Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Automated assignment of graph-set descriptors for crystallographically symmetric molecules

W. D. Samuel Motherwell,* Gregory P. Shields and Frank H. Allen

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England

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

Algorithms for the automatic assignment of graph-set notation for intermolecular networks have been extended to molecules having internal crystallographic symmetry, for patterns up to the second level. This provides a means of achieving systematic and consistent assignments for networks containing symmetric molecules. These methodologies have been implemented in the program *RPLUTO*. Examples are given of the application of the method to a number of molecules with hydrogen-bonded and other intermolecular networks, illustrating the diversity of the patterns that occur.

Received 28 October 1999
Accepted 31 January 2000

1. Introduction

The graph-set approach to describing intermolecular networks in terms of **C**, **R**, **S** and **D** patterns is now well established (Bernstein *et al.*, 1995). This approach was initially suggested by Kuleshova & Zorkii (1980) and developed in its present form by Etter and Bernstein (Etter, 1990; Etter *et al.*, 1990). The graph-set method was originally applied to hydrogen-bonded networks, particularly in cases of polymorphism (Bernstein, 1991; Bernstein *et al.*, 1990, 1995; Blagden *et al.*, 1998). However, it is equally applicable to any directional intermolecular contact, *i.e.* those which can be considered as comprising a 'donor' and 'acceptor' (*e.g.* $\mathrm{Sb}\cdots\mathrm{Cl}$, but not $\mathrm{I}\cdots\mathrm{I}$ interactions; *e.g.* Starbuck *et al.*, 1999).

Subsequently, the mathematical foundation for this type of analysis was established, placing the procedure on a rigorous graph-theoretical footing (Bernstein *et al.*, 1997; Grell *et al.*, 1999). This introduced a number of new concepts. In particular, a constructor graph may be used to describe the networks, in which molecules are collapsed to points and only the intramolecular bonds (H edges) are shown, labelled as **a**, **b**, **c**. Paths may then be described by a directed label sequence, *e.g.* $\overrightarrow{a}\ \overrightarrow{a}\ \overrightarrow{b}\ \overleftarrow{a}\ \overleftarrow{a}\ \overleftarrow{b}$, and this description is unique (a significant labelling) where there is only one H edge associated with a given molecule having a particular directed label. In practice, this consideration is obeyed where no molecules or atoms lie on crystallographic special positions, *i.e.* the molecules are crystallographically asymmetric. Grell *et al.* (1999) showed that in order to achieve a significant labelling where molecules have crystallographic symmetry, it was necessary to describe the structure in a lower-symmetry space group; in particular, the group with only translational symmetry elements maintained.

Originally, the assignment of graph-set descriptors was undertaken by visual inspection of the intermolecular network, which is a tedious process that is prone to human error. As a result, we have recently applied the concepts of Grell *et al.* (1999) in the development of computer algorithms for assigning graph sets (up to the second level, *i.e.* involving one or two symmetry-independent H edges) for asymmetric molecules (Motherwell *et al.*, 1999). This was implemented in the molecular visualization program *RPLUTO*.[1]

In this paper, we describe how our approach has been extended to molecules with internal crystallographic symmetry, by considering the possible relative symmetry relationships between H edges having the same label, without the need for a transformation to a lower-symmetry normal subgroup. This allows possible paths to be generated in a systematic manner for symmetric, as well as for asymmetric molecules, and has also been implemented in the *RPLUTO* program. We have included a number of examples in this paper (Table 1); additional examples and references for the Cambridge Structural Database (CSD) entries described are available as supplementary material.[2] These assignments may be verified by the reader using the *RPLUTO* program.

2. First-level graph sets

2.1. Simple first-level patterns

Consider a structure containing a type of molecule or ion **A**, with no internal symmetry, associated with one type of H edge a, *i.e.* molecule $\mathbf{A}_1$ (Fig. 1*a*) which is involved in only one unique interaction. The H edge a^{I} (the ^{I} labelling is arbitrary), starting at a donor atom D_1, has an associated symmetry operator, termed the contact symmetry operator S_c. This operator defines the transformation of an acceptor X_1 in the same asymmetric unit as the donor D_1 into an acceptor position X_2 at which this acceptor forms a contact with the donor D_1. In the absence of internal symmetry, there is only one H edge which enters the molecule and one which leaves it (Fig. 1*a*). In this case, there is only one possible path, $\overrightarrow{a}$, through the network (*i.e.* $\overrightarrow{a}^{\,I}$, $\overrightarrow{a}^{\,II}$ *etc.* in Fig. 1*a*). Successive acceptors (*e.g.* X_1 and X_2) and donors (*e.g.* D_1 and D_2) in the path are related by the contact symmetry operator S_c, and successive molecules are also related by this operator, since both the donor and acceptor D_1 and X_1 are in the same asymmetric unit of the molecule $\mathbf{A}_1$.

The situation is more complicated when molecules possess internal symmetry, since more than one H edge $\overrightarrow{a}$ may both enter molecule $\mathbf{A}_1$ and leave molecule $\mathbf{A}_1$ (*i.e.* more than one H edge may emanate from molecule $\mathbf{A}_1$ in the same direction, *e.g.* $D \rightarrow X$). As a result, the acceptor X through which the path $\overrightarrow{a}$ enters $\mathbf{A}_1$ may be in a different asymmetric unit from the donor D through which the path $\overrightarrow{a}$ leaves $\mathbf{A}_1$. The possible paths with a period of 1 comprise successive H edges for which the donor and acceptor are related to those in a previous edge by the same symmetry operator for each edge. In order to derive all such paths, it is sufficient to consider just one of the symmetry-related H edges $\overrightarrow{a}^{\,I}$, starting from a donor D_{11} at the identity operator, as the H edge leaving a given molecule $\mathbf{A}_1$ at the start of the path (Fig. 1*b*). A path may enter $\mathbf{A}_1$ from the acceptor X_{11} in the same asymmetric unit as D_{11}, or a different asymmetric unit (*e.g.* X_{12} or X_{13}). In the latter case, it is necessary to take into account the internal symmetry operator S_i which relates the acceptor X_{13} (for a path entering *via* edge a^{II}) to the acceptor X_{11} in the same asymmetric unit as D_{11}. Successive acceptors X_{13} and X_{23} in the path, and hence donors D_{11} and D_{21} and molecules $\mathbf{A}_1$ and $\mathbf{A}_2$, are related by the combined operator $S_t = S_cS_i$. All possible paths may be found by considering all internal symmetry operators S_i, including the identity, of the special equivalent position on which the molecule or ion **A** lies.

