DALHOUSIE UNIVERSITY

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Crystal-Energy Landscapes of Active Pharmaceutical Ingredients Using Composite Approaches†

Luc M. LeBlanc,* and Erin R. Johnson*

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

The crystal structure prediction (CSP) of organic molecular solids remains challenging, as the demand to predict more complex crystal structures increases. Low-cost (semi-)empirical methods are now more commonly being used to unburden the computational bottlenecks in intermediate stages of a CSP protocol, but remain inadequate for the energy ranking of crystal structures. With the use of "composite approaches", however, these relative energies can be refined with higher levels of theory at the cost of a single-point energy calculation, provided that the low-level geometries are amenable to such purposes. Herein, a composite method making use of the B86bPBE-XDM density functional, and combining a low-level small-basis method using finite-support numerical orbitals with high-level plane-wave calculations, is applied to predict crystal-energy landscapes of four active pharmaceutical ingredients: 5-fluorouracil, naproxen, carbamazepine, and olanzapine. Results show that this composite method can aid in resolving realistic energy landscapes of drug-like molecules, consistently placing the experimentally isolable polymorphs as the lowest-energy structures and, in addition, providing a sound stability ordering of polymorphs, in reasonable agreement with available experimental data. This is in stark contrast to the energy rankings provided by previously reported refined anisotropic force-field data or results obtained from other small-basis set approaches, such as sHF-3c. While the B86bPBE-XDM composite method is generally more expensive than (semi-)empirical approaches, it does not rely on any parameter fitting or tuning with regards to a given system of interest, making it a suitable and more generally applicable alternative for CSP.

1 Introduction

The phenomenon of polymorphism is ubiquitous in nature and is exhibited by many organic solids of importance to the development of active pharmaceutical ingredients,[1,2] microporous materials,[3] semiconductors,[4,5] explosives,[6–8] pigments and dyes.[9,10] The ability to predict which polymorph will be formed under a given set of experimental conditions, as well as defining its stability with respect to other crystalline forms, has been of growing interest and concern in many scientific communities,[11–22] most notably in the pharmaceutical industries.[1,18,23,24] However, predicting the molecular crystal structure of a given compound solely from computational grounds, or 'crystal structure prediction' (CSP), is currently a great challenge in computational chemistry.[22,25]

A practical CSP algorithm must efficiently and accurately sample a high-dimensional crystal-energy landscape of the atomic positions and crystal lattice parameters to locate the minimum-energy structures. Stringent limitations on the choice of computational methods[22] for CSP are imposed by the increasing demand to study larger and more complex systems.[26,27] If executed properly, one can benefit from using multilevel methodologies to perform CSP studies more efficiently. In this fashion, one can gradually increase the accuracy of the computational method used, while concomittantly reducing the initial pool of candidate structures to more manageable numbers, to hone in on the low-energy parts of the landscape. While density-functional tight-binding (DFTB)[28] or small-basis-set (e.g., HF-3c)[29,30] approaches are popular low-cost methods for CSP purposes, their energetic accuracy for the intermediate stages of a search is often sub-optimal.[31] This requires a larger subset of structures to be brought forward in the CSP protocol[32] to avoid experimentally isolable structures being thrown out during the refinement steps.[22] However, low-cost methods can yield reasonably accurate geometries,[31,33,34] particularly if the dispersion correction parameters are fitted not only to energies, but also to reproduce benchmark geometries of molecular dimers[32,33] and/or geometries of molecular crystals.[30,34]

Department of Chemistry, Dalhousie University, 6274 Coburg Road, PO Box 15000 Halifax, Nova Scotia, Canada B3H 4R2. E-mail: erin.johnson@dal.ca, luc.leblanc@dal.ca

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

In previous work,[35] we made use of composite approaches with small-basis-set methods (sHF-3c and PBE-D2/DZP), followed by high-level plane-wave DFT single-point energy calculations (e.g., B86bPBE-XDM) to examine and assess their ability to provide absolute and relative lattice energies of small molecular crystals. These composite approaches were then applied to the previously reported CSP for a chiral organic semiconductor, 1-aza-[6]-helicene, and were found to reduce the computational cost of generating accurate electronic-energy landscapes, relative to full plane-wave DFT calculations. From our previous experience on small molecular solids, and moreso for the systems presented herein, the composite approach (including the high-level single-point energy calculation) can reduce the computational time by one to two orders of magnitude relative to full plane-wave DFT calculations alone, depending on the choice of low-cost method. The composite methods were found to work best when the geometries were transferable between the low- and high-level methods. This prompted us to implement B86bPBE-XDM in the numerical-basis SIESTA code,[36,37] to allow use of the same functional for both low- and high-level calculations.[38]

Others have also recently applied similar ideaologies to perform CSP studies. In the work by Iuzzolino and co-workers,[32] a reparametrized version of the D3-dispersion-corrected DFTB method allowed for low-cost optimization of several thousands of candidate structures generated during CSP studies of drug-like molecules (compounds **XX**, **XXIII**, and **XXVI** of the most recent CCDC blind tests,[22,39] two tautomers of mebendezole, and a lead compound in a pharmaceutical study). While DFTB did not improve on the initial ranking from the CrystalPredictor-generated structures, the geometries were amenable to higher-level single-point calculations using a multipole-based atom-atom force field. These single-point energy calculations aided in placing the experimentally isolated polymorphs in the low-energy regions of the CSP landscapes. The success of this DFTB reparametrization relies on refitting dispersion damping parameters to minimize errors in geometries of small molecular dimers or molecular crystals within the X23 set,[40,41] rather than minimizing energies.[32–34] It would seem that this reparametrization of DFTB would be the best choice for a low-cost method in a composite approach, given the speedup it can offer in pre-optimizing geometries. However, several challenges to its use in CSP remain. For instance, in the study by Iuzzolino and co-workers,[32] DFTB was unable to provide an improved energy landscape relative to the force fields used in crystal structure generation. While DFTB improved the agreement of computed crystal geometries with experimental X-ray data, it generally caused the experimentally isolated polymorphs to be destabilized significantly with respect to other candidate structures on the energy landscape, meaning that a larger number of structures would need to be considered using the composite approach. It should be noted that the DFTB method also altered the covalent bonding in some situations.

In another study,[42] the use of the sHF-3c method as a cheap alternative to full DFT geometry optimizations allowed the unit-cell volumes of zeolites to be reproduced within 2% of experimental values. Additional high-level dispersion-corrected DFT energy calculations were then used to improve upon the sHF-3c results, reducing the errors in formation enthalpies from ca. 5.9 kJ/mol to 0.8 kJ/mol per silica unit for the quartz-based zeolites. Other fragment-based approaches using a combination of low- and high-cost methods to efficiently compute lattice energies of the small molecular crystals of the X23 set,[31] or applying machine-learning methods to correct energies of crystal structures on landscapes generated via anisotropic atom-atom force-field methods,[43] have been developed. In the latter study, two-body interaction energies within a fragment-based method initially computed by force fields were corrected with high-level methods (such as DFT or MP2) for a series of polymorphic structures of small organic molecules (oxalic acid, maleic hydrazide, and tetrolic acid, to name a few). However, the large number of dimer calculations and corrections applied with high-level methods prompted the authors to develop a machine-learned variant, capable of reproducing the high-level corrections much more cheaply, and improving the accuracy of the relative energies of these polymorphs on generated crystal-energy landscapes.

Herein, the robustness of composite approaches utilizing small-basis set methods (in particular, B86bPBE-XDM/DZP and sHF-3c) in combination with plane-wave DFT calculations (B86bPBE-XDM/PAW) to efficiently produce accurate energy landscapes is validated for a selection of active pharmaceutical ingredients (APIs): 5-fluorouracil, naproxen, carbamazepine, and olanzapine. Each of these have several experimentally characterized polymorphs and have been the subject of previous CSP studies.[44–47] The work presented herein focuses solely on energy-ranking methods, and not on the generation of candidate structures. The initial sets of structures were obtained from the Control and Prediction of the Organic Solid State (CPOSS) database,[47] courtesy of Prof. Sally L. Price and Dr. Louise S. Price, University College London. The results presented herein show that the use of composite approaches can lead to the generation of more realistic energy landscapes for pharmaceutical-like molecules, in contrast to previous force-field calculations. The experimentally isolated polymorphs of the drugs are generally placed as the lowest in energy and the proper experimental stability ordering between polymorphs is well described (within the tolerance for thermal effects). Ultimately, these composite approaches could also serve as a means to identify the limitations of (semi-)empirical methods, by further examining the candidate structures which are most affected by reranking.

2 Computational Methods

CPOSS database structures: Three of the four APIs selected, 5-fluorouracil,[44] naproxen,[45] and olanzapine,[46] have had their crystal-energy landscapes published, while the data for carbamazepine is currently unpublished. All of the force-field energy rankings were obtained from empirically fitted dispersion-repulsion potentials, supplemented with distributed multipole analysis (DMA) for the description of electrostatics, derived from post-Hartree-Fock (e.g., MP2) wavefunction or DFT methods.[48] Details can be found in each of the references pertaining to the initial CSP studies of the APIs. Briefly, (i) 5-fluorouracil:[44] crystal structures were generated with MOLPAK, using MP2/6-31G(d,p) optimized molecular structures; lattice energy calcula-

tions were performed with DMAREL keeping the molecules completely rigid throughout. (ii) Naproxen:[45] a rigid molecule search in MOLPAK was conducted, and further refinements were performed with DMAflex-2, which allows minimization with some molecular flexibility, i.e., alternating between geometry optimizations and charge density calculations with GAUSSIAN and lattice-energy minimizations with DMACRYS. (iii) Olanzapine:[46] a flexible molecule search was performed with CrystalPredictor-I, then lattice energies were computed and refined with CrystalOptimizer using a DMA-refined force-field. (iv) Carbamazepine:[47] a similar procedure to what was used for olanzapine was performed, and is detailed in Ref. 49.

