Computational predictions of cocrystal formation: A benchmark study of 28 assemblies comparing five methods from high-throughput to advanced models

Item Type	Article
Authors	Fox, Robert;Klug, Joaquin;Thompson, Damien;Reilly, Anthony
Citation	Journal of Computational Chemistry
Publisher	Wiley Periodicals LLC
Download date	2026-01-18 17:05:17
Item License	https://creativecommons.org/licenses/by-nc-sa/4.0/
Link to Item	https://doi.org/10.34961/researchrepository-ul.26243900

Received: 18 April 2024 | Revised: 10 June 2024 | Accepted: 12 June 2024
DOI: 10.1002/jcc.27454

RESEARCH ARTICLE

Journal of COMPUTATIONAL CHEMISTRY WILEY

Computational predictions of cocrystal formation: A benchmark study of 28 assemblies comparing five methods from high-throughput to advanced models

Robert Fox[1] | Joaquin Klug[2] | Damien Thompson[3] | Anthony Reilly[1]

[1]School of Chemical Sciences, Dublin City University, Dublin, Ireland

[2]Department of Life Sciences, Faculty of Sciences, Atlantic Technological University, ATU Sligo, Sligo, Ireland

[3]Department of Physics, Bernal Institute, University of Limerick, Limerick, Ireland

Correspondence
Damien Thompson, Department of Physics, Bernal Institute University of Limerick, Limerick V94 T9PX, Ireland.
Email: damien.thompson@ul.ie

Funding information
Science Foundation Ireland, Grant/Award Number: 12/RC/2275_P2

Abstract

Cocrystals are assemblies of more than one type of molecule stabilized through non-covalent interactions. They are promising materials for improved drug formulation in which the stability, solubility, or biocompatibility of the active pharmaceutical ingredient (API) is improved by including a coformer. In this work, a range of density functional theory (DFT) and density functional tight binding (DFTB) models are systematically compared for their ability to predict the lattice enthalpy of a broad range of existing pharmaceutically relevant cocrystals. These range from cocrystals containing model compounds 4,4′-bipyridine and oxalic acid to those with the well benchmarked APIs of aspirin and paracetamol, all tested with a large set of alternative coformers. For simple cocrystals, there is a general consensus in lattice enthalpy calculated by the different DFT models. For the cocrystals with API coformers the cocrystals, enthalpy predictions depend strongly on the DFT model. The significantly lighter DFTB models predict unrealistic values of lattice enthalpy even for simple cocrystals.

1 | INTRODUCTION

Cocrystals are formed when two or more neutral molecules coassemble in a single lattice.[1] Typically, a functional molecule and a second nonsolvent molecule, termed the coformer, are crystallized in a stoichiometric ratio, with ordered supramolecular packing of the two components.[2,3] Other separate classes of multicomponent solid forms include hydrates with water, solvates with organic solvents, salts with charged compounds,[4] and racemic mixtures of chiral enantiomers.[5] Originally discovered by Wöhler in 1844 when quinhydrone was crystallized with quinone,[6] known cocrystals have experienced significant and rapid growth over the last 30 years.[6] This growth in the number of cocrystals is due to their ability to improve solubility, mechanical properties, density, and bioavailability[7] by molecular and crystal engineering. Cocrystals have provided useful materials solutions in fields ranging from pharmaceuticals[8] to semiconductors[9] and other value-added chemicals.[10] The potential to improve the solubility of active pharmaceutical ingredients (APIs) using cocrystals significantly impacts the pharmaceutical sector because 40% of approved APIs and 90% of developmental drugs have poor solubility.[11] Another important issue for the (bio)pharma sector is the poor permeability of APIs.[12] For example, cocrystals formulated with permeation enhancing coformers enable the delivery of peptides and related small molecule drugs.[13]

Molecules in a crystal structure can be arranged in multiple ways within the lattice, creating polymorphism. Polymorphism complicates the rational engineering of cocrystals, especially when polymorphs are near iso-energetic but have different properties, meaning multiple polymorphs can be produced under standard synthesis conditions. A well-known example is the HIV drug ritonavir, marketed under the brand name Norvir®, which morphed into a lower solubility structure over time. This was discovered 2 years after its release to the market when Norvir® capsules started failing dissolution tests, and the drug was removed from the market.[14]

Previous work has shown that cocrystals are thermodynamically stable, with 95% of the cocrystals studied having favorable negative lattice enthalpy.[1] This was established through dispersion-inclusive density functional theory (DFT-D) calculations incorporating a van der Waals correction to the main exchange-correlation function to predict the lattice enthalpy. Uncorrected DFT can accurately predict simple salts, hydrates and solvates, and hydrogen-bonded assemblies but becomes less useful for cocrystals and crystals with significant van der Waals (vdW) contacts. Previous work predicted the mean absolute percentage error (MAPE) ranging from 97% presented by the B3LYP to 49% by the B971 exchange functional for crystals with significant vdW contacts in contrast to the much lower values for hydrogen bonded systems ranging from 18% using the HCTH407 exchange functional to 3% with B971.[15] The difference between the measured and calculated lattice energies emphasizes the importance of incorporating dispersion corrections for more accurate description of the interactions between nonpolar and weakly polar molecular groups.[16] Another example of the importance of dispersion corrections was the discovery of the previously unknown polymorph of aspirin, which was found using computational models with dispersion interactions and experiments.[17]

Dispersion corrections can scale to progressively longer-range and computationally more demanding calculations, with options to calculate over pairwise, three-body or many-body interactions.[16,18–22] Here, we have assessed the predictive power of different DFT-based methods to identify and quantify the stabilities of a broad range of cocrystals spanning four different functional molecules, namely, oxalic acid, 4,4′-bipyridine, aspirin and paracetamol, with multiple coformers. Oxalic acid and 4,4′-bipyridine were selected due to the simplicity of their molecular structures, with primarily hydrogen bonding or van der Waals intermolecular interaction in their single crystal form, respectively. The cocrystals containing the APIs,[23] aspirin or paracetamol as the functional molecule were selected to test the limits of the predictive power of the various dispersion correction methods, including the pairwise Tkatchenko-Scheffler (TS)[24] dispersion correction and the many-body dispersion (MBD) schemes.[25,26]

One way to reduce the computational cost of DFT calculations and develop high throughput screens is to introduce approximations to the Kohn–Sham DFT model, and one of these approximations is density functional based tight binding (DFTB).[27] In DFTB, a minimal atomic basis set is used, and the exchange correlation energy derived from the Kohn–Sham DFT is expanded to a third-order Taylor series to calculate the total energy.[28] This allows faster calculations on more complex crystals but comes with reduced accuracy that may or may not significantly reduce the predictive power of the calculations.