Where the donor or acceptor atom itself lies on a special position, certain symmetry operators will generate the same donor site (*e.g.* if molecule $\mathbf{A}_1$ in Fig. 1*b* lies on a mirror plane), but a different resulting operator S_t, such that the overall path is not necessarily the same. Thus, all internal symmetry operators must be considered, even if they generate the same atom position. However, if all donor and acceptor atoms in a ring path lie on a mirror plane, paths may be generated for which the apparent size of the ring is doubled, since each atom may be considered twice, with mirror-related operators, if the order of the rotation axis is odd (*e.g.* threefold axis). Such spurious rings may be eliminated by ensuring that the atomic coordinates for the second half of the ring are not simply a repeat of those in the first half, within a suitable tolerance value which determines when atom sites are considered identical. Where donors and/or acceptors do not actually lie in the mirror plane, such rings would represent distinct paths.

As for the case of molecules without internal symmetry (Motherwell *et al.*, 1999), the nature of the motif (**R** or **C**) for symmetric molecules may be identified by considering the nature of the operator S_t relating successive molecules linked by the H edge. In some instances, this procedure will generate paths which are symmetrically equivalent, but these are retained at this stage. The number of donors and acceptors and the path repeat unit (period) may be derived in the same way as for asymmetric molecules. However, the covalent distance table (Grell *et al.*, 1999) must be modified to provide the distances between acceptor and donor, and *vice versa*, for each internal symmetry operator S_i.

1,4-Dibenzoic acid (terephthalic acid, TEPHTH), which lies on an inversion centre, provides a good example of this situation (Fig. 2, Table 1). If the path involves a donor (HO) and acceptor (O) in the same asymmetric unit, *i.e.* the internal symmetry operator is C_1, and the covalent distance is four bonds, then since the contact symmetry operator is $C_i + t$, the resulting motif is a centrosymmetric **R**2,2(8) dimer. Alternatively, if the internal symmetry operator S_i is C_i (*i.e.* D and X are in different asymmetric units related by an inversion centre), the path enters and leaves the molecule from different halves, with a covalent distance of nine bonds, and the resultant operator S_t is translation t producing a **C**1,1(9) chain. Similarly, where there is more than one molecule involved in an H edge a, the pattern designator is necessarily **D** (discrete), since even if there is more than one symmetry-equivalent H

---

[1] This program is available free of charge for non-commercial use and may be downloaded from the CCDC website at http://www.ccdc.cam.ac.uk/prods/rpluto.

[2] Supplementary data for this paper are available from the IUCr electronic archives (Reference: HA0194). Services for accessing these data are described at the back of the journal.

edge emanating from the same molecule, it necessarily does so in the same direction.

This treatment has also been applied to intramolecular (self, **S** motifs), rather than simply taking the shortest path between donor and acceptor as in the asymmetric case. Consider a molecule $\mathbf{B}_1$ in Fig. 1(*c*), with three symmetry-equivalent intramolecular H edges $\overrightarrow{a}^{i}$, $\overrightarrow{a}^{ii}$ and $\overrightarrow{a}^{iii}$. A path $\overrightarrow{a}^{i}$ may pass from donor D_{11} to the acceptor X_{11} in the same asymmetric unit, having started at this acceptor X_{11}, giving a simple intramolecular **S**1,1(*n*) ring (here the contact operator S_c is the identity, but this is not necessarily so). Alternatively, the path could have started from acceptors X_{12} or X_{13}, *i.e.* at acceptors related by an internal symmetry operator S_{i2} and S_{i3} to acceptor X_{11}. Combining the contact operator S_c and the internal symmetry operators, again gives the overall operator S_t relating successive acceptors and donors in the path. For molecule $\mathbf{B}_1$, operators S_{i2} and S_{i3} lead to the paths $\overrightarrow{a}^{i}\overrightarrow{a}^{iii}\overrightarrow{a}^{ii}$ and $\overrightarrow{a}^{i}\overrightarrow{a}^{ii}\overrightarrow{a}^{iii}$, respectively. N{CH[C(O)OMe][NHC(O)Ph]}$_3$, YIRYAZ (Fig. 3, Table 1), provides a good example of this situation, the molecule lying on a threefold axis and exhibiting the chemically relevant **S**3,3(12) amide ring motif. Similarly, the cation in HIWQEJ exhibits an **S**3,3(6) (O—H$\cdots$)$_3$ ring motif (Table 1).

2.2. Complex first-level patterns

In the absence of internal symmetry, a path of the type ($\overrightarrow{a}\ \overleftarrow{a}$) is not possible, since it would require the same H edge to be traced consecutively, in opposite directions (*e.g.* $\overrightarrow{a}^{i}\ \overleftarrow{a}^{i}$ in Fig. 1*a*). We previously chose not to consider such paths (Motherwell *et al.*, 1999). The condition that a path should not follow the same H edge more than once consecutively is considerably less stringent than the rule suggested by Bernstein *et al.* (1997) that only *M-simple* paths, passing through each molecule once, should be considered. However, we have found that the latter rule excludes motifs which would appear to be intuitive from a chemical standpoint, so we have adopted the rule that paths may not retrace the same H edge consecutively. Note that a path may thus pass through the same donor or acceptor atom more than once, *via* a different H edge, and such a path may represent the symmetry of a pattern more effectively.

However, for internally symmetric molecules, such ($\overrightarrow{a}\ \overleftarrow{a}$) type paths are possible, since the path may leave *via* an edge which is symmetry related to that through which it entered (Table 1). We have considered the simplest of such

Table 1
Graph-set assignments for selected CSD structures.

Contacts are N/O—H$\cdots$N/O hydrogen bonds unless otherwise indicated. References for the CSD RefCodes are included in the supplementary data for this paper, which are available from the IUCr electronic archives (Reference: HA0194). Services for accessing these data are described at the back of the journal.

