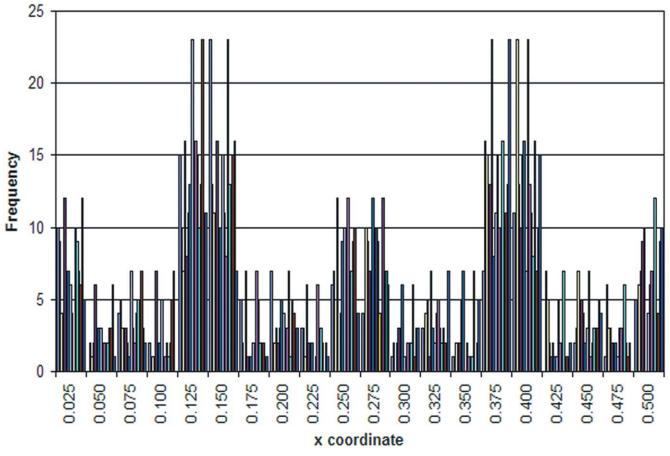
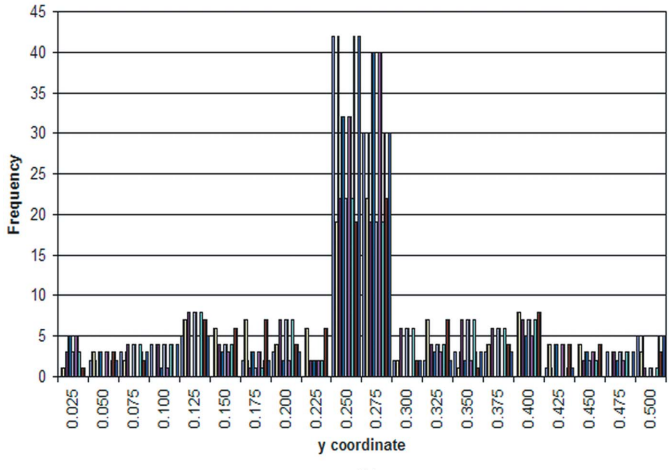


Figure 19
Three-dimensional frequency plot for the positions of pair centroids for $Z' = 2$ structures in the xy projection in $Pca2_1$.



(*a*)



(*b*)

Figure 20
Histogram of the positions of pair centroids in $Z' = 2$ structures by axis in $Pca2_1$. (*a*) x coordinate; (*b*) y coordinate.

It is worth emphasizing at this point that the graphs shown are for all $Z' = 2$ structures in $Pca2_1$ in the CSD; no filtering has been applied for the similarity of the conformation of the two molecules.

These results make for an interesting comparison with the findings of Marsh *et al.* (1998); the distributions are very similar in the xy projection, although here we observe many more structures where the two molecules are centred along a line of $x = \frac{1}{4}$, rather than being restricted to the point $(x, y) = (0, \frac{1}{4})$. We suggest that this is because the independent molecules are related by operators other than centres of inversion.

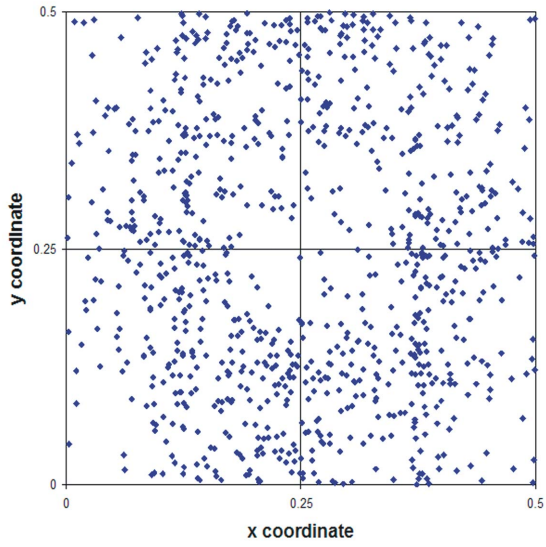


Figure 21
Scatterplot for the positions of molecular centroids in $Z' = 1$ structures in the zx projection in $Pna2_1$.

