research papers

Automatic detection of molecular symmetry in the Cambridge Structural Database

Jason C. Cole,[a] Jing Wen Yao,[a] Gregory P. Shields,[a] W. D. S. Motherwell,[a] Frank H. Allen[a]* and Judith A. K. Howard[b]

[a]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England, and [b]Department of Chemistry, University of Durham, South Road, Durham DH1 3LE, England

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

A method for the detection of approximate molecular symmetry in crystal structures has been developed. The point-group symmetry is assigned to each molecule and the relevant symmetry elements can be visualized, superimposed on the molecule. The method has been validated against reference structures with exact symmetry subjected to small random perturbation.

1. Introduction

The relationship of molecular shape to the observed crystal packing arrangements is governed primarily by the Kitaigorodskii principle of close packing (Kitaigorodskii, 1973). Some basic rules for the packing of molecules in periodic arrays have been summarized more recently (Brock & Dunitz, 1994). It should be possible to derive more sophisticated packing rules as more data accumulates in the Cambridge Structural Database (hereafter, CSD; Allen & Kennard, 1993). Relationships between retained molecular symmetry and space-group selection have been identified in the CSD (Wilson, 1993) and in order to further such research, it is desirable to detect molecular symmetry from the given atomic coordinates. This paper reports an algorithm for this purpose.

Molecular symmetry does sometimes relate directly to the crystallographic space-group symmetry when considering the crystal packing. In such cases, an ordered symmetric molecule can be located on a special position in a space group, if the special position has a point group that either corresponds directly to the molecular point group or to a sub-group of the molecular point group. In these cases, the molecular symmetry and crystallographic symmetry coincide and the molecular point group must correspond exactly to at least the point group of the special position. If this is not the case, the molecule must be crystallographically disordered on that site. Studies of space group frequencies have shown that certain space groups occur more frequently when a particular one of its special positions is occupied. The most striking example being mirror symmetric space groups which never occur for ordered crystal structures unless the mirror plane is occupied by at least one residue (Wilson, 1993; Brock & Dunitz, 1994; Cole, 1995).

Most molecules, regardless of their molecular symmetry, are located in general positions in their crystal structures. It has been shown that molecular centres frequently lie at certain preferred general positions (Motherwell, 1997), with no obvious correlation with molecular symmetry.

It would be interesting to analyse the relationship between molecular point group, space-group selection and special position occupation in crystal structures. For this purpose a

Table 1
Topological index for each node in the given molecular graph of 13 atoms with the group partitions.

Nu is the number of unique index values $A(i,j)$ in the set of atoms $j = 1,13$. Ng is the number of partitions of atoms into topological groups based on the topological index $A(i)$ for cycle i of the application of the Morgan algorithm. The basic Morgan algorithm terminates at cycle 4 because Nu does not change. The extended algorithm terminates at cycle 6 because Ng does not change.

	1	2	3	4	5	6	7	8	9	10	11	12	13	Nu	Ng	Partitioned groups
$A(1)$	1	2	3	2	3	2	2	3	2	2	1	2	1	3	3	{3,5,8},{2,4,6,7,9,10,12},{1,11,13}
$A(2)$	2	4	6	6	6	5	5	6	6	4	2	4	2	4	4	{3,5,8},{4,9},{6,7},{2,10,12},{1,11,13}
$A(3)$	4	8	16	12	15	11	11	15	12	8	4	8	4	6	6	{3},{5,8},{4,9},{6,7},{2,10,12},{1,11,13}
$A(4)$	8	20	32	31	31	26	26	31	31	19	8	19	8	6	7	{3},{5,8},{4,9},{6,7},{2},{10,12},{1,11,13}
$A(5)$	20	40	82	76	76	57	57	76	63	39	19	39	29		8	{3},{5,8},{4,9},{6,7},{2},{10,12},{1},{11,13}
$A(6)$	40	102	166	158	159	133	133	159	158	95	39	95	39		8	

rapid method for analysing CSD entries for both exact and approximate molecular symmetry is needed. The method reported here is a combination of a two-dimensional topological symmetry check followed by three-dimensional symmetry perception. An alternative algorithm has been proposed recently by Ivanov & Schüürmann (1999). There are also similarities between our method of symmetry perception and that of Pilati & Forni (1998).

