CrystEngComm

ARTICLE

Explaining crystallization preferences of two polyphenolic diastereoisomers by crystal structure prediction

Marta K. Dudek,*[a,b] and Graeme M. Day[a]

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Despite their structural similarity, two naturally occuring polyphenols, epicatechin and catechin, display significantly distinct crystallization behaviour. Epicatechin crystallizes only as a pure compound, and appears to be monomorphic, whereas no pure crystalline form of catechin is known, but it can form a variety of solvates with polar solvents. This work aims to explain these experimentally observed differences using the results of crystal structure prediction calculations. The entire conformational space of both molecules has been included in the crystal structure prediction study, which also explored the crystal structure landscapes of the pure crystals, 1:1 solvates and 2:1 solvates. From the computational results, we were able to show that the experimental observations with regard to crystallization behavior are the result of a trade-off between intra- and intermolecular energy contributions to the total energy of a crystalline system. In the case of epicatechin, conformers with low gas phase energies are at the same time the ones able to form crystal structures with favourable intermolecular interactions. In contrast, only high energy gas phase conformers of catechin were found to be able to pack efficiently. Consequently, the unfavorable intramolecular energy contribution has to be compensated by stabilizing intermolecular interactions. The calculations for 2:1 and 1:1 methanol solvates of both molecules demonstrate that such compensation can be readily provided by solvent molecules able to form hydrogen bonds with catechin.

Introduction

In recent years Crystal Structure Prediction (CSP) has been one of the most rapidly developing areas of computational chemistry.[1] While CSP has proved to be very successful in predicting *de novo* crystalline structures of fairly rigid organic compounds,[2] and there has been good progress in addressing conformationally flexible molecules,[3] some systems still remain difficult to tackle. This applies especially to multicomponent systems, such as crystals with multiple molecules in the asymmetric unit, including hydrates and solvates, and even more so if the component molecules have multiple degrees of conformational (intramolecular) freedom.[4,5] For example, in the recent 6th blind test on crystal structure prediction one polymorph with Z' = 2 of a pharmaceutical molecule was not predicted by any of the 25 groups taking part in the test, whereas a structure of a salt hydrate has been predicted by only one group.[6] It is therefore not surprising that much effort has been directed at pushing the applicability boundaries of CSP still further to tackle more and more complicated systems.

As a part of this effort, the prediction of solvate and hydrate formation of pharmaceutically important molecules has been recently undertaken by a number of research groups. Among others, CSP was employed to predict the stoichiometry and structure of acetic acid solvates of theobromine,[7] caffeine and urea,[8] hydrates of dihydrobenzoic acids,[9] hydrates of a 5-HT_{2a} antagonist B5HCl[10] and hydrates of dapsone.[11] Solvate formation propensity was also tested for alkaloids, strychnine and brucine which, despite structural similarity, behave differently in terms of the formation of solvates and hydrates.[12] In the latter work, after exhaustive calculations of the crystal energy landscapes for the two alkaloids, the authors attributed the differences in behaviour to the ability of brucine to form some loosely packed, but energetically competitive structures to a more tightly packed structure.

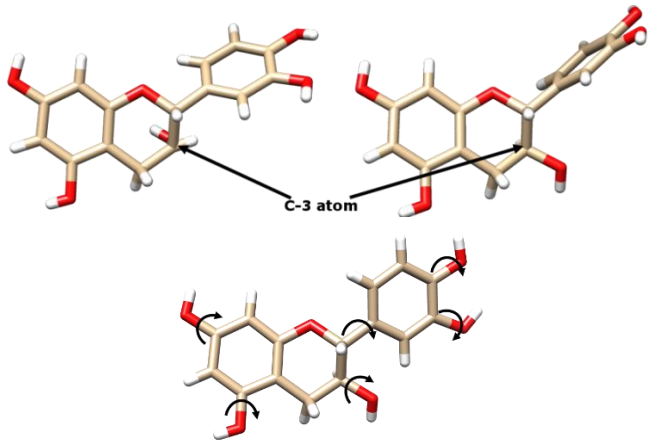
In this work we employ CSP to explain the crystallization preferences of two diastereomeric flavan-3-ol derivatives, (+)-catechin (CAT) and (-)-epicatechin (EPI). Both substances are naturally occurring polyphenols with many reported biological activities,[13] such as antioxidant,[14] antimicrobial,[15] neuroprotective,[16] and Ca^{2+} antagonist.[17] Although they have been known to chemists for over 100 years, and a lot of effort was made to recognize their chemistry, their crystallization behaviour and propensity to exist in certain crystalline forms remains unclear. Despite their structural similarity (with the only difference in their chemical structure being stereochemistry at the C-3 atom, see Scheme 1) EPI and CAT display profoundly distinct behaviour as to the formation of crystalline structures. The crystal structure of anhydrous EPI has been known for over 30 years, and no other crystalline structure has been found so far, be it another polymorphic form, a solvate or a hydrate. By contrast, CAT does not yield any crystalline

[a] Computational Systems Chemistry, School of Chemistry, University of Southampton, SO17 1BJ, UK
[b] Center of Molecular and Macromolecular Studies PAS, Sienkiewicza 112, 90-363 Lodz, Poland.

Electronic Supplementary Information (ESI) available: Details of experimental crystallization screening; further crystal structure prediction data. See DOI: 10.1039/x0xx00000x

forms by itself, and to date one methanol solvate with two methanol molecules per CAT molecule found in the asymmetric unit cell[18] and two non-stoichiometric hydrates, with 4.5 and 2.5 water molecules per CAT molecule[19,20] have been described at the atomic level.

In the course of our studies on CAT solvate formation, we have discovered that these two solvent systems are not the only ones to form solvates with CAT. In fact, CAT tends to form solvates with any encountered polar solvent, regardless of the crystal preparation method, be it solution crystallization, mechanochemical grinding or solvent vapour diffusion (see Supporting Information for a detailed description of the solvents and methods used in crystallization screening). It may be expected from these experimental observations that the formation of CAT solvates will be more energetically favourable than the formation of the unsolvated crystalline forms. To test this hypothesis we have performed crystal structure prediction calculations for the unsolvated structures of EPI and CAT, as well as for methanol solvates with two methanol molecules per CAT/EPI molecule (2:1 solvates), as in these two tested sets we are able to verify the applied methodology with the experimental results for one of the polyphenols in each case. Additionally, the capability of the CSP methods to predict solvate stoichiometry is tested, by including in the CSP search 1:1 methanol solvates for both systems. This paper aims also at understanding the reason behind the observed differences between EPI and CAT in their tendency to crystallize in such a different manner. As a result, a broader understanding of solvate formation may be reached. Finally, since both tested compounds are fairly flexible, each with six rotatable bonds, this work may serve as a test for CSP performance in treating multicomponent and flexible systems.



Scheme 1. Molecular structures of the studied diastereosiomers: EPI (upper right) and CAT (upper left), and CAT structure with marked rotatable torsions (lower). The only structural difference between the two compounds is the C-3 stereogenic centre.

Methods

Conformational search.

The conformational search was performed with the Spartan'16[21] software using a grid search and the MMFF force field.[22] All the rotatable (exocyclic) bonds of EPI and CAT were included in the search, with 3-fold or 2-fold (for aromatic hydroxyl groups) rotation. As a result, 96 conformers were generated for each of the polyphenols. Subsequently, each conformer was subjected to geometry optimization using the Gaussian09 software,[23] at the B3LYP[24]/6-311G** level of theory with the GD3BJ dispersion correction scheme,[25] followed by frequency calculations to ensure that an energetic minimum was achieved for each conformer. As some starting structures optimized to the same minimum, the number of distinct conformers was fewer than the number of starting geometries; 56 and 68 conformers were located for EPI and CAT, respectively. Note, that the methodology used here takes only local minima from conformational search as starting points for CSP, and therefore it is not as thorough as allowing for full flexibility of a molecule during crystal structure generation. However, we believe that with sufficient number of distinct local gas phase minima, and by allowing all atoms to relax at the last stage of the CSP process (see DFT-D calculations), most relevant structures should be found, using at the same time less amount of time and resources for the search. All the obtained relative energies as gathered in the Supporting Information.

CSP search for unsolvated structures.