The low-cost and high-level methods used for the composite approaches taken in this work are now detailed.

SIESTA calculations: B86bPBE[50,51]-XDM,[52] with the DZP basis set, as implemented in an in-house version of the 4.0b-485 SIESTA code,[36,37] was used for the low-cost calculations. Similar parameters to what were used in previous work[35,38] are also used here. The "energy shift" parameter controlling the the confinement radius of the orbitals was set to 0.001 Ry, the real-space integration-grid cutoff value was set to 400 Ry, and Trouillier-Martins-type[53,54] norm-conserving pseudopotentials[55,56] generated with the ATOM code[57] were utilized. The XDM damping parameters, a_1 and a_2, were set to 0.5000 and 2.5556 Å, respectively.[38] The Broyden optimizer was used for geometry relaxations, and convergence thresholds for forces and stresses were set to 0.04 eV/Å and 0.02 GPa, respectively.

CRYSTAL calculations: sHF-3c,[30,58,59] as implemented in the CRYSTAL17[60] code, was also used for the low-cost calculations. sHF-3c makes use of a minimal basis set (MINI(x)[58,59]) and is supplemented with Grimme-D3 dispersion,[61,62] a geometrical counterpoise correction,[63,64] and a correction for short-range basis-set incompleteness errors.[30,58,59] Becke-Johnson damping was used for the D3 dispersion model,[62,65] three-body interactions were included, and a fit parameter of $s_8 = 0.6143$ was applied.[30]

Plane-wave calculations: High-level plane-wave B86bPBE-XDM calculations using projector-augmented wave (PAW) data sets[66] were performed with the Quantum ESPRESSO[67] code, version 5.1. For single-point energy calculations, energy cutoffs for the wavefunction and density were set to 60 Ry and 600 Ry, respectively. For the full geometry relaxations of 5-fluorouracil, these two parameters were set to 80 Ry and 800 Ry, respectively. The XDM damping parameters, a_1 and a_2, were set to 0.6512, and 1.4633 Å, respectively. When structures were fully relaxed (e.g., in the case of 5-fluorouracil), convergence thresholds for energies and forces of 10^{-5} Ry and 10^{-4} Ry/bohr, respectively, were used.

k-point grid sampling: A $4\times4\times4$ MP **k**-point mesh sampling of the Brillouin zone was used for all calculations.

Similarity metric between crystal-energy landscapes: In order to compare the degree of similarity between crystal-energy landscapes a similarity index, S, is indicated on each of the panels of the figures presented. The value of this index is given by

$$S = \frac{1}{N}\sum_{i=1}^{N}\sqrt{(\Delta E_{\text{low}} - \Delta E_{\text{high}})^2}, \quad (1)$$

where the sum runs over the N data points (ca. 50-60) present on a crystal-energy landscape. Overall, the index represents a mean deviation between ΔE's on two different energy landscapes, from low and high levels of theory, with the high level serving as the reference landscape. ΔE_{low} and ΔE_{high} are the energy differences between a given data point on the landscape and the mean energy of all the points on this same landscape. The mean energy is chosen to eliminate dependance (or sensitivity) on an arbitrary choice of reference point. The larger this index, the more dissimilar the points are with respect to their corresponding points on the reference landscape. A value of zero indicates an exact match.

Relative energies and densities of crystal structures from the generated energy landscapes for 5-fluorouracil, naproxen, carbamazepine, and olanzapine can be found in the Electronic Supplementary Information (ESI).†

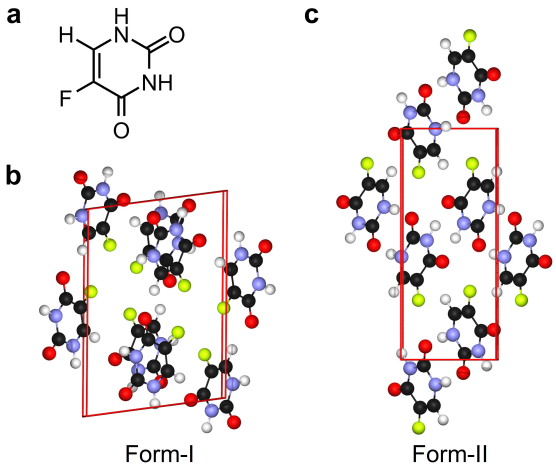
3 Results and Discussion

3.1 5-Fluorouracil

5-Fluorouracil, first synthesized in 1957 and used as an anticancer API for over forty years,[68,69] forms two known polymorphs (Figure 1) to date. Its first crystal structure (Form-I) was solved in 1973,[70] but a second form (Form-II) was predicted by computational means and subsequently isolated thirty years later.[44] From this initial CSP study,[44] Form-II was predicted via force-field computations to be the thermodynamic minimum, being ca. 6 kJ/mol more stable than Form-I at 0 K. However, experimental evidence from thermal analysis measurements, reported in that same study, pointed to the opposite stability ordering and supported a monotropic relationship between the two forms. That is, Form-I was found to be more stable than Form-II, and no phase transitions were observed upon differential scanning calorimetry measurements between 298 K and the melting point of each form. The fitted force-field potential was able to reproduce the geometries of crystal structures at room temperature within reasonable limits, given that thermal expansion was neglected. However, the neglect of thermal expansion was not deemed to be the main reason for the disagreement between the experimental and computational results, as the quality of the computed lattice energies did not improve upon comparing to low-temperature (150 K) experimental crystal structures. Instead, it was suggested that the force field would not necessarily describe the relative lattice energies of these two forms of 5-fluorouracil properly, because it was fitted for geometries using a set of molecules containing only a small number of perfluorohydrocarbons.

The prediction of a stable Form-II of 5-fluorouracil prompted several investigations into why only Form-I had been isolable for several decades past. In particular, as this second form could only be prepared in dry nitromethane, it was postulated that the solvation of 5-fluorouracil could affect the mechanisms of nucleation and crystal growth of both forms.[44] Subsequent molecular dynamics simulations showed that the presence of water in organic solvents, or of water as a solvent itself, could affect which polymorph of 5-fluorouracil would be kinetically favoured.[71,72] Ultimately, the hydrogen bonding of water

Fig. 1 5-Fluorouracil and its two known polymorphs. Form-I is the more stable experimentally at 298 K.



molecules to the C=O and N–H sites leaves the 5-fluorouracil molecules open to self-associate via F–F interactions only (as in Form-I, Figure 1(b)). Conversely, in the absence of water, more weakly-interacting solvents (such as nitromethane) do not disallow formation of hydrogen-bonded dimers between the parent 5-fluorouracil molecules, which translates to yield Form-II (Figure 1(c)) in the solid state.

The computational evidence for kinetically-controlled prenucleation of 5-fluorouracil has offered a more complete understanding of crystallisation of 5-substitued uracils.[73] Yet, despite additional efforts in establishing free-energy differences between the two polymorphs, and accounting for thermal expansion through the use of classical molecular-dynamics computations, Form-II has continually been predicted to be more stable than Form-I.[72,74] This remains in disagreement with the experimental evidence reported when Form-II was first isolated/solved.[44]

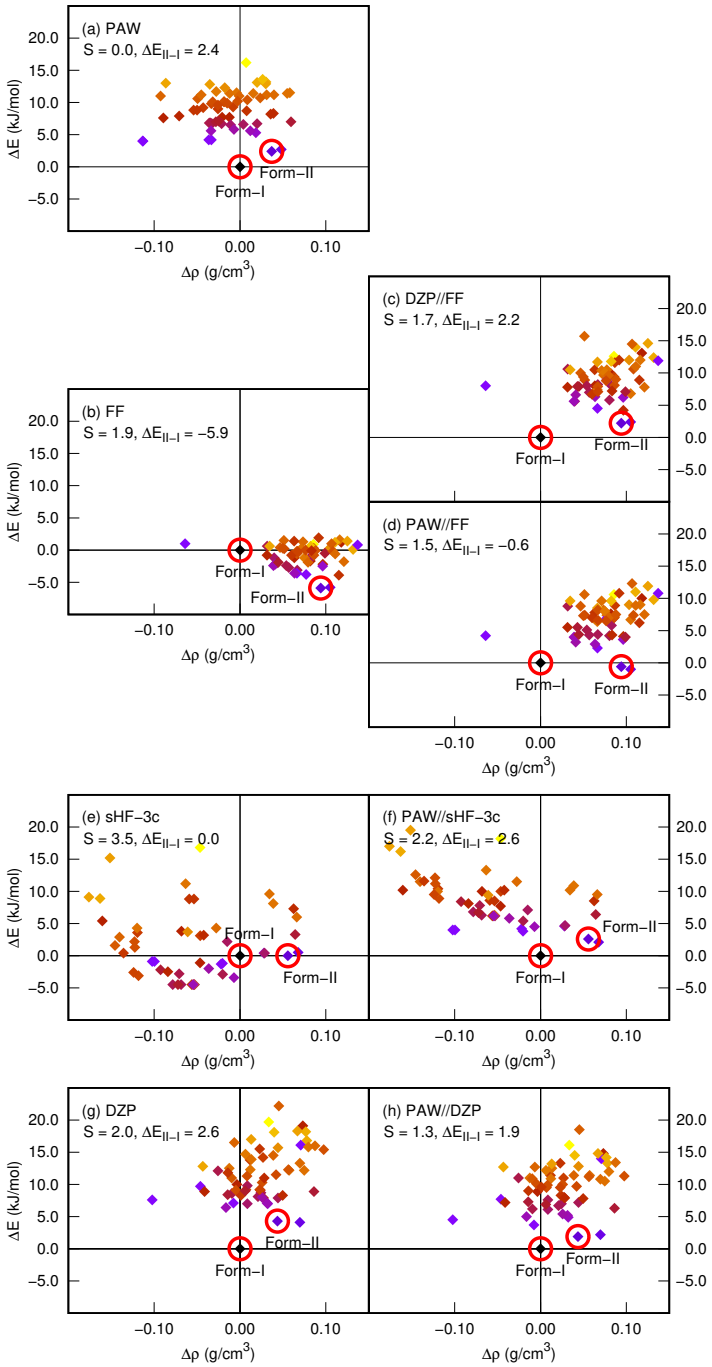
The force-field ranking results reported in Ref. 44 for the crystal-energy landscape are presented in Figure 2(b). Figure 2(a), on the other hand, shows the reranking of crystal structures with fully-relaxed geometries obtained from plane-wave DFT calculations (B86bPBE-XDM). In contrast to the force-field results, Form-I is now predicted to be more stable than Form-II by 2.4 kJ/mol, which (while these results neglect thermal/entropic effects) recovers the same relative stability of the crystals forms as experiment.[44] A possible reason that the force fields utilized in earlier studies predict the incorrect relative energy ordering of Forms I and II could be an inability to describe halogen–halogen interactions,[75–77] which impact the stability of Form-I of 5-fluorouracil.[78]