2 | METHODS

2.1 | Cocrystal selection

We systematically survey organic cocrystals of four selected molecules (oxalic acid, 4,4′-bipyridine, paracetamol, and aspirin) of increasing complexity to benchmark the computational methods, where paracetamol and aspirin are two well-characterized single component crystals. The cocrystal sets and their individual component crystals were mined from the Cambridge structural database (CSD)[29] using the ConQuest code, limiting the atom list to C, H, N, O, F, S, and Cl. When duplicate structures were available, the lowest temperature structure was selected to better complement the 0 K DFT calculation. A limitation on the number of atoms in the unit cell was used to make the calculations feasible for the PBE0 hybrid exchange-correlation functional[30] that mixes the Perdew–Burke–Ernzerhof (PBE)[31] exchange energy and Hartree–Fock exchange energy in a set 3:1 ratio, along with the full PBE correlation energy. We investigate five oxalic acid cocrystals, thirteen 4,4′-bipyridine cocrystals, six paracetamol cocrystals, and four aspirin cocrystals, with a broad range of commonly used, readily available coformers (Tables 1 and 2) more information on the conquest search on the Data S1.

2.2 | Computational methods

All the selected experimental cocrystal and single component crystal structures were optimized using planewave DFT as implemented in the CASTEP (CAmbridge Serial Total Energy Package) code.[32] The PBE exchange-correlation functional[31] was used for both standard DFT and dispersion-inclusive DFT along with the PBE0 variant, and we compared the predictive power of DFT[33] utilizing two different dispersion corrections, the Tkatchenko-Scheffler (TS)[24] and the MBD schemes.[25,26] The PBE0 functional aims to offset the delocalization error seen in the PBE functional.[34]

In general, the dispersion interactions are incorporated in a second step following calculation of the base exchange-correlation functional energy.[16] The popular TS dispersion correction method uses a simple pairwise summation of the dispersion energy (Equation 1)[35]:

$$E_{\mathrm{disp}} = -\frac{1}{2}\sum_{\mathrm{A,B}} \frac{C_{6\mathrm{AB}}}{r_{\mathrm{AB}}^{6}} f_{\mathrm{damp}}(r_{\mathrm{AB}}), \qquad (1)$$

where r_{AB}^{6} is the interatomic distance between atoms A and B, and $f_{\mathrm{damp}}(r_{\mathrm{AB}})$ is the dampening function. With this method, the $C_{6\mathrm{AB}}$ uses the relationship between polarizability and volume, allowing it to factor in the chemical environment.

The MBD method attempts to calculate the many-body effects beyond two-body effects. This many-body energy is included within the long-range correlation energy.[16] This method has previously reported an improvement in the MAPE from 9% with PBE-TS to 5% with the PBE-MBD when using the S22 database.[25]

The unit cells were optimized with a cutoff energy of 900 eV, using an electronic energy tolerance of 1×10^{-7} eV per atom and geometry energy tolerance of 5×10^{-7} eV per atom to generate final structures with all forces below 5×10^{-3} eV Å^{-1}. A tight k-point spacing of 0.06 Å^{-1} was used together with ultra-soft pseudopotentials (USP) to describe the electronic structure.[36]

To test the effect of the functional choice apart from the dispersion corrections, we performed some high-level single point energy PBE0-MBD calculations on the optimized PBE-MBD structures.

TABLE 1 The calculated cocrystal compositions.

CSD reference code			Composition A_mB_n		H-bonding in the cocrystal structure	
Cocrystal	Coformer A	Coformer B	*m* (A)	*n* (B)	Inter-	Intra-
GUDSUV	OXALAC05	BACJUN	1	1	Yes	No
UMINAF	OXALAC05	EVESIJ04	1	1	Yes	No
UROXAM01	OXALAC05	UREAXX12	1	1	Yes	No
XAPMIK	OXALAC05	PHENAZ11	1	1	Yes	No
XEJWUF	OXALAC05	BAPLOT04	1	2	Yes	No
GIPQEB01	HIQWEJ03	FUMAAC01	1	1	Yes	No
GOKCEQ	HIQWEJ03	FORMAC01	1	2	Yes	No
LOYRIC	HIQWEJ03	BENZDC11	1	2	Yes	No
MEWNUA	HIQWEJ03	ADOGUW	1	2	Yes	No
NUJFEF	HIQWEJ03	HOPHAL11	1	1	Yes	No
PAVXAN	HIQWEJ03	TEPHTH14	1	1	Yes	No
ROQYED	HIQWEJ03	ROQXUS	1	2	Yes	No
RUXMAZ	HIQWEJ03	ETYNBZ01	1	1	No	No
SITDIJ	HIQWEJ03	ACETAC03	1	2	No	No
SOVFOY01	HIQWEJ03	TGLYCL01	1	1	Yes	No
UCEXIJ	HIQWEJ03	PAXNIL	1	1	Yes	No
VEXQOE	HIQWEJ03	CYURAC14	1	2	Yes	No
XOWKEB01	HIQWEJ03	XYANAC06	1	2	Yes	Yes
KIGLUI01	HXACAN13	BAPLOT04	1	1	Yes	No
LUJSOZ	HXACAN13	PHENAZ11	1	2	Yes	No
CUQKAC	HXACAN13	WEMWEQ15	1	1	Yes	No
LUJSIT	HXACAN13	NAPHTA15	2	1	Yes	No
LUJTAM	HXACAN13	OXALAC05	1	1	Yes	No
LUJTAM	HXACAN13	OXALAC11	1	1	Yes	No
DIPJAQ	ACSALA05	BAPLOT04	1	1	Yes	No
HUNJEH	ACSALA05	JAKGEH	1	1	Yes	Yes
SIBYUA	ACSALA05	HIQWEJ03	2	1	Yes	No
TAZRAO	ACSALA05	CBMZPN21	1	1	Yes	No

Note: The hydrogen bonds were identified using the CSD mercury code from the experimental crystal structure with H-bonds defined as a distance and angle within 0.05–0.95 quantiles with intramolecular hydrogen bonds separated by >3 bonds.

The crystal structures were also optimized using the DFTB3+ software version 19.1 for all DFTB calculations[28] method to gage the balance between model cost and accuracy. The DFTB calculations were run with the Grimme D3 dispersion correction[37] and three different dampening methods, standard Becke-Johnson, the H5[38,39] and Boese[40] models. The reduced computational cost of the DFTB method comes with minimal basis sets[28] that generally require corrections, such as the H5 method that improves the description of hydrogen bonded systems.[38,39] The DFTB method can also be improved by fitting the geometries with the Boese method, aiming to capture dispersion effects.[40] These different methods of improving the predictive power of the DFTB method can be incorporated along with the dispersion corrections as described above (in this case, DFTB-D), while keeping the computational overhead at a fraction of that of full DFT. Further details on the DFT and DFTB optimization methods are provided in Data S1.