The CSP search was performed in two stages. In the first stage, for each conformer of EPI and CAT separately, 2000 valid (successfully lattice energy minimized) crystal structures in each of the two most common chiral space groups, namely $P2_1$ and $P2_12_12_1$, were generated using the Global Lattice Energy Explorer code.[26] Since in this work only enantiomerically pure systems are considered, only chiral space groups were considered. Crystal structures were optimized with respect to the intermolecular energy contribution of the lattice energy using DMACRYS 2.2.1.0,[27] which treats molecules as rigid entities. For each conformer, atom-centred distributed multipoles up to rank 4 (hexadecapoles) were calculated from the B3LYP-D3BJ/6-311G** charge density using a distributed multipole analysis (DMA)[28] with GDMA 2.2.11.[29] Intermolecular repulsion-dispersion interactions were calculated using the FIT potential, [30] using a 25Å cut-off on van der Waals interactions. The resulting crystal structures were clustered using similarity of PXRD patterns for structures within the density range of 0.05 g/cm^3 and within an energy range of 0.1 kJ/mol, to eliminate duplicates. Finally the lattice energies for all the best structures were recalculated using multipoles calculated from a polarized electron density of the molecule, using the PCM model to account for the dielectric constant in a solid phase, which we took to have a value of 3. The total relative energy for the *i*-th structure was calculated according to the equation:

$$E_{tot} = E_i^{inter} - (E_i^{intra} - E_{min}^{intra}) \quad \text{(eq. 1)}$$

where E^{intra} denotes the gas phase energy obtained from the DFT molecular calculations, E^{inter} denotes intermolecular lattice energy from the DMACRYS calculations, and the 'min' subscript denotes values obtained for the conformer with the lowest gas phase energy.

In the second stage of CSP, we selected all conformers that had led to any crystal structures within 20 kJ/mol of the structure with the lowest total energy. A fuller search was performed for these conformers, comprising the generation and optimization of 10000 crystal structures for each of the four most common chiral space groups ($P2_1$, $P2_12_12_1$, $C2$ and $P1$). As no differences between the 2000 structures and 10000 structures landscapes in terms of the best 20 kJ/mol structures were observed, we focus on the initial 2000 structures set, obtained for all conformers. For the comparison of both sets see Supporting Information.

CSP search for methanol solvates.

The solvated structures have considerably more degrees of freedom, due to two additional methanol molecules, each of which has its own degrees of freedom with respect to their location and orientation in the unit cell. As a result, more sampling is needed to cover the whole crystal energy landscape. The search was carried out until 10000 crystal structures were generated and successfully lattice energy minimized for each of the conformers, which resulted in 560000 and 680000 crystal structures in total for EPI and CAT, respectively, first in each of the two tested space groups, $P2_1$ and $P2_12_12_1$. For those conformers that were found to be able to form energetically favourable crystal structures in one of those two tested space groups, crystal structure searches in two more space groups, *i.e.*$C2$ and $P1$, were performed. The structure generation and optimization procedures were the same as for the unsolvated structures, and were performed separately for 1:1 and 2:1 methanol solvates.

For comparison of the calculated solvate lattice energies of solvates with the pure crystal forms of EPI and CAT, we corrected the calculated lattice energy by the energetic cost of removing N (N = 1 for 1:1 methanol solvates and N = 2 for 2:1 solvates) molecules of methanol from its pure phase. For the energy of pure methanol, lattice energy calculations using DMACRYS and the same force field as used in CSP were performed for the α and β polymorphs of methanol.[31] The calculated lattice energies were equal to 44.44 and 43.29 kJ/mol, for the α and β polymorphs, respectively. Finally, to account for the change in the internal energy of methanol in the liquid and solid state, an energy correction equal to 3/2RT at 300K was applied, as the equipartition estimate of the internal energy change due to molecular rotations, translations and intermolecular vibrations in the solid.[32,33] Thus, the final equation used for the calculations of the comparable total energy of the solvates is as follows:

$$\boldsymbol{E_{comp}^{tot} = E_{solv}^{tot} - N * (E_{methanol}^{intra} + E_{methanol}^{inter} - 3/2 * RT)} \quad \text{(eq. 2)}$$

where E_{solv}^{tot} is the total energy of the solvated crystal, calculated as in equation 1, while N denotes the number of solvent molecules, and is equal to 1 or 2 for 1:1 and 2:1 methanol solvates, respectively.

DFT optimization.

To evaluate the obtained energy rankings, all solvated and unsolvated structures found within the lowest 15 kJ/mol on each landscape (separately for pure forms, 1:1 solvates and 2:1 solvates) were recalculated at the periodic DFT level of theory with the CASTEP code.[34] These optimizations were performed in two steps: first, all atomic positions were optimized with the cell parameters kept fixed at the force field values, then a second optimization was performed allowing unit cell parameters to relax. All periodic DFT calculations were performed with the PBE functional, Grimme-D2 dispersion correction scheme,[35] ultrasoft pseudopotentials, and a plane wave basis set cut-off energy of 600 eV, which was chosen based on convergence tests. The number of electronic k-points used depended on the structure (more precisely on the unit cell dimensions) and was chosen to provide a k-point spacing of approximately 0.07 $Å^{-1}$ in all directions of the Brillouin zone.

Results & Discussion

Conformational search

Both, EPI and CAT have considerable molecular flexibility, at least in terms of what can be considered during a CSP search, with six rotatable bonds defining the hydroxyl orientations and the relative orientations of the two ring systems. An exhaustive conformational search of the isolated molecules is required to identify the conformer space that must be included in the CSP studies used to understand the crystallization preferences of these polyphenols. All six dihedrals were included in the search and, after DFT optimization, 56 and 68 unique conformers were obtained for EPI and CAT, respectively. An important feature of the resulting set of conformers is their small energy range, with ca. two thirds of the conformers found within a DFT energy range of 20 kJ/mol, and the gas phase energy of the least energetically favourable conformer being 38.5 kJ/mol above the most stable (for full list of conformers and their energies see Supporting Information). It is usually assumed that stable crystal structures may be formed by conformers with relative gas phase energy not higher than about 20 kJ/mol,[36] as up to that amount the gas phase energy may be compensated by favourable intermolecular interactions. Nevertheless, in our further studies we have decided to include all the obtained conformers. This was for three reasons: (i) in previous studies performed to determine the preferred cut-off energy level for the gas phase conformers,[36] molecules with intramolecular hydrogen bonding were deliberately excluded from the studied set of structures; meanwhile both CAT and EPI have neighbouring hydroxyl groups, which can form such an intramolecular hydrogen bond, whose energetic contribution could broaden the conformational energy range; (ii) the mentioned level of lattice energy compensation was derived from the computational studies of unsolvated structures, so there is no clear evidence that the same also apply for the solvated ones; (iii) since in the studied systems conformers of EPI and CAT have unusually small gas phase energy differences, some conformers could have been omitted by chance by

applying too low cut-off level, especially accounting for the possible errors in the DFT methods.

Crystal energy landscapes for unsolvated structures

Figure 1 presents the CSP results for the unsolvated crystalline structures of EPI and CAT, including crystal structures generated using all conformers of each molecule. The predicted landscape obtained for EPI features one particularly stable crystal structure, whose total energy is 9 kJ/mol lower than that of the next most stable structure. This lowest energy structure (the global minimum) of EPI corresponds to the experimentally observed crystal structure: according to the Crystal Packing Similarity Tool,[37] an overlay of a 20-molecule cluster from the CSP global minimum matches the experimental structure of EPI, with an RMSD in atomic positions of 0.383 Å or 0.440 Å, when comparing structures without and with hydrogens, respectively. Comparison of unit cell parameters of the predicted and experimental structure is shown in Table 1. After re-optimizing both crystal structures (the experimental one and the global minimum from the CSP search) with DFT (PBE-D2) under periodic boundary conditions, and including cell parameter optimization, the RMSD drops to 0.023Å and 0.024Å for a comparison excluding and including the positions of protons, respectively. This confirms that the obtained CSP global lattice energy minimum structure may be regarded as identical to the experimental one.

As can be seen in Figure 1, the two polyphenols lead to noticeably different landscapes of structures, particularly in terms of the number and distribution of structures found in the lowest 20 kJ/mol of their landscapes. For EPI, only a few conformers are able to form energetically favourable crystal structures, whereas for CAT a variety of conformers can form low energy crystal structures, resulting in approximately 40 crystal structures within the 20 kJ/mol low energy window. The inclusion of polarization effects in the final stage of the CSP calculation, via calculation of multipoles using the PCM model, enhanced this difference between the two molecules. We also note that only crystal structures in space groups $P2_1$ and $P2_12_12_1$ contribute to the low energy regions of the landscapes for both molecules; all crystal structures predicted in $C2$ and $P1$ have significantly higher lattice energies (see Supporting Information).