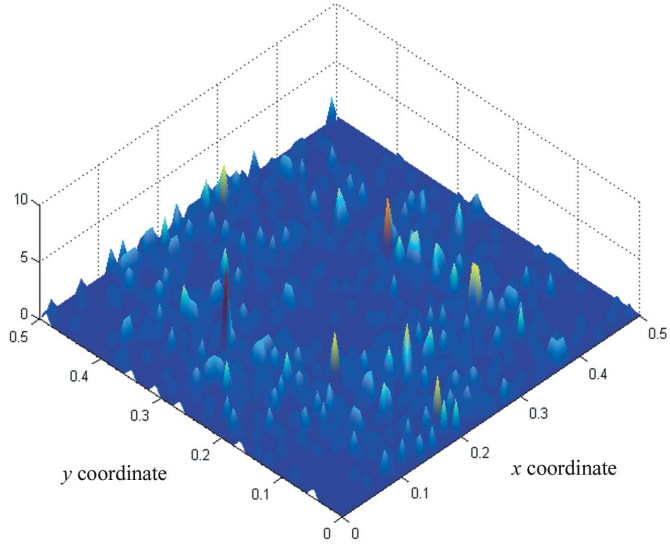


Figure 22
Three-dimensional frequency plot for the positions of molecular centroids for $Z' = 1$ structures in the xy projection in $Pna2_1$.

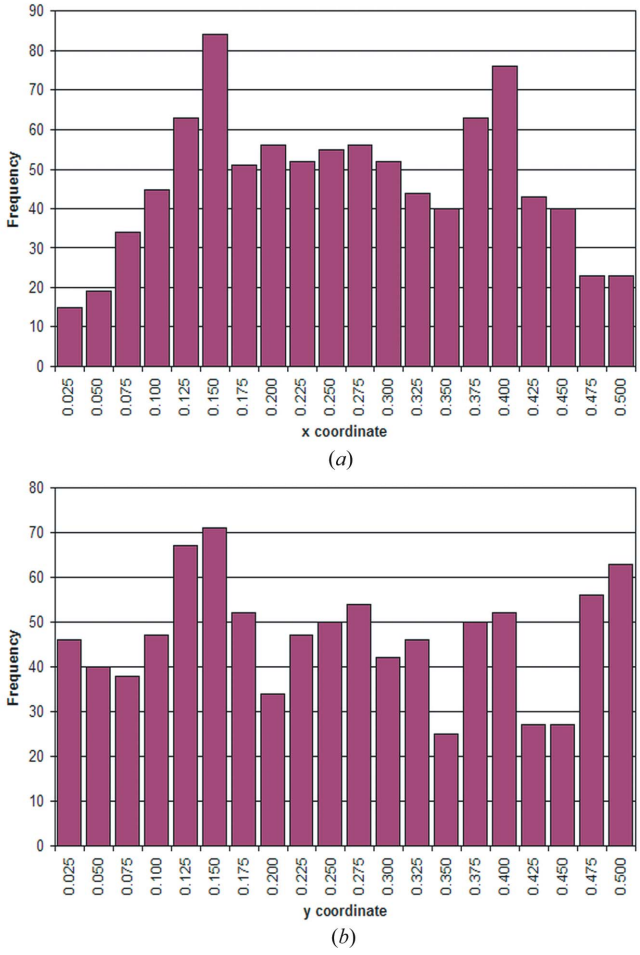


Figure 23
Histogram of the positions of molecular centroids in $Z' = 1$ structures by axis in $Pna2_1$. (*a*) *x* coordinate; (*b*) *y* coordinate.

