

## On the significance of weak hydrogen bonds in crystal packing: a large databank comparison of polymorphic structures

Leonardo Lo Presti<sup>\*a,b,c</sup>Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

[www.rsc.org/](http://www.rsc.org/)

While there is general consensus on the fundamental role played by strong hydrogen bonds (HB) in crystal packing, the significance of weak CH...X (X = N, O, S, F, Cl) interactions is still debated. Here, ~ 250 polymorph pairs of small molecules with no strong HB donors were retrieved from the Cambridge Structural Database. Statistical analysis tools were applied to look for conserved features among chemically analogue compounds, in terms of crystal packing, self-recognition energetics and lattice cohesion. The occurring frequency of weak HB is significantly higher than that expected for a random distribution of close H...X contacts, but they are seldom conserved in different polymorphs and do not correlate with neither the molecule-molecule nor the lattice energies. Comparison of interaction energies of closest molecular pairs in the solid state shows that weak HB do not generally provide a significant thermodynamic drive toward crystallization. Accordingly, lattice energies of different polymorphs are often dominated by the dispersive/repulsive balance, pointing out the importance of steric and shape factors in determining their self-assembling. This likely indicates that, in most cases, the preference for a particular self-recognition mode arises due to the interaction of the whole charge density distributions, rather than to specific weakly attractive atom-atom contacts. Weak HB, however, might assist this process, providing extra stabilizing terms and advantaging some interaction modes over other ones. Implications on thermodynamics and kinetics of crystallization are discussed.

### 1 Introduction

Experimentally observed crystal structures are the product of an elusive equilibrium of intermolecular forces,<sup>1–3</sup> which rely on potential energy terms of very different physical origins, such as dispersions, “Pauli repulsions”, and polarizations / electrostatics.<sup>4</sup> Developing consistent general rules to recognize structure determinants, *i.e.* specific chemical features that are decisive to produce a certain crystal packing, is likely the major unresolved problem of solid-state organic chemistry. The quest is discouragingly complex, as intermolecular non-covalent interactions (NCI) are up to 2–3 orders of magnitude weaker than covalent bonds,<sup>5</sup> and thus intrinsically prone to compete<sup>3,6</sup> with each other. Difficulties further escalate when “weak bond” interactions<sup>7,8</sup> come into play. Half a century ago, Sutor<sup>9</sup> noticed that CH...O contacts shorter than the sum of the van der Waals radii are ubiquitous in organic crystals, raising a long-standing debate<sup>10–12</sup> on the nature and importance of such interactions, with controversial outcomes.<sup>13–15</sup> Nowadays, it is largely accepted that CH...X interactions, X being an electronegative acceptor, are true hydrogen bonds.<sup>16</sup> Indeed, it has been demonstrated there are no fundamental chemical differences

between strong and weak HB, neither in terms of contact geometry,<sup>17</sup> nor from the viewpoint of the electron density distribution.<sup>6,18–20</sup> Nevertheless, a question of genuine crystallographic interest remains open, namely whether and to what extent such interactions are true structure determinants, *i.e.* which is their significance, if any, in *predicting* and *controlling* crystal structures.

In a recent work,<sup>21</sup> Gavezzotti and myself applied principal component analysis methods to single out multivariate correlations among packing descriptors and lattice energies in a selected pool of crystal structures containing only C–H donors. CH...X (X = N, O) contacts showed some correlations with Coulombic contributions to molecule-molecule interaction energies, but they belonged in most cases to weakly bounded pairs, and generally they did not determine neither the pair energy, nor the lattice one. We thus concluded<sup>21</sup> that such interactions “cannot be taken a priori as reliable and solid crystal building blocks nor can they be entrusted with a general status of reproducible chemical bonds.” By return of post, Taylor rebutted our thesis.<sup>22</sup> On the basis of surface area considerations, he studied the frequencies of expected *vs.* observed CH...X intermolecular contacts in the same pool of crystal structures and pointed out that such interactions are significantly more numerous than in the case of random packing. Therefore, his conclusion<sup>22</sup> was that “X...H interactions [...] are relevant in stabilizing crystal packing arrangements, including for the contentious cases when X = Cl and, especially, X = F”.

The present contribution aims at shedding light on this issue, providing some suggestions that may help to reconcile such apparently conflicting views. Hydrogen bond propensity and

<sup>a</sup> Università degli Studi di Milano, Department of Chemistry, Via Golgi 19 I–20133 Milano (Italy).

<sup>b</sup> Centre for Materials Crystallography, Århus University, Langelandsgade 140, DK–8000 Århus C. (Denmark).

<sup>c</sup> Istituto di Scienze e Tecnologie Molecolari, Italian CNR, Via Golgi 19 I–20133 Milano (Italy).

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

relative energetics are compared for ~ 250 pairs of polymorphs of small–medium size organic molecules with no strong HB donors, all retrieved from the Cambridge Structural Database (CSD).<sup>23</sup> By definition, polymorphs differ just by their crystal packing, offering the opportunity of disclosing how a change in number and type of close intermolecular contacts influences the lattice structure and energetics. If specific weak HB are indispensable to supramolecular arrangements in a crystal, some of them should be conserved within a given polymorph pair, at least in the presence of similar molecule–molecule interaction modes. At the same time, to have a chance of using weak HB as practical “molecular hinges” toward the synthesis of desired crystal structures, a 1:1 correspondence should exist among different polymorphs and different hydrogen bond patterns. This study should give some guiding principles for understanding whether weak HB can act as true structure determinants, and what is their role in polymorphism.

## 2 Methods

### 2.1 Crystal structure search.

A total of 497 crystal structures (hereinafter, “POLY–all” dataset) were retrieved from the 5.39 release of CSD<sup>23</sup> (2017) using the CCDC software Conquest.<sup>24</sup> The following specifications were applied: just 1 or 2 molecules in the asymmetric unit (ASU), *i.e.*  $Z' = 1$  or 2; one chemical residue per structure (no co–crystals, no solvates); no disorder, no errors, no ions, no polymers, no coordination complexes, no powder structures. Only C, H, N, O, S, F and Cl atoms were allowed, in conjunction with complete 3D coordinates and a crystallographic  $R$  factor lower than 5.6 %. Structures bearing either –NH, –OH or –SH HB donors were excluded, to get rid of competing stronger HB interactions. We looked for CSD records explicitly citing the keyword “polymorph”, selecting those for which at least two distinct polymorphs were actually available. Duplicates were eliminated by hand: we kept the structures with the lowest  $R$  factor or, when very similar  $R$ 's were available, those determined at the most similar temperatures. C–H bond lengths were renormalized to the usual neutron–derived estimate of 1.08 Å<sup>25</sup> through the Retic module of the CLP/PIXEL program package.<sup>26</sup> Finally, whenever a structure bore unrealistically repulsive molecule–molecule interactions, it was removed *a posteriori*, together with its related polymorph(s). Upon application of all the filters, ~ 36 % of the retrieved polymorph pairs/triplets were deemed of sufficient quality for the accurate study of crystal packing and lattice energetics. We followed a prudence criterion, as including uncertain or suspicious data would likely add noise – or, worse, systematic biases – to the database without providing relevant information. The current dataset is nevertheless large enough to achieve accurate conclusions (see below). The full list of CSD refcodes in POLY–all can be found in the ESI (Table S1).

### 2.2 Analysis of close contacts.

Intermolecular contacts A...B in POLY–all were searched on a geometrical basis, with the only requirements that (i) the involved atoms were on “line–of–sight” (*i.e.* without a third

atom between them)<sup>15,22</sup> and (ii) their distance,  $R_{AB}$ , was lower than the sum of standard atomic radii (SAR), multiplied by a tuning factor,  $P$ , to highlight increasingly shorter contacts.<sup>21</sup>  $P$  values of 0.90, 0.95, and 1.00 were selected to focus, respectively, on “extremely short”, “very short” and “short” contacts.<sup>21</sup> Each A...B interaction thus unequivocally bridges a pair of molecules in the crystal.

Three sets of SAR were checked, namely those proposed by Rowland & Taylor,<sup>27</sup> Alvarez<sup>28</sup> and Bondi.<sup>29</sup> As expected, different choices of SAR led just to minor quantitative differences, which did not alter the general trends and outcomes. According to previous works,<sup>21,22</sup> the results based on the radii by Rowland & Taylor<sup>27</sup> are here discussed. As in ref. 21, we define a “shortness index”,  $B_s$ :

$$B_s = \sum_k 100 \cdot \frac{(R_{AB}^0 - R_k)}{R_{AB}^0} \quad (1)$$

$R_k$  is the  $k$ –th distance satisfying the above described contact conditions, and  $R_{AB}^0$  the corresponding sum of SAR.  $B_s$  thus summarizes the total percent reduction in specific contact distances for a given structure, and can be taken as a reasonable estimator of the frequency and significance of atom–atom close contacts.

Following Taylor,<sup>15,22</sup> we also computed the  $R_F$  metrics for POLY–all. In a nutshell, if molecules in the crystal were present in totally random orientations, the expected number of A...B close contacts in the  $i$ –th structure,  $E_i(A...B)$ , should be

$$E_i(A...B) = N_A \cdot S_B / S_T \quad (2)$$

Where  $N_A$  is the number of A atoms that satisfy the close contact conditions and  $S_B$  the extent of total molecular surface ( $S_T$ ) spanned by B atoms. Being  $O_i(A...B)$  the observed number of A...B contacts in the  $i$ –th structure,  $R_F$  is defined as

$$R_F(A...B) = \frac{\sum_i O_i(A...B)}{\sum_i E_i(A...B)} \quad (3)$$

and measures whether and to what extent a given A...B interaction in the data pool occurs more (or less) often than expected if the packing was completely random. Surface areas were computed with the method of Infantes & Motherwell,<sup>30</sup> and 95 % confidence intervals on  $R_F$  were calculated by applying the bootstrapping procedure suggested by Taylor.<sup>15</sup>

### 2.3 Energy calculations.

Molecule–molecule interaction energies ( $E_{mol}$ ) and lattice cohesive energies ( $E_{coh}$ ) in POLY–all were obtained with the Atom–Atom Coulomb–London–Pauli (AA–CLP) approach.<sup>31,32</sup> A statistically representative subset of 105 structures (“POLY–pix” dataset) was also selected to perform more accurate energy calculations with PIXEL.<sup>33</sup> To this end, charge densities of symmetry–independent molecules at their in–crystal geometries were computed at the MP2/6–31G(p,d) level by the Gaussian16 package.<sup>34</sup> The usual grid condensation factor of 4 was used throughout. See ESI (Table S2) for the full list of refcodes within the POLY–pix subset of structures.