The observed differences in the crystal energy landscapes for unsolvated forms of CAT and EPI can be attributed to the fact that for the former only high-energy gas conformers are able to pack most efficiently in terms of intermolecular contributions to the crystal lattice energy, while for the latter mostly conformers with lower intramolecular energy are able to form crystal structures with favourable (low) intermolecular energies. This can be clearly seen after splitting the total energy of the predicted crystal structures into its two components, namely intra- and intermolecular interactions. The first value is represented by the DFT relative energy of each of the isolated conformers, whereas the second, intermolecular contribution is calculated from the atom-atom force field. The two contributions are visualized in Figure 2, where the intermolecular contribution to the energy is plotted for each crystal structure against its density, with the colour bar indicating the intramolecular energy of the conformer found in each structure. For CAT, the lowest (most stabilizing) intermolecular energy was found for one of the highest energy conformers, and all of the predicted crystal structures with intermolecular energy in the lowest 15 kJ/mol region correspond to high energy conformers (Fig. 2, right); the most energetically favourable conformer within this range of intermolecular energy (marked in Figure 2 with a black square) has an intramolecular energy 24.0 kJ/mol above CAT's most stable gas phase conformer. This hinders the stability of crystal structures that are available to CAT in its pure form (unsolvated structures). In contrast, for EPI the predicted crystal structure with the lowest intermolecular energy was found for a conformer whose intramolecular energy is only 5.05 kJ/mol above the most stable EPI conformer (marked in Figure 2 with a black square). Furthermore, there are a number of other energetically favourable conformers which form crystal structures within the range of the most stabilizing intermolecular interactions (Fig. 2, left).

Figure 2 also highlights conformers of EPI and CAT with (circles) and without (crosses) intramolecular hydrogen bond between two OH groups from dihydroxyphenyl ring. As can be seen, for CAT the conformers without intramolecular hydrogen bonds dominate the low intermolecular energy region, in contrast to EPI, for which many conformers having intramolecular hydrogen bonds also have favourable intermolecular energy, The breaking of this bond reflects itself in the hydrogen bonding motifs present in the most stable crystal structures of the two polyphenols. In Figure 3 these motifs are shown for the global minima of pure EPI and CAT (Figure 3a and b), as well as for the structure of CAT with the lowest intermolecular energy (Figure 3c). The global minimum of CAT is the structure built by the most stable gas phase conformer, possessing intramolecular hydrogen bond. On the other hand, this is the structure with relative intermolecular energy equal to 24.8 kJ/mol, which can be the result of the noticeably longer distances found for its H…OH pairs, as compared to those found for the structure with the lowest intermolecular energy. This latter structure has no intramolecular hydrogen bond, and therefore its gas phase conformer has an intermolecular energy 36 kJ/mol above the gas phase energy minimum. We conclude that part of the reason why CAT does not form any stable unsolvated crystal structure may lay in the inability of its low energy conformers to pack efficiently in a crystal lattice.

CrystEngComm

ARTICLE

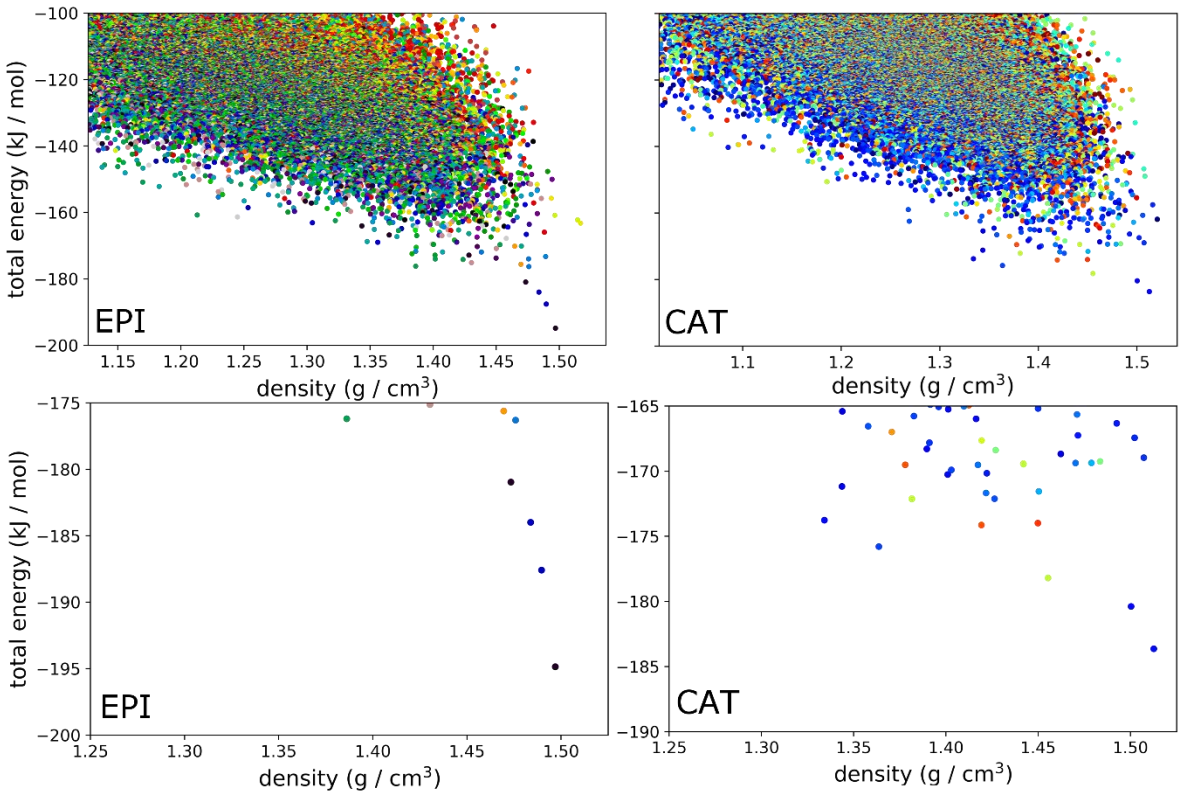


Figure 1. CSP results for epicatechin (left panel) and catechin (right panel). All low energy structures are located in space groups $P2_1$ and $P2_12_12_1$. Energies are calculated with the FIT + DMA force field and PCM polarization model. The colours of data points distinguish crystal structures containing different molecular conformers. Note, that the y axes in the lower panel have different scales.

Table 1. Comparison of unit cell parameters of the experimental (single crystal X-Ray diffraction) and CSP-calculated structure (optimized with force field or after DFT-D re-optimization) for epicatechin (EPI) and catechin (CAT) methanol solvate.

structure	cell parameter	experimental	CSP (force field)	PBE-D2
EPI	a / Å	6.708	6.369	6.648
	b / Å	13.291	13.274	13.246
	c / Å	14.262	15.234	13.687
CAT	a / Å	13.637	13.694	13.306
	b / Å	5.547	5.602	5.837
	c / Å	12.625	12.914	11.873
	β / °	63.64	63.48	65.14