2. Methodology

2.1. Topological analysis

A molecule can be said to have a specific point group if every symmetry element within the point group maps each atom within the molecule either to itself or to another atom in a chemically identical environment. This requirement of atoms being in chemically identical environments gives rise to the first part of this perception algorithm. A topological analysis of the chemical diagram treated as a two-dimensional molecular graph is used to subdivide the molecule into groups of topologically identical atoms. Atoms in chemically identical environments must be topologically equivalent in the two-dimensional substructure and so the topological groups represent sets of atoms that can possibly be related by three-dimensional symmetry operations. The size of the largest topological group also indicates the maximum order of any symmetry element within the molecular point group with the noted exceptions of linear and planar molecules.

The topological analysis is applied using a variant of an algorithm due to Morgan (1965) which is commonly used for enumerating chemical graphs. The basic Morgan algorithm attempts to assign unique numbers to the atomic nodes of a graph by comparing the extended connectivity of each node in a stepwise manner. Its operation is exemplified in Table 1 for molecule (I). In the first cycle of the algorithm the j nodes of the graph are numbered arbitrarily, as in (I), and the number of connections, $A(1,j)$, to each of these nodes is then calculated as shown in the first line of Table 1. The number of unique $A(1,j)$ values, Nu, is also calculated. In subsequent cycles, i, values of $A(i,j)$ are calculated as the sum of the values of $A(i-1,k)$ for the set of nodes $\{K\}$ which are bonded to node j. The new value of Nu is calculated for each cycle. This process

is repeated until the value of Nu for the current cycle is less than or equal to the Nu value from the previous cycle, or until unique values of $A(i,j)$ are generated for every node in the graph, i.e. $Nu = j$. For molecule (I), the algorithm terminates after cycle 4 (Table 1, $Nu = 6$) and the extended connectivity values from the $(i-1)$th cycle are used to assign the canonical (Morgan) atom numbering: atom 3 in the original arbitrary scheme has the highest $A(3,j)$ value (16), and is assigned as atom 1 in the canonical scheme. The three atoms bonded to 3 are then renumbered as 2, 3 and 4 according to their $A(3,j)$ values, and so on.
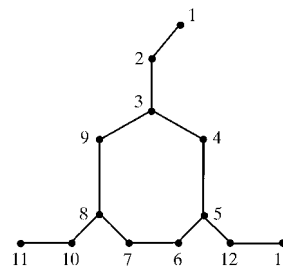
The important factor to note about the basic algorithm, as far as topological symmetry is concerned, is that atoms in identical environments cannot be distinguished. Thus, symmetrically related atoms, e.g. 4 and 9 in (I), will always be assigned the same $A(i,j)$ values, irrespective of how many cycles are carried out. However, the basic algorithm does not fully partition atoms into symmetry-related subsets. Thus, atoms 1, 11 and 13 have identical $A(3,j)$ values in Table 1, but only atoms 11 and 13 exist in identical environments, the $A(3,1)$ value being an accidental equivalence.

For the detection of topological symmetry, significant improvements can be obtained by using each cycle of the basic Morgan algorithm to index atoms in a hierarchical manner. At each step of the extended algorithm, nodes are partitioned into groups $\{l,m,n,...\}$ which have the same connectivity index $A(i,j)$, and the number of unique groups, Ng, is calculated, as illustrated in the rightmost columns of Table 1. This process is repeated on each cycle until all atoms are assigned to a unique group, $Ng = j$, or the cycle yields no further partitioning, $Ng(i) \leq Ng(i-1)$. For molecule (I), Table 1 shows that no further partitioning occurs at cycle 6 and the eight partitions of cycle 5 (or 6) reveal the eight topologically unique sets of atoms in the molecule.

Both the basic and extended Morgan algorithms can be improved further by incorporating chemical information for each node (atom) of the molecular graph, so that fewer steps are required to achieve complete partitioning. In the extended algorithm, the groups can be partitioned initially so that only atoms of the same type, *e.g.* same element symbol and primary connectivity, can be in the same group. Bond type information can also be used to weight the connectivity scores in the first cycle so as to obtain an improved primary partition.