DFT simulations were carried out with Gaussian16 using the Minnesota–class functional M06,<sup>35</sup> in conjunction with the

triple zeta 6–311G(p,d) basis set,<sup>36</sup> to estimate the strength of selected hydrogen-bonded molecular pairs as a function of their H-Acceptor distance (Section 3.7). M06 takes into account non-locality in the exchange–correlation potential by including terms that depend on the kinetic energy density of the electron gas reference.<sup>35</sup> Thus, it is able to retrieve, at least partially, correlation effects related to dispersion interactions.<sup>37,38</sup>

#### 2.4 Quality assessment and reproducibility

The commercial CCDC data mining utilities Conquest<sup>24</sup> and Mercury<sup>39</sup> were used to retrieve and analyse the structures. The full list of polymorphs included in the databank is deposited as supplementary materials for this paper (see ESI). Calculations of short atom–atom distances, hydrogen bond geometries, molecular surface and  $B_s$  and  $R_F$  indices were carried out with in-house software. All the quantum simulations were performed using commercial programs. The AA-CLP and PIXEL packages are available free of charge from <http://www.angelogavezzotti.it/>.

### 3. Results

#### 3.1 Database composition and properties

Table 1 summarizes the main properties of the databases employed in this work. The POLY-*all* data pool contains 243 distinct chemical species, spanning 232 polymorph pairs and 11 polymorph triplets (Table 1). On average, donor groups are roughly 3–4 times the number of acceptors, mirroring the expected H / heteroatoms ratios in organic crystals. However, there are always more than four acceptors per molecule (Table 1, last row), ensuring that the casuistry of H...A interactions is large enough to provide meaningful statistics. Frequency considerations suggest that H...O interactions should occur more than twice H...N and H...(S, halogen) ones. Most structures contain just 1 molecule in the asymmetric unit (86–87 %), but a non-negligible minority (13–14 %) of  $Z'=2$  structures is also present (Table 1).

Both databanks contain a majority of aromatic compounds, as it can be inferred from the large amount of C- $sp^2$  atoms. It is worth noting that 2/3 of the molecules included in the data pool contain more than 35 atoms; POLY-*all* spans a very wide range of molecular weights, from EKIPUL (111 atoms) to YOLDAF (8 atoms). Just molecules of intermediate size were included in the POLY-*pix* subset, going from CINYEE (64 atoms) to DCLBQN (12 atoms). The similarity of average parameters between the databases (Table 1) ensures that POLY-*pix* gathers a statistically representative sample of POLY-*all* structures; the t-test ensures that the differences are not significant at a very high confidence level (> 99 %).

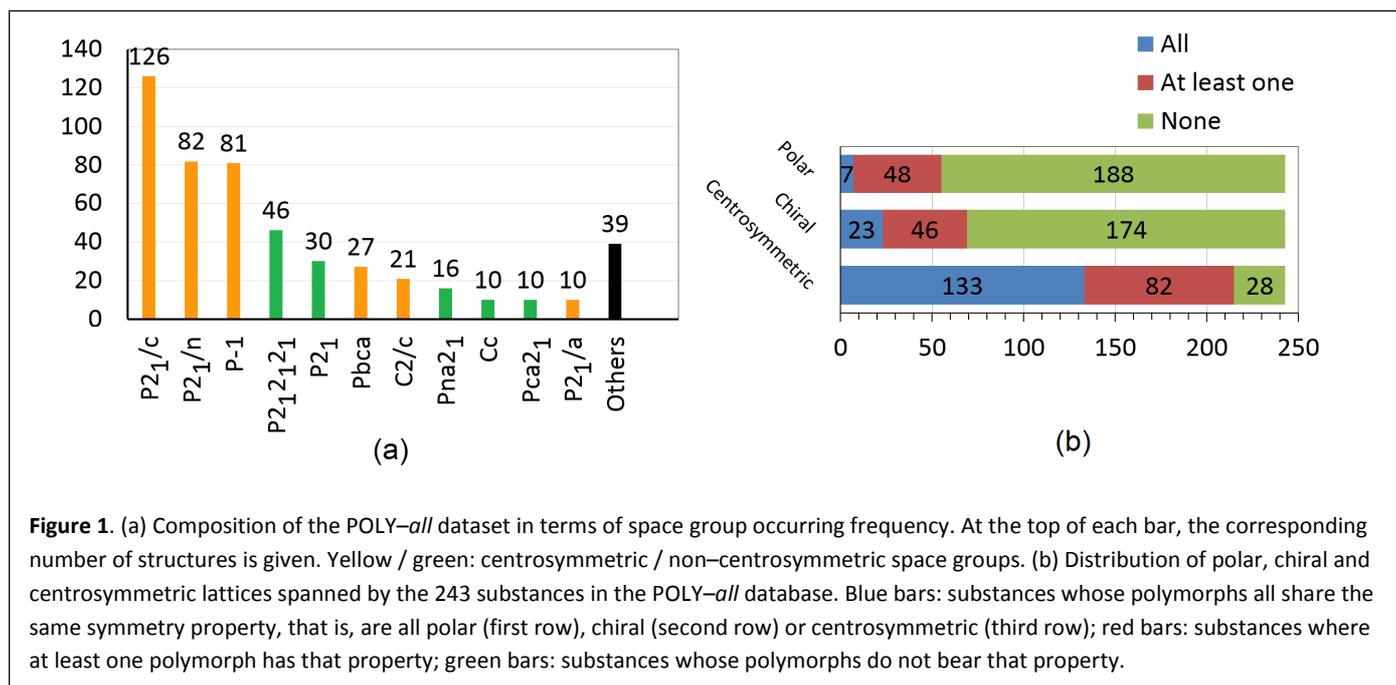
Figure 1 shows the composition of the database in terms of general crystallographic parameters, such as frequencies of occurring space groups and their symmetries. The  $P2_1/c$  symmetry (whatever is the choice of the crystallographic reference frame) occurs most frequently (43.9 %), followed by the triclinic  $P\bar{1}$  (16.3 %) one. As expected, the most recurrent chiral groups are the monoclinic  $P2_1$  (6.0 %) and the orthorhombic  $P2_12_12_1$  (9.3 %) ones. Higher cell symmetries are seldom represented (< 2 %). Most substances (54.7 %)

crystallize in forms that are invariably centrosymmetric, but it is also frequent that one polymorph is centrosymmetric, while the other is not (33.7 %).

**Table 1.** Chemical composition of the POLY-*all* and POLY-*pix* databases (see text).  $N_{\text{atom}}$  is the atom count. All the C–H groups are considered as possible donors, while acceptors can be O, N, F, S and Cl.

	POLY- <i>all</i>	POLY- <i>pix</i>
Number of structures	497	105
Polymorph pairs / triplets	232 / 11	51 / 1
Total molecules	569	118
Structures with $Z' = 1 / 2$	425 / 72	92 / 13
Total atoms	21722	4129
Donor / Acceptor ratio	41.3 % / 12.6 %	39.2 % / 12.2 %
$0 < N_{\text{atom}} \leq 20$ (small)	6.1 %	3.8 %
$20 < N_{\text{atom}} \leq 35$ (medium)	33.8 %	34.3 %
$35 < N_{\text{atom}}$ (large)	60.0 %	61.9 %
Average number of donor / acceptor groups per molecule		
<i>Donors:</i>		
H-[C≡], acetylenic	0.01	0.02
H-[C=], aromatic or vinylic	7.95	7.86
H-[C-], aliphatic	7.81	5.85
C $sp$	0.17	0.19
C $sp^2$	12.17	12.10
C $sp^3$	3.91	2.75
<i>Acceptors:</i>		
=O (any)	1.58	1.58
–O– (ether)	0.74	0.77
<i>Total O</i>	2.32	2.36
>N- $sp^3$	0.56	0.47
=N- $sp^2$	0.51	0.32
≡N $sp$	0.15	0.15
<i>Total N</i>	1.22	0.94
=S (any)	0.07	0.02
–S– (tioether)	0.39	0.17
–F	0.19	0.21
–Cl	0.29	0.47
<i>Total acceptors</i>	4.49	4.14

Fewer substances (11.5 %) avoid centrosymmetric space groups at all, often due to chirality requirements; accordingly, chiral space groups are well represented (17.9 %). In contrast, racemates are rare (4.2 %), as the dataset includes a large majority of aromatic achiral compounds. Interestingly, even though 17.6 % of the 497 structures in the POLY-*all* dataset crystallize in polar point groups, compounds where both polymorphs have polar unit cells are sporadic (2.9 %); in contrast, cases where one polymorph is polar and the other is not are almost ten times more frequent (19.8%).



**Figure 1.** (a) Composition of the POLY-all dataset in terms of space group occurring frequency. At the top of each bar, the corresponding number of structures is given. Yellow / green: centrosymmetric / non-centrosymmetric space groups. (b) Distribution of polar, chiral and centrosymmetric lattices spanned by the 243 substances in the POLY-all database. Blue bars: substances whose polymorphs all share the same symmetry property, that is, are all polar (first row), chiral (second row) or centrosymmetric (third row); red bars: substances where at least one polymorph has that property; green bars: substances whose polymorphs do not bear that property.

### 3.2 On the nature and frequency of close HB contacts

The POLY-all databank represents a minor subset of the general ensemble of CH...X (X: HB acceptor)-containing structures in the CSD. Therefore, we verified whether the distribution of such contacts mirrors the general trends deduced from larger databases.<sup>21</sup>

The only relevant HB donors in POLY-all are either sp<sup>2</sup> (aromatic) or sp<sup>3</sup> (aliphatic) C-H groups. Acetylenic hydrogens are present here just in HEHWOJ and YAXCEH, and produce close contacts only with linear nitrile groups. In any case, the rarity of ≡C-H donors reflects the low amount of structures with terminal acetylenic groups (0.36 %) that are deposited in the current CSD release. Table 2 displays the number and average H...X distances of HB contacts in POLY-all as a function of *P* (Section 2.2), depending on the chemical nature of the acceptor group. The correspondent distributions for aromatic and aliphatic donor H atoms are very similar to each other, with differences comparable in most cases with the statistical noise (Table S3 ESI).

As expected,<sup>21</sup> aliphatic N atoms are poor HB acceptors, even though they are slightly more frequent than N sp<sup>2</sup> ones (Table 1). Indeed, N sp<sup>3</sup> atoms are often sterically screened by their aliphatic substituents and thus less available to intermolecular H...N contacts. In general, steric and shape issues should always be kept in mind when discussing crystal packing from a statistical viewpoint, as they might be, and in fact are, as much important as directional attractive interactions in producing ordered packing modes (see *infra*).