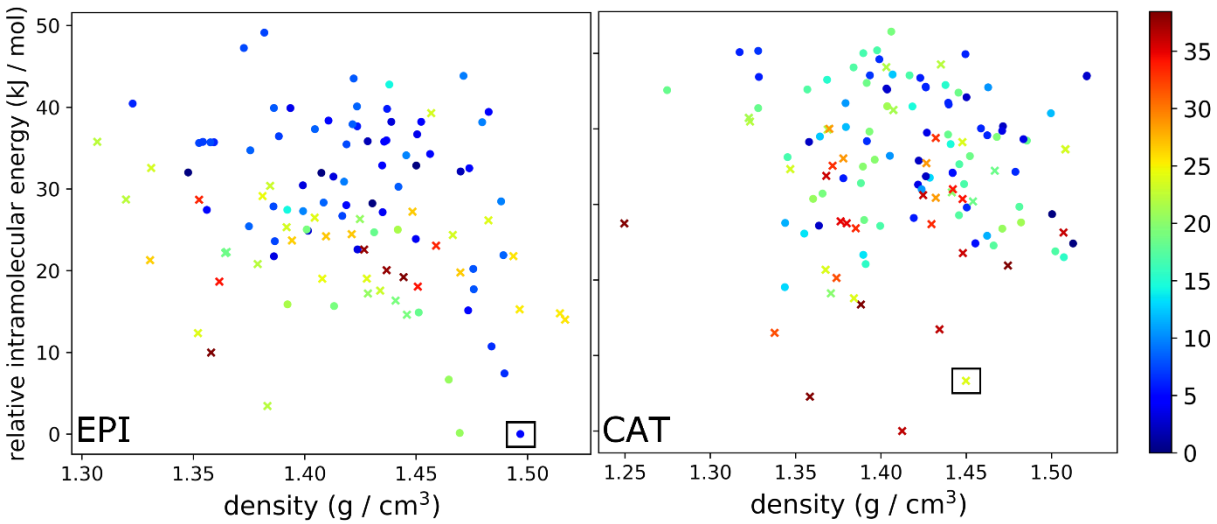


Figure 2. Intermolecular contributions to lattice energies for the best crystal structures of each of the conformers of EPI (left panel) and CAT (right panel) in space groups $P2_1$ and $P2_12_12_1$. The colour bars indicate the intramolecular energy calculated at the B3LYP/6-311G** level of theory (in kJ/mol) for the conformer in each crystal structure. Structures built by conformers with intramolecular hydrogen bond are marked with circles, and those without by crosses. Black squares mark the lowest energy predicted crystal structures having a conformer with the lowest gas phase energy found in the 20 kJ/mol intermolecular energy window. Intermolecular energies are calculated with the FIT + DMA force field and PCM polarization model. Note that here the respective colours represent gas phase energy of a conformer, and are different from those used in Figure 1.

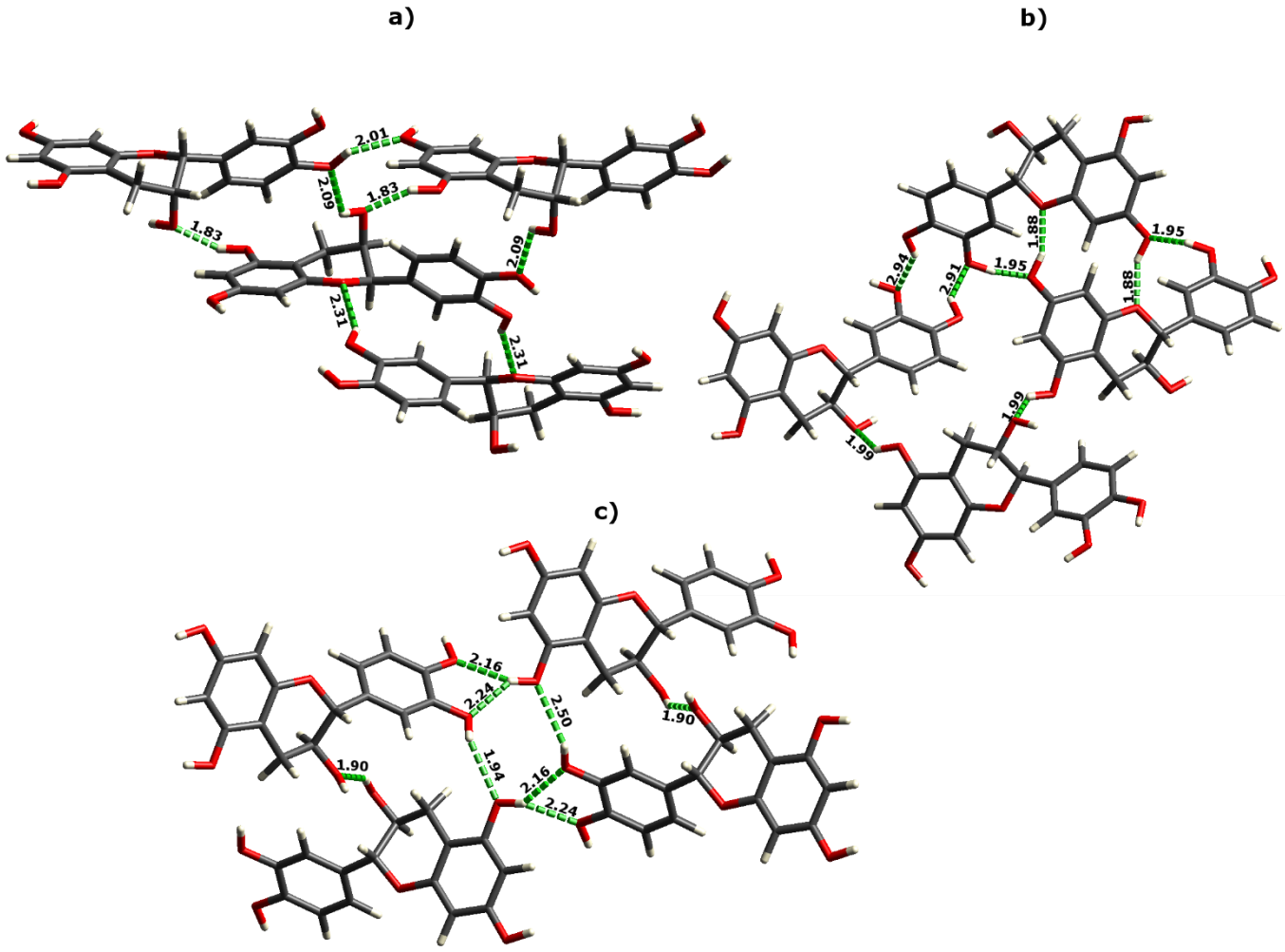


Figure 3. Hydrogen bonding motifs found in the global minima structures of EPI (a) and CAT (b), as well as in the structure of CAT with the lowest intermolecular energy contribution (c). The H…OH distances are given in Å.

Crystal energy landscapes for 2:1 methanol solvates

The crystal energy landscapes found for the solvated crystal structures of all studied conformers of EPI and CAT are shown in Figure 4. As was the case for the unsolvated structures, the lattice energies obtained for predicted 2:1 solvates in space groups *C*2 and *P*1 were significantly higher than those in $P2_1$ and $P2_12_12_1$, but here the differences between these space group pairs were smaller. Consequently three structures with total energies found within the lowest 15 kJ/mol energy window were found in the less common space groups: one in *P*1 with relative energy (with respect to the best 2:1 solvate energy) equal to 11.57 kJ/mol, and two in *C*2, with relative energies of 6.87 and 14.41 kJ/mol. Those structures were included in the final DFT-D re-optimization. However, a general trend of $P2_1$ and $P2_12_12_1$ space groups being more energetically favourable for methanol solvates of EPI and CAT, is noticeable.

These 2:1 solvate landscapes show a large number of low energy crystal structures and lack the large energy gap that was observed for the unsolvated EPI landscape (Fig. 1). In order to determine whether the experimental structure of the methanol solvate of CAT has been found among the predicted structures, those structures generated for the most similar conformers to the one present in the experimental structure were compared to the experimental structure using the Crystal Packing Similarity Tool.[37] None of these CSP structures displayed a perfect match with the experimental one, when comparing directly the DMACRYS results with the crystallographic structure. There was however one structure, which had the position of molecules in common with the reference, but in which the orientation of one methanol molecule was misaligned with the experimental crystal structure. This structure was found to be 10.6 kJ/mol higher in total energy than the lowest energy predicted 2:1 methanol solvate structure of CAT.

To further investigate the energies of the lowest energy predicted crystal structures, all the structures within a 15 kJ/mol energy window of the global energy minimum, from both the solvated and unsolvated landscapes, were re-optimized with the dispersion-corrected plane wave solid-state DFT (DFT-D). The re-ranking at the DFT-D level of theory is described in a later section. The results of periodic DFT-D re-optimization confirmed the match between predicted and experimental structures; after re-optimization, the matching crystal structure gave an excellent match to experiment, with an RMSD in atomic positions of 0.538 Å for a 20-molecule cluster taken from the predicted and experimental crystal structures.

It is noteworthy that the conformer found in the experimental crystal structure of CAT has one of the highest intramolecular energies of all the possible conformers, and is 33.1 kJ/mol higher in energy than the gas phase global minimum conformer. The high intramolecular energy of this conformer is mainly due to the lack of an intramolecular hydrogen bond between hydroxyl groups in a dihydroxyphenyl moiety, which is present in the low energy conformers of CAT.