2.2. Summary of procedure

2.2.1. Three-dimensional symmetry perception.
The topological analysis described above detects possible molecular symmetry in a two-dimensional graph and sub-divides the graph into groups of atoms that could possibly be related by a symmetry operation. This information is then passed to a geometrical rule-based approach which tests if a molecule possesses actual symmetry within certain user-defined tolerances. The three-dimensional symmetry perception algorithm works as follows.

(i) Transform the coordinate set so that the eigenvectors of the inertial tensor of the coordinate set coincide with the axis system. In lower-order point groups, symmetry elements coincide with the directions of these eigenvectors and will thus be detected in stage 2. In higher-order point groups, some or all of the eigenvectors are degenerate and lie in arbitrary directions.

If two eigenvalues are less than a given tolerance value, the distance between each atom and the principal inertial axis is calculated, and if the distances of all atoms from the axis are less than 0.1 Å the molecule is considered to be linear.

(ii) Test along the coordinate axes for symmetry of maximum order N; the method of symmetry testing being given in detail below. N is defined as the largest number of atoms in any topological group, or 2 if no topological group has greater than two atoms, which ensures that flat molecules are detected as having mirror symmetry. If all eigenvectors of the inertial tensor are non-degenerate within a given tolerance, the highest point symmetry possible is $D_{2h}(mmm)$ so only axes up to the order two are considered.

(iii) Calculate the centroids of all pairs of topologically equivalent atoms and test every origin–centroid direction for symmetry up to order N.

(iv) If any symmetry axes of the same type are found then check that they are not within an angular difference of one another. If they are, average the overall direction of the equivalent vectors and include only the operator that corresponds to the average direction.

(v) If (a) the maximum symmetry order detected so far is less than 3, (b) twofold axes are present within the final set and (c) the eigenvalues of the inertial tensor are threefold degenerate, then test the direction that corresponds to the sum of each selection of three axes from this subset. This test ensures that the threefold axes in T_h symmetric molecules are detected.

Table 2
The most commonly occurring molecular point groups in the CSD together with the number of symmetry operations, *Nmats*.

Each operation corresponds to a matrix derived by the program.

Nmats	Symmetry
1	$C_1(1)$
2	$C_2(2)$, $C_s(m)$, $C_i(\bar{1})$
3	$C_3(3)$
4	$C_{2v}(2mm)$, $C_{2h}(2/m)$, $D_2(222)$, $C_4(4)$, $S_4(\bar{4})$
5	$C_5(5)$
6	$C_{3v}(3m)$, $C_{3h}(3/m)$, $C_{3i}(\bar{3})$, $D_3(32)$, $C_6(6)$
7	$C_7(7)$
8	$D_{2h}(mmm)$, $D_{2d}(\bar{4}2m)$, $C_{4v}(4mm)$, $C_{4h}(4/m)$, $D_4(422)$, $C_8(8)$, $S_8(\bar{8})$
10	$C_{5v}(5m)$, $C_{5h}(5/m)$, $D_5(52)$, $C_{10}(10)$, $C_{5i}(\bar{5})$
12	$C_{6v}(6mm)$, $C_{6h}(6/m)$, $D_{3d}(\bar{3}m)$, $D_{3h}(\bar{6}m2)$, $D_6(622)$, $T(23)$, $C_{12}(12)$, $S_{12}(\bar{12})$
14	$C_{7v}(7m)$, $C_{7h}(7/m)$, $D_7(72)$, $C_{14}(14)$, $C_{7i}(\bar{7})$
16	$D_{4h}(4/mmm)$, $D_{4d}(\bar{8}2m)$, $C_{8v}(8mm)$, $C_{8h}(8/m)$, $D_8(822)$, $C_{16}(16)$, $S_{16}(\bar{16})$
20	$D_{5h}(5/m2m)$, $D_{5d}(\bar{5}m)$
24	$D_{6h}(6/mmm)$, $D_{6d}(\bar{12}2m)$, $T_h(m3)$, $T_d(\bar{4}3m)$, $O(432)$
28	$D_{7h}(7/m2m)$, $D_{7d}(\bar{7}m)$
32	$D_{8h}(8/mmm)$, $D_{8d}(\bar{16}2m)$
48	$O_h(m3m)$
60	$I(235)$
120	$I_h(m\bar{3}5)$

(vi) Expand the final set of operations by cross multiplying each operator with itself and other operators, and comparing the resultant matrices.

