Evaluating the energetic driving force for co-crystal formation

Christopher R. Taylor and Graeme M. Day*

School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom

E-mail: g.m.day@soton.ac.uk

Abstract

We present a periodic density functional theory (DFT) study of the stability of 350 organic co-crystals relative to their pure single-component structures, the largest study of co-crystals yet performed with high-level computational methods. Our calculations demonstrate that co-crystals are on average 8 kJ mol^{-1} more stable than their constituent single-component structures, and are very rarely ($< 5\%$ of cases) less stable; co-crystallization is almost always a thermodynamically-favorable process. We consider the variation in stability between different categories of systems – hydrogen-bonded, halogen-bonded, and weakly-bound co-crystals – finding that, contrary to chemical intuition, the presence of hydrogen or halogen bond interactions is not necessarily a good predictor of stability. Finally, we investigate the correlation of the relative stability with simple chemical descriptors: changes in packing efficiency and hydrogen bond strength. We find some broad qualitative agreement with chemical intuition – more densely-packed co-crystals with stronger hydrogen bonding tend to be more stable – but the relationship is weak, suggesting that such simple descriptors do not capture the complex balance of interactions driving co-crystallization. Our conclu-

sions suggest that while co-crystallization is often a thermodynamically-favorable process, it remains difficult to formulate general rules to guide synthesis, highlighting the continued importance of high-level computation in predicting and rationalizing such systems.

Introduction

Co-crystals, where two or more neutral molecules crystallize together within a single crystal lattice, continue to receive high levels of interest in the field of crystal engineering. A major driver for co-crystal studies is the increased opportunity for the modification of solid form properties that is introduced by the combination of two different chemical entities, rather than chemical modification of the original molecules. This has been exemplified by the modification of dissolution rate,[1] mechanical properties,[2] optical properties,[3] stability to hydration[4] and melting point[5] through co-crystallization.

Like polymorphism, co-crystallization is a phenomenon that is straightforward in principle, but where the definition of guiding rules and the development of predictive methods is challenging. In the area of polymorphism, the use of computational modelling has helped understand the typical energy difference between observed structures: most pairs of polymorphs are separated by lattice energy differences of less than 2 kJ mol⁻¹, and 95% by less than 7.2 kJ mol⁻¹,[6] and we are gaining a better understanding of magnitude of entropic contributions and their impact on polymorph free energy differences.[6,7]

We are not aware of a general guideline for the difference in energy between a typical co-crystal and the pure crystal structures of its individual components. This is despite the utility of such a characteristic value; most obviously, it would indicate whether co-crystallization in general is likely to be spontaneous (i.e. thermodynamically favorable with respect to formation of the separate, pure structures). A better quantitative understanding of the typical energetic driving force for co-crystallization also provides context for the introduction or modification of intermolecular interactions in attempts to induce the co-crystallization of

two molecules.

Computational chemists have developed a range of methods for predicting co-crystallization. These include the use of simplified models for the possible interactions between molecules,[8] specific criteria based on electrostatic potentials between donor and acceptor atoms on co-formers,[9] models trained from crystallographic databases for predicting the propensity of competing hydrogen bond donor-acceptor combinations,[10] the adaptation of liquid-phase models to predict mixing enthalpies between components[11] and, recently, the use of machine learning based on molecular descriptors.[12] Furthermore, the methods developed for crystal structure prediction (CSP) have been applied to co-crystal prediction by exploring the complete space of crystal packing possibilities of the single components and their co-crystals and an evaluation of the possible energetic gain of co-crystallization.[13–15] The possibilities for different stoichiometric ratios of co-crystal components must be considered in predicting co-crystallization; in the case of CSP, this involves predicting the landscapes of possible structures at a range of stoichimetries.[13,16] From a computational CSP perspective, a quantification of the characteristic energy of observed co-crystals with respect to their separate single components would help define the relevant energy range within which a CSP study should explore possible co-crystal structures.

Discussions of whether a co-crystal is thermodynamically more or less stable than the corresponding pure structures has largely been limited to studies of specific sets of co-crystal systems. Computationally, these typically involve a small closely-related set of target molecules[17,18] rather than a broad, chemically-diverse space, making the conclusions drawn from such studies useful for those specific systems but difficult to apply to co-crystallization generally. A recent effort to investigate co-crystallization as a general phenomenon by Gavezzotti *et al.*[19] considered a large set of around 1500 co-crystal structures, but only calculated energies relative to the single-component structures for a small fraction (97) of those, using the efficient but physically-approximate PIXEL method.[20] That being said, both their work and the work of others on co-crystals of specific composition indicate that co-crystallization is

likely a thermodynamically-driven process; observed co-crystals are usually found to be more stable than the pure structures, but more general and quantitative conclusions regarding the magnitude of this stability cannot be reached in these studies.

The aim of the present work is to accurately compute the thermodynamic stabilities of a large set of experimentally-known co-crystals with respect to the pure structures of the molecules concerned. Ideally, in order to reach generally-applicable conclusions, the set of structures must be both sufficiently large and chemical diverse to make generally representative conclusions. While the relative stabilities will likely show considerable variation depending on the physical nature of each system, it is our hypothesis that co-crystallization is in general thermodynamically driven to a degree that is resolvable with current computational methods.

We also explore whether the calculated co-crystal stability could be linked with a straightforward structural descriptor. We consider several relatively simple descriptors of both the crystal structures and the molecules within them. In particular, we calculate and compare descriptors for two typical strong, directional intermolecular interactions – hydrogen bonding and halogen bonding – as well as the packing efficiency of the structures. We assess whether the change in these descriptors between the pure structures and the co-crystal can be linked to the observed stability of the co-crystal relative to its single component structures.

Considering a hypothetical binary co-crystal with composition $A_m B_n$, we hereafter refer to the constituent molecular species A and B as the *components* of the co-crystal. Hence, we term crystal structures of purely constituent A or B as the corresponding *single-component structures* of the co-crystal. We define the relative thermodynamic stability of the co-crystal as

$$\Delta E_{\text{co-cryst}}(A_m B_n) = E_{\text{tot}}(A_m B_n) - \left[m E_{\text{tot}}(A) + n E_{\text{tot}}(B) \right] \tag{1}$$

where E_{tot} is the total energy of a crystal structure per formula unit of constituent(s): $E_{\text{tot}}(A_m B_n)$ is the calculated energy of the co-crystal; $E_{\text{tot}}(A)$ and $E_{\text{tot}}(B)$ are energies of the single-component crystal structures of co-formers A and B, respectively. Where poly-

4

morphism has been observed in the single-component structures, we present the co-crystal stability relative to the single-component structures with the lowest total calculated energy per formula unit.

The cost-effective energy evaluation and structural optimization of molecular crystal structures is itself a difficult problem and an area of considerable research. Here, we apply periodic density functional theory (DFT) calculations, performed in a plane wave basis and supplemented by a dispersion correction (DFT+D). Thus, the energy differences calculated using equation 1 account for both intermolecular energy differences and changes in intramolecular energy between the single component and co-crystal structures. The best DFT+D methods have been demonstrated to obtain values correct to within a few kJ mol[-1] relative to experimentally-derived lattice energies for a molecular crystal benchmark set of 23 organic crystal structures.[21] While such periodic DFT calculations are quite computationally demanding compared to force field based methods, they remain tractable for large sets of structures using large scale high-performance computing.

Method and Computational Details

General co-crystal structure set

Our aim was to select a representative set of organic co-crystal structures, from which general observations on the energetics of co-crystallization can be made. The set of co-crystals and their corresponding single-component structures were obtained from the Cambridge Structural Database (CSD)[22] using the ConQuest software package. For the general co-crystal set, constituent elements were limited to C, H, N, O, F, S, and Cl. We define co-crystals for the purposes of this work as crystal structures:

- containing at least two chemically different polyatomic units,

- containing only overall neutral molecules (i.e. non-ionic structures), and

- not containing any of a set (defined in the Supplementary Information) of common solvents or small molecules known to be liquid/gaseous at room temperature, ruling out solvates and clathrates.

To avoid duplicate and low-quality structures, all searches were initially restricted to the pre-defined "best hydrogens" subset[23] of the CSD. Matching of molecules between co-crystals and possible single component structures was handled using SMILES strings. Further details of our search procedure are provided in the Supplementary Information. For brevity, we will refer to structures in the text by their CSD reference codes.

Halogen-bonded co-crystal set

The general set of co-crystal structures was supplemented by a specific search performed to identify halogen-bonded co-crystals and their corresponding single-component structures. We identified all co-crystals containing at least one possible halogen bond, which we defined as a contact D$\cdots$X—A, where: D is one of N, O, S, or Cl; X is either Br or I; the D$\cdots$X distance is less than the sum of the van der Waals radii of the two atoms plus a tolerance of 0.1 Å and the angle formed by D$\cdots$X—A is greater than 160°.

This process, combined with the additional restriction that crystal structures were also available for the individual components of each co-crystal, yielded 3303 unique co-crystal structures matching the general co-crystal criteria, and 34 unique co-crystal structures for the halogen-bonded set. Of the latter, only 28 successfully completed our optimization procedure – we then randomly selected co-crystals from the general set until 322 (and their corresponding single component structures) were successfully geometry optimized. This gave a total combined set of 350 co-crystals for our computational study. The assembled set of single component structures include all crystal polymorphs available within the CSD of each molecule found in the co-crystal set.

While our set excludes any structures that are reported as salts in the CSD, such an unambiguous definition is not always possible due to dynamic proton transfer or disorder.

6

Indeed, the experimental determination of co-crystal versus salt can be sensitive to conditions, as described for the co-crystal target of the most recent (sixth) Blind Test of CSP methods.[24] Since our optimization procedure allows all structural degrees of freedom to relax, it is feasible for proton transfer between species to occur if the co-crystal configuration is unstable at the level of theory used in our calculations, potentially changing structures defined in the CSD as being neutral co-crystals into ionic salts.

Periodic DFT optimizations

All experimental co-crystal and single-component structures in our sets were optimized using plane-wave-based periodic DFT using the CASTEP[25] and VASP[26–29] packages. Structural optimization was performed in two steps to aid convergence – an initial optimization in which only atomic positions were optimized (i.e. the lattice parameters are fixed) followed by a second optimization in which both atomic positions and cell parameters are fully flexible. Such a procedure has been noted to improve convergence of DFT optimizations of experimentally-obtained organic crystal structures.[30]

Throughout our calculations, the PBE exchange-correlation functional[31] was employed, supplemented with the Becke-Johnson-damped Grimme dispersion corrections[32,33] – the D2 version in CASTEP and the more recent D3 version in VASP. Except where otherwise noted, we present only the final resulting structures and energies from our VASP calculations (PBE+D3) due to the more accurate dispersion correction.

All calculations in VASP made use of the projector-augmented wave (PAW) method[34] and the standard supplied pseudopotentials.[35] The convergence criteria for all VASP calculations were as follows: a plane-wave energy cut-off of 500 eV, an energy tolerance in convergence of the electronic minimization of 1×10^{-7} eV per atom, and a force tolerance in the geometry optimization of 3×10^{-2} eV Å^{-1}. These tolerances were found to yield sufficiently accurate energies (and converged minimizations) while moderating the considerable total cost of the many optimizations to be performed. The computational settings used are also in line with

those used in other work employing DFT in VASP for accurate energy evaluations of organic crystal structures.[17,30,36,37]

Further details regarding our optimization procedure are given in the SI.

Calculation of descriptors

Hydrogen bonding

Analysis of the hydrogen-bonding environments in crystal structures was carried out using in-house code in conjunction with the CSD Python API.[38] A molecular shell was constructed around each unique species in a crystal structure to count all hydrogen bonds in which the species participates (i.e. ignoring crystal symmetry). In cases of chemically-equivalent but symmetrically-unrelated moieties (e.g. a $Z' = 2$ single-component crystal), hydrogen bond counts were averaged over the distinct moieties. Contacts were considered hydrogen bonds if they satisfied all of the following criteria:

- the H$\cdots$A contact distance is less than the sum of the van der Waals' radii of the two atoms,

- the D—H$\cdots$A contact angle is greater than or equal to 130 degrees,

- the donor atom D is any of N, O, or S, and

- the acceptor atom A is any of N, O, or S.