Unlike in the CSP results for the unsolvated structures, the landscapes for 2:1 methanol solvates of EPI and CAT are quite similar, with a large number of structures found in the lowest energy range for both polyphenols (Fig. 4), though for CAT this number is somewhat lower than for EPI. A detailed analysis of the intra- and intermolecular contribution to the total energy of the predicted crystal structures also indicates that for both compounds the crystal structures with the lowest intermolecular energies are formed by conformers with the highest intramolecular energies, and with no intramolecular hydrogen bond (structures marked with crosses in Figure 5). However, we note that, compared to the unsolvated CAT results, the inclusion of methanol molecules in the crystal structure allows lower energy conformers of CAT to produce crystal structures with intermolecular stabilization energies close to the structure with the best intermolecular interactions (Fig. 5). For unsolvated CAT, the lowest energy conformer that led to crystal structures within 20 kJ/mol of the best intermolecular interaction energies was 24.0 kJ/mol above the gas phase lowest energy conformer. For CAT methanol solvates, conformers with noticeably lower energies (as low as 11.5 kJ/mol above the lowest energy gas phase conformer) led to crystal structures within 20 kJ/mol of the predicted solvates with the most stabilizing intermolecular interactions.

Figure 6 features two hydrogen bonding motifs found for the global minimum structure of 2:1 methanol solvate of CAT (6a), and for the experimental methanol solvate of CAT (6b). Both motifs are highly similar, even though the experimental structure is the one without the intramolecular hydrogen bond. Usually it is assumed that in alcoholic and aqueous solutions intramolecular hydrogen bond of polyphenols are retained, and only solvents with strong electron donating properties, such as pyridine or DMSO, have the capacity to break them.[38,39] On the other hand, it is possible that upon concentrating the solution during crystallization methanol – polyphenol interactions become stronger, thus making the conformations without an intramolecular hydrogen bond more feasible. This hypothesis is partially confirmed by the fact that in the experimental CAT solvate methanol molecules interact *via* hydrogen bonds with both hydroxyl groups, which otherwise may form this intramolecular hydrogen bond (Figure 6b). Interestingly, the observed hydrogen bonding motif in both structures of CAT solvates shown in Figure 6 is also similar to that of the unsolvated global minimum structure of CAT, but in the case of the solvates the H…OH distances are smaller, due to the presence of methanol molecules, and thus these interactions can contribute more towards the intermolecular energy. As a result it can be concluded that, although low gas phase energy conformers of CAT are still not able to pack efficiently, the interactions with solvent molecules within the crystal lattice helps to compensate for this intramolecular energy penalty.

Comparison of crystallization preferences of EPI and CAT

As CAT and EPI differ structurally only in the stereochemistry at one stereogenic centre, namely C-3 (Scheme 1), the total

energies obtained for their CSP-generated structures can be compared directly, and the same also applies for the predicted solvates. For 2:1 methanol solvates, the CSP results demonstrate that CAT is able to form more stable crystal structures than EPI solvates (Fig. 7, right), whereas for unsolvated crystal structures it is clear that EPI is able to form more energetically favourable structures (Fig. 7, left). In light of the results presented in the previous section, these differences can be associated with the capability of methanol molecules to stabilize the crystal structure of CAT, and therefore to saturate all the possible hydrogen-bonding sites, which compensates for the unfavourable intramolecular energy of the conformers that lead to the best crystal packing. For EPI, such compensation is not necessary and the inclusion of methanol molecules merely leads to the substitution of one hydrogen bond for another, without much gain in total energy.

Towards predicting stoichiometric ratio of solvates

As was shown above, there is a clear energetic gain from the formation of 2:1 methanol solvates by CAT, which is in agreement with the experimental observations that this molecule forms solvates in preference to an unsolvated crystal structures. At this point it is worthwhile to also compare the possible solvates with different stoichiometric ratio, *i.e.* 2:1 and 1:1 methanol solvates, and to examine whether one methanol molecule per CAT could be enough to overcome the unfavourable intramolecular energy. In Figure 8, we show a direct comparison of the relative total energies of the best predicted crystal structures of methanol solvates in two stoichiometric ratios. These comparisons account for the energetic cost of removing methanol from their pure phase, and for the internal energy change associated with the change in dynamics of the methanol molecules from the liquid phase to a crystalline structure (see equation 2). Note that these relative energies are shown with respect to the most stable solvate, separately for EPI and CAT.

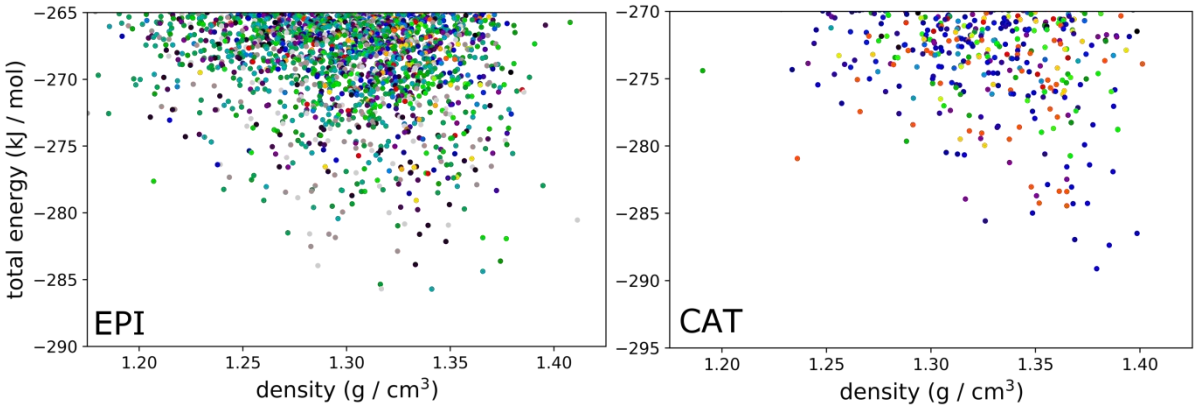


Figure 4. The best 20 kJ/mol part of the crystal energy landscapes for the 2:1 methanol solvates of all conformers of EPI (left panel) and CAT (right) in space groups $P2_1$, $P2_12_12_1$, $C2$ and $P1$. The colours of data points distinguish crystal structures containing different molecular conformers. Note, different scales at the y axes for EPI and CAT.

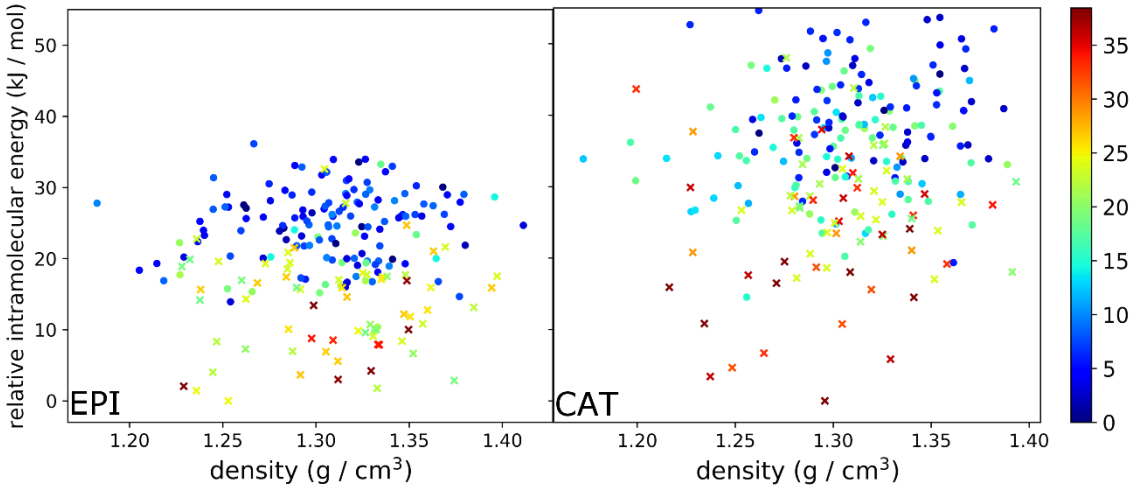


Figure 5. Intermolecular contributions to lattice energies for the best crystal structures of 2:1 methanol solvates of each of the conformers of EPI (left panel) and CAT (right panel) in space groups $P2_1$, $P2_12_12_1$, $C2$ and $P1$. The colour bars indicate the intramolecular energy calculated at the B3LYP/6-311G** level of theory (in kJ/mol) for the conformer in each crystal structure. Intermolecular energies are calculated with the FIT + DMA force field and PCM polarization model. Structures built by

conformers with intramolecular hydrogen bond are marked with circles, and those without by crosses. Note that here the respective colours represent gas phase energy of a conformer, and are different from those used in Figure 4.

