Acta Crystallographica Section B
Structural Science
ISSN 0108-7681

Persistent hydrogen bonding in polymorphic crystal structures

Peter T. A. Galek,[a,b]* László Fábián[a,b] and Frank H. Allen[a,b]

[a]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, England, and [b]Pfizer Institute for Pharmaceutical Materials Science, Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, England

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

Received 11 July 2008
Accepted 11 December 2008

The significance of hydrogen bonding and its variability in polymorphic crystal structures is explored using new automated structural analysis methods. The concept of a *chemically equivalent* hydrogen bond is defined, which may be identified in pairs of structures, revealing those types of bonds that may *persist*, or not, in moving from one polymorphic form to another. Their frequency and nature are investigated in 882 polymorphic structures from the Cambridge Structural Database. A new method to compare conformations of equivalent molecules is introduced and applied to derive distinct subsets of conformational and packing polymorphs. The roles of chemical functionality and hydrogen-bond geometry in persistent interactions are systematically explored. Detailed structural comparisons reveal a large majority of persistent hydrogen bonds that are energetically crucial to structural stability.

1. Introduction

Structural polymorphism describes permutations in the mode of packing of equivalent covalently bonded units in the crystalline state. A polymorph represents one of a set of alternative crystalline phases; polymorphs frequently display contrasting thermochemical and mechanical properties such as melting and sublimation temperature, solubility and dissolution rate, heat capacity, conductivity and hygroscopicity (Haleblian & McCrone, 1969; Haleblian, 1975; Giron, 1995; Threlfall, 1995). Such variety promotes a necessarily high interest from many fields including the pharmaceutical, dyestuffs, food and energetic materials (*e.g.* explosives) industries (Bernstein, 1993; Grant, 1999).

In the pharmaceutical industry in particular, as many as half of all crystallized drug substances exist in multiple polymorphic forms (Henck *et al.*, 1997) and thus identification of the most desirable formulation of a drug candidate forms a key stage in product development (Singhal & Curatolo, 2003). For a given crystalline candidate, polymorph screening and selection techniques attempt to find all potential forms and to continue with the most appropriate, based on its materials properties (Stahly, 2007). Owing to the time-intensive nature of the screening process, assistance is increasingly being sought in the form of informatics-based analysis (Chisholm *et al.*, 2006). For the crystalline state, *structural informatics* offer collective knowledge built using analysis of databases such as the Cambridge Structural Database (CSD; Allen, 2002), which contains structural information for more than 423 000 small molecule organic and metal-organic crystal structures (Version 5.29, November 2007).

Hydrogen bonding is crucial to structural stability as it provides strong, consistent intermolecular interactions, and

plays a central role in the occurrence of polymorphism (Bernstein, 1993). A knowledge-based method, named the Logit Hydrogen Bonding Propensity (LHP) model, to predict the propensity for individual hydrogen-bonding interactions has been presented recently (Galek *et al.*, 2007). It has been shown that hydrogen bonding in CSD structures is highly consistent and predictable in individual examples. However, due to the variation in packing, polymorphic structures can exhibit highly contrasting hydrogen bonding. Although polymorphic structures represented in the CSD are relatively rare (3.31%), their number is sufficient to allow assessment of underlying trends using a statistical approach. This study will therefore investigate how much variation exists between one polymorphic form and another, and will examine what characteristics might underpin these structural similarities and diversities.

2. Methodology

2.1. Definition: persistent hydrogen bonds

Polymorphic crystal structures exhibit multiple alternative packing arrangements of comparative stability at a given temperature and pressure. In the present work, the structural variation encountered requires a flexible description of a hydrogen-bonding interaction that may be identified in diverse crystal forms. The concept of *chemically equivalent* hydrogen bonds will be used. Chemically equivalent bonds are formed between donor–acceptor pairs of the same chemical type. This definition includes hydrogen bonds that are symmetry-equivalent in one structure, between equivalent atoms of a compound in multiple structures, and also between non-equivalent atoms, but that are formed by equivalent chemical groups. The flexibility of this description is that the varied potential chemical nature of a hydrogen bond is described without location-specific details. A *persistent* hydrogen bond may thus be defined as a hydrogen bond for which a chemically equivalent bond is identified in multiple polymorphic forms. This definition can then be used to assess pairs of polymorphic structures for the extent of similar and dissimilar interactions.

Technically, each hydrogen bond is described by its donor and acceptor atoms, their functional group membership and the hydrogen-bond type (inter- or intramolecular). Functional group membership is determined using a substructure-search algorithm (Bruno *et al.*, 2002) and a library of 85 functional groups most commonly found in organic compounds (listed in the supplementary material;[1] Galek *et al.*, 2007). Once a donor or acceptor atom is identified, a group is selected based on matching atom type, and the extent of matched atoms, bond descriptions and other constraints (*e.g.* cyclic/acyclic atom constraints may have been specified in the substructure). An observed hydrogen bond may then be represented as a character string:

```
D (Functional Group) [H — Bond Type]
A (Functional Group)                    (1)
```

To illustrate how hydrogen bonding can persist or not, a dimorphic system is selected as an example from the CSD, refcodes FAZTAC and FAZTAC01. The structures are visualized in Fig. 1 using *MERCURY* (Macrae *et al.*, 2008). FAZTAC has $Z' = 1$, whereas FAZTAC01 has $Z' = 2$; the two molecules in the asymmetric unit of the latter differ by a rotation about the aniline–thiadiazine bond. This conformational change is associated with alternative intermolecular interactions, however, a degree of similarity also persists. In both dimorphs a hydrogen bond exists between the secondary amine and the carbonyl group of the pyrazolone ring. An intramolecular bond is also common to both forms, between the NH of the thiadiazine ring and the pyrazolone carbonyl.

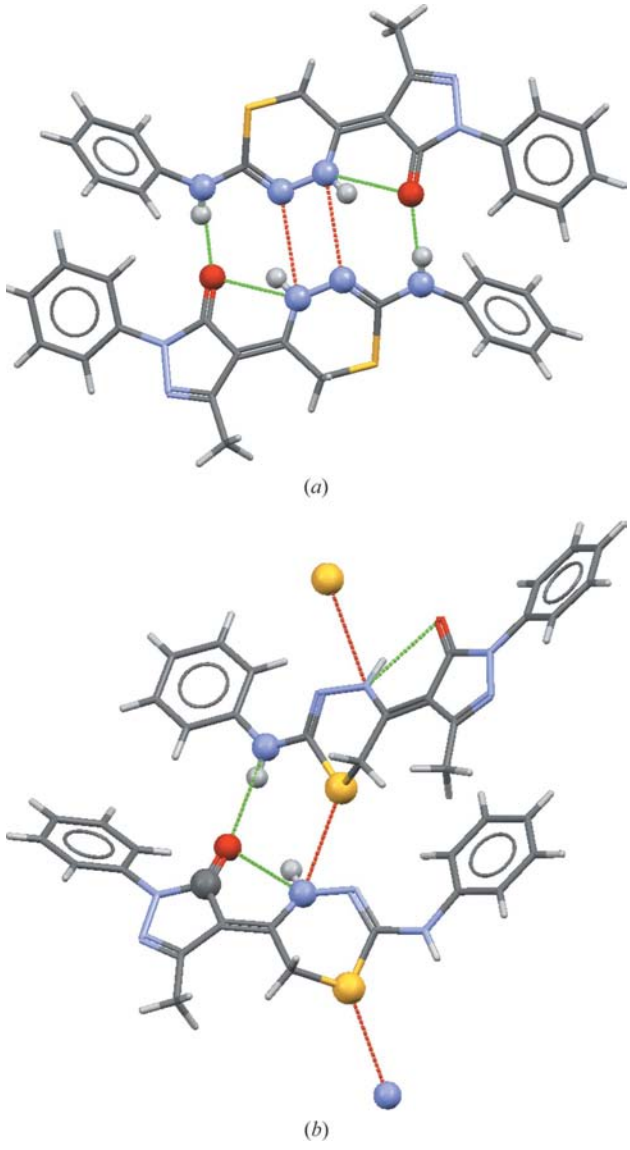


Figure 1
Hydrogen bonding in (*a*) FAZTAC and (*b*) FAZTAC01 crystal structures (Liu *et al.*, 2004). Hydrogen bonds that persist between forms are coloured green, unique hydrogen bonds to a form are coloured red.

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: BK5078). Services for accessing these data are described at the back of the journal.

One unique hydrogen bond also exists in each form: in FAZTAC the thiadiazine NH group donates to an N acceptor of the equivalent ring of a second molecule, and in FAZTAC01 the same donor group forms a bond with the sulfur acceptor. In summary, two persistent hydrogen bonds are found in the system and one unique bond is found in each polymorph.

In a minority of cases, chemically equivalent hydrogen bonds are found despite their specific donor–acceptor atoms being unique to one polymorph. The refcodes CUYQUI and CUYQUI01 are illustrated in Fig. 2. The compound has three independent hydroxyl groups; OH···OH hydrogen bonds are formed between different pairs in the two polymorphs. Despite this *topological* difference in hydrogen bonding, we consider the chemical nature of the interactions to be equivalent. Neither form displays any of three potential hydroxyl–furan hydrogen bonds, which would be considered a significantly different alternative. It is noted that polymorphic compounds such as this, which present multiple equivalent donors/acceptors, are found amongst particular chemical classes, most notably carbohydrates and sugars. Such examples are not common as polymorphic structures in the CSD [*e.g.* a search for polymorphic carbohydrates with three-dimensional coordinates returned just 12 compounds (26 structures)], out of more than 800 compounds, as described below in §2.4.

The persistence of chemically equivalent hydrogen bonds has been demonstrated, notwithstanding a large change in crystal structure; however, interactions unique to one form nonetheless occur alongside. By investigating the frequency and underlying features of these hydrogen bonds that may or may not persist, we hope to learn about their effects on crystal formation, and ultimately to better appreciate the occurrence of polymorphism.

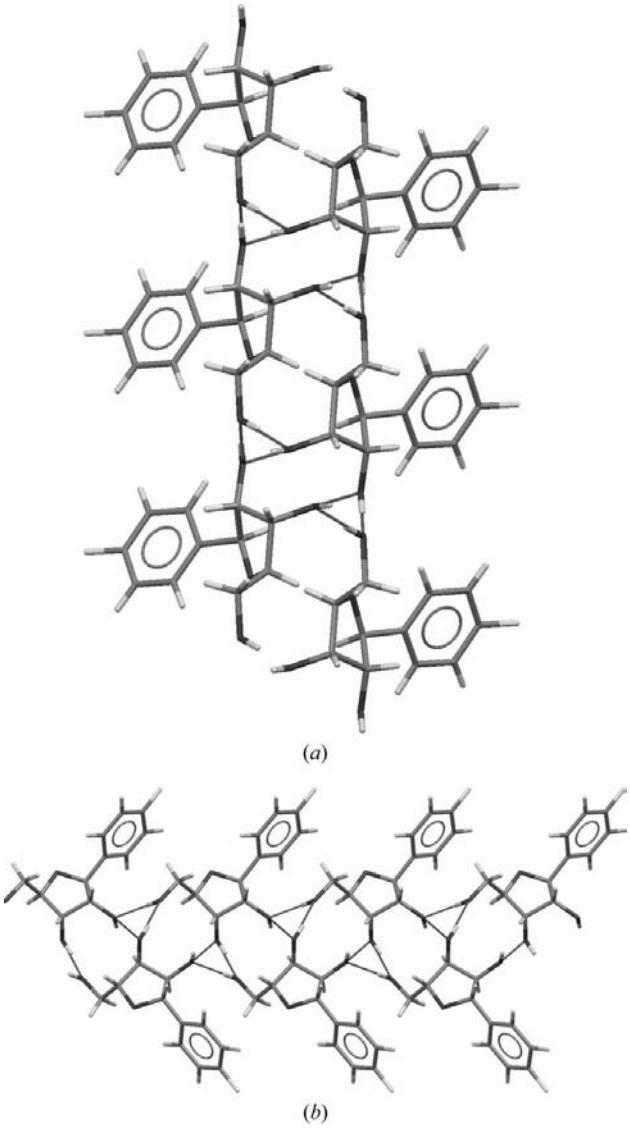


Figure 2
Hydrogen bonding in (*a*) CUYQUI, viewed down the crystallographic *b* axis, and (*b*) CUYQUI01, viewed down the crystallographic *a* axis (Bats *et al.*, 2000). Both structures display three chemically equivalent hydroxyl–hydroxyl hydrogen bonds which form threefold spiral catemers perpendicular to the figures.

2.2. Identifying hydrogen bonds

Hydrogen bonds in a crystal structure are identified using new software based on the current CSD system using the following criteria

$$d_{\rm DA} \leq r_{\rm vdW}(D) + r_{\rm vdW}(A) + t_{\rm d} \qquad (2a)$$

$$t_\theta \leq \theta_{\rm DHA} \leq 180^\circ. \qquad (2b)$$

The criteria define the presence of short atom–atom inter- and intramolecular contacts that involve a mediating H atom. $d_{\rm DA}$ and $\theta_{\rm DHA}$ are the observed heavy-atom distance and hydrogen-bond angle, respectively, $t_{\rm d}$ is the upper-distance tolerance, and t_θ is the lower-angle tolerance. Tolerances of $t_{\rm d}$ = 0.1 Å and $t_\theta = 90^\circ$ are used by default, however, $t_{\rm d}$ ranges from +0.3 to −0.2 Å in some studies herein, and t_θ ranges from 90 to 160° (see below).