2.3 | Estimation of cocrystal stability

Assuming that the entropic contribution to the free energy is negligible going from the single component crystals to the cocrystal, which is reasonable for many sets of similarly sized, neutral molecules, the driving force for the cocrystal formation can be approximated using Equation (2)[1]:

$$\Delta E_{\mathrm{cocryst}}(\mathrm{A}_n\mathrm{B}_m) = E_{\mathrm{tot}}(\mathrm{A}_n\mathrm{B}_m) - [mE_{\mathrm{tot}}(\mathrm{A}) + nE_{\mathrm{tot}}(\mathrm{B})], \qquad (2)$$

here, $E_{\mathrm{tot}}(\mathrm{A}_n\mathrm{B}_m)$ is the calculated enthalpy of the cocrystal, and $E_{\mathrm{tot}}(\mathrm{A})$ and $E_{\mathrm{tot}}(\mathrm{B})$ are the enthalpies of the single-component crystal structures of functional molecule A alone and coformer B alone, respectively, with *m* and *n* defining the stoichiometric ratio of A and B in the cocrystal. Table 1 shows the list of cocrystals used in this study,

TABLE 2 The reference single component crystals.

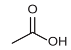
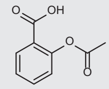
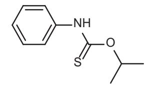
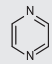
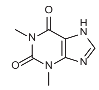
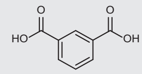
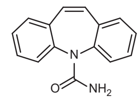
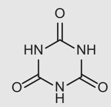
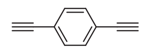
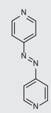
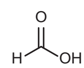
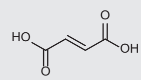
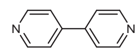
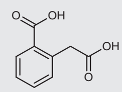
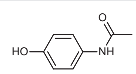
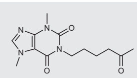
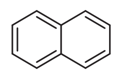
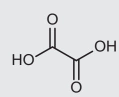
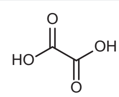
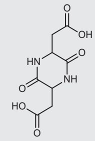
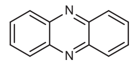
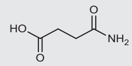
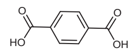
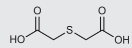
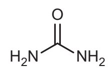
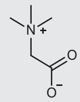
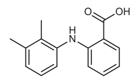
CSD reference code	Name	H-bond[a] Inter	H-bond[a] Intra	Chemical structure
ACETAC03	Acetic acid	Yes	No	
ACSALA05	Aspirin	Yes	Yes	
ADOGUW	O-isopropyl N-phenyl thiocarbamate	No	No	
BACJUN	Pyrazine	No	No	
BAPLOT04	Theophylline	Yes	No	
BENZDC11	Isophthalic acid	Yes	No	
CBMZPN21	Carbamazepine	Yes	No	
CYURAC14	1,3,5-triazinane-2,4,6-trione	Yes	No	
ETYNBZ01	1,4-diethynylbenzene	No	No	
EVESIJ04	Azapyridine	No	No	
FORMAC01	Formic acid	Yes	No	
FUMAAC01	Fumaric acid	Yes	No	
HIQWEJ03	4,4′-bipyridine	No	No	
HOPHAL11	2-(carboxymethyl) benzoic acid	Yes	No	
HXACAN13	Paracetamol	Yes	No	
JAKGEH	Pentoxifylline	No	No	
NAPHTA15	Naphthalene	No	No	
OXALAC05	Oxalic acid alpha polymorph	Yes	No	
OXALAC11	Oxalic acid beta polymorph	Yes	No	

TABLE 2 (Continued)

CSD reference code	Name	H-bond[a] Inter	H-bond[a] Intra	Chemical structure
PAXNIL	Cyclo-diaspartic acid	Yes	Yes	
PHENAZ11	β-Phenazine	No	No	
ROQXUS	Succinamic acid	Yes	No	
TEPHTH14	Terephthalic acid	Yes	No	
TGLYCL01	Thiodiglycolic acid	Yes	No	
UREAXX12	Urea	Yes	No	
WEMWEQ15	Betaine	No	No	
XYANAC06	2-[(2,3-dimethylphenyl) amino] benzoic acid	Yes	Yes	

[a]Presence of intermolecular and intramolecular H-bonds in the crystal structure as measured by the CSD mercury code.

including their CSD reference codes, molecular composition in the cocrystal, and the indication of the presence or absence of H-bonds in the cocrystal structure. Meanwhile, Table 2 has the complete list of the single component crystals, which are included in Table 1 with their CSD code, chemical name, chemical structure, and presence or absence of both intermolecular and intramolecular H-bonds in the crystal structure.

2.4 | Crystal structure similarity

To check the optimized crystal structures, the crystal packing similarity[41] function, $RMSD_{20}$, of the CSD Mercury 4.0[42] was used. This is done by comparing 20 molecules from the starting molecule to the calculated one. This was given a 30% distance tolerance and 30° tolerance, for all heavy, nonhydrogen atoms. These figures are included in the supplemental material Tables S2–S5.

3 | RESULTS AND DISCUSSION

3.1 | DFT results

We start our study with the alpha polymorph of oxalic acid (OXALAC05) cocrystals, including five different common coformers from small molecules of urea and pyrazine to larger molecules including phenazine, theophylline, and azapyradine (Tables 1 and 2). From Table 1, it can be noted that all oxalic acid cocrystals have hydrogen bond interactions.

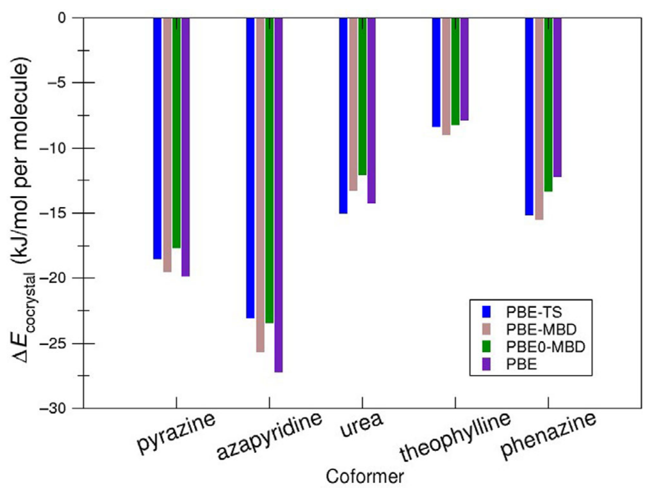


FIGURE 1 Predicted relative stabilities of five oxalic acid-based cocrystals lattice enthalpy, $\Delta E_{cocrystal}$, calculated using three dispersion-inclusive DFT models and one uncorrected DFT model.

