# Facts and factors in the formation and stability of binary crystals
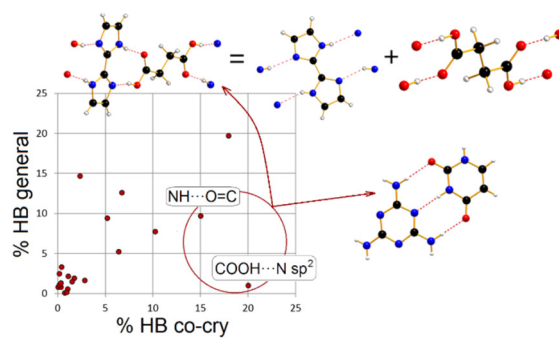
A. Gavezzotti,[a] V. Colombo,[a] and L. Lo Presti[*a,b,c]

[a]Università degli Studi di Milano, Dipartimento di Chimica, Via Golgi 19, 20133 Milano, Italy;

[b] Center for Materials Crystallography, Aarhus University, Langelandsgade 140, 8000 Aarhus, Denmark;

[c]CNR-ISTM, Istituto di Scienze e Tecnologie Molecolari, Via Golgi 19 I−20133 Milano, Italy.

E-mail: *leonardo.lopresti@unimi.it*

**Synopsis TOC:** Toward consistent guidelines for co-crystal design: favourable molecular recognition patterns in co-crystals derive from mixing HB features of individual co-formers.

**Abstract.**

Despite significant ongoing experimental and computational efforts, factors involved in the choice between homomolecular and heteromolecular recognition remain elusive. Here, a large-database study of cohesive energy and intermolecular non-covalent interactions (NCI) in co-crystals from the Cambridge Structural Database has been undertaken. Centrosymmetric space groups (especially P $\bar{1}$) are largely more frequent than unary crystals, while the frequency of chiral space groups is halved. Overall close-packing is observed, but the relative sizes of the two co-formers can vary widely. 86% of extant co-crystals are hydrogen-bonded, all of which include bonding between the two co-formers. Carbonyl oxygens and aromatic nitrogens are the most consistent acceptors, while the donor activity decreases according to COOH > NH >> R–OH series, so that COOH···N(aromatic) is the favorite H-bond. $\pi \cdots \pi$ stacking is another recurring interaction. The lattice energy of the binary crystal is nearly always more stabilizing than the sum of the lattice energies of pure co-formers. When sub-lattices are considered, the AB one is mostly more stabilizing than the AA+BB sum; moreover, in most cases the A-B hetero-pair also ranks first in energy. Finally, it has been demonstrated that co-crystallization mainly involves the evolution to stronger hydrogen bonds than those found in the co-former crystals, implying that hetero-recognition provides a thermodynamic drive to co-crystal formation. Existing co-crystals are a collection of successful attempts at co-crystallization, and conversely their common properties may provide valid suggestions along the path to success.

# 1. Introduction

The crystallization of organic molecules from solution usually involves the association of identical partners to form one-component crystals. This fact is often exploited by organic chemists as a simple and inexpensive method of purification. Much less common is the association of different chemical species to form multi-component crystals. Recently, and in the past, there has been an increasing interest in such systems[1] and techniques are developed for the intentional, predictable preparation of these complex entities.[2] As is often the case with matters related to crystal formation, there is presently a limited understanding and little control of the thermodynamic and kinetic factors involved in the choice between a homomolecular and a heteromolecular recognition process. The consequence is a large amount of serendipity in the outcome of the related chemical experiments. There is an obvious conceptual interest for a fundamental principle of molecular recognition, but also a promise for practical applications, from the preparation of materials with unusual electronic or electro-optical properties from charge transfer[3] to the now more fashionable idea of improving the flexibility of pharmaceutical formulation.[4]

The quest for an answer to the basic question - homomolecular or heteromolecular association? - has been pursued by a number of theoretical approaches[5–7] and experimental screenings.[8–11] The subject has been also frequently reviewed.[1,12–14] Exploiting the richness of structural data present in the Cambridge Structural Database[15] we first establish in this work a fully characterized, error-free sample of 1500-plus two-component crystal structures and a set of 97 crystal structure triads comprising a binary crystal and its two co-formers. Small-molecule crystals (see Section 2 below) are selected to correlate on the most general grounds, where possible, structure and energetics of co-crystals with those of the corresponding co-formers. We then perform a theoretical analysis of these co-crystals, using statistical studies of crystal properties, of the structural patterns of aggregation and of hydrogen-bonding features. In an attempt to improve over concepts based only on geometry, among which hydrogen bonding itself, $\pi$-stacking, or even weak interactions[16], often based on subjective choices of the principal determinants, we report a systematic and quantitative study of the cohesion energies in hundreds or thousands of these crystals. We compare dimerization energies in homodimers and heterodimers, and lattice energies of pure and mixed crystals, also using a partitioning into Coulombic, polarization and hyperpolarization or dispersive energy contributions (admittedly a simplified, but useful model), gaining insight into the types of electronic rearrangements that provide the bonding between different organic molecules. The results provide theoretical and experimental crystal chemists with a panoramic view of the structural and energetic co-crystallization landscape, out of which one may

derive also hints about the essential dualism between homo- and hetero-recognition and, possibly, suggestions about the choice of promising co-formers to guide the synthetic approach.


## 2. Material and methods

This work is focused on the electronic factors and relative ease of homo- *vs.* hetero-recognition. We therefore consider only crystalline systems that result from the aggregation of different organic molecules of approximately the same size and with approximately the same binding abilities, and in which the heterogeneity either has been introduced on purpose or is the result of efforts in that direction. We therefore exclude crystals with unintentional inclusion of small-molecule crystallization solvents, hydrates and solvates, as well as clathrates. We simply call our systems binary, or tertiary, etc. crystals, refraining to enter the somewhat useless debate over nomenclature.[1,17]

The primary search was conducted on the Cambridge Structural Database (CSD; v5.36, 2015 release with nov2014 updates) using CCDC programs Conquest[18] and Mercury.[19] The search parameters were: number of atoms with 3-D atomic coordinates, including H atoms, $\leq 50$; number of chemical units = 2; R-factor < 7.5%; no hydrate, solvate, clathrate in text entries; no disordered or powder diffraction structures; heaviest element allowed Cl, but excluding P, Se, B, Si (for which no reliable potential energy schemes are available); no ions, only organic compounds; allowed elements are thus C, H, N, O, F, S, Cl. This gave 2252 primary hits. Further screening was done by in-house software (the CLP package[20]) as follows. When multiple determinations are available, the one with lowest R-factor is kept, irrespective of temperature of the X-ray determination. X-ray H-atom positions are discarded and hydrogen atoms are assigned at standard C-H, O-H or N-H distances (1.08, 1.00 and 1.00 Å, respectively) according to the standard CLP renormalization procedures.[20] Structures with unknown H-atom positions in alcoholic OH or amino NH groups must be discarded, as these positions cannot be reassigned by automatic recipes. The geometry of the COOH groups have been reset (see SI for details) for structures in which the carboxyl group attached to a phenyl ring showed a disorder with single and double C-O bond distances (due to formation of OH···O hydrogen bonds).

When dealing with crystals with more than one molecular species the crystal structures often show some ambiguities in what concerns the number of molecules in the asymmetric unit and in the stoichiometric ratio of the components, especially when some of them are in a crystallographic special position. Therefore, once the full unit cell content has been tentatively reconstructed, other checks include an evaluation of the crystal packing coefficient (see Section 3.3 below), $C_k$, which

must be around 0.7, and a preliminary estimation by atom-atom parameters of the lattice energy, which must be reasonably stabilizing as an approximate function of molecular size.