Gas-phase geometry optimizations were carried out for the tetrameric arrangement of molecules present in Form-I of 5-fluorouracil, and the equivalent arrangement for uracil molecules (Figure 3). To retain the same symmetry found in the crystal, the geometry of the tetramers was fixed to the S_4 point group. The optimizations were performed using the BHandHLYP-XDM functional[79–81] and aug-cc-pVDZ basis set, with the GAUSSIAN 09 software package[82] and the postg code.[83] This functional has previously been shown to provide a good description of halogen bonding.[84] Subsequent single-point energy calculations with BHandHLYP-XDM/aug-cc-pVTZ indicate binding energies (per molecule) of -35.4 and -29.1 kJ/mol for these two systems, respectively. Only a small energy difference ($<$ 3 kJ/mol) in intermolecular binding was computed for the 5-fluorouracil tetramer when comparing the optimized gas-phase geometry and the rigid, crystallographic arrangement (-38.4 kJ/mol). The majority of the binding interactions are due to the base functional used, with the XDM dispersion correction representing only -7.2 and -5.2 kJ/mol per molecule, respectively. It is rather unlikely that the ca. 6 kJ/mol difference in binding energy between the 5-fluorouracil and uracil systems is due to halogen bonding between the fluorine atoms, given that these interactions are weakest for fluorine,[85] and that the interatomic distance between fluorines (ca. 2.94 Å) is at the edge of what is typically observed for fluorine–fluorine halogen bonding (ca. 2.95 Å).[78] Instead, the larger binding energy observed in 5-fluorouracil is likely due to the C-H bonds being more acidic, due to the fluorine atom withdrawing electron density, which results in stronger C–H$\cdots$O interactions between 5-fluorouracil molecules.

Having shown that plane-wave DFT calculations can afford a crystal-energy landscape for 5-fluorouracil that agrees with available experimental findings, focus is now shifted to whether composite approaches can reproduce similar-quality landscapes. As has been shown in earlier work,[35,38] it is important to choose low-cost methods that will produce geometries of similar quality to the higher-level methods (here, plane-wave DFT), so that the application of a single-point energy calculation on the “cheaply" optimized crystal structures yields reliable relative energies. The best low-cost methods for use in composite approaches with plane-wave B86bPBE-XDM methods were previously found to be sHF-3c and PBE-D2/DZP.[35] This prompted the implementation of B86bPBE-XDM in SIESTA.[38] It is expected that these geometries will generally be more compatible with plane-wave B86bPBE-XDM, given the improved energetics obtained over PBE-D2 with DZP and because there will be consistency in the choice of functional and dispersion correction between low- and high-level calculations. The small-basis B86bPBE-XDM method alone (Figure 2(g)) affords a more realistic energy landscape than that offered by the force-field calculations (Figure 2(b)), given that both experimentally observed crystal forms fall within the lower-energy regions. However, this is not the case for sHF-3c (Figure 2(e)), which gives many lower-energy structures than the experimentally isolated forms at larger cell volumes. An improved description of the crystal-energy landscapes is obtained by performing single-point energy calculations with B86bPBE-XDM using a plane-wave basis set, as shown in Figure 2(d,f,h). In all three cases, performing these single-point energy calculations restores energy differences between the two isolable crystal forms in very close agreement to the full plane-wave DFT results.

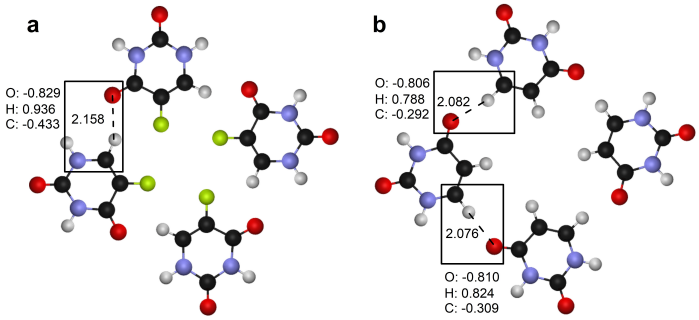
The overall agreement between energy landscapes can be observed qualitatively by taking into account the heat map, where each data point in Figure 2(a)-(h) has its color fixed to that of the corresponding point for the full plane-wave results in Figure 2(a). Alternatively, a distance metric is also presented for each panel

Fig. 2 Crystal-energy landscapes of 5-fluorouracil. Energy landscapes were generated via (a) plane-wave B86bPBE-XDM calculations (PAW); (b) a force-field (FF, data obtained from the CPOSS database[44,47]); and two small- and/or minimal-basis low-cost methods, (e) sHF-3c and (g) B86bPBE-XDM/DZP (DZP). Crystal structures within panels (b,e,g) were then reranked with single-point energy calculations using plane-wave B86bPBE-XDM to give panels (d,f,h), respectively. The force-field structures within panel (b) have also been reranked with single-point energy calculations using B86bPBE-XDM/DZP to give panel (c). Data points that are encircled are labelled according to the experimentally isolated polymorphs with which they are equivalent, i.e., Forms I and II. The computed energies, ΔE, and densities, $\Delta\rho$, are expressed relative to Form-I for all methods. The similarity index, S, and the heat-map colouring are relative to the crystal-energy landscape in panel (a). $\Delta E_{\mathrm{II-I}}$ is the energy difference between the two isolable polymorphs, relative to Form-I.



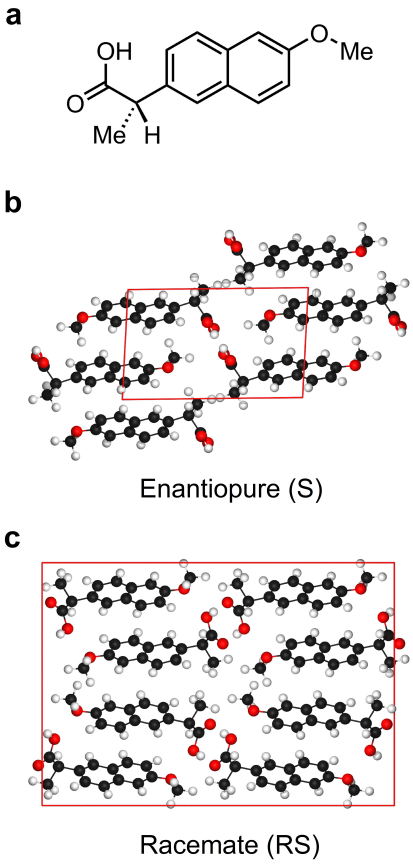
to give a more quantitative descriptor of similarity between the low-level, or composite, and full plane-wave crystal-energy landscapes. Interestingly, it is clear from this distance metric that the force-field energy ranking alone yields a similar-quality landscape to that of plane-wave DFT calculations, given that the relative energies are only 1.9 kJ/mol away from the reference landscape on average. This is on par with the average deviation computed for the landscape generated with the small-basis B86bPBE-XDM method (2.0 kJ/mol). The major issue encountered with the force-field ranking is clearly its inability to describe the proper

Fig. 3 Tetramers of (a) 5-fluorouracil and (b) uracil obtained from geometry optimizations using the BHandHLYP-XDM functional[79–81] (see text). Bond lengths (in Å), indicated by dashed lines, and Mulliken charges (in atomic units), indicated for the boxed atoms, are shown.



energies for systems where close halogen–halogen contacts are important, as in Form-I.[44] The sHF-3c landscape is comparatively of poorer quality than the other two methods, with a distance metric from the reference landscape of 3.5 kJ/mol. However, application of DFT single-point energy calculations, in general, does reduce deviations from the reference plane-wave DFT landscape.

Fig. 4 Naproxen and its known enantiopure and racemate crystal forms. The racemate form is the more stable experimentally.



Overall, although more computationally expensive than the force field and sHF-3c methods, B86bPBE-XDM/DZP gives geometries that seem to be most compatible with plane-wave B86bPBE-XDM, making it more amenable to be combined with this particular high-level method in a composite approach. That being said, in order to remain competitive with other approaches making use of the faster DFTB or sHF-3c methods, one could instead resort to pre-screening, by performing single-point energy calculations with B86bPBE-XDM/DZP on the force-field geometries (Figure 2(c)). The force-field geometries in this case are amenable to a composite approach with the small-basis DZP method, and reproduce an energy landscape of similar quality to the one where full DZP optimizations were performed (Figure 2(g)), with Form-I again predicted to be more stable than Form-II. Then, a subset of low-energy structures from the composite DZP//FF energy landscape (Figure 2(c)) could be subjected to full optimization with DZP (Figure 2(g)) and subsequent single-point energy calculations with plane-wave basis sets (Figure 2(h)).