The predicted lattice enthalpies of the oxalic acid cocrystals are presented in Figure 1. All four methods, PBE, PBE-TS, PBE-MBD, and PBE0-MBD, rank the azapyridine cocrystal (UMINAF) as the most stable cocrystal with the pyrazine cocrystal (GUDSUV) being the 2nd most stable and identify the cocrystal with theophylline (XEJWUF) as the least stable. The stability rank of the remaining two oxalic acid cocrystals, with phenazine (XAPMIK), and urea (UROXAM01), depends on the method implemented. All the DFT computational methods used show negative values for the enthalpy of oxalic acid cocrystal. As shown in Figure 1, the four different DFT computational methods provide similar values for the enthalpy of cocrystal for each oxalic acid cocrystal. A clear message from Figure 1 is the lack of dependence on the dispersion method used on the $\Delta E_{cocryst}(A_nB_m)$ estimation for oxalic acid cocrystal. This lack of dependence on the dispersion method is expected due to the presence of hydrogen bond interactions in all the oxalic acid cocrystals (Table 1), as hydrogen bond interactions are well described by the exchange correlation functionals.[43]

The next subset of cocrystals studied are for 4,4′-bipyridine. The stability of the single component crystal of 4,4′-bipyridine (HIQWEJ03) is primarily driven by dispersive interactions, as the molecule does not have any H bond donor group, and two nitrogen atom H bond acceptor groups (Table 2). These hydrogen bond acceptor nitrogen atoms have potential for strong hydrogen bond interactions in cocrystals with coformers with hydrogen bond donor groups. Figure 2 shows the calculated lattice enthalpy of the cocrystals for 4,4′-bipyridine.

For all but two of the 13 cocrystals (Table 3), the four DFT computational methods agree on large negative values for the cocrystals formation energy (Figure 2). The two cocrystals where the cocrystal enthalpy is dependent upon the computational DFT method are O-Isopropyl N-phenylthiocarbamate (no. 4, MEWNUA) and 2-[(2,3-dimethylphenyl) amino]benzoic acid (no. 13, XOWKEB01). The discrepancy between the

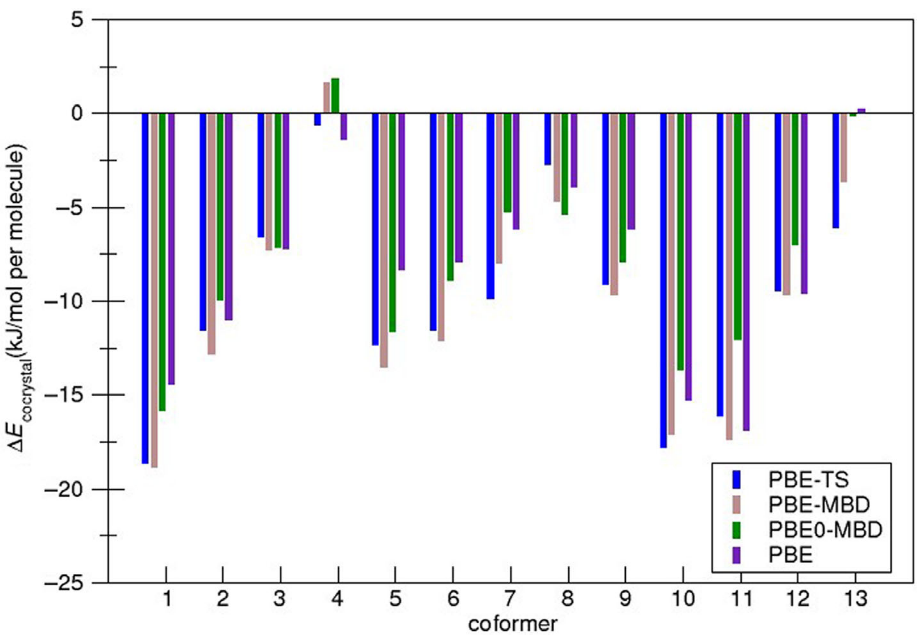


FIGURE 2 Predicted relative stability for thirteen 4,4′-bipyridine-based cocrystals lattice enthalpy, $\Delta E_{cocrystal}$, calculated using three DFT-D models and uncorrected DFT model. Coformers 1–13 on the horizontal axis are identified in Table 3.

TABLE 3 List of 4,4′-bipyridine coformers.

	Coformer name	Coformer CSD reference code	Cocrystal CSD reference code
1	1,3,5-triazinane-2,4,6-trione	CYURAC14	VEXQOE
2	Fumaric acid	FUMAAC01	GIPQEB01
3	Thiodiglycolic acid	TGLYCL01	SOVFOY01
4	O-isopropyl N-phenylthiocarbamate	ADOGUW	MEWNUA
5	1,4-diethynylbenzene	ETYNBZ01	RUXMAZ
6	Acetic acid	ACETAC03	SITDIJ
7	Isophthalic acid	BENZDC11	LOYRIC
8	Terephthalic acid	TEPHTH14	PAVXAN
9	2-(carboxymethyl)benzoic acid	HOPHAL11	NUJFEF
10	Formic acid	FORMAC01	GOKCEQ
11	Cyclo-diaspartic acid	PAXNIL	UCEXIJ
12	Succinamic acid	ROQXUS	ROQYED
13	2-[(2,3-dimethylphenyl)amino]benzoic acid	XYANAC06	XOWKEB01

small negative values with uncorrected PBE and PBE-TS and the small positive values with PBE-MBD and PBE0-MBD in the case of MEWNUA may be numerical noise but nevertheless this cocrystal is predicted to be weak. For no. 13, the near-zero prediction for uncorrected PBE and PBE0-MBD but modest negative values for PBE-TS and PBE-MBD is difficult to rationalize, though we note that PBE-MBD predicts the largest negative enthalpy for eight of the cocrystals, consistent with more complete treatment of dispersive interactions by MBD.[44] The cocrystal with 1,4-diethnylbenzene (no. 5, RUXMAZ) is the only cocrystal where there are no hydrogen bond interactions in the cocrystal nor in their single component crystals, as reflected in the most significant and consistent percentage increase in magnitude of negative lattice enthalpy on switching from uncorrected PBE to the three dispersion-inclusive models.

The degree of dependence of results on the choice of model is increased as the complexity of the functional molecule is increased, as demonstrated by the calculated stabilities of cocrystals for aspirin or paracetamol. Both APIs engage in both hydrogen bonding and π–π stacking with van der Waals stabilization. Previous calculations on aspirin have shown that the crystal packing relies heavily on dispersive interactions.[45] The paracetamol results, in general, show favorable, negative values for the formation of the six available cocrystals (Table 1 and Figure 3). Only the naphthalene (LUJSIT) and phenazine (LUJSOZ) cocrystals are predicted as unfavorable in some cases by two out of four models.