A first database (the NOCOR database) consists of 1941 binary or ternary crystal structures with unique and correct space group and compound names, but without full check of atomic coordinates; it is used for statistical screens that do not require a calculation of energies. A restricted database that passes all the screening procedures is the DIM database with 1522 entries, that can be used for the calculation of lattice energies by the cheap and fast CLP atom-atom energy scheme.[20] A subset of 148 binary crystal structures with smaller molecules (the PIX dataset) were selected for lattice energy calculations with the more time-consuming PIXEL procedure.[21] Correspondingly, the COF dataset includes homocrystals of individual co-formers of the dimer structures; this allowed the preparation of a set of 97 triads for which the energy of the dimer and the energy of the separate co-formers are available. To this end, when multiple polymorphic choices of co-formers were possible, just the most stable ones were selected for the composing the triad. At the same time, when X-ray determinations at multiple temperatures were available, preference was given to structures reported at the most similar as possible temperatures. Finally, we also considered a control set (CON) consisting of 2322 general one-component organic crystal structures for which the correct PIXEL energy terms were computed in previous work[22] under the same restraints as discussed above.

The PIXEL approach allows the calculation of the total lattice energy $E_t$ and a separate evaluation of Coulombic ($E_c$), polarization ($E_p$), dispersion ($E_d$) and repulsion ($E_r$) terms. It is much more accurate than simple atom-atom schemes, as it resides on electron densities of isolated molecules, computed at the MP2/6-31G** theory level[16,22]. The same decomposition scheme can be applied to in-crystal molecular pairs to rank them in descending order of interaction energy.[16]  In this way it is possible to assess the relative weight of intermolecular interactions between hetero- or homo- co-formers (in an AB crystal, A···B or A···A / B···B) in determining the overall lattice stability.

The Supplementary Information contains lists of all the refcodes in all the mentioned databases (Tables S1-S5 SI). After all the CSD Database searches and control checks, only about 60% of the retrieved entries were deemed of sufficient quality for the further accurate calculation of lattice energies. In view of the wealth of data anyway available, it was decided to keep on the safe side; by long experience with statistical studies on the CSD it is better have a smaller number of data than taking the chance of polluting the database with uncertain or suspicious data.

## 3. Results and discussion

## 3.1 Space group symmetry

Crystal symmetry and co-crystal stoichiometry have a deep influence on the physical and thermodynamic properties of the solid mixture and on the chemistry of multi-component crystal formation. Full characterization of the materials require *i*) the content of the ASU, *ii*) the relative component stoichiometry, and *iii*) the independent space-group symmetry operations that are needed to reproduce the entire contents of the unit cell, as necessary for a fast calculation of the lattice energy. Each molecular component may be in a general crystallographic position, or on a crystal special position of inversion, mirror or twofold axis (*i*, *m*, 2) symmetry coinciding with its point-group symmetry. Table 1 shows the comparison of space group frequencies in binary crystals (NOCOR and DIM databases) with those in the reference set of organic homocrystals (CON database). In binary crystals, there is a notable increase of $P2_1/c$ and a doubling of $P\bar{1}$ frequencies due to the popularity of centrosymmetric co-formers in special positions; in the DIM database, $P2_1/c$, $P\bar{1}$, $C2/c$ and *Pbca* space groups together account for more than 80% of the data.

**Table 1**. Space group frequency distributions in the two-component crystal databases, as compared with those in general organic structures.

| space group[a] | % frequency, unary crystals control database | binary crystals (NOCOR database) | binary crystals (DIM database) |
|---|---|---|---|
| $P2_1/c$ | 36.0 | 41.4 | 43.1 |
| $P\bar{1}$ | 13.7 | 27.0 | 28.1 |
| $C2/c$ | 6.6 | 9.1 | 8.5 |
| $P2_12_12_1$ | 11.6 | 3.7 | 3.8 |
| $P2_1$ | 6.7 | 3.8 | 3.9 |

[a]Just the most frequent space groups (> 3.0 %) are here considered.

There is also an increase of $C2/c$, due to its ability to accommodate twofold-symmetric co-formers in the corresponding special position. The frequency of chiral space groups $P2_1$ and $P2_12_12_1$ is impressively reduced. The most common lattices, with the only exception of $P\bar{1}$, all have glide- or screw-translational symmetry elements, due to their well known ability to accommodate small and symmetric molecules under steric or electrostatic complementarity requirements. Structures with higher-order axes are rare, as in general organic structures (there are only 2 $P4_32_12$, 3 $P4_3$ and one $P3_22$ space groups in DIM).

The frequency variations in Table 1 most likely stem from prejudicial choices of co-formers, rather than from physical reasons connected with intermolecular bonding. The marked decrease of

the chiral space groups can be simply explained: people who struggle to synthesize or characterize chiral compounds (among which nearly all natural compounds) which must crystallize in a chiral space group, always want a crystal structure determination, while workers that aim at binary crystals have no interest in using chiral compounds, more expensive and perhaps more troublesome. This suggests that it is the percentage of chiral space groups of general organics that is abnormally high in the Database, for social (policy) reasons. To the sake of comparison, from the 2005 release (v5.26) of the CSD, it was found[23] that non-chiral organics strongly prefer crystallization in centrosymmetric space groups, but in a non-negligible minority of cases ($\approx$ 11 %, 1:4 in proportion) they choose to crystallize in either the $P2_12_12_1$ or $P2_1$ Sohncke group. The frequency differences with respect to Table 1 seem to indicate that chiral compounds are indeed underrepresented in NOCOR and DIM datasets.

## 3.2 Chemical composition and lattice site symmetries

Binary crystals exhibit a wide variety of symmetry combinations. A first subset of the DIM database (1212 structures) has both components in general position or both components in a special position; mostly (95.6 %) the stoichiometry is 1:1, for the rest the stoichiometry is 1:2 with only one 1:3 case (CSD refcode PEFGIS). For the 1:1 stoichiometry the ASU usually contains a full A and a full B molecule, both in general position (formally, the number of "chemical units" in the asymmetric unit is then $Z'$=1). In a few cases the ASU has ½ A and ½ B molecule, both in special positions (formally $Z' = $ ½): for example, centrosymmetric space groups with both molecules sitting on lattice inversion centers, or, in space groups $Pnma$ or $P2_1/m$, one molecule on an inversion center and one on a mirror plane (e.g. AWUVEU and IPEWEG, respectively). Exceptionally, AMILUD01 is in space group $C2/c$ with both molecules sitting on a twofold axis. In just 4 cases the ASU contains two molecules of A and two molecules of B, having thus formally $Z' = 2$.

In a second subset of 1522-1212= 310 crystal structures, for 85.5 % of the entries the stoichiometry is 1:2 with one molecule in a special position (mostly an inversion center) and one in a general position. The rest all exploit $C2/c$ symmetry with 1:2 stoichiometry. Given the complex structure of C2/c symmetry elements, there are many possible combinations, but in all cases one component is on a twofold axis while the other may be in general position or in either of the two different lattice inversion centers. In the above examples, Z' is undefined.

Even though it is difficult to establish in general whether point symmetry elements will be maintained upon crystallization or not, and to what extent, it is worth noting that the frequent occurrence of some specific high-symmetry small molecules in our datasets, such as pyrazine and symmetric dicarboxylic acids, might influence the occurrence frequency of space groups, especially

when inversion centers in the molecular point group are also conserved in the crystal point symmetry.

### 3.3 Molecular sizes and packing efficiency

In order to compare and discuss the relative sizes of components in a binary crystal, we resort to a simple indicator, the packing coefficient $C_k$, or the sum of molecular volumes in the unit cell divided by cell volume. Overall $C_k$ values for organic crystals are in the 0.65-0.85 range peaking at 0.7. For two-component crystals one can define a separate value for each molecular species, reflecting the relative space occupation, so that $C_{k1} + C_{k2} = C_k$.