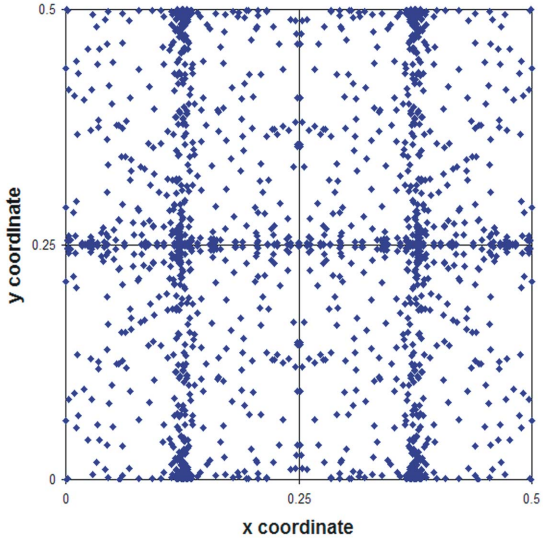


Figure 24
Scatterplot for the positions of pair centroids in $Z' = 2$ structures in the *zx* projection in $Pna2_1$.

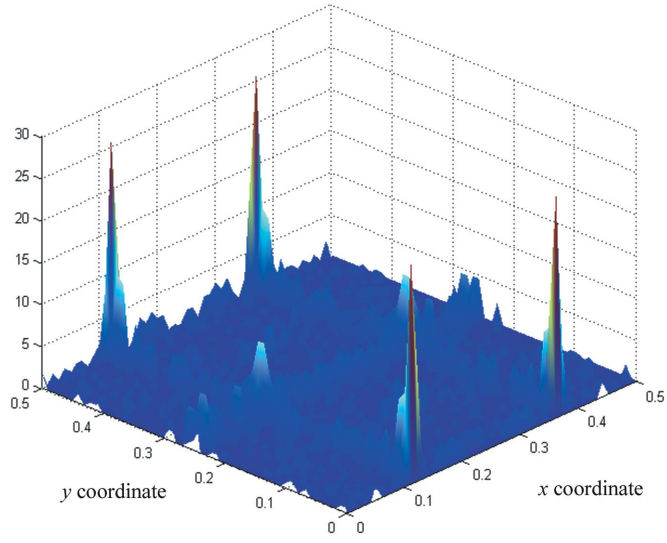


Figure 25
Three-dimensional frequency plot for the positions of pair centroids for $Z' = 2$ structures in the *xy* projection in $Pna2_1$.