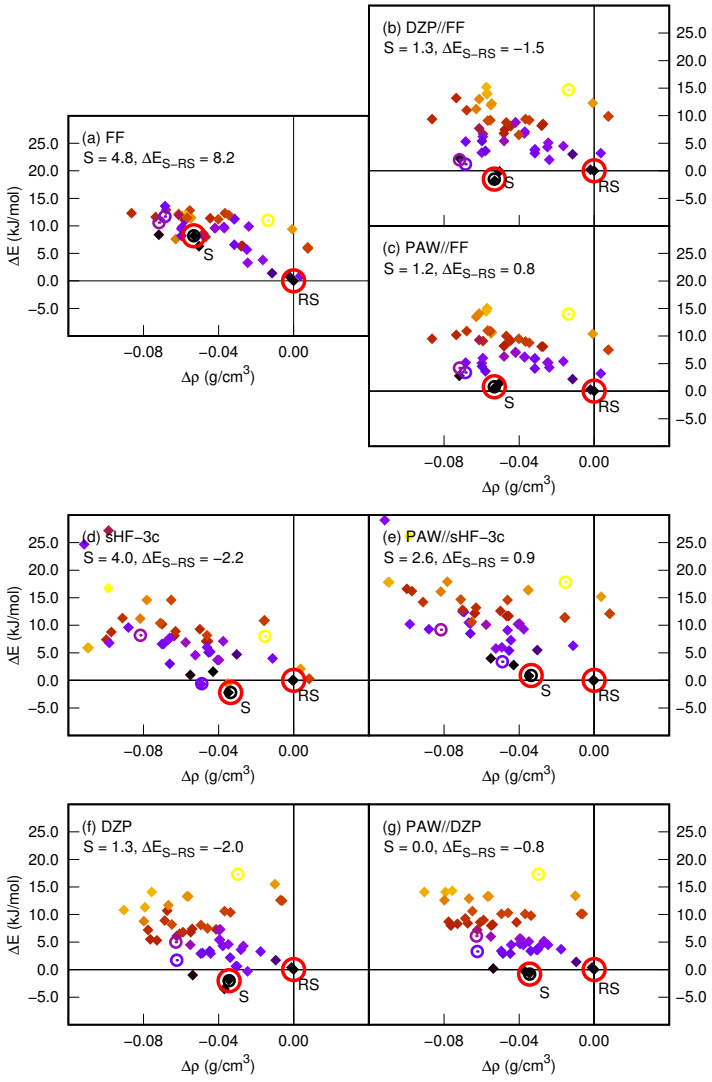
Because the B86bPBE-XDM composite approach, making use of DZP geometries followed by plane-wave single-point energies, has afforded the best results so far (including the study of small molecular solids),[38] this method will be used as a reference for establishing energy landscapes for the following APIs: naproxen, olanzapine and carbamazepine. Performing full plane-wave DFT calculations for these molecules' crystal-energy landscapes would be extremely costly, as their unit cells contain several hundred atoms. Thus, it is now shown that physically reasonable energy landscapes, agreeing with available experimental evidence, can be obtained via composite approaches, without having to rely on high-level geometry optimizations.

3.2 Naproxen

The (S)-enantiomer of naproxen is marketed as a non-steroidal anti-inflammatory drug. Its enantiopure crystal structure was first reported in 1985,[86] and refined two years later (Figure 4).[87] In 2011, the existence of a racemate crystal structure for naproxen was predicted by CSP and experimentally isolated.[45] The racemate and enantiopure crystal structures were ultimately found to have similar stabilities: (i) differential scanning calorimetry measurements determined the racemate form to have a similar melting point (within $< 1^{\circ}$C), while having a higher heat of fusion than the enantiopure form (by ca. 1.5 kJ/mol), and (ii) solubility measurements determined the relative heats of solution to be in favor of the racemate by ca. 2.4 kJ/mol. Force-field computations of the relative lattice energies estimated the racemate form as being 6.1 to 9.2 kJ/mol more stable than the enantiopure crystal form, depending on the quality of the method used to describe intra- and intermolecular forces. This overestimation of the relative energy difference has been attibuted to neglect of thermal expansion and entropic contributions.[45]

The crystal-energy landscape obtained from the previous force-field computations, reported in Ref. 45 and reproduced in Figure 5(a), predicts both the experimental racemate and enantiopure crystal forms as the overall minima of all candidate racemic and enantiopure crystal structures, respectively. An even more dramatic improvement in the quality of the energy landscape to what was observed for 5-fluorouracil is observed for naproxen when only B86bPBE-XDM/DZP single-point energy calculations,

Fig. 5 Crystal-energy landscapes of naproxen. Energy landscapes were generated via (a) a force-field (FF, data obtained from the CPOSS database[45,47]) and two small- and/or minimal-basis low-cost methods, (d) sHF-3c and (f) B86bPBE-XDM/DZP (DZP). Crystal structures within panels (a,d,f) were then reranked with single-point energy calculations using plane-wave B86bPBE-XDM (PAW) to give panels (c,e,g), respectively. The force-field structures within panel (a) have also been reranked with single-point energy calculations using B86bPBE-XDM/DZP to give panel (b). Racemate crystal structures are indicated by filled diamonds and enantiopure structures by open circles. Data points that are encircled are labelled according to the experimentally isolated polymorphs with which they are equivalent, i.e., the racemate and enantiopure forms. The computed energies, ΔE, and densities, $\Delta\rho$, are expressed relative to the racemate form, "RS", for all methods. "S" is the enantiopure form. The similarity index, S, and the heat-map colouring are relative to the crystal-energy landscape in panel (g). ΔE_{S-RS} is the energy difference between the two isolable crystalline forms, relative to the racemate form.



instead of full geometry optimizations, are performed on force-field geometries (Figure 5(b)). This indicates that an initial composite DZP//FF approach could be taken on a larger number of structures prior to moving forward only a subset of low-energy structures to the PAW//DZP composite approach. Making use of a composite approach with plane-wave DFT using the force-field geometries (Figure 5(c)) also gives the enantiopure and racemate crystal structures as the lowest-energy structures and maintains the correct ordering. While the energy difference between these two minima is overestimated with the force field,[45] the composite approach brings both forms into closer proximity (ca. 0.8 kJ/mol apart), in good agreement with the experimental measurements.

The energy landscape generated solely from the low-cost B86bPBE-XDM/DZP method (Figure 5(f)) produces a similar-quality energy landscape to the reference (Figure 5(g)), showing that relative energies need not be corrected for basis-set superposition and or incompleteness errors (e.g., with counterpoise corrections), which are typically needed to compute accurate absolute lattice energies.[35,38,88] The sHF-3c calculations (Figure 5(d)), again do not yield as good results as the B86bPBE-XDM/DZP method (Figure 5(f)), when compared to the reference landscape, as quantified by the larger S value (*viz.* S = 4.0 vs. 1.3 kJ/mol). Applying the composite approach on force-field or sHF-3c geometries leads to the racemate structure being slightly more favored than the enantiopure (Figures 5(c,e), respectively), in agreement with experiment, whereas with the

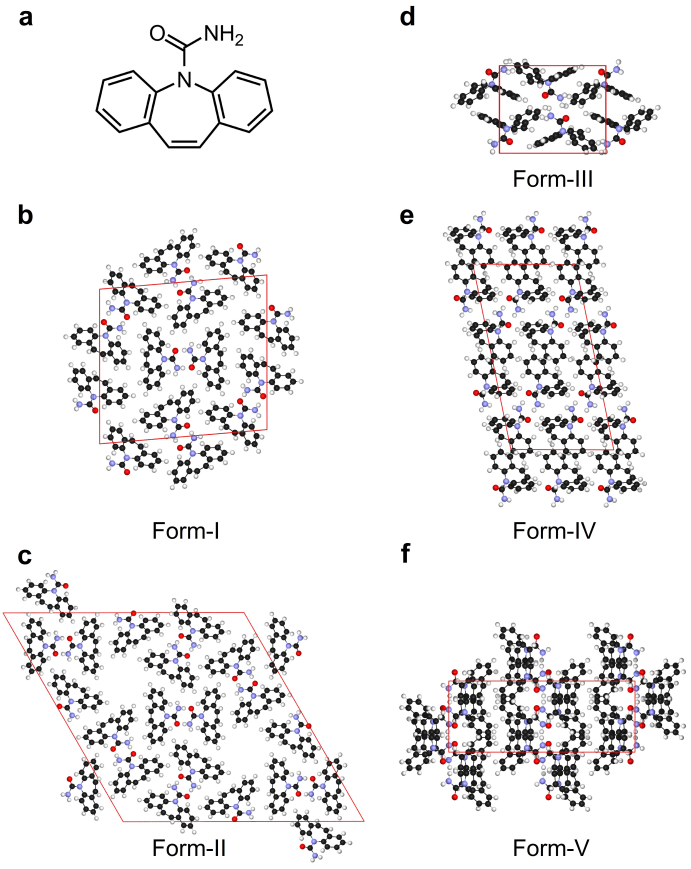
B86bPBE-XDM/DZP geometries, the reverse is true, and thus is no longer in agreement with experiment. However, the enantiopure structure is predicted to be more stable by only ca. 1 kJ/mol, so thermal effects (typically found to affect relative energies of molecular crystals on the order of, at most, 2-4 kJ/mol[89,90]) could potentially reverse the stability ordering of the two forms. Conversely, full B86bPBE-XDM/PAW geometry optimizations performed on the racemic and enantiopure crystal structures places the racemate form as 4.2 kJ/mol more stable than the enantiopure form (without any thermal correction); this matches the experimental ordering, but slightly overestimates the energy difference between the two forms. This overestimation could due to the use of a semi-local base density functional[91] or, less probably, the choice of dispersion correction. However, given that there are large differences in crystallographic symmetry between the previously established lattice-energy minima and experimentally determined structure,[45] thermal effects are likely the predominant source for the disagreement between the computed energy differences and those measured at finite temperatures.