Additionally, we calculated a basic measure of the strength of a hydrogen bond based on the Coulombic attraction between the partial charges of the (positively charged) hydrogen and (negatively charged) acceptor atom involved (the H$\cdots$A electrostatic interaction). We computed these charges (q_H and q_A respectively) based on fitting to the molecular electrostatic potential *in vacuo* of each unique moiety in the optimized crystal structure using the CHELPG method[39] in Gaussian09.[40] The strength of an individual hydrogen bond is then straightforwardly calculated as:

$$E_{\text{H-bond}}^{\text{H-A,electr}} = \frac{1}{4\pi\varepsilon_0} \frac{q_{\text{H}}q_{\text{A}}}{r_{\text{HA}}} \tag{2}$$

where r_{HA} is the hydrogen-acceptor distance. A total hydrogen-bond attraction energy was then calculated for a given crystal structure C as:

$$E_{\text{H-bond}}^{\text{H-A,electr}}(\text{C}) = \sum_i^{\text{H-bonds}} E_{\text{H-bond},i}^{\text{H-A,electr}} \tag{3}$$

where the sum runs over all hydrogen bonds present in the unit-cell.

Finally, the change in this measure of hyrogen bond strengh upon forming the co-crystal can be calculated in analogy with equation 1:

$$\Delta E_{\text{H-bond}}^{\text{H-A,electr}}(\text{A}_m\text{B}_n) = E_{\text{H-bond}}^{\text{H-A,electr}}(\text{A}_m\text{B}_n) - \left[m E_{\text{H-bond}}^{\text{H-A,electr}}(\text{A}) + n E_{\text{H-bond}}^{\text{H-A,electr}}(\text{B}) \right]. \tag{4}$$

Packing coefficient

To determine the efficiency of packing of a given crystal structure, we use the packing coefficient C_K as defined by Kitaigorodsky.[41] Each molecule is assumed to exclude a volume V_{mol} based on the van der Waals' radii of its constituent atoms. The packing coefficient for a crystal structure with unit cell volume V_{cell} is therefore

$$C_K = \frac{1}{V_{\text{cell}}} \sum_i^{\text{molecules}} V_{\text{mol},i} \tag{5}$$

and this value was calculated for each of our optimized crystal structures using the PLATON package,[42] with a spherical probe of radius 1.2 Å, and an integration grid of 0.1 Å spacing.

Analogously to equation 1, the change in packing coefficient upon forming the co-crystal can be defined as the difference between C_K in the co-crystal and the stoichiometrically-normalised sum of C_K values of the single components:

$$\Delta C_K(\mathrm{A}_m\mathrm{B}_n) = C_K(\mathrm{A}_m\mathrm{B}_n) - \frac{1}{m+n}\Big[mC_K(\mathrm{A}) + nC_K(\mathrm{B})\Big]. \tag{6}$$

Results and Discussion

Distribution of the energetic stability of co-crystals

We first analyze the distribution of co-crystal stabilities in the overall co-crystal set, comprising 350 co-crystal structures, relative to the single-component structures of their constituent co-formers (Figure 1). The results are presented in two ways: as an energy gain per co-crystal formula unit (blue bars) and as an energy gain per molecule (orange bars, the energy gain per formula unit divided by the number of molecules in the formula unit).

It is clear from Fig. 1 that the calculations generally and reliably predicts observed co-crystals to be more stable than their corresponding single-component structures. The mean value of the data is -8.0 kJ per mole of molecules or -19.1 kJ per mole of co-crystal formula unit – the average gain in stability achieved by co-crystallization is greater than the typical energy difference between crystal polymorphs of organic molecules (around 2 kJ mol^{-1}).[6] The median stability is -6.7 kJ mol^{-1} of molecules (-16.2 kJ mol^{-1} of formula unit), and the distribution is skewed towards increasing co-crystal stability.

Note the magnitude of the energy per formula unit depends on the definition of the latter (obtained from the co-crystal's CSD entry), hence the obvious outlier value at -129 kJ mol^{-1}; this co-crystal, CSF refcode RAWFAW, has an unusually large formula unit containing 7 molecules (tris(piperazine) tetrakis(1,1,1-tris(4-hydroxyphenyl)ethane)). Thus, the stability per molecule of RAWFAW relative to the single component structures is a much less extreme value: -18.4 kJ mol^{-1}.

Of the 350 co-crystals considered in Figure 1, only 18 (5.1%) are found to be less stable than the corresponding single-component structures – i.e. our level of theory indicates that 95% of experimentally-observed co-crystals in the CSD are thermodynamically favorable.
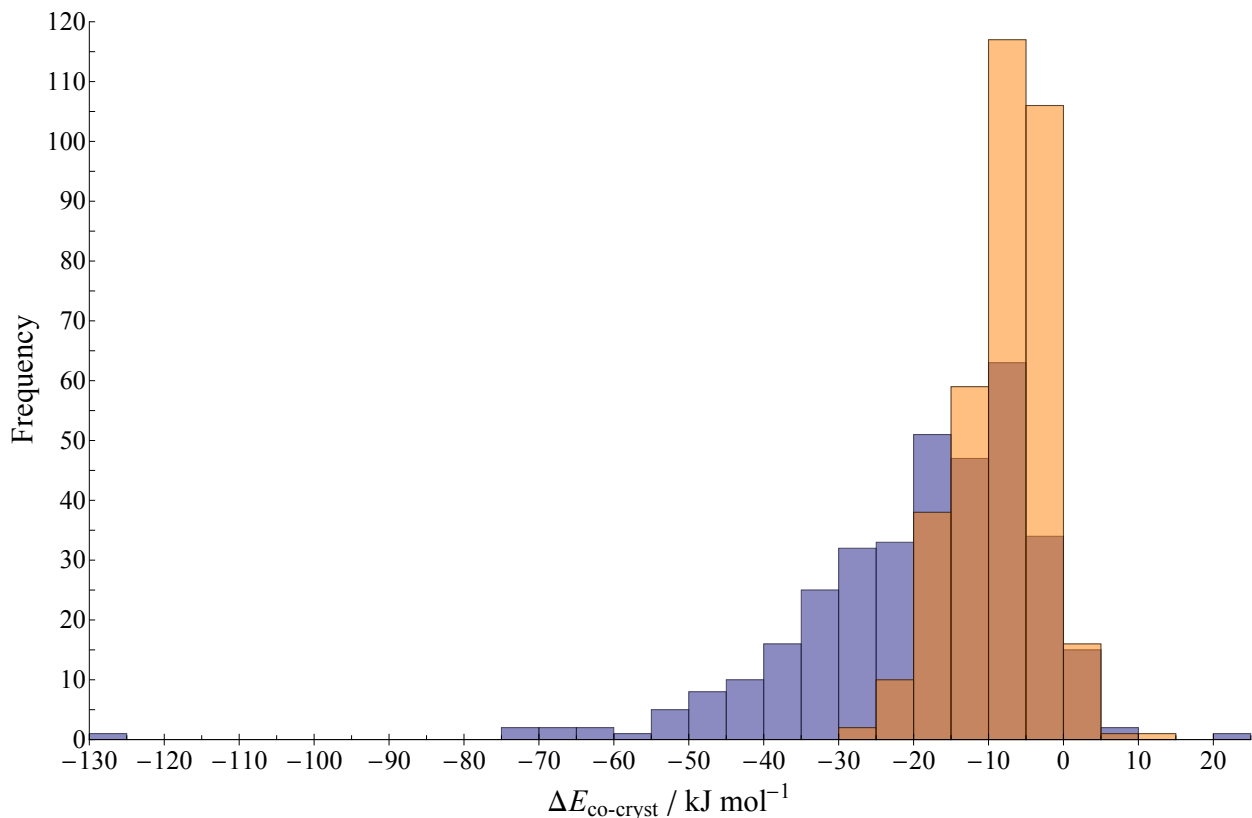
Figure 1: Distributions of calculated (PBE+D3) relative stabilities (defined in Equation 1) for the full set of 350 co-crystals relative to their corresponding simple-component structures. Blue bars are the energy per formula unit of the co-crystal (as defined within its CSD entry), while orange bars are the energy per molecule (the dark regions of bars indicate overlap). Values are expressed in kJ mol^{-1} and collected into bins of width 5 kJ mol^{-1}.

Of those that are found to be less stable, the instability never exceeds $+10.2$ kJ mol^{-1} of molecules ($+20.5$ kJ mol^{-1} of formula unit); the range of positive (in)stabilities is considerably smaller than the range of negative stabilities.

The mean relative stability per formula unit of co-crystal in our dataset, -19.1 kJ mol^{-1}, is in reasonably good agreement with the work of Chan *et al.*,[17] who obtained a value of -11.5 kJ mol^{-1} also using periodic DFT calculations in VASP (with the PW91 functional and a custom dispersion correction). Chan *et al.* considered only co-crystals of a few select molecules, so the distribution of co-crystal stabilities that we compute is likely to be more generally representative of organic co-crystals. We also agree extremely well with Chan *et al.* concerning the proportion of co-crystal structures that are predicted to be more stable

than the single-component structures; 94.9% in both cases (though their figure also includes several structures that were known *a priori* to be salts).

Other efforts to compute relative stabilities of specific sets of co-crystals have used the force field and electrostatic models common in CSP, with varying success. Issa *et al.*[18] used a combination of electrostatic multipoles (derived from MP2 charge densities) and "exp-6" repulsion-dispersion potentials to compute relative stabilities in DMAREL[43] for a set of 26 co-crystals of three molecules of interest with assorted co-formers. This approach yielded stabilities of similar magnitude to our work in cases where co-crystallization was predicted to be favorable; however, it did not consistently predict co-crystals to be more stable, and at times strongly disfavored the co-crystal.

It is difficult to fully quantify the uncertainty in calculated relative stabilities. Based on a benchmark comparing to lattice energies derived from measured sublimation enthalpies, the PBE+D3 energy model achieves a mean absolute error of 4.5 kJ mol^{-1} with a random error (one standard deviation) of 5.6 kJ mol^{-1}.[37] This is a relatively small error compared to magnitude of the co-crystal stability for the majority of systems; approximately one third of the co-crystals studied have a stability relative to the single component structures within about 5 kJ mol^{-1} of zero - this is the proportion of co-crystal structures where the uncertainty in the energy calculation makes it unclear whether they are truly relative to their pure components. However, given that our relative stabilities are differences in energies between crystal structures, we expect that fortuitous cancellation of error might reduce the size of this uncertainty, particular with regards to systematic errors in the PBE+D3 description of the intramolecular physics. It has also been shown[6] that within an empirical force-field and electrostatic multipole description of organic crystals, the contribution of vibrational energy terms to free energy differences between polymorphs rarely exceeds 2 kJ mol^{-1}. Errors of this magnitude, due to us not including vibrational energies, are small relative to the static lattice energy differences.

We therefore suggest that while we are unable to fully quantify the uncertainty in our

calculated stabilities, we remain confident that the distribution of stabilities is broadly correct and certainly that the qualitative findings – co-crystallization is very often energetically favorable, and very rarely significantly unfavorable – are reliable.

Distribution subsets: the influence of hydrogen and halogen bonding

Given our specific inclusion of the halogen-bonding co-crystal subset and the large number of more general co-crystals considered, we can categorize our systems into three different types of co-crystal; hydrogen-bonded, halogen-bonded, and those without one of these specific strong, directional interactions, which for convenience we will refer to as "weakly-bound".

Hydrogen-bonded co-crystals

The vast majority (82%) of co-crystals in our set exhibit hydrogen bonding, despite our making no particular efforts to favor such systems. This likely reflects the conventional synthetic routes to the formation of co-crystals being constructed around the complementary pairing of hydrogen bond donor and acceptor species. The distribution of hydrogen-bonded relative stabilities seen in Figure 2a is therefore unsurprisingly similar to our overall distribution in Figure 1. The range in $\Delta E_{\text{co-cryst}}$ in this set is from -24.6 kJ mol^{-1} (the most stable co-crystal, relative to the corresponding single-component structures) to $+10.2$ kJ mol^{-1} (the least stable).

