

Composite and Low-Cost Approaches for Molecular Crystal Structure Prediction

Luc M. LeBlanc,^{*,†} Alberto Otero-de-la-Roza,^{*,‡} and Erin R. Johnson^{*,†}

[†]*Department of Chemistry, Dalhousie University, 6274 Coburg Road, PO Box 15000
Halifax, Nova Scotia, Canada B3H 4R2*

[‡]*Department of Chemistry, University of British Columbia, Okanagan, 3247 University
Way, Kelowna, British Columbia, Canada V1V 1V7.*

E-mail: luc.leblanc@dal.ca; aoterodelaroz@gmail.com; erin.johnson@dal.ca

1 Abstract

Molecular crystal structure prediction (CSP) requires evaluating differences in lattice energy between candidate crystal structures accurately and efficiently. In this work, we explore and compare several low-cost alternatives to dispersion-corrected density-functional theory (DFT) in the plane-waves/pseudopotential approximation, the most accurate and general approach used in CSP at present. Three types of low-cost methods are considered: DFT with a small basis set of finite-support numerical orbitals (the SIESTA method), dispersion-corrected Gaussian small or minimal-basis-set Hartree-Fock and DFT with additional empirical corrections (HF-3c and PBEh-3c), and self-consistent-charge dispersion-corrected density-functional tight binding (SCC-DFTB3-D3). In addition, we study the performance of composite methods that comprise a geometry optimization using a low-cost approach followed by a single-point calculation using the accurate but comparatively expensive B86bPBE-XDM functional. All methods were tested for their abilities to produce absolute lattice energies, relative lattice energies, and crystal geometries. We show that assessing various methods by their ability to produce absolute lattice energies can be misleading, and that relative lattice energies are a much better indicator of performance in a CSP context. The EE14 set of relative sol-

ubilities of homochiral and heterochiral chiral crystals is proposed for relative lattice-energy benchmarking. Our results show that PBE-D2 plus a DZP basis set of numerical orbitals coupled with a final B86bPBE-XDM single-point calculation gives excellent performance at a fraction of the cost of a full B86bPBE-XDM calculation. The B86bPBE-XDM//PBE-D2/DZP method was subsequently tested in a practical CSP application based on our recent CSP study on the crystal structure of the enantiopure and racemate forms of 1-aza[6]helicene, a chiral organic semiconductor. Our results show that this multilevel method is able to correctly reproduce the energy ranking of both crystal forms.

2 Introduction

The computational prediction of a molecular crystal structure (crystal structure prediction, CSP) from the two-dimensional molecular diagram is currently a great challenge in computational chemistry. CSP has many technologically relevant applications in the pharmaceutical industry,¹⁻³ manufacturing of dyes⁴ and explosives,⁵ and organic electronics,^{6,7} since it provides a way to predict the structure and stability of all polymorphic forms of a given compound. A number of CSP strategies have been used over the past few decades, and progress in the field has been tracked by a series of

blind test competitions held by the Cambridge Crystallographic Data Center (CCDC).^{8–13} In these blind test contests, participants are given the molecular diagrams for a few organic compounds and are asked to predict the experimental crystal structure, which is unknown to the them. The sixth and most recent blind test was completed last year.¹³ Many excellent reviews have highlighted the progress made in the development of computational protocols for CSP, while outlining the difficulties that remain.^{3,13–18}

All current CSP protocols work under the assumption that the experimentally observed crystal structure coincides with the thermodynamically stable phase¹⁹ (i.e. kinetic effects are neglected). A CSP protocol must sample a complex highly-dimensional energy landscape efficiently and must also be able to rank reliably candidate crystal structures by their free energy.²⁰ Usually, the electronic lattice energies are used instead of free energies due to the cost and the difficulty of evaluating the vibrational free-energy contributions accurately. These vibrational terms are typically added only in the final candidate ranking step, if at all.^{18,21} Even if one focuses on calculating electronic lattice energies alone, it is estimated that over 80% of all isolable polymorphic forms are separated by less than ca. 4.2 kJ/mol.^{1,21} This is a very strict requirement that poses a challenge to current computational methods regarding their ability to produce accurate relative lattice energy differences in molecular crystals.

In recent years, some of us have highlighted the good performance of the B86bPBE exchange-correlation functional^{22,23} paired with the exchange-hole dipole moment (XDM) model^{24,25} in modeling intermolecular interactions. We have shown that B86bPBE-XDM yields lattice energies of molecular organic crystals^{26,27} within an accuracy neighboring that of the experimental uncertainty, i.e., 4.2 kJ/mol.^{28,29} B86bPBE-XDM was also successful when applied as the final ranking function on the candidate structures proposed by various groups in the first five blind tests.^{30,31} Similar dispersion-corrected density functionals can be successfully applied to CSP

as well.^{32–35} Two examples currently in use as part of CSP protocols are Grimme’s D3 dispersion correction^{36,37} and Tkatchenko *et al.*’s many-body dispersion (MBD) model.^{38,39} The X23 set of experimental back-corrected lattice energies⁴⁰ (an extension of the C21 set²⁶) is typically used to assess the accuracy of computed lattice energies and geometries for organic molecular crystals. Mean absolute errors (MAE) in the predicted lattice energies for the X23 molecular-crystal benchmark set^{26,40} obtained with the B86bPBE-XDM, PBE0-MBD//PBE0-TS, and PBE-D3 methods are 3.6 kJ/mol,²⁷ 3.9 kJ/mol,⁴⁰ and 4.5 kJ/mol,⁴¹ respectively.