(*a*) One independent H edge

	$\overrightarrow{a}$	$\overrightarrow{a}\ \overleftarrow{a}$
YIRYAZ	**S**1,1(8) **S**3,3(12) **S**3,3(24)	
HIWQEJ	**S**3,3(6) **S**1,1(4) **S**3,3(12)	
TETSSU (Te$\cdots$O)	**C**1,1(4)	**C**2,1(8)
PAPSII	**C**1,1(2) **C**1,1(8)	**C**2,2(16)
JOFHER	**C**1,1(4) **C**1,1(6)	**R**2,2(12)
YUZPAK	**C**1,1(4)	**R**2,2(16)
PELBIR	**C**1,1(2) **C**1,1(8)	**C**2,2(16) **R**6,6(48)
KELCAF	**R**4,4(12) **R**4,4(20)	**R**2,2(10)
RUHWUM	**R**2,2(22) **R**6,6(30) **R**6,6(66)	**C**2,2(22) **R**6,6(66)
TEPHTH	**C**1,1(9) **R**2,2(8)	**C**2,2(18)
BASVUM01	**C**1,1(6) **R**6,6(42)	**C**2,2(16)
PERYTO03	**C**1,1(6) **R**4,4(8) **R**2,2(12) **R**4,4(24)	**C**2,2(12) **R**4,4(24) **R**8,8(48)
FETRUR	**D**1,1(2)	
HIMGAL	**D**1,1(2)	**D**2,2(9)
BOCNEM	**D**1,1(2)	**C**1,2(8)
JASXUW	**D**1,1(2)	**R**1,2(7)
TAMGUC01	**D**1,1(2)	**C**1,2(6) **R**3,6(18)

(*b*) Two independent H edges

	$\overrightarrow{a}$	$\overrightarrow{a}\ \overleftarrow{a}$	$\overrightarrow{b}$	$\overrightarrow{b}\ \overleftarrow{b}$	$\overrightarrow{a}\ \overrightarrow{b}$	$\overrightarrow{a}\ \overleftarrow{b}$
MERYOL03	**C**1,1(2) **R**4,4(28)	**C**2,2(14)	**C**1,1(5) **R**4,4(8)	**C**2,2(10)	**C**2,2(11) **R**2,2(11) **R**4,4(20) **R**4,4(24)	**C**2,2(10) **C**2,2(11) **C**2,2(12) **R**4,4(24)
YISTEZ	**D**1,1(2)	**D**2,2(10)	**D**1,1(2)	**D**2,2(8)	**C**2,2(6) **C**2,2(11)	
HIWQEJ (OH$\cdots$F)	**D**1,1(2)	**C**2,2(8)	**D**1,1(2)	**C**2,2(8)		**C**2,1(4) **C**2,2(8) **R**6,3(12)
ZUKKAR	**D**1,1(2)	**C**2,2(30)	**D**1,1(2)	**C**2,2(30)	**C**2,2(8) **R**2,2(30) **R**4,4(38)	
AMTBTZ (NH$\cdots$Cl)	**D**1,1(2)	**R**1,2(8) **R**2,4(16)	**D**1,1(2)	**C**1,2(8) **R**2,4(16)		**D**2,2(5) **D**2,2(9)
VOBXEP	**D**1,1(2)	**R**1,2(10)	**D**1,1(2)	**C**2,2(12)		**D**2,2(5) **D**2,2(11)
YINJOU	**D**1,1(2)	**D**2,2(13)	**C**1,1(10) **R**2,2(8)	**C**2,2(20) **R**8,8(80)	**C**($\overleftarrow{b}\ \overrightarrow{a}\ \overrightarrow{b}$)2,3(16) **C**($\overleftarrow{b}\ \overrightarrow{a}\ \overrightarrow{b}$)3,3(32)	
SAKYOS10	**R**6,6(12)		**D**1,1(2)	**D**2,2(5)	**C**($\overleftarrow{a}\ \overrightarrow{b}\ \overrightarrow{a}$)3,3(8)	
HMTAAB	**D**1,1(2)		**D**1,1(2)	**C**2,2(6) **R**6,6(18)		**C**2,2(6)

Table 1 (continued)

(*c*) Three independent H edges

	$\overrightarrow{a}$	$\overrightarrow{a}\ \overleftarrow{a}$	$\overrightarrow{b}$	$\overrightarrow{b}\ \overleftarrow{b}$	$\overrightarrow{c}$	$\overrightarrow{c}\ \overleftarrow{c}$
BUYSET10	**C**1,1(7)		**D**1,1(2)	**D**2,2(7)	**C**1,1(6)	
VOJFEF	**D**1,1(2)	**D**2,2(23)	**D**1,1(2)	**D**2,2(23)	**D**1,1(2)	**D**2,2(21)

	$\overrightarrow{a}\ \overrightarrow{b}$	$\overrightarrow{a}\ \overleftarrow{b}$	$\overrightarrow{a}\ \overrightarrow{c}$	$\overrightarrow{a}\ \overleftarrow{c}$	$\overrightarrow{b}\ \overrightarrow{c}$	$\overrightarrow{b}\ \overleftarrow{c}$
BUYSET10	**C**($\overrightarrow{a}\ \overrightarrow{a}\ \overleftarrow{b}$)3,3(17)		**C**2,2(7)	**C**2,2(13)	**C**($\overrightarrow{c}\ \overrightarrow{b}\ \overleftarrow{b}$)3,3(14)	
VOJFEF		**D**1,2(3) **D**2,2(23)	**D**2,2(6) **D**2,2(22)		**R**2,2(8) **R**4,4(48) **R**8,8(96)	