Once a hydrogen bond has been located it is characterized according to the three-part hydrogen bond description presented above (1), enabling equivalent hydrogen bonds to be compared. Secondary supramolecular structural features (*e.g.* hydrogen-bond motifs) or other extended structural descriptions are not directly considered in this analysis. We believe their influence on hydrogen-bond persistence, *e.g.* through correlations between different hydrogen bonds in the same motif, to be small relative to other factors. To remain as general as possible, the method takes no account of factors such as crystallization conditions or selected solvents even though both can affect the crystal phase obtained, and perhaps influence the hydrogen bonding observed. Collectively, such influences are expected to average out over the complete dataset.

2.2.1. Varying the hydrogen-bond criteria. The $t_{\rm d}$ and t_θ thresholds essentially control how strictly an interaction must resemble a hypothetically optimal geometry. Better (stronger) hydrogen bonds are commonly found alongside less than optimal bonds, therefore relaxing these criteria affords a greater number of individual hydrogen bonds to be observed on average per structure. When altering the acceptance

criteria, it is important to analyse changes, if any, on observations of hydrogen-bond persistence, so subsequent investigations are performed under various tolerance settings. t_d will be varied (0.3 to −0.2 Å in steps of 0.1 Å) with t_θ fixed at 90° (it is observed that persistent hydrogen bonding is not as sensitive to changes in t_θ). The upper limit of 0.3 Å is chosen as it includes the majority of the potential interactions, despite some of these being significantly weaker. Characteristics of the longer interactions will be discussed below in §3.1. Studies will also be performed varying the distance and angle criteria in combination: first t_d = 0.1 Å and t_θ = 90°; second, t_d = −0.2 Å and t_θ = 160°.

2.3. Comparison methods

2.3.1. Hydrogen-bond similarity index. To provide an initial description of the extent of equivalent interactions observed, a similarity index S_{HB} is defined which assesses pairs of polymorphs in turn. The index is a function of the total number of hydrogen bonds in each form and the number they share, with values ranging between 1 and 0. A score of S_{HB} = 1.0 arises for polymorphs in which all observed hydrogen bonds persist, that is, the numbers of hydrogen bonds are the same and all are found to have chemically equivalent donor and acceptor groups and hydrogen-bond type. S_{HB} = 0 arises where no equivalent hydrogen bonds are observed. The index takes the following form

$$S_{HB}(A, B) = \frac{1}{2}\left(\frac{m_A}{n_A} + \frac{m_B}{n_B}\right) \quad (3)$$

$$= \left(\frac{m_A n_B + m_B n_A}{2 n_A n_B}\right), \quad (4)$$

in which m_A denotes the number of persistent hydrogen bonds, and n_A is the total number of symmetry-irreducible hydrogen bonds in form A. m_A and m_B are mostly, but not always, identical (in the case of $Z' > 1$ structures, for example, there exist multiples of the same donors and acceptors that may or may not hydrogen bond in equivalent ways but that are not crystallographically equivalent). The index enables a hierarchy of hydrogen-bonding similarity in polymorphs in the CSD to be constructed.

To illustrate its use, the index is calculated for the dimorphs FAZTAC and FAZTAC01 (see also §2.1, Fig. 1). The number of symmetry-independent hydrogen bonds in each structure is n_{FAZTAC} = 3 and $n_{FAZTAC01}$ = 6 (recall FAZTAC01 has Z' = 2). The number of persistent hydrogen bonds in each structure is m_{FAZTAC} = 2 and $m_{FAZTAC01}$ = 4, accounting for the equivalent intermolecular secondary amine–carbonyl and intramolecular thiadiazine–carbonyl hydrogen bonds. Hence, S_{HB} may be calculated using (4) as $[(3 \times 4) + (6 \times 2)]/(2 \times 3 \times 6) = 0.666$.

2.3.2. Classification by persistent hydrogen-bond count. To begin to investigate their hydrogen bonding in greater depth, polymorphs are classified according to the number of observed persistent hydrogen bonds. A variable limit, n_{match}, is set to categorize every structure pair as *similar* or *dissimilar*, based on a prerequisite number of persistent bonds. Compounds may then be assessed according to the similarity of their known forms. With changing n_{match}, the level of required similarity can be adjusted.

Polymorphic structures of each compound are compared in pairs to reveal the hydrogen bonds per system that persist in multiple forms, and which are unique. All combinations of possible polymorph pairings are constructed, for example, a tetramorphic system, ABCD, can be paired AB, AC, AD, BC, BD, CD. Each structure's hydrogen bond set is then cross-referenced with that of the corresponding structure in the pair until all itemized hydrogen bonds have or have not been classified as persistent or not. The analysis then involves an assessment of the total number of persistent hydrogen bonds, n_{match}, between all pairs in turn.

The outcome of the comparison is either:

(i) Similar: At least n_{match} hydrogen-bonded contacts are common to the polymorph pair.

(ii) Dissimilar: Fewer than n_{match} hydrogen-bonded contacts are common to the polymorph pair.

Crystal structures in which there are no *dissimilar* hydrogen bonds (all are found to reoccur in multiple forms), but in which there are fewer than n_{match} hydrogen bonds per asymmetric unit are also classified as *similar*. Following the paired comparisons, all crystal structures of the polymorph group can be categorized by their observed hydrogen bonding. One of three outcomes is obtained per system:

(i) all polymorphs are similar;

(ii) all polymorphs are dissimilar;

(iii) given a compound with > 2 polymorphic forms, one or more pairs are similar and one or more are dissimilar. This possibility is given the label *both*.

The procedure may be repeated with a variation in n_{match} to study the changing proportion of similar *versus* dissimilar structures. This analysis is conducted in two parts: applying first a persistent bond threshold, n_{match} = 2, and secondly, n_{match} = 5. Because a higher persistent bond threshold is a stricter criterion for similarity, it might be expected that the similarity of a polymorph pair will be more sensitive to the distance criterion with higher n_{match} limit, and so this effect will be studied.

2.3.3. The nature of persistent hydrogen bonds. Hydrogen bonds are well understood in energetic terms (Pimentel & McClellan, 1960; Scheiner, 1997; Desiraju & Steiner, 1999; Steiner, 2002; Novoa & d'Oria, 2007). They are consistent, localized and directional interactions which can be well categorized by properties such as bond geometry and chemical nature. In an approach contrasting the polymorph similarity classification described above, the hydrogen bonds themselves are categorized based on their persistence in multiple polymorphs. The analysis is repeated using two sets of criteria: first, t_d = 0.1 Å and t_θ = 90°; and secondly, t_d = −0.2 Å and t_θ = 160°, thus retaining only the shorter and more linear hydrogen bonds. Any specific chemical and geometrical trends associated with persistence may be revealed and described by comparing the persistent and unique bonds. For example, unusually persistent interactions between certain pairs of functional groups may be revealed.

Table 1
Frequency of polymorphic system size for the polymorphic screened CSD dataset (1677 structures) and the screened hydrogen-bonding dataset (882) structures.

No. of polymorphs	No. of compounds CSD	Hydrogen bonding
2	724	372
3	58	33
4	11	7
5	2	1
6	0	0
7	1	1
Total	796	414

2.4. Dataset preparation

2.4.1. Polymorphism in the CSD. The complete set of chemical compounds with two or more known forms was extracted from the CSD (Version 5.28). A polymorph in the CSD is related to the alternative known structures of that compound by its reference code, whose first six letters denote a unique compound, and subsequent numbering denotes alternative structural determinations, including polymorphic forms (*e.g.* FAZTAC and FAZTAC01). Appropriate screening was thus conducted to obtain suitable polymorphic structures for automated analysis. To remove duplicate structures, redeterminations *etc.* a best representative CSD subset was used comprising structures with known H atom locations (van de Streek, 2006). In total the subset retains 3944 unique crystal structures of 1894 unique chemical compounds. Further screening retained the 1736 organic structures having error-free coordinates, R factor < 0.1, no disorder, no polymeric bonding and determined using single-crystal methods. This subset was then edited to remove any orphan structures whose polymorphs had been removed by the automated screening, leaving 1677 polymorphic structures of 796 unique chemical compounds.

Table 1 shows polymorphic system size *versus* frequency in the CSD (the screened subset of 1677 structures). The title of 'most polymorphic CSD entry' is awarded to the pigment known as ROY (red–orange–yellow), which has 13 database entries in the QAXMEH family. Seven unique structures pass the screens into the final dataset; the remaining six CSD entries do not have three-dimensional coordinates. Due to its extensive polymorphism, this compound has previously been examined in detail (*e.g.* Dunitz & Gavezzotti, 2005). *p*-Dichlorobenzene (DCLBEN) has 12 database entries, five of which pass our screens, and sulfathiazole (SUTHAZ) has seven database entries, five of which are retained. Such extensive polymorphism would appear to be exceptional: the majority of the systems in the CSD are dimorphic (724 compounds = 91%).

2.4.2. Hydrogen-bonded polymorphs. Polymorphic structures with no hydrogen-bonding functionality were removed from the analysis. Hydrogen bonding was then identified per structure, according to the default definitions in §2.2. Most compounds were either hydrogen bonded or not, however, 13 compounds of the 796 were also removed for which only one polymorph displayed hydrogen bonding according to the default criteria. The resultant hydrogen-bonding subset of CSD polymorphs has 882 structures or 414 polymorphic systems.

Table 1 also shows polymorph system size *versus* frequency amongst the hydrogen-bonded polymorphic structures. A very similar trend is observed as for the superset of polymorphic structures in the CSD. Note that the relative number of systems with three or more polymorphs is higher for hydrogen-bonding structures (33/414) than for the full set (58/796). Notice also that the non-hydrogen-bonding pentamorphic *p*-Dichlorobenzene (DCLBEN) has been lost from the set.

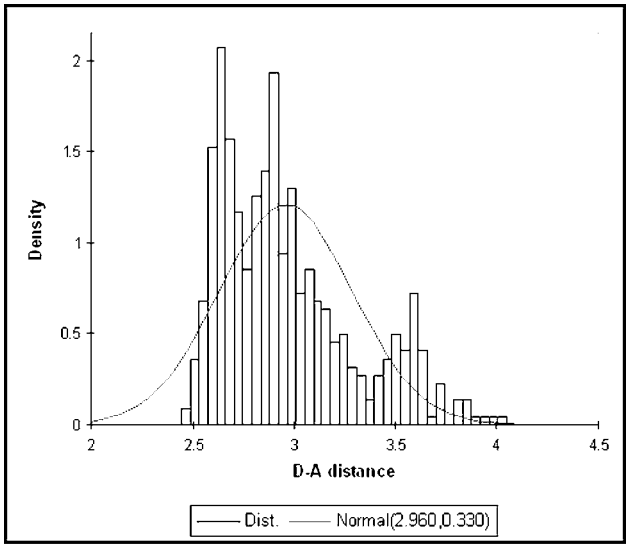
2.4.3. Conformational *versus* packing polymorphs. Conformational polymorphism is a well known phenomenon in which alternative intermolecular interactions and crystal packing may be realised alongside a change in molecular conformation. Improved lattice energies are often found as a result of stabilizing hydrogen bonds, despite distortions from a stable solution-state molecular conformation (Threlfall, 2003). It has been observed in individual cases that certain preferential modes of interaction (*i.e.* hydrogen-bond motifs) are nonetheless retained in multiple forms (Fábián & Kálmán, 2004). The alternative, which we term packing polymorphism, occurs where equivalent molecular conformations are packed differently to form contrasting three-dimensional structures. We decided to investigate any contrast in hydrogen-bond persistence between the two types of polymorphism and therefore devised a method to classify the set of polymorphic crystal structures into two subsets of packing and conformational polymorphs. The method compares molecules using their set of non-terminal proper torsion angles, with a difference in conformation being characterized by a threshold in the difference of any non-terminal proper torsion angle, $\Delta\tau_i$. Technical details can be found in *Appendix* A1.

To establish the torsion angle threshold, a standardized torsion angle difference is employed, denoted $|\Delta\tau^*|$ (see *Appendix* A1), and a frequency distribution of $|\Delta\tau^*|$ in polymorph pairs was studied to reveal how often and to what extent variations in torsion angles occur. A choice of $|\Delta\tau^*| = 40^\circ$ was made for the threshold value, based on the analysed distribution. (Appendix A2 presents the $|\Delta\tau^*|$ analysis). It is found that the threshold corresponds to 55% of all the polymorphs in the CSD (918 of 1677 structures), but only 14% of the total number of individual comparisons have $|\Delta\tau^*| \geq 40^\circ$, *i.e.* it is sufficient for one torsion angle in a molecule to differ significantly to characterize a change in conformation. Similar proportions of compounds with significant conformational change have also been observed by Weng *et al.* (2008) using a method for conformational clustering of complete molecules in the CSD based on r.m.s. atom distances, both for polymorphs and for all organic structures in the CSD.