Fig. 1 shows that binary crystals can be obtained from co-formers of widely different sizes, and also that on the whole the average overall packing coefficient is on the expected range: averages are 0.752(1), 0.751(2), and 0.753(7) for 1:1, 1:2 and 2:2 stoichiometries, respectively. The thickness of the distribution cluster reflects the minor effect of different molecular shapes; for example, a partner whose partial coefficient is 0.35 can accept partners with a coefficient of 0.35 to 0.45. This mostly reflect the number of atoms in the two coformers, so that large differences between individual $C_{k1}$ and $C_{k2}$ values are associated with uneven number of atoms (Fig. S1 SI). Taken at face value this result reveals that relative size is not an issue in the likelihood of co-crystal formation.
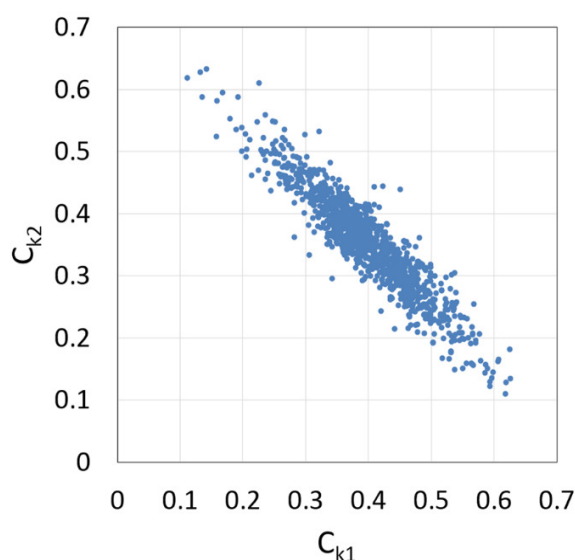


**Figure 1.** Scatterplot of partial packing coefficients $C_{k1}$, vs. $C_{k2}$ for binary crystals of 1:1 stoichiometry. See Figure S2 SI for the whole plot with other stoichiometries.

More generally, this is a confirmation of the well known principle according to which packing efficiency is a necessary but not sufficient condition for crystal stability. A more thorough discussion of the case, including plots for crystals with 1:2 and 2:2 stoichiometries that just confirm the above conclusions, can be found in Fig. S2 SI.

**3.4 Hydrogen bonding**

**3.4.1 Co-former frequency: chemical types.** We analyze in this section the chemical composition and in particular the hydrogen bonding abilities of co-crystal-forming compounds. The preliminary conclusions coming from these analyses are qualitatively informative, but quantitative interpretations are doubtful because all samples are to some extent biased by chemical prejudices affecting the operators that design the co-crystallization experiments. Besides, there is obviously no information about experiments with negative results and/or failed co-crystallizations.

Table 2 shows the names and frequencies of appearance of compounds or chemical functions in existing co-crystals; hydrogen bonding is predominant. For a detailed analysis, the following atoms are considered as H-bond acceptors: amino $N(H_2)$ or $N(HR)$, $sp^2$ N, mostly in heterocyclic rings, and nitrile N; ether R-O-R', carbonyl (C)=O, alcohol O(H), nitro (N)=O, and (S)=O; thioether, thienyl, and thiol sulfur. H-bond donors are amino (N)-H, amide (CON)-H, acid (COO)-H and alcohol (O)-H hydrogens. A hydrogen bond is recognized when the donor-acceptor separation is smaller than 0.8 times the sum of accepted random radii (H 1.10, N 1.64, O 1.58, S 1.81 Å). A different choice of any of the above parameters would only marginally affect the main results.

**Table 2.** Compound or group names found in the NOCOR database (1941 entries)

| Name | Structure | Type | % frequency |
|---|---|---|---|
| acid | –COOH | Donor | 49.3 |
| aza-aromatics[a] | ⊃>N– | Acceptor | 22.8 |
| hydroxy+phenol | –OH, ⊃>C–OH | Donor | 15.4 |
| amide | –CO–NH– | Donor | 13.2 |
| nitro | –NO$_2$ | Acceptor | 11.1 |
| chloro | –Cl | Acceptor | 8.4 |
| fluoro | –F | Acceptor | 5.2 |
| urea+thiourea | >N–CO–N< >N–CS–N< | Acceptor | 4.3 |

[a]*Pyridine, pyridyl, bipyridine, pyrazine, etc.*

In order to keep the analysis as general as possible, no functional group-based criteria were adopted in selecting structures to be included in the examined datasets. Therefore, in this study multi-function co-formers, more closely related to pharmaceutical applications, are likely slightly underrepresented.

Table 3 shows the possible combination of H-bonding schemes, from a simple count of the number of donor and acceptor atoms in the above list, along with the percent of those actually occurring out of the possible ones. In about half (48%) of the cases the systems can form all

combinations of H-bonding, but, out of these, only relatively few (15.2%) actually do. What mostly happens is only inter-co-former bonds (43%) or inter-co-former bond plus bond over one co-former only (41%). When one of the two partners is unable to form H-bonds with itself, the large majority of the cases (74.2%) has only the inter-co-former bond. There are virtually no instances of co-crystals where only one of the co-formers is strongly H-bonding, because in that case separation is much more likely than co-crystallization, due to the very different binding strengths.

**Table 3.** Possible combinations of H-bonding in A-B, 1:1 co-crystals, established from the presence of donor and acceptor atoms, and percent of actually occurring ones. Sample of 1153 crystal structures. See SI (Supplementary Information to Table 3) for details of relevant HB interactions and some examples.

| possible H-bonds [a] | AA, AB and BB 48.0% | AA and AB 37.4% | AA only 6 | none 14.1% |
|---|---|---|---|---|
| *of which actually formed[b]* | | | | |
| AA, AB and BB | 15.2% | - | - | - |
| AA and AB | 40.5% | 20.6% | - | - |
| AB only | 42.5% | 74.2% | - | - |
| AA only[c] | 2 | 6 | 2 | - |
| AA and BB[c] | 4 | - | - | - |
| none[c] | 4 | 16 | 4 | 100% |

[a] *AB only hydrogen bonding is chemically impossible (would require the presence of donors without acceptors and vice versa).*
[b] *Number or percent of each actual combination within the possible ones.*
[c] *Integers are number of sporadic cases from inaccurate location of hydrogen atoms; weak N-H···O interactions at long distance; intramolecular H-bond formed, especially in trinitrophenol. See SI for details.*

A significant 15% of the co-crystals in the database do not involve hydrogen bonding that therefore is not an indispensable feature for co-crystallization. Most of these occurrences involve stacking of large aromatic moieties, like in the numerous examples of arene-perfluoroarene interaction[24] or incipient charge-transfer between π-donor and π-acceptor electronic systems (fulvalenes, cyanobenzenes, etc.). The subject deserves a separate treatment, which is the subject of work in progress.