The most stable of our hydrogen-bonded co-crystals is RADHEL, a 1:2 co-crystal of 1,4-di[bis(4'-hydroxyphenyl)methyl]benzene (RADGUA) and 1,4-diazanaphthalene (HEYJOK01). Two additional hydrogen bonds are formed per molecule in the co-crystal, as the diazanaphthalene molecule cannot form hydrogen bonds on its own.

The least stable hydrogen-bonded co-crystal, which is also the least-stable co-crystal overall, is CSD refcode AWAJOX02 - a co-crystal of 2,2'-diethoxy-α-truxilic acid and trans-2-ethoxycinnamic acid. This particular case seems to be due to conformational strain in one
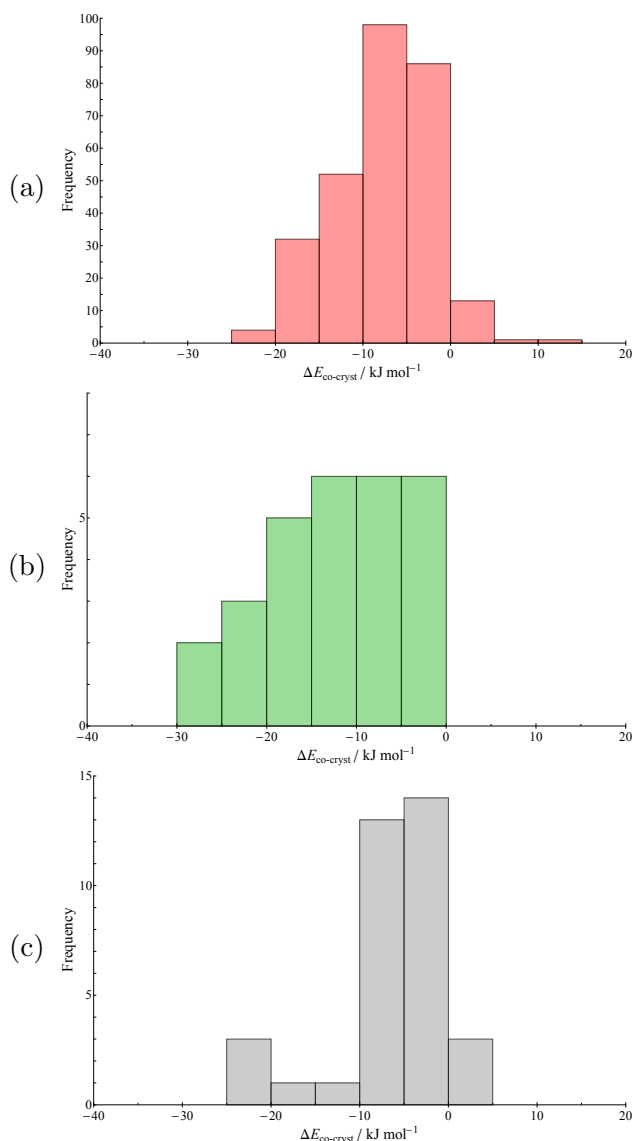
Figure 2: PBE+D3 relative stabilities of our test set of 350 co-crystals, broken down into subsets of (a) hydrogen-bonded (red), (b) halogen-bonded (green) and (c) "weakly-bound" (grey) systems. Each histogram spans the same energy range on the horizontal axis, but note the different frequency ranges on the vertical axes, due to the different subset sizes.

of the component molecules, as discussed in a later section.

Halogen-bonded co-crystals

As there are fewer structures than in any of the other co-crystal subsets, the energies of the individual structures are also presented as a pair of bar charts in Figure 3.

The most stable halogen-bonded co-crystal is also the most stable of all the co-crystals
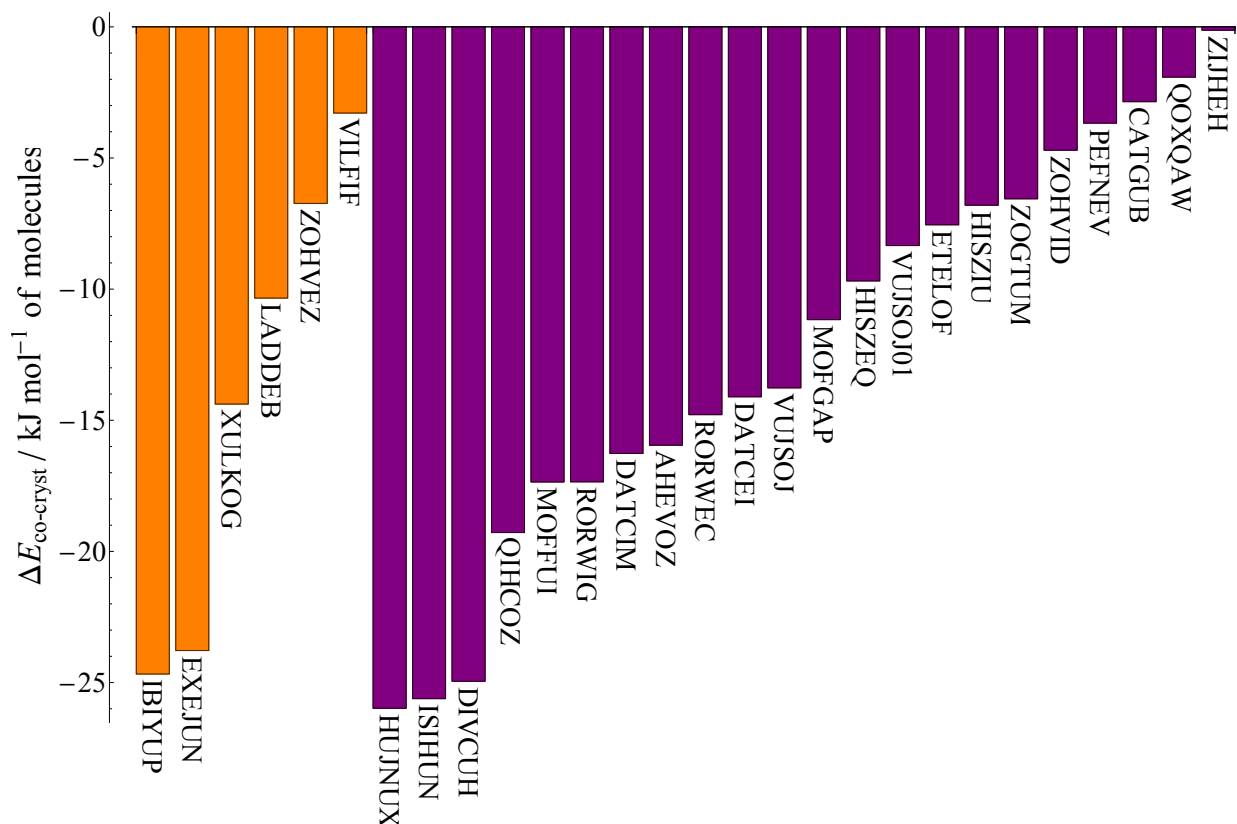
14

Figure 3: PBE+D3 relative stabilities of the subset of halogen-bonded co-crystals. Orange bars indicate structures in which bromine forms the halogen bond, while purple bars indicate iodine. Each structure is labeled by the CSD refcode of the co-crystal.

that we consider in this work; HUJNUX, a 1:1 co-crystal of tetraiodo-*para*-benzoquinone and tetrathiafulvalene (TTF), with a relative stability of -26.0 kJ mol^{-1} (Figure 4a). The co-crystal gains 2 halogen bonds per molecule, though this is also true of a number of other, less-stable co-crystals in our set. As well as the I$\cdots$S halogen bonds that are formed in the co-crystal, the large stability of this co-crystal is most likely related to charge transfer between the stacked, planar donor and acceptor molecules.[44] The next most stable halogen bonded co-crystals (Figure 4b-d) have no other obvious strong intermolecular interactions, so most likely gain their stability through the halogen bonds themselves. These involve highly polarized halogen atoms, such as the iodine in 1,4-diiodo-2,3,5,6-tetrafluorobenzene (ISIHUN, DIVCUH) or the bromine in N-bromosuccinimide (IBIYUP).[45]

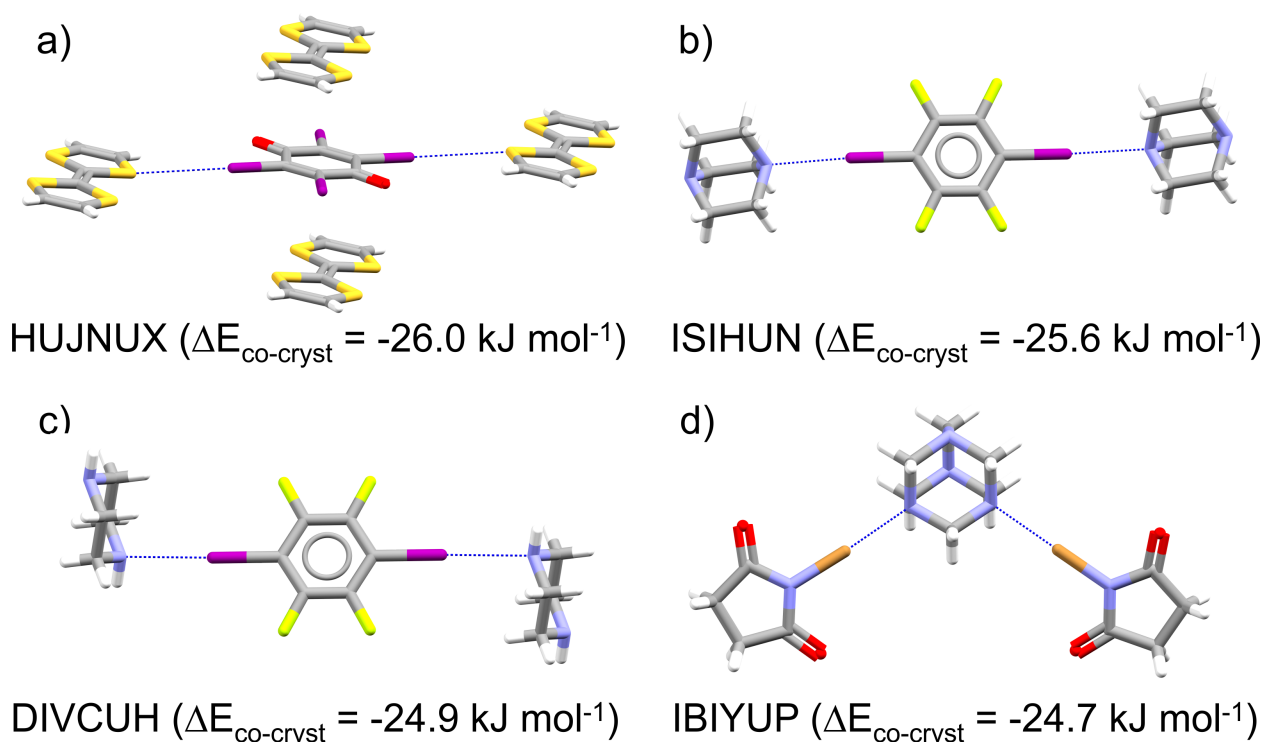Meanwhile, the least stable halogen-bonded co-crystal, ZIJHEH, is a 1:1 system of 3,6-

Figure 4: The four halogen bonded co-crystals found to be most stable with respect to their single component structures: a) HUJNUX, tetraiodo-*para*-benzoquinone : tetrathiafulvalene (TTF), b) ISIHUN, c) DIVCUH and d) IBIYUP. Atoms are colored by element type: grey = carbon; white = hydrogen; blue = nitrogen; red = oxygen, light yellow = fluorine, yellow = sulfur; purple = iodine; brown = bromine. Dashed blue lines indicate halogen bonds. Co-crystal energies are given in kJ mol^{-1} per molecule with respect to the single-component crystals of their components.

diiodotetrachlorobenzene (ZZZAVM02) and 4-methylbenzenesulfinamide (ZIJGOQ). This structure gains only one halogen bond per molecule, among the fewest in our set, and $\Delta E_{\text{co-cryst}}$ is only -0.1 kJ mol^{-1}.