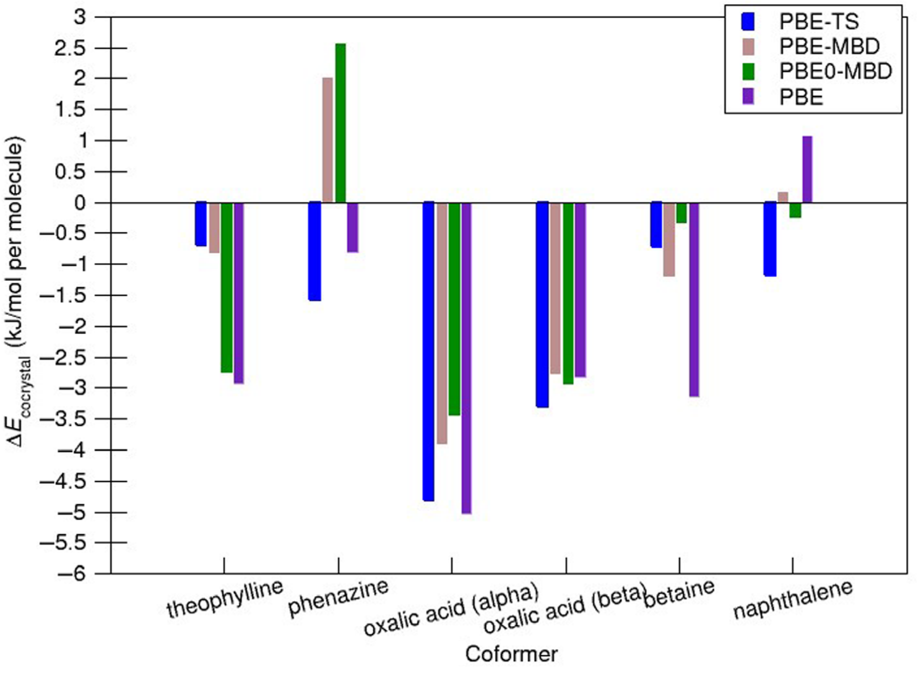


FIGURE 3 Predicted relative stabilities for six paracetamol-based cocrystals lattice enthalpy, $\Delta E_{\text{cocrystal}}$, calculated using three DFT-D models and uncorrected DFT model.

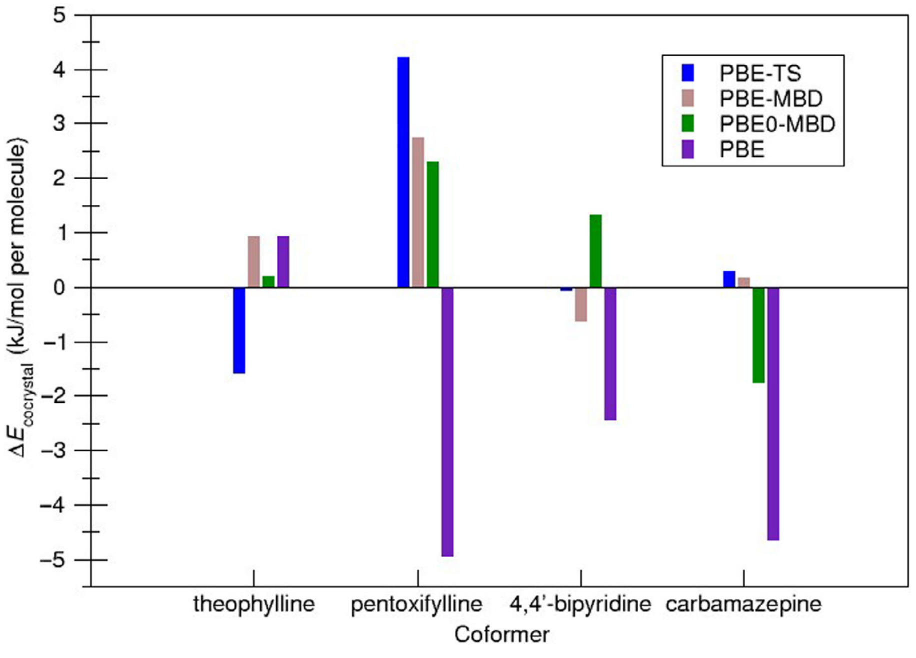


FIGURE 4 Predicted relative stabilities for four aspirin-based cocrystals lattice enthalpy, $\Delta E_{\text{cocrystal}}$, calculated using three DFT-D models and uncorrected DFT model.

Paracetamol oxalic acid (LUJTAM) is the most stable cocrystal, reflecting the strong hydrogen bonding character of the coformer, which may also influence its compressibility in paracetamol tabletting.[46]

The final cocrystal dataset studied using the full DFT methodology are the aspirin cocrystals. The predicted enthalpy of formation for the aspirin cocrystals using uncorrected and dispersion-inclusive DFT is presented in Figure 4. Aspirin cocrystals are clearly challenging to predict, with some enthalpies of cocrystals predicted as positive, some negative and some close to zero. Uncorrected PBE performs best and only fails for the cocrystal with theophylline. This result highlights the challenge of predicting the stability using DFT models for these complex cocrystals with mixed hydrogen bonding and van der Waals contacts, with dispersion corrections likely overestimated and leading to unphysical cocrystal structures.

Overall, the DFT results show that the enthalpy of cocrystallization is negative and so thermodynamically favorable for the majority of the cocrystals. The PBE-TS method predicts 93% of cocrystals with negative enthalpy. When PBE-MBD and PBE0-MBD methods are used, negative enthalpy is predicted for 83% of the cocrystals. It is important to highlight that the entropy contribution to the free energy was not calculated in our study, therefore cocrystals structures with positive values of enthalpy of formation may be stabilized through entropic contributions.

3.2 | DFTB results

Full DFT calculations represent a medium to low throughput method of predicting cocrystal stability for assemblies of low to moderate

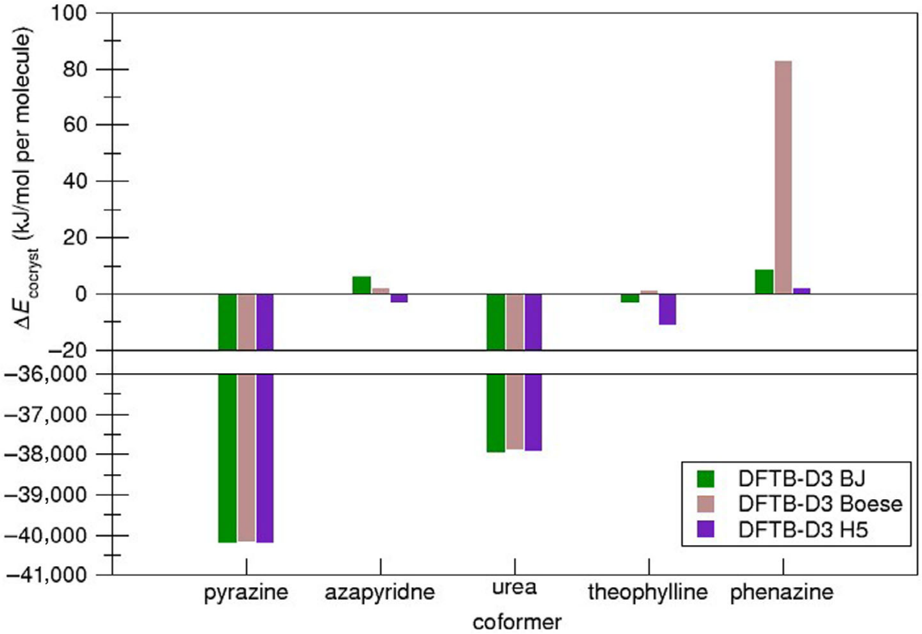


FIGURE 5 Predicted relative stabilities for five oxalic acid-based cocrystals lattice enthalpy, $\Delta E_{\text{cocrystal}}$, calculated using three DFTB+ parameterizations.