While these dispersion-corrected DFT methods provide lattice energies accurate enough for reliable CSP (except in specific cases^{30,31}), their computational cost is quite high.¹³ This is especially problematic in the early stages of the search, when it is necessary to discard the less stable candidate structures.⁴¹ To address this problem, one can resort to using a multistep approach, in which computationally inexpensive (in the following “low-cost” or “cheap”) methods are used to prune the list of candidate structures.⁴² These low-cost methods are often empirical or semi-empirical in nature (e.g. force fields) or resort to drastic approximations, such as the use of empirically corrected Gaussian minimal-basis-set calculations.⁴³ Low-cost methods also speed up the energy calculation by several orders of magnitude and allow for more points on the crystal energy landscape to be surveyed. However, it is critical that they are able to identify reliably whether a candidate is within a certain energy range above the global minimum. Otherwise, low-energy polymorphic forms will be lost in the pruning stage.⁴²

Several low-cost methods have been developed recently. Density-functional tight-binding^{44–46} (DFTB) paired with the D3 dispersion correction was shown to yield a MAE on the X23 lattice energies of 10.4 kJ/mol,⁴⁷ although it is important to note that cell parameters were kept fixed during optimization. The PBEh-3c⁴⁸ and HF-3c^{41,49,50} methods use small or minimal Gaussian basis sets coupled with *ad hoc* corrections, and have MAEs of

5.4 kJ/mol⁴⁸ and 8.4 kJ/mol,⁵¹ respectively. In the case of HF-3c, the MAE can be reduced to 6.7 kJ/mol if some of the correction parameters are tuned such that the resulting geometries are closer to the experimental crystal structures.⁵¹

In a recent study,⁵² Carter and Rohl evaluated the performance of the vdW-DF1⁵³ and vdW-DF2⁵⁴ non-local functionals with locally-supported numerical orbitals (SIESTA^{55,56}) on the C21 set,²⁶ and compared their results to the same calculation using plane-waves/pseudopotentials.⁵⁷ The authors found that, by using vdW-DF2 in SIESTA with a double- ζ plus polarization (DZP) basis set and a counterpoise correction, the resulting MAE (3.8 kJ/mol) is similar to B86bPBE-XDM (3.6 kJ/mol)²⁷ and much smaller than either using vdW-DF2 in ESPRESSO (8.2 kJ/mol) or the same method without the counterpoise correction (22.6 kJ/mol). Other options, such as using a triple- ζ plus polarization (TZP) basis set at the DZP geometries (MAE = 10.2 kJ/mol) were also explored.^{52,58}

While low-cost methods are typically assessed based on their ability to produce accurate lattice energies, this is a flawed measure of the accuracy and reliability of a CSP protocol; instead, relative lattice energies would be a more suitable measure. For instance, the enantiomeric excesses (ee) of a scalemic solution in contact with the racemate and enantiopure phases of various organic compounds can be used. This measure is accessible experimentally, and provides an excellent benchmark tool for energy differences between crystal structures, as shown in our recent work on a set of amino acid crystals.⁵⁹

It is also interesting to find which methods yield accurate crystal geometries in addition to absolute or relative lattice energies.⁶⁰ For instance, a low-cost method able to generate crystal geometries comparable to B86bPBE-XDM would prevent having to run geometry optimizations with the latter, resulting in a large saving in computational cost. This composite approach, which we call a multilevel method in the rest of this article, is common in molecular quantum chemistry (e.g. the Gn series of methods⁶¹⁻⁶⁵). A multilevel approach (TPSS-

D3 at the HF-3c geometries) has been previously tested for polymorphs by Brandenburg and Grimme⁶⁶ on the POLY59 benchmark set. This benchmark set consists of 9 experimental crystal structures from the sixth blind test, 5 of which are polymorphic, supplemented with 10 low-energy candidate structures generated for the blind test. While Brandenburg and Grimme did demonstrate an improvement in the polymorph energy ranking using their multilevel TPSS-D3//HF-3c approach (as opposed to HF-3c alone), the results were not accurate enough to predict the experimental crystal structure as being the lowest-lying energy structure for each system. Note, however, that this result does not necessarily reflect the quality of their approach, given that more than just thermodynamics, i.e., kinetics,^{19,67} can dictate what phases are observed experimentally. Even so, testing low-cost methods for their ability to reproduce equilibrium crystal geometries calculated at a higher level of theory is as important as the relative lattice energies themselves when considering a multilevel method.

In this work, we aim to assess the ability of various low-cost methods (localized Gaussian and numerical basis sets, HF-3c, DFTB) to produce accurate absolute and relative lattice energies and geometries, with the idea of building a multilevel approach using one these methods with a final B86bPBE-XDM single-point energy calculation. The X23 benchmark set,^{26,40} and the set of ten relative solubilities of chiral amino acids reported previously⁵⁹ supplemented with four additional chiral compounds (dubbed herein the EE14 set) are used as test sets for our analysis. As a practical CSP example, the resulting methods are applied to the crystal structure prediction of 1-aza[6]helicene, an organic semiconductor on which a B86bPBE-XDM-based CSP study was recently performed.⁶

3 Methods

The high- and low-cost methods used in this work are now described.

SIESTA (Spanish Initiative for Electronic

Simulations with Thousands of Atoms) is a density-functional-theory method that uses finite-support numerical orbitals to achieve linear asymptotic scaling with system size.^{56,68–70} The homonymous SIESTA program (version 4.0b) was used in this work. The PBE²³ semi-local functional, and the vdW-DF1 and vdW-DF2 non-local functionals^{53,54} are considered here. PBE was coupled with Grimme’s D2 dispersion correction,⁷¹ with the functional- and basis-set-specific fitting parameters equal to those reported by Peverati and co-workers⁷² (e.g. for PBE with a DZP-quality basis set, $s_6 = 0.5$ and $s_r = 1.1$). Troullier-Martins norm-conserving pseudopotentials^{73–75} were generated using the ATOM code.⁷⁶

Double- ζ plus polarization (DZP) basis sets were used for the SIESTA calculations, except in occasional triple- ζ plus polarization (TZP) basis-set calculations. DZP is the standard basis set implemented in SIESTA, while TZP was constructed and optimized by Louwse and Rothenberg⁵⁸ for H, N and O atoms, and further extended to the C atom by Carter and Rohl.⁵² The effect of changing the confinement radius of the finite-support orbitals using SIESTA’s “energy shift” parameter—the energy increase experienced by the orbital when it is confined—was also explored. By lowering the energy shift (0.02 Ry is SIESTA’s default), the orbital confinement radii are extended, increasing the computational cost of the calculation, but reducing errors due to basis set superposition (BSSE).⁷⁷ In previous studies, an energy shift of 0.001 Ry proved to give converged results with respect to relative lattice or binding energies for various systems.^{52,77} Unless otherwise indicated, this value was used in this work. The fineness of the real-space integration grid for charge densities and potentials was set by having plane waves of kinetic energy of 200 Ry or less be represented without aliasing.