(vii) Use the trace and determinant of the operator matrices to identify the 'character' of the operator and then use this to generate a count of the operators observed, *Nmats*.

(viii) Compare this count with reference values of *Nmats* from point-group character tables to identify the molecular point group (see Table 2).

2.3. Symmetry testing

The following algorithm is applied to test an axis for symmetry. Firstly, we define two tolerances: TOL1: a geometric distance, set to 0.1 Å by default; TOL2: an angle, set to 5.0° by default. A similar distance tolerance definition can be found in the program *SYMMOL* of Pilati & Forni (1998).

(i) Generate the symmetry operations that correspond to a symmetry element of type S around the prescribed axis (see Giacovazzo, 1993).

(ii) Ignore atoms that are within TOL1 Å of the fragment centroid, since these must map back onto themselves chemically.

(iii) Apply each operator $\mathbf{g}$ in turn to every other atom $\mathbf{X_i}$ in the fragment, generating new coordinates $\mathbf{g(X_i)}$.

(iv) Loop through all the original coordinates to search for an atom $\mathbf{Y_i}$ (optionally of the same type or in the same two-dimensional topological group established by the extended Morgan algorithm) that lies within tolerances of the transformed coordinates.

Using these steps, the following geometrical quantities are calculated: (a) $d(\mathbf{Y_i})$ and $d\mathbf{g(X_i)}$, the distances of the atoms $\mathbf{Y_i}$

and the symmetry-derived positions $g(X_i)$ from the origin; (b) the distances $d(Y_i, g(X_i))$; (c) the angle $Ang(Y_i, g(X_i))$ subtended by Y_i and $g(X_i)$ at the origin. These values are all tested against the tolerances using equation (1)

$$|d(Y_i) - dg(X_i)| < f(TOL1) \text{ and}$$
$$Ang(Y_i, g(X_i)) < TOL2 \text{ or}$$
$$d(g(X_i), Y_i) < TOL1, \qquad (1)$$

where the first tolerance is a linear function, $f(TOL1)$ of the distance $d(Y_i)$, as depicted in (2). The variation of $f(TOL1)$ with $d(Y_i)$ is shown in Fig. 1.
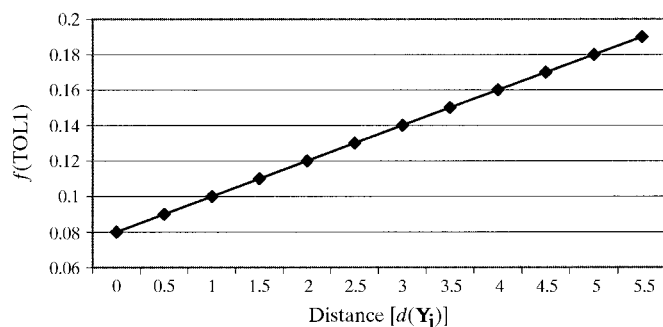
$$f(TOL1) = [(d(Y_i) - 1) \cdot TOL1 \cdot 0.2] + TOL1 \qquad (2)$$

From (1), two atomic positions are regarded as equivalent if either: (a) their distances from the origin are within $f(TOL1)$ of one another and the angle subtended at the origin by the atom pair is less than TOL2 or (b) the distance between the pair of atoms is less than TOL1.

This equivalence definition is used to minimize the effect of molecular size on the effectiveness of the algorithm. In a large molecule, two equivalent atoms on the periphery can be further apart than a pair of atoms near the molecular centroid, and yet still be approximately related by symmetry. Using the first tolerance, $f(TOL1)$, gives an unbiased estimate of such situations for all atoms that lie in shells at fixed distances increasing around the origin, which increases in thickness linearly with distance from the origin.