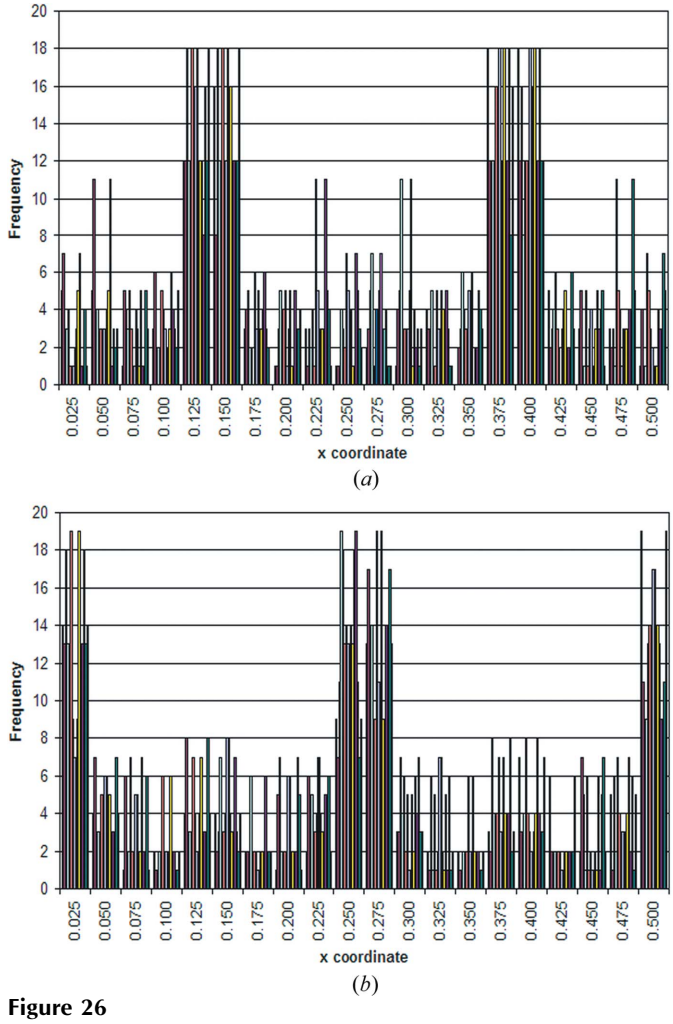


Figure 26
Histogram of the positions of pair centroids in $Z' = 2$ structures by axis in $Pna2_1$. (*a*) *x* coordinate; (*b*) *y* coordinate.

3.6. *Pna*2_1

There were 934 $Z' = 1$ structures in *Pna*2_1 (Figs. 21–23) and 105 structures with $Z' = 2$ (Figs. 24–26).

The forms of the $Z' = 1$ and $Z' = 2$ centroid distributions are quite different; these differences are stronger in both x and y than was the case for *Pca*2_1. In the $Z' = 1$ case, there are no values of the x or y coordinates that clearly appear to be preferred (Fig. 23); however, the areas $(x, y) = (0, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{2})$ and $(\frac{1}{4}, \frac{1}{4})$ are obviously unfavourable and the xy scatterplot shows very clear exclusion zones of molecular centroid positions at these values (Fig. 21). These positions correspond to the positions of the 2_1 screw axis $[(0, 0), (0, \frac{1}{2}), (\frac{1}{2}, 0), (\frac{1}{2}, \frac{1}{2})]$ and the intersection of the two glide planes at point $(\frac{1}{4}, \frac{1}{4})$. The resulting xy distribution therefore resembles a ring of occupied centroid positions.

In contrast, there is evidently a greater degree of ordering in the x and y coordinate distributions for $Z' = 2$ structures, with strong peaks at $x = \frac{1}{8}$ and $x = \frac{3}{8}$ (Fig. 26). In the y coordinate distribution, peaks occur at $y = 0, \frac{1}{4}$ and $\frac{1}{2}$; as can be seen from the xy scatterplot the peak at $y = \frac{1}{2}$ is particularly strong (Fig. 24). The xy scatterplot and the three-dimensional frequency plot (Figs. 24 and 25) show clearly that there is clustering in the distribution of pair centroids in $Z' = 2$ structures. There is some evidence for the annular distribution of pair centroids, like that observed for $Z' = 1$ structures; this is particularly evident from the apparent curvature of the occupied positions close to $(x, y) = (\frac{1}{4}, \frac{1}{8})$ and $(\frac{1}{4}, \frac{3}{8})$.

Again, comparison of these results with those of Marsh *et al.* (1998) shows that the distributions are similar, but with a larger spread of centroid positions in this study resulting from the presence of local operators other than inversion centres.

4. Conclusions

The analyses of positions of pair centroids for structures in $P\overline{1}$, $P2_1$, $P2_1/c$, $P2_12_12_1$, *Pca*2_1 and *Pna*2_1, and molecular centroids for structures in *Pca*2_1 and *Pna*2_1, show that the general trends for $Z' = 2$ structures are the same as for molecular centroids in $Z' = 1$ structures, *i.e.* broadly speaking the pairs of molecules tend to behave like a single unit. This may be useful for predicting structures with $Z' > 1$. In $Z' = 1$ structures, molecular centroids show where the molecules are. In $Z' = 2$ structures, pair centroids, composed of two independent molecules, show where the molecules are not located. In those cases where the $Z' = 1$ molecular centroid distributions are similar to the $Z' = 2$ pair centroid distributions, it follows that the $Z' = 2$ molecular centroids must be in different environments from the $Z' = 1$ molecular centroids.