3.3 Carbamazepine

Carbamazepine, a drug used in the treatment of epilepsy and trigeminal neuralgia, has five known polymorphs,[92–94] as depicted in Figure 6. The most stable form at room temperature is Form-III,[93] and both Forms I and II are related enantiotropically to it at higher temperatures, meaning that they become more stable than Form-III at a temperature higher than 298 K.[92] The stability of Form-IV has been reported to lie between that of Forms I and II.[95] Forms II-IV have been shown to have an enantiotropic relationship, and display conversion to Form-I at high temperatures, prior to the melting point of Form-I.[96] Nevertheless, this means that, at 0 K, one would expect Form-III to be the most stable polymorph. Several computational studies[97–99] are generally in qualitative agreement with experimental findings,[92,93,95] giving the following stability ordering of the first four known polymorphs: III $<$ I,IV $<$ II, with the latter three forms being separated by no more than ca. 3-4 kJ/mol. Force-field lattice-energy calculations have stipulated Form-V to be of similar stability to the other carbamazepine polymorphs, within ca. 2 kJ/mol from Forms I and IV,[94,100] but no experimental measurement of its stability relative to the other forms has been reported to date. Presumably, this is because of its later prediction, and eventual isolation, than the first four forms.[94,100]

The relative lattice-energy rankings derived via force-field computations reported in the CPOSS database[47] are presented in Figure 7(a), together with the data obtained herein using the sHF-3c (Figure 7(d)) and B86bPBE-XDM/DZP (Figure 7(f)) low-cost methods. Applying plane-wave DFT single-point energy calculations to all of these geometries yields the energy landscapes depicted in Figure 6(c,e,g), respectively, while the DZP//FF results are depicted in Figure 6(b). The force-field computations again place many of the isolable polymorphs as high-energy structures, whereas reranking from the plane-wave DFT calculations place them all, except for Form-II, in the lower-energy regions of the crystal-energy landscape. The small-basis B86bPBE-XDM/DZP single-point energy calculations perform less well than the plane-wave calculations, but are significantly cheaper and, in addition, generate an energy landscape on par with the fully optimized DZP case, Figure 7(f). Similar results as seen for 5-fluorouracil and naproxen are obtained for sHF-3c, showing poorer performance than B86bPBE-XDM/DZP. Reranking via the composite approach from the fully-optimized B86bPBE-XDM/DZP geometries, however, affords energy landscapes giving the same stability ordering as previously described by other force-field calculations[97–99] and, more importantly, through experiment.[92,93,95] While the low-cost methods disagree on the relative stability of Form-V with respect to Forms IV and II, results from the composite approaches all suggest that it should be more stable than Form-II only. The overall advantage of using composite approaches remains that most isolable polymorphs observed through experiments are predicted to be thermodynamic minima, instead of being of similar relative energy as close to 50 other candidate structures.

Fig. 6 Carbamazepine and its five known polymorphs.



3.4 Olanzapine

Olanzapine, an anti-psychotic drug used in the treatment of schizophrenia,[101,102] currently has three characterized anhydrous polymorphic forms, one other uncharacterized form, and many isolated hydrates and solvates.[46,103] The anhydrous solid-state forms have been inconsistently reported as Forms I-IV in the

literature. The two first forms that were isolated have since been found to be the same and are now both labeled as Form-I.[104] The third form (Form-III) remains uncharacterized and not isolable as a pure crystal.[105] The fourth form has been isolated as a second distinct polymorph (Form-II),[106] although it has been questioned whether it too can be isolated without contamination from Forms I and/or III.[105] That being said, Forms II and III of olanzapine are known to be metastable with regards to Form I and solid-solid phase transitions can be observed through hot-stage microscopy experiments.[105] Very recently, a fourth distinct crystalline form of olanzapine (Form-IV) has been succesfully isolated and characterized.[103] This form differs from other known crystal structures of olanzapine (including the known hydrates and solvates), in that it does not contain the same centrosymmetric dimer stacking motif of olanzapine molecules within its unit cell. Rather, olanzapine molecules in Form-IV interact via hydrogen bonds between neighboring molecules.[103] The structures of the characterized olanzapine polymorphs are depicted in Figure 8.

Previous force-field calculations had predicted the new form (Form-IV) as a potentially isolable structure, given its near degeneracy with Form-I,[46] but conventional methods of crystallization had not been able to obtain polymorphs that incorporated anything other than centrosymmetric dimers.[103] The recent report on olanzapine's Form-IV[103] shows that alternative crystallization conditions to solution-based methods (e.g., using the drug-polymer dispersion system) can yield low-energy structures predicted by CSP methods,[46] but never previously isolated. Through differential scanning calorimety measurements combined with powder X-ray diffraction data, Form-IV was found to be less stable than Form-I, given its lower melting point (ca. 189 vs. 194°C) and lower heat of fusion (ca. 36.7 vs. 40.5 kJ/mol), and no conversion between the forms was observed.[103]

Equally recently, the Gibbs free-energy landscape of olanzapine has been determined through embedded-fragment quantum mechanical methods and Form I has been confirmed to be more stable than Form-II by ca. 2.8 kJ/mol and 4.8 kJ/mol at 5 K and 350 K, respectively.[107] These results are in qualitative agreement with the earlier reported force-field computations of Bhardwaj *et al.*[46] (ca. 6.7 kJ/mol in favor of Form-I at 0 K).

The force-field energy ranking for the 50 lowest-energy structures reported in the previous CSP study[46] is depicted in Figure 9(a). The force-field computations predict both Forms I and IV to be in the lower region of the energy landscape, whereas Form-II is predicted to be significantly less stable. Form-II has nearly 40 or so forms more stable than it, making it quite unlikely that this form and not others would be isolated experimentally. Reranking of the force-field crystal structures using composite approaches with B86bPBE-XDM/DZP and B86bPBE-XDM/PAW (Figures 9(b,c), respectively) now predicts Form II to be significantly lower in energy than most other candidate structures, making these landscapes more realistic. However, Form-IV is still predicted to be lower in energy than Form-I, contrary to what experimental data suggests.[103] The sHF-3c and the B86bPBE-XDM/DZP methods (Figure 9(d,f)), along with the corresponding composite approaches (Figure 9(e,g)), all generate more plausible energy landscapes than did the force field computations. The reference energy landscape (Figure 9(g)) reproduces an energy difference between Forms I and II in good agreement with high-level theory results[107] and, in addition, now predicts Form-I to be slightly more stable than Form-IV by 1.0 kJ/mol, which is in qualitative agreement with experimental data, barring any inclusion of thermal effects.

As for Form-III, its structure has been hypothesized based on structural and energetic similarities present between it and Form-II.[46] Preparing Forms II and III in pure form can be problematic, with mixtures of these forms often resulting instead.[105] The suggested structure by Bhardwaj and co-workers[46] is a likely possibility for this form. The composite-method data presented herein support this, given that the proposed Form-III is now brought into the lower-energy region of the landscape, along with Form-II. It is satisfying to see that all four known polymorph structures are the lowest-energy candidate structures on the crystal-energy landscape generated with the most accurate composite approach presented in this work (Figure 9(g)).

4 Conclusions

The work presented herein aimed at applying small-basis set methods, and composite approaches, to the CSP of four APIs. The results obtained were compared to anisotropic force-field data from the CPOSS database. In general, the force fields typically did not yield accurate energy landscapes, as many of the isolable polymorphs were higher in energy than other candidate structures. In addition, the relative stability ordering of the polymorphs did not always agree with the experimental data. This was also the case for the sHF-3c results, most notably for 5-fluorouracil and carbamazepine. In contrast to this, the small-basis B86bPBE-XDM/DZP method implemented in the SIESTA code provided more realistic energy landscapes, consistently placing isolable polymorphs in the lowest-energy regions of the crystal-energy landscapes.

Composite methods, in which high-level single-point energies are calculated at the low-level geometries, yielded improved results in agreement with full plane-wave DFT calculations (for 5-fluorouracil) and, most importantly, in agreement with available experimental evidence. Specifically, for 5-fluorouracil, composite methods recovered the correct experimental ordering, confirming that the previous error is not due to thermal effects, but is a limitation of the force field. For naproxen, it is conjectured that thermal effects may be important to recover the correct ordering, since the known racemic and enantiopure forms are nearly degenerate and the order flips depending on the choice of geometries employed in the composite methods. However, in a CSP protocol, high-level plane-wave DFT optimizations could be performed, together with phonon calculations, to resolve the thermodynamic minimum. For carbamazepine, the composite approaches place most isolable forms within the low-energy region of the crystal-energy landscape. This work adds computational evidence that Form-V is slightly less stable than Forms I and IV, but remains close in energy to the other experimentally isolable polymorphs overall. Inclusion of thermal effects, which remains challenging to do accurately and/or efficiently,[108,109] would again be beneficial when comparing to experimental data, or in determining

Fig. 7 Crystal-energy landscapes of carbamazepine. Energy landscapes were generated via (a) a force-field (FF, data obtained from the CPOSS database[47]) and two small- and/or minimal-basis low-cost methods, (d) sHF-3c and (f) B86bPBE-XDM/DZP (DZP). Crystal structures within panels (a,d,f) were then reranked with single-point energy calculations using plane-wave B86bPBE-XDM (PAW) to give panels (c,e,g), respectively. The force-field structures within panel (a) have also been reranked with single-point energy calculations using B86bPBE-XDM/DZP to give panel (b). Data points that are encircled are labelled according to the experimentally isolated polymorphs with which they are equivalent, i.e., Forms I-V. The computed energies, ΔE, and densities, $\Delta\rho$, are expressed relative to Form-I for all methods. The similarity index, S, and the heat-map colouring are relative to the crystal-energy landscape in panel (g). Relative energies between the isolable polymorphs are tabulated in Table 1.