The number of close HB contacts sharply decreases as the *P* cut-off becomes tighter. At the *P* = 1.0 level, 10.6 % of the 497 structures in the POLY-all database bear no short contacts at all. This quantity roughly doubles for any 5 % decrease in *P*, becoming as large as 27.7 % for *P* = 0.95 and 60.1 % for *P* = 0.90. As *P* is lowered, also the average H...Acceptor contact distances,

*R*<sub>av</sub>, decrease, as less and less long interactions are taken into account.

**Table 2.** Number of close contacts in POLY-all, as a function of the chemical nature of the acceptor group and the distance cutoff *P*.

Acceptor	<i>N</i> <sub>c</sub> <sup>a</sup>			<i>R</i> <sub>av</sub> /Å <sup>b</sup>		
	<i>P</i> = 0.90	0.95	1.00	0.90	0.95	1.00
Aliphatic N	1	6	26	// <sup>c</sup>	2.54(2)	2.64(1)
Aromatic N	16	89	204	2.39(2)	2.515(8)	2.601(7)
-CN or -N=N	16	61	128	2.39(2)	2.50(1)	2.587(9)
-O- ether	22	83	210	2.36(1)	2.451(8)	2.553(7)
C=O (Ox.)	175	424	751	2.335(5)	2.419(4)	2.503(4)
N=O (Ox.)	53	157	271	2.339(7)	2.439(7)	2.512(7)
S=O (Ox.)	24	55	93	2.33(1)	2.42(1)	2.50(1)
-S- (Sulphur)	0	15	64	// <sup>c</sup>	2.726(7)	2.815(8)
C=S (Sulphur)	0	3	15	// <sup>c</sup>	2.68(3)	2.82(2)
O=S (Sulphur)	0	1	2	// <sup>c</sup>	// <sup>c</sup>	2.79(9)
-F	0	6	34	// <sup>c</sup>	2.38(1)	2.48(1)
-Cl	0	9	56	// <sup>c</sup>	2.689(8)	2.780(8)
Total	307	909	1854			

<sup>a</sup> Number of contacts satisfying the close contact requirements detailed in Section 2.2.

<sup>b</sup> Average H...A distance in Å, with the estimated standard deviation in parentheses.

<sup>c</sup> No averages are performed if the number of contacts is lower than 2.

This also reduces the data variability: *R*<sub>av</sub> of very short HB are much similar to each other with respect to the distributions for large *P* (Table 2). In other words, strong HB have more similar contact distances, no matter the chemical nature of the donor

and acceptor atoms, possibly mirroring the fact that very strong CH...X interactions are more similar to their classical OH...X and NH...X counterparts.<sup>6,20</sup>

The most consistent HB acceptors are the carbonyl oxygen and terminal -NO<sub>2</sub> substituents, likely due to concurrent steric and electronegativity reasons. As for the S acceptor, thioether is preferred over sp<sup>2</sup> S on absolute grounds, but this clearly reflects the higher amount of sp<sup>3</sup> S atoms in the POLY-*all* databank (Table 1).

**Table 3.** Acceptor capability: total number of CH...X (X=N, O, S, F, Cl) contacts set up by each chemically different acceptor, divided by the total number of that acceptor atoms in the POLY-*all* database.

Acceptor	P=	0.90	0.95	1.00
Aliphatic Nitrogen		0.00	0.02	0.08
N aromatic		0.06	0.31	0.70
-CN or -N=N		0.18	0.69	<b>1.46</b>
-O- ether		0.05	0.20	0.50
C=O (Oxygen)		0.36	0.86	<b>1.52</b>
N=O (Oxygen)		0.15	0.44	0.76
S=O (Oxygen)		0.44	<b>1.02</b>	<b>1.72</b>
-S-	// <sup>b</sup>		0.07	0.29
C=S (Sulphur)	// <sup>b</sup>		0.27	<b>1.36</b>
O=S (Sulphur)	// <sup>b</sup>		0.03	0.07
-F	// <sup>b</sup>		0.05	0.28
-Cl	// <sup>b</sup>		0.06	0.34

<sup>a</sup> Entries corresponding, on average, to more than 1 contact per atom are highlighted in bold.

<sup>b</sup> Too few data for a meaningful statistics.

A better estimator of the acceptor capability is the mean number of contacts per acceptor, that is, the ratio between the total number of contacts for a given acceptor and the total number of that acceptor species in the databank (Table 3). Inspection of Table 3 confirms that sp<sup>2</sup> oxygen and sulphur acceptors are preferred over the sp<sup>3</sup> ones, while for N atoms the preference follows the expected sequence sp<sup>2</sup> > sp<sup>3</sup>. C=O, C=S, S=O and C≡N groups set up, on average, more than one contact per atom at P = 1.00, and their contact frequency remains relatively high upon reduction of the cut-off parameter. The only exception is C=S, whose acceptor capability becomes immaterial below P = 0.95. However, this is likely also due to the very low amount of sp<sup>2</sup> S atoms in the databank. In contrast, despite being even more represented than sp N (Table 1), halogen atoms display low acceptor capabilities, which rapidly fade away upon lowering P.

These results nicely reflect our previous findings, both qualitatively and quantitatively.<sup>21</sup> It can be thus concluded that the POLY-*all* database mirrors the expected relative frequencies of HB close contacts, as determined from the analysis of a larger ensemble of structures.

### 3.3 R<sub>F</sub> metrics

Being X a non-H atom, Table 4 shows the R<sub>F</sub> metrics, and their 95 % confidence intervals (see Section 2.2), for H...X, H...H and X...X contacts as a function of the cut-off parameter P.

**Table 4.** R<sub>F</sub> metrics<sup>a</sup> for H...X, H...H and X...X (X = N, O, S, F, Cl) primary interactions with contact distance lower than P · (sum of SAR)<sup>b</sup>.

H...X	P=	0.90	0.95	1.00
H...O		5.1 (5.1–5.1)	4.4 (4.4–4.5)	3.6 (3.5–3.6)
H...N		2.4 (2.4–2.5)	2.9 (2.8–2.9)	2.6 (2.5–2.6)
H...S	// <sup>c</sup>		1.2 (1.2–1.3)	1.4 (1.4–1.5)
H...F	// <sup>c</sup>		1.8 (1.6–1.9)	1.8 (1.6–1.9)
H...Cl	// <sup>c</sup>		0.8 (0.8–0.8)	1.8 (1.7–1.8)
H...H		0.1 (0.1–0.1)	0.1 (0.1–0.1)	0.1 (0.1–0.1)
X...X	P=	0.90	0.95	1.00
C...C		0.1 (0.0–0.1)	0.6 (0.6–0.6)	1.1 (1.1–1.1)
O...O		0.1 (0.1–0.1)	0.3 (0.2–0.3)	0.3 (0.3–0.3)
S...S	// <sup>c</sup>		1.1 (1.1–1.1)	1.1 (1.0–1.1)
Cl...Cl	// <sup>c</sup>		1.1 (1.1–1.2)	0.9 (0.8–1.0)
F...F	// <sup>c</sup>		1.0 (1.0–1.1)	0.9 (0.8–1.0)

<sup>a</sup> 95 % confidence interval in parentheses.

<sup>b</sup> Sum of standard atomic radii, according to Rowland & Taylor.<sup>27</sup>

<sup>c</sup> Too few data for meaningful stats.

At the full sum of SAR (P = 1.00), all the close H...X contacts have R<sub>F</sub> >> 1, meaning that in POLY-*all* they occur roughly from 1.5 to 3.5 times more frequently than it would be expected if packing were totally random. These results are in good qualitative agreement with the Taylor's ones,<sup>22</sup> taking into account that the present study refers to a totally different data pool. H...O and H...N contacts are again the most frequent ones, and their trend is different with respect to all the other interactions. Actually, while H...S and H...halogen frequencies either undergo a net reduction, or remain quite unchanged at P = 0.95, the H...O and H...N ones experience a significant increase. At P = 0.90, H...O contacts are even 5 times more frequent than expected on the basis of surface area consideration. This is due to the fact that the number of atoms in close contact, N<sub>A</sub>, is smaller at very short threshold distances, resulting in a lower amount of expected interactions for each *i*-th structure, E<sub>i</sub> (equation (2)), and thus in a smaller denominator in equation (3). If the number of observed contacts, O<sub>i</sub> (equation (3)), remains comparatively high, as in the case of oxygen acceptors (see also Table 2), R<sub>F</sub> is meant to increase as well.

In agreement with Taylor,<sup>22</sup> interactions between non-H atoms are much less significant. Halogen...halogen and S...S contacts invariably have R<sub>F</sub> very close to 1, meaning that their average observed frequency is not much different to that expected for random interactions. However, R<sub>F</sub> estimates provide just an average picture, and values close to unity do not preclude the occurrence of halogen bonded contacts<sup>40</sup> in some crystal forms. In this respect, favorable Cl...Cl geometries are present in

roughly 1/3 of the structures bearing chlorine atoms, but they are definitely less frequent than the H...X ones. It should be also noted that, at  $P=1.00$ , ~ 80 % of the hydrogen atoms present in the databank are “naked”, *i.e.* they are not involved in intermolecular HB contacts. The relative proportion rapidly increases for stricter cut-offs and becomes of ~ 96 % for  $P = 0.90$ . This, however, also partly reflects the fact that the ratio between potential donors and acceptors is biased toward an excess of the former (Table 1).

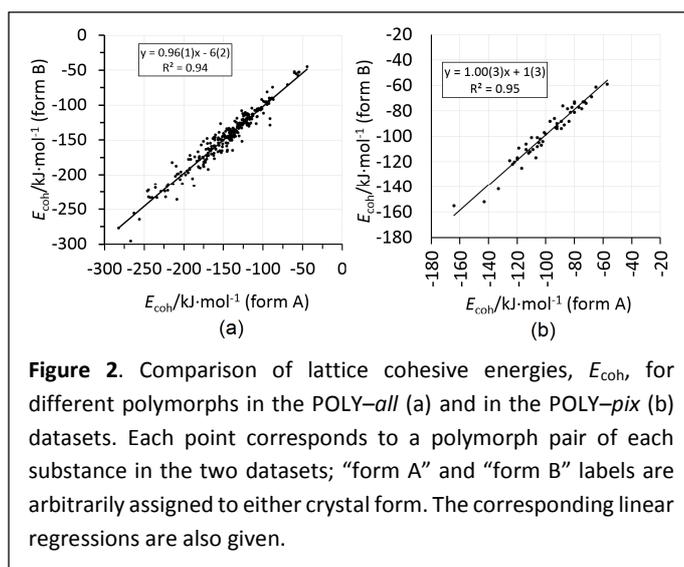
C...C interactions have recurring frequencies also close to 1 at the  $P = 1.00$  level, but they become increasingly rarer at lower distance thresholds. On average, no particularly short C...C contacts are present, possibly indicating that stacking interactions are not significant in our data pool. Indeed, no flatness constraints were imposed on the molecular shape during the CSD search. Various examples exist of structures that, albeit aromatic, are far from being planar (*e.g.* BOLMIZ and WOVGIZ), and thus less prone to produce stacking motifs<sup>41</sup>.