paths, those which have a period of two H edges, both where $\overrightarrow{a}$ associates crystallographically equivalent molecules $\mathbf{A}_1$, $\mathbf{A}_2$ *etc.* (Fig. 1*b*) or different molecules $\mathbf{A}_1$ and $\mathbf{B}_1$ (Fig. 1*d*). In the latter case, the first edge $\overrightarrow{a}^{\,\mathrm{I}}$ links D_{11} (in $\mathbf{B}_1$) and X_{12} (in $\mathbf{A}_1$) *via* operator S_c (which relates X_{12} to X_{11}, the position of X when it is transformed by the same operator as D_1 relative to the original coordinates; while not in the same molecule, X_{11} and D_{11} are in the same asymmetric unit). The previous edge $\overleftarrow{a}$ in the path may enter the molecule $\mathbf{B}_1$ through either $\overleftarrow{a}^{\,\mathrm{II}}$ or $\overleftarrow{a}^{\,\mathrm{III}}$, *i.e. via* donor D_{12}, or D_{13}, related to D_{11} by an internal symmetry operator S_{ib} (Fig. 1*d*). Here it is necessary to reject the identity, and any other operators which do not affect the position D_{11}, where D_{11} itself is on a special position, in order to avoid retracing the same edge $\overleftarrow{a}^{\,\mathrm{I}}$ in the other direction. Similarly, the internal symmetry operator S_{ia} relating the acceptor atom X_{12}, through which $\overrightarrow{a}^{\,\mathrm{I}}$ enters $\mathbf{A}_1$, with the acceptor atom X_{13} or X_{11}, *via* which the path leaves by H edge $\overleftarrow{a}^{\,\mathrm{V}}$ or $\overleftarrow{a}^{\,\mathrm{V}}$, respectively. Again, it is necessary to reject the identity, and any other operators which do not affect the position of the acceptor X. The repeat operator S_t generating the path $\overleftarrow{a}^{\,\mathrm{II}}\ \overrightarrow{a}^{\,\mathrm{I}}\ \overleftarrow{a}^{\,\mathrm{V}}$ *etc.* is that relating D_{13} and D_{52} (or equivalently X_{21} and X_{13}). This is the combination of four operators $S_t = S_cS_{ia}S_cS_{ib}$, and determines whether the motif is **R**, **C** or **D**. The method is the same when molecule **A** is crystallographically identical to **B**, successive

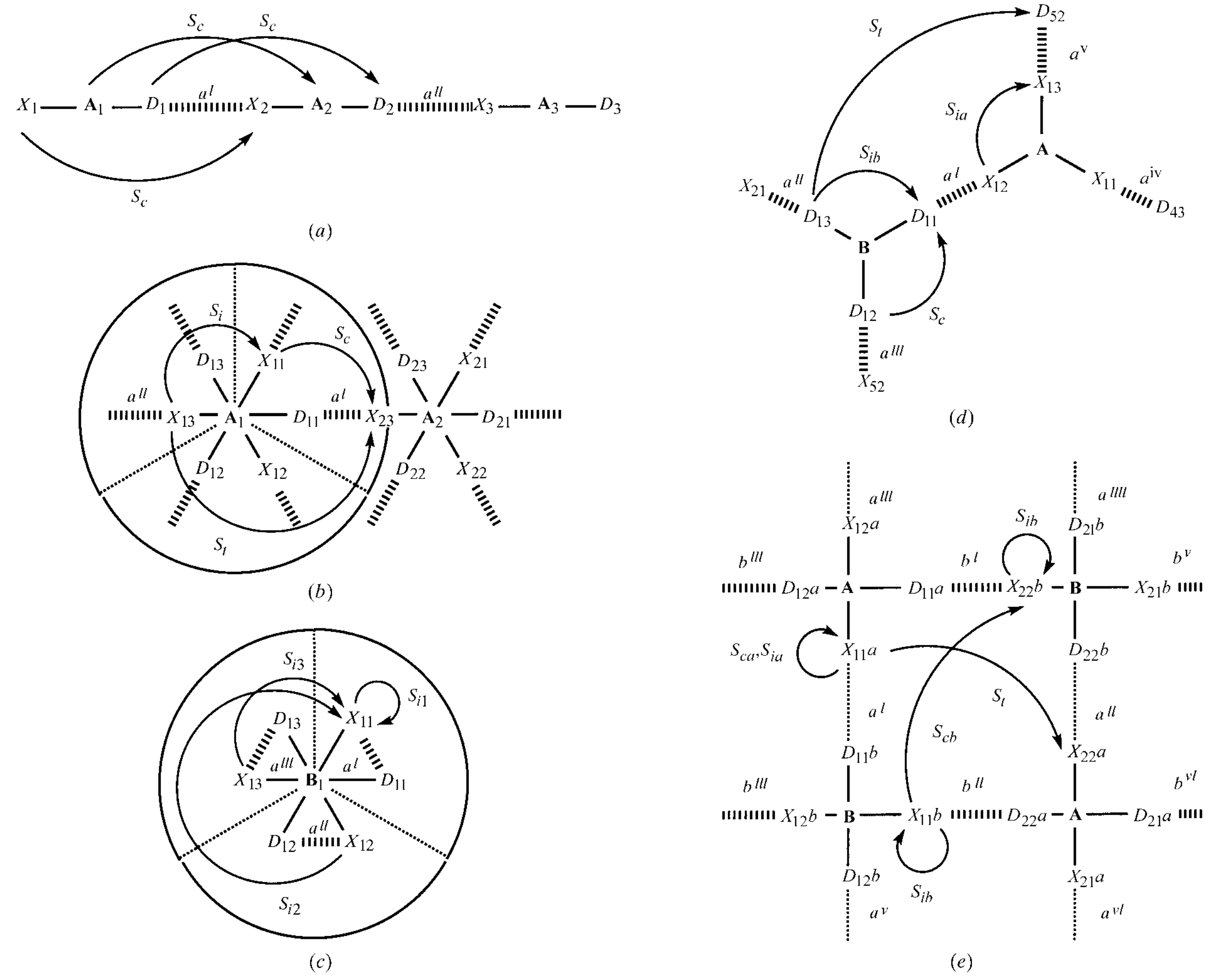

Figure 1
Schematic representations of paths in structures composed of (*a*) asymmetric and (*b*)–(*e*) symmetric molecules; in (*b*) and (*c*), the segments enclosed by circles represent asymmetric units.

molecules in the path having only acceptors or only donors involved in the path.