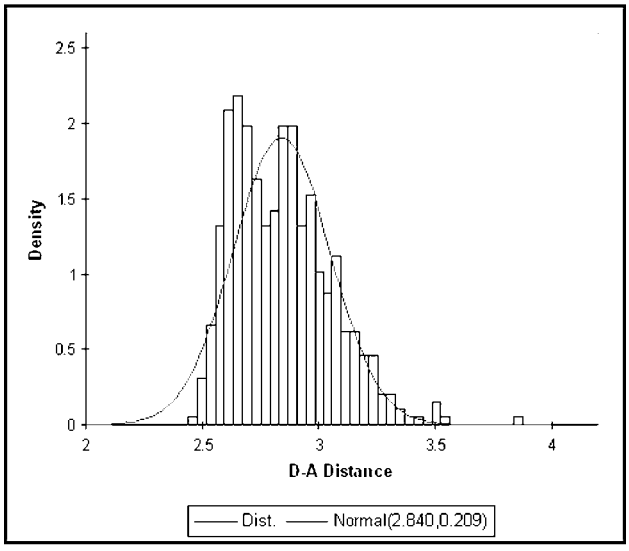
The distribution of $|\Delta\tau^*|$ frequencies in the subset of polymorphs with hydrogen bonding (882 structures) is qualitatively the same as the entire set of polymorphs. Thus, the presence of hydrogen bonding does not appear to influence the degree of conformational polymorphism. With a bound of $|\Delta\tau^*| = 40^\circ$, 12% of the torsion-angle differences indicate a

significant change, originating from 513 structures (58%). Thus, the analysis provides two sets of hydrogen-bonded polymorphic structures: 513 structures that exhibit a significant conformational change and 369 structures having conformers classified as 'similar'.

2.4.4. Further restrictions. To more fully investigate any inherent differences between the behaviour of packing and conformational polymorphs, two comparable samples from each set were prepared. First, it was ensured that all systems contained structures determined under the same experimental conditions, *i.e.* at temperatures within a 20 K difference and equivalent experimental pressures (room temperature, 283–303 K, and ambient pressure was assumed for the entries where no such information was recorded). Secondly, all packing polymorphs whose components are entirely rigid, owing to their chemical structure, were removed from the analysis to leave packing polymorphs whose conformers have the potential to differ even if they do not. Rigid molecules were identified by counting the number of flexible or rotatable non-terminal bonds, n_{tor}, and any fully rigid structures with $n_{tor} = 0$ were discarded. It is acknowledged that some effectively constrained ring systems may remain. Thirdly, a statistical bias favouring the occurrence of conformational polymorphs was observed through an appreciable correlation between n_{tor} and molecular mass, and a contrast in the distribution of molecular mass in the two sets was noted; a mean molecular mass of 273 ($\pm$110) g mol^{-1} for conformational polymorphs, and 201 ($\pm$111) g mol^{-1} for packing polymorphs. To correct this, an equal sample size of packing and conformational polymorphs was generated by random selection after discarding structures with extreme molecular mass values, providing 180 structures of each polymorphic type (from 87 and 84 compounds for packing and conformational polymorphs, respectively). It is the results of contrasting the hydrogen-bond persistence in these two subsets that will be presented below, however, comparisons will also be made with equivalent results obtained from the unrestricted sets.



(*a*)



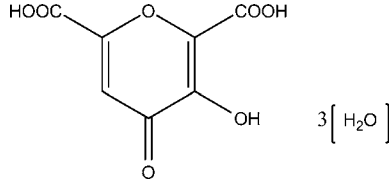
(*b*)

Figure 3
Distributions of donor–acceptor atom distance observations for hydrogen bonds of all polymorphic CSD structures with (*a*) hydrogen-bond criteria t_d = 0.3 Å and (*b*) t_d = −0.2 Å. A normal distribution with the mean and variance of each sampled distribution is also plotted, denoted *normal*(μ,σ^2).

3. Results

3.1. Sensitivity to hydrogen-bond criteria

To investigate any inherent sensitivity of hydrogen-bonding observations to the choice of hydrogen-bond criteria t_d and t_θ (as defined in §2.2), contrasts in the frequency of observed hydrogen-bond distances and angles were analysed for the set of hydrogen-bonding polymorphs. Donor–acceptor (*D*–*A*) distances and donor–hydrogen–acceptor (*D*—H—*A*) angles were recorded for hydrogen bonds in all structures of the set. The number of observed hydrogen bonds per asymmetric unit is observed to depend significantly on the distance criterion: t_d = 0.3 Å gives an average in the dataset of 2.91 hydrogen bonds per asymmetric unit, whilst t_d = −0.2 Å yields an average of 2.22 bonds. The difference in specific structures can be more striking: ZZZSSY01 (meconic acid trihydrate) has the highest number of individual hydrogen bonds per asymmetric unit in the dataset studied. There are 13 lone pairs available to accept hydrogen bonds and nine donor H atoms in the asymmetric unit. 17 hydrogen bonds may be identified with t_d = 0.3 Å, in contrast to 10 with t_d = −0.2Å, thus 7 potential hydrogen bonds are lost from the comparison as t_d is decreased.



To look more closely at observed hydrogen-bond geometries, we present two distributions of observed *D*—*A* distances, at long and short distance criteria of t_d = 0.3 and −0.2 Å, respectively. The effect of varying t_θ was also investigated, but

contrasts were not as significant and are not discussed here. Further effects of varying the criteria are studied in §3.3 when classifying compounds by persistence, and an investigation into complete hydrogen-bond geometry for persistent and unique interactions is presented in §3.4.2. Fig. 3 shows the comparison at the two distance thresholds. The distribution using t_d = 0.3 Å (Fig. 3*a*) shows a multimodal behaviour with a significant amount of observations around 3 Å, with a second significant peak at ∼ 3.5–3.6 Å. That the distance distribution is not smooth is suggestive of different underlying behaviours; the less common behaviour at longer distances implies a weaker interaction type of significantly different character to that of the primary peak. With t_d = −0.2 Å (Fig. 3*b*), the behaviour causing the added peak at ∼ 3.5 Å has been lost, while the distribution at shorter distances is largely unaffected. Infrequent longer hydrogen bonds may still be found, these can be traced to sulfur as the donor or acceptor {van de Waals radius [r_{vdW}] (S) = 1.80 Å; *cf.* r_{vdW} (O) = 1.52 Å, r_{vdW} (N) = 1.55 Å}.

In summary, the relative underlying behaviour in hydrogen-bond geometries does not appear grossly sensitive to the choice of distance and angle thresholds. Absolute changes in observed numbers of hydrogen bonds are seen, and a loss of longer hydrogen is found with stricter criteria. Alongside the results below, any observed effects of choices in the criteria to hydrogen-bond persistence will also be discussed.

3.2. Hydrogen-bond similarity index

The extent of persistent interactions in all polymorph pairs was quantified using the similarity index, S_{HB}, described in §2.3.1. We observe that chemically equivalent hydrogen bonding is highly likely to persist in polymorphic crystal structures. The distribution of S_{HB} scores is plotted in Fig. 4 (t_d = 0.1 Å, t_θ = 90°). Only 78 polymorph pairs (15%) are completely dissimilar: S_{HB} = 0, and there are only eight pairs with a low, but non-zero score ($0.0 < S_{HB} < 0.5$). The majority (83%) have $S_{HB} > 0.5$, which corresponds, on average, to at least half of the available chemically equivalent hydrogen bonds being conserved in the polymorph pair. In fact 332 pairs (66%) have perfect similarity, S_{HB} = 1.0. The trend is for a greater number of pairs to be observed as the similarity increases over the range $0.4 < S_{HB} < 1.0$. The trends are also qualitatively the same for the separate packing and conformational polymorph subsets.

3.3. Classification by persistent hydrogen-bond count

Every polymorph pair was classified as similar or dissimilar on the basis of the number of chemically equivalent hydrogen bonds (n_{match}). Firstly, hydrogen bonds were identified per polymorphic system with default settings of t_d and t_θ at 0.1 Å and 90°, respectively. n_{match} was systematically adjusted from 1 to 8 (Fig. 5, Table 2). Three categories for each compound are possible: *similar*, *dissimilar* and *both*. At n_{match} = 1, a large majority (95.6%) of the polymorphs are classified as similar and share at least one hydrogen bond. The result drops only slightly (93.7%) if at least two shared hydrogen bonds are required, corresponding to a difference of eight polymorphic systems. There is more of a change from n_{match} = 2 to n_{match} = 3 (similar proportion = 84.1%). Hydrogen bonds which co-exist as dimer motifs may account for the comparatively higher proportion of polymorphs that share two hydrogen bonds. Hydrogen-bond dimers in polymorphic systems have been investigated previously (Bernstein *et al.*, 1995; Blagden *et al.*, 1998). The effects of consistently occurring motifs are indirectly observed in this analysis as they afford a favourable likelihood for the occurrence of certain combinations of hydrogen bonds.

As the n_{match} criterion is increased, the proportion of polymorphs with similar hydrogen bonding decreases further. There is increasing scope for those structures with greater than n_{match} potential hydrogen bonds to show variation in the bonds which form, however, there is also a limit on the number of structures with > n_{match} potential hydrogen donors, and so the curve's gradient declines. A dominant proportion of polymorphic forms with similar hydrogen bonding persists across the range of analyses undertaken.

3.3.1. Effect of the hydrogen-bond criteria. The distance criterion t_d was decreased from 0.3 to −0.2 Å in six steps

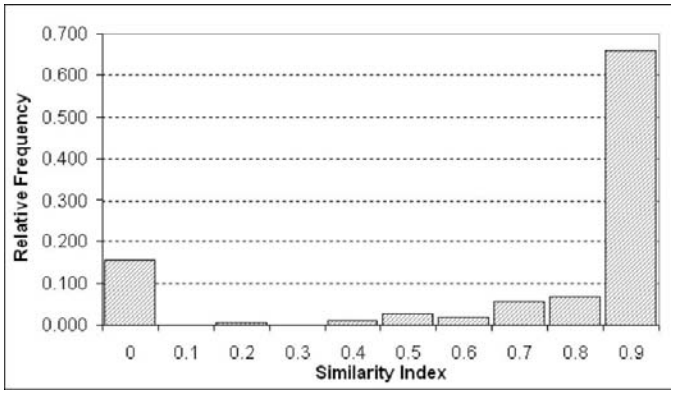


Figure 4
Histogram showing the distribution in frequency of similarity index, S_{HB}, for hydrogen-bonding polymorph pairs. The majority achieve $0.9 < S_{HB} \leq 1.0$.

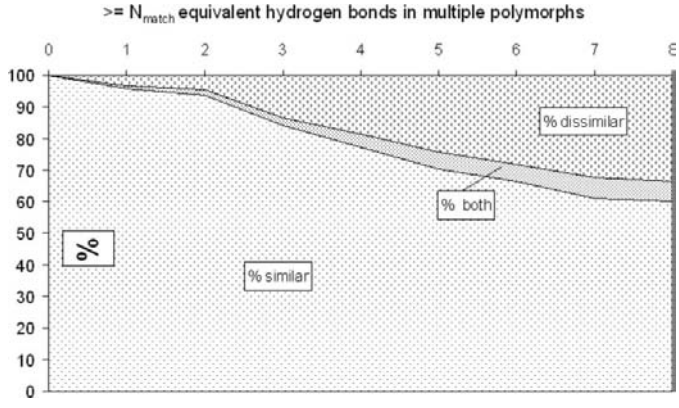


Figure 5
Categorization of CSD polymorphs according to number of common hydrogen bonds per polymorphic system.

Table 2
The number of structures classified as similar or dissimilar with changing n_{match} count as classifier.

n_{match}	Total similar compounds	% similar	Total dissimilar compounds	% dissimilar	Total both observations	% both
0	414	100	0	0	0	0
1	396	95.6	13	3.1	5	1.2
2	388	93.7	19	4.6	7	1.7
3	348	84.1	55	13.3	11	2.6
4	320	77.3	77	18.5	17	4.1
5	291	70.2	101	24.4	22	5.3
6	274	66.2	116	28.0	24	5.8
7	253	61.1	134	32.4	27	6.5
8	248	60.0	139	33.5	27	6.5

Table 3
Number of structures classified as similar or dissimilar with changing hydrogen-bond upper distance tolerance, with n_{match} = 2 as classifier.

Distance tolerance (Å)	Total similar compounds	% similar	Total dissimilar compounds	% dissimilar	Total compounds with both	% both
0.3	388	93.7	19	4.6	7	1.7
0.2	387	93.5	21	5.1	6	1.4
0.1	383	92.5	24	5.8	7	1.7
0	382	92.3	25	6.0	7	1.7
−0.1	381	92.0	26	6.3	7	1.7
−0.2	382	92.3	27	6.5	5	1.2

Table 4
Number of structures classified as similar or dissimilar with changing hydrogen-bond upper distance tolerance, with n_{match} = 5 as classifier.