CRYSTAL^{78,79} is a code for electronic-structure calculations in periodic solids using localized Gaussian orbitals. In this work, CRYSTAL was used to run Hartree-Fock (HF) and DFT calculations. Specifically, the HF-3c^{41,49,51} and PBEh-3c⁴⁸ methods implemented in CRYSTAL17 were utilized.⁷⁹ Both

of these methods make use small or minimal basis sets (MINI(x)^{41,49} for HF-3c and def2-mSVP⁴⁸ for PBEh-3c) and are supplemented with Grimme’s D3 dispersion^{36,37} and a geometrical counterpoise correction (gCP).^{80,81} HF-3c has an additional correction for short-range basis-set incompleteness errors (SRB).^{41,49,51} The HF-3c calculations used the Becke-Johnson damping function in D3,^{37,82} three-body interactions, and a fit parameter of $s_8 = 0.6143$ as described elsewhere.⁵¹ For comparison, results for HF combined with the MINI(x) basis set plus Grimme’s D2 dispersion as implemented in CRYSTAL14⁷⁸ and no further corrections are presented to evaluate the importance of additional BSSE corrections. In this case, the s_6 fit parameter for the dispersion correction was set to 1.

DFTB+,⁸³ version 1.3, and its implementation of the third-order self-consistent-charge density-functional tight-binding method (SCC-DFTB3-D3(BJ)) was used as another low-cost method. The DFTB semiempirical method is based on the n^{th} -order expansion of local density fluctuations with respect to a reference superposition of atomic electron densities.^{45,46} DFTB uses precomputed two-center electron integrals; the 3OB parametrization set was used in this work.^{84,85} Damping of the hydrogen pair potentials was set to $\zeta = 4.2$ in order to obtain an improved description of hydrogen-bonding.^{47,84} The method was also supplemented with Grimme’s D3 dispersion correction,³⁶ with Becke-Johnson damping,^{37,82} and fit parameters set to $s_8 = 0.5883$, $a_1 = 0.5719$, and $a_2 = 3.6017$ as described elsewhere.⁸¹

Quantum ESPRESSO⁵⁷ (version 5.1), a plane-waves/pseudopotentials code, was used for the high-level reference calculations in this work. Periodic-boundary DFT calculations employed the B86bPBE exchange-correlation functional^{22,23} supplemented with the exchange-hole dipole moment (XDM) model.^{24,25} The damping function parameters in XDM, a_1 and a_2 , were set to 0.6512, and 1.4633Å, respectively. The projector-augmented wave (PAW) method was used.⁸⁶ Wavefunction and density cutoffs were set to 80 Ry and 800 Ry, respectively. Structure

relaxations were performed with tighter thresholds for convergence of the energies and forces, i.e., 10^{-5} Ry and 10^{-4} Ry/bohr, respectively.

Structure Relaxations Atomic coordinates and cell parameters for all crystals were fully relaxed to obtain the minimum-energy structures. For isolated molecules, a sufficiently large simulation cell was used to avoid image interactions, and only the atomic coordinates were relaxed (the exception being with CRYSTAL14/17, where the “MOLECULE” keyword was used to treat molecules as non-periodic systems).

Brillouin-Zone Sampling Unless otherwise stated, a Monkhorst-Pack (MP) $4 \times 4 \times 4$ k-point mesh sampling of the Brillouin zone was used throughout for crystal structures, and sampling at the Γ -point was used for molecules.

Geometry Comparison The POWDIFF utility in CRITIC2⁸⁷ was used to measure the degree of similarity between low-cost and B86bPBE-XDM crystal structures. POWDIFF is based on the comparison of powder diffraction patterns using a cross-correlation function.⁸⁸ A POWDIFF result of zero is an exact match, and one represents maximum dissimilarity.

The computational cost of each method above depends on the type of crystal and, in the case of geometry optimizations, also on the starting geometry. According to our tests, a rough guide is that SIESTA methods employing a DZP basis set and PBEh-3c are about 5 times faster than a plane-waves/pseudopotentials (QE) calculation. In contrast, HF-3c and DFTB are roughly around 50 and 500 times faster than QE. However, crystal geometry optimizations using SCC-DFTB3-D3(BJ) as implemented in DFTB+ require approximately two orders of magnitude more steps than the other methods, which is detrimental to its performance. More detailed timing information is given in the supplementary material (SI).

In the rest of the article, multilevel methods built from a geometry optimization using a low-cost method (A), followed by single-point energy calculation (B; almost always B = B86bPBE-XDM) have been considered. The usual notation for composite methods in chem-

istry (B//A) is used in the rest of the article.

4 Results and Discussion

4.1 Absolute Lattice Energies

We begin by examining the performance of all methods on the crystals of the X23 benchmark set. All of the individually computed lattice energies and geometrical data for structures within the set are presented in the SI. The relevant statistics are summarised in Table 1. The best-performing expensive method (QE block) is B86bPBE-XDM, with a MAE of 3.6 kJ/mol. Similar in performance, although with slightly higher MAEs, are other XDM-corrected methods: PW86PBE-XDM (3.7 kJ/mol), PBE-XDM (4.7 kJ/mol), and BLYP-XDM (5.5 kJ/mol). The popular PBE-D2 functional and the non-local vdW-DF1 and vdW-DF2 functionals have MAEs around 6 kJ/mol, with a computational cost similar to B86bPBE-XDM. Finally, alternative expensive methods such as the rVV10 (as implemented in QE) are clearly not suitable for lattice energies in molecular crystals.

In agreement with previous studies, the use of low-cost methods to determine lattice energies of molecular crystals showed a less than desirable accuracy. The MAE ranges from 5.7 kJ/mol for the relatively expensive PBEh-3c to up to 22.3 kJ/mol for vdW-DF1/DZP. The best-performing low-cost methods are PBEh-3c and HF-3c, although PBEh-3c failed to converge its SCF in some cases (anthracene and naphthalene) due to their small band gaps.⁴⁸ These values are still relatively far from the 3.6 kJ/mol obtained using B86bPBE-XDM, although the most accurate of them, HF-3c and PBEh-3c, are reasonably accurate, and similar in performance to some expensive methods available in QE.^{13,66} All low-cost methods except for HF-3c and PBEh-3c also show a very strong tendency to overestimate lattice energies. This is reasonable if the source of the error is basis set incompleteness, but it is also the case for SCC-DFTB3-D3(BJ).