Where there are two crystallographically independent molecules $\mathbf{A}_1$ and $\mathbf{B}_1$, it may not be possible to start a path with both $(\overrightarrow{a})$ and $(\overleftarrow{a})$ H edges, *i.e.* from a contact $X\cdots D$ as well as $D\cdots X$, since one of the molecules may only be associated with one H edge $(\overrightarrow{a})$ if it lies on a general position, or the donor or acceptor atom lies on the same special position as the molecule as a whole, giving a $\mathbf{D}(\overleftarrow{a}\ \overrightarrow{a})$ or $\mathbf{D}(\overrightarrow{a}\ \overleftarrow{a})$ pattern, but not both.

A number of possibilities exist for $(\overrightarrow{a}\ \overleftarrow{a})$ patterns where $(\overrightarrow{a})$ is discrete (**D**). If both donor and acceptor themselves lie on the same special position as their respective molecules (but not a lower-symmetry special position), no $(\overrightarrow{a}\ \overleftarrow{a})$ pattern exists (*e.g.* FETRUR, Table 1), as in the asymmetric case. If one of the molecules lies on a general position, or either the acceptor or the donor (but not both) lies on the same special position as the molecule to which it belongs, there is only one H edge $\overrightarrow{a}$ or $\overleftarrow{a}$ emanating from one of the molecules, and the composite motif is also discrete, $\mathbf{D}(\overrightarrow{a}\ \overleftarrow{a})$ {*e.g.* [1,3,5-$(OH)_3C_6H_3$].3[4-MeC_5H_4N], HIMGAL, Fig. 4}. Otherwise, where both molecules have more than one H edge $\overrightarrow{a}$ or $\overleftarrow{a}$, **R** (*e.g.* JASXUW), **C** (*e.g.* BOCNEM) or both **R** and **C** {*e.g.* $[C(NHNH_2)_3]^+.Cl^-$, TAMGUC01, Fig. 5) may exist.

For consistency, we have also considered complex first-level $(\overrightarrow{a}\ \overleftarrow{a})$ paths where the motifs $(\overrightarrow{a})$ are **R** and/or **C**. There are a large number of possible combinations of **R** and **C** for $(\overrightarrow{a})$ and $(\overrightarrow{a}\ \overleftarrow{a})$ motifs. For example, all the $(\overrightarrow{a})$ and $(\overrightarrow{a}\ \overleftarrow{a})$ motifs may be chains [*e.g.* TETSSU, formed *via* Te$\cdots$O contacts, and PAPSII, Table 1, comprising symmetry-related parallel (O—)H$\cdots$O(—H) chains]; all the $(\overrightarrow{a})$ motifs may be chains and all the $(\overrightarrow{a}\ \overleftarrow{a})$ may be rings (*e.g.* JOFHER) or all the $(\overrightarrow{a})$ may be chains and $(\overrightarrow{a}\ \overleftarrow{a})$ may be rings and chains (*e.g.* PELBIR). Similarly, all the $(\overrightarrow{a})$ and $(\overrightarrow{a}\ \overleftarrow{a})$ may be rings (*e.g.* KELCAF), $(\overrightarrow{a})$ **R** and $(\overrightarrow{a}\ \overleftarrow{a})$ **R** and **C** (*e.g.* RUHWUM). Alternatively, $(\overrightarrow{a})$ may be **R** and **C** and $(\overrightarrow{a}\ \overleftarrow{a})$ **C** (this is commonly observed for dicarboxylic acids and dioximes lying on C_i or C_2 special positions; *e.g.* [1,4-$(COOH)_2C_6H_4$], TEPHTH, Fig. 2; BASVUM01, Table 1), or $(\overrightarrow{a}\ \overleftarrow{a})$ also **C** and **R** (*e.g.* PERYTO03).

Which of these possibilities is adopted depends on the symmetry elements relating the molecules as well as on the internal molecular symmetry. We have not found examples where $(\overrightarrow{a})$ are **R** and $(\overrightarrow{a}\ \overleftarrow{a})$ **C** or where $(\overrightarrow{a})$ are **R** and **C** and $(\overrightarrow{a}\ \overleftarrow{a})$ are all **R** in the compounds studied so far. Currently, we have not applied this treatment to complex intramolecular motifs (self, **S** motifs), concentrating on the intermolecular patterns which are more important in describing network architectures.

3. Second-level graph sets

3.1. Simple second-level patterns

Simple second-level patterns with a period of two H edges, *i.e.* $(\overrightarrow{a}\ \overleftarrow{b})$ and $(\overrightarrow{a}\ \overrightarrow{b})$ paths, may be derived in a similar manner to first-level patterns, *e.g.* Fig. 1(*e*) where different molecules **A** and **B** each have crystallographic symmetry. Paths of the type $(\overrightarrow{a}\ \overrightarrow{b})$ are evaluated by repeating the process described for $(\overrightarrow{a})$ motifs for the second H edge. The resulting operator which relates acceptors and donors in successive H edges of the same type is $S_t = S_{ca}S_{ib}S_{cb}S_{ia}$ (*e.g.* for the path $\overrightarrow{b}^{\,\imath}\ \overrightarrow{a}^{\,\imath}\ \overrightarrow{b}^{\,\imath\imath}\ \overrightarrow{a}^{\,\imath\imath}$, S_{ca}, S_{ia} and S_{ib} are the identity, so $S_t = S_{cb}$ relates successive acceptors X_{11a} and X_{22a}). Similarly, $(\overrightarrow{a}\ \overleftarrow{b})$ patterns are found in the same way as $(\overrightarrow{a}\ \overleftarrow{a})$ motifs, except in this case the identity operator is allowed, since the H edges to which the donors (or acceptors) in the same molecule belong are different. The resulting operator is the same as for $(\overrightarrow{a}\ \overrightarrow{b})$ patterns. Again, each internal symmetry operator is considered which relates the asymmetric unit in which one H edge enters the molecule $\mathbf{A}_1$ to the asymmetric unit from which the next H edge in the path leaves molecule $\mathbf{A}_2$, and similarly for the other molecule **B**.

We have not yet extended this treatment to second-level patterns incorporating **S** motifs. Where **S** motifs are combined with intermolecular patterns, they simply provide an alternative intramolecular path, and thus offer little insight beyond the first-level motifs themselves. Similarly, combinations of **S** motifs are rarely useful in providing a clear description of intramolecular contacts.