It is not possible to state that either halogen leads to more stable co-crystals than the other – though the bromine-bonded sample is considerably smaller, both samples display similar distributions of $\Delta E_{\text{co-cryst}}$. This is particularly noticeable when considering the ZO-HVEZ/ZOHVID pair, 1:1 co-crystals of 3,4,5-trichlorophenol and *para*-haloaniline featuring halogen$\cdots$Cl interactions, that are chemically identical apart from the halogen atom present (Br in ZOHVEZ and I in ZOHVID). Despite the change in halogen atom, the $\Delta E_{\text{co-cryst}}$ values obtained are -6.7 and -4.7 respectively; the difference between these is small and within

the estimated uncertainty of the energy calculations. The two co-crystal structures are isostructural with very small differences in atomic coordinates, indicating that the considerable increase in size of the I atom relative to the Br atom does not affect the packing of the co-crystal. This is despite the fact that the single-component structures of the haloanilines pack differently. At least in this case, substituting Br with I does not substantially affect the relative stability of the resulting co-crystal.

Weakly-bound co-crystals

Our "weakly-bound" subset of co-crystals (Figure 2c) is perhaps the most visibly distinct from the overall distribution of relative stabilities (Figure 1). A larger proportion of weakly-bound structures have small or slightly-positive relative stability values compared with the overall set or the other subsets. The least-stable weakly-bound structure is IFIMAL, a 2:1 (defined in the CSD as being 1:$\frac{1}{2}$) co-crystal of endo-4a,5,8,8a-tetrahydro-5,8-ethano-1,4-naphthoquinone (HETNQU) and *para*-benzoquinone (BNZQUI).

However, there is also a larger proportion of very stable (25-30 kJ mol^{-1}) co-crystals in the weakly-bound set compared to the other subsets. This is perhaps counterintuitive, as one might expect the lack of strong, directional interactions to reduce the likelihood of the co-crystal being significantly more stable. As we found in the halogen bond set, the most stable co-crystal structure is likely a charge transfer complex. This most stable structure is PINHID, a 1:1 co-crystal of 6,12-dioxa-anthanthrene (or peri-xanthenoxanthene, QUPMUJ10) and tetracyanoquinodimethane (TCNQ, CSD refcode TCYQME), with $\Delta E_{\text{co-cryst}}=-23.6$ kJ mol^{-1} (per molecule).

In fact, many of the more-stable half of the weakly-bound co-crystal subset are found to contain TCNQ, tetracyanoethylene (TCNE, CSD refcode TCYETY), or trinitrobenzene (TNBENZ) (Figure 5), which all have electron-deficient π-systems and are typical acceptor species in charge-transfer systems, paired with planar aromatic donors. All of the low energy co-crystals (up to $\Delta E_{\text{co-cryst}}=-4.2$ kJ mol^{-1}) in this set contain aromatic donor and acceptor
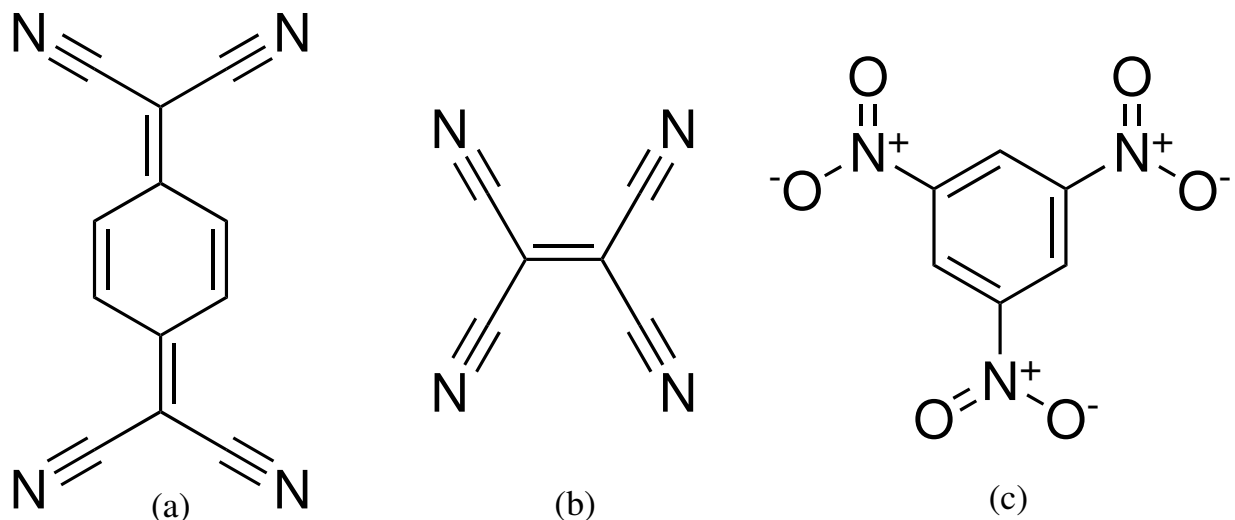
Figure 5: Three examples of common molecules in our weakly-bound co-crystal set: (a) tetracyanoquinodimethane (TCYQME), (b) tetracyanoethylene (TCYETY02), and (c) trinitrobenzene (TNBENZ13). All three species are electron-deficient and commonly used as acceptor species in charge-transfer complexes.

molecules, all with face-to-face packing in their co-crystal structures and the possibility for charge transfer between coformers. Therefore, it is likely that many of the unusually-stable "weakly-bound" systems are in fact charge-transfer co-crystals with strong interactions between donor and acceptor species.

The most stable co-crystal in this subset that lacks face-to-face packing of aromatic donor and acceptor molecules is the 1:1 co-crystal of 4,4'-bipyridine with 1,4-diethynylbenzene (RUXMAZ, $\Delta E_{\text{co-cryst}}=-4.2$ kJ mol^{-1}), which contains C-H$\cdots$N interactions (C-H$\cdots$A interactions were not classed as hydrogen bonds in our definition of co-crystal subsets). As we did not specifically target our search to include or exclude particular types of systems in this case (unlike for the halogen bonds), this indicates a considerable experimental bias in the CSD towards charge-transfer systems among co-crystals that lack hydrogen bonds or halogen bonds. It is possible this is precisely because the lack of other sources of strong, directional interactions in weakly-bound co-crystals discourages attempts at their synthesis.

Comparison of subsets

Table 1 presents a summary of the calculated energies categorized by the type of interactions present. We preface our analysis by highlighting that the considerably-smaller halogen-bonded and weakly-bound sets may be less representative of their overall category than the hydrogen-bonded set.

Table 1: Minimum, mean, and maximum relative stabilities (kJ mol^{-1} per molecule) calculated using PBE+D3 for our subsets of experimental co-crystal structures, categorised by the presence or absence of strong interactions. Also shown is the proportion of structures with relative stabilities greater than zero (indicating instability compared to the single-component structures).

Type	Structures	$\Delta E_{\text{co-cryst}}$ / kJ mol^{-1}			$N(\Delta E_{\text{co-cryst}} > 0)$
		Min	Mean	Max	
Hydrogen-bonded	287	−24.6	−7.7	+10.3	15 (5.2%)
Halogen-bonded	28	−26.0	−12.6	−0.1	0
Weakly-bound	35	−23.6	−6.5	+1.1	3 (8.6%)

Regarding the energetics of the three different categories, we see that there is only a slight difference in the average relative stability between the hydrogen-bonded and weakly-bound subsets, but the average relative stability of halogen-bonded systems appears to be markedly greater (by around 5 kJ mol^{-1}). Given their relative scarcity in the CSD (we obtained approximately 1% as many halogen-bonded co-crystals compared to general co-crystals in our searches), they are overrepresented in our overall set of structures, and their greater average relative stability could therefore bias our overall relative stability values slightly. It is clear that, at least for the structures considered in our searches, halogen-bonding leads, on average, to considerably more stable co-crystals relative to the single component structures.

The difference between the average values for the hydrogen-bonded and weakly-bound sets, on the other hand, is considerably smaller, only 1.2 kJ mol^{-1} in favour of hydrogen-bonded co-crystals. The sign of this difference agrees with chemical intuition that co-crystal synthesis via formation of more complementary (or stronger) hydrogen bonding relative to the single component might lead to more stable co-crystals. In turn, the energetics of

weakly-bound systems might be expected to be less sensitive to their composition due to the unspecificity of the interactions, leading to less strongly-favored co-crystals on average. As discussed above, all of the most stable co-crystals int he "weakly bound" are composed of charge donor-acceptor molecules that probably lead to charge-transfer complexes. Apart from likely charge-transfer complexes, the $\Delta E_{\text{co-cryst}}$ values in the "weakly bound" subset are -4.2 kJ mol^{-1} or smaller.

The minimum values of $\Delta E_{\text{co-cryst}}$ across the three sets show relatively little variation (no more than 1.5 kJ mol^{-1}), suggesting that the greatest possible stability is also not necessarily intrinsic to a particular category, and therefore that the most stable of all co-crystals are not necessarily those with strongly directional, specific interactions. However, the maximum value of $\Delta E_{\text{co-cryst}}$ (i.e. the greatest instability) is considerably larger for the hydrogen-bonded set. The least stable co-crystal is AWAJOX02, with $\Delta E_{\text{co-cryst}} = 10.3$ kJ mol^{-1}; this structure, which is discussed further below, is a result of a photodimerization reaction of *ortho*-ethoxy-*trans*-cinnamic acid, leading to a 1:1 co-crystal of the photodimerization product with unreacted *ortho*-ethoxy-*trans*-cinnamic acid.[46] Given the unusual route to this co-crystal, it is unsurprising that the energy of the co-crystal structure is an outlier in our data set. Apart from AWAJOX02, the next least stable co-crystal is ABUCIP, a co-crystal of the agrochemical thiophanate-ethyl and 2,2'-bipyridine with $\Delta E_{\text{co-cryst}} = 5.8$ kJ mol^{-1}, which is still noticeably higher than any structures in the halogen bonding or "weakly bound" subsets. This higher maximum $\Delta E_{\text{co-cryst}}$ may simply be due to greater sampling of that population, or perhaps a greater systematic error in the description of hydrogen-bonding using the PBE+D3 level of theory (though the latter might be expected to largely cancel out when considering energy differences). Alternatively, it may be that hydrogen bonding does occasionally result in co-crystals that are thermodynamically unfavorable – perhaps due to the hydrogen bonds encouraging nucleation pathways that yield the co-crystal (i.e. perhaps kinetic effects driving structures to higher-energy local minima are more significant in hydrogen-bonded systems).

On the other hand, when looking at the proportion of co-crystals that are unstable in a given subset, this quantity is greatest for weakly-bound co-crystals, which suggests these combinations of systems are more likely to yield less-stable co-crystals, as might be expected from chemical intuition. However, the effect of small sample size is particularly noticeable here – if just one of these structures with $\Delta E_{\text{co-cryst}} > 0$ were excluded, this proportion would move much closer to the hydrogen-bonded set. While our halogen-bonded subset is also quite small, it is a much larger proportion of all halogen-bonded co-crystal systems in the CSD than the other subsets are of their categories. Hence, we can more confidently state that the lack of any significantly unstable halogen-bonded systems indicates that these interactions are particularly useful for synthesis favoring co-crystals.

To improve our understanding and rationalize the distributions we observe, we now consider further specific cases, as well as to attempt to correlate the distributions we obtain with chemical descriptors.

Individual co-crystal energetics of note

Instability caused by molecular strain

The most unstable co-crystal in our set is AWAJOX02, a hydrogen-bonded co-crystal of 2,2'-diethoxy-α-truxilic acid (XOSKEV) and *ortho*-ethoxy-*trans*-cinnamic acid (ZZZNQS05), which our procedure predicts to be 10.3 kJ mol^{-1} higher in energy than the single components. For both component molecules, the single-component structures display molecular geometries that resemble the gas-phase minima that chemical intuition would suggest. However, in the co-crystal, both molecules are noticeably distorted from those geometries – the 2,2'-diethoxy-α-truxilic acid geometry is compressed such that the overall shape is flattened (Figure 6), while *ortho*-ethoxy-*trans*-cinnamic acid's conjugated backbone is noticeably non-planar (approximately 10° deviation, while the molecule in the single-component case is nearly perfectly planar).