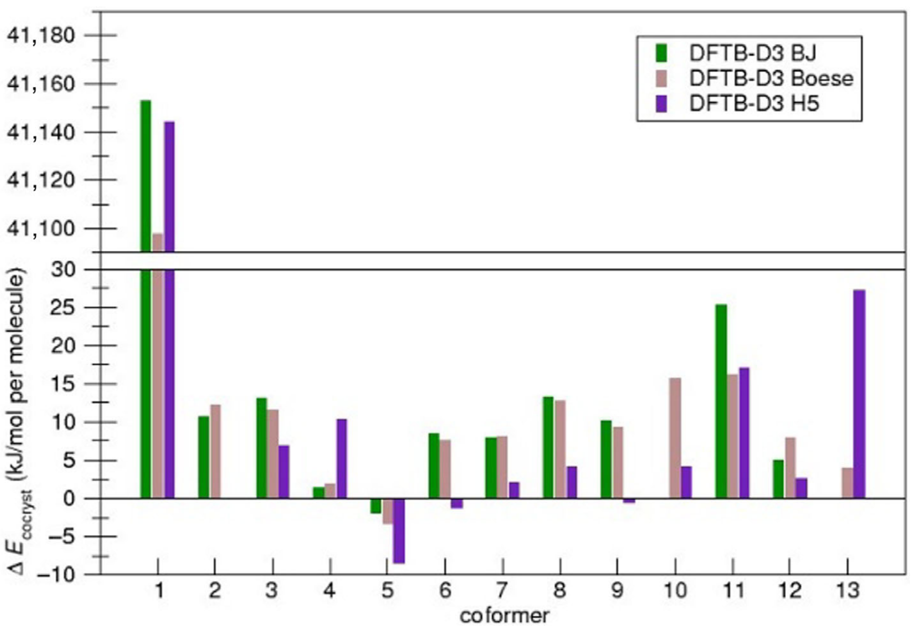


FIGURE 6 Predicted relative stabilities for thirteen 4,4′-bipyridine-based cocrystals lattice enthalpy, $\Delta E_{\text{cocrystal}}$, calculated using three DFTB+ parameterizations.

complexity in terms of crystal size and molecular weights. Faster high-throughput methods more suitable for large datasets and/or large crystals, may be in principle obtained using DFTB methods. Here, to directly compare with full DFT, we test the predictive power of DFTB methods for the same set of cocrystals. We compared the effectiveness of three methods of DFTB-D, one with standard Becke-Johnson (BJ) dampening and two using the Boese method and H5 parametrisation.

Compared with DFT, the computationally more expedient but approximate DFTB-D method showed significantly less predictive power. For cocrystals with oxalic acid (Figure 5), the assemblies formed with pyrazine or urea are predicted to be stable. Note the very large absolute values of the energies in some cases, reflecting numerical instability and poor quantitative power of DFTB-D to predict energies. Of the three DFTB methods tested, H5 parameterization performed best, predicting stable cocrystals for all but the with cocrystal phenazine, reflecting its ability to model hydrogen bonding more effectively than the alternative schemes.

One limitation of the DFTB-D method is the possibility of large errors in the calculation of weak dispersive interactions, such as van der Waals packing, as a result of the semi-local origins of the DFTB approximation.[28]

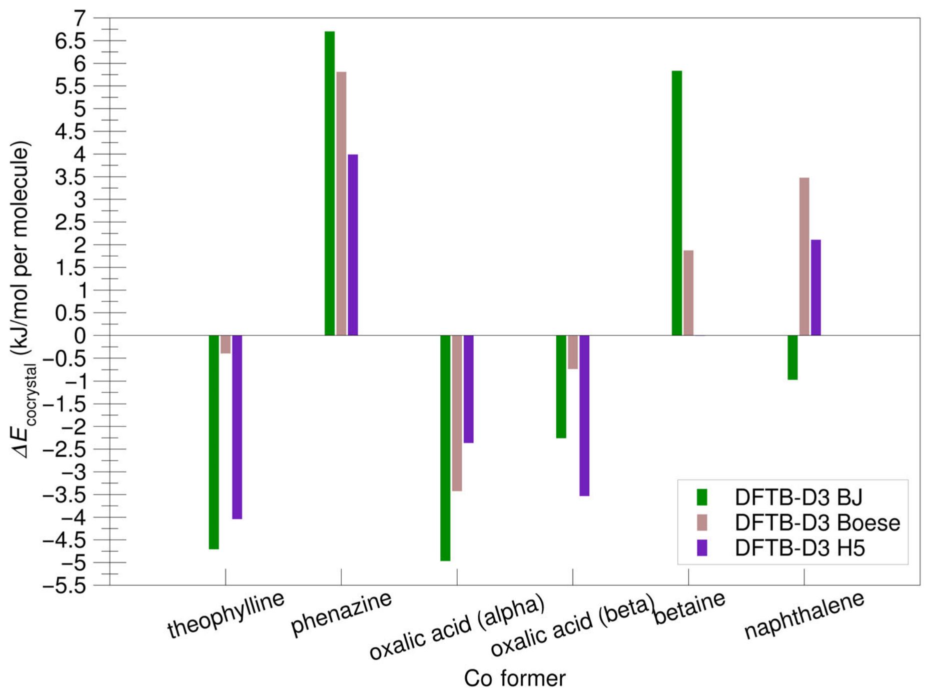


FIGURE 7 Predicted relative stabilities for six paracetamol-based cocrystals lattice enthalpy, $\Delta E_{cocrystal}$, calculated using three alternative DFTB+ parameterizations.