More interesting is the scarceness of O...O and H...H contacts, even within the usual van der Waals cut-off (Table 4).  $R_F \ll 1$  indicates that some contacts exist, which are much less frequent than expected on the basis of surface area considerations, implying deviations from the null hypothesis (random packing) that are as much significant as those in the opposite direction ( $R_F \gg 1$ ). As already noticed by Taylor,<sup>22</sup> this is unsurprising, as negatively charged O atoms tend to avoid each other, and at the same time (meta)stable crystal lattices minimize the amount of possible H...H steric clashes. Indeed, packing is *not* random, and close contacts both more and less frequent than expected under the null hypothesis are set up, depending on the interplay of molecular shape and crystal field. In a stable arrangement of close packed molecules, no net intermolecular forces are acting on nuclei, but those that oppose the nuclear kinetic energy due to thermal effects and restore the atomic and molecular equilibrium positions.<sup>42</sup> The null resultant, however, is due to a rather subtle balance among competing intermolecular forces, in turn determined by the potential crystal field.<sup>1,3,5,6</sup> In this respect, short-range repulsions can be structure-determining factors as well as attractive dispersions or hydrogen bonds. The  $R_F$  metrics clearly highlights these two opposite tendencies, namely toward a frequency increase ( $R_F \gg 1$ ) of specific intermolecular contacts, and a concurrent decrease of other ones ( $R_F \ll 1$ ). Thus, the real problem is to clarify the role of an ensemble of weakly attractive HB in determining the overall thermodynamic stability of the crystal lattice, with respect to competing repulsions, as well as other attractive interactions. Likely, a universal answer does not exist, due to the huge chemical diversity of organic compounds. Nevertheless, a close look at the interactions that are conserved among different polymorphs, as well as at their energetics, should provide at least some clues on the general trends. This topic is explored in the next Sections.

### 3.4 Packing energies

Lattice cohesive energies,  $E_{\text{coh}}$ , of different polymorphs are strongly correlated (Figure 2). This is true both at the AA-CLP and PIXEL computational levels<sup>†</sup> (Section 2.3), as in the large

majority of cases the differences between the lattice energies of chemically related crystal forms lie within 5–10 % of their average  $E_{\text{coh}}$ .

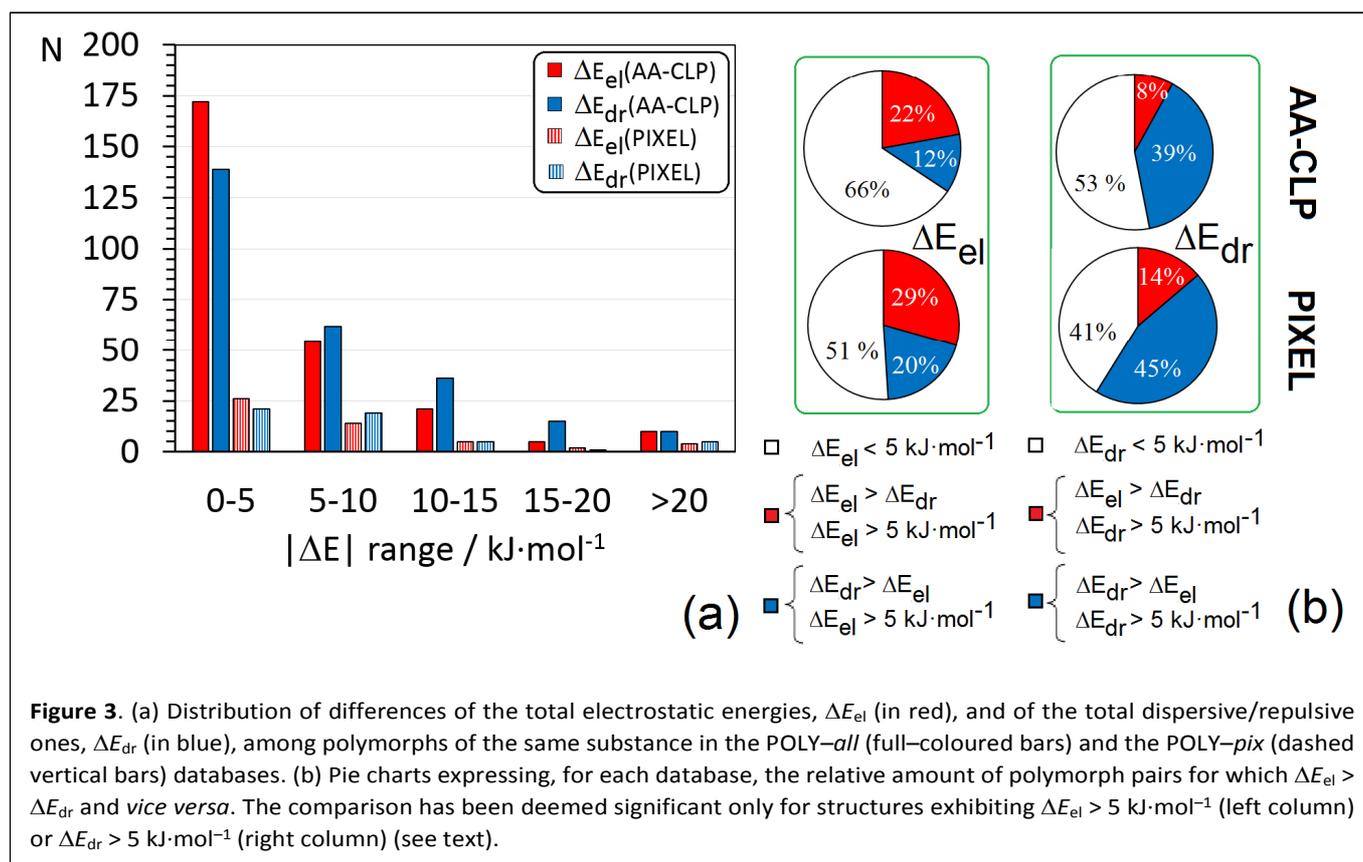


**Figure 2.** Comparison of lattice cohesive energies,  $E_{\text{coh}}$ , for different polymorphs in the POLY-*all* (a) and in the POLY-*pix* (b) datasets. Each point corresponds to a polymorph pair of each substance in the two datasets; “form A” and “form B” labels are arbitrarily assigned to either crystal form. The corresponding linear regressions are also given.

As expected, cases where a crystal form is either much more or much less stable than its chemically equivalent counterparts are rare. This holds true independently on the computational method, and thus on the absolute magnitude of the predicted cohesive energies (Figure 2).

In both the AA-CLP and PIXEL frameworks, the total lattice energy can be decomposed into Coulomb ( $E_c$ ), polarization ( $E_p$ ), dispersion ( $E_d$ ) and repulsion ( $E_r$ ) terms (see also Tables S4–S5 ESI). As for the attractive contributions, dispersive energies  $E_d$  are invariably more negative than the sum of the electrostatic ones: the  $E_d/(E_c+E_p)$  ratio is always well higher than 1.0, with average values of 3.4(1) and 1.9(1), respectively, for the POLY-*all* and the POLY-*pix* datasets.

Figure 3 summarizes the differences detected in the total electrostatic part of the lattice energy ( $E_{\text{el}} = E_c + E_p$ ), as well as those affecting the dispersive–repulsive balance ( $E_{\text{dr}} = E_d + E_r$ ), when different polymorphs are compared to each other. For most substances, the absolute differences in  $E_{\text{el}}$  and  $E_{\text{dr}}$  are very small ( $< 5 \text{ kJ}\cdot\text{mol}^{-1}$ ) and comparable in magnitude with the intrinsic uncertainty threshold of the PIXEL method.<sup>4,5</sup> This confirms that, in most cases, the preference for a certain crystal structure emerges due to a non-obvious balance among different classes of NCI, none of which neatly prevails in determining either the total energy, or its evolution across diverse polymorphic forms. At the same time, a non-negligible minority of structures bear  $\Delta E_{\text{el}} > 0$  and/or  $\Delta E_{\text{dr}} > 0$  (Figure 3a), indicating that a change in the crystal packing can imply a substantial rearrangement of electrostatic and/or dispersive/repulsive NCI. Considering a significance threshold of  $5 \text{ kJ}\cdot\text{mol}^{-1}$ , in both databases the number of structures in which  $\Delta E_{\text{dr}}$  prevails over  $\Delta E_{\text{el}}$  is comparable, if not even larger, with respect to that in which  $\Delta E_{\text{el}} > \Delta E_{\text{dr}}$  (Figure 3b). This stresses the importance of the interplay of dispersive and steric factors, even though electrostatic-driven interactions, including weak HB, are not negligible on absolute grounds.



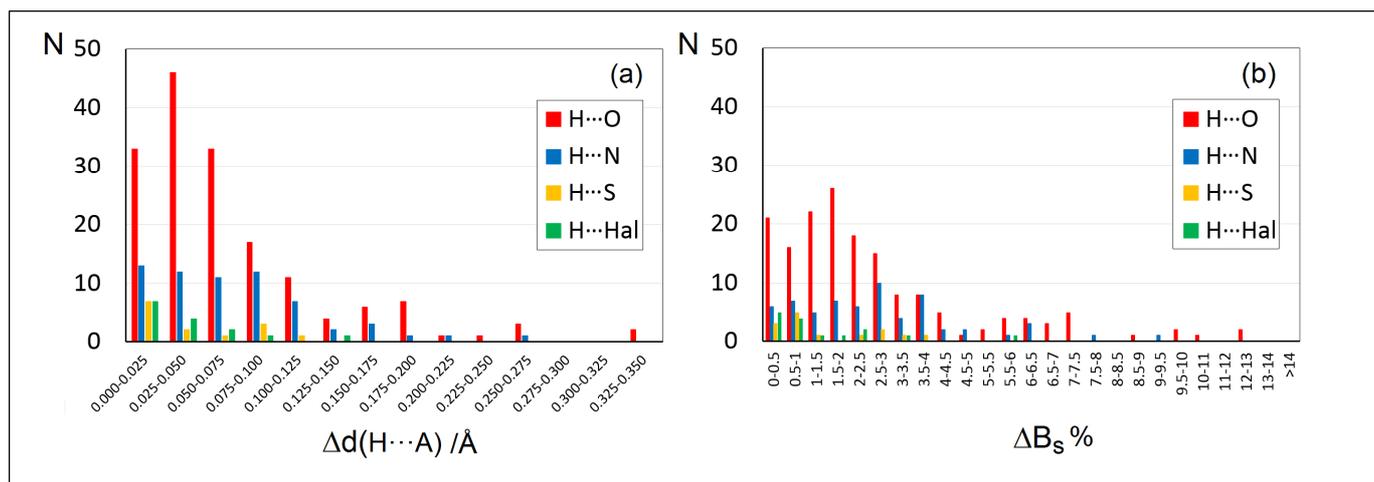
Hydrogen bonds influence Coulomb and polarization terms; in fact, a weak correlation exists among their shortness parameter,  $B_s$ , and  $E_{el}$  (see Figure S1 ESI). The present analysis demonstrates that the dispersive/repulsive potential,  $E_{dr}$ , is often (but not exclusively) more sensitive to a change in the crystal structure than the electrostatic one. This implies that interactions that do not require specific atom–atom contacts, such as dispersions and low–order electrostatics, which are more long–range (“non–local”) in nature, can effectively compete with repulsions, high–order electrostatics and HB (“local” ones), all active at shorter range.