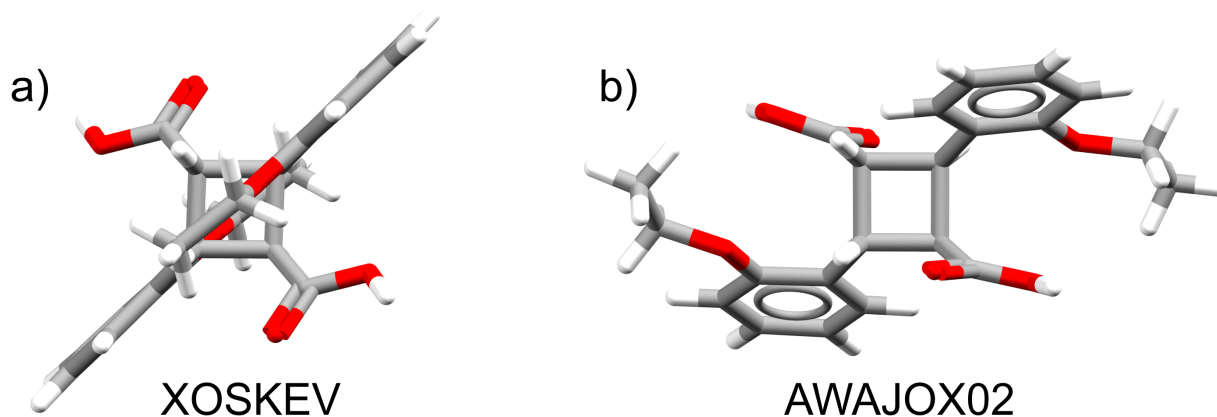Molecular geometry optimizations at a comparable level of theory to our crystal cal-

Figure 6: Molecular conformations of 2,2'-diethoxy-α-truxilic acid in a) its single-component crystal structure (XOSKEV) and b) the co-crystal (AWAJOX02) with *ortho*-ethoxy-*trans*-cinnamic acid .

culations (PBE+D3/6-311G** in Gaussian09[40]) confirm this hypothesis; in the co-crystal, AWAJOX02, the strain energy (the energy required to adopt the in-crystal conformation relative to the nearest local gas-phase minimum) is 48 kJ mol^{-1} for the truxilic acid and 20 kJ mol^{-1} for the cinnamic acid. In comparison, the pure truxilic acid crystal XOSKEV exhibits a molecular strain energy of 37 kJ mol^{-1}, and the pure cinnamic acid ZZZNQS05 a strain energy of 17 kJ mol^{-1} – the single-component structures place the molecules (particularly the truxilic acid) under markedly less strain than the co-crystal.

Note also that these strain energies are relative to the nearest gas-phase minimum to the observed in-crystal conformer – the XOSKEV gas-phase minimum is 6 kJ mol^{-1} higher in energy than that of the same molecule in AWAJOX02, meaning that the formation of the AWAJOX02 in-crystal conformation is further disfavored. Interestingly but perhaps fortuitously, the sum of these strain energies relative to the lowest energy conformer and averaged over the two molecules in AWAJOX02 is about +10 kJ mol^{-1}, i.e. very similar to the overall relative stability we calculate. The unusual energetics of this co-crystal likely relate to its formation as the product of photodimerization of the α' ($Z' = 3$) polymorph of *ortho*-ethoxy-*trans*-cinnamic acid, in which two out of three independent molecules in the asymmetric unit combine to form the product, leaving one molecule unreacted in an

ordered 1:1 co-crystal. Thus, the co-crystal exists as a result of this solid-state reaction, rather than the stabilization gained from combining the two molecules in a crystal. Our observations regarding intramolecular strain are in agreement with the observed build-up of molecular strain as the photodimerization reaction proceeds.[46] The molecules are strained because they are constrained by the packing of the reactant *ortho*-ethoxy-*trans*-cinnamic acid polymorph.

Indications of entropic contributions to small stabilities

Another interesting, specific case is the co-crystal VIGDAR01, of celecoxib (DIBBUL) and nicotinamide (NICOAM), which recent experimental work by Zhang *et al.*[47] has determined to be entropy-stabilized – the difference in enthalpy between the co-crystal and the single-component structures is positive. Being a co-crystal of nicotinamide, this is one of the systems previously investigated by Chan *et al.*,[17] who predicted it to be slightly less stable than the single component structures in advance of the experimental work. (However, Chan *et al.* used the only available structure in the CSD at the time, VIGDAR, a very similar structure but one which they proposed had assignment errors that they corrected[17]).

Our work predicts a slight energetic favorability of -1.9 kJ mol^{-1} for VIGDAR01, contrary to experiment. However, given that the uncertainty in our calculated energies is likely to be on the order of a few kilojoules per mole, we cannot definitively state that a value of this magnitude is indicative in either direction. Indeed, Chan *et al.* predict a similarly marginal value of $+1.5$ kJ mol^{-1} using PW91 with a dispersion correction (after a proposed structural reassignment), and it is known that periodic DFT can produce considerable variation in such small energy differences between crystal structures based on the functional employed (see e.g. qualitative reordering of stability of polymorphs of organic semiconductors[48]).

While co-crystals with such small magnitude (in)stabilities cannot be firmly classified as more or less stable, such small computed values likely indicate cases where the co-crystal is potentially experimentally accessible, due to (for example) entropic effects at finite tem-

perature reversing the relative stability computed at zero temperature. In particular, a small calculated stability of a co-crystal which contains at least one molecule with marked conformational freedom within the crystal structure might indicate potentially entropically-stabilized cases like VIGDAR01. In this way, computing the relative stability can still be a powerful tool for predicting and validating experimental co-crystallization outcomes, even in cases where the co-crystal is not definitively more stable.

Co-crystal packing efficiency

We first consider as a descriptor the change in the packing efficiency (defined in Equation 6) upon forming the co-crystal. Given that, according to our results in Figure 1 and Table 1, co-crystals are usually thermodynamically more stable than their single-component counterparts, we consider the possibility that this is purely down to improved efficiency of packing in the co-crystal (and therefore improved non-specific interactions in general).

One significant finding is that the majority of co-crystals (245 co-crystals, 70% of the set) pack *less* efficiently than their single-component counterparts – evidently the energetic drivers of co-crystallization are more complex than simply improved packing in the co-crystal. One interpretation of this observation is that the directionality of the strong, specific inter-actions (such as hydrogen or halogen bonds) used to form co-crystals forces their structures to sacrifice close packing to improve the geometry of these interactions.

Unfortunately, our calculations reveal no clear correlation between relative stability and the change in packing efficiency. The overall set displays virtually no relationship; examining the subsets, only the halogen-bonded subset displays even a qualitative relationship, which in this case is negative – less densely-packed co-crystals tend to have greater relative stabilities. Even this, however, is subject to a number of exceptions.

This is perhaps unsurprising, as it has been observed that the packing efficiency alone is a very poor predictor for many physical properties of crystals.[49] It can clearly be said that co-crystallization occurs despite generally resulting in less favourable packing, but no significant
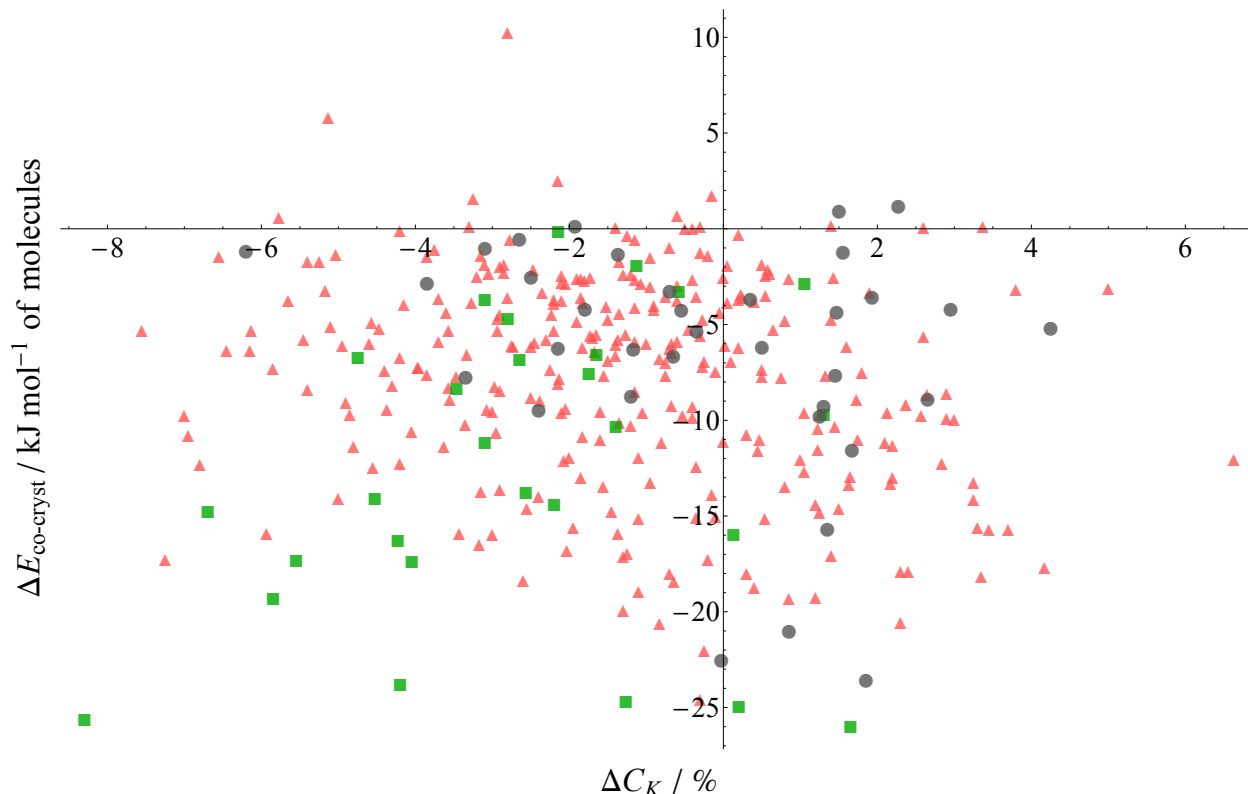
Figure 7: PBE+D3 relative stabilities of co-crystals plotted against the change in packing coefficient upon forming the co-crystal. A negative value of ΔC_K indicates less efficient packing in the co-crystal. Red triangles are from our hydrogen-bonded set, green squares are halogen-bonded, and grey circles are weakly-bound co-crystals.

trend, either overall or within the subsets of interaction types, can be seen. Evidently, packing information alone is insufficient to rationalize the relative stabilities we obtain.

Hydrogen-bonding analysis

As discussed, chemical intuition often leads to synthetic routes to co-crystals that attempt to either create new hydrogen bonds that are not possible in the single-component structures, or to improve the quality of the hydrogen bonds that do form. However, our results (Figure 1 and Table 1) indicate that the presence alone of hydrogen bonding is not sufficient to yield more-stable co-crystals compared to the other sets of co-crystals. We now consider whether the change in hydrogen-bonding in the co-crystal can be correlated with its stability – in the first instance by simply considering the number of hydrogen bonds gained or lost, and in the

second by quantifying the quality of the hydrogen bonds formed. The presence of hydrogen bonding is not confined to our "hydrogen-bonding" subset, however; 11 of the co-crystals in our halogen-bonded subset also display hydrogen bonding as well. Hence, we include these structures in the following analysis.

It was found in the course of analyzing the hydrogen bonding in our structures that 6 of our co-crystals underwent proton transfer during the periodic DFT optimizations. All 6 structures feature O—H$\cdots$N hydrogen bonds, along which coordinate the hydrogen atom transfers to protonate the N atom during the optimization. Evidently, at this level of theory, the experimental structures as presented in the CSD were unstable with respect to proton transfer between the species, resulting in these structures forming salts. It would, of course, be interesting to revisit these structures experimentally to verify the proton positions and monitor whether proton tranfer occurs at low temperature (recalling that our calculations are temperature-free), as well as examining the sensitivity of proton transfer to the computational method (e.g. DFT functional). These structures are listed in the Supplementary Information; their energies have been retained in the data in Figure 1 and Table 1, but they are omitted from the hydrogen-bond analysis.