However, although the general trends are similar, there are fundamental differences between the distributions of pair centroids and molecular centroids for $Z' = 2$ and $Z' = 1$ structures, respectively. Where the two-dimensional scatterplots are similar, there is frequently a sharpening of the distributions in structures with $Z' = 2$ compared with $Z' = 1$. $Z' = 2$ structures show a greater tendency for the centroids to cluster at coordinates as far removed as possible from existing space-group operators; this indicates a trend towards pseudosymmetry or missed symmetry. In general, the $Z' = 2$ distributions are more ordered than the distributions for $Z' = 1$ structures, and this is particularly true for the higher-symmetry space groups. Although each $Z' = 2$ structure contributes more than one point to the scatterplots and three-dimensional graphs, constructing the histograms from each contribution separately indicates that there is very little difference in patterns arising from different choices of pair centroids. Further differences between the distributions of $Z' = 1$ molecular centroids and $Z' = 2$ pair centroids arise if the individual molecules in the $Z' = 2$ structure favour the positions furthest away from existing operators. A predicted crystal structure with $Z' = 2$ where the pair centroid of the independent molecules lies in the area that is found to be sparsely populated for that space group would be suspect.

APPENDIX *A*
Combinations of molecular positions giving rise to possible pair centroid positions for $P2_1/c$

The general positions for individual molecules are given for $P2_1/c$. These are then combined and used to derive all possible combinations for the pair centroids for that space group.

The general positions of the individual molecules are:

$$x, y, z \qquad X, Y, Z$$

$$\overline{x}, y+\tfrac{1}{2}, \overline{z}+\tfrac{1}{2} \qquad \overline{X}, Y+\tfrac{1}{2}, \overline{Z}+\tfrac{1}{2}$$

$$\overline{x}, \overline{y}, \overline{z} \qquad \overline{X}, \overline{Y}, \overline{Z}$$

$$x, \overline{y}+\tfrac{1}{2}, z+\tfrac{1}{2} \qquad X, \overline{Y}+\tfrac{1}{2}, Z+\tfrac{1}{2}.$$

There are 16 pair centroids:

$$\frac{(x+X)}{2}, \frac{(y+Y)}{2}, \frac{(z+Z)}{2}, \tag{1}$$

$$\frac{(x-\overline{X})}{2}, \frac{(y+Y+\frac{1}{2})}{2}, \frac{(z+\overline{Z}+\frac{1}{2})}{2}, \tag{2}$$

$$\frac{(x+\overline{X})}{2}, \frac{(y+\overline{Y})}{2}, \frac{(z+\overline{Z})}{2}, \tag{3}$$

$$\frac{(x+X)}{2}, \frac{(y+\overline{Y}+\frac{1}{2})}{2}, \frac{(z+Z+\frac{1}{2})}{2}, \tag{4}$$

$$\frac{(\overline{x}+X)}{2}, \frac{(y+Y+\frac{1}{2})}{2}, \frac{(\overline{z}+Z+\frac{1}{2})}{2}, \tag{5}$$

$$\frac{(\overline{x}+\overline{X})}{2}, \frac{(y+Y)}{2}, \frac{(\overline{z}+\overline{Z})}{2}, \tag{6}$$

$$\frac{(\overline{x}+\overline{X})}{2}, \frac{(y+\overline{Y}+\frac{1}{2})}{2}, \frac{(\overline{z}+\overline{Z}+\frac{1}{2})}{2}, \tag{7}$$

$$\frac{(\overline{x}+X)}{2},\frac{(y+\overline{Y})}{2},\frac{(\overline{z}+Z)}{2}, \quad (8)$$

$$\frac{(\overline{x}+X)}{2},\frac{(\overline{y}+Y)}{2},\frac{(\overline{z}+Z)}{2}, \quad (9)$$

$$\frac{(\overline{x}+\overline{X})}{2},\frac{(\overline{y}+Y+\tfrac{1}{2})}{2},\frac{(\overline{z}+\overline{Z}+\tfrac{1}{2})}{2}, \quad (10)$$