The MAEs obtained using B86bPBE-XDM

Table 1: Statistics for the X23 set of lattice energies using various methods. Mean error (ME), mean absolute error (MAE), and maximum absolute error (MAX) relative to electronic energies back-corrected from experimentally-measured sublimation enthalpies (EXP)^{26,40} or to fully-relaxed B86bPBE-XDM results (DFT). “A” and “B//A” refer, respectively, to the results obtained from the low-cost optimizations directly or to a B86bPBE-XDM single-point energy calculation at the low-cost equilibrium geometry. All values are in kJ/mol per molecule, except POWDIFF similarity measures (POW), which are dimensionless.

Method (A)	EXP				DFT				
	ME	A MAE	MAX ^c	B//A MAE	A MAE	POW	B//A ME	MAE	MAX ^c
QE									
B86bPBE-XDM	0.5	3.6	13.4 (cyt)	3.6	0.0	0.0000	0.0	0.0	0.0
PW86PBE-XDM	0.4	3.7	14.0 (cyt)	3.7	1.2	0.0191	0.5	0.6	3.1 (ox α)
vdW-DF2	4.2	6.0	14.6 (cyt)	3.5	5.4	0.2136	-1.1	1.1	3.0 (suc)
PBE-XDM	-3.2	4.7	17.9 (cyt)	3.7	2.6	0.2168	0.3	0.5	3.0 (ox β)
rVV10	15.8	15.8	27.5 (ada)	3.7	16.4	0.2168	-0.6	0.6	1.7 (ox β)
PBE-D2	3.7	5.8	18.4 (ada)	4.2	4.3	0.2688	-1.4	1.4	5.5 (ada)
BLYP-XDM	4.1	5.5	15.4 (ada)	4.2	4.6	0.4207	-1.5	1.8	4.8 (ant)
vdW-DF1	4.4	6.3	19.1 (ada)	5.0	6.3	0.6031	-3.5	3.5	7.1 (ant)
SIESTA									
PBE-D2/DZP	7.4	11.4	25.2 (ant)	3.6	10.8	0.0102	0.8	0.8	1.5 (ant)
PBE-D2/DZP ^a	7.3	11.3	26.0 (ant)	6.3	10.6	0.1130	-2.2	3.0	31.6 (ura)
vdW-DF2/DZP	20.1	20.1	37.0 (suc)	3.4	20.6	0.1406	-0.2	0.6	2.0 (ada)
vdW-DF1/DZP	22.3	22.3	47.2 (suc)	4.4	22.8	0.2363	1.1	1.8	5.1 (ox α ,suc)
CRYSTAL									
PBEh-3c/def2mSVP ^b	-1.3	5.7	13.8 (amm)	3.1	4.7	0.1660	-1.0	1.1	2.9 (ure)
HF-3c	-0.5	6.5	17.2 (ox α)	5.8	7.0	0.4146	-3.8	3.8	11.3 (suc)
HF-D2/MINI(x)	21.3	21.9	51.8 (ox β)	5.4	23.5	0.5737	-4.0	4.1	15.1 (suc)
DFTB+									
SCC-DFTB3-D3(BJ)	7.1	12.8	34.6 (ox β)	14.0	12.8	0.6689	-13.4	13.4	30.1 (ura)

^a SCC-DFTB3-D3 geometries were used as the starting point for the PBE-D2/DZP optimization.

^b Anthracene and naphthalene, which have small band gaps, have been excluded from the statistics for PBEh-3c because the SCF cycle failed to converge. This is in agreement with previous reports.⁴⁸

^c Labels in parentheses identify the X23 system that yields the maximum absolute error (ada: adamantane, amm: ammonia, ant: anthracene, cyt: cytosine, ox α : α -oxalic acid, ox β : β -oxalic acid, suc: succinic acid, ura: uracil, ure: urea).

single-point energy calculations at the geometry resulting from any of the low-cost methods tested are, in general, smaller than using the energies from those same methods, except in the case of SCC-DFTB3-D3(BJ). The MAEs correlate very roughly with the methods’ ability to produce crystal structures similar to the B86bPBE-XDM equilibrium geometries, as shown in the POW column of Table 1. Importantly, the lattice energies improve substantially for all methods where the primary source of error is BSSE, indicating that BSSE affects very strongly the calculation of absolute lattice energies but only has a minor effect on crystal geometries. This seems to be particularly true for the finite-support numerical orbital calculations in SIESTA, which grossly overestimate the absolute lattice energies but yield surprisingly good crystal geometries. Table 1 shows that the B86bPBE-XDM//PBE-D2/DZP multilevel method would be particularly effective in reproducing pure B86bPBE-XDM results, both giving the same MAE. The powder diffraction similarity measure indicates that the crystal geometries of these two methods are essentially coincident (POW = 0.0102) and the maximum absolute deviation between both methods in the whole X23 set is only 1.5 kJ/mol.

Slightly lower MAEs can be obtained using vdW-DF2 or PBEh-3c for the geometry optimization, but the POW similarity measure is not as good as PBE-D2/DZP. It is also interesting to note that HF-D2/MINI(x), which is essentially HF-3c with a different dispersion and minus the gCP and SRB corrections, achieves a smaller MAE than HF-3c when combined with B86bPBE-XDM in a multilevel method, indicating that the gCP and SRB corrections are not necessarily beneficial when HF-3c is supplemented with a final single-point calculation, as in the method proposed by Brandenburg and Grimme.⁶⁶

While the structures obtained with the PBE-D2/DZP method (SIESTA) are quite comparable to the B86bPBE-XDM geometries, this could be an artifact of the optimization procedure, since for all entries in Table 1, we used B86bPBE-XDM as the starting geometry. To test this, we re-ran the PBE-D2/DZP

and B86bPBE-XDM//PBE-D2-DZP calculations starting from the worst geometries available (SCC-DFTB3-D3(BJ)). Although the pure PBE-D2/DZP results are relatively independent of the starting geometry, the multilevel method is penalized, with a MAE of 6.3 kJ/mol, relative to experiment, when the poor starting geometries are used. The discrepancy between the two sets of B//A results is a consequence of the fairly loose default convergence thresholds for geometry optimization within SIESTA. The MAE using the DFTB starting geometries and tighter optimization convergence criteria (0.01 eV/Å force and 0.02 GPa stress convergence thresholds) decreases to 4.8 kJ/mol. Thus the choice of convergence thresholds has a non-negligible effect on the composite lattice energies, particularly when the starting geometries are quite dissimilar to those obtained with B86bPBE-XDM, but the looser thresholds are retained in practice to lower the computational cost.