Simple hydrogen-bond count

Presented in Figure 8 is a plot of the computed relative stability of each co-crystal against the overall change in the average number of hydrogen bonds per molecule. For obvious reasons, here we only include structures within the hydrogen-bonded subset plus any from the halogen-bonded subset which also feature hydrogen bonding. Immediately apparent is the considerable majority (246, 84.3% of the set) of co-crystals which have exactly zero net change in the number of hydrogen bonds present compared to the single-component structures. Those co-crystals which do change the number of hydrogen bonds per molecule are nearly five times more likely to gain than lose them – 13.0% of all structures considered compared to 2.7%, respectively – as might be expected if synthetic procedures are guided
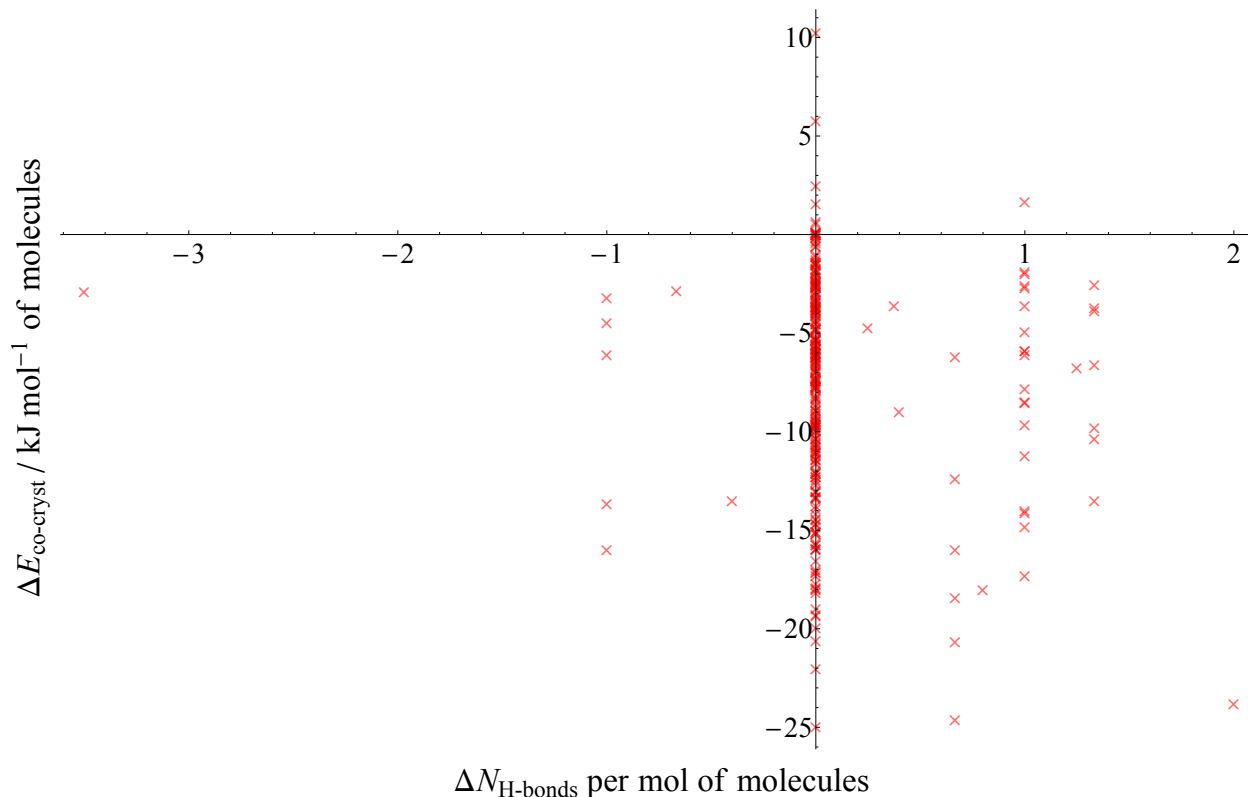
Figure 8: PBE+D3 relative stabilities of co-crystals plotted against the overall change in the number of hydrogen bonds per molecule upon forming the co-crystal. Positive values indicate more hydrogen bonds in the co-crystal. Only the 292 co-crystals from our set in which hydrogen bonding occurs are plotted here; 281 from the hydrogen-bonded subset ($287 - 6$ that underwent proton transfer) and 11 additional structures in the halogen-bonded subset.

by complementary pairing of donor and acceptor molecules.

More surprisingly, Figure 8 indicates that, even in the minority of cases in which the hydrogen bond count does change, there is no significant correlation with the computed relative stability. The overall picture is that an increased number of hydrogen bonds per molecule has no systematic effect on the relative stability of the co-crystal. Indeed, there is a slight trend for structures which do gain in hydrogen bonds to be *less* stable on average as the number of hydrogen bonds increases above 0.5 per molecule, though this is unlikely to be significant. Interesting, both the most extreme gain ($+2$) and the most extreme loss ($-3\frac{1}{2}$) in hydrogen bonds occur in the halogen-bonded set.

Examining the systems that behave contrary to the presumed relationship that more hy-

drogen bonds leads to greater stability does not reveal anything unusual. Consider VEKKEB, a 2:1 co-crystal of acetophenone (ACETPH) and trans-9,10-dihydroxy-9,10-diphenyl-9,10-dihydroanthracene (DIKCOP01), and one of the co-crystals with the largest gain in hydrogen bonds per molecule ($1\frac{1}{3}$ on average). Yet the relative stability of VEKKEB is only -2.5 kJ mol^{-1}, well below the average value. Its single-component structures exhibit no hydrogen bonding, and both molecules are reasonably rigid, so there are no obvious indicators that either of the single-component structures are unusually stable and would therefore lead to only a marginally-stable co-crystal. On the other hand, CEKKOU, a 1:1 co-crystal of benzene-1,2,3-triol (PYRGAL02) and 1,10-phenanthroline (OPENAN), has a considerable and above-average relative stability of -15.9 kJ mol^{-1} despite losing an average of one hydrogen bond per molecule upon forming the co-crystal. While OPENAN cannot form hydrogen bonds in the pure crystal due to only possessing acceptor atoms, PYRGAL02 has three hydroxyl groups and forms six hydrogen bonds per molecule in the crystal, which one might assume would make it unusually stable and thus disfavour formation of the co-crystal CEKKOU, the opposite of what we calculate.

Contrary to chemical intuition, it appears that the simple change in hydrogen bond count does not correlate with the stability of the resulting co-crystal. One weakness of this descriptor is that it considers all H-bonds equally – no chemical or geometric information is involved (apart from the criteria used to define what contacts are H-bonds in the first place). Hence, the H-bonds are binary – they either exist or do not – and there is no measure of their quality or strength. Clearly, a descriptor which very often shows a change of exactly zero between co-crystal and single components is of minimal benefit.

Weighting by hydrogen-bond strength

A more physically-meaningful descriptor instead considers the change in both the number and quality of hydrogen bonds, using the electrostatic H-bond attraction measure we define in Equation 4. While this now requires specific geometric information from the crystal

structures, it allows us to quantify the strength of a hydrogen bond and therefore numerically describe the change in strength of H-bonding between structures. Figure 9 shows the co-crystal relative stability as a function of the total change in the hydrogen bond electrostatic attraction upon forming the co-crystal. Note that the magnitudes of the hydrogen bond energies are large because we have not included the repulsion energy that typically cancels much of the electrostatic stabilization of hydrogen bonds.
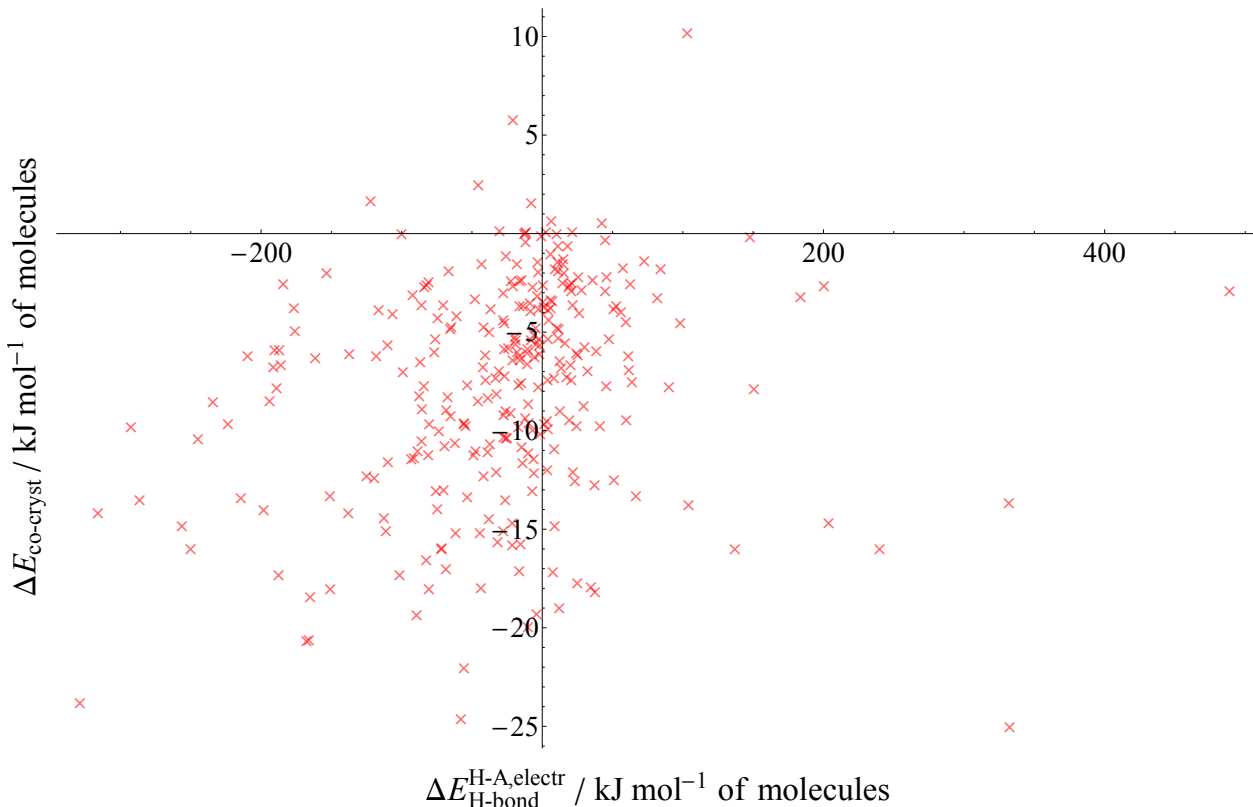
Figure 9: Relative stabilities of co-crystals as a function of the change in the H-bond strength measure defined in Equation 4. Note that this measure considers only the Coulombic attraction between H and A, hence the magnitude of the values along the x-axis.

Although the correlation between this descriptor and the relative stability is very weak ($R^2 = 0.05$), there is a statistically-significant proportional relationship (P-value $= 4 \times 10^{-4}$, relative to a random normal distribution of points). Broadly speaking, a gain in overall hydrogen bonding strength usually leads to a gain in co-crystal stability, though the magnitude of the latter is very weakly correlated with the gain in strength.

As an attempt to incorporate greater physical accuracy into our hydrogen bond strength measure, we also re-calculated the changes in hydrogen bond strength including the acceptor-donor Coulombic interaction as well. The strength measure therefore considers a system akin to a model dipole, with the acceptor Coulombically interacting with both the hydrogen atom and the donor atom. Unfortunately, this model not only failed to improve the relationship observed between the change in strength of the hydrogen bond and the co-crystal, it qualitatively altered the character of the hydrogen bond measure (around 20% of all our crystal structures had a net positive, i.e. repulsive hydrogen bond energy when the $D \cdots A$ interaction was included). We present this data in the Supplementary Information, but here simply state that this attempt to improve our energetic description of the hydrogen bonding was unsuccessful.

Halogen-bond counts

None of the corresponding single-component structures of the halogen bond co-crystal subset exhibit halogen bonding, so when analyzing the effect of halogen bonding on the co-crystal stability, we may count those in the co-crystals alone. Figure 10 presents the relative stabilities of the halogen-bonded co-crystals as a function of the number of halogen bonds gained per molecule in the co-crystal structure.