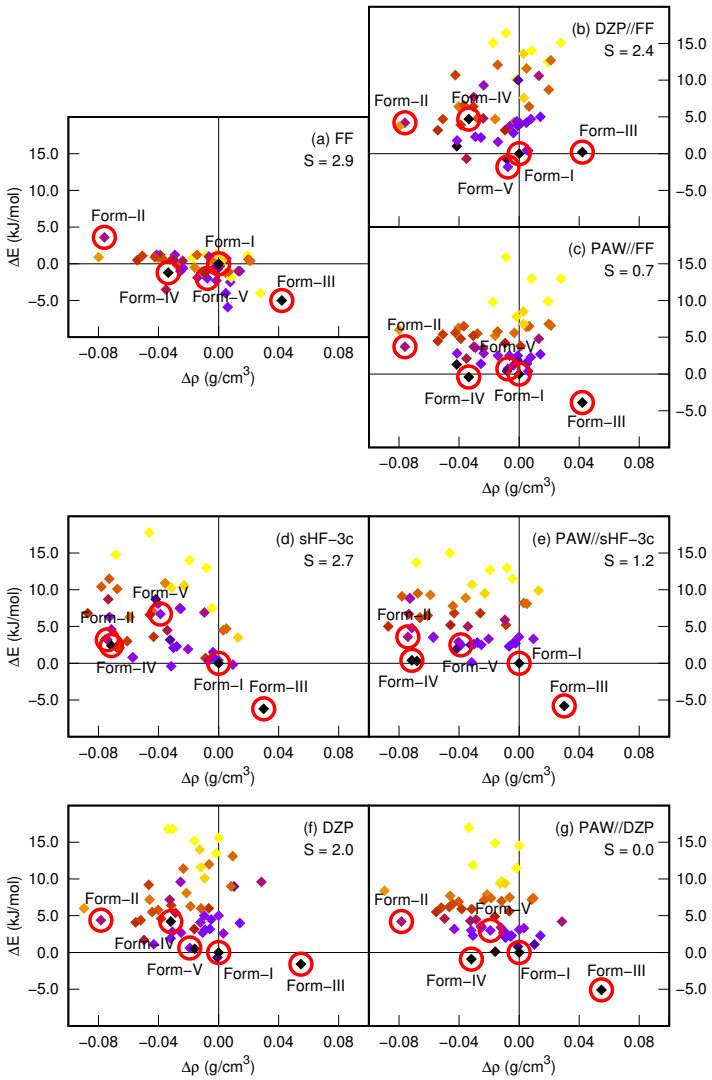
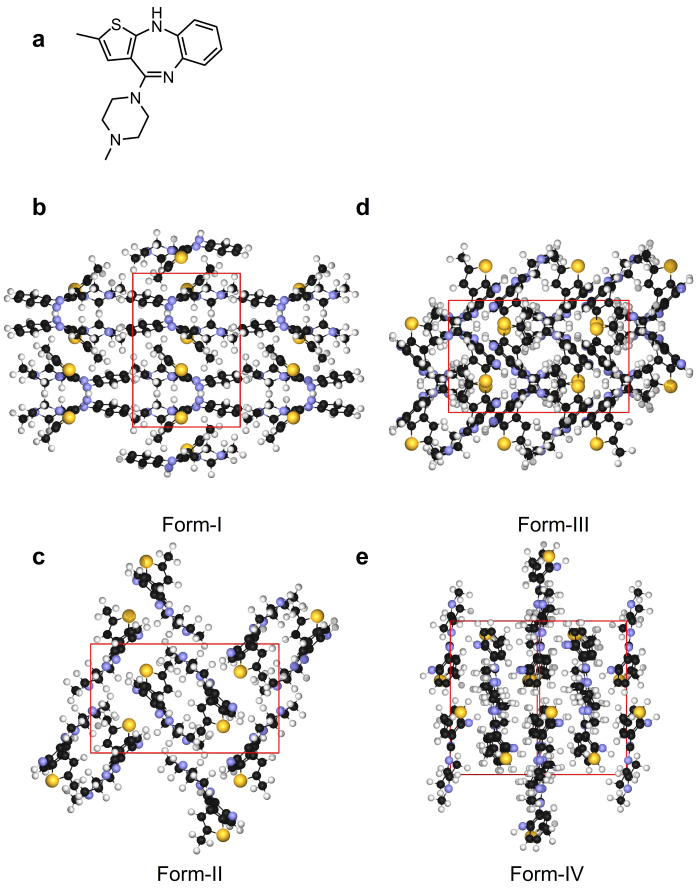


Table 1 Relative energies, in kJ/mol, for the experimental polymorphs of carbamazepine computed with low-level and composite approaches. Energies are expressed relative to the most stable form of carbamazepine at 0 K, Form-III. B//A: corresponds to the use of the low-level and composite approach, where A = the lower-level method indicated in the table header, B = the higher-level B86bPBE-XDM/DZP and B86bPBE-XDM/PAW methods (DZP and PAW, respectively). The established experimental ordering is III $<$ I,IV $<$ II.

	FF			sHF-3c		DZP	
	A	DZP//A	PAW//A	A	PAW//A	A	PAW//A
Form-III	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Form-I	5.0	-0.2	3.9	6.2	5.8	1.6	5.1
Form-IV	3.8	4.5	3.5	8.6	6.2	5.8	4.2
Form-V	3.0	-2.0	4.6	12.9	8.3	2.2	8.1
Form-II	8.6	4.0	7.6	9.4	9.4	6.0	9.3

whether other competitive forms still remain to be discovered. For olanzapine, the data obatined via composite methods confirm that there should be four isolable polymorphs, given that four low-energy structures are predicted from the given candidates. Form-I is predicted to be the most stable polymorph, in agreement with experiment, while Form-IV is slightly less stable than

Fig. 8 Olanzapine and its three characterized polymorphs, along with the hypothesized structure of Form-III.[46] Form-I is the most stable experimentally.



the thermodynamic minimum, and Forms II and III are nearly degenerate in energy. This work also confirms the proposed structure for Form-III from the CSP study of Bhardwaj and co-workers, given that it is one of the four lowest-energy structures identified on the crystal-energy landscape.

Overall, the small-basis B86bPBE-XDM/DZP method yielded the geometries most amenable to further single-point energy calculations with B86bPBE-XDM/PAW, followed by the force-field and sHF-3c geometries, respectively. The poorer performance of the force-field and sHF-3c methods is most likely due to the high levels of (semi-)empiricism involved in these methods' constructions. However, while force-field calculations do not always yield the proper energetic ordering, the use of a composite approach could be useful in identifying crystal-packing motifs where force fields are biased and subject to systematic errors. The results confirm that, for composite approaches to be most succesful and reliable, the low- and high-level methods should be compatible, i.e., be based on similar levels of theory. Finally, while the cost of performing small-basis set calculations with B86bPBE-XDM/DZP is slightly more expensive than sHF-3c, and both of these are relatively more expensive than density-functional tight-binding methods, it should be noted that no empiricism or parameter fitting (beyond the damping function) is involved in B86bPBE-XDM/DZP. This is in stark contrast to the other low-cost methods (force fields, DFTB, and sHF-3c) considered for use in CSP, which makes it a more reliable and generally applicable method for systems where DFT can provide an appropriate description of the electronic structure. Finally, to make the B86bPBE-XDM composite approach competitive with lower-cost semiempirical methods, one could first utilize B86bPBE-XDM/DZP single-point calculations to refine the preliminary energy landscape produced by a force-field. Then, taking only the lower-energy candidate structures, one could further refine the energy ranking and geometries of candidate structures with the composite B86bPBE-XDM/PAW//B86bPBE-XDM/DZP approach.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The authors acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support and Compute Canada for computational resources. L.M.L. would also like to acknowledge the Walter C. Sumner Foundation for financial support. The authors would like to further thank Prof. Sally L. Price and Dr. Louise S. Price, University College London, for providing access to the CPOSS database structures for the APIs studied in this work.

Notes and references

1 S. L. Price, D. E. Braun and S. M. Reutzel-Edens, *Chem. Commun.*, 2016, **52**, 7065–7077.

2 S. L. Price and S. M. Reutzel-Edens, *Drug Discovery Today*, 2016, **21**, 912–923.

3 E. O. Pyzer-Knapp, H. Thompson, F. Schiffmann, K. E. Jelfs, S. Y. Chong, M. A. Little, A. I. Cooper and G. M. Day, *Chem. Sci.*, 2014, **5**, 2235–2245.

4 Y. Yang, B. Rice, X. Shi, J. R. Brandt, R. Correa da Costa, G. J. Hedley, D. Smilgies, J. M. Frost, I. Samuel, A. Otero-de-la-Roza, E. R. Johnson, K. E. Jelfs, J. Nelson, A. J. Campbell and M. J. Fuchter, *ACS Nano*, 2017, **11**, 8329–8338.

5 M. K. Ravva, C. Risko and J. L. Brédas, in *Non-covalent Interactions in Quantum Chemistry and Physics*, ed. A. Otero-de-la-Roza and G. DiLabio, Elsevier, Amsterdam, Netherlands, 2017, ch. 9, pp. 277–298.