### 3.5 Similarities among HB contacts

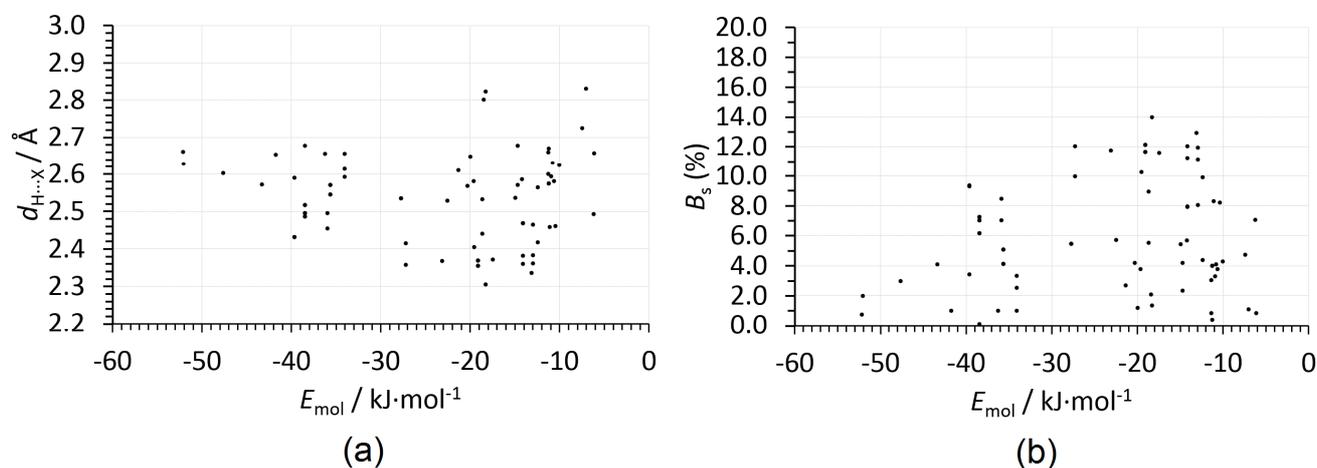
Figure 4 compares different groups of polymorphs, from the viewpoint of the geometrical contact distance of  $\text{H}\cdots\text{X}$  atom–atom interactions. For each class of  $\text{H}\cdots\text{X}$  contacts and each structure in POLY–all, we computed the average values of the  $d_{\text{H}\cdots\text{X}}$  distances and the bond shortness estimator,  $B_s$ , within the full sum of SAR ( $P = 1.0$ ). Then, we obtained the absolute differences of  $\langle d_{\text{H}\cdots\text{X}} \rangle$  and  $\langle B_s \rangle$  among the polymorphs of a substance,  $\Delta d$  and  $\Delta B_s$ . The histograms represent the number distributions of  $\Delta d$  and  $\Delta B_s$  for different classes of X acceptors ( $X = \text{O}, \text{N}, \text{S}, \text{halogen}$ ); the corresponding tabular entries, as well as graphs where all the possible HB acceptors are summed together, are available in the ESI (Tables S6–S7 and Figure S2). In almost 50 % of cases, different polymorphs have similar average  $\langle d_{\text{H}\cdots\text{O}} \rangle$  distances (Figure 3a), with differences not exceeding 0.05 Å. The cumulative probability rapidly increases

to 78.6 % when the limit of 0.1 Å is considered.  $\text{H}\cdots\text{N}$  contacts show a very similar behaviour, with a cumulative probability of 76.2 % of having  $\Delta d \leq 0.1$  Å. Being significantly less represented, contacts with sulphur and halogen atoms do not allow to draw safe conclusions. However, also for  $X = \text{S}, \text{F}$  and  $\text{Cl}$  acceptors  $\Delta d > 0.1$  Å is embodied just by a minority of cases.  $\Delta B_s$  estimates provide the same information as  $\Delta d$  (see Figure S2 ESI), but from the viewpoint of the  $\text{H}\cdots\text{X}$  closeness relative to the sum of reference SAR (Figure 3b). It turns out that  $\sim 50$  % of  $\text{H}\cdots\text{O}$  and  $\text{H}\cdots\text{N}$  contacts have identical  $B_s$  within, respectively, 2.0 and 2.5 absolute percent points.

In conclusion, considering just contact distances, a certain propensity toward the conservation of chemically similar weak HB interactions exists among polymorphs of the same substance. However, such a tendency is not quantitative. A change of 2.0 or 2.5 absolute percent points in  $B_s$  means that, for example,  $\text{H}\cdots\text{O}$  and  $\text{H}\cdots\text{N}$  contact distances can differ, on average, by up to 0.055 Å, which might be far from being irrelevant for their interaction energetics (see *infra*). Moreover, several examples also exist, where  $\Delta d$  and  $\Delta B_s$  are largely significant, with a minority of polymorphs differing from each other by up to 0.3 Å in  $\langle d_{\text{H}\cdots\text{X}} \rangle$ , corresponding to  $\sim 12$ –13 % percent points in  $\langle B_s \rangle$ . Such large changes likely imply that both crystal forms bear some kind of  $\text{CH}\cdots\text{O}$  or  $\text{CH}\cdots\text{N}$  interactions, but either these are set up among different pairs of donor/acceptor groups, or the identities of the donor and acceptor groups are the same, but the crystal packing is markedly diverse.



**Figure 4.** Distribution of differences between average contact distances (a),  $\Delta d$ , and shortness index (b),  $\Delta B_s$ , for  $H\cdots X$  interactions among the polymorphs of each substance within the POLY-*all* databank. Red:  $X = O$ ; Blue:  $X = N$ ; Yellow:  $X = S$ ; Green:  $X = \text{halogen}$ .



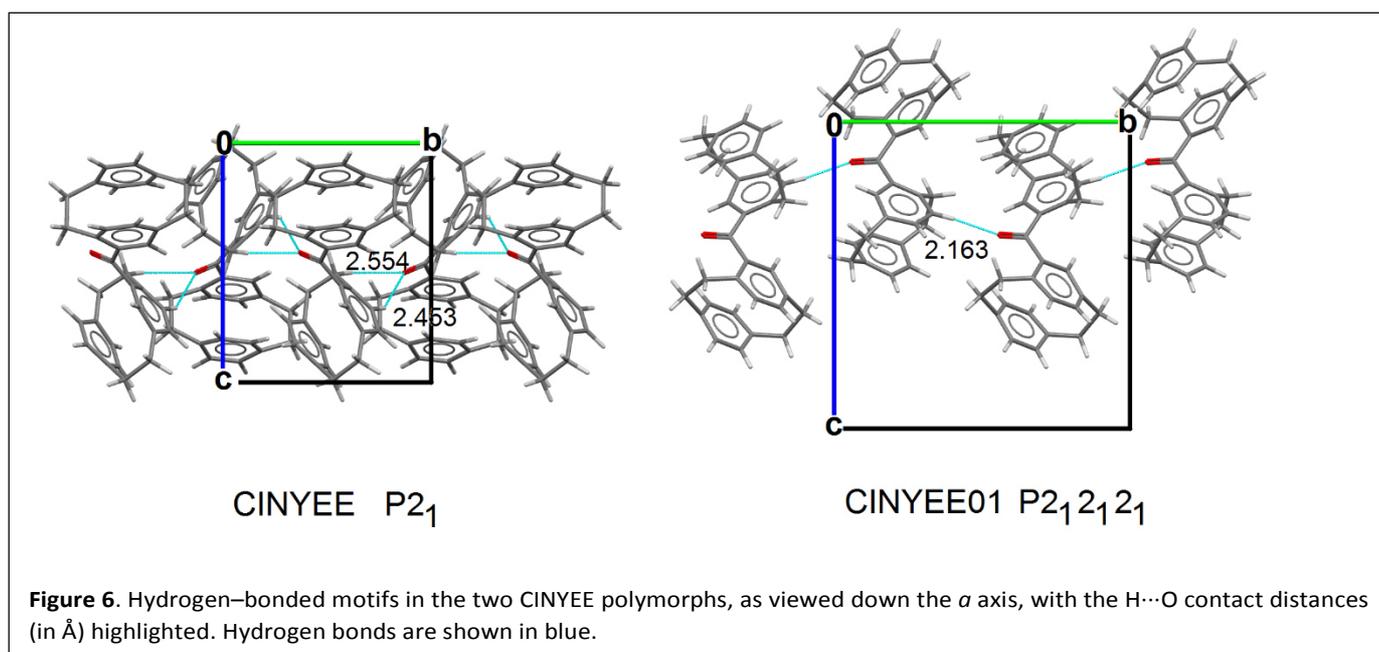
**Figure 5.** In-crystal PIXEL interaction energies,  $E_{\text{mol}}$  ( $\text{kJ}\cdot\text{mol}^{-1}$ ) of molecular pairs bonded by weak HB that are conserved between polymorphs of the same substance vs. (a)  $\text{CH}\cdots\text{X}$  ( $X = \text{N, O, S, Cl}$ ) distances and (b) shortness parameter,  $B_s$ , of individual HB.