(v) If any atomic position fails the test in step 4 for any of the generated operators, then that symmetry element is rejected.

After all the possible operations are tested over all possible directions, the operator matrices for any symmetry elements detected in the molecule are derived. Finally, these operator matrices are expanded into a full set of operators for the point group by multiplying matrix pairs and storing any new matrices which are generated. A new matrix M_{new} is retained if no previously stored matrix M_{stored} can be found that has all elements $M_{stored}(i,j)$ within 0.1 of $M_{new}(i,j)$. Each new matrix is subsequently multiplied by both original and other new matrices in the same way. This process may introduce symmetry elements that were not detected in steps 4 and 5 because they failed the tolerance tests. However, it is neces-

sary to include all operators found in the multiplication step so that the full set comprises a point group.

For each new matrix found, the rotation angle $360°/n$ is derived along with the rotation vector. The matrix is idealized to correspond to an exact rotation by regenerating the matrix from the derived rotation vector and the rational rotational angle closest to the derived angle according to certain criteria (i.e. the axis order must be consistent with the operators already found). This avoids the propagation of cumulative errors in the multiplication procedure. Thus, a complete set of matrices representing all possible symmetry operations existing in a molecule is derived.

2.4. Assignment of point group

Using the method given above, the symmetry elements of any molecule can be detected. Hence, the molecular point group can be determined from the derived number of symmetry operations and a set of transformation matrices.

There is a characteristic number of symmetry operations for each point group, e.g. for a simple C_n point group there are $n - 1$ rotational symmetry operations and the identity operator, so that n discrete symmetry operations are involved in this point group. Table 2 lists the number of symmetry operations in point groups which occur most frequently in observed molecular structures. Moreover, each matrix represents a symmetry operation by a specific symmetry element. Any atom, when transformed by this matrix, should result in another equivalent atom in the molecule. If a symmetry element, for instance a twofold axis, is coincident with an axis of the defined coordinate system, the matrix usually has a simpler form that is directly related to the rotation angle $360°/n$. It is $180°$ for the twofold axis and 120, $240°$ for the threefold axis, and so on. However, if the symmetry element is not directly referred to its principal axes of inertia x, y and z, the matrix can be expressed in a more general form (Giacovazzo, 1993). The matrices can be derived from the vectors corresponding to the symmetry elements.

The symmetry elements can be identified from certain characteristics of the matrices, i.e. there is a characteristic value for the trace of a matrix for a particular rotation and the determinant of a matrix will have a value of +1 or −1 depending on whether the operation is a proper or improper (inversion) axis, respectively (see *International Tables for Crystallography*, 1992). The trace of a matrix T for a symmetry operation can be expressed as

$$T = 1 + 2\cos[m(2\pi)/n], \qquad (3)$$

where n is the order of the symmetry operation, $m < n$; m and n are integers without common factors (except 1).

From the number and nature of the derived matrices, a molecular point group can be assigned to each molecule under investigation. For example, if n symmetry elements are detected and all matrices represent one rotation axis (principal axis), the point group is C_n. Similarly, when there are n reflection planes accompanying a principal axis, the molecule is assigned as C_{nv}. All the basic principles for the assignment of

Figure 1
Graph of the tolerance function $f(TOL1)$ (Å) versus $d(Y_i)$ (Å), the distance of atom Y_i from the origin at the molecular centroid.

point groups can be applied, so that when the number of matrices generated indicates more than one possible point group, it is important to locate particular special symmetry elements which are characteristic of a specific point group. For example, when 24 symmetry elements are detected, there are five possible point groups (Table 2): D_{6h}, D_{6d}, T_h, T_d and O. Here D_{6h} and D_{6d} can be distinguished from the others by the presence of a sixfold axis and the inversion centre distinguishes D_{6h} from D_{6d}. Of the remaining three, all have multiple threefold and $\bar{4}$ or fourfold axes, but the presence of an inversion centre **i** will indicate T_h and mirror planes will distinguish T_d from O.

3. Implementation in *RPLUTO*

Molecular symmetry detection for CSD entries has been implemented in *RPLUTO*,[1] a program for the visualization of molecules and crystals.