The extraordinary agreement between B86bPBE-XDM and PBE-D2/DZP geometries prompts the question of whether it is possible to further reduce the cost of these SIESTA calculations by adjusting one or more input parameters, while maintaining a similar quality in the predicted geometries. In a practical CSP context, it is undesirable to explore the convergence of these parameters for each individual candidate structure, so there is value in studying their effect beforehand. Specifically, we explored the effect of reducing the energy-shift parameter (i.e. the “size” of the atomic orbitals), and basis set size. Table 2 shows the results of these calculations. There is a small impact from using a DZP, rather than a TZP, basis set; the computational savings from using the former outweigh the 0.7 kJ/mol improvement in the agreement with the reference B86bPBE-XDM energies. Likewise, a $2 \times 2 \times 2$ k-point mesh seems to be enough for the crystals in the X23 set and, since these are sufficiently small compared to the typical molecular crystal sizes in CSP, it is reasonable to assume that this k-point mesh can be used routinely. In contrast, the value of the “energy-shift” parameter, which is linked to the spatial extent of

Table 2: The MAE for the X23 set using PBE-D2/DZP (A) and B86bPBE-XDM//PBE-D2/DZP (B//A) compared to the B86bPBE-XDM lattice energies, in kJ/mol per molecule. The POW similarity measure referred to the B86bPBE-XDM equilibrium structures is also shown.

Test	A	B//A	POW
Reference ^a	10.8	0.8	0.0102
Tighter opt. convergence ^b	10.8	0.7	0.0382
Reference, DFTB geom. ^a	10.6	3.0	0.1130
Tighter opt. convergence, DFTB geom. ^b	10.8	1.8	0.0811
k-point mesh 2×2×2 ^c	10.5	0.8	0.0358
Energy shift (0.02 Ry) ^d	124.2	6.9	0.4876
TZP basis set ^e	4.7	0.1	0.0253

^a The PBE-D2/DZP method used in Table 1 and the rest of the article, with maximum force and stress components convergence criteria of 0.04 eV/Å and 1.0 GPa, a 0.001 Ry energy shift parameter, and a 4×4×4 k-point mesh starting from B86bPBE-XDM geometries, or where indicated, from SCC SCC-DFTB3-D3 geometries.

^b Tighter force and stress convergence thresholds (0.01 eV/Å and 0.02 GPa for maximum component values),

^c a 2×2×2 k-point mesh,

^d the default SIESTA energy shift parameter (0.02 Ry), and

^e a TZP basis set were used.

the numerical atomic orbitals in SIESTA calculations,^{52,77} is crucial in obtaining good-quality geometries and accurate single-point energies. A 0.001 Ry cutoff was found to be sufficient, but the default 0.02 Ry value used in SIESTA results in unacceptable errors.

4.2 Relative Lattice Energies

We focus now on whether low-cost and multi-level methods can describe lattice energy differences using our enantiomeric excess (ee) data for chiral compounds. For the subset of these compounds in which the racemate is more stable than the enantiopure crystal, the relative free energy between the two phases can be directly calculated from the experimental ee of a solution in contact with the enantiopure and racemic crystals (the eutectic point). The ee is calculated using:^{59,89}

$$ee = \frac{[L] - [D]}{[L] + [D]} \times 100 = \frac{(1/4) - \beta^2}{\beta^2 + (1/4)} \times 100, \quad (1)$$

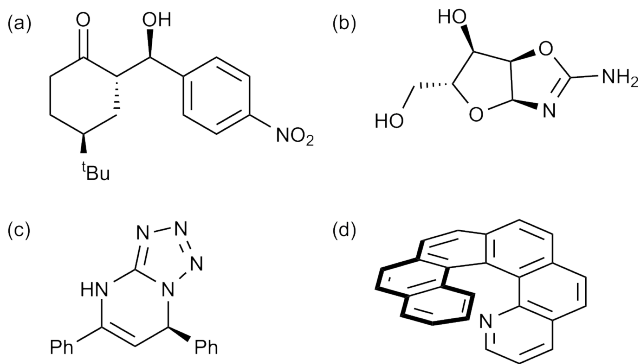
where [L] and [D] are concentrations in solution and:

$$\beta = \exp\left(-\frac{\Delta G_{L-DL}}{RT}\right) \approx \exp\left(-\frac{\Delta E_{L-DL}}{RT}\right) \quad (2)$$

with ΔG_{L-DL} the Gibbs free energy difference between the enantiopure (L) and the racemate (DL) crystals at the experimental temperature and ΔE_{L-DL} the corresponding lattice energy difference. In the last term, vibrational free energy contributions are neglected. The (1/4) factors in equation 1 are different from the equation in Ref. 59 in that an $RT \ln 2$ contribution proposed by Price et al.⁸⁹ to account for the entropy of mixing in solution was added to ΔE . This formula is valid only in cases where the racemate (DL) is more stable than the enantiopure (L) phase by at least the entropy of mixing term ($\Delta E_{L-DL} > RT \ln 2$). Otherwise, a conglomerate of L and D crystals is formed instead, and the ee in solution is zero.

We supplement the ee information for the 10 pairs of amino acid crystals in Ref. 59 with four additional compounds for which experimentally measured ee's are available,^{6,90-92} shown in Figure 1. Given the exponential dependence of

Figure 1: Chemical diagrams for (a) aldol⁹⁰ (4-(*tert*-butyl)-2-[hydroxy(4-nitrophenyl)methyl]-(2*S*,4*S*)-*rel*-cyclohexanone), (b) oxazoline⁹¹ (ribo-amino-oxazoline), (c) tetrazole⁹² (5,(7*R*)-diphenyl-4,7-dihydro-tetrazolo[1,5-*a*]pyrimidine), and (d) N-helicene⁶ (*P*-1-aza[6]helicene) added to the set of ten amino acids to form the EE14 set.



the ee on the computed electronic-energy differences and the precision with which it can be experimentally determined, these measures provide an excellent benchmark for the relative energy differences between crystal forms.