Distance tolerance (Å)	Total similar compounds	% similar	Total dissimilar compounds	% dissimilar	Total compounds with both	% both
0.3	291	70.3	101	24.4	22	5.3
0.2	288	69.6	105	25.4	21	5.1
0.1	288	69.6	105	25.4	21	5.1
0	280	67.6	112	27.1	22	5.3
−0.1	280	67.6	114	27.5	20	4.8
−0.2	278	67.1	117	28.3	19	4.6

of 0.1 Å to study how grouping polymorphs by the numbers of persistent hydrogen bonds might be affected by this upper distance limit, with a required number of matches, n_{match} of 2 and 5. The results are given in Tables 3 and 4. The ratio of similar to dissimilar polymorph pairs is largely unchanged, again reflecting a lack of sensitivity to the criteria. A small loss in the *similar* subset as t_d decreases is of note; such a change occurs if an equivalent hydrogen bond in a pair of structures has a large disparity in the $D-A$ distance. This trend is also shown when repeating calculations of S_{HB} using the longer t_d limit, as in the previous section, with little difference in the distribution of scores.

Requiring five persistent hydrogen bonds means the number of similar polymorphs is reduced. With this requirement the effect of a shorter distance cut-off is also stronger, and the number of similar polymorphs becomes fewer still. In either case, the change is not substantial: there is $\sim 2.5\%$ difference in category assignment with n_{match} = 5 as the distance limit is decreased, and $\sim 1.4\%$ difference when n_{match} = 2.

3.4. The nature of persistent hydrogen bonds

For the complete set of hydrogen-bonding polymorphs, 5760 persistent hydrogen bonds (62% of 9318 observed hydrogen bonds) and 3558 unique bonds (38%) are observed using t_d = 0.1 Å and t_θ = 90°. If only hydrogen bonds with ideal geometry ($t_d \leq -0.2$ Å and $t_\theta \leq 160°$) are considered then the proportion of bonds that persist in multiple forms increases: 1804 persistent hydrogen-bond pair observations (71%) and 753 unique (29%). Indeed, the assertion that the stronger hydrogen bonds are retained across multiple structural forms is reinforced.

3.4.1. Chemical group influence. The observed influences of different chemical groups were qualitatively the same for the packing and conformational polymorphs; subsequent discussion is therefore based on analysis of all polymorphic structures collectively. More than 80 independent chemical groups were observed in these structures, and there are many possible $D-A$ pair combinations. Despite this, certain highly persistent and unique interactions may be seen. Table 5 displays the observed frequency of common $D-A$ pairs. Differences in the frequency of persistent and unique pairs were calculated, together with a Poisson error dependent on the number of samples. Results that differed by less than 30% of the frequency were ignored, and differences with a Poisson error of greater than 20% of the difference were also discarded. We note that the frequencies observed in this analysis are particular to the chemistry of the polymorphic structures of our sample: crystal structures in the CSD are present as a result of scientific relevance and/or interest, and related compounds have been studied systematically, the so-called 'social bias'. With this in mind, care has been taken to compare chemical groups represented with sufficient frequency in the CSD, but such an analysis cannot perfectly cover the complete chemical space.

The % difference may be interpreted as the excess proportion of observations in the persistent (%> 0) or unique (% < 0) category. Four different hydrogen-bonded interactions are seen to persist in > 80% of cases, *i.e.* if they are observed in one polymorph, in > 80% of cases they are observed in a second. Amines feature strongly as donors in this list. There are some highly polar interactions in the list, *e.g.* ammonium–carboxylate (80.8%: persistent unique) indicating a stronger and more stabilizing interaction. Considering the unique pairs list, eight hydrogen-bonding pairs are > 60%

Table 5
Comparison of the frequency of complete hydrogen-bonded contacts between persistent *versus* unique hydrogen bonds in >1 polymorph.

(*a*) Most frequent persistent contacts; (*b*) most frequent unique contacts.

Rank	Hydrogen-bonded contact ($D-A$)	N_{obs} persistent hydrogen bonds	N_{obs} unique hydrogen bonds	% Δ (persistent-unique) (Poisson Error)
(*a*)				
1	Hydroxyl, *sp*2 carbon–acetyl	36	3	84.6 (16.0)
2	Hydroxyl, *sp*2 carbon–carboxylate	32	3	82.9 (16.9)
3	Amino, secondary–methoxycarbonyl	76	8	81.0 (10.9)
4	Ammonium–carboxylate	508	54	80.8 (4.2)
5	Amino, secondary–nitro, uncharged	142	17	78.6 (7.9)
6	Amino, secondary–hydroxyl, aliphatic	28	4	75.0 (17.7)
7	Hydroxyl, aliphatic–hydroxyl, aliphatic	2246	445	66.9 (1.9)
8	Carbamoyl–carbamoyl	228	49	64.6 (6.0)
9	Hydroxyl, aromatic–methoxycarbonyl	28	8	55.6 (16.7)
10	Hydroxyl, aromatic–nitro, uncharged	24	7	54.8 (17.9)
11	Amino, aromatic–ketone, aliphatic	22	7	51.7 (18.6)
12	Amino, aromatic–nitro, uncharged	110	35	51.7 (8.3)
13	Amido–amido	464	153	50.4 (4.0)
14	Carboxylic acid–carboxylic acid	244	99	42.2 (5.4)
15	Amino, secondary–ketone, aliphatic	28	13	36.6 (15.6)
17	Hydroxyl, aromatic ketone, aliphatic	98	51	31.5 (8.19)
(*b*)				
1	Carboxylic acid–water	8	51	−72.9 (13.0)
2	Monosub'd sulfonamide–nitrogen, aromatic	6	37	−72.1 (15.2)
3	Hydroxyl, aliphatic–ether	40	224	−69.7 (6.1)
4	Water–water	10	46	−64.3 (13.4)
5	Monosub'd sulfonamide–monosub'd sulfonamide	6	27	−63.6 (17.4)
6	Hydroxyl, aliphatic–nitrogen, aromatic	14	59	−61.6 (11.7)
7	Water–nitrogen, aromatic	6	25	−61.3 (17.9)
8	Amino, sec.–monosub'd sulfonamide, charged	12	48	−60.0 (12.9)
9	Water–hydroxyl, aliphatic	36	131	−56.9 (7.7)
10	Hydroxyl, aromatic–water	6	21	−55.5 (19.2)
11	Hydroxyl, aliphatic–water	30	96	−52.4 (8.9)
12	Amino, aromatic–nitrogen, aromatic	12	37	−51.0 (14.3)
13	Hydroxyl, aliphatic–carboxylic acid	18	54	−50.0 (11.8)
14	Carboxylic acid–carboxylate	18	52	−48.6 (12.0)
15	Hydroxyl, aliphatic–amido	10	28	−47.0 (16.2)
16	Water–carboxylic acid	20	55	−46.7 (11.5)
17	Carboxylic acid–hydroxyl, aliphatic	10	26	−44.4 (16.7)
18	Amino, secondary–amido	10	24	−41.1 (17.1)
19	Hydroxyl, aromatic–nitrogen, aromatic	18	43	−41.0 (12.8)
20	Amido–carboxylic acid	8	19	−40.7 (19.2)
21	Water–nitro, uncharged	8	19	−40.7 (19.2)
22	Hydroxyl, aromatic–carboxylic acid	18	41	−39.0 (13.0)
23	Hydroxyl, aromatic–amido	26	56	−36.6 (11.0)
24	Water-hydroxyl, aromatic	16	30	−30.4 (14.7)

more likely to be unique than persistent, however no unique contacts are >73% likely. Water features strongly as both donor and acceptor in the unique pairings, and is discussed further below.

Next, hydrogen donor and acceptor groups are considered individually according to their relative likelihoods of appearing in the persistent or the unique set. Table 6 presents the frequency data for significant identified groups. Poisson errors are again considered: samples were discarded with an error in the frequency of > 20%. The ammonium group is by far the most persistent donor with matches found in 71% of polymorphic systems in which the group is present. The effect of charge on the hydrogen bonding of amines can perhaps be seen, as aromatic amines are only 8% more likely to be involved in persistent bonding, secondary amines 7% and aliphatic amines are 28% *less* favoured. Hydroxyl groups are also commonly persistent donors and acceptors (aliphatic hydroxyl = 42% persistent as donor and 56% as acceptor). This statistic may be partly explained by multiple hydroxyl groups being frequently observed for compounds in the dataset, providing more opportunities for chemically equivalent but topologically different interactions to occur in different polymorphs.

Equivalent hydrogen bonds involving a water molecule are unlikely to be found in pairs of polymorphs (−50% persistent observations as donor and −59% as acceptor), which complements the observation of water in unique donor–acceptor pairs. It suggests alternative competitive interactions including water are available, which appear to be accessible and of comparable stabilizing ability. Crystal hydrates have been widely reported containing an array of common aggregation states (*e.g.* Mascal *et al.*, 2006). Furthermore, a study of the molecular geometry of water in aqueous organic crystal structures (Krygowski *et al.*, 1998) has shown that water–solute and water–water interactions occur more readily than might be expected because of a high degree of elasticity in the water conformation.

The weakly accepting ether moiety less readily forms equivalent hydrogen bonds in multiple forms by 59.2%. The least likely donor group to be found in persistent hydrogen bonds is the monosubstituted sulfonamide (−77%), however, there is no discrimination as an acceptor (−3 ± 7%). This fragment has many potential hydrogen-bond acceptor sites which would suggest an added flexibility in its behaviour.

3.4.2. Hydrogen-bond geometry. Either the relaxed (default) or restricted geometry tolerances showed qualitatively similar observed geometrical behaviour and so the results presented subsequently concern the *persistent* and *unique* sets obtained with the default hydrogen-bonding criteria. The distributions of both the *D*–*A* hydrogen-bond distances and the $D-H-A$ hydrogen-bond angles show significantly contrasting behaviours of persistent and unique hydrogen bonds. The distance distributions are shown in Figs. 6(*a*) and (*b*). Fig. 6(*c*) shows a plot of the difference between

Table 6
Comparison of the frequency of donor and acceptor groups between persistent *versus* unique hydrogen bonds in < 1 polymorph.

Those groups with a difference < 1% are not shown.

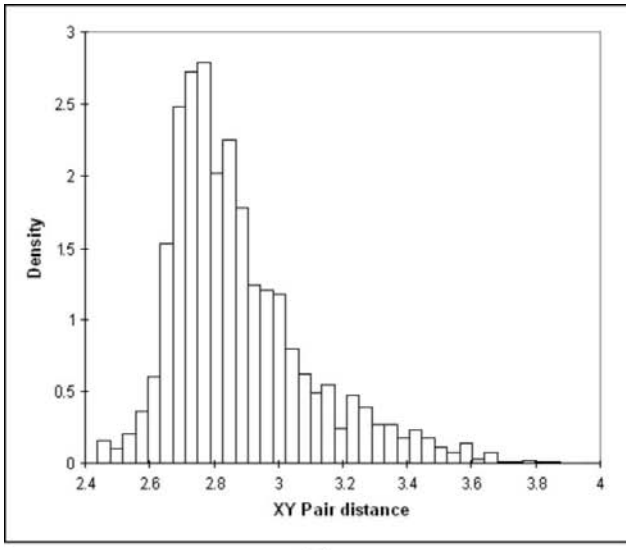
				Δ (persistent-unique)		
Type	Group	N_{obs} persistent hydrogen bonds	N_{obs} unique hydrogen bonds	Frequency	%	Poisson error (%)
Donor	Ammonium	526	90	436	70.8	4.0
	Hydroxyl, aliphatic	2406	974	1432	42.4	1.7
	Carbamoyl	264	108	156	41.9	5.2
	Hydroxyl, *sp*2 carbon	88	36	52	41.9	9.0
	Amido	630	365	265	26.6	3.2
	Hydroxyl, aromatic	462	393	69	8.1	3.4
	Amino, aromatic	384	327	57	8.0	3.8
	Amino, secondary	456	394	62	7.3	3.4
	Carboxyl	304	299	5	0.8	4.1
	Amino, uncharged aliphatic	32	57	−25	−28.1	10.6
	Water	116	350	−234	−50.2	4.6
	Monosub'd sulfonamide	12	92	−80	−76.9	9.8
Acceptor	Carbamoyl	232	62	170	57.8	5.8
	Hydroxyl, aliphatic	2298	648	1650	56.0	1.8
	Carboxylate	576	183	393	51.8	3.6
	Acetyl	52	20	32	44.4	11.8
	Methoxycarbonyl	38	16	22	40.7	13.6
	Nitro, uncharged	244	118	126	34.8	5.3
	Ketone, aliphatic	214	108	106	32.9	5.6
	Amido	544	339	205	23.2	3.4
	Lactone	20	13	7	21.2	17.4
	Formyl	26	18	8	18.2	15.1
	Hydroxyl, aromatic	198	157	41	11.5	5.3
	Acetylamino	20	17	3	8.1	16.4
	Cyano	20	18	2	5.3	16.2
	Carboxyl	368	366	2	0.3	3.7
	Nitrogen, aromatic	338	354	−16	−2.3	3.8
	Monosub'd sulfonamide	88	94	−6	−3.3	7.4
	Sulfonylimine	268	311	−43	−7.4	4.2
	Thioether	8	20	−12	−42.9	18.9
	Monosub'd sulfonamide charged	28	76	−48	−46.2	9.8
	Ureido	8	26	−18	−52.9	17.1
	Water	62	241	−179	−59.1	5.7
	Ether	64	250	−186	−59.2	5.6

the two distributions (persistent $D-A$ distance density *minus* unique $D-A$ distance density), enabling an effective visual comparison of the distributions. A three-bin average is plotted to emphasize underlying trends without sampling effects. We find that both very short and longer hydrogen bonds are much less likely to persist than medium range bonds. From the distribution plot, up to a 32% deviation based on a three-sample difference average and distance maxima may be seen. A large positive area around a distance of ~ 2.7 Å indicates a predominance of hydrogen bonds that are conserved at this length, whereas a predominance of unique hydrogen bonds is observed below 2.6 Å. A sample density favouring unique hydrogen bonds is also seen at longer distances, although the trend is less pronounced. We point out that the cumulative densities for both positive and negative samples are equal, due to the symmetry in sample density (and thus the scale of the distributions), and therefore regions of positive and negative dominance are to be expected.