6 H. G. Gallagher and J. N. Sherwood, *J. Chem. Soc., Faraday Trans.*, 1996, **92**, 2107–2116.

7 R. M. Vrcelj, J. N. Sherwood, A. R. Kennedy, H. G. Gallagher and T. Gelbrich, *Cryst. Growth Des.*, 2003, **3**, 1027–1032.

8 C. Wei, H. Huang, X. Duan and C. Pei, *Propellants Explos. Pyrotech.*, 2011, **36**, 416–423.

9 Z. Hao and A. Iqbal, *Chem. Soc. Rev.*, 1997, **26**, 203–213.

10 T. Zykova-Timan, P. Raiteri and M. Parrinello, *J. Phys. Chem. B*, 2008, **112**, 13231–13237.

11 W. McCrone, in *Physics and Chemistry of the Organic Solid State*, ed. D. Fox, M. M. Labes and N. Weissenberg, Wiley Interscience, New York, 1965, ch. 8, pp. 725–767.

12 J. Maddox, *Nature*, 1988, **335**, 201.

Fig. 9 Crystal-energy landscapes of olanzapine. Energy landscapes were generated via (a) a force-field (FF, data obtained from the CPOSS database[46,47]) and two small- and/or minimal-basis low-cost methods, (d) sHF-3c and (f) B86bPBE-XDM/DZP (DZP). Crystal structures within panels (a,d,f) were then reranked with single-point energy calculations using plane-wave B86bPBE-XDM (PAW) to give panels (c,e,g), respectively. The force-field structures within panel (a) have also been reranked with single-point energy calculations using B86bPBE-XDM/DZP to give panel (b). Data points that are encircled are labelled according to the experimentally isolated polymorphs, or proposed structure (Form-III), with which they are equivalent, i.e., Forms I-IV. The computed energies, ΔE, and densities, $\Delta\rho$, are expressed relative to Form-I for all methods. The similarity index, S, and the heat-map colouring are relative to the crystal-energy landscape in panel (g). Relative energies between the isolable polymorphs are tabulated in Table 2.

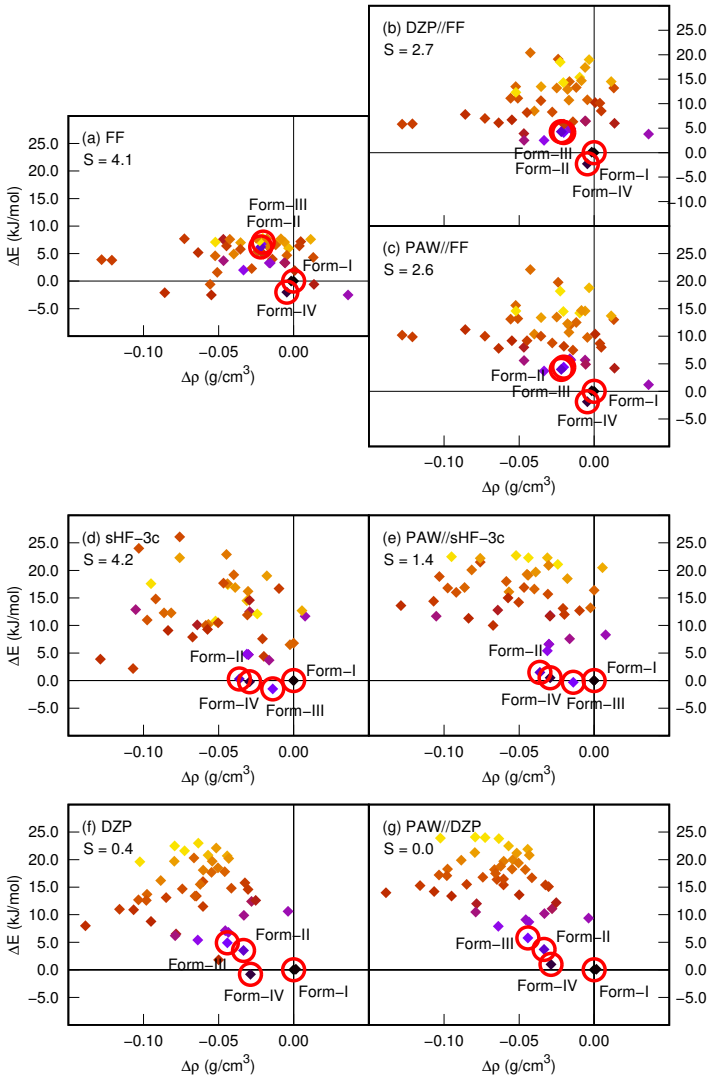


Table 2 Relative energies, in kJ/mol, for the experimental polymorphs of olanzapine computed with low-level and composite approaches. Energies are expressed relative to the most stable form of olanzapine, Form-I. B//A: corresponds to the use of the low-level and composite approach, where A = the lower-level method indicated in the table header, B = the higher-level B86bPBE-XDM/DZP and B86bPBE-XDM/PAW methods (DZP and PAW, respectively).

	FF			sHF-3c		DZP	
	A	DZP//A	PAW//A	A	PAW//A	A	PAW//A
Form-I	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Form-II	6.2	4.3	4.0	0.3	1.5	3.5	3.7
Form-III	7.1	4.1	4.4	-1.5	-0.3	4.9	5.8
Form-IV	-2.0	-2.3	-1.9	-0.2	-0.5	-0.8	1.0

13 A. Gavezzotti, *Acc. Chem. Res.*, 1994, **27**, 309–314.

14 G. M. Day, *Crystallogr. Rev.*, 2011, **17**, 3–52.

15 S. L. Price, *Science*, 2014, **345**, 619–620.

16 S. L. Price, *Chem. Soc. Rev.*, 2014, **43**, 2098–2111.

17 G. Beran, *Angew. Chem. Int. Ed.*, 2015, **54**, 396–398.

18 A. J. Cruz-Cabeza, S. M. Reutzel-Edens and J. Bernstein, *Chem. Soc. Rev.*, 2015, **44**, 8619–8635.

19 T. S. Thakur, R. Dubey and G. R. Desiraju, *Annu. Rev. Phys.*

Chem., 2015, **66**, 21–42.

20 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallographica Section B: Structural Science*, 2016, **72**, 171–179.

21 G. Beran, *Chem. Rev.*, 2016, **116**, 5567–5613.

22 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. Janckiewicz, D. T. de Jong, J. Kendrick, N. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. 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. Z. Yang, Q. Zhu and C. R. Groom, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2016, **72**, 439–459.

23 J. Haleblian and W. McCrone, *J. Pharm. Sci.*, 1969, **58**, 911–929.

24 Y. A. Abramov, *Org. Process Res. Dev.*, 2013, **17**, 472–485.

25 S. L. Price, *Faraday Discuss.*, 2018, **211**, 9–30.

26 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.

27 J. Yang, S. De, J. E. Campbell, S. Li, M. Ceriotti and G. M. Day, *Chem. Mater.*, 2018, **30**, 4361–4371.

28 J. G. Brandenburg and S. Grimme, *J. Phys. Chem. Lett.*, 2014, **5**, 1785–1789.

29 J. G. Brandenburg, M. Hochheim, T. Bredow and S. Grimme, *J. Phys. Chem. Lett.*, 2014, **5**, 4275–4284.

30 M. Cutini, B. Civalleri, M. Corno, R. Orlando, J. G. Brandenburg, L. Maschio and P. Ugliengo, *J. Chem. Theory Comput.*, 2016, **12**, 3340–3352.

31 O. A. Loboda, G. A. Dolgonos and A. D. Boese, *J. Chem. Phys.*, 2018, **149**, 124104.

32 L. Iuzzolino, P. McCabe, S. L. Price and J. G. Brandenburg, *Faraday Discuss.*, 2018, **211**, 275–296.

33 M. Mortazavi, J. G. Brandenburg, R. J. Maurer and A. Tkatchenko, *J. Phys. Chem. Lett.*, 2018, **9**, 399–405.

34 G. A. Dolgonos and A. D. Boese, *Chem. Phys. Lett.*, 2019, **718**, 7–11.

35 L. M. LeBlanc, A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Theory Comput.*, 2018, **14**, 2265–2276.

36 J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón and D. Sánchez-Portal, *J. Phys.: Condens. Matter*, 2002, **14**, 2745–2779.

37 E. Artacho, J. D. Gale, A. García, J. Junquera, R. M. Martin, P. Ordejón, J. M. Pruneda, D. Sánchez-Portal and J. M. Soler, *J. Phys.: Condens. Matter*, 2008, **20**, 064208.

38 L. M. LeBlanc, J. A. Weatherby, A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Theory Comput.*, 2018, **14**, 5715–5724.

39 D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle, G. R. Desiraju, B. P. van Eijck, J. C. Facelli, M. B. Ferraro, D. Grillo, M. Habgood, D. W. M. Hofmann, F. Hofmann, K. V. J. Jose, P. G. Karamertzanis, A. V. Kazantsev, J. Kendrick, L. N. Kuleshova, F. Leusen, A. V. M. ad A. J. Misquitta, S. Mohamed, R. J. Needs, M. A. Neumann, D. Nikylov, A. M. Orendt, R. Pal, C. C. Pantelides, C. J. Pickard, L. S. Price, S. L. Price, H. A. Scheraga, J. van de Streek, T. S. Thakur, S. Tiwari, E. Venuti and I. K. Zhitkov, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2011, **67**, 535–551.

40 A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Phys.*, 2012, **137**, 054103.

41 A. M. Reilly and A. Tkatchenko, *J. Chem. Phys.*, 2013, **139**, 024705.

42 M. Cutini, B. Civalleri and P. Ugliengo, *ACS Omega*, 2019, **4**, 1838–1846.

43 D. McDonagh, C. Skylaris and G. M. Day, *J. Chem. Theory Comput.*, 2019, **15**, 2743–2758.

44 A. T. Hulme, S. L. Price and D. A. Tocher, *J. Am. Chem. Soc.*, 2005, **127**, 1116–1117.

45 D. E. Braun, M. Ardid-Candel, E. D'Oria, P. G. K. anf J.-B. Arlin, A. J. Florence, A. G. Jones and S. L. Price, *Cryst. Growth Des.*, 2011, **11**, 5659–5669.