### 3.6 Conserved HB contacts

Usually, the definition of conserved close contacts is made on purely geometrical grounds, based on *a priori* selected more or less tight cut-offs. To label a HB as “conserved”, we further require that the pair energy,  $E_{\text{mol}}$ , be comparable between the crystal forms, as it is reasonable to suppose that consistent structural building blocks preserve their interaction energetics as well as their geometry. We thus require that a HB contact (i) is set up between the same chemical type of donor and acceptor groups (not necessarily just the same atoms), and (ii) that  $E_{\text{mol}}$ ,  $H\cdots X$  distance ( $d_{\text{HX}}$ ) and  $C-H\cdots X$  angle ( $\alpha_{\text{DHX}}$ ) are equal within tolerances of 10%, 10% and 6%. These thresholds roughly correspond to the tightest ones employed by Mercury<sup>39</sup> to search for structure similarity among different compounds. 75 conserved  $H\cdots X$  contacts were found in the POLY-*pix* databank and 312 in the POLY-*all* one, corresponding to  $\sim 17$ – $19$  % of the total number of HB contacts within the full sum of SAR.

However, no obvious correlations exist among the  $H\cdots\text{Acceptor}$  distance of conserved contacts and the corresponding pair interaction energy,  $E_{\text{mol}}$  (Figure 5). Moreover, very short HB are not generally associated to strongly bonded molecular pairs. Weak HB are flexible enough to adapt their geometry to the interaction requirements dictated by other interactions, but also mean that they can be hardly considered as the main actors underlying a given packing mode. Analogue considerations hold true even though the above defined thresholds are relaxed to include more weak HB into the pool of conserved interactions (Figure S3 ESI).

On the other hand, if specific patterns of weak HBs would be associated to well-defined packing modes, it might be expected that they should be different in different polymorphs. Thus, it could be not surprising that weak HB are poorly conserved, especially when the interaction energies  $E_{\text{mol}}$  come into play. This should be evident is one focus on those structures



**Table 5.** Interaction geometries and PIXEL molecular interaction energies<sup>a</sup> for close neighbours of crystal forms shown in Figures 5 and 6. Values in Å, deg and kJ·mol<sup>-1</sup>.

Crystal	$d_{\text{H}\cdots\text{X}}$	$d_{\text{C}\cdots\text{X}}$	$\alpha_{\text{CH}\cdots\text{X}}$	$d_{\text{CM}}^b$	Symmetry <sup>c</sup>	$E_c$	$E_p$	$E_d$	$E_r$	$E_{\text{mol}}$
CINYEE	2.554 <sup>d</sup>	3.515	147.8	7.4	1-x, 1/2+y, 1-z	-12.9	-6.4	-43.6	34.3	-28.6
	2.453 <sup>e</sup>	3.355	140.2	7.4						
CINYEE01	2.163 <sup>e</sup>	3.236	171.3	7.6	1-x, 1/2+y, 1/2-z	-17.3	-7.5	-44.3	39.1	-30.0
DIWKOK	2.380 <sup>e</sup>	3.296	141.7	8.2	1+x, -1+y, z	-11.7	-5.1	-10.0	12.8	-14.1
	2.468 <sup>e</sup>	3.516	163.3	8.2						
DIWKOK01	2.358 <sup>e</sup>	3.271	141.3	8.2	-1+x, 1+y, z	-12.2	-5.4	-10.2	13.6	-14.1
	2.467 <sup>e</sup>	3.518	163.9	8.2						

<sup>a</sup> The total molecule-molecule interaction energy,  $E_{\text{mol}}$ , is decomposed into Coulomb ( $E_c$ ), polarization ( $E_p$ ), dispersion ( $E_d$ ) and repulsion ( $E_r$ ) contributions.

<sup>b</sup> Centre of mass distance, in Å.

<sup>c</sup> Symmetry operation generating the HB acceptor.

<sup>d</sup> Donor: aliphatic CH<sub>2</sub>. Acceptor: C=O.

<sup>e</sup> Donor: aromatic C-H. Acceptor: C=O

exhibiting high differences in their  $\langle d_{\text{H}\cdots\text{X}} \rangle$  and  $\langle B_c \rangle$  parameters (Figure 4, Section 3.5).

CINYEE and CINYEE01 provide an example of polymorphs with one CH⋯O interaction that involves exactly the same pair of donor/acceptor groups (Figure 6), but can be hardly classified “conserved”, as the C-H⋯O angle varies by more than 30°, and the H⋯O distance by ~0.3 Å (Table 5). In both the crystal forms, infinite CH⋯O ribbons are set up along the *b* axis. In the monoclinic crystal, relatively long contacts are formed among the C=O acceptor and a pair of aliphatic and aromatic donors (Figure 6). The orthorhombic form prefers to set up a single, much shorter CH⋯O interaction that enhances the proximity of the phenyl hydrogen atom with oxygen (Figure 6, Table 5). A neat correlation is apparent (Table 5) between the Coulombic contributions,  $E_c$ , and the strengthening of the Ph-H⋯O bond.<sup>1,21</sup> At the same time, the enhanced electrostatic stability is compensated by an increase in the repulsive term  $E_r$ , thus the slight  $|E_{\text{mol}}|$  increase in the orthorhombic form is due to the

marginally more attractive dispersion and polarization terms. When the overall lattice cohesion is considered, though, the CINYEE crystal turns out to be more stable than the CINYEE01 one by 4-5 kJ·mol<sup>-1</sup> (Table 6 and Table S5 ESI).

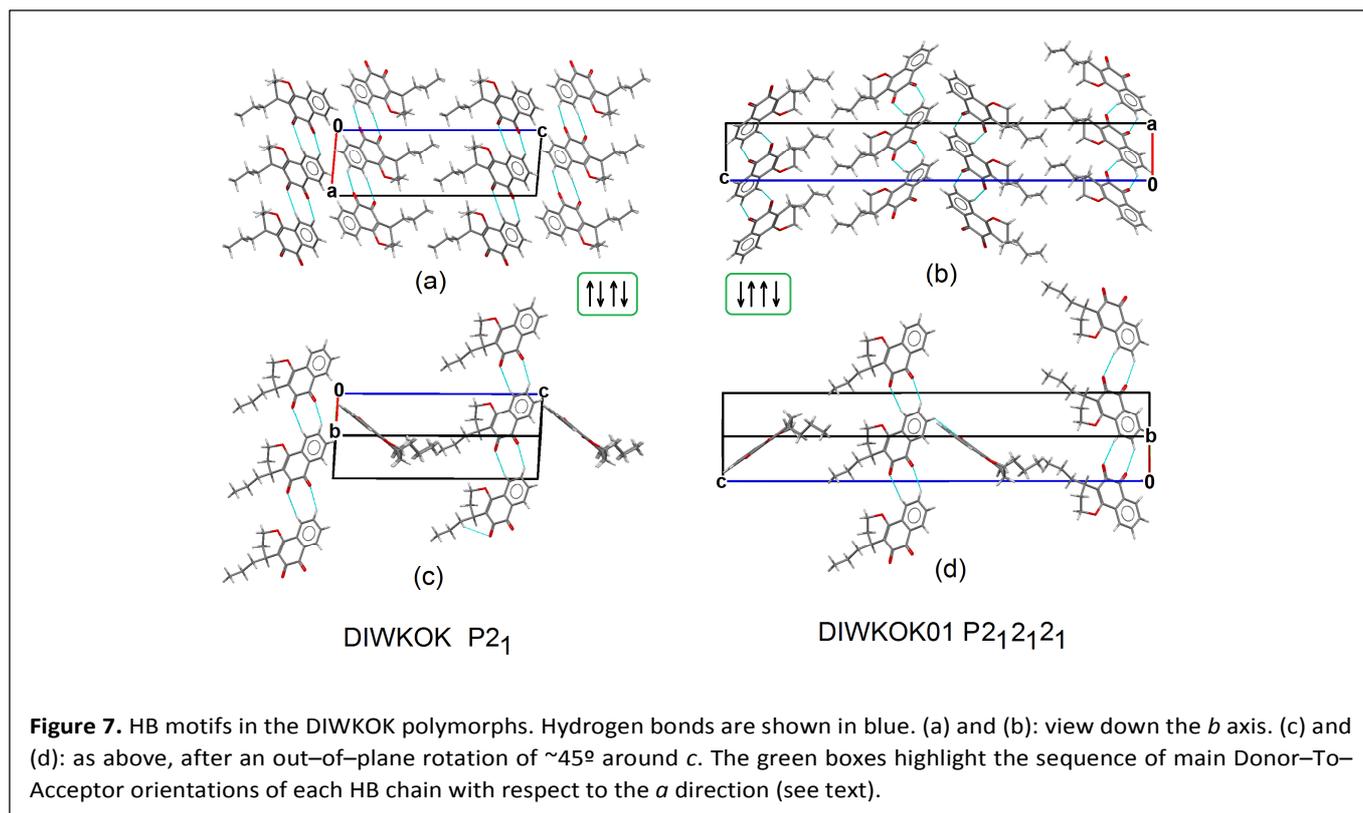
**Table 6.** PIXEL cohesive energies,  $E_{\text{coh}}$ , of CINYEE and DIWKOK polymorphs.<sup>a,b</sup>

Crystal	$E_c$	$E_p$	$E_d$	$E_r$	$E_{\text{dip}}^c$	$E_{\text{coh}}$
CINYEE	-41.1	-19.4	-164.8	120.1	-0.9	-106.0
CINYEE01	-42.7	-21.1	-157.1	119.2	//	-101.7
DIWKOK	-36.4	-24.8	-111.3	94.3	-2.6	-80.8
DIWKOK01	-35.4	-23.5	-112.1	97.2	//	-73.8

<sup>a</sup> The decomposition of  $E_{\text{coh}}$  into a sum of Coulombic ( $E_c$ ), polarization ( $E_p$ ), dispersion ( $E_d$ ) and repulsion ( $E_r$ ) terms is also given.

<sup>b</sup> Values in kJ·mol<sup>-1</sup>.