The bromine-bonded systems exhibit no particular relationship between their relative stabilities and bromine bond counts; there are only 6 structures amd it is not possible to discern a meaningful trend. However, the iodine-bonded systems, while displaying considerable variation, do exhibit a noticeable trend towards increasing stability with increasing iodine bond count. The distributions for each value of the bond count $(1, 1\frac{1}{3},$ and 2 per molecule) shift towards increasing stability as the bond count increases – the minimum, maximum, and average values all become more negative. (Note that the non-integer change in halogen bonds of $1\frac{1}{3}$ results from cases where a system of 2:1 stoichiometry, such as the co-crystal LADDEB, gains one halogen bond to the first molecule and two to the second, giving a total
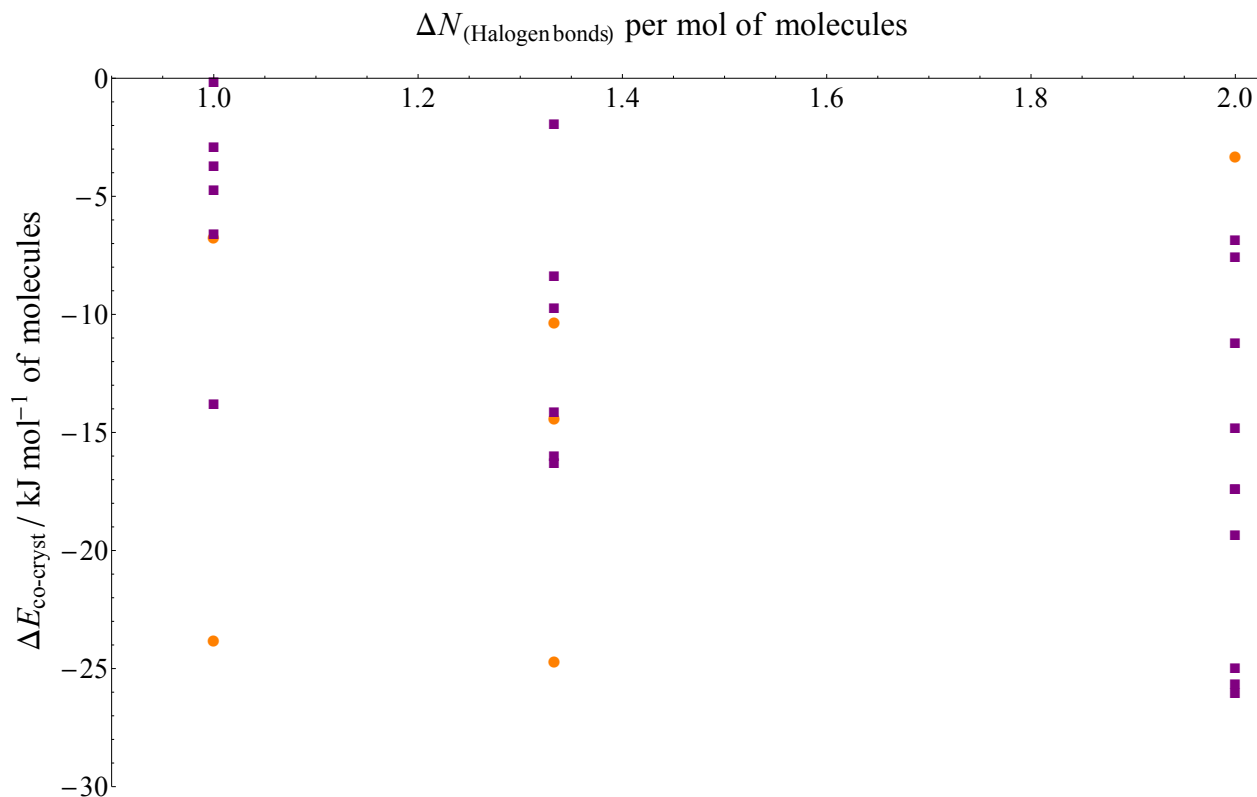
Figure 10: PBE+D3 relative stabilities of halogen-bonded co-crystals plotted against the gain in halogen bonds per molecule. Orange circles indicate structures in which bromine forms the halogen bond, while purple squares indicate iodine.

of 4 halogen bonds averaged over 3 molecules.)

The reasons for this relationship applying only to the iodine-bonded cases are uncertain, though it is possible that iodine bonds are sufficiently strong compared to bromine bonds that they make up a larger proportion of the change in energy between the co-crystal and single-component structures. Regardless, the number of new iodine bonds created in the co-crystal appears to offer modest predictive power as to the degree to which it is more stable than the single-component structures.

Perhaps the clearest outlier is the bromine-bonded co-crystal EXEJUN (relative to components FATLER01 and HXMTAM07), the fifth-most stable halogen-bonded co-crystal ($\Delta E_{\text{co-cryst}} = -23.8\,\text{kJ mol}^{-1}$) despite only gaining one halogen bond per molecule. However, upon inspection, this co-crystal's stability may be due to a considerable change in hydrogen

bonding; EXEJUN contains two hydrogen bonds to each molecule in addition to the halogen bond, while no hydrogen bonding is present in either of the single component structures. (Note that this is already a more extreme change in hydrogen bonding per molecule than exhibited by any of the hydrogen-bonded co-crystal set.)

Combining hydrogen bonding and packing efficiency

The most informative description is obtained by simultaneously considering the change in packing coefficient in the co-crystal and the measure of the change in hydrogen bond strength, as is shown in Figure 11, where points are colored according to $\Delta E_{\text{co-cryst}}$.

While the data presented still show considerable variation, there is a general agreement with chemical intuition when our descriptors are combined. More stable co-crystals tend to be those that both increase the strength of the H-bonds present and increase their efficiency of crystal packing (i.e. those points in the upper-left quadrant) relative to the single component structures. Conversely, the less-stable co-crystals tend to be those that show little gain or a net loss in H-bond strength, as well as having less efficient packing.

As mentioned, a majority of the co-crystals considered pack less efficiently than their single-component structures. This is despite the overwhelming proportion of co-crystals that our PBE+D3 optimizations predict to be more stable. In some cases, this less-efficient packing is offset by a gain in H-bond strength. However, the spread of data in Figure 11 indicate that a simple balance of these two factors is insufficient to explain the variation in the relative stabilities observed; several comparatively stable co-crystals demonstrate both a loss in packing efficiency and a considerable loss in H-bond strength. Moreover, two of our most stable co-crystals either suffer a large loss in hydrogen-bonding (and very little change in packing) or a very large loss in packing efficiency (and have no hydrogen bonding). Evidently, the complexities of intermolecular interactions driving co-crystallization cannot be completely described by such comparatively simple descriptors.
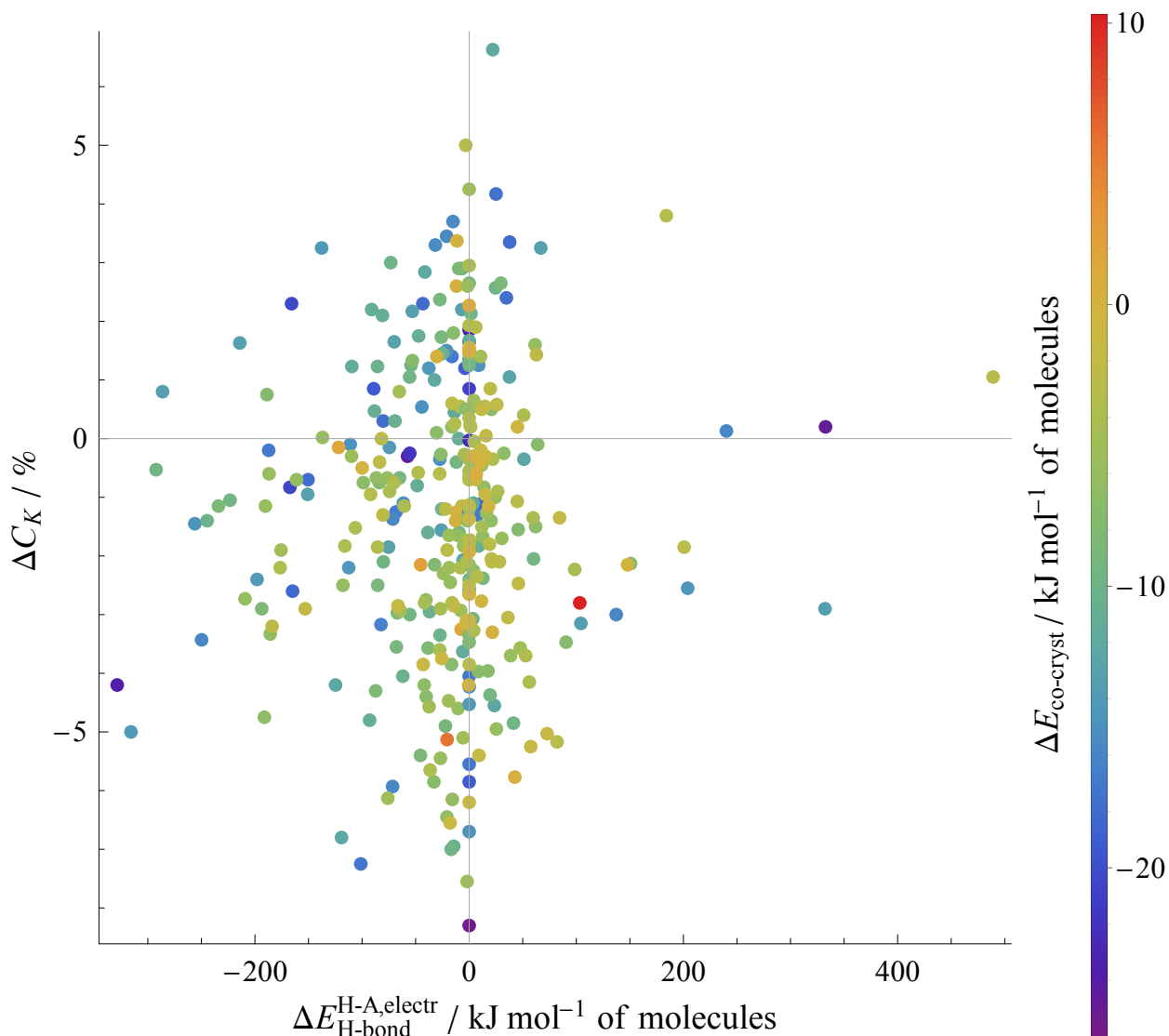
Figure 11: Relative stabilities of co-crystals (indicated by colour) as a function of both the change in packing coefficient and the change in the H-bond strength measure. Note that a more negative value for $\Delta E_{\text{H-bond}}$ indicates a net gain in strength in the co-crystal. This plot includes all of our co-crystals except for the 6 structures found to undergo proton transfer in their structural optimizations.

Limitations of descriptors

A major drawback of both the packing efficiency and the hydrogen bond strength descriptor is that they are dependent on the crystal structure, not merely molecular information. (The straightforward count of hydrogen bonds, while also dependent the crystal structure, is at

least constrained by the structure of the individual molecules and the stoichiometry in the co-crystal.) This reduces the utility of these descriptors for predicting the stability of co-crystals based solely on the molecules involved – a desirable goal for crystal engineering purposes.

However, these descriptors still can be useful to provide a qualitative model for the relative stability of a hypothesized structure without requiring an expensive periodic DFT optimization, and the relative difference in descriptors may be more informative than the absolute values. If a hypothetical co-crystal structure demonstrates both reduced packing efficiency and weaker H-bonding compared to its single components, it is unlikely to be considerably more stable according to our calculations. Such a structure may therefore be less worthy of synthetic targeting than another co-crystal structure in which both descriptors improve.

It is notable that these descriptors are not necessarily independent of each other – the H-bond strength measure (depending as it does on r_{HA}) can in principle be improved through denser packing without any substantial rearrangement of molecules, e.g. in the trivial case of compression of a given crystal structure. However, there is no systematic physical link between the two descriptors, and the data in Figure 11 show the values obtained tend to be largely independent of each other in practice.

It should also be mentioned that our method for computing the partial charges for measuring the H-bond strength acts on the isolated gas-phase molecules in their crystal geometry – the only influence of the surrounding crystal environment on these charges is indirect, in enforcing the precise molecular conformation adopted. Our H-bond strength measure lacks any description of polarization due to neighboring molecules (even the other molecule involved in the H-bond), and is therefore only a basic description of this interaction.