Where both intermolecular H edges *a* and b involve the same molecule **A**, the $\overrightarrow{a}$, $\overrightarrow{a}\ \overleftarrow{a}$, $\overrightarrow{b}$ and $\overrightarrow{b}\ \overrightarrow{b}$ patterns are all **C** or **R**, both $\overrightarrow{a}\ \overrightarrow{b}$ and $\overrightarrow{a}\ \overleftarrow{b}$ second-level patterns exist and they themselves are **C** and/or **R** (*e.g.* butane-1,2,3,4-tetrol, MERYOL03, Fig. 6). If two independent molecules **A** and **B** are involved in $\overrightarrow{a}$ and $\overrightarrow{b}$, *i.e.* both $\overrightarrow{a}$ and $\overrightarrow{b}$ are discrete, a number of possibilities exist for the $\overrightarrow{a}\ \overrightarrow{b}$ and $\overleftarrow{a}\ \overrightarrow{b}$ patterns. Where $\overrightarrow{a}\ \overleftarrow{a}$ and $\overrightarrow{b}\ \overleftarrow{b}$ are also **D** (see above), $\overrightarrow{a}\ \overrightarrow{b}$ or $\overrightarrow{a}\ \overleftarrow{b}$ (but not both, since each of *a* and *b* may only leave each of the molecules in one direction) will exist as **R** and/or **C** (*e.g.* YISTEZ). The situation is similar if $\overrightarrow{a}\ \overleftarrow{a}$ and $\overrightarrow{b}\ \overleftarrow{b}$ are **R** and/or **C**, and the second-level patterns of type $\overrightarrow{a}\ \overrightarrow{b}$ or $\overrightarrow{a}\ \overleftarrow{b}$ (depending on whether the H edges between molecules **A** and **B** are in the same, or opposing, directions) may be **R** and/or **C** (*e.g.* ZUKKAR, HIWQEJ, Table 1). As discussed for first-level motifs, an $\overrightarrow{a}\ \overleftarrow{a}$ motif may not exist (*e.g.* HMTAAB). Alternatively, one $\overrightarrow{a}$ may be **D** and $\overrightarrow{b}$ **C** or **R**, or *vice versa* (*e.g.* YINJOU, SAKYOS10, Table 1).

If there are three molecules involved, and both $\overrightarrow{a}$ and $\overrightarrow{b}$ are **D** motifs associated with a central symmetric molecule **A**, either $\overrightarrow{a}\ \overrightarrow{b}$ or $\overrightarrow{a}\ \overleftarrow{b}$ patterns exist and will all be **D** {*e.g.* $[N(CH_2CH_2NH_3)_3]^{3+}.3[ClO_4]^-$, VOBXEP, Fig. 7; AMTBTZ, Table 1}.

3.2. Composite second-level patterns

Where both $\overrightarrow{a}\ \overrightarrow{b}$ and $\overleftarrow{a}\ \overrightarrow{b}$ patterns exist and are of type **D**, conventionally these are combined to form the complex graph set **bab**, with degree $n_c = n_{\overrightarrow{a}\,\overrightarrow{b}} + n_{\overleftarrow{a}\,\overrightarrow{b}} - 2$, $d_c = d_{\overrightarrow{a}\,\overrightarrow{b}} + d_{\overleftarrow{a}\,\overrightarrow{b}} - 1$ and $a_c = a_{\overrightarrow{a}\,\overrightarrow{b}} + a_{\overleftarrow{a}\,\overrightarrow{b}} - 1$, since the contact atom pair $X\cdots D$ occurs in both **ab** and **ba**. Such composite $\overleftarrow{b}\ \overrightarrow{a}\ \overrightarrow{b}$ motifs may be derived in a similar manner for molecules with internal symmetry. However, there is a further complication if pattern $\mathbf{D}(\overleftarrow{b}\ \overrightarrow{b})$ also exists, since the composite path is in fact a chain, $\mathbf{C}(\overleftarrow{b}\ \overrightarrow{a}\ \overrightarrow{b})$. The period of the chain is the sum of the periods of $\mathbf{D}(\overleftarrow{b}\ \overrightarrow{a}\ \overrightarrow{b})$ and $\mathbf{D}(\overleftarrow{b}\ \overrightarrow{b}) - 4$ (two of these four represent atoms replicated in the patterns, together with two atoms at the end of the pattern which represent atoms in the adjacent repeat units of the chain). Similarly, $N_d = N_{d(\overleftarrow{b}\,\overrightarrow{a}\,\overrightarrow{b})} + N_{d(\overleftarrow{b}\,\overrightarrow{b})} - 2$ and $N_a = N_{a(\overleftarrow{b}\,\overrightarrow{a}\,\overrightarrow{b})} + N_{a(\overleftarrow{b}\,\overrightarrow{b})}$ (since one donor and one acceptor overlap between the $\overleftarrow{b}\ \overrightarrow{a}\ \overrightarrow{b}$ and $\overleftarrow{b}\ \overrightarrow{b}$ motifs, and one donor and one acceptor represent atoms in the adjacent chain repeat units).

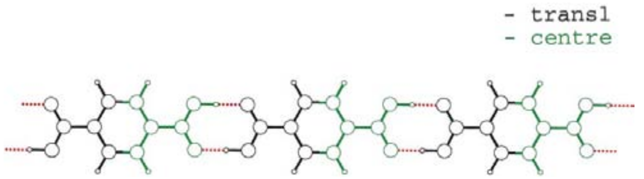

Figure 2
Centrosymmetric molecules of 1,4-dibenzoic acid, TEPHTH, showing the chain-of-rings first-level motif, coloured according to the type of operator relating atoms to the base asymmetric unit: green for inversion and black for translation.