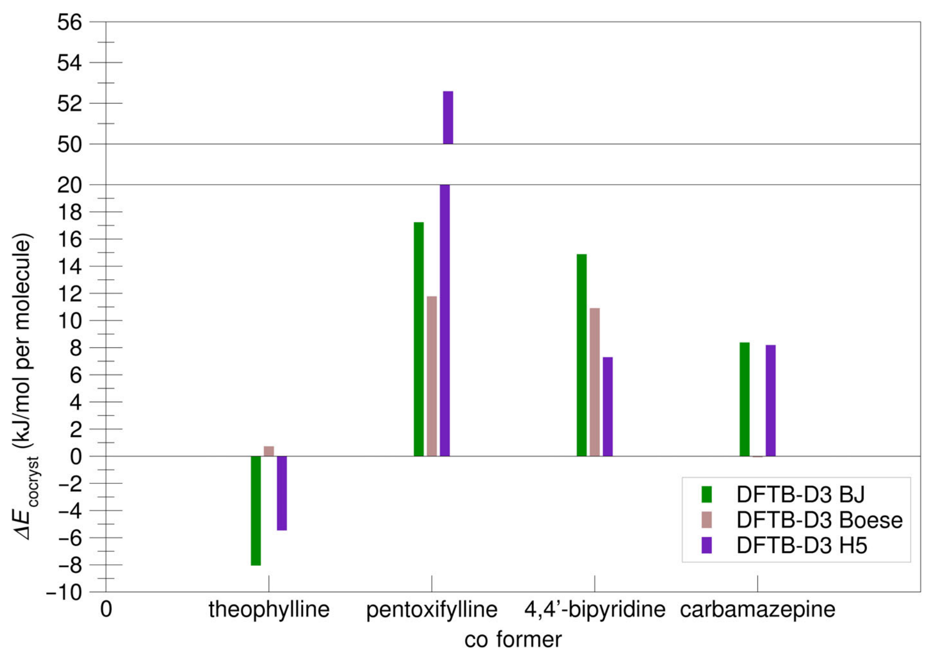


FIGURE 8 Predicted relative stabilities for four aspirin-based cocrystals lattice enthalpy, $\Delta E_{cocrystal}$, calculated using three DFTB+ parameterizations.

The limitations of the approximations made in the DFTB-D methods are clearly evident for the 4,4′-bipyridine cocrystal dataset (Figure 6), for which DFTB-D fails. Note that when the DFTB-D3 H5 parameterization was used for the fumaric acid (no. 2, GIPQEB01) cocrystal and DFTB-D3 BJ for the formic acid cocrystal (and no. 10, GOKCEQ), numerical instabilities prevented the successful completion of the calculation.

Not surprisingly then, DFTB-D also fails for the more complex cocrystals with APIs paracetamol or aspirin. For paracetamol cocrystals (Figure 7), DFTB-D3 BJ performs better than the others but is still unreliable, erroneously predicting unstable structures for two out of the six cocrystal lattices. Aspirin cocrystals, which also challenged DFT to some extent (Figure 4) cannot be predicted by DFTB (Figure 8).

Overall, the DFTB-D method do not predict stable lattice energies for these known, synthesized cocrystals, highlighting the poor predictability of these putative high-throughput methods for cocrystal design. This is consistent with the initial study of the Boese parameterization as this significantly improved the unit-cell volume mean average error with a detriment to the mean average error of the lattice enthalpy.[40] This could be the reason for the poor prediction of the oxalic acid and 4,4′-bipyridine cocrystals.

4 | CONCLUSIONS

Our results show that the prediction of cocrystal stability is heavily dependent upon the computational model used. For simple cocrystals, such as the oxalic acid and 4,4′-bipyridine datasets, the different DFT models largely agreed. However, for more complex cocrystals, such as the paracetamol and aspirin datasets with more diverse interactions, the enthalpy of formation depends on the DFT model, as seen most dramatically in the case of aspirin-based cocrystals.

Implementation of the computationally more demanding PBE0-MBD calculations did not significantly change the predicted lattice enthalpy of formation of the chosen cocrystals, relative to the less computationally demanding PBE functional. The dispersion corrections TS and MBD performed similarly well overall.

Our dataset confirms previous findings that cocrystallization is a thermodynamically favorable process, with the PBE-TS method predicting favorable assemblies for 93% of cocrystals. The alternative PBE-MBD and PBE0-MBD methods predict 83% to be favorable. Structures with predicted positive values of enthalpy of formation may be stabilized through entropic contributions in crystallization experiments.

Within this cocrystal dataset, the vibrational energy was not included, which could be reflected in some cocrystal structures being predicted to have marginally positive relative stability when applying the 0 K DFT methodology, with previous work indicating that molecular crystals have on average 1 kJ/mol vibrational energy contribution.[47] With continuing improvements in hardware and software, and increasing access to high-performance computing, computation of the vibrational contribution and corresponding entropic contribution to the predicted formation free energies is becoming more feasible for large datasets of cocrystals with diverse functional and coformer molecules.

By contrast, the computationally lighter DFTB-D method predicts a high population of unstable cocrystals with large positive and values of enthalpy. This failure of prediction is particularly prominent for the more complex systems, such as the aspirin cocrystals. The difficulty in predicting these systems could be due to the semi-local origins of this method, resulting in the lack of dispersion forces,[32] which are only marginally compensated by dampening corrections. Hence, using DFTB-D for cocrystal research at a significantly lower computational cost, which would allow for the study of larger crystal structures, requires careful preparametrization and testing. Any such preparametrization will need benchmarking against full DFT to ensure that the DFTB-D model is used only in cases where it can overcome, or be improved to overcome, the semi-local approximation. Perhaps the development of more advanced dispersion/dampening schemes that do not carry prohibitive computational overhead could achieve the required balance between accuracy and computational cost, with integration of AI methods also promising high predictive power for large, complex datasets.[48–51]

ACKNOWLEDGMENTS

D.T. and R.F. thank Science Foundation Ireland (SFI) for financial support under grant number 12/RC/2275_P2 (SSPC), and for provision of computing resources at the SFI/Higher Education Authority Irish Centre for High-End Computing (ICHEC).

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Robert Fox https://orcid.org/0000-0003-4591-1396
Joaquin Klug https://orcid.org/0000-0002-6820-0899
Damien Thompson https://orcid.org/0000-0003-2340-5441

REFERENCES

[1] C. R. Taylor, G. M. Day, *Cryst. Growth Des.* **2018**, *18*, 892.
[2] P. A. Wood, N. Feeder, M. Furlow, P. T. A. Galek, C. R. Groom, E. Pidcock, *CrystEngComm* **2014**, *16*, 5839.
[3] S. Aitipamula, R. Banerjee, A. K. Bansal, K. Biradha, M. L. Cheney, A. R. Choudhury, G. R. Desiraju, A. G. Dikundwar, R. Dubey, N. Duggirala, P. P. Ghogale, S. Ghosh, P. K. Goswami, N. R. Goud, R. R. K. R. Jetti, P. Karpinski, P. Kaushik, D. Kumar, V. Kumar, B. Moulton, A. Mukherjee, G. Mukherjee, A. S. Myerson, V. Puri, A. Ramanan, T. Rajamannar, C. M. Reddy, N. Rodriguez-Hornedo, R. D. Rogers, T. N. G. Row, P. Sanphui, N. Shan, G. Shete, A. Singh, C. C. Sun, J. A. Swift, R. Thaimattam, T. S. Thakur, R. K. Thaper, S. P. Thomas, S. Tothadi, V. R. Vangala, N. Variankaval, P. Vishweshwar, D. R. Weyna, M. J. Zaworotko, *Cryst. Growth Des.* **2012**, *12*, 2147.
[4] E. Grothe, H. Meekes, E. Vlieg, J. H. Horst, R. D. Gelder, *Cryst. Growth Des.* **2016**, *16*, 3237.
[5] J. W. Steed, *Trends Pharmacol. Sci.* **2013**, *34*, 185.
[6] S. L. Childs, M. J. Zaworotko, *Cryst. Growth Des.* **2009**, *9*, 4208.
[7] A. M. Healy, Z. A. Worku, D. Kumar, A. M. Madi, *Adv. Drug Delivery Rev.* **2017**, *117*, 25.