<sup>c</sup> Correction to the coulomb sums in polar space groups, performed according to Kroon and van Eijck.<sup>43</sup>



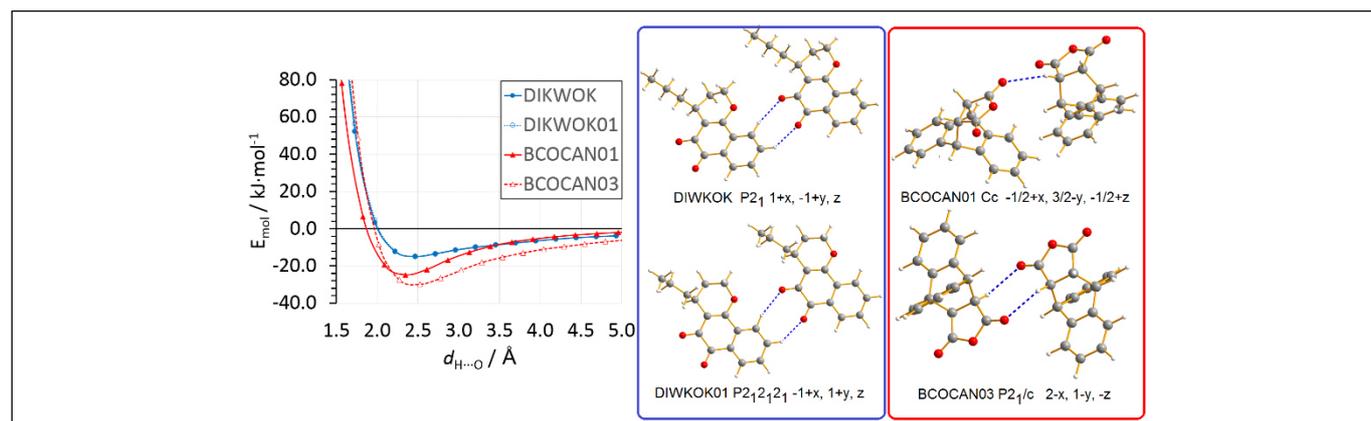
This is likely the consequence of significant conformational differences between the two asymmetric units, which involve the mutual orientation of the bulky paracyclophane substituents with respect to the pivot carbonyl and produce a more favourable dispersive/repulsive balance in the monoclinic form. Conformational changes, as well as cohesive energies, depend on the whole interacting charge densities. Shape/steric factors are likely at stake in determining the crystal packing differences in CINYEE and CINYEE01, even though the Ph–H...carbonyl HB is clearly related to a preferential molecule–molecule interaction mode in both crystals. Other examples where similar considerations hold true are reported in the ESI (Figure S4, Table S8). In conclusion, even when different crystal lattices are associated to different weak HB patterns, the latter are so flexible and adaptable that such differences are likely the consequence, and not the cause, of more striking packing requirements.

Conversely, if packing motifs of different polymorphs are similar, it is easy to foresee that weak HB will be more or less preserved, resulting in much lower  $\Delta d$  and  $\Delta B_s$  estimates (Section 3.5). This might produce highly conserved contacts, from both the geometrical and energetic perspectives. Figure 7 displays the crystal packing of the two crystalline forms of DIWKOK, which are characterized by a highly conserved double CH...O pattern (Table 5). The two polymorphs have almost identical packing features, namely, molecules are arranged in infinite planar hydrogen-bonded 1–D ribbons along the  $(a,b)$  diagonals, while their side aliphatic chains extend in the free space along the  $c$  axis. As a consequence, neighbouring ribbons run roughly along perpendicular directions (Figure 7c,d). What

changes is the sequence, along  $c$ , of the relative orientations of the donor and acceptor groups. Taking as a reference the main Donor–To–Acceptor orientation of each ribbon along the  $a$  direction in the  $(a,c)$  plane (Figure 7a,b and green boxes therein), in the monoclinic form an antiparallel sequence is set up. In the orthorhombic crystal, any second pair of antiparallel ribbons is reversed with respect to the first one, resulting in alternating arrangements of parallel and antiparallel 1–D H-bonded chains. As expected, the interaction energy  $E_{\text{mol}}$  for symmetry-related DIWKOK pairs in the ribbon is essentially identical in the two polymorphs (Table 5). Nevertheless, the difference in the lattice cohesive energies is not negligible, with the monoclinic form being more stable than the orthorhombic one by  $7.0 \text{ kJ}\cdot\text{mol}^{-1}$  (Table 6 and Table S5 ESI). The reasons are clearly not related with the stability of hydrogen-bonded molecular pairs; rather, they should be looked for in the general interactions among different molecular chains. The space group of DIWKOK is polar, thus it bears an extra attractive term due to the nonvanishing unit cell dipole ( $\sim 2.5 \text{ kJ}\cdot\text{mol}^{-1}$ ); at the same time,  $E_r$  is also lower by  $\sim 3 \text{ kJ}\cdot\text{mol}^{-1}$  in the monoclinic cell (Table 6). Taken together, these effects account for roughly the 80 % of the extra lattice stabilization; what remains is due to slightly more attractive electrostatic contributions.

### 3.7 HB strength / pair energy correlation

Sometimes, hydrogen bonds very similar in terms of interaction energy and geometry are formed between molecular pairs with different interaction geometries. Figure 8 compares two CH<sub>2</sub>...O=C interactions in BCOCAN01/BCOCAN03 with the already discussed highly conserved double Ph–H...O=C ones in



**Figure 8.** On the left: evolution of quantum mechanical  $E_{\text{mol}}$  ( $\text{kJ}\cdot\text{mol}^{-1}$ ) as a function of the  $\text{H}\cdots\text{O}$  distance of weak HBs shown in the insets on the right for BCOCAN (triangles) and DIWKOK (dots). Refer to Tables 5 and 7 for geometrical and energetic parameters. The lines serve just as a guide to the eye.

**Table 7.** Interaction geometries and PIXEL molecular interaction energies<sup>a</sup> for close neighbours of BCOCAN crystal forms shown in Figures 8. Values in Å, deg and  $\text{kJ}\cdot\text{mol}^{-1}$ .

Crystal	$d_{\text{H}\cdots\text{X}}$	$d_{\text{C}\cdots\text{X}}$	$\alpha_{\text{CH}\cdots\text{X}}$	$d_{\text{CM}}^b$	Symmetry <sup>c</sup>	$E_c$	$E_p$	$E_d$	$E_r$	$E_{\text{mol}}$
BCOCAN01	2.608 <sup>d</sup>	3.374	127.3	6.9	$-1/2+x, 3/2-y, -1/2+z$	-9.5	-4.7	-20.6	13.5	-21.3
BCOCAN03	2.527 <sup>d</sup>	3.239	122.5	7.0	$2-x, 1-y, -z$	-19.9	-6.4	-16.3	20.1	-22.5

<sup>a</sup> The total molecule–molecule interaction energy,  $E_{\text{mol}}$ , is decomposed into Coulomb ( $E_c$ ), polarization ( $E_p$ ), dispersion ( $E_d$ ) and repulsion ( $E_r$ ) contributions.

<sup>b</sup> Centre of mass distance, in Å.

<sup>c</sup> Symmetry operation generating the HB acceptor.

<sup>d</sup> Donor: aliphatic  $\text{CH}_2$ . Acceptor:  $\text{C}=\text{O}$ .

DIWKOK/DIWKOK01 (Section 3.6). The energy trends were computed at the M06/6–311G(p,d) theory level and corrected for basis–set superposition error with the method of Boys and Bernardi.<sup>44</sup> BCOCAN03 is centrosymmetric and forms cyclic HB patterns between inversion–related molecules, while BCOCAN01 exploits a glide symmetry to form chain motifs along the [101] direction. Interestingly, the geometrical parameters of the HB contact are identical within 3–4 %, and the overall  $E_{\text{mol}}$  estimates are very similar as well (Table 7), despite the patent diversities in the molecular interaction modes (red inset in Figure 8). As expected, the Coulomb contribution is more favourable for the cyclic pattern in BCOCAN03, but a significantly more repulsive  $E_r$  frustrates the energy gain.

To clarify the relative stability of the various molecular pairs as a function of the interaction distance, we computed the evolution of  $E_{\text{mol}}$  as a function of the linear  $\text{H}\cdots\text{O}$  separation in the molecular pairs shown in Figure 8. Obviously,  $E_{\text{mol}}$  includes contributions from the whole charge density distributions of the interacting molecular pairs, and does not depend only on the weak HBs. In particular, BCOCAN is more globular than DIWKOK, and more atoms are thus closer to each other while the two molecules approach along the  $\text{H}\cdots\text{O}$  directrix. Indeed, van der Waals interactions are more effective in BCOCAN01 (see the  $E_d$  value in Table 7), where the approaching mode implies a larger sharing of the molecular surface (Figure 8). The contribution of HBs reveals in the more attractive in–crystal electrostatic contributions of BCOCAN03 (Table 7), even though the  $\text{H}\cdots\text{O}$  equilibrium distance is slightly longer in the isolated pair (Figure

8), likely due to the occurrence of stronger repulsions. As expected, the two molecular pairs in DIWKOK polymorphs follow a trend identical to each other, as their basic supramolecular repeating units are identical as well (Section 3.6).

Therefore, even the  $\text{H}\cdots\text{O}$  contacts that imply conserved interaction energetics are not always in 1:1 correspondence with specific molecular recognition modes. This poses into question the prospect of consistently use them as structural building blocks to achieve the desired supramolecular patterns. The problem is even more striking when the whole crystal structure is considered. BCOCAN01, with a patent less attractive  $\text{H}\cdots\text{O}$  contact, has  $E_{\text{coh}} = -94.5 \text{ kJ}\cdot\text{mol}^{-1}$ , which is  $\sim 5 \text{ kJ}\cdot\text{mol}^{-1}$  more negative than in BCOCAN03 (Table S5 ESI). The thermodynamic stability of the lattice depends on non–obvious interactions among all the forces acting in the unit cell. In the present case, dispersions are neatly prevailing in BCOCAN01, and electrostatics in BCOCAN03. Weak HB certainly play a role as well, but the crystal structures tend to take advantage to the intrinsic flexibility of such interactions to satisfy more strict crystal field requirements.

## Conclusions

In this work, a statistical study has been carried out on conserved hydrogen bonds in  $\sim 250$  pairs of polymorphic organic compounds not bearing  $-\text{OH}$  and  $-\text{NH}$  donor groups. The purpose was to understand the role of weak HB in

determining the preference for a specific crystal form over the other possible ones.

In most cases, different polymorphs have very similar cohesive energies, meaning that preference toward a given packing arises due to subtle and non-obvious interplay of competing potentials. The dispersive/repulsive contributions are often, but not exclusively, more sensitive to a change in the crystal structure than the electrostatic ones, which weakly correlate with the HB strength. Accordingly, weak HB are seldom conserved among different polymorphs, even though the majority of crystal structure tend to set up H...X interactions with similar contact distances in their first coordination shell. In any case, no obvious correlations exist among the hydrogen-acceptor distance of conserved contacts and the molecule-molecule interaction energy. One could object that, if there were a 1:1 correspondence between specific HB patterns and packing modes, weak HBs would not be conserved in different polymorphs. We discussed the case of CINYEE as representative of structures where comparable short CH...O HBs are set up, which are not strictly conserved in different crystal forms. We noted that cohesive energies are not directly related to the HB modes; rather, conformational differences might come into play, resulting in a higher stabilization of the polymorph with weaker hydrogen bonds. Even when conserved, very short HBs are not generally related to strongly bonded molecular pairs, and might appear both in similar (DIWKOK) and different (BCOCAN) packing modes. There is generally not a 1:1 correspondence among weak HBs and specific molecular recognition patterns. This poses into question the possibility of consistently use weak HBs as structural building blocks to achieve the desired supramolecular patterns.