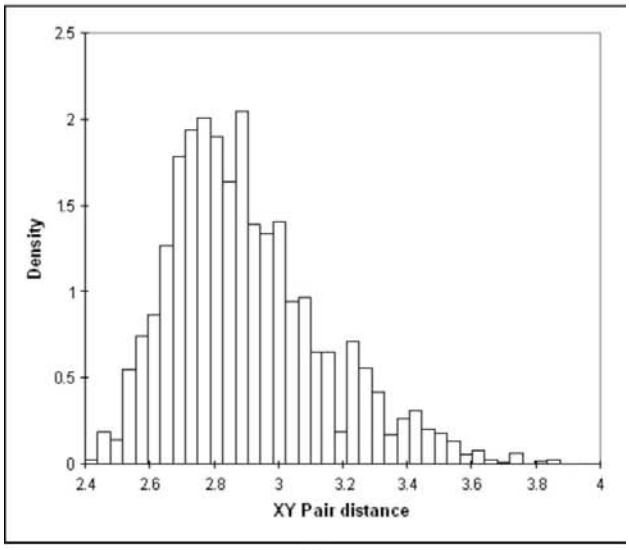
The distributions of $D-H-A$ angles are shown in Figs. 7(*a*) and (*b*). The presentation and interpretation of the angle statistics uses a cone correction (Kroon *et al.*, 1975; Lommerse *et al.*, 1996; Ciunik & Desiraju, 2001) which avoids an artificial up-weighting of observations at lower angles. These histograms are plotted *versus* the relative sample frequency (*cf.* density, used above) which simplifies the renormalization after weighting with the cone correction. Visual comparison shows very little contrast in the distributions. Fig. 7(*c*) shows a plot of the difference between the distributions together with a three-sample average. The absolute differences are small (there is at most a 12% deviation), but bent geometries are more frequent amongst the unique hydrogen bonds. In particular a positive area between 160 and 170° is noted, indicating a proliferation of persistent hydrogen bonds in this range, while bonds unique to one polymorph are more likely to exhibit an angle between 140 and 160°.

Visualizing $D-A$ distances and $D-H-A$ angles simultaneously in the two sets of data provides a more complete view of variation in the interaction geometries. Representation may be achieved by way of bivariate scatterplots. Alternatively, when the dataset size becomes large, density plots allow any underlying data structure to be revealed more clearly, especially where, due to the number of observations, datapoints virtually overlap. Kernel density estimation (KDE; Scott, 1992) has been applied using a Gaussian kernel of flexible bandwidth in each variable to model the density of the hydrogen-bond data. Source code has been written that produces a two-dimensional grid of arbitrary resolution of interpolated density estimates from a bivariate dataset with the Gaussian KDE method. An added advantage of the method is that the required cone correction for the $D-H-A$ distribution may be automatically included in the algorithm, whereas unmodified scatterplots cannot represent such a correction.

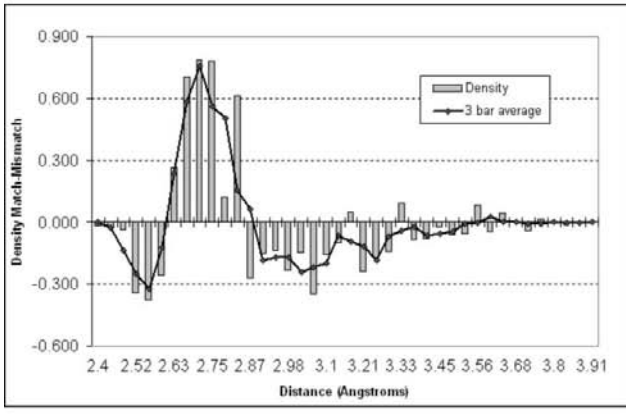
Density plots are shown in Fig. 8 comparing the correlation between $D-A$ distance and $D-H-A$ angle for the persistent and unique hydrogen bonds. Maxima at 175–180° and 2.7–2.9 Å appear in both distributions, characterizing an optimal short and linear interaction. The shapes of the peaks are in marked contrast: that of the persistent hydrogen bonds is



(*a*)



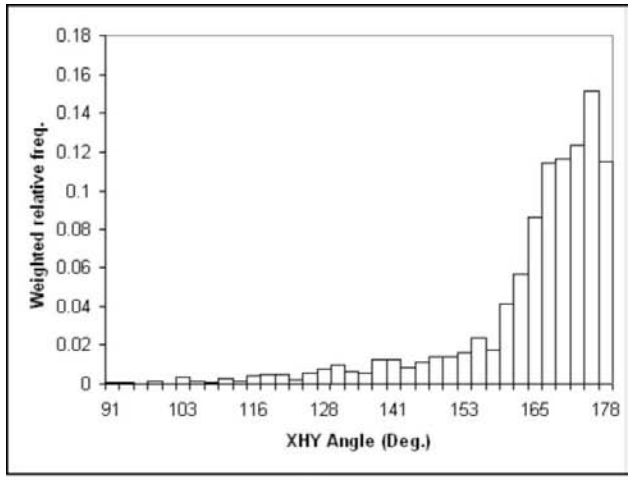
(*b*)



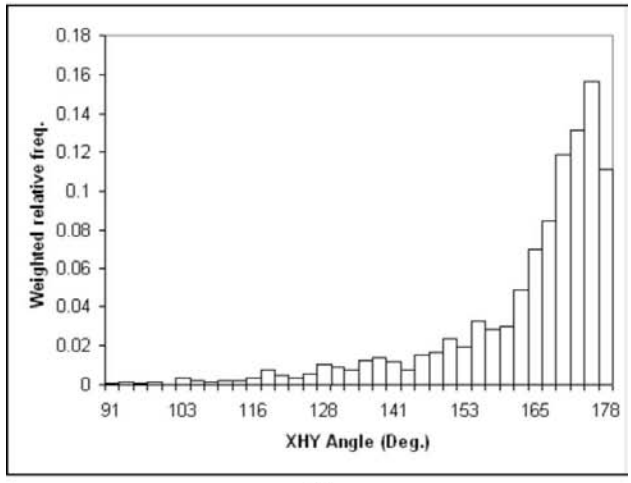
(*c*)

Figure 6
D—A distance distributions for hydrogen bonds that are (*a*) persistent and (*b*) unique in polymorph pairs. (*c*) Difference in the distributions, with three-bin average.

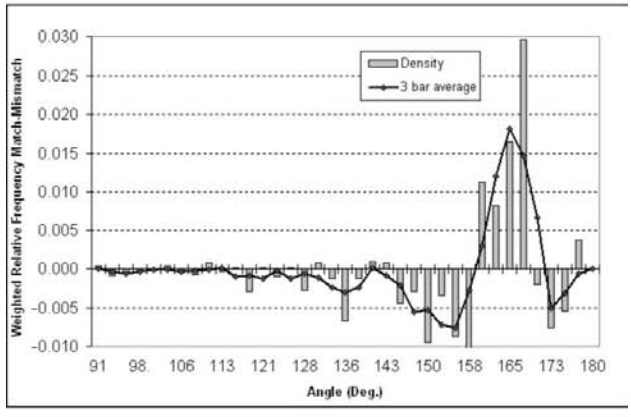
more dense and falls off more rapidly away from the maximum, showing persistent hydrogen bonds nearer the optimum geometry are more frequent. The unique hydrogen bonds are more likely to occupy an alternative point in



(*a*)



(*b*)



(*c*)

Figure 7
D—H—A angle distributions for hydrogen bonds that are (*a*) persistent and (*b*) unique in polymorph pairs. (*c*) Difference in the distributions, with three-bin average.

geometrical space. Despite θ_{DHA} samples obtained above t_θ = 90°, observations below 120° are insignificant in both samples.

Comparisons of packing and conformational polymorphs were first performed with the complete subsets as determined in §2.4.3 (513 conformational polymorphs and 369 packing polymorphs). Following that, the restricted subsets defined in §2.4.4 (180 structures of each type) were analysed in the same way. Both comparisons showed qualitatively the same behaviour, despite a small increase in statistical noise, and so the following results concern the restricted sets. Persistent hydrogen bonding remains roughly the same (∼ 60% persistent hydrogen bonds) for both conformational and packing polymorphs. Density plots for both classes are plotted in Fig. 9. Just as in the complete set, the peak in density for persistent hydrogen bonds for either polymorph type is distinct: bond distance between 2.7–2.9 Å and angle around 180°. The decay in density is rapid away from the maximum. The maximum for persistent hydrogen bonds occurs at a slightly shorter $D—A$ distance with conformational polymorphs. Hydrogen-bond geometries away from the mode are again more prevalent for unique bonds, particularly in the conformational polymorphs, and hence a greater correlation between persistent hydrogen bonding and optimal geometry can be identified for conformational polymorphs.

3.4.3. Intramolecular hydrogen bonds. Throughout these studies we have observed similar persistence behaviour of intra- and intermolecular hydrogen bonding. For this reason, and because intramolecular bonds are more rare, all hydrogen bonds thus far have been analysed collectively. Here, we further explore the observed intramolecular interactions in isolation. One motivation is that persistent intramolecular bonds are quite frequent in the packing polymorphic structures, but not in conformational polymorphs.

109 polymorphic compounds (24%) have been observed with intramolecular hydrogen bonds in at least one structure (for intramolecular bonds, D and A belong to the same covalently bonded unit, and are separated by more than three covalent bonds). Approximately 20% (22) of these compounds have at least one form where the intramolecular bond is lost, meaning ∼ 80% of compounds display a persistent interaction. It is noted that the persistence might have been expected to be even higher for intramolecular hydrogen bonds, which can be highly probable in certain chemical structure types (Bilton *et al.*, 2000). Conjugated molecular