[8] O. N. Kavanagh, D. M. Croker, G. M. Walker, M. J. Zaworotko, *Drug Discovery Today* **2019**, *24*, 796.
[9] A. A. Dar, S. Rashid, *CrystEngComm* **2021**, *23*, 8007.
[10] K. Honer, E. Kalfaoglu, C. Pico, J. Mccann, J. Baltrusaitis, *ACS Sustainable Chem. Eng.* **2017**, *5*, 8546.
[11] S. Kalepu, V. Nekkanti, *Acta Pharm. Sin. B* **2015**, *5*, 442.
[12] D. Sun, W. Gao, H. Hu, S. Zhou, *Acta. Pharma. Sin. B*, **2022**, *12*, 3049.
[13] N. Pandey, A. Ghosh, *J. Drug Delivery Sci. Technol.* **2022**, *76*, 1773.
[14] S. L. Morissette, S. Soukasene, D. Levinson, M. J. Cima, Ö. Almarsson, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2180.
[15] E. R. Johnson, G. A. DiLabio, *Chem. Phys. Lett.* **2005**, *419*, 333.
[16] A. Otero-De-la-Roza, L. M. LeBlanc, E. R. Johnson, *Phys. Chem. Chem. Phys.* **2020**, *22*, 8266.
[17] A. G. Shtukenberg, C. T. Hu, Q. Zhu, M. U. Schmidt, W. Xu, M. Tan, B. Kahr, *Cryst. Growth Des.* **2017**, *17*, 3562.
[18] A. M. Reilly, A. Tkatchenko, *J. Chem. Phys.* **2013**, *139*, 024705.
[19] J. Hermann, A. Tkatchenko, *Phys. Rev. Lett.* **2020**, *124*, 146401.
[20] P. Xu, M. Alkan, M. S. Gordon, *Chem. Rev.* **2020**, *120*, 12343.
[21] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, S. Grimme, *J. Chem. Phys.* **2019**, *150*, 154122.
[22] S. Grimme, A. Hansen, J. G. Brandenburg, C. Bannwarth, *Chem. Rev.* **2016**, *116*, 5105.
[23] G. Sun, Y. Jin, S. Li, Z. Yang, B. Shi, C. Chang, Y. A. Abramov, *J. Phys. Chem. Lett.* **2020**, *11*, 8832.
[24] A. Tkatchenko, M. Scheffler, *Phys. Rev. Lett.* **2009**, *102*, 073005.
[25] A. Tkatchenko, R. A. Distasio, R. Car, M. Scheffler, *Phys. Rev. Lett.* **2012**, *108*, 236402.
[26] A. Ambrosetti, A. M. Reilly, R. A. Distasio, A. Tkatchenko, *J. Chem. Phys.* **2014**, *140*, 1818A508.
[27] J. G. Brandenburg, S. Grimme, *J. Phys. Chem. Lett.* **2014**, *5*, 1785.
[28] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitric, A. Dominguez, S. Ehlert, M. Elstner, T. Van Der Heide, J. Hermann, S. Irle, J. J. Kranz, C. Köhler, T. Kowalczyk, T. Kubař, I. S. Lee, V. Lutsker, R. J. Maurer, S. K. Min, I. Mitchell, C. Negre, T. A. Niehaus, A. M. N. Niklasson, A. J. Page, A. Pecchia, G. Penazzi, M. P. Persson, J. Řezáč, C. G. Sánchez, M. Sternberg, M. Stöhr, F. Stuckenberg, A. Tkatchenko, V. W. Z. Yu, T. Frauenheim, *J. Chem. Phys.* **2020**, *152*, 124101.
[29] F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *B58*, 380.
[30] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, *105*, 9982.
[31] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
[32] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, M. C. Payne, *Z Kristallog* **2005**, *220*, 567.
[33] S. Grimme, *J. Comput. Chem.* **2004**, *25*, 1463.
[34] B. Rana, G. J. O. Beran, J. M. Herbert, *Mol. Phys.* **2023**, *121*, e2138789.
[35] W. A. Al-Saidi, V. K. Voora, K. D. Jordan, *J. Chem. Theory Comput.* **2012**, *8*, 1503.
[36] D. Vanderbilt, *Phys. Rev. B* **1990**, *41*, 7892.
[37] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* **2010**, *132*, 154104.
[38] J. Řezáč, *J. Chem. Theory Comput.* **2017**, *13*, 4804.
[39] V. M. Miriyala, J. Řezáč, *J. Comput. Chem.* **2017**, *38*, 688.
[40] G. A. Dolgonos, A. D. Boese, *Chem. Phys. Lett.* **2019**, *718*, 7.
[41] J. A. Chisholm, S. Motherwell, *J. Appl. Crystallogr.* **2005**, *38*, 228.
[42] C. F. Macrae, I. Sovago, S. J. Cottrell, P. T. A. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler, P. A. Wood, *J. Appl. Crystallogr.* **2020**, *53*, 226.
[43] L. Kronik, A. Tkatchenko, *Acc. Chem. Res.* **2014**, *47*, 3208.
[44] A. M. Reilly, A. Tkatchenko, *Chem. Sci.* **2015**, *6*, 3289.
[45] G. S. Nichol, W. Clegg, *Cryst. Growth Des.* **2009**, *9*, 1844.
[46] S. Karki, T. Friščić, L. Fabián, P. R. Laity, G. M. Day, W. Jones, *Adv. Mater.* **2009**, *21*, 3817.
[47] J. A. Weatherby, A. F. Rumson, A. J. A. Price, A. Otero De La Roza, E. R. Johnson, *J. Chem. Phys.* **2022**, *156*, 114108.
[48] J. J. Devogelaer, H. Meekes, P. Tinnemans, E. Vlieg, R. de Gelder, *Angew. Chem., Int. Ed.* **2020**, *59*, 21711.
[49] R. Birolo, F. Bravetti, E. Alladio, E. Priola, G. Bianchini, R. Novelli, A. Aramini, R. Gobetto, M. R. Chierotti, *Cryst. Growth Des.* **2023**, *23*, 7898.
[50] C. V. Essen, D. Luedeker, *Drug Discovery Today* **2023**, *28*, 103763.
[51] P. Kim, I.-S. Lee, J.-Y. Kim, M. E. Mswahili, Y.-S. Jeong, W.-J. Yoon, H.-S. Yun, M.-J. Lee, G. J. Choi, *CrystEngComm* **2022**, *24*, 3938.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: R. Fox, J. Klug, D. Thompson, A. Reilly, *J. Comput. Chem.* **2024**, 1. https://doi.org/10.1002/jcc.27454