Several important enhancements to *RPLUTO* have resulted from recent research activities (Motherwell *et al.*, 1999). One of many features in *RPLUTO* is the display of planes or vectors, defined by clicking on relevant atoms, and the visualization of molecular symmetry elements has been developed using the vector and plane graphic display objects. The assigned point-group and the major symmetry elements of the point group are displayed superimposed on the molecule, with colour coding for the symmetry types. Two examples are shown in Fig. 2. Structure BIPHEN08 (Badour *et al.*, 1986) has D_{2h} symmetry and BEYXUY (Lindeman *et al.*, 1982) is D_3.

For crystal structures with multiple molecules and ions, symmetry detection is carried out on each molecule in turn or on one molecule or ion selected by the user. The program indicates whether the molecular fragment is a linear molecule or a single atom in these cases.

H atoms are ignored by default, but there is an option to include them where necessary for correct symmetry detection (*e.g.* where the molecule with H atoms removed has a higher symmetry). Symmetry is sometimes broken by the H atoms and these atoms are in practice often inaccurately positioned in X-ray crystal structures.

The symmetry matrices represent transformations of sets of topologically equivalent atoms onto their equivalent sites, within certain practical tolerances for distances and angles given above. In general, the default values, TOL1 = 0.1 Å for distance and TOL2 = 5.0° for angles, work well for most structures. However, for some poorly refined structures or structures refined in an incorrect space group, a lower symmetry than expected may be observed. An interesting example is the structure AACRUB, originally determined in space group Cc (Bino *et al.*, 1979). Our algorithms indicated a C_2 symmetric cation $(C_8H_{18}O_{10}Ru_2)^+$ and C_s for the symmetric BF_4^- anion. If the distance tolerance is slightly increased, D_{4h} symmetry can be detected for the cation. For the re-refinement of the same structure in space group $C2/c$ (AACRUB01;

[1] *RPLUTO* may be downloaded for non-commercial research purposes from www.ccdc.cam.ac.uk/prods/rpluto.

Table 3
Validation results for randomly perturbed structures.

The number of atoms N_{atom} and the maximum distance MD_{cen} (Å) of an atom in the structure from the molecular centroid of the reference structure are also given (see text, D_{max} is the maximum displacement randomly generated).

Model structures			Symmetry found			
Symmetry	N_{atom}	MD_{cen} (Å)	D_{max} (Å)	0.125	0.100	0.075
C_i	12	3.13	C_i	551	841	989
			C_1	449	159	11
C_3	13	2.60	C_3	770	976	1000
			C_1	230	24	
C_{4h}	9	2.92	C_{4h}	874	990	1000
			C_{2h}	91	8	
			C_s	35	1	
			C_1	0	1	
C_{2v}	11	3.03	C_{2v}	931	998	1000
			C_s	63	1	
			C_2	2		
			C_1	4	1	
C_{5v}	11	2.60	C_{5v}	736	978	1000
			C_5	9		
			C_s	193	19	
			C_1	62	3	
D_{2h}	13	4.50	D_{2h}	931	997	1000
			C_{2v}	69	3	
D_{7d}	15	2.69	D_{7d}	1000	1000	1000
O_h	13	3.00	O_h	999	1000	1000
			D_{4h}	1		

Marsh & Schomaker, 1981), the expected D_{4h} symmetry is detected for the cation and T_d for the anion. The inversion centre of each cation lies on the crystallographic inversion centre of $C2/c$ and one of the twofold axes in T_d coincides with the twofold axis in $C2/c$.

RPLUTO may be used automatically to process a retrieved subset of CSD entries giving an output file of the perceived molecular symmetry. The output file includes the CSD refcode, residue number, the detected molecular point group, space group, Z and Z' values. In addition, *RPLUTO* will read data in CIF, *SHELX* or free-format (orthogonal or fractional coordinates).

4. Validation

When applying the algorithm to detect approximate molecular symmetry with the observed molecules recorded in the CSD, it is important to assess the reliability of the results. It can be seen from AACRUB, discussed above, that deviations of atoms from their idealized/correct positions can result in the assignment of symmetry that is lower than expected.