$$\frac{(\overline{x}+\overline{X})}{2},\frac{(\overline{y}+\overline{Y})}{2},\frac{(\overline{z}+\overline{Z})}{2}, \quad (11)$$

$$\frac{(\overline{x}+X)}{2},\frac{(\overline{y}+\overline{Y}+\tfrac{1}{2})}{2},\frac{(\overline{z}+Z+\tfrac{1}{2})}{2}, \quad (12)$$

$$\frac{(x+X)}{2},\frac{(\overline{y}+Y+\tfrac{1}{2})}{2},\frac{(z+Z+\tfrac{1}{2})}{2}, \quad (13)$$

$$\frac{(x+\overline{X})}{2},\frac{(\overline{y}+Y)}{2},\frac{(z+\overline{Z})}{2}, \quad (14)$$

$$\frac{(x+\overline{X})}{2},\frac{(\overline{y}+\overline{Y}+\tfrac{1}{2})}{2},\frac{(z+\overline{Z})}{2}, \quad (15)$$

$$\frac{(x+X)}{2},\frac{(\overline{y}+\overline{Y})}{2},\frac{(z+Z)}{2}, \quad (16)$$

The author thanks the EPSRC for funding (grant No. GR/R75250/01) and D. J. Watkin for comments on the manuscript.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Brock, C. P. (1996). *J. Res. Natl Stand. Technol.* **101**, 321–325.
Brock, C. P. & Dunitz, J. D. (1994). *Chem. Mater.* **6**, 1118–1127.
Dalhus, B. & Görbitz, C. H. (2000). *Acta Cryst.* B**56**, 715–719.
Day, G. M., Motherwell, W. D. S., Ammon, H. L., Boerrigter, S. X. M., Della Valle, R. G., Venuti, E., Dzyabchenko, A., Dunitz, J. D., Schweizer, B., van Eijck, B. P., Erk, P., Facelli, J. C., Bazterra, V. E., Ferraro, M. B., Hofmann, D. W. M., Leusen, F. J. J., Liang, C., Pantelides, C. C., Karamertzanis, P. G., Price, S. L., Lewis, T. C., Nowell, H., Torrisi, A., Scheraga, H. A., Arnautova, Y. A., Schmidt, M. U. & Verwer, P. (2005). *Acta Cryst.* B**61**, 511–527.
Filippini, G. & Gavezzotti, A. (1992). *Acta Cryst.* B**48**, 230–234.
Gavezzotti, A. & Flack, H. (2005). (*IUCr*) *Teaching Pamphlets: Crystal Packing.* http://www.iucr.org/iucr-top/comm/cteach/pamphlets/21/21.html.
Gillespie, R. J., Robinson, E. A. & Heard, G. L. (1998). *Inorg. Chem.* **37**, 6884–6889.
Gruber, B. (1973). *Acta Cryst.* A**29**, 433–440.
Heine, V. & Price, S. L. (1985). *J. Phys. C*, **18**, 5259–5278.
Karthe, P., Sadasivan, C. & Gautham, N. (1993). *Acta Cryst.* B**49**, 1069–1071.
Kitaigorodskii, A. I. (1961). *Organic Chemical Crystallography*. New York: Academic Press.
Koutentis, P. A., Haddon, R. C., Oakley, R. T., Cordes, A. W. & Brock, C. P. (2001). *Acta Cryst.* B**57**, 680–691.
Lehmler, H. J., Parkin, S. & Brock, C. P. (2004). *Acta Cryst.* B**60**, 325–332.
Marsh, R. E., Schomaker, V. & Herbstein, F. H. (1998). *Acta Cryst.* B**54**, 921–924.
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., Leuson, 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.
Pidcock, E. (2006). *Acta Cryst.* B**62**, 268–279.
Pidcock, E. & Motherwell, W. D. S. (2004). *Cryst. Growth Des.* **4**, 611–620.
Steed, J. W. (2003). *CrystEngComm*, **5**, 169–179.
Wilson, A. J. C. (1993). *Acta Cryst.* A**49**, 210–212.