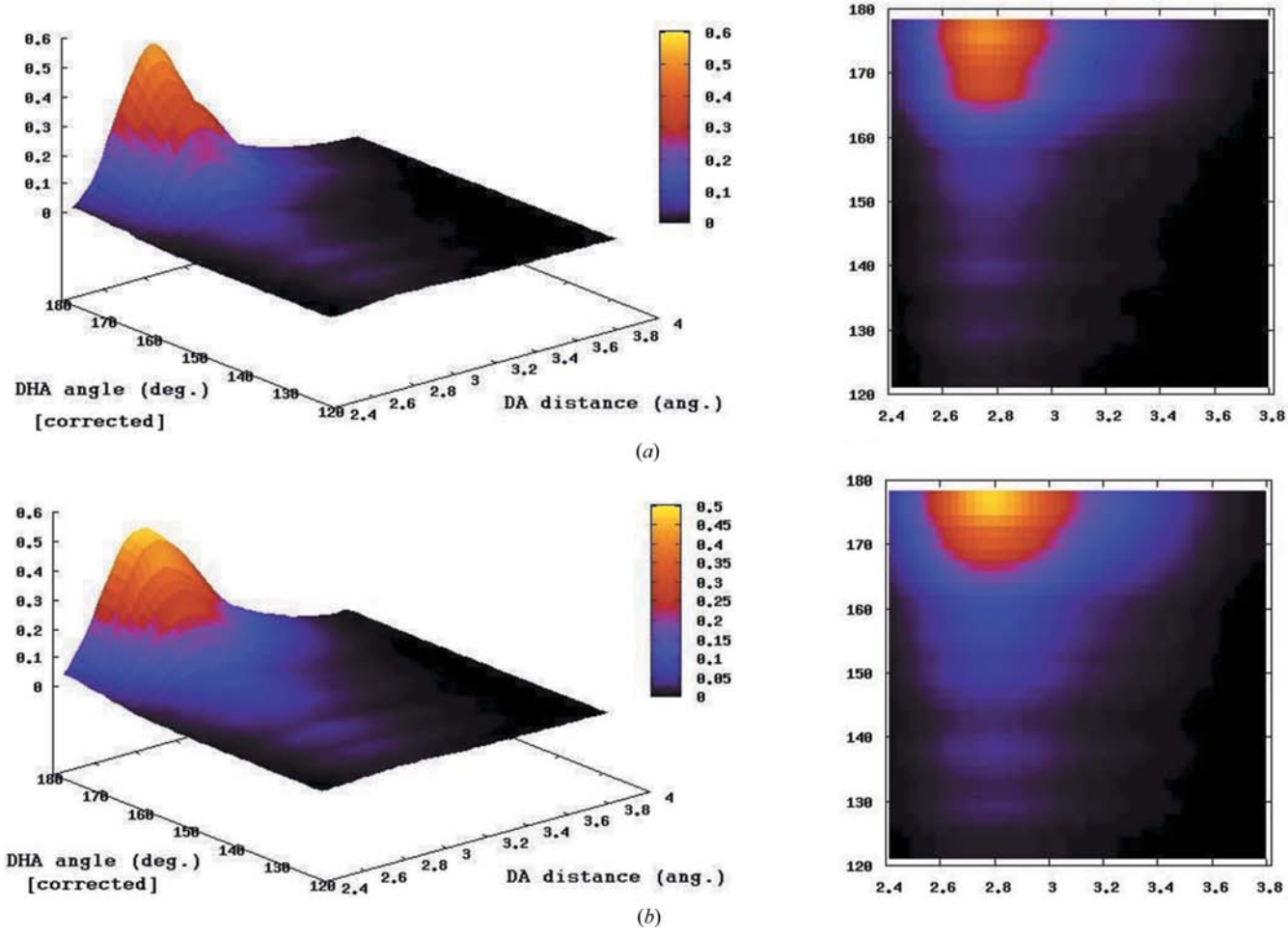


Figure 8
$D—A$ distance–$D—$H$—A$ angle density surfaces and corresponding two-dimensional maps for (*a*) persistent and (*b*) unique hydrogen bonds.

fragments are most likely to form a planar six-membered intramolecular hydrogen-bonding motif (85–100% occurrence), however, it was also observed that some fully saturated potential five- and six-membered motifs are much less likely (< 10% occurrence). In this way, our results appear to reflect the mixed chemistry in the sample of polymorphic structures available. ROY for example (discussed in §2.4.1), with seven polymorphs, displays a persistent six-membered intramolecular interaction between a secondary amine and a nitro group throughout all its modifications. On the other hand, *o*-(acetoamido)benzamide (ACBNZA) has a six-membered intramolecular hydrogen-bond ring between two amide moieties, which is not found in ACBNZA01. The hydrogen bonding in this system is displayed in Fig. 10.

Of the molecules that have at least one variable torsion angle, $n_{tor} \geq 1$ (see §2.4.3), many retain a degree of planarity by way of particular persistent intramolecular hydrogen bonds, such that they show little conformational change. The many ways of layering planar components together in three-dimensional structural assemblies would seem to account for their proliferation in the packing polymorphs. Such packing follows the principle that a minimal external area is retained upon molecular packing (Pidcock & Motherwell, 2005). Examples are DODDAB and APUSAF01 (Figs. 11*a* and *b*). Another refcode family, MARBEM (Fig. 11*c*), illustrates, nevertheless, that conformational persistence need not exclusively involve planar molecules. An intramolecular hydrogen bond appears to restrict bond rotation in this case such that conformations are identical in MARBEM and MARBEM01.

A minority of flexible molecules retain planarity without the aid of intramolecular hydrogen bonds. The effect of

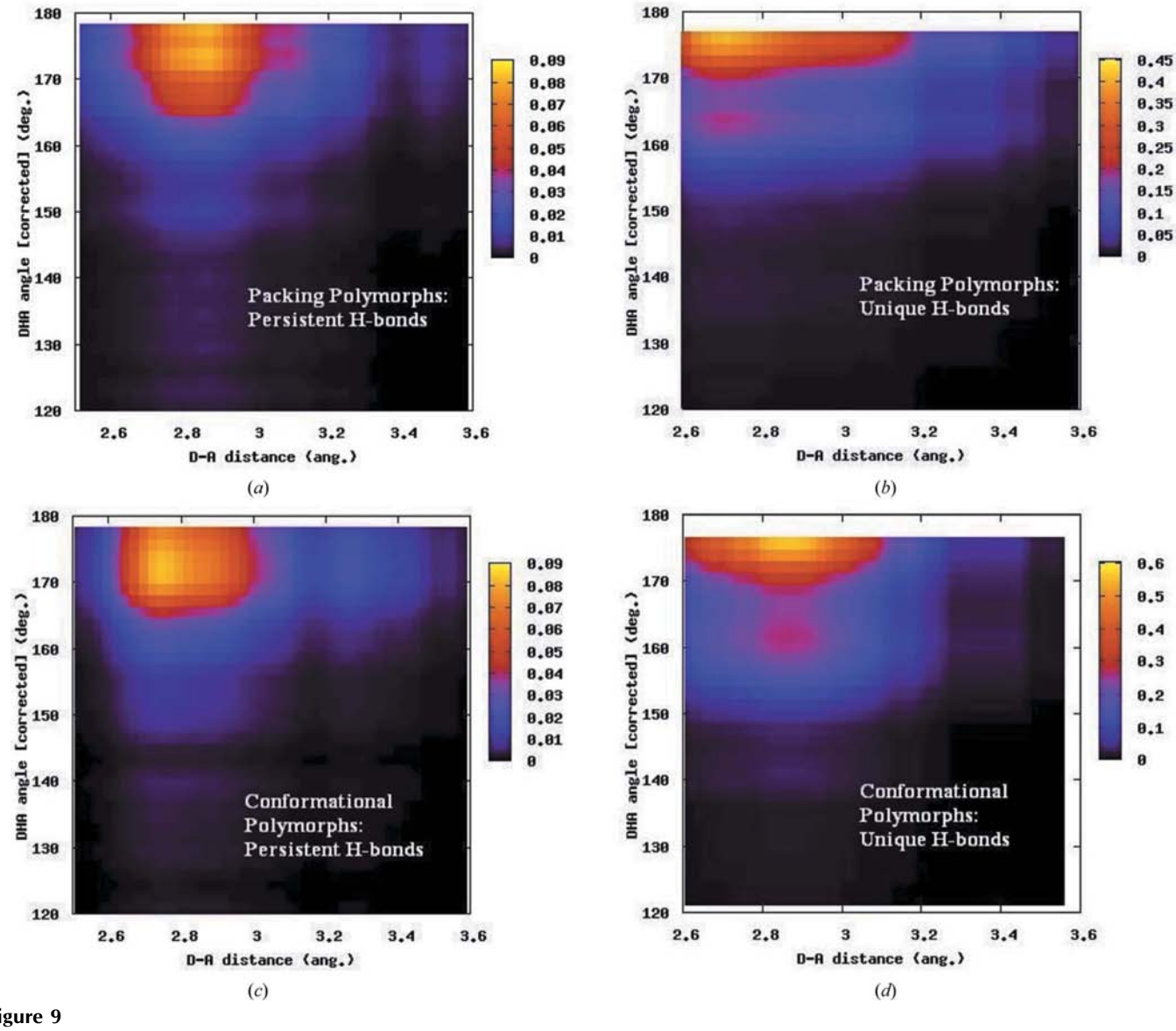


Figure 9
D—*A* distance–*D*—H—*A* angle density maps for (*a*) persistent and (*b*) unique hydrogen bonds in conformational polymorphs; (*c*) persistent and (*d*) unique hydrogen bonds in packing polymorphs.

decreased free energy by favourable packing is well understood (Brock & Minton, 1989; Brameld *et al.*, 2008), even if strained molecular conformations are involved. These tend to be those structures whose flexible parts are hydrogen bonding and they are locked into a preferred conformation by the motif or cooperative interaction they have formed, *e.g.* PYRZIN01 (Fig. 12*a*). In the following section, two further examples will also be discussed. Finally, note that molecules that are almost entirely rigid, *e.g.* with $n_{tor} = 1$, remain in the sample. Those with perhaps one rotatable functional group are again often held in one conformer by cooperative hydrogen bonds, *e.g.* MALEHY and MALEHY01 (Fig. 12*b*).

3.5. Outlying examples

Hydrogen bonds that persist between polymorphs are very common. The high propensity for these bonds to reoccur is also seen to be affected by the chemical groups acting as the donor and acceptor, and also by the extent to which an energetically optimal geometry can be obtained. As observed in §3.3.1, the proportion of persistent hydrogen bonds decreases with the amount of hydrogen bonds present, which implies competition between donors and acceptors is another crucial factor. Thus far, the methods in this study have relied on statistical interpretation; in contrast, the focus is now on a selection of examples that fall outside the observed trends.

Referring again to §3.3, < 6% of polymorphs were seen to display completely different hydrogen bonds (*i.e.* no persistent hydrogen bonding between pairs of polymorphs). Do any of these compounds have few competing groups, but no persistent interactions, nonetheless? Polymorphs of this kind are denoted group A. Conversely, 66% compounds were found in §3.2 with a perfect similarity score (S_{HB}). Do any of these have many competing donors and acceptors? Polymorphs of this kind are denoted group B.

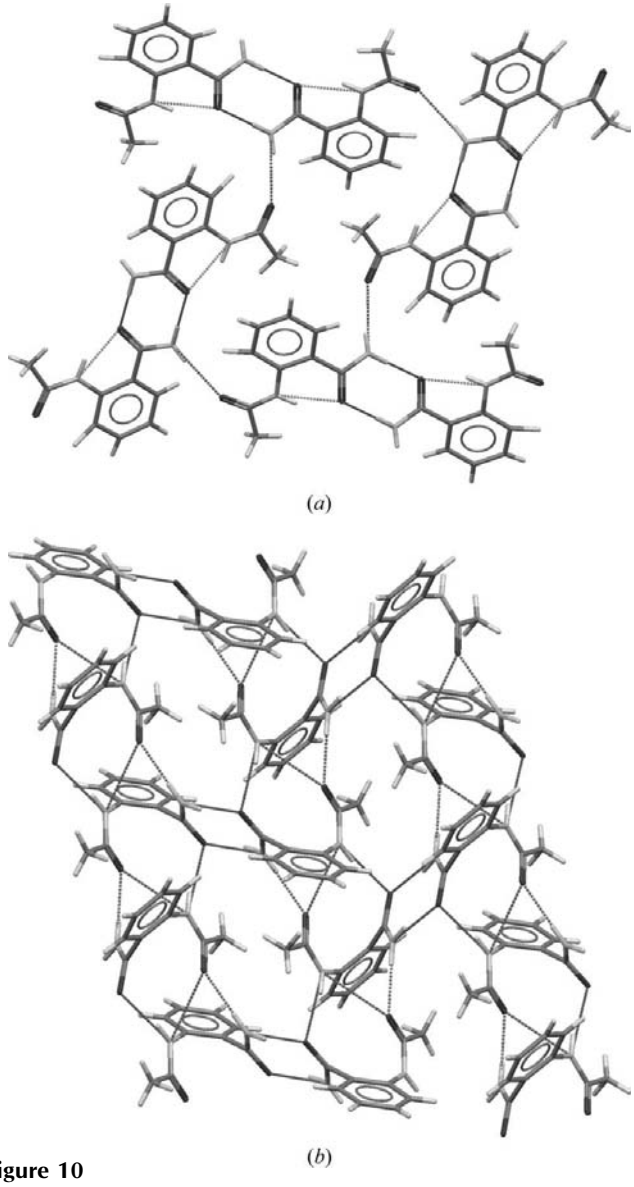


Figure 10
An intramolecular interaction that does not persist: the *o*-acetamidobenzamide system (*a*) ACBNZA and (*b*) ACBNZA01.