**3.4.2. Recognition frequencies of chemical groups.** The data summarized in condensed form in Table 3 can be further analyzed to show the relative frequencies of each type of donor-acceptor pairing in co-crystals. For simplicity, and without loss of generality, the analysis was carried out on a subset of 1212 structures with 1:1 stoichiometry from the DIM dataset (2226 hydrogen bonds, for an average of 1.8 hydrogen bonds per structure). Results are shown in Fig. 2. The most consistent acceptors are the carbonyl oxygen and the aromatic nitrogen, followed at a distance by aminic nitrogen and alcohol or nitro oxygen.
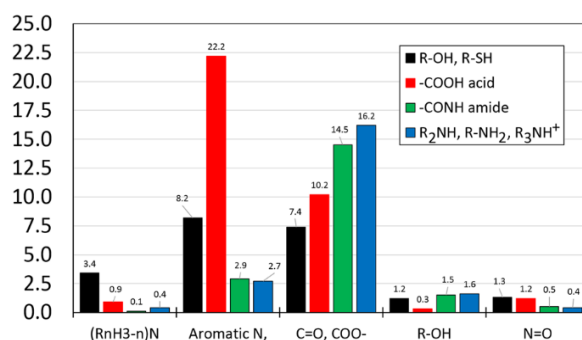
**Figure 2.** HB pairing frequency for various HB donors (alcohols, black; acids, red; amide, green; amine, blue) and acceptors.

The order of donor activity is carboxylic hydrogen by far first, followed by NH hydrogen and, at a distance, by alcoholic hydrogen. The resulting pairing modes are, in order of decreasing frequency, COO-H···N(aromatic), amide or amino N-H···O=CO, followed by alcoholic O-H to carboxyl group or to amino nitrogen, but curiously very seldom to alchol oxygen. Note once again that these results do not necessarily reflect the relative strength of the linkages, but may be influenced to an unknown extent by biases in the synthetic routes.

In a complement to Fig. 2 and Table 3, Fig. 3 shows the frequency of appearance of each kind of hydrogen bond in co-crystals, compared with crystals of general unary organic compounds. The frequency of alcohol O-H···O hydrogen bonds is largely reduced in binary crystals; while the general organic dataset contains a large number of saccharide crystal structures, sugars are presumably very impractical for co-crystallization. On the contrary, the frequency of COOH to heterocyclic, mostly pyridyl nitrogen is greatly enhanced due to the popularity of aza-aromatics as co-formers. Correspondingly, the two top co-former names are 'acid' and 'pyridine' (Table 2). The advantage of such COO-H···N hydrogen bonding is well illustrated by the crystal structure of nicotinic (3-pyridinecarboxylic) acid, which forms that bond rather than the classic acid-acid cyclic double hydrogen bond. Concomitantly, and, perhaps, consequently (many nitrogen heterocycles contain both N and NH groups) also the NH···O=C frequency is higher in binary compounds. In a recent survey,[25] the COO-H···N hydrogen bond energy has been estimated at 49 kJ mol$^{-1}$ against about 35 kJ mol$^{-1}$ for the COO-H···O=C hydrogen bond.
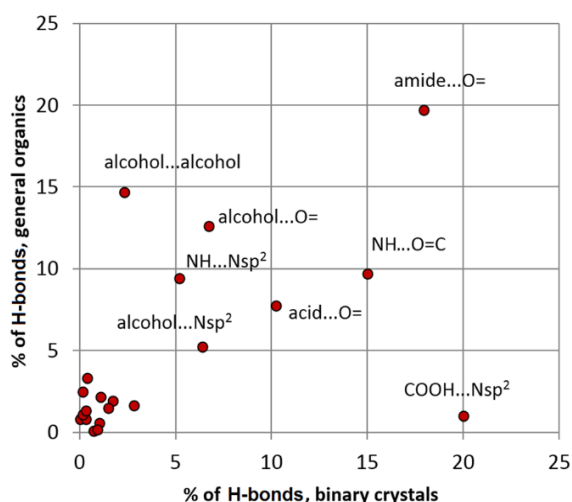
**Figure 3.** The relative frequency of hydrogen bond formation in binary crystals (3922 total hydrogen bonds from the DIM database) and the corresponding frequency in unary crystals of general organic compounds (CON database, 13866 total H-bonds).

### 3.5. Energy calculations

**3.5.1. Lattice and sub-lattice energy calculations; "coupling" energies.** The lattice energies of 1153 structures of 1:1 binary crystals out of the DIM database were evaluated by the CLP procedure in the atom-atom formulation. This requires a negligible computing effort, while more accurate PIXEL calculations would have been far too expensive. Besides, the many-body character of the PIXEL polarization energy prevents the partitioning of total lattice energies into contributions from separate molecular entities.

The CLP procedure allows an estimation of Coulombic ($E_c$), polarization ($E_p$), dispersion ($E_d$) and repulsion ($E_r$) two-body additive terms. These are all empirically calibrated quantities, but are nevertheless reliable enough for general considerations of the kind exposed in the following. Consider a unary crystal made of component A. The lattice energy is given by:

$$E_t(\text{unary}) = E_c(\text{A}) + E_p(\text{A}) + E_d(\text{A}) + E_r(\text{A}) \tag{1}$$

$E_t$ is the energy difference between one mole of free A molecules in the gas phase and one mole of crystalline material. This can be compared with the sublimation energy which is in many cases also available experimentally. For an A-B binary crystal, the two-body character of the atom-atom approach allows a partitioning into contributions from each of the components:

$$E_t(\text{binary}) = E_c(\text{XY}) + E_p(\text{XY}) + E_d(\text{XY}) + E_r(\text{XY}) \tag{2}$$

where X, Y = A, B. Therefore the total lattice energy can be represented by a sum of $E(\text{AA})$, $E(\text{BB})$ and $E(\text{AB})$ sub-lattice energies. Here, $E_t$ is the energy difference between one mole of component A plus one mole of component B in the gas phase, and one mole of AB crystalline material. This is analogous to a sublimation energy, but, to the best of our knowledge, there are no experimental

determinations available for co-crystals, also because it is difficult to imagine an appropriate experimental setup for such a measurement. Quantitative comparisons between calculation and experiment are thus impossible and one must rely on the accuracy of the potentials, resulting from calibration on unary crystals.

Total lattice energies of co-crystals are in the same range and of the same order of magnitude as lattice energies of unary crystals. There is no sign of global gain or loss in overall packing forces. $E_t(AA)$ and $E_t(BB)$ can be very different (Fig. S3 SI), as just 9 % of the structures with 1:1 stoichiometry have $E_t(AA) = E_t(BB)$ within 5 kJ·mol$^{-1}$, a sign that co-crystallization can occur between molecular units of very different packing abilities.

Fig. 4 plots the $E_t(AA)$, $E_t(BB)$ and $E_t(AB)$ sublattice contributions against the difference of the A and B molecular volumes $\Delta V$. As previously mentioned, the cohesive energies of co-crystals are in the same range as those of unary crystals. At a glance, $E(AB)$ is mostly more stabilizing than both $E_t(AA)$ and $E_t(BB)$. The broad trends are such that the each sublattice energy becomes more stabilizing with increasing volume of its molecular component; this rough correlation is a somewhat obvious consequence of the general increase in cohesive power with increasing molecular size and number of interactive electrons, modulated by the presence (enhancing) or absence (depleting) of hydrogen bonding, as well as by the fact that with increasing quantum number, more polarizable electrons can be cumulated under a just minor increase in volume (e.g. in the series of halogens). The $E_t(AB)$ cross-interaction is insensitive to the volume difference, being related to the sum of the two volumes; its distribution is therefore roughly globular in the plot of Fig. 4.
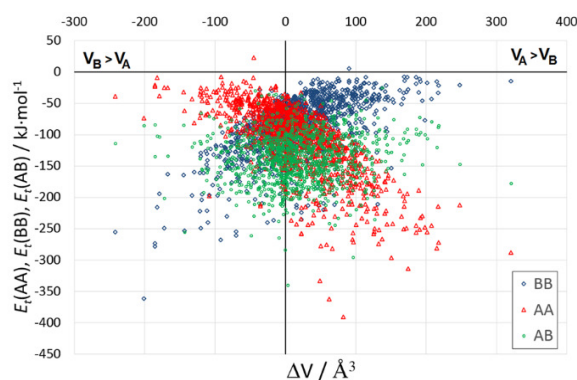


**Figure 4.** Sublattice energies $E_t(AA)$ (red; increasingly stabilizing left to right), $E_t(BB)$ (blue; increasingly stabilizing right to left) and $E_t(AB)$ (green) *vs.* molecular volume difference $\Delta V = V(A)-V(B)$. kJ·mol$^{-1}$ and Å$^3$ units.