A stronger correlation between the descriptors and stability might be also seen in a set which does not focus solely on observed structures. Our use of structures from the CSD restricts our test set to systems which are already experimentally known, and there are physical

factors that could promote the formation of a co-crystal beyond the packing and specific interactions – for example, kinetic or solvent effects during the crystallization process – such that it is observed regardless of the change in packing efficiency or the H-bond interaction network. However, given our focus on rationalizing co-crystal formation for experimentally-observed cases, and the expense of considering a chemically- as well as crystallographically-diverse set of hypothetical computationally-generated structures, we leave such exploration to future efforts.

Conclusions

We have demonstrated that periodic DFT using the PBE+D3 level of theory predicts the vast majority of organic co-crystals to be energetically more stable than their single-component counterparts. Co-crystallization seems to almost always be a thermodynamically-driven process, with 95% of co-crystals more stable than their single-component structures. The characteristic gain in stability upon co-crystallization averages 8 kJ mol^{-1} (per molecule); a value considerably greater than that associated with crystalline polymorphism but well within the region of relevance for computational CSP studies.

Both the sign and magnitude of this stability are in very good agreement with previous, smaller studies of specific, chemically-related sets of molecules. We find that considering our own specific set of halogen-bonded co-crystals reveals a slightly greater stability than the general lighter-element organic case, and that *all* of the halogen-bonded systems we consider are more stable as the co-crystal than as separate single-component crystals.

We have also shown that the gain in stability upon co-crystallization is a complex phenomenon that is not necessarily well-described by changes in basic environmental or crystalline descriptors like molecular hydrogen bond count, strength of hydrogen bonding, or crystalline packing efficiency. Qualitative overall trends can be observed that match chemical intuition – more strongly hydrogen-bonded and efficiently-packed co-crystals tend to be

more stable. However, the considerable variance and significant number of exceptions in the general data set support the idea that co-crystallization is driven by a wide range and balance of physical factors that are not encapsulated by these descriptors. That said, when looking at extremes of stability or at groups of specific, more closely-related systems such as halogen-bonded co-crystals, it appears that these basic descriptors can be of some utility in rationalizing the observed stability. Such descriptors may therefore be more appropriately used in narrowly-targeted surveys of related systems. This is particularly true of, for example, the increase in number of halogen bonds, which tracks well with the typical relative stability in the specific case of iodine-bonded systems.

The size of our test set and the expense of periodic DFT limit us to employing only one exchange-correlation functional and dispersion correction scheme, but we suggest that the wide variety of co-crystal systems considered and the sheer size of the set (at least three times larger, and much more chemically diverse, than any other DFT study yet undertaken) make our conclusions reasonably applicable to the general case. Even if the precise value of the stability may vary due to a systematic inaccuracy of the functional, we emphasize that the size and breadth of the test set make our study more illuminating of co-crystallization generally than conclusions obtained from a considerably smaller test set assessed with a more expensive method or a range of functionals.

We have made our set of optimized structures available in the supplementary information, along with the identification of the relevant single-component structures for each co-crystal. It is our hope that, beyond our evaluation of the characteristic relative stability, this data set enables further work exploring the phenomenon of co-crystallization in general.

The availability of our benchmark value for the relative stability of organic co-crystal systems will also inform computational CSP studies, both for the prediction of specific co-crystal structures as well as for indicating when single-component structures may compete with co-crystals in the configuration space. Experimental work can also be assured that co-crystallization is predicted by theory to generally be a thermodynamic process, which

may help to provide routes for the simplification or streamlining of co-crystal screening procedures. Along these lines, a particularly relevant avenue for further computational work is consideration of the effects of thermal motion on the relative co-crystal stability via the calculation of e.g. lattice free energies as has been performed in our group for crystal polymorphs.[6]

The range of calculated values of the stability highlights the importance of using the most physically-accurate possible description of systems that is still computationally tractable, as attempting to explain or rationalize such complex and multi-faceted interactions using simple descriptors or highly-approximate interaction models can be unrepresentative of the true energetics of co-crystallization. The current work demonstrates the power of high-level quantum mechanical methods to provide insights that should be valuable in interpreting and guiding the experimental study of co-crystals, as well as computational efforts aimed at predicting co-crystallization.

Supporting Information Available

Description of CSD search, list of common solvents, structures that undergo proton transfer upon optimization and results of hydrogen bond calculation including donor-acceptor

Coulombic repulsion. Tables of co-crystal Cambridge Structural Database refcodes with their corresponding single-component structures, calculated relative stabilities and specific interaction counts. Optimized PBE+D3 structures for all co-crystals and their single-component structures, provided in Crystallographic Information File (.cif) format. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Thakuria, R.; Delori, A.; Jones, W.; Lipert, M. P.; Roy, L.; Rodríguez-Hornedo, N. *Int. J. Pharm.* **2013**, *453*, 101 – 125.

(2) Karki, S.; Friščić, T.; Fábián, L.; Laity, P. R.; Day, G. M.; Jones, W. *Adv. Mater.* **2009**, *21*, 3905–3909.

(3) Yan, D.; Delori, A.; Lloyd, G. O.; Friščić, T.; Day, G. M.; Jones, W.; Lu, J.; Wei, M.; Evans, D. G.; Duan, X. *Angew. Chem. Int. Ed.* **2011**, *50*, 12483–12486.

(4) Trask, A. V.; Motherwell, W. D. S.; Jones, W. *Cryst. Growth Des.* **2005**, *5*, 1013–1021.

(5) Fleischman, S. G.; Kuduva, S. S.; McMahon, J. A.; Moulton, B.; Bailey Walsh, R. D.; Rodríguez-Hornedo, N.; Zaworotko, M. J. *Cryst. Growth Des.* **2003**, *3*, 909–919.

(6) Nyman, J.; Day, G. M. *CrystEngComm* **2015**, *17*, 5154–5165.

(7) Nyman, J.; Day, G. M. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31132–31143.

(8) Musumeci, D.; Hunter, C. A.; Prohens, R.; Scuderi, S.; McCabe, J. F. *Chem. Sci.* **2011**, *2*, 883–890.

(9) Aakeroy, C. B.; Wijethunga, T. K.; Desper, J. *New J. Chem.* **2015**, *39*, 822–828.

(10) Delori, A.; Galek, P. T. A.; Pidcock, E.; Patni, M.; Jones, W. *CrystEngComm* **2013**, *15*, 2916–2928.

(11) Klamt, A. *J. Cheminform.* **2012**, *4*, O14.

(12) Wicker, J. G. P.; Crowley, L. M.; Robshaw, O.; Little, E. J.; Stokes, S. P.; Cooper, R. I.; Lawrence, S. E. *CrystEngComm* **2017**, *19*, 5336–5340.

(13) Cruz-Cabeza, A.; Day, G. M.; Jones, W. *Chem. - Eur. J.* **2008**, *14*, 8830–8836.

(14) Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. *J. Chem. Theory Comput.* **2009**, *5*, 1432–1448.

(15) Bučar, D.-K.; Day, G. M.; Halasz, I.; Zhang, G. G. Z.; Sander, J. R. G.; Reid, D. G.; MacGillivray, L. R.; Duer, M. J.; Jones, W. *Chem. Sci.* **2013**, *4*, 4417–4425.

(16) Cruz-Cabeza, A. J.; Karki, S.; Fabian, L.; Friščić, T.; Day, G. M.; Jones, W. *Chem. Commun.* **2010**, *46*, 2224–2226.

(17) Chan, H. C. S.; Kendrick, J.; Neumann, M. A.; Leusen, F. J. J. *CrystEngComm* **2013**, *15*, 3799–3807.

(18) Issa, N.; Karamertzanis, P. G.; Welch, G. W. A.; Price, S. L. *Cryst. Growth Des.* **2009**, *9*, 442–453.

(19) Gavezzotti, A.; Colombo, V.; Lo Presti, L. *Cryst. Growth Des.* **2016**, *16*, 6095–6104.

(20) Gavezzotti, A. *J. Phys. Chem. B* **2002**, *106*, 4145–4154.

(21) Otero-de-la Roza, A.; Johnson, E. R. *J. Chem. Phys.* **2012**, *137*, 054103.

(22) Allen, F. H. *Acta Cryst. Sect. B* **2002**, *58*, 380–388.

(23) van de Streek, J. *Acta Cryst. Sect. B* **2006**, *62*, 567–579.

(24) Reilly, A. M. et al. *Acta Cryst. Sect. B* **2016**, *72*, 439–459.

(25) Clark, S.; Segall, M.; Pickard, C.; Hasnip, P.; Probert, M.; Refson, K.; Payne, M. *Z. Kristall* **2005**, *220*, 567–570.

(26) Kresse, G.; Hafner, J. *Phys. Rev. B* **1993**, *47*, 558–561.

(27) Kresse, G.; Hafner, J. *Phys. Rev. B* **1994**, *49*, 14251–14269.

(28) Kresse, G.; Furthmller, J. *Computational Materials Science* **1996**, *6*, 15–50.

(29) Kresse, G.; Furthmüller, J. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(30) van de Streek, J.; Neumann, M. A. *Acta Cryst. Sect. B* **2010**, *66*, 544–558.

(31) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(32) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*.

(33) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comp. Chem.* **2011**, *32*, 1456–1465.

(34) Blöchl, P. E. *Phys. Rev. B* **1994**, *50*, 17953–17979.

(35) Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(36) Neumann, M. A.; Perrin, M.-A. *J. Phys. Chem. B* **2005**, *109*, 15531–15541.

(37) Moellmann, J.; Grimme, S. *J. Phys. Chem. C* **2014**, *118*, 7615–7621.

(38) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. *Acta Cryst. Sect. B* **2016**, *72*, 171–179.

(39) Breneman, C. M.; Wiberg, K. B. *J. Comp. Chem.* **1990**, *11*, 361–373.

(40) Frisch, M. J. et al. Gaussian 09 Revision D.01. Gaussian Inc. Wallingford CT 2016.

(41) Kitaigorodsky, A. *Molecular crystals and Molecules*; Physical Chemistry; Elsevier Science, 2012.

(42) Spek, A. L. *Acta Cryst. Sect. D* **2009**, *65*, 148–155.

(43) Willock, D. J.; Price, S. L.; Leslie, M.; Catlow, C. R. A. *J. Comp. Chem.* **1995**, *16*, 628–647.

(44) Horiuchi, S.; Kobayashi, K.; Kumai, R.; Minami, N.; Kagawa, F.; Tokura, Y. *Nature Comm.* **2015**, *6*, 7469.

(45) Raatikainen, K.; Rissanen, K. *CrystEngComm* **2011**, *13*, 6972–6977.

(46) Fernandes, M. A.; Levendis, D. C. *Acta Cryst. Sect. B* **2004**, *60*, 315–324.

(47) Zhang, S.-W.; Brunskill, A. P. J.; Schwartz, E.; Sun, S. *Cryst. Growth Des.* **2017**, *17*, 2836–2843.

(48) Taylor, C. R.; Bygrave, P. J.; Hart, J. N.; Allan, N. L.; Manby, F. R. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7739–7743.

(49) Gavezzotti, A. *Crystallogr. Rev.* **1998**, *7*, 5–121.

For Table of Contents Use Only

Evaluating the energetic driving force for co-crystal formation

Christopher R. Taylor and Graeme M. Day
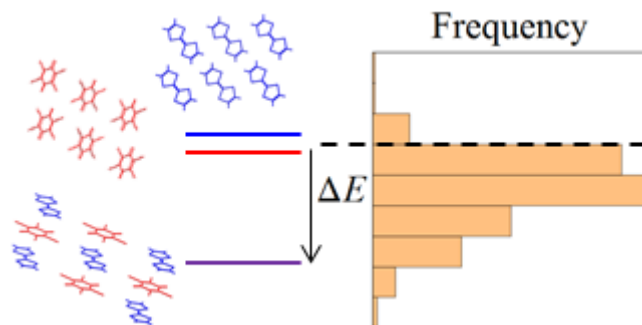
synopsis:

Solid state DFT is used to evaluate the energetic stability of 350 observed co-crystals with respect to the crystal structures of the separate coformers. The results demonstrate that the vast majority of known co-crystals are more stable than the stoichiometrically weighted sum of their component crystals. We analyze the distribution of energies and the influence of hydrogen and halogen bonding.