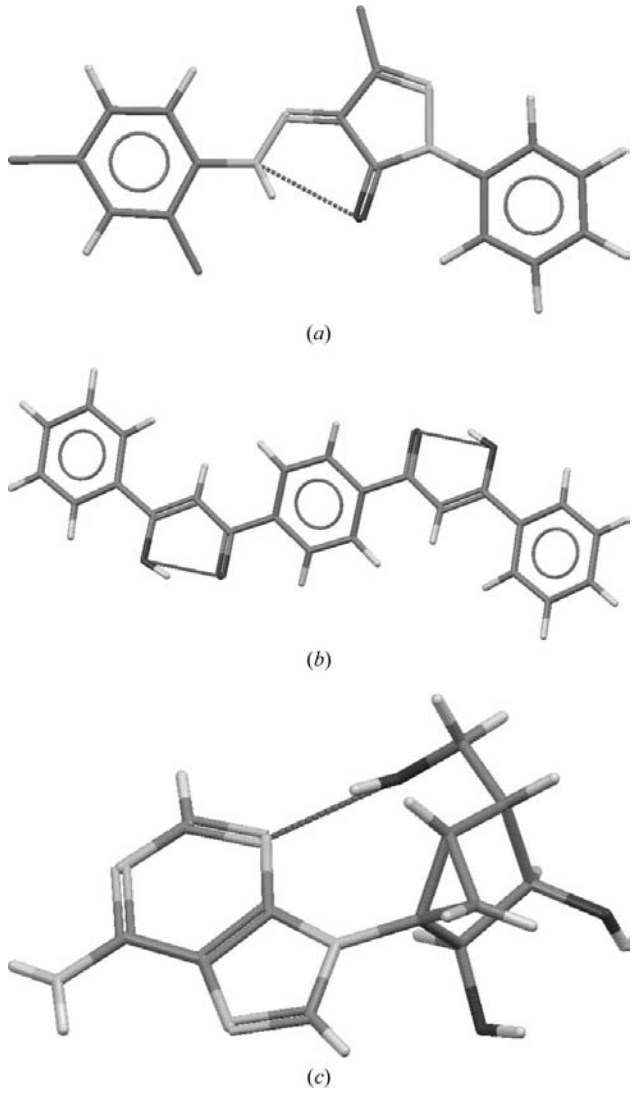


Figure 11
Intramolecular hydrogen bonding which retains certain molecular conformations in packing polymorphs. (*a*) DODDAB, (*b*) APUSAF01 and (*c*) MARBEM.

Only 13 compounds belong to group A. These polymorphs show two distinct types:

(i) structures in which two acceptor groups compete for a lone donor, an example of which are the dimorphs HOHRIF and HOHRIF01 (shown in Fig. 13); and

(ii) structures in which two groups, *e.g. acid*, *amide* having both donating and accepting ability, can form homo or hetero interactions: *acid–acid*, *amide–amide* or *acid–amide*. Polymorphs with this relation are rarer still than those of type (i). The dimorphs BIDLOP, BIDLOP01 and SIHVOU, SIHVOU01 exemplify that situation. BIDLOP and BIDLOP01 display competing carboxyl and hydroxyl moieties, whereas SIHVOU and SIHVOU01 have carboxyl and amide moieties in competition, which are all common, strong interactions (Leiserowitz & Nader, 1977; Allen *et al.*, 1999). In polymorphs in general, competing hydrogen bonds as in type (ii) are more frequent than those of type (i), however, also with simultaneous persistent hydrogen bonds. These extra interactions would seem necessary to these observed polymorphic structures.

In a purely statistical sense, polymorphs belonging to group B are expected to be rare: there naturally exists a high potential for alternative interactions when many combinations of donor and acceptor groups are possible. Persistence against the odds in these polymorphs points towards two types of behaviour. One is for compounds which contain a number of chemically highly similar donors and/or acceptors, where hydrogen bonds are classified as chemically equivalent (potentially between non-equivalent atoms). Refer again to the example in §2.1 and Fig. 2. Persistent hydrogen bonds in these compounds are almost inevitable, given the perceived lack of discrimination between potential hydrogen bond sites.

Persistent interactions despite an array of alternative pairings suggest cooperative structural features, *i.e.* hydrogen-bond motifs. It would seem these features are retained, despite reordering of the crystalline lattice into a new phase, and is suggestive of molecular ordering prior to the crystal phase, *e.g.* in the melt or solution. The dimorphic structures MUROXA and MUROXA01 show this possibility. It is observed that the hydrogen bonding between the two structures is identical, and features a hydrogen-bonded two-dimensional layer. Such two-dimensional isostructurality is not uncommon amongst certain polymorphic structures (Fábián & Kálmán, 2004). Fig. 14 displays the equivalent hydrogen bonding. The two structures differ by alternative stacking of their equivalent layers. In MUROXA the structure repeats every third layer (*i.e.* in the scheme ABA . . .), whilst in MUROXA01 the structure repeats every fifth layer (*i.e.* in the scheme ABCDA . . .).

4. Conclusions

In this report, a thorough study has been carried out on the nature of hydrogen bonding in polymorphic crystal structures in the CSD. Using a new automated analysis, the significance of chemically equivalent hydrogen bonds found in multiple forms was investigated: hydrogen-bond persistence. Persistent hydrogen bonds are very common: in 83% of polymorphic compounds at least half of their hydrogen bonds persist, whilst in 66% all hydrogen bonds persist. Classifying structures based on the number of persistent bonds shows that 93.7% of

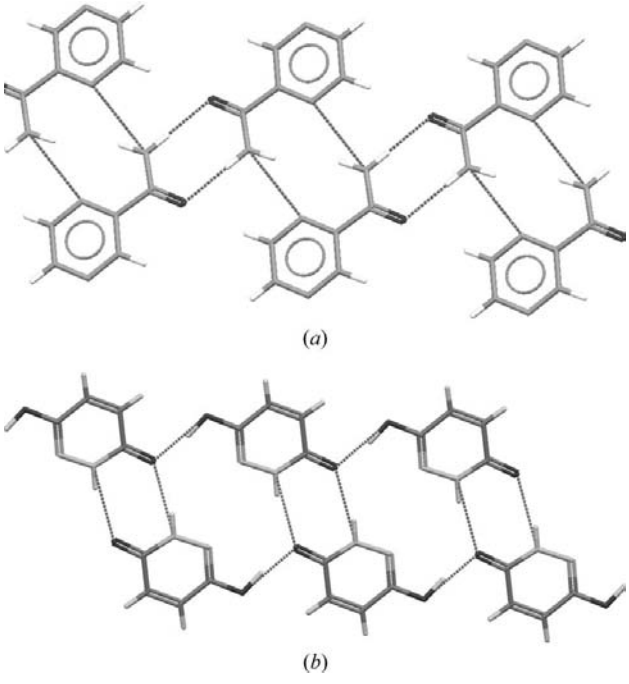


Figure 12
Polymorphs with unchanged molecular conformations without stabilizing intramolecular hydrogen bonds. (*a*) Pyrazine-2-carboxamide in PYRZIN01; (*b*) 1,2-dihydropyridazine-3,6-dione in MALEHY01.

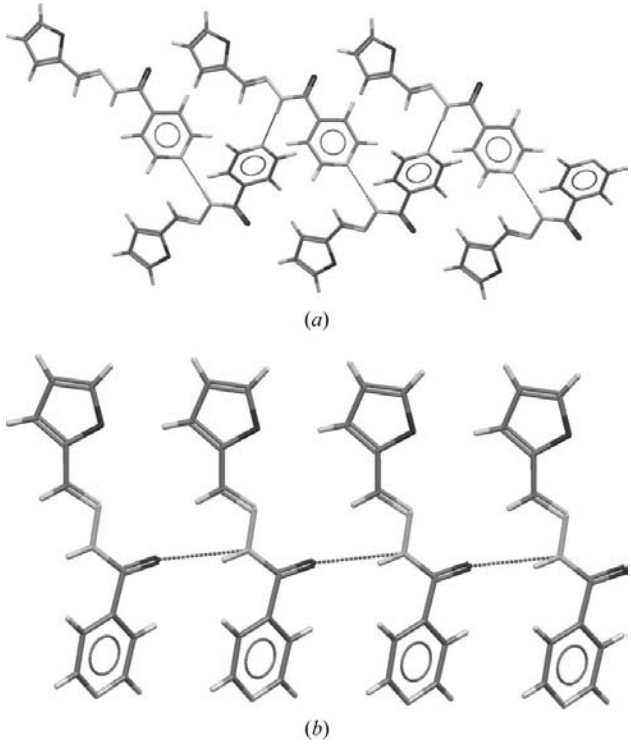


Figure 13
Amide-pyridyl hydrogen bonds in HOHRIF *versus* amide–amide chains in HOHRIF01.

polymorph pairs share at least two hydrogen bonds. A lesser proportion share more than this number but the trend decays slowly as the number of shared bonds increases. Geometric criteria were used to systematically identify hydrogen bonding in crystal structures and it has been shown that the choice of the criteria has only a small influence on the results.

Chemical influences on the persistence of hydrogen bonds proved to be significant: individual chemical groups and certain donor–acceptor pairings strongly persist, *e.g.* the ammonium–carboxylate interaction is persistent in 80.8% (±4.2%) of cases, pointing to the relationship between stronger interactions and their likelihood. This link to potential stabilization carries through when comparing the geometries of the hydrogen bonds. The well known preference for linear and short hydrogen bonds is observed, and we find that this trend is followed more strongly for those interactions that persist in polymorphic structures.

Two structural subsets, conformational and packing polymorphs, were obtained with a new procedure to compare molecular conformations, and analysed as separate groupings. 55% of pairs of polymorphic structures were classified as having a significant change in molecular conformation. The hydrogen-bond geometry considerations reveal that conformational polymorphs show a greater trend for both stronger persistent and weaker unique interactions. It would seem that suboptimal hydrogen bonds, in the place of strong, persistent interactions, are less likely to remain under conditions of conformational change.

Having found a link between persistent interactions and their potential stabilizing effect, our observations agree well with the accepted role of hydrogen bonding as being energetically crucial to crystal structures. Both identifying the extent of, and emphasizing the influences on common hydrogen bonding can be of value to the study and prediction of crystal forms, be they polymorphs, cocrystals or solvates. The results presented here could be used, for example, to assess a particular polymorphic system and guide a drug-development scientist considering the possible modifications to a known form. To illustrate, a hydrogen bond between groups X and Y is observed with distance d and angle θ. Given this information our results can provide an indication, based on the CSD, of how likely this type of interaction is to reoccur in an alternative crystal form. Extensions may be envisaged in which predicted crystal structures for a compound are compared with known form(s) based on properties of their persistent and unique hydrogen bonds, leading to an assessment of the potential for undiscovered forms. Exploring these avenues is beyond our current scope; however, we are confident, based on our results, that the presence or absence of particular hydrogen bonds can be applied successfully in the assessment of polymorphic compounds.

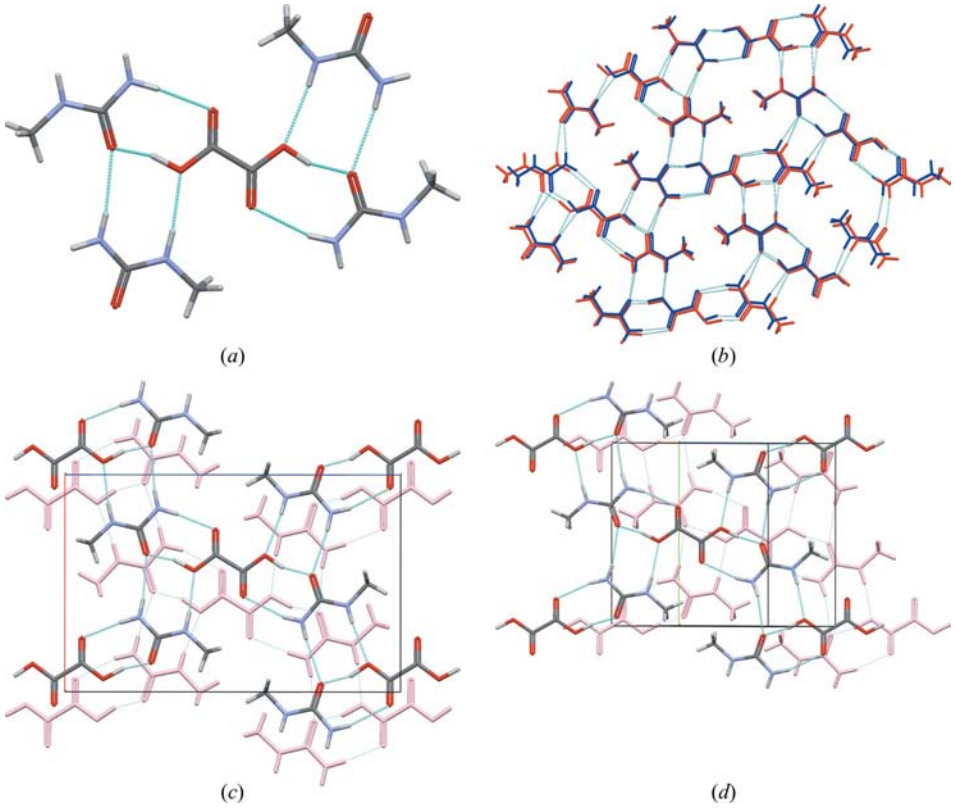


Figure 14
Hydrogen bonding in the MUROXA and MUROXA01 crystal structures. (*a*) The bis(*N*-methyl-urea) oxalic acid motif; (*b*) overlay of the MUROXA (blue) and MUROXA01 (red) hydrogen-bonded sheets in their respective crystal packing arrangements. (*c*) Projections showing the inter-layer packing in the structures of MUROXA and (*d*) MUROXA01. Hydrogen-bonded layers are in the plane of the figure; a 2_1 symmetry operator relates adjacent layers in both cases, corresponding to a shift parallel to the (100) plane in MUROXA, and along the $[20\bar{1}]$ vector in MUROXA01.