Indicators other than total lattice energies must be devised for an appreciation of the balance between hetero- and homo-interactions in co-crystals. To this end, within a given binary crystal structure one may define the "coupling energy", $E_{coup}$, as

$$E_{coup} = E(AB) - 1/2[E(AA) + E(BB)]$$

(3)

All lattice energies being negative numbers, $E_{coup} < 0$ means that the contribution to co-crystal stabilization due to the A-B sub-lattice is larger than the average contribution from the sub-lattices of the co-formers. $E_{coup}$ can thus be assumed as a first approximation to the thermodynamic driving force toward the formation of a stable co-crystal. However, $E_{coup}$ (eq. 3) does not correlate with the total packing energy $E_t$ of eq. 2 (see Fig. S4 SI). Indeed, there is no reason why it should. Figure 5 shows the results. As expected, the larger $E(AB)$, the higher the chances that $E_{coup}$ be a negative number, and the scatterplot shows a pronounced southwest-northeast slanting. Notably, in the large majority of structures $E_{coup} < 0$.
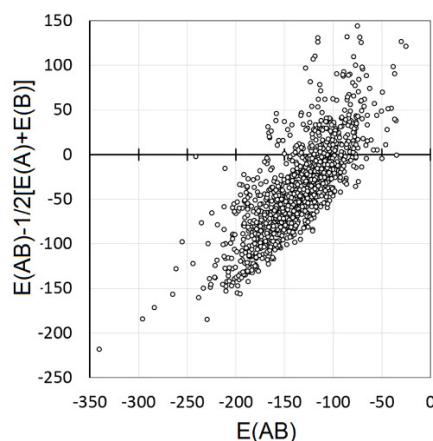


**Figure 5.** A scatter-plot of $E_{coup}$ for 1153 binary crystals. Atom-atom, CLP-scheme energies. kJ·mol$^{-1}$ units.

Notwithstanding the many approximations, this result underscores the importance of hetero-interaction in the formation of co-crystals and implicitly points to a significant thermodynamic drive in their stabilization.

Fig. 6 shows some typical examples of largely stabilizing $E_{coup}$, where the packing of the A···B pair dominates the lattice cohesion for different reasons. The co-crystal in Fig. 6a (CSD refcode NOQFUV) has a flat and rigid heteroaromatic moiety, coupled by very tight π···π stacking to a tetrachlorophenol unit, with extensive advantage in dispersion energy; the structure in Fig. 6b (CSD refcode QOGMII) has two polyphenols joined by six strong OH···O hydrogen bonds, with a very large excess in stabilizing Coulombic interaction. At the other extreme, the structure of SORBPY (Fig. 6c), has $E_{coup} \gg 0$. The sorbitol units act as strong hydrogen bond donors and acceptors and segregate into helicoidal H-bonded motifs, so that one self-term strongly prevails

over cross-interactions. The pyridine is allocated in the free space between helices with only a weak OH⋯N interaction. NOQFUV and QOGMII structures may be called real molecular complexes, while SORBPY is more like a solvate where the included molecule plays a lesser role.
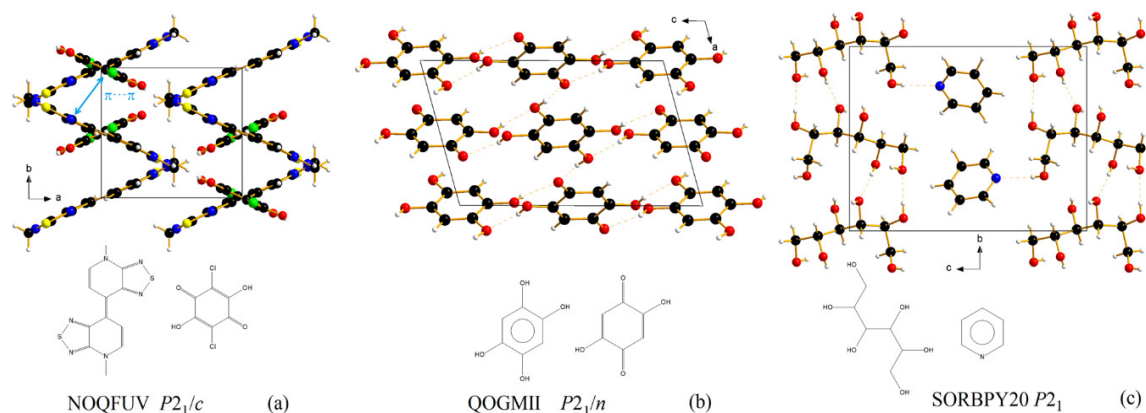


**Figure 6.** Co-crystals with (a-b) very large and negative $E_t(AB)$ and (c) very large and positive $E_{coup}$. (a), blue arrow: $\pi\cdots\pi$ stacking interaction. (b,c), dashed lines: hydrogen bonds. Color code: C: black; H: white; O: red; N: blue; S: yellow; Cl: green (graphics by Diamond v3.2k, (c) 1997–2014 Crystal Impact GbR, Bonn, Germany).

**3.5.2. Molecular pair energies.** In a molecular crystal, it is straightforward to single out molecular pairs, related by some symmetry element or unrelated when the crystal has more than one molecule in the asymmetric unit. Moreover, in a co-crystal, each pair may involve AA, BB or AB partners. These pairs can be computationally isolated out of the crystal and their interaction energy can be evaluated by PIXEL, with its Coulombic polarization and dispersion components. The pairs can thus be assigned an energetic rank according to the descending order of their interaction energies. If the drive to co-crystallization is given by a very strong A-B affinity, A-B pair interaction energies should be among the top-ranking ones in the crystal structure. Fig. 7 shows the frequencies with which the AB pairs have a given rank in the PIX dataset; for comparison, the Figure also shows the same frequencies for homomolecular, but non-symmetry related pairs, in crystals with Z' > 1 from the CON dataset. In the co-crystal dataset, one A⋯B pair is top-ranking in more than 70 % of the instances, or among the first five top-ranking pairs in 60-50% of the instances. In our PIX database more than 98 % of the A⋯B pairs are H-bonded (Table 3), so that all top-ranking A⋯B pairs must also be H-bonded (this results from conscious planning on the part of the experimenters). On the one hand, the high proportion of top-ranking AB pairs confirms the importance of hetero-recognition, especially driven by H-bonding, in co-crystallization, but, at the same time, one is warned that a strong A-B drive is not indispensable. The trend is similar for the homomolecular asymmetric pairs, as previously noted,[26] where (presumably) symmetry constraints would hamper the optimal pairing of complementary chemical functions. In general, structures where a homomolecular pair ranks first usually exploit favorable stacking interactions between inversion-

related molecules (see for example AMILEN and AMILUD). However, it must be stressed that this does not necessarily imply weak A···B interactions on absolute grounds. For example, in ASAXUN (4-fluorobenzoic acid + isonicotinamide), heteromolecular pairs are set up which are strongly related through COOH···amide hydrogen bonds. However, a very large repulsive contribution overwhelms the favorable Coulombic and polarization terms, resulting in a second-ranking A···B pair. The difficulty of properly weighing all these considerations stems from the extreme flexibility of organic crystal packing, where intermolecular factors are always multiform and variable, and the resulting structure is always a compromise among several strains.
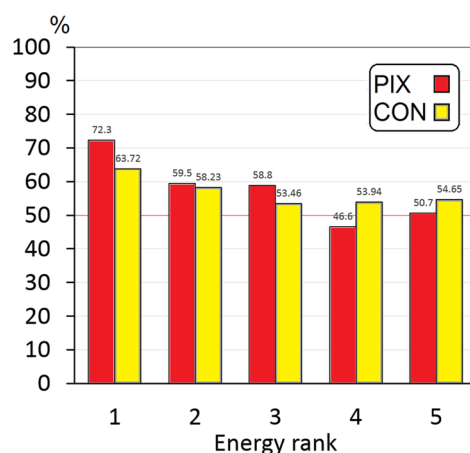


**Figure 7.** Percent of structures in which the cohesive energy of one A-B molecular pair has the rank marked on the horizontal axis (first to fifth). Percents add to more than 100 as each structure has many such pairs. Red: heteromolecular pairs in co-crystals of the PIX dataset. Yellow: asymmetric pairs in crystals with more than one molecule in the asymmetric unit from the CON dataset (419 structures).