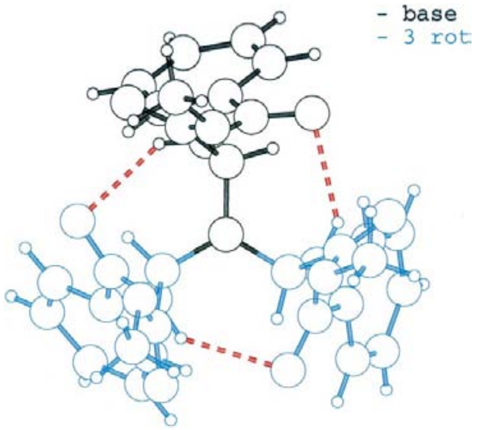

Figure 3
Molecule of YIRYAZ, showing the intramolecular **S**3,3(12) ring; cyan atoms are related to atoms in the base asymmetric unit (black) by a threefold rotation axis.

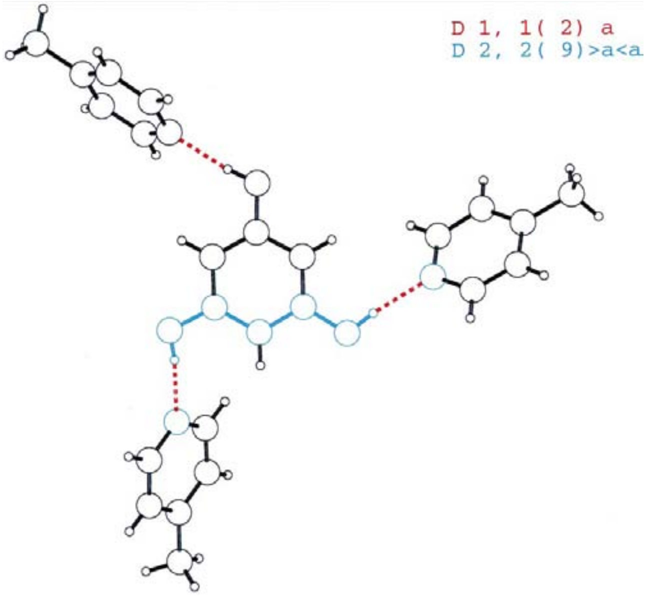

Figure 4
Structure of HIMGAL, highlighting the complex first-level **D**2,2(9) pattern in blue.

The structure $2[C_{11}H_{11}O_7].[C_4H_4O_3]$, BUYSET10, provides a good example of a $\mathbf{C}(\overleftarrow{b}\ \overrightarrow{a}\ \overrightarrow{b})$ type pattern (Fig. 8). One molecule lies on an twofold axis and is hydrogen-bonded to an asymmetric molecule *via* a **D**2,2(7) motif. The asymmetric molecule forms chains through two independent hydrogen bonds; each of these forms a composite **C**3,3 pattern in combination with the **D**2,2(7) motif. Similarly, in YINJOU (Table 1) there are **D**2,3(7) and **D**3,3(23) $\overleftarrow{b}\ \overrightarrow{a}\ \overrightarrow{b}$ patterns, which, when combined with the $\overrightarrow{a}\ \overleftarrow{a}$ **D**2,2(13) motif, give overall **C**2,3(16) and **C**3,3(32) chain paths, respectively, while in SAKYOS10 (Table 1), the combined path is simpler, namely **C**3,3(8).

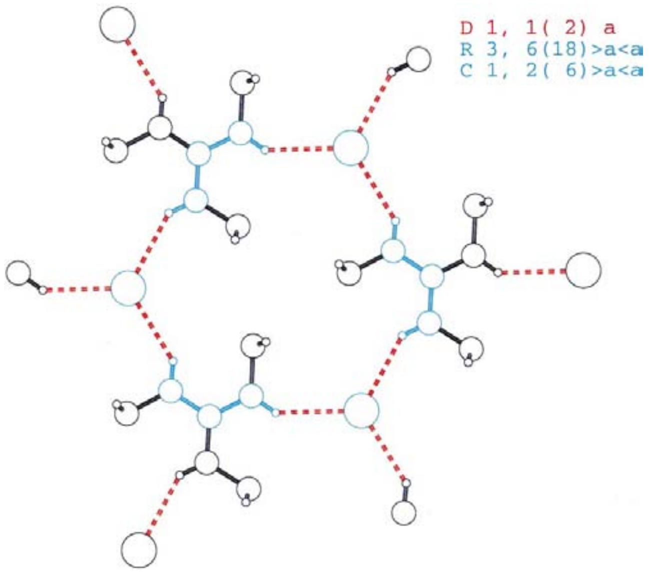

Figure 5
The complex first-level **R**3,6(18) ring (blue atoms) in the structure of TAMGUC01.

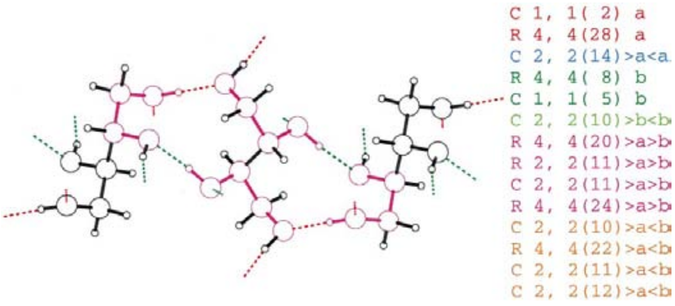

Figure 6
Hydrogen-bonded structure of butane-1,2,3,4-tetrol, MERYOL03, showing the second-level **R**4,4(20) ring in magenta.

4. Graph-set matrix

Initially, the graph sets for all paths of type $\overrightarrow{a}$, *i.e.* period 1, are summarized in one matrix, and those of period 2 in another ($\overrightarrow{a}\ \overrightarrow{b}$ in the upper right, $\overrightarrow{a}\ \overleftarrow{b}$ in the lower left, and $\overrightarrow{a}\ \overleftarrow{a}$ on the leading diagonal; $\overrightarrow{a}\ \overrightarrow{a}$ paths are not currently considered). This includes any symmetry-equivalent paths found in the analysis. The graph sets are subsequently analysed for composite **D** and **C** patterns as described above. The conventional lower-triangular graph-set matrix is derived using the same rules as for asymmetric molecules, graph sets being included in the priority **C** > **R**, **S** > **D**, with **R**, **S** motifs being included in square parentheses where **C** patterns take precedence (Motherwell *et al.*, 1999). Table 2 gives these matrices for VOJFEF (Table 1) with three independent H edges.