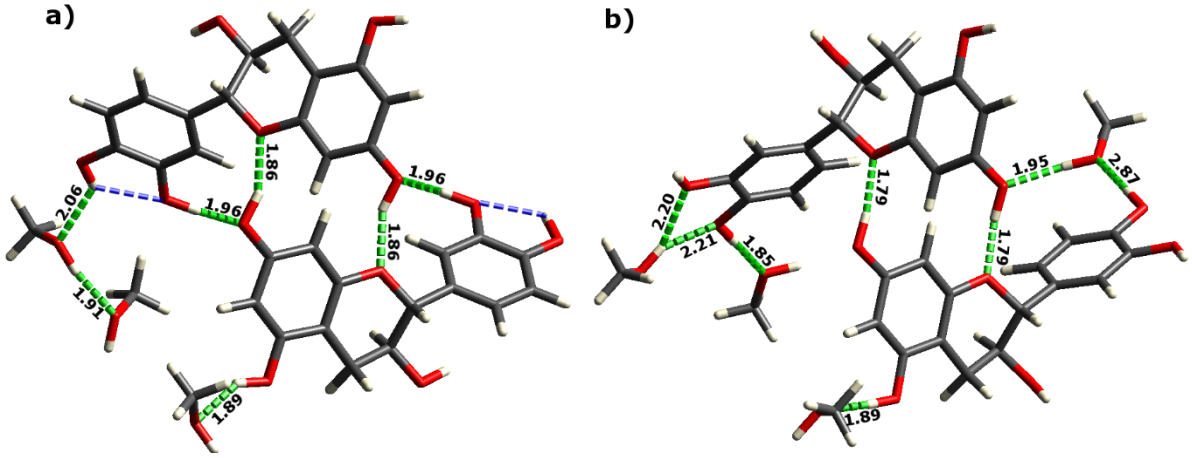


Figure 6. Hydrogen bonding motifs found for the global minimum structure of 2:1 methanol solvate of CAT (a), as well as for the experimental methanol solvate of CAT (b). The H…OH distances are given in Å, the intramolecular hydrogen bonds are marked with blue.

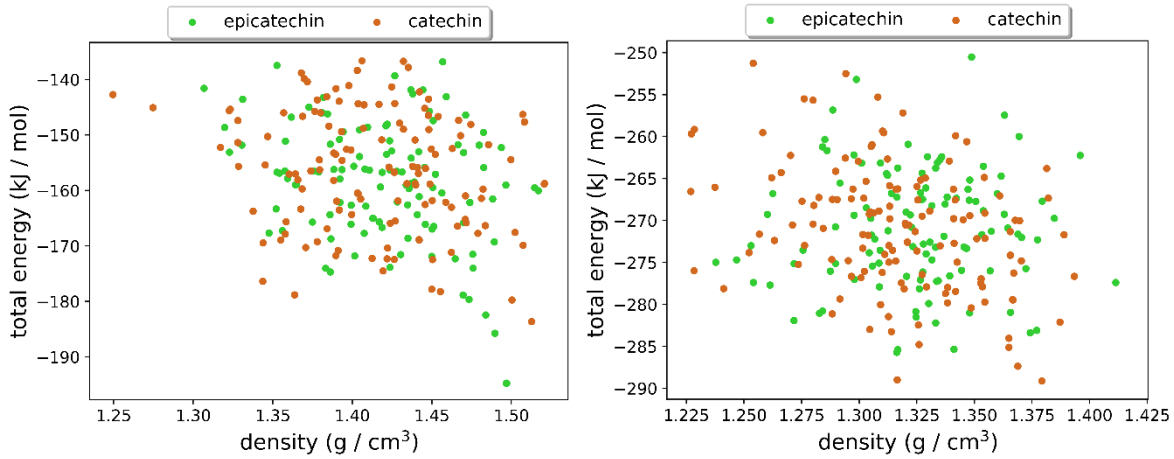


Figure 7. A comparison of crystalization preferences for CAT and EPI: unsolvated forms (left), methanol 2:1 solvates (right). Energies are calculated with the FIT + DMA force field and PCM polarization model.

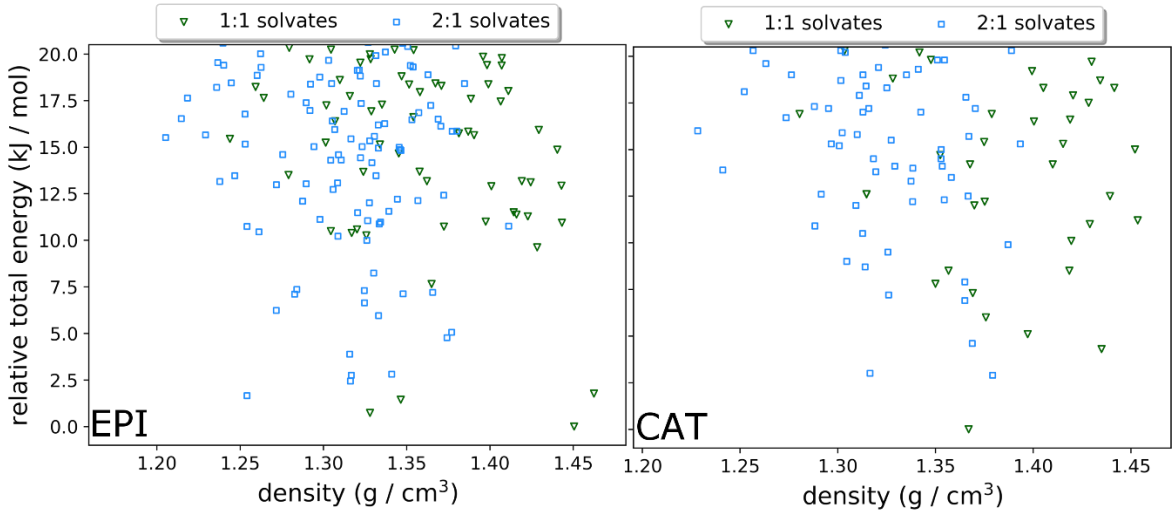


Figure 8. Best crystal structures (found within the best 20 kJ/mol energy window) of 2:1 and 1:1 methanol solvates of EPI (left) and CAT (right). Energies are calculated with the FIT + DMA force field and PCM polarization model, and are given relative to the lowest energy of the most stable solvate of the pair.

Inspection of Figure 8 shows that, while the 2:1 stoichiometry leads to slightly more low energy crystal structures, for both EPI and CAT, the 2:1 and 1:1 stoichiometries lead to crystal structures with similar stabilities. For both molecules, 1:1 solvates were found to be slightly more energetically favourable than 2:1 solvates, but the energy difference between the best 1:1 and 2:1 solvates in either case does not exceed 3 kJ/mol, which is well within the uncertainty in the force field method used to assess the lattice energies.[40]

Final comparison of the 1:1 and 2:1 solvates with the unsolvated crystal structures

A comparison of the relative energies of the solvated and unsolvated structures (Figure 8) clearly indicates that indeed there should be a difference in the crystallization preferences of EPI and CAT, in agreement with the experiment. In the case of EPI, the most stable structure is definitely the unsolvated one, which was found to be identical to the experimentally observed structure. Its total energy is 9.0 kJ/mol lower than the next best structure, which is also an unsolvated structure. The best solvated structure of EPI has relative total energy of 10.45 kJ/mol compared to the best unsolvated structure. These results explain the lack of any experimental tendency of EPI to form methanol solvates. Quite the opposite observations can be made for CAT, for which the most stable structure is a 1:1 solvate, followed by a 2:1 solvate. The best unsolvated crystal structure of CAT has relative total energy of 4.5 kJ/mol, which is only 1.6 kJ/mol above the best 2:1 solvate. The energy differences between unsolvated, 1:1 solvate and 2:1 solvates for CAT are small enough that uncertainties in the model related to the force field accuracy and rigid-molecule approximation make it difficult to assess the energetic preference for CAT. It is also concerning that the relative total energy found for the experimental structure of CAT is 13.5 kJ/mol higher than the most stable solvate structure. On the other hand, at this stage of the CSP no relaxation of the molecular conformation in each crystal structure was allowed, so that the observed energy difference may well be the result of the importance of small molecular adjustments, in particular in the positions of some of the protons, which can be adjusted at the final stage of the energy evaluation. This latter step not only allows for all the structural parameters to be relaxed, but also employs solid state DFT methods, which can give more accurate relative energies. Such calculations were performed for all CAT and EPI structures (both solvated and unsolvated) found in the best 15 kJ/mol energy window. The results are presented in Figure 10.