APPENDIX *A*
Conformational analysis

Molecules in pairs of polymorphic structures were analysed for conformational change. There are various methods for characterizing a conformational change (see *e.g.* Brameld *et al.*, 2008; Gavezzotti, 2008; Weng *et al.*, 2008). Here a comparison of equivalent non-terminal proper torsion angles is applied, which may be quickly interpreted as a 'significant' variation to allow the classification of conformers as similar or dissimilar. Indeed, only one torsion angle need vary by a defined threshold to indicate a conformational change, even though many other torsions may vary as well. The algorithm identifies pairs of equivalent non-terminal proper torsion angles by persistent equivalent atoms, using an atom canonicalization procedure (Morgan, 1965) with an

improved algorithm (Cole *et al.*, 2001). The angles can then be identified and subsequently compared, even if the atomic labelling in each separate database entry is not equivalent. For every molecule pair, the set of torsion-angle differences is computed and stored for later assessment. A further requirement is that the computed torsion differences must be standardized for correct interpretation, the origins of which lie with their periodicity over 360°. Details of these origins and how the standardization is achieved are set out below.

A1. Standardized torsion-angle differences

Torsion angle measurements have an inherent phase property that can complicate correct statistical representation (Allen & Johnson, 1991). Torsion angles (τ_i) may be considered as individual values of a periodic function with a phase restriction (Klyne & Prelog, 1960). The issue encountered may be illustrated by considering two angles, $i = 150°$ and $j = -110°$. Their absolute difference is $|i - j| = 260°$, which due to the underlying phase (modulo 360° or 2π radians), is equivalently 100°. That is, in general terms torsion angles are a periodic function f of the form

$$f(\tau) = f(\tau + 2n\pi)\ n = 1, 2, 3\ldots . \tag{5}$$

100° is clearly a more appropriate description of the difference: in effect the torsion angles can be no more dissimilar than 180° before they start to approach one another again. The remaining part of the section outlines a standardization method for all torsion angle differences that ensures measurements are represented correctly.

We define the torsion angle difference

$$|\Delta\tau_{i,ab}| \equiv |\delta\tau_i| \tag{6a}$$

$$= |\tau_{i,a} - \tau_{i,b}| \in [0, 360°], \tag{6b}$$

where $\tau_{i,a}$ is the ith torsion angle in polymorph a, and likewise for polymorph b.

The values of $18 < |\Delta\tau_i| \leq 360°$ may however be represented more appropriately for accurate statistics by their circular complement. The following equations define these requirements

$$|\tau_i^*| = |\Delta\tau_i \text{ for } 0 < |\Delta\tau_i| \leq 180° \tag{7a}$$

$$|\Delta\tau_i^*| = 360° - |\Delta\tau_i| \text{ for } 180 < |\Delta\tau_i| \leq 360°. \tag{7b}$$

The same solution is achieved systematically for all observations with some use of trigonometry by mapping $\Delta\tau_i$ from polar to a Cartesian representation, and then applying an inverse mapping

$$\Delta\tau_i^* = 2\tan^{-1}\left[\frac{\sin\left(\frac{\Delta\tau_{i,ab}}{2}\right)}{\cos\left(\frac{\Delta\tau_{i,ab}}{2}\right)}\right], \tag{8}$$

where

$$|\Delta\tau_i^*| \in [0, 180°]. \tag{9}$$

Because $\tan^{-1}(x)$ is a multi-valued function centred at 0° with period 180°, *i.e.*

$$\tan^{-1}(x) \in [-90, 90°] \text{ for } -\infty < x < \infty, \tag{10}$$

a factor of $\Delta\tau_i/2$ maps the input to within the same bounds over which $\tan^{-1}(x)$ is defined. The result is then rescaled by a factor of 2 to return the correct mapping. Using (8), $|\Delta\tau_i^*|$ may easily be computed for every $\Delta\tau_{i,ab}$ value to give a standardized difference in torsion angles.

A2. Distribution of $|\Delta\tau^*|$

The distribution of absolute torsion-angle differences, $|\Delta\tau^*|$, is investigated in pairs of polymorphs to reveal underlying trends. The entire polymorphic CSD subset (1677 structures) is first considered, and second the hydrogen-bonding polymorphic subset (882 structures). In the entire structure set $|\Delta\tau^*|$ ranges over 0–180°, but a very high majority of $|\Delta\tau^*|$ lies around 0°: equivalent torsion angles are very common in pairs of polymorphs: 32 471 of 46 741 observations (69.5%) lie between the bounds of 0 and 5.7°. A significant proportion of these angles are constrained in ring systems (rigid π systems being completely restrictive, although any other type of cyclicity restricts torsion angle variability to an extent). The true number of freely rotatable non-terminal torsions that do not vary between pairs of polymorphs were not identified as the current focus is on selecting those torsion angles which significantly *differ*, and not on the proportion that do so.

A histogram ignoring the low $|\Delta\tau^*|$ frequencies is required for greater resolution of the distribution at higher angles. Fig. 15 plots a histogram of $|\Delta\tau^*|$ *versus* frequency density between $|\Delta\tau^*| = 40$ and 180°, the inset shows the distribution over a $|\Delta\tau^*| = 10$ and 180° range. A minimum in the distribution at 40° may be identified, suggesting a lower bound to isolate the torsion angle differences corresponding to significant alternative conformations. Note that classically syn- or anti-periplanar conformations are defined with up to a 30° deviation from the ideal planar angle (0 or 180°), hence the

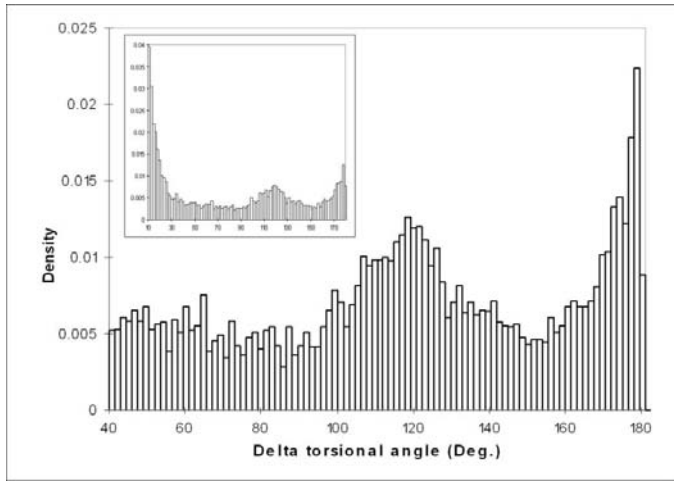


Figure 15
Distribution of equivalent non-terminal $|\Delta\tau_i^*|$ for pairs of polymorphs between 40 and 180°. (Inset shows same distribution over the 10–180° range).

chosen 40° threshold would also seem to suggest a significant deviation in the $|\Delta\tau^*|$ measurements.

Regions of relatively high density in the distribution can be observed: (*a*) at 180° corresponding to complete half-rotations (often the flipping of a single chemical substituent), or *e.g. syn* to antiperiplanar transformations; (*b*) 120° corresponding to *e.g.* (+)-*gauche* to (−)-*gauche* rotations, or synperiplanar to anticlinal transformations. Recently, Weng *et al.* (2008) also observed substituent flipping and *syn*/*anti* rotations as very common conformational variations, using a method for conformational clustering of complete molecules in the CSD based on r.m.s. atom distances.

PTAG and LF would like to acknowledge funding from the Pfizer Institute for Pharmaceutical Materials Sciences.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Allen, F. H. & Johnson, O. (1991). *Acta Cryst.* B**47**, 62–67.
Allen, F. H., Motherwell, W. D. S., Raithby, P. R., Shields, G. P. & Taylor, R. (1999). *New J. Chem.* **23**, 25–34.
Bats, J. W., Parsch, J. & Engels, J. W. (2000). *Acta Cryst.* C**56**, 201–205.
Bernstein, J. (1993). *J. Phys. D Appl. Phys.* **26**, B66–B76.
Bernstein, J., Davis, R. E., Shimoni, L. & Chang, N.-L. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 1555–1573.
Bilton, C., Allen, F. H., Shields, G. P. & Howard, J. A. K. (2000). *Acta Cryst.* B**56**, 849–856.
Blagden, N., Davey, R. J., Liebermann, H. F., Williams, L., Payne, R., Roberts, R., Rowe, R. & Docherty, R. (1998). *J. Chem. Soc. Faraday Trans.* **94**, 1035–1044.
Brameld, K. A., Kuhn, B., Reuter, D. C. & Stahl, M. (2008). *J. Chem. Inf. Model.* **48**, 1–24.
Brock, C. P. & Minton, R. P. (1989). *J. Am. Chem. Soc.* **111**, 4586–4593.
Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *Acta Cryst.* B**58**, 389–397.
Ciunik, Z. & Desiraju, G. R. (2001). *Chem. Commun.* pp. 207–704.
Chisholm, J., Pidcock, E., Streek, J. van de, Infantes, L., Motherwell, S. & Allen, F. H. (2006). *CrystEngComm*, **8**, 11–28.
Cole, J. C., Yao, J. W., Shields, G. P., Motherwell, W. D. S., Allen, F. H. & Howard, J. A. K. (2001). *Acta Cryst.* B**57**, 88–94.
Desiraju, G. R. & Steiner, T. (1999). *The Weak Hydrogen Bond in Structural Chemistry and Biology.* Oxford University Press.
Dunitz, J. D. & Gavezzotti, A. (2005). *Cryst. Growth Des.* **5**, 2180–2189.
Fábián, L. & Kálmán, A. (2004). *Acta Cryst.* B**60**, 547–558.
Galek, P. T. A., Fábián, L., Motherwell, W. D. S., Allen, F. H. & Feeder, N. (2007). *Acta Cryst.* B**63**, 768–782.
Gavezzotti, A. (2008). *CrystEngComm*, **10**, 389–398.
Giron, D. (1995). *Thermochim. Acta*, **248**, 1–59.
Grant, D. J. W. (1999). *Polymorphism in Pharmaceutical Solids*, edited by H. G. Brittain, pp. 1–31. New York: Marcel Dekker, Inc.
Haleblian, J. & McCrone, W. (1969). *J. Pharm. Sci.* **58**, 911–929.
Haleblian, J. K. (1975). *J. Pharm. Sci.* **64**, 1269–1288.
Henck, J. O., Griesser, U. J. & Burger, A. (1997). *Pharm. Ind.* **59**, 165–169.
Klyne, W. & Prelog, V. (1960). *Endeavour*, **16**, 521–528.
Kroon, J., Kanters, J. A., van Duijneveldt-van de Rijdt, J. G. C. M., van Duijneveldt, F. B. & Vleingenthart, J. A. (1975). *J. Mol. Struct.* **24**, 109–129.
Krygowski, T. M., Grabowski, S. J. & Konarski, J. (1998). *Tetrahedron*, **54**, 11311–11316.
Leiserowitz, L. & Nader, F. (1977). *Acta Cryst.* B**33**, 2719–2733.
Lommerse, J. P. M., Stone, A. J., Taylor, R. & Allen, F. H. (1996). *J. Am. Chem. Soc.* **118**, 3108–3116.
Liu, G.-F., Lui, L., Jia, D.-Z. & Yu, K.-B. (2004). *J. Chem. Cryst.* **34**, 835–841.
Macrae, C. F., Bruno, I. J., Chisholm, J. A., Edgington, P. R., McCabe, P., Pidcock, E., Rodriguez-Monge, L., Taylor, R., van de Streek, J. & Wood, P. A. (2008). *J. Appl. Cryst.* **41**, 466–470.
Mascal, M., Infantes, L. & Chisholm, J. (2006). *Angew. Chem. Int. Ed.* **44**, 2–7.
Morgan, H. L. (1965). *J. Chem. Soc.* **5**, 107–113.
Novoa, J. J. & d'Oria, E. (2007). *Engineering of Crystalline Materials Properties*, edited by J. J. Novoa, D. Braga & L. Addadi, pp. 307–332. Dordrecht, The Netherlands: Springer.
Pidcock, E. & Motherwell, W. D. S. (2005). *Cryst. Growth Des.* **5**, 2322–2330.
Pimentel, G. C. & McClellan, A. L. (1960). *The Hydrogen Bond.* San Francisco, USA: W. H. Freeman.
Scheiner, S. (1997). *Hydrogen Bonding. A Theoretical Perspective.* Oxford University Press.
Scott, D. W. (1992). *Multivariate Density Estimation: Theory, Practice and Visualization.* New York, Chichester: John Wiley and Sons.
Singhal, D. & Curatolo, W. (2003). *Adv. Drug. Del. Rev.* **56**, 335–347.
Stahly, G. P. (2007). *Cryst. Growth Des.* **7**, 1007–1026.
Steiner, T. (2002). *Angew. Chem. Int. Ed.* **41**, 48.
Streek, J. van de (2006). *Acta Cryst.* B**62**, 567–579.
Threlfall, T. L. (1995). *Analyst*, **120**, 2435–2460.
Threlfall, T. L. (2003). *Org. Process Res. Dev.* **7**, 1017–1027.
Weng, Z. F., Motherwell, W. D. S., Allen, F. H. & Cole, J. M. (2008). *Acta Cryst.* B**64**, 348–362.