**3.5.3. Co-crystals *vs.* crystals of individual co-formers: energies.** The search for clues about co-crystal stability may proceed through a comparison between the mixed structure of a given co-crystal and the pure structures of its separate co-formers. We have assembled a co-former database (the COF database, 165 crystal structures), preparing a list of 97 triads including the three implied structures: co-crystal from the PIX database plus its two co-formers, for which also the PIXEL energies were computed. Lists of the corresponding CSD refcodes are given in Tables S3-S4 SI. When many structure determinations for a co-former are available, the one with the most negative PE is selected, compatibly with the requirement of having the three structures determined at the same temperature.

Essential question are: (*i*) do co-crystals somehow gain stability over crystals of individual co-formers, and (*ii*) if yes, what are the relationships in terms of symmetry elements, packing motifs or contributions to the packing energy, between the co-crystal and the corresponding co-former

crystals? An immediate measure of energetic advantage is the excess energy, $PE_{exc}$, the difference between the total packing energy (PE) of the binary system and the sum of the PE's of the separate one-component crystals of the two co-formers:

$$PE_{exc} = PE(AB) - [PE(A) + PE(B)] \qquad (4)$$

These energies refer to the three different, full crystal structures and should not be confused with the energies defined in eq. (3) which refer to sub-lattices within the same co-crystal structure. When $PE_{exc} < 0$, the lattice energy of the binary crystal is calculated to be more stabilizing than the sum of the separate components, also offering suggestions about the ease of co-formation. Fig. 8a shows a scatterplot of excess lattice energies in co-crystals. At first sight, the excess energy is nearly always negative, confirming that the lattice energy of the co-crystal is more stabilizing than the sum of the lattice energies of the co-formers. For a more conservative analysis, 37.1% of the structures have an excess energy within a range of $\pm 5$ kJ·mol$^{-1}$, which we consider as an intrinsic uncertainty threshold. However, even with this proviso the majority (55.7 %) of the structures have $PE_{exc} \ll 0$ while only 7.2 % of the structures show $PE_{exc} \gg 0$. The breakdown into components (Fig. 8b) shows that large stabilization comes mainly from a gain in Coulombic terms balanced by an increase of repulsion that is from the formation of stronger (shorter) hydrogen bonds. In comparison, dispersion energies play a much lesser role.
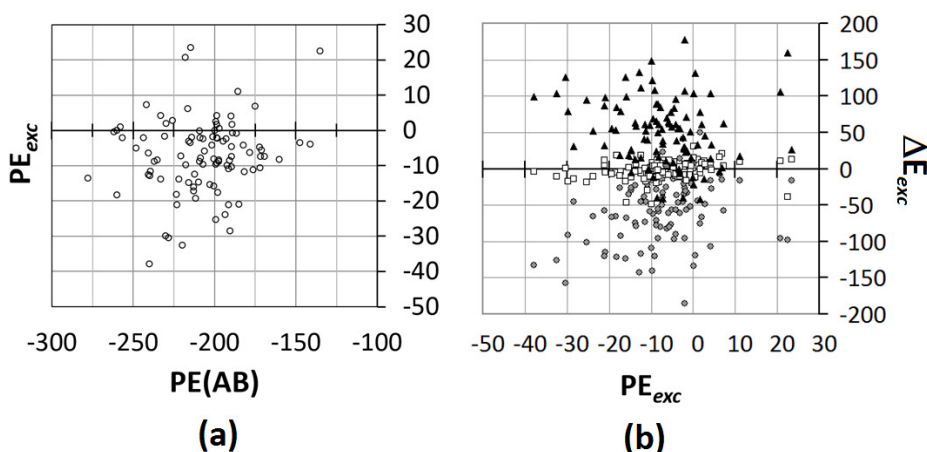


**Figure 8** (a) A scatter-plot of the excess energy, eq. (4), versus the total lattice energy in 97 co-crystal structures. (b) Coulombic + polarization (gray circles), dispersion (white squares) and repulsion (black triangles) energy contributions against excess energy. PIXEL energies, kJ·mol$^{-1}$ units. Figure S5 SI shows the same result in the form of a PE(A)+PE(B) versus PE(AB) graph whose slope is y = 0.965(5) x.

It should be noted that data in Figure 8 reflect the criteria mentioned in Section 2 above, *i.e.* we adopted the prudent choice of computing $PE_{exc}$ from the most stable co-formers when multiple polymorphic choices were available. This means that polymorphism, when present, implies at best a

slight lowering of some $PE_{exc}$ estimates, but does not change the overall picture. Conformational rearrangements of the co-former may also occur from its own crystal to the co-crystal, involving an energy gain or penalty, $\Delta E_{conf}$. In principle, $\Delta E_{conf}$ can be supplied by an *ab-initio* calculation of the total electronic energy of molecules extracted from the crystal structure. However, as already noted,[27] these numbers are highly sensitive to minor detail in the quality of the X-ray work, especially for the positions of crucial H-bonding hydrogen atoms; therefore energy differences are usually largely overestimated. In our case, out of 190 $\Delta E_{conf}$ values only 21 are $> 15$ kJ mol$^{-1}$; when total = lattice + conformational energies are considered, 27% of the co-crystals are less stable than the sum of their co-formers.

The predominance of cases in which the co-crystal is substantially more stable than the sum of its co-formers is another clear indication of added thermodynamic stability from hetero-recognition in the observed co-crystals.

**3.5.4. Co-crystals *vs.* crystals of individual co-formers: structure.** When $PE_{exc} \approx 0$, the pattern of intermolecular non-covalent interactions (NCI) of individual co-formers is nearly preserved in the co-crystal. For example, in the bis-imidazole-succinic acid co-crystal (Fig. 9) all three crystals form extended, chain-like H-bonded motifs with two symmetric NH (OH) donors and two N (O) acceptors. In the co-crystal, the number of donors and acceptors and the organization of the chains do not change, but donors and acceptors are swapped and XH$\cdots$Y (X, Y = N, O) cross-interactions are formed. Thus, $PE_{exc} \approx 0$ stems from preserved features and compensating effects: stronger OH$\cdots$O hydrogen bonds and weaker NH$\cdots$N ones change into OH$\cdots$N and NH$\cdots$O bonds of intermediate strength.[24]
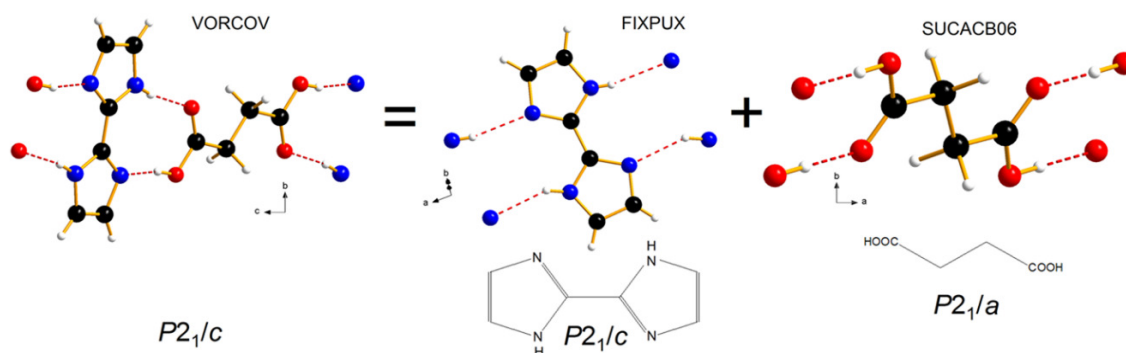