The  $R_F$  metrics was introduced by Taylor<sup>15,22</sup> to determine whether a given A...B atom-atom contact occurs more or less frequently than expected under the null hypothesis of random packing. When applied to the current databank, it clearly highlights two opposite tendencies. On the one hand, H...Acceptor contacts always occur with higher frequencies than in the null hypothesis. At the same time, other atom-atom contacts, such as the H...H and O...O ones, have  $R_F \ll 1$ , meaning that they occur less frequently than if the packing were random. Recently, Taylor<sup>22</sup> rightly stated that *"any explanation of the crystal packing of the structures discussed herein must account for the fact that they contain many more X...H interactions than would be expected by chance"*. Crystal structures, indeed, are not based on random packing. Molecules are arranged such that they maximize the number of HB contacts, while minimizing the number of possible steric clashes; at the same time, however, competing steric and shape effects often prevail over local atom-atom contacts in determining the overall thermodynamic stability.

In our opinion, these apparently conflicting results can be reconciled by taking into account the natural tendency of interacting molecules to maximize the attractive contacts and minimize the repulsive ones. These effects are important already in the first stages of the nucleation event, when crystal embryos start forming from either a supersaturated solution or the melt. CH...X interactions always provide stabilizing

contributions to the electronic energy of the system; indeed, both recently and in the past, charge density studies demonstrated that weak HBs can even compete with stronger OH...O ones, if allowed by the crystal packing.<sup>3,6,20</sup> Hydrogen bonds, however, are intrinsically "short-range" in nature. They might assist the molecular recognition process, providing extra stabilizing terms that can advantage some interaction modes over other ones and drive nucleation toward either crystal form. In other words, they might significantly influence the crystallization kinetics, especially during the very first elementary acts, when they could still provide non-negligible contribution to the (self-)recognition energetics. Later, when the crystal gains mass and grows, "long-range" interactions do not require specific atom-atom contacts (van der Waals, low-order electrostatic moments) might come into play, governing the overall thermodynamics. Collective interactions among growing patterns of molecules, such as supramolecular clusters, sheets and chains, can confine weak HBs to a less central role. In fact, metadynamics simulations recently demonstrated<sup>45</sup> that crystal nuclei could contain different proto-polymorphic forms, which compete to each other until one prevails and determines the structure of the nascent solid phase. Molecular recognition is just the first step of a series of complex self-assembly events at various scales, some of which might be dominated by collective interactions among structurally different patterns.

In conclusion, there are at least two reasons explaining the observed high frequency of H...X interactions in molecular crystals. On the one hand, the packing is obviously reminiscent of the supramolecular synthons that were present in the early stages of nucleation;<sup>46</sup> on the other hand, the system tends toward close packing to maximize attractive interactions, according to Kitaigorodskii rules.<sup>47</sup> Thus, further stabilizing short H...X contacts might arise, due to conformational and shape constraints. To check the validity of the above sketched conjectures, one might select some crystal structures where H...X interactions could be significant, such as the CINYEE, DIWKOK and BCOCAN test cases discussed above, and try to crystallize derivatives where the involved C-H donors are somewhat "switched off". For example, deuterium substitution in suitable chemical sites could be exploited, to further weaken the corresponding hydrogen bonds according to the Ubbelohde effect,<sup>48</sup> without significantly perturbing the rest of the system. If no changes are detected in the crystallization output, one should conclude that the studied CH...O short contact is poorly significant in determining the lattice equilibrium structure. Analogue crystallization experiments, carried out as a function of the thermodynamic boundary conditions, might help to disentangle kinetics effects from the purely thermodynamics ones.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The Danish National Research Foundation is acknowledged for partial funding of this work through contract grant number DNRF93. This research was also partially funded by the Unimi Development Plan–Line 2, Action B, project NOVAQ, n° PSR2015–1716FDEMA\_08. Prof. A. Gavezzotti is warmly thanked for fruitful discussions. Thanks are also due to Mr. M. Frigerio for explorative calculations.

## Notes and references

‡ A unique outlier, namely the NUJLIN / NUJLINO1 pair, was detected in the POLY–*all* databank and consequently removed from the statistics on the packing energy. The empirical AA–CLP

method predicts an abnormally low  $E_{\text{coh}}$  for NUJLIN, due to a patent overestimation of the coulomb repulsions. However, the problem is no more apparent when the energies are computed through the more accurate PIXEL procedure; this substance was thus included into the POLY–*pix* databank.

- A. Gavezzotti, V. Colombo and L. Lo Presti, *Cryst. Growth Des.* 2016, **16**, 6095–6104.
- L. Lo Presti, M. Sist, L. Loconte, A. Pinto, L. Tamborini and C. Gatti, *Cryst. Growth Des.* 2014, **14**, 5822–5833.
- R. Destro, E. Sartirana, L. Loconte, R. Soave, P. Colombo, C. Destro and L. Lo Presti, *Cryst. Growth Des.* 2013, **13**, 4571–4582.
- A. Gavezzotti, *New J. Chem.*, 2011, **35**, 1360–1368.
- A. Gavezzotti, *Molecular Aggregation. Structural Analysis and Molecular Simulations of Crystals and Liquids*. In *IUCr book series*; Oxford University Press: Oxford, 2007. ISBN-13: 9780198570806
- L. Lo Presti, R. Soave and R. Destro, *J. Phys. Chem. B* 2006, **110**, 6405–6414.
- D. A. Leigh, C. C. Robertson, A. M. Z. Slawin and P. I. T. Thomson, *J. Am. Chem. Soc.* 2013, **135**, 9939–9943.
- J. Bernstein, *Cryst. Growth Des.* 2013, **13**, 961–964.
- D. J. Sutor, *Nature*, 1962, **195**, 68–69.
- J. Donohue, in *Structural Chemistry and Molecular Biology*, ed. A. Rich and N. Davidson, W. H. Freeman, San Francisco, 1968, pp. 459–463.
- F. A. Cotton, L. M. Daniels, G. T. Jordan IV and C. A. Murillo, *Chem. Commun.*, 1997, 1673–1674.
- T. Steiner and G. R. Desiraju, *Chem. Commun.*, 1998, 891–892.
- J. D. Dunitz and A. Gavezzotti, *Angew Chem Int Ed Engl.* 2005, **44**, 1766–1787.
- J. D. Dunitz, *IUCrJ* 2015, **2**, 157–158.
- R. Taylor, *CrystEngComm*, 2014, **16**, 6852–6865.
- E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem.*, 2011, **83**, 1637–1641.
- R. Taylor and O. Kennard, *Acc. Chem. Res.* 1984, **17**, 320–326.
- E. May, R. Destro and C. Gatti, *J. Am. Chem. Soc.*, 2001, **123**, 49, 12248–12254.
- E. Espinosa, M. Souhassou, H. Lachekar and C. Lecomte, *Acta Crystallogr., Sect. B: Struct. Sci.* 1999, **55**, 563–572.
- G. Macetti, L. Loconte, S. Rizzato, C. Gatti and L. Lo Presti, *Cryst. Growth Des.* 2016, **16**, 6043–6054
- A. Gavezzotti and L. Lo Presti, *Cryst. Growth Des.* 2016, **16**, 2952–2962
- R. Taylor, *Cryst. Growth Des.* 2016, **16**, 4165–4168.
- C. R. Groom and F. H. Allen, *Angew. Chem. Int. Ed.* 2014, **53**, 662–671.
- I. J. Bruno, J. C. Cole, P. R. Eddington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 389–397.
- F. H. Allen and I. J. Bruno, *Acta Crystallogr., Sect. B: Struct. Sci.* 2010, **66**, 380–386.
- A. Gavezzotti, CLP User's manual, 2016. Manual, executables, and source codes are freely available at <http://www.angelogavezzotti.it>.
- R. S. Rowland and R. Taylor, *J. Phys. Chem.* 1996, **100**, 7384–7391.
- S. Alvarez, *Dalton Trans.* 2013, **42**, 8617–8636.
- A. Bondi, *J. Phys. Chem.* 1964, **68**, 441–451.
- L. Infantes and S. Motherwell, *Struct. Chem.* 2004, **15**, 173–184.
- Deposited as Appendix I to: A. Gavezzotti, *Mol. Phys.* 2008, **106**, 1473–1485. Executables and source codes are also freely available at <http://www.angelogavezzotti.it>
- A. Gavezzotti, *New J. Chem.* 2011, **35**, 1360–1368.
- A. Gavezzotti, *J. Phys. Chem. B* 2003, **107**, 2344–2353
- M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheesman, J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, *et al.* *Gaussian16*; Gaussian Inc.: Wallingford, CT, 2017.
- Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.* 2008, **120**, 215–241.
- R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.* 1980, **72**, 650–654.
- J. M. Skelton E. Lora da Silva, R. Crespo-Otero, L. E. Hatcher, P. R. Raithby, S. C. Parker and A. Walsh, *Faraday Discuss.* 2015, **177**, 181–202.
- M. Walker, A. J. A. Harvey, A. Sen and C. E. H. Dessent, *J. Phys. Chem. A*, 2013, **117**, 12590–12600.
- C. F. Macrae, I. J. Bruno, J. A. Chisholm, P. R. Eddington, P. McCabe, E. Pidcock, L. Rodriguez-Monge, R. Taylor, J. van de Streek and P. A. Wood, *J. Appl. Cryst.*, 2008, **41**, 466–470.
- G. Cavallo, P. Metrangolo, R. Milani, T. Pilati, A. Priimagi, G. Resnati and G. Terraneo, *Chem. Rev.* 2016, **116**, 2478–2601.
- V. Colombo, L. Lo Presti, A. Gavezzotti, *CrystEngComm* 2017, **19**, 2413–2423.
- R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*. International Series of Monographs on Chemistry, Oxford, 1990, ISBN: 9780198558651
- J. Kroon, and B. P. van Eijck, *J. Phys. Chem. B* 1997, **101**, 1096.
- S. F. Boys and F. Bernardi, *Mol. Phys.* 1970, **19**, 553–566.
- F. Giberti, M. Salvalaglio and M. Parrinello, *IUCrJ*, 2015, **2**, 256–266.
- G. R. Desiraju, *Angew. Chemie Int. Ed.* 1995, **17**, 2311–2327.
- A. Kitaigorodskii, *The Theory of Crystal Structure Analysis*, 1961, Springer US, ISBN 978–1–4757–0340–5
- A. R. Ubbelohde and K. J. Gallagher, *Acta Cryst.* 1955, **8**, 71–83.

## TABLE OF CONTENT GRAPHICS

**Synopsis:** Weak hydrogen bonds control initial molecular recognition modes, but the structure is determined by the interactions among larger supramolecular assemblies.