46 R. M. Bhardwaj, L. S. Price, S. L. Price, S. M. Reutzel-Edens, G. J. Miller, I. Oswald, B. F. Florence and A. J. Florence, *Cryst. Growth Des.*, 2013, **13**, 1602–1617.

47 *The CPOSS (Control and Prediction of the Organic Solid State) database holds details of the hypothetical crystal structures generated in the computational searches carried out at University College London, and is maintained by Dr. Louise S. Price, and led by Prof. Sally L. Price. See http://www.chem.ucl.ac.uk/cposs/index.htm*.

48 S. L. Price, M. Leslie, G. Welch, M. Habgood, L. S. Price, P. G. Karamertzanis and G. M. Day, *Phys. Chem. Chem. Phys.*, 2010, **12**, 8478–8490.

49 V. K. Srirambhatla, R. Guo, S. L. Price and A. J. Florence, *Chem. Commun.*, 2016, **52**, 7384–7386.

50 A. D. Becke, *J. Chem. Phys.*, 1986, **85**, 7184–7187.

51 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

52 E. R. Johnson, in *Non-Covalent Interactions in Quantum Chemistry and Physics*, ed. A. Otero-de-la-Roza and G. DiLabio, Elsevier, Amsterdam, Netherlands, 2017, ch. 5, pp. 169–192.

53 N. Troullier and J. L. Martins, *Phys. Rev. B*, 1991, **43**, 1993–2006.

54 N. Troullier and J. L. Martins, *Phys. Rev. B*, 1991, **43**, 8861–8869.

55 D. R. Hamann, M. Schlüter and C. Chiang, *Phys. Rev. Lett.*, 1979, **43**, 1494–1497.

56 G. B. Bachelet, D. R. Hamann and M. Schlüter, *Phys. Rev. B*, 1982, **26**, 4199–4228.

57 *ATOM, a program for DFT calculations in atoms and pseudopotential generation, maintained by Alberto Garcia and distributed as part of the SIESTA package. See http://www.icmab.es/siesta/atom*.

58 R. Sure and S. Grimme, *J. Comput. Chem.*, 2013, **34**, 1672–1685.

59 J. G. Brandenburg and S. Grimme, *Top. Curr. Chem.*, 2014, **345**, 1–24.

60 R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rerat, S. Casassa, J. Baima, S. Salustro and B. Kirtman, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1360.

61 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

62 S. Grimme, S. Ehrlich and L. Goerigk, *J. Comput. Chem.*, 2011, **32**, 1456–1465.

63 H. Kruse and S. Grimme, *J. Chem. Phys.*, 2012, **136**, 154101.

64 J. G. Brandenburg, M. Alessio, B. Civalleri, M. F. Peintinger, T. Bredow and S. Grimme, *J. Phys. Chem. A*, 2013, **117**, 9282–9292.

65 E. R. Johnson and A. D. Becke, *J. Chem. Phys.*, 2006, **124**, 174104.

66 P. E. Blöchl, *Phys. Rev. B*, 1994, **50**, 17953–17979.

67 P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. Buongiorno Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. Dal Corso, S. de Gironcoli, P. Delugas, R. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast and S. Baroni, *J. Phys.: Condens. Matter*, 2017, **29**, 465901.

68 C. Heidelberger, N. R. Chaudhuri, P. Danneberg, D. Mooren and L. Griesbach, *Nature*, 1957, **179**, 663–666.

69 J. L. Grem, *Invest. New Drugs*, 2000, **18**, 299–313.

70 L. Fallon III, *Acta. Cryst. B*, 1973, **29**, 2549–2556.

71 S. Hamad, C. Moon, C. Catlow, A. T. Hulme and S. L. Price, *J. Phys. Chem. B*, 2006, **110**, 3323–3329.

72 S. L. Price, S. Hamad, A. Torrisi, P. G. Karamertzanis, M. Leslie and C. Catlow, *Mol. Simul.*, 2006, **32**, 985–997.

73 S. A. Barnett, A. T. Hulme, N. Issa, T. C. Lewis, L. S. Price, D. A. Tochter and S. L. Price, *New J. Chem.*, 2008, **32**, 1761–1775.

74 P. G. Karamertzanis, P. Raiteri, M. Parrinello, M. Leslie and S. L. Price, *J. Phys. Chem. B*, 2008, **112**, 4298–4308.

75 T. Bui, S. Dehaoui, C. Lecomte, G. R. Desiraju and E. Espinosa, *Angew. Chem. Int. Ed.*, 2009, **48**, 3838–3840.

76 R. J. Baker, P. E. Colavita, D. M. Murphy and J. D. Wallis, *J. Phys. Chem. A*, 2012, **116**, 1435–1444.

77 P. Politzer, J. S. Murray and T. Clark, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11178–11189.

78 K. N. Jarzembska, M. Kubsik, R. Kamiński, K. Woźniak and P. M. Dominiak, *Cryst. Growth Des.*, 2012, **12**, 2508–2524.

79 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

80 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

81 A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2007, **127**, 154108.

82 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, J. M. 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, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox.

83 A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Phys.*, 2013, **138**, 204109.

84 A. Otero-de-la Roza, E. R. Johnson and G. A. DiLabio, *J. Chem. Theory Comput.*, 2014, **10**, 5436–5447.

85 P. Metrangolo, J. S. Murray, T. Pilati, P. Politzer, G. Resnati and G. Terraneo, *CrystEngComm*, 2011, **13**, 6593–6596.

86 K. Ravikumar, S. S. Rajan and V. Pattabhi, *Acta Cryst. C*, 1985, **41**, 280–282.

87 Y. B. Kim, H. J. Song and I. Y. Park, *Arch. Pharm. Res.*, 1987, **10**, 232–238.

88 D. J. Carter and A. L. Rohl, *J. Chem. Theory Comput.*, 2014, **10**, 3423–3437.

89 J. Nyman and G. M. Day, *CrystEngComm*, 2015, **17**, 5154–5165.

90 Y. N. Heit and G. Beran, *Acta Cryst. B*, 2016, **72**, 514–529.

91 S. R. Whittleton, A. Otero-de-la-Roza and E. R. Johnson, *J. Chem. Theory Comput.*, 2017, **13**, 5332–5342.

92 C. Rustichelli, G. Gamberini, V. Ferioli, M. C. Gamberini, R. Ficarra and S. Tommasini, *J. Pharm. Biomed. Anal.*, 2000, **23**, 41–54.

93 M. Lang, J. W. Kampf and A. J. Matzger, *J. Pharm. Sci.*, 2002, **91**, 1186–1190.

94 J. Arlin, L. S. Price, S. L. Price and A. J. Florence, *Chem. Commun.*, 2011, **47**, 7074–7076.

95 A. L. Grezsiak, M. D. Lang, K. Kim and A. J. Matzger, *J. Pharm. Sci.*, 2003, **92**, 2260–2271.

96 A. E. Clout, A. Buanz, S. Gaisford and G. R. Williams, *Chem. Eur. J.*, 2018, **24**, 13573–13581.

97 P. G. Karamertzanis and S. L. Price, *J. Chem. Theory Comput.*, 2006, **2**, 1184–1190.

98 A. J. Florence, A. Johnston, S. L. Price, H. Nowell, A. R. Kennedy and N. Shankland, *J. Pharm. Sci.*, 2006, **95**, 1918–1930.

99 A. J. Cruz-Cabeza, G. M. Day, W. D. S. Motherwell and W. Jones, *Cryst. Growth Des.*, 2006, **6**, 1858–1866.

100 A. J. Florence, C. K. Leech, N. Shankland and A. Johnston, *CrystEngComm*, 2006, **8**, 746–747.

101 B. Fulton and K. L. Goa, *Drugs*, 1997, **53**, 281–298.

102 G. D. Tollefson, C. M. Beasley Jr., P. V. Tran, J. S. Street, J. A. Krueger, R. N. Tamura, K. A. Graffeo and M. E. Thieme, *Am. J. Psychiatry*, 1997, **154**, 457–465.

103 S. Askin, J. K. Cockcroft, L. S. Price, A. D. Gonçalves, M. Zhao, D. A. Tocher, G. R. Williams, S. Gaisford and D. Craig, *Cryst. Growth Des.*, 2019, DOI:10.1021/acs.cgd.8b01881.

104 S. M. Reutzel-Edens, J. K. Bush, P. A. Magee, G. A. Stephenson and S. R. Byrn, *Cryst. Growth Des.*, 2003, **3**, 897–907.

105 C. G. Testa, L. D. Prado, R. N. Costa, M. L. Costa, Y. G. Linck, G. A. Monti, S. L. Cuffini and H. Rocha, *Int. J. Pharm.*, 2019, **556**, 125–135.

106 R. Thakuria and A. Nangia, *Acta Cryst. C*, 2011, **67**, o461–o463.

107 H. Luo, X. Hao, Y. Gong, J. Zhou, X. He and J. Li, *Cryst. Growth Des.*, 2019, DOI:10.1021/acs.cgd.9b00068.

108 J. G. Brandenburg, J. Potticary, H. A. Sparkes, S. L. Price and S. R. Hall, *J. Phys. Chem. Lett.*, 2017, **8**, 4319–4324.

109 J. Hoja, H. Ko, M. A. Neumann, R. Car, R. A. DiStasio Jr. and A. Tkatchenko, *Sci. Adv.*, 2019, **5**, eaau3338.