**Figure 9.** Hydrogen bonds (red dashed lines) for a representative co-crystal system with $PE_{exc} \approx 0$. CSD refcodes and space groups are also shown. Other examples are the chloranilic acid-pyrazine and resorcinol-urea co-crystals (Figures S6-S7 SI).

When PE(AB) prevails over the sum of PE(A) and PE(B) ($PE_{exc} \ll 0$), some cooperative mechanism operates to provide more favorable NCI networks in the co-crystal. This may be due to

a reinforcement of previously existing interactions, or, more likely, to the formation of new stabilizing interactions. In its crystal, the melamine molecule forms eight NH···N donor and five N···HN acceptor sites; uracil forms two NH···O and two O···HN interactions, but one potential acceptor carbonyl oxygen remains free (Fig. 10). In the co-crystal four NH···N and four H···O bonds appear together with three N···HO and four O···N short contacts with both molecules acting as H-bond acceptors. Strong cooperative networks, resembling the cytosine-guanine pairing motif, are thus formed in a strictly planar layer (Fig. 10) giving the co-crystal its enhanced stability.
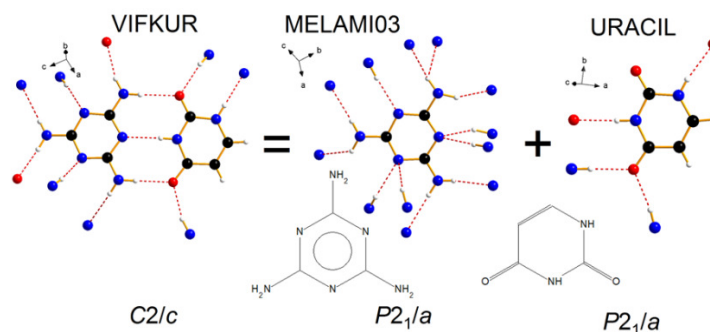


**Figure 10.** Hydrogen bonds (red dashed lines) for a representative co-crystal system with PE$_{exc}$ << 0. CSD refcodes and space groups are also shown. Other examples are in Supplementary Information (Figure S8 SI).

When PE$_{exc}$ >> 0, some breakup of favorable NCI patterns present in the lattice of one or both co-formers must occur, an example is the aminopyrimidine-phthalic acid system (Fig. 11). The co-formers can form extended OH···O or NH···N networks in their crystal. The co-crystal has, instead, cyclic arrangements containing two pairs of co-formers (tetramers), saturating all H-bonding sites. Therefore, the crystal can be onsidered as composed by strongly bound (phtalic acid)$_2$:(aminopyrimidine)$_2$ units, unable to establish further 3-dimensional liaisons, being actually a scarcely effective supramolecular synthon.
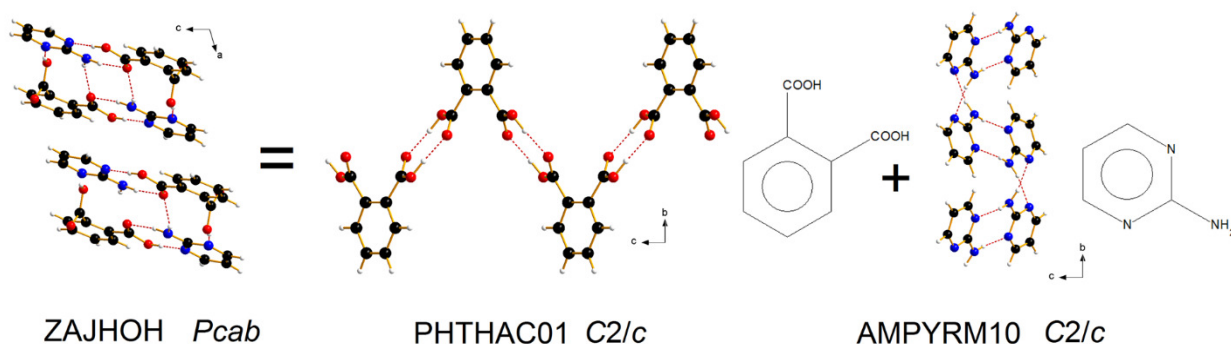


**Figure 11**. Hydrogen bonds (red dashed lines) for a representative co-crystal system with PE$_{exc}$ > 0. CSD refcodes and space groups are also shown. Other examples are in the Supporting Information (Figs. S9 and S10).

In general, as the molecular MP2 charge densities employed to compute the packing energies at the PIXEL theory level (Section 2) are evaluated at $T = 0$ K, $PE_{exc}$ should directly determine the thermodynamic driving force, in terms of Gibbs free energy changes, toward the co-crystal, which in turn might be related with the equilibrium composition of the ternary A, B, AB mixture at $T = 0$ K. We expect that $PE_{exc}$ should correlate with the easiness of obtaining a given co-crystal upon grinding in a mortar crystals of the corresponding co-formers. To a certain extent, it should also help in determining the likelihood of obtaining the co-crystal by mixing reagents in solutions, especially in conjunction with negative $E_{coup}$ estimates (Section 3.5). Unfortunately, no experimental information are available from the Cambridge Structural Database on how a given co-crystal was obtained, or whether it is stable or metastable. We believe that further speculations on this topic should be substantiated by experimental observations: the simple lattice energy calculations here proposed at the more accurate PIXEL and at the less accurate atom-atom CLP theory levels provide an useful guidance at best.

### 3.6. Conclusions

The structural and energetic properties of crystals made of two molecular species (co-crystals) have been studied over a large database, also in comparison with a very large control database of unary organic crystals. We assume that existing co-crystals sample imply deliberate, or at least successful attempts at co-crystallization, so that conversely their common average properties may provide valid suggestions along the path to success. We summarize here the main results, from which hints to synthetic routes can be derived:

1) We have obtained a database of 1522 co-crystal structures with fully corrected and explicit unit cell content, stoichiometry, and atomic coordinates including renormalized hydrogen-atom positions. The stoichiometry is 1:1 (76%) or 1:2, with only one 1:3 structure. We also provide 97 triads with the crystal structure of the co-crystal and those of its two co-formers (Table S4). These databases can be obtained upon request and can be useful for further studies.

2) 80% of co-crystals are in centrosymmetric space groups (57% in the control; Table 1), with $P\,\overline{1}$ frequency doubled. Chiral space group frequency is 7% (19% in the control). Special positions are commonplace, there being a considerable population of centrosymmetric (and to a lesser extent, twofold-symmetric) co-formers residing in the lattice sites of corresponding symmetry. Using co-formers with point-group symmetry seems to be a promising route to stability.