Figure 2 shows the performance of different low-cost and multilevel methods in the calculation of relative lattice energies, while Table 3 gives the relevant statistics relative to computed B86bPBE-XDM data. The detailed list of energy differences between both crystal phases and the predicted and experimentally-measured enantiomeric excesses, as well as differences in simulated powder diffraction spectra relative to the B86bPBE-XDM equilibrium geometries for all crystals can be found in the Supporting Information. Overall, the agreement with the theoretical model (black curve) is improved when final B86bPBE-XDM single-point calculations are performed, which is consistent with our discussion in Section 4.1. In this case, however, the performance of the methods based on a DZP basis set of numerical orbitals (PBE-D2/DZP and vdW-DF1,2/DZP) is surprisingly good, comparable to that of B86bPBE-XDM. This is in striking contrast with the poor results for absolute lattice energies shown in Table 1. This is an indication that, while BSSE originating from the finite nature of the orbitals

in SIESTA methods affects absolute lattice energies very strongly, its impact on the relative lattice energies, which are the quantity of interest in CSP, is comparatively small. The use of tighter convergence criteria again provided similar results to those discussed for the X23 set in the previous section.

Conversely, HF-3c and PBEh-3c give worse results than the DZP methods, particularly if a final B86bPBE-XDM calculation is used. Pure HF-3c has a MAE similar to the DZP methods and, surprisingly, so does SCC-DFTB3-D3(BJ). However, the relatively poor HF-3c and PBEh-3c geometries make the corresponding multilevel method very inaccurate, which may explain the relatively poor performance of the TPSS-D3//HF-3c multilevel method in the sixth blind test.⁶⁶ It is important to note that PBEh-3c and the DZP methods are one order of magnitude slower than HF-3c. Furthermore, as in the case of the X23 crystals, PBEh-3c could not be applied in some cases due to SCF convergence failures (N-helicene and tetrazole) for similar reasons as in anthracene and naphthalene (band gap closing). The performance of the multilevel method based on HF-D2/MINI(x) is only slightly worse than HF-3c, which is another indication that the two corrections other than dispersion in 3c are not very effective at improving the accuracy of the corresponding multilevel method. In contrast, the MAE of SCC-DFTB3-D3(BJ) is on par with other methods, which is notable considering the enormous average error in the calculation of absolute lattice energies (Table 1). Contrarily, single-point calculations at the HF-3c, HF-D2/MINI(x), and SCC-DFTB3-D3(BJ) result in higher average errors than using the low-cost methods alone. This observation highlights the importance of assessing low-cost methods regarding their ability to produce accurate equilibrium geometries, and not just lattice energies.

4.3 Crystal-Structure Prediction

Finally, we consider a practical CSP application using the best-performing multilevel approach in the previous sections, B86bPBE-

Figure 2: Experimental enantiomeric excess as a function of calculated energy differences between the enantiopure and racemate crystal structures, ΔE_{L-DL} . The solid black line represents the ee as a function of calculated ΔE_{L-DL} (Eq. 1). Accurate relative energies yield good agreement between the experimental ee and the black curve (vibrational contributions are neglected). Data points for energies resulting from the geometry relaxation using a low-cost method, and those obtained from additional B86bPBE-XDM single-point energy calculations are represented by open circles and closed diamonds, respectively.