After DFT re-optimization, the landscape for EPI does not change much, at least in terms of structure ranking. What changes though, is the energy range, and as a result also the energy difference between particular structures. On the PBE-D2 landscape, the experimental structure remains the lowest in energy and is found to be 8.75 kJ/mol more stable than the next best structure, which is a predicted 1:1 methanol solvate, while the best 2:1 methanol solvate has total energy 21.4 kJ/mol above the experimentally observed (unsolvated) structure. These results further support the thermodynamic preference of the unsolvated structures of EPI over the solvated ones, regardless of the solvent ratio. As for the landscape of CAT, the change upon DFT re-optimization is more significant (Fig. 10, lower panel), mostly due to much smaller energy differences between the structures. After re-ranking, a large number of solvated structures are found noticeably more stable than the lowest energy unsolvated CAT crystal structure, with the energy difference between the best solvate and the best unsolvated structure equal to 10.5 kJ/mol. The energy ranking was also improved after DFT re-optimization; the experimental structure of CAT solvate (marked with a cross in Figures 9 and 10) is now ranked as the third most stable structure including all stoichiometries, 3.0 kJ/mol higher in total energy than the best structure. That this structure is not the energetic global minimum may relate to the observation that the experimental methanol solvate of CAT seems to be metastable and converts very rapidly to a differently solvated form of CAT, containing half a molecule of methanol and one molecule of water per CAT molecule. The exact structure of this new solvate – hydrate remains unclear, as it does not form crystals appropriate for the single crystal X-Ray diffraction measurements, and it will be a subject of further studies.

In terms of the prediction of solvate stoichiometry, the obtained results do not present as clear a picture as the comparison of solvated and unsolvated structures. The global minimum for the DFT landscape for CAT is a 2:1 solvate (although it does not correspond to the experimentally observed structure), but is only 0.82 kJ/mol lower in energy than the next best structure, which is a 1:1 solvate, with the experimental 2:1 solvate, as mentioned before, being the third lowest energy structure. There are several possible explanations for the remaining slight disagreement between experiment and prediction: (i) the inaccuracy of the DFT method, especially the D2 dispersion correction scheme, (ii) the approximate nature of the energy correction related to moving methanol from its pure liquid state to the solvate crystal structures (eq. 2), (iii) neglect of entropy contributions to relative energies, which can give free energy contributions as large as the lattice energy differences observed here,[41] and (iv) the aforementioned metastability of the experimental methanol solvate.

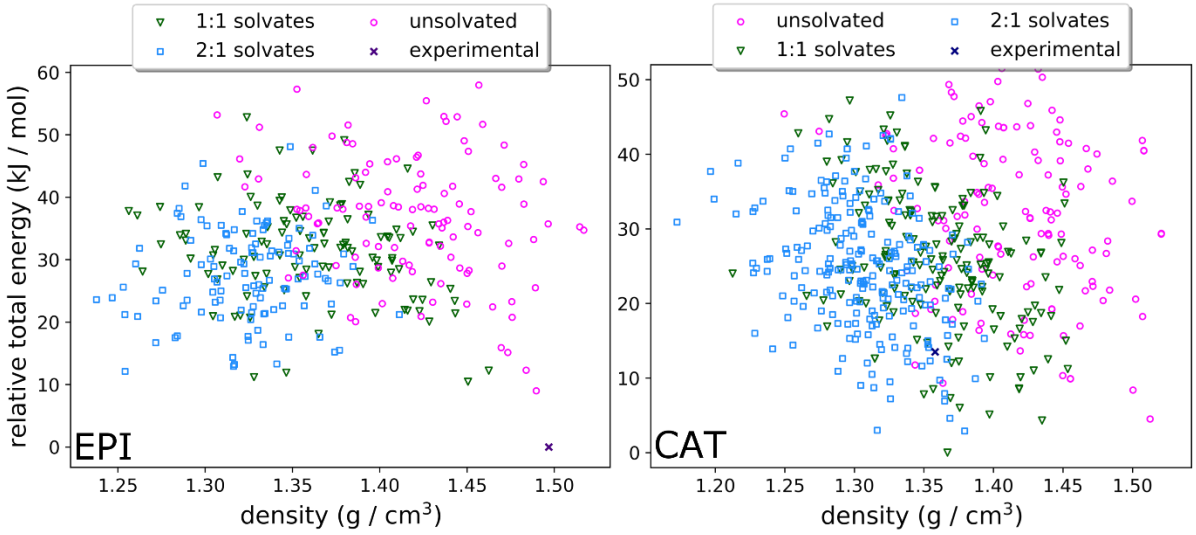


Figure 9. A comparison of crystalization preferences for EPI (left) and CAT (right).Energies are calculated with the FIT + DMA force field and PCM polarization model. Solvate structures are corrected for their energetic cost of removing methanol from its pure liquid phase (see eq. 2). Please note slightly different scales at the y axes of EPI and CAT.

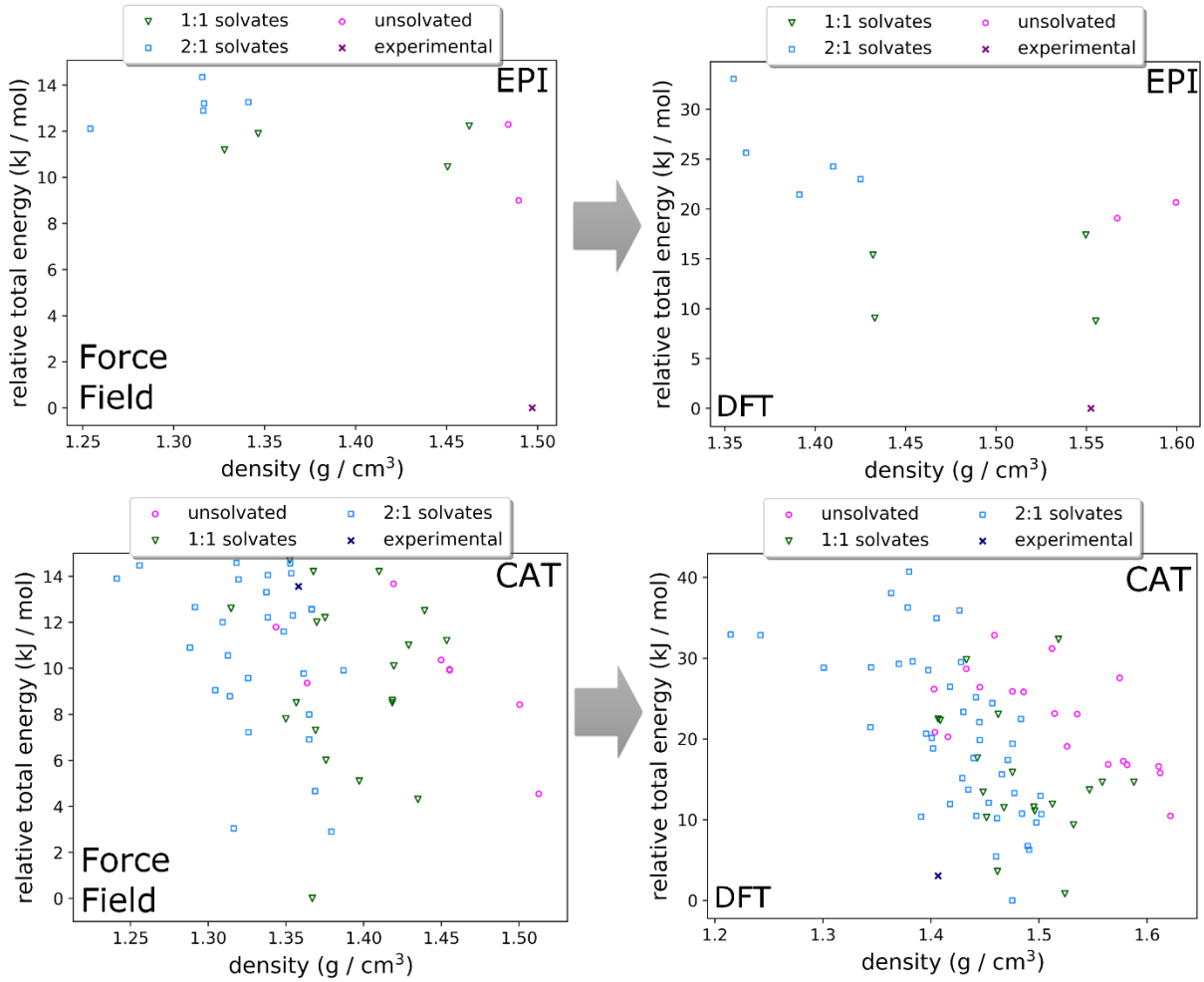


Figure 10. Re-ranking of the best predicted EPI (upper panel) and CAT (lower panel) crystal structures upon DFT (solid state PBE-D2) re-optimization.