In order to assess the extent of molecular deformation which the algorithm can tolerate, reference structures in representative point groups, *i.e.* ideal molecules possessing exact symmetry, were subjected to random shifts in the three Cartesian coordinates (x, y, z) of each atom, such that the total displacement of each atom did not exceed a maximum displacement randomly generated, D_{max}. For each reference structure, 1000 deformed structures were generated for a given D_{max} and the symmetry detection applied. The results are summarized in Table 3. These show that for displacements

less than TOL1 (0.1 Å), the symmetry of the ideal reference structure is found in all cases except C_i. For displacements comparable to the tolerance limit, lower symmetries are detected in < 10% of the perturbed structures for most cases. No correlation was found between the point group detected and either the RMS (root mean square) distortion of the molecule from the original reference or the maximum displacement of any atom in the molecule, *i.e.* a lower symmetry was not always detected for the molecules which have the largest RMS deviations. This is probably because the detection of symmetry is dependent on the correlation of displacements of one or more pairs of atoms. If sets of atoms are displaced in the same direction, symmetry may be main-

tained (*e.g.* along the fivefold axis direction in C_{5v}). This occurs most frequently in lower-symmetry structures; in higher-symmetry point groups, even if some of the symmetry elements are not detected initially they may be generated in the matrix multiplication step. In C_{5v}, for example, detection of two mirror planes approximately 72° apart will lead in the expansion step to the full set of operators for C_{5v}. This is not possible in C_i and C_3, with only one rotation or rotation–inversion axis, if no symmetry is found in the first detection step.

Examination of Table 3 shows that in structures containing an inversion centre, this symmetry is more easily destroyed by perturbation compared with other symmetry elements. The results from the tests should also be dependent on the shape and size of selected molecules, but it is difficult to assess these factors due to the difficulty of constructing similar molecules with different symmetries but comparable numbers of atoms.

5. Conclusions

A combination of two-dimensional topological graph analysis and three-dimensional perception in a symmetry detection program provides a useful tool for the recognition of molecular symmetry in experimental molecular structures. The validation analysis using molecules with small perturbations from exact symmetry in most chemically relevant point groups shows that the expected symmetry can be found correctly in most cases using acceptable tolerance values. Detection of higher molecular symmetries seems to be more stable than for lower symmetries within the perturbation tests.