3) Co-crystals are ordinarily close packed (packing coefficient average of 0.75; Fig. 1). The relative sizes of the two co-formers can vary widely, the distribution of cell space occupancy peaking at 50:50 but extending to as much as 80:20. Rather obviously, molecular size is only very broadly

related to intermolecular cohesive properties (the heat of sublimation of acetic acid is the same as that of naphthalene).

4) 86% of extant co-crystals are hydrogen-bonded (Tables 2-3). Appearing chemical functions are, in descending order of frequency, acid, aza-aromatic, hydroxy-phenol, and amide (Table 2). All H-bonded co-crystals include a H-bond between the two co-former units (A-B), but a good percent include also H-bonding between molecules of one co-former (A-A) and a significant minority include all three possible combinations (A-A, A-B and B-B). Carbonyl oxygen atoms and aromatic nitrogen atoms are the most consistent acceptors, while the donor activity decreases according to COOH > NH >> R–OH series (Fig. 2-3). An absolute favorite is the COOH···N(aromatic) hydrogen bond, particularly strong also from theoretical considerations and presumably also a favored by the deliberate choice of pyridyl acceptors by synthetists. Also frequent is H-bonding of all kinds of donors to the acceptor C=O oxygen. In comparison with the control distribution, the alcohol O-H donor goes to all acceptors except its own O(H) acceptor; this apparently surprising result can be explained by the very high frequency of sugars (where OH···OH is the only chance) in the control database. Unless large π-stacking units are present, hydrogen bonding is a must for co-crystallization; using the information in Fig.3, for example coupling a co-former with alcohol functions only to give it a chance to form a stronger H-bond might be a favorable synthetic route.

5) Non H-bonded crystals invariably exploit stacking between flat aromatic systems. Reverse polarization seems indispensable, as occurs for example between hydrocarbons and perfluorohydrocarbons.

6) Lattice energies of co-crystals are in the same range as those of unary crystals of comparable size and chemical nature. For an A-B system, A-A, A-B and B-B sub-lattice energies can be calculated. The inter-dimer sublattice contribution, A-B, tends to be larger than either A-A or B-B (Fig. 4); a first indication that co-crystal formation has a thermodynamic drive in hetero-recognition. Nevertheless, A-B drive is not indispensable; at the other extreme, a strongly H-bonded co-former may include a weakly binding partner (Fig. 6c) in what can be called an inclusion compound rather than a co-crystal system (energy calculations may provide a quantitative gauge to nomenclature issues).

7) Cohesion energies of molecular pairs singled out of the crystal structures show that mostly (72%) the A-B hetero-pair is the top-ranking one, and that the number of A-B pairs in the first three top ranks is very high (Fig. 7). This is another indication that there must be an energetic drive in hetero-recognition, even at dimerization stage, possibly with consequences on early stages of crystal nucleation. Nowadays, quantum chemical calculations can provide *a priori* estimates of dimerization energies and hence hints for synthetic approaches.5

8) The lattice energy of the co-crystal is nearly always more stabilizing than the sum of the lattice energies of the separate co-formers ($PE_{exc} < 0$, Fig. 8). This stabilization comes essentially from switching to more favorable H-bonding patterns, as revealed by the parallel increase of Coulombic attraction and overlap repulsion. Enthusiasm about this apparently decisive finding is mitigated by the difficulty of accounting for the accompanying differences in conformational energies. An estimate of the latter leaves however a good (73%) majority of cases where the co-crystal is more stable than the sum of co-formers. Also, a favorable $PE_{exc}$ need not correlate with a large A-B sublattice energy because the final lattice energy of the co-crystal results from a reshuffling of several interactions also in the A-A and B-B contacts. To provide more information on this crucial issue, workers in the co-crystallization area should be encouraged to carry out very accurate crystal structure determinations for co-crystal and for the separate co-formers as well.

**References and Notes**

(1) Stahly, G. P. *Cryst. Growth Des.* **2009**, 9, 4212.

(2) Braga, D.; Grepioni, F.; Maini, L. *Chem. Commun.*, Feature Article **2010**, 46, 6232.

(3) Gujrati, M. D.; Saleesh Kumar, N. S.; Brown, A. S.; Captain, B.; Wilson, J. N. *Langmuir* **2011**, 27, 6554.

(4) Wishweshwar, P.; McMahon, J. A.; Peterson, M. L.; Hickey, M. B.; Shattock, T. R.; Zaworotko, M. J. *Chem. Commun.* **2005**, 4601.

(5) Musumeci, D.; Hunter, C. A.; Prohens, R.; Scuderia, S.; McCabe, J. F. *Chem. Sci.* **2011**, 2, 883.

(6) Grecu, T.; Hunter, C. A.; Gardiner, E. J.; McCabe, J. F. *Cryst. Growth Des.* **2014**, 14, 165.

(7) Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W. A.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. *J. Chem. Theory Comput.* **2009**, 5, 1432.

(8) Issa, N.; Barnett, S. A.; Mohamed, S.; Braun, D. E.; Copley, R. C. B.; Tocher, D. A.; Price, S. L. *CrystEngComm* 2012, 14, 2454.

(9) Braga, D.; Grepioni, F.; Maini, L.; Polito, M.; Rubini, K.; Chierotti, M. R.; Gobetto, R. *Chem.*

*Eur. J.* **2009**, 15, 1508.

(10) Braga, D.; Maini, L.; de Sanctis, G.; Rubini, K.; Grepioni, F.; Chierotti, M. R.; Gobetto, R. *Chem. Eur. J.*, **2003**, 9, 5538.

(11) Friščić, T.; Jones, W. *Crystal Growth Des.* **2009**, 9, 1621.

(12) Rodríguez, S. B.; Price, C. P.; Jayasankar, A.; Matzger, A. J.; Rodríguez, H. N. *Adv. Drug Delivery Rev.* **2004**, 56, 241.

(13) Braga, D.; d'Agostino, S.; Dichiarante, E.; Maini, L.; Grepioni, F. *Chem. Asian J.* **2011**, 6, 2214.

(14) Braga, D.; Grepioni, F.; Maini, L. *Chem. Commun.* **2010**, 46, 6232.

(15) Allen, F. H. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, 58, 380.

(16) Gavezzotti, A.; Lo Presti, L. *Crystal Growth Des*. **2016**, 16, 2952.

(17) Aitipamula, S.; Banerjee, R.; Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.; Duggirala, N.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetti, R. R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.; Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Guru Row, T. N.; Sanphui, P.; Shan, P.; Shete, G.; Singh, A.; Sun, C. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Thaper, R. K.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Variankaval, N.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J. *Cryst.Growth Des.* **2012**, 12, 2147.

(18) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. *Acta Cryst., B*, **2002**, 58, 389.

(19) 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. *J. Appl. Cryst.* **2008**, 41, 466.

(20) Gavezzotti, A. *CLP User's manual* **2016**. Manual, executables, and source codes are freely available at http://www.angelogavezzotti.it

(21) Gavezzotti, A. *Mol. Phys.* **2008**, 106, 1473.

(22) Gavezzotti, A. *Acta Cryst. B* **2010**, 66, 396.

(23) Pidcock, E. *Chem. Comm.* **2005,** 3457.

(24) Collings, J. C.; Roscoe, K. P.; Robins, E. G.; Batsanov, A. S.; Stimson, L. M.; Howard, J. A. K.; Clark, S. J; Marder, T. B. *New J. Chem.* **2002**, 26, 1740.

(25) Gavezzotti, A. *New J. Chem.*, 2016, in press. doi: 10.1039/C6NJ01087C

(26) Gavezzotti, A. *CrystEngComm*, **2008**, 10, 389.

(27) Gavezzotti, A.; Lo Presti, L. *Cryst Growth Des*. **2015**, 15, 3792.