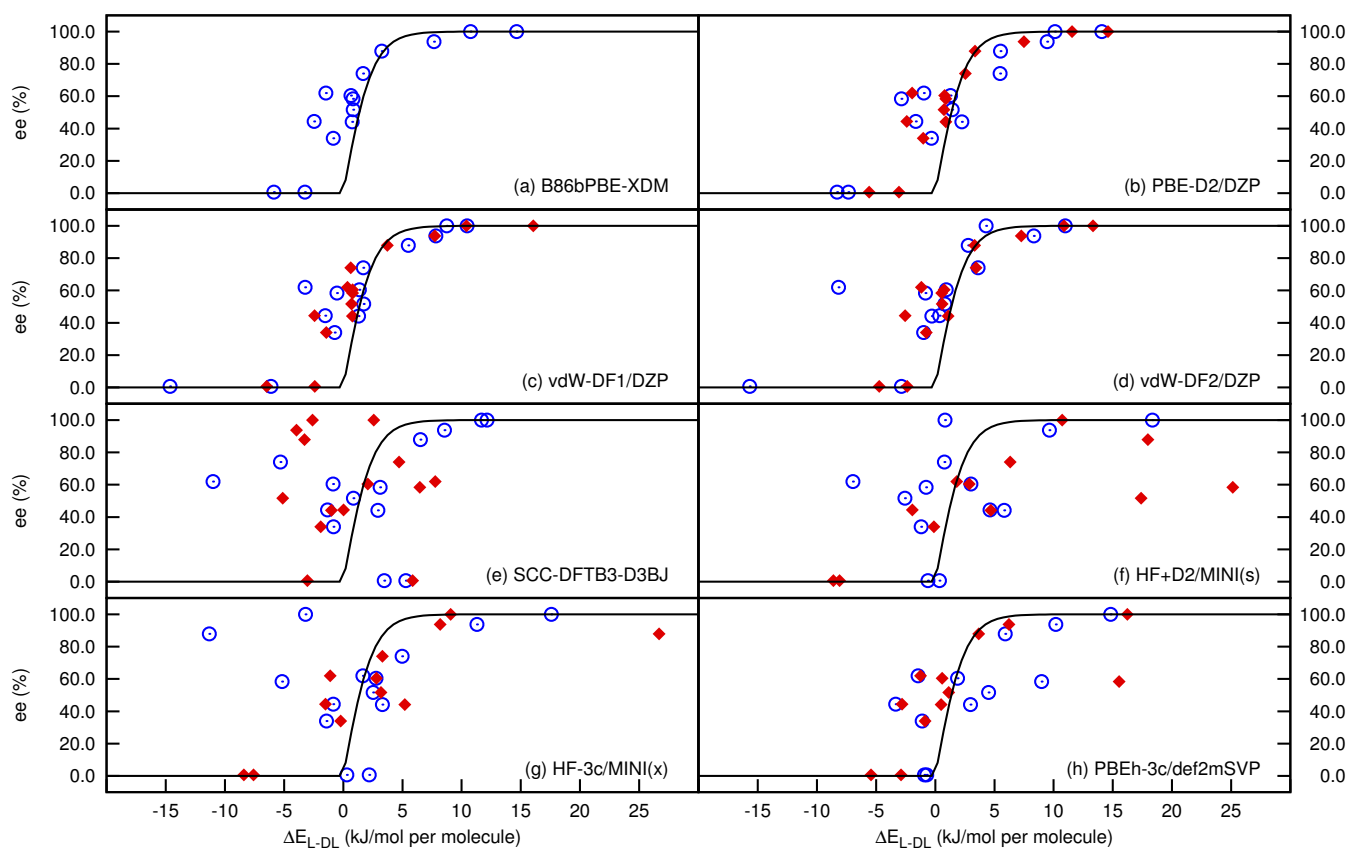


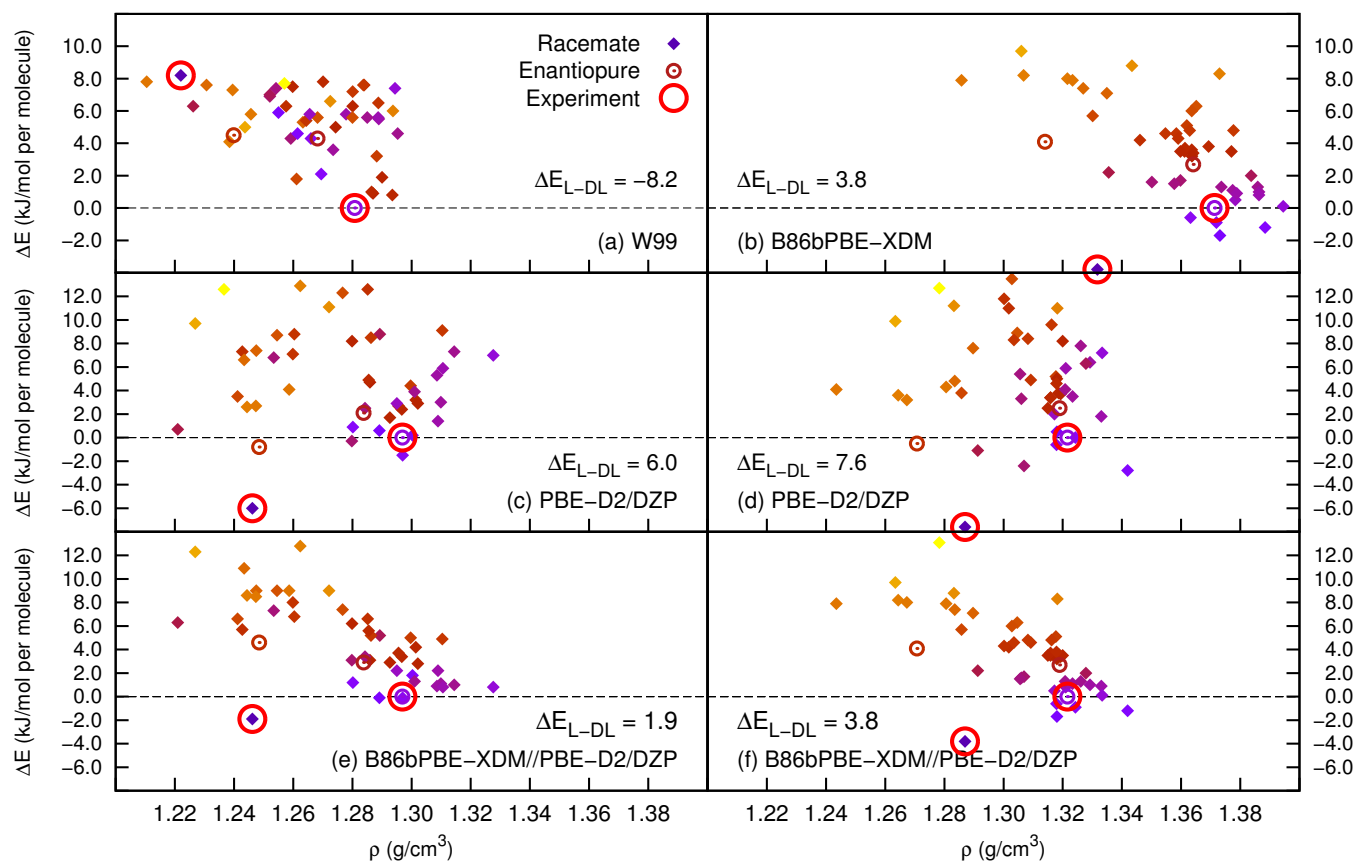
Table 3: Statistics for the EE14 set of relative lattice energies using various low-cost methods. Mean error (ME), mean absolute error (MAE), and maximum absolute error (MAX) relative to experiment (EXP) or fully-relaxed B86bPBE-XDM results (DFT). “A” and “B//A” refer, respectively, to the results obtained from the low-cost optimizations directly or to a B86bPBE-XDM single-point energy calculation at the low-cost equilibrium geometry. All values are in kJ/mol per molecule, except POWDIFF similarity measures (POW), which are dimensionless.