5. Visualization

As implemented in the *RPLUTO* program, each symmetry-inequivalent graph set is included in the colour key ($\overrightarrow{a}$ being represented as a>, *etc.*), with the exception of second-level **D**($\overrightarrow{a}\ \overrightarrow{b}$) and **D**($\overrightarrow{a}\ \overleftarrow{b}$) motifs, which are replaced by the corresponding composite patterns, or omitted altogether if **C**,**R**($\overrightarrow{a}\ \overrightarrow{b}$ / $\overleftarrow{b}$) patterns also exist. Motifs may be highlighted as described previously, any atom forming a symmetry-related path being highlighted, by selecting the graph set from the key. Alternatively, just the atoms forming one symmetry-unique **R**, **S** or **D** pattern, or the basic repeat unit for **C** patterns, may be shown, the molecules comprising this path being selected automatically. This provides a convenient means of visualizing a single instance of the particular path. As a result, extra link atoms are no longer added for atoms comprising shortest intramolecular paths in adjacent molecules. Graph sets may be derived in the current version for up to 24 independent H edges.

Table 2
Graph-set matrices for VOJFEF.

(*a*) Initial graph set matrix: period 1 patterns

a	b	c
D1,1(2)	**D**1,1(2)	**D**1,1(2)

(*b*) Initial graph set matrix: period 2 patterns

	a	b	c
a	**D**2,2(23)		**D**2,2(6) **D**2,2(22)
b	**D**1,2(3) **D**2,2(23)	**D**2,2(23)	**R**2,2(8) **R**4,4(48) **R**8,8(96)
c			**D**2,2(21)

(*c*) Final graph set matrix

	a	b	c
a	**D**2,2(23)		
b	**D**1,2(3)	**D**2,2(23)	
c	**D**2,2(6)	**R**2,2(8)	**D**2,2(21)

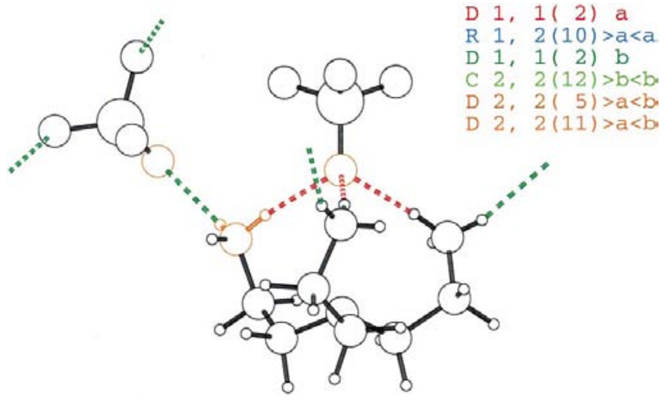

Figure 7
Hydrogen-bonded network in VOBXEP, highlighting the second-level **D**2,2(5) pattern in orange.

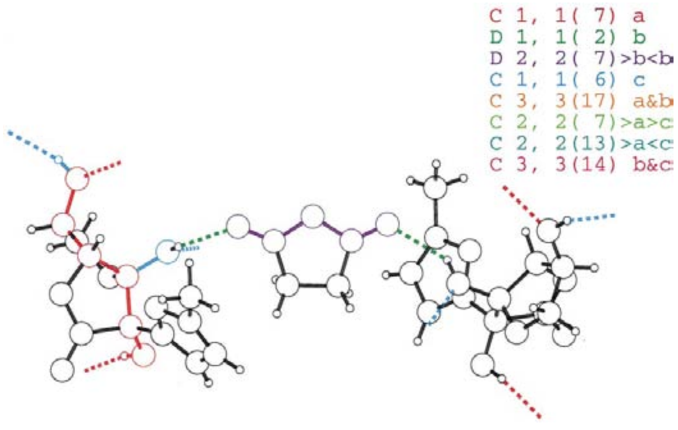

Figure 8
Portion of the hydrogen-bonded network in BUYSET, emphasizing the first-level **C**1,1(7) (red) and **C**1,1(6) (blue) chains in the asymmetric molecule, and **D**2,2(7) motif (violet) linking these chains *via* the symmetric succinic anhydride moiety.

6. Conclusions

Our algorithm for the automatic computation of graph-set descriptors, up to the second level, has been extended to symmetric molecules. As a result, graph sets can be assigned for the majority of networks linked by directional (donor–acceptor) interactions, being no longer limited to the asymmetric molecule case. This has been achieved without transformation to a lower-symmetry crystallographic sub-group in order to achieve a unique numbering of the constructor graph, by considering the internal symmetry of the molecules explicitly. The routine availability of this procedure, albeit only up to the second level, should extend the range of applicability of the graph-set method to further chemical systems, with applications in crystal structure comparison, prediction and modelling.

References

Bernstein, J. (1991). *Acta Cryst.* B**47**, 1004–1010.

Bernstein, J., Davis, R. E., Shimoni, L. & Chang, N.-L. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 1555–1573.

Bernstein, J., Etter, M. C. & MacDonald, J. M. (1990). *J. Chem. Soc. Perkin Trans. II*, pp. 695–704.

Bernstein, J., Ganter, B., Grell, J., Hengst, U., Kuske D. & Pöschel, R. (1997). Technische Universität Dresden Technical Report MATH-AL-17-1997.

Blagden, N., Davey, R. J., Liebermann, H. F., Williams, L., Payne, R., Roberts, R., Rowe, R. & Docherty, R. (1998). *J. Chem. Soc. Faraday Trans.* pp. 1035–1044.

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

Etter, M. C., MacDonald, J. C. & Bernstein, J. (1990). *Acta Cryst.* B**46**, 256–262.

Grell, J., Bernstein, J. & Tinhofer, G. (1999). *Acta Cryst.* B**55**, 1030–1043.

Kuleshova, L. N. & Zorkii, P. M. (1980). *Acta Cryst.* B**36**, 2113–2115.

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

Starbuck, J., Norman, N. C. & Orpen, A. G. (1999). *New J. Chem.* **23**, 969–972.