CrystEngComm

ARTICLE

Conclusions

According to the results of crystal structure prediction calculations, the differences in crystallization preferences of two diastereosiomers examined in this work, EPI and CAT, can be atributted to different packing abilities of their high and low gas phase energy conformers. Our findings show a different aspect of crystallization preferences from the one published previously by Braun and Griesser,[12] for alkaloids strychnine and brucine. In the case of the polyphenols studied in this work, both seem to have similar propensities for forming solvated and unsolvated structures of similar densities, but for CAT only high energy molecular conformers are able to pack in such a manner. As a result, an additional energetic compensation is needed, and this can be provided by solvate formation. We note, however, that for one of the molecules, CAT, the energetic balance between unsolvated, 1:1 solvates and 2:1 solvates is finely balanced and computational results that are consistent with experimental observations only became fully visible after DFT re-optimization of the predicted crystal structures. In terms of the prediction of solvate stoichiometry, although it still remains challenging to correctly rank the best structures, a general tendency of 2:1 solvates being more energetically favourable than 1:1 solvates was found for CAT.

Our results also demonstrate the importance of including high energy molecular conformations in crystal structure prediction studies where improved intermolecular interactions can compensate for high conformational energies. This is the case here, where loss of an intramolecular hydrogen bond can be counterbalanced by more extensive intermolecular hydrogen bonding. This applied in particular to the methanol solvates of CAT, where inclusion of methanol within the crystal lattice is required to exploit the better intermolecular hydrogen bonding capacity of its high energy conformations, leading to a stable 2:1 methanol solvate formed by a molecular conformer 33 kJ/mol less stable than the most stable CAT conformer.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was carried out with the kind financial support of the Polish Ministry of Science and Higher Education, as a part of the fellowship 'Mobility Plus' (grant No 1664/MOB/V/2017/0). Interdisciplinary Centre for Mathematical and Computational Modelling (computational grant G 67-11) is kindly acknowledged for providing computational resources. The IRIDIS High Performance Computing Facility, with associated support services at the University of Southampton are gratefully acknowledged for providing computational resources. We are also grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1).

Notes and references

1 S. M. Woodley and R. Catlow, *Nature Materials*, 2008, **7**, 937–946.
2 G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2005, **5**, 1023–1033.
3 M. Habgood, I. J. Sugden, A. V. Kazantsev, C. S. Adjiman, and C. C. Pantelides, *J. Chem. Theory Comput.*, 2015, **11**, 1957–1969.
4 J. van de Streek and M. A. Neumann. *CrystEngComm*, 2011, **13**, 7135–7142.
5 H. C. Stephen Chan, J. Kendrick, and F. J. J. Leusen, *Angew. Chem. Int. Ed.*, 2011, **50**, 2979–2981.
6 A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio Jr, A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu and C. R. Groom, *ActaCrystallogr. Sect. B*, 2016, **72**, 439–459.
7 A. J. Cruz-Cabeza, S. Karki, L. Fábián, T. Friščić, G. M. Day and W. Jones, *Chem. Commun.*, 2010, **46**, 2224–2226.
8 A. J. Cruz-Cabeza, G. M. Day and W. Jones, *Chem. Eur. J.*, 2008, **14**, 8830–8836.
9 D. E. Braun, P. G. Karamertzanis and S. L. Price, *Chem. Commun.*, 2011, **47**, 5443–5445.
10 D. E. Braun, S. R. Lingireddy, M. D. Beidelschies, R. Guo, P. Müller, S. L. Price and S. M. Reutzel-Edens. *Cryst. Growth Des.* 2017, **17**, 5349−5365.
11 D. E. Braun and U. J. Griesser, *Front Chem,* 2018, 6:31. doi: 10.3389/fchem.2018.00031.
12 D. E. Braun and U. J. Griesser, *Cryst Growth Des*, 2016, **16**, 6405−6418.

13 D. Ferreira and R. Bekker, *Nat. Prod. Rep.*, 1996, *13*, 411–433
14 C. A. Rice-Evans, N. J. Miller and G. Paganga, *Free Radic. Biol. Med.*, 1996, **20**, 933–956.
15 T. P. T. Cushnie and A. J. Lamb, *Int. J. Antimicrob. Agents* 2005, **26**, 343–356.
16 S. Mandel and M. B. H. Youdim, *Free Radical Biol. Med.*, 2004, **37**, 304–317.
17 M. N. Ghayur, H. Khan, and A. H. Gilani. *Arch. Pharm. Res.*, 2007, **30**, 970–975.
18 H. Tsutsumi, Y. Kinoshita, T. Sato and T. Ishizu, *Chem. Pharm. Bull.*, 2011, **59**, 1008—1015.
19 J. K. Harper, J. A. Doebbler, E. Jacques, D. M. Grant and R. B. Von Dreele, *J. Am. Chem. Soc.*, 2010, **132**, 2928–2937.
20 M. K. Dudek, A. Jeziorna, M. J. Potrzebowski, CrystEngComm 2016, **18**, 5267–5277.
21 Spartan16, Wavefunction, Inc, 2016, Irvine USA.
22 T. A. Halgren, *J. Comp. Chem.*, 1996, **17**, 490–519.
23 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, GAUSSIAN 09 (Revision D.01), Gaussian, Inc., Wallingford, CT, 2010.
24 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
25 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comp. Chem.*, 2011, **32**, 1456–1465.
26 D. H. Case, J. E. Campbell, P. J. Bygrave and G. M. Day, *J. Chem. Theory Comput.*, 2016, **12**, 910–924.
27 S. L. Price, M. Leslie, G. W. A. Welch, M. Habgood, L. S. Price, P. G. Karamertzanis and G. M. Day, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8478–8490.
28 A. J. Stone, *J. Chem. Theory Comput.*, 2005, **1**, 1128–1132.
29 A. J. Stone, Distributed multipole analysis of Gaussian wavefunctions. GMDA version 2.2.11
30 D. S. Coombes, S. L. Price, D. J. Willock and M. Leslie, *J. Phys. Chem.*, 1996, **100**, 7352–7360.
31 M. T. Kirchner, D. Das and R. Boese, *Cryst Growth Des*, 2008, **8**, 763–765.
32 D. P. Mcmahon, A. Stephenson, S. Y. Chong, M. A. Little, J. T. A. Jones, A. I. Cooper and G. M. Day, *Faraday Discussions*, 2018, DOI: 10.1039/C8FD00031J
33 A. Pulido, L. Chen, T. Kaczorowski, D. Holden, M. A. Little, S. Y. Chong, B. J. Slater, D. P. McMahon, B. Bonillo, C. J. Stackhouse, A. Stephenson, C. M. Kane, R. Clowes, T. Hasell, A. I. Cooper and G. M. Day, *Nature*, 2017, **543**, 657–664.
34 S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson and M. C. Payne, *Zeitschrift fuer Kristallographie* 2005, **220**, 567–570.
35 S. J. Grimme, *Comput. Chem.*, 2006, **27**, 1787-1799.
36 H. P. G. Thompson and G. M. Day, *Chem. Sci.*, 2014, **5**, 3173–3182.
37 J. A. Chisholm and S. Motherwell, *J. Appl. Cryst.*, 2005, **38**, 228–231.
38 M. K. Dudek (Jamróz), S. Kaźmierski, K. Stefaniak, V. B. Gliński and Jan. A. Gliński, *Org. Biomol. Chem.*, 2014, **12**, 9837–9844
39 N. Hayashi and T. Ujihara, *J. Nat. Prod.*, 2017, **80**, 319–327
40 J. Nyman, O. S. Pundyke and G. M. Day, *Phys. Chem. Chem. Phys.,* 2016, **18**, 15828–15837.
41 J. Nyman and G. M. Day, *CrystEngComm*, 2015, **17**, 5154-5165.