Method	EXP				DFT				
	ME	A MAE	MAX ^b	B//A MAE	A MAE	POW	B//A ME	MAE	MAX ^b
QE									
B86bPBE-XDM	-1.5	2.1	5.9 (glu)	2.1	0.0	0.0000	0.0	0.0	0.0
SIESTA									
PBE-D2/DZP	-1.3	3.2	8.3 (glu)	2.0	1.7	0.0169	0.1	0.3	0.9 (Nhe)
vdW-DF2/DZP	-2.7	3.5	15.7 (asp)	1.9	3.0	0.1112	0.2	0.5	1.8 (Nhe)
vdW-DF1/DZP	-2.3	3.3	14.7 (asp)	2.1	1.9	0.1930	0.1	0.5	1.8 (oxa)
CRYSTAL									
PBEh-3c/def2mSVP ^a	4.2	4.5	11.9 (his)	4.1	2.4	0.1411	1.3	1.7	14.7 (cys)
HF-3c	-0.8	3.5	14.7 (leu)	8.3	5.4	0.2011	6.7	7.9	42.0 (cys)
HF-D2/MINI(x)	2.6	6.2	28.8 (leu)	9.4	6.2	0.2872	8.0	9.1	25.8 (his)
DFTB+									
SCC-DFTB3-D3(BJ)	-0.7	3.9	12.8 (oxa)	4.2	3.6	0.3611	-1.3	6.2	17.3 (ser)

^a N-helicene and tetrazole encountered similar SCF convergence problems as anthracene and naphthalene, and have thus been excluded from the statistics for PBEh-3c.⁴⁸

^b Labels in parentheses identify the system that yields the maximum absolute error (cys: cysteine, his: histidine, Nhe: N-helicene, ser: serine).

Figure 3: Lowest-energy enantiopure and racemate crystal structures of 1-aza[6]helicene obtained from a previous CSP study on chiral helicenes⁶ with the W99 force field (a) and reranked using B86bPBE-XDM (b). Panels (c) and (d) depict the energy profiles obtained when performing full relaxations with PBE-D2/DZP starting from either W99 or QE geometries, respectively. Panels (e) and (f) show results from further re-ranking the PBE-D2/DZP structures using B86bPBE-XDM single-point energy calculations. The color scheme follows the relative energetic ordering obtained with full B86bPBE-XDM relaxations (b) and is kept constant for all other panels in order to compare how the energies shift when using the force fields and low-cost methods. Experimentally observed structures are encircled. The experimentally isolated enantiopure crystal is set as the zero of energy.



XDM//PBE-D2/DZP. In a recent article, we have shown how the electronic properties of an organic semiconductor based on the chiral 1-aza[6]helicene molecule⁶ (see Figure 1(d)) can be dramatically affected by the racemic or enantiopure nature of the material through differences in the racemate and enantiopure crystal structures. In the CSP part of the study, an initial set of candidate structures were ranked using the W99 repulsion-dispersion force field,⁹³ with electrostatics described by a distributed multipole analysis.⁹⁴ Fifty candidate crystal structures (47 racemate and 3 enantiopure crystals) were reoptimized and reranked with B86bPBE-XDM. The set of candidates includes the experimentally-observed racemic and enantiopure crystal structures. The initial ranking by the W99 is shown in Figure 3(a), and the reranking with B86bPBE-XDM is shown in panel (b). Detailed data can be found in the Supporting Information.

The initial ranking from the W99 force field proved to be inadequate as the experimentally isolated form of the racemate crystal was ranked highest in energy and the wrong relative ordering of the enantiopure and racemate crystal phases was predicted. According to the experimental measurements, the ee in solution for this crystal is 74%, which corresponds to a racemate more stable than the enantiopure crystal by an energy difference of 4.1 kJ/mol at room temperature.⁶ Figure 3 shows that B86bPBE-XDM not only recovers the correct relative ordering of both phases, but also gives the experimentally-observed racemate as the phase with the lowest electronic energy among all the candidates. The energy difference between racemate and enantiopure crystals is 3.8 kJ/mol, in excellent agreement with the EE14 value.

Figure 3(c) and (d) show how the energy profile of this crystal energy landscape is affected by the reoptimization using PBE-D2/DZP. The equilibrium crystal structures are different depending on whether the W99 or B86bPBE-XDM geometries are used as starting points of the optimization. As in previous sections, this disagreement is caused by the loose geometry optimization convergence criteria (0.04 eV/Å

for the force and 1.0 GPa for the stress components). Subsequent tests on the minimum-energy racemate structure (circled full point in Figure 3) showed that the true PBE-D2/DZP minimum-energy cell dimensions fall between the two sets of results. Although this disagreement could be resolved using tighter convergence thresholds for forces and stresses, this would increase the computational cost by approximately a factor of three.

As expected, the relative energies from using the low-cost PBE-D2/DZP method do not reproduce the B86bPBE-XDM results in Figure 3(b), but simply performing an additional single-point energy calculation on each crystal structure, as shown in Figure 3(e) and (f), recovers crystal energy landscapes in good agreement with the reference high-level method (the relative energies for all crystals can be found in the SI). These plots show that, although the optimization convergence thresholds are having an effect on the calculated energies, the error introduced is tolerable, and the B86bPBE-XDM//PBE-D2/DZP method predicts the correct energy ordering even if the poor W99 starting geometries are used. For instance, the MAE between the B86bPBE-XDM//PBE-D2/DZP and the B86bPBE-XDM results for all 50 relative energies is 0.3 kJ/mol when the B86bPBE-XDM geometries are used as the starting point of the PBE-D2/DZP optimizations and 1.6 kJ/mol when the W99 geometries are used instead. Nonetheless, this accuracy is sufficient to correctly identify the most-promising candidate structures for subsequent high-level structure refinement. Thus, given that both energy landscapes are similar and the proper relative ordering between enantiopure and racemate phases is recovered, we consider our multilevel approach to be suitable for crystal structure prediction purposes.

5 Conclusions

In this article, we studied the applicability and performance of computationally inexpensive (cheap) and multilevel approaches to crystal structure prediction (CSP). Multilevel methods

are a composite of two techniques in which the cheap method is used for geometry optimization and a more expensive and accurate method is used in a final single-point calculation. Multilevel methods are very popular in molecular quantum chemistry, as they furnish an accuracy similar to the expensive method with a much reduced computational cost. In particular, we have studied several cheap methods: PBE-D2, vdW-DF1, and vdW-DF2 with a double-zeta basis set of numerical orbitals (SIESTA); minimal or small Gaussian basis set calculations (PBEh-3c, HF-3c, HF-D2/MINI(x)), and self-consistent-charge dispersion-corrected density-functional tight binding (SCC-DFTB3-D3). These cheap methods have been evaluated both alone and in a multilevel method where the final single-point calculation is run with B86bPBE-XDM, which