A systematic study of molecular symmetry derived using this method for CSD structures in specific space groups is now in hand and will be presented in a future publication.
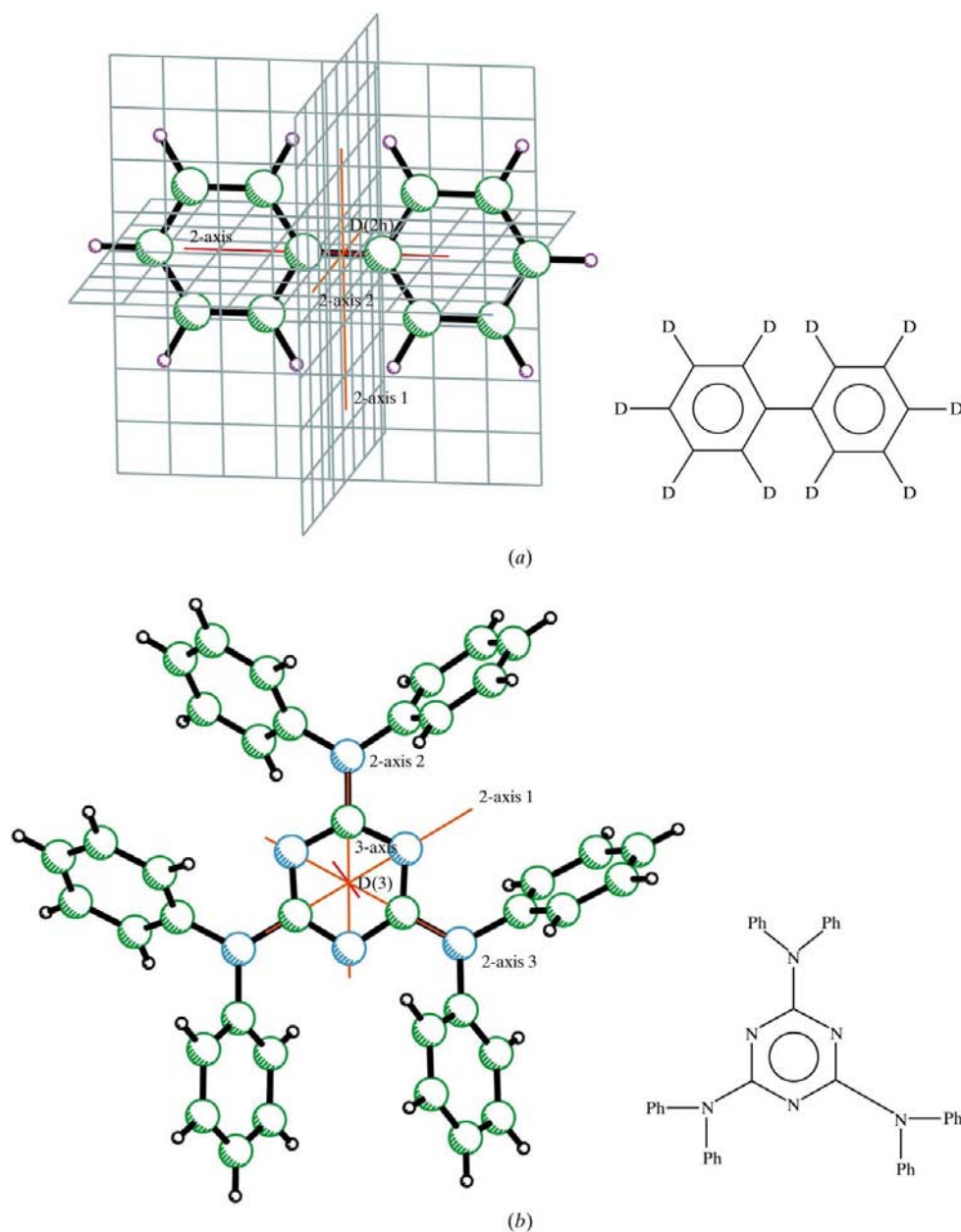
Figure 2
Three-dimensional display of molecules with detected symmetry elements in the *RPLUTO* program. The chemical diagram is also printed here for: (a) BIPHEN08, detected as D_{2h}, space group $P2_1/a$, viewed from one of the detected twofold rotation axes. Mirror planes are shown in grey. (b) BEYXUY, detected as D_3, space group $R\bar{3}c$, viewed from the detected threefold rotation axis.

References

Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 1, 31–37.

Baudour, J. L., Toupet, L., Delugeard, Y. & Ghemid, S. (1986). *Acta Cryst.* C**42**, 1211–1217.

Bino, A., Cotton, F. A. & Felthouse, T. R. (1979). *Inorg. Chem.* **18**, 2599–2604.

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

Cole, J. C. (1995). Ph.D Thesis. University of Durham, England.

Giacovazzo, C. (1993). *Fundamentals of Crystallography*, p. 71. Oxford University Press.

Ivanov, J. & Schüürmann, G. (1999). *J. Chem. Inf. Comput. Sci.* **39**, 728–737.

Kitaigorodskii, A. I. (1973). *Molecular Crystal and Molecules.* New York: Academic Press.

Lindeman, S. V., Shklover, V. E., Struchkov, Yu. T., Kuznetzov, S. N. & Pankratov, V. A. (1982). *Kristallografiya*, **27**, 65–73.

Marsh, R. E. & Schomaker, V. (1981). *Inorg. Chem.* **20**, 299–303.

Morgan, H. L. (1965). *J. Chem. Documt.* **5**, 107–113.

Motherwell, W. D. S. (1997). *Acta Cryst.* B**53**, 726–736.

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

Pilati, T. & Forni, A. (1998). *J. Appl. Cryst.* **31**, 503–504.

Wilson, A. J. C. (1993). *Acta Cryst.* A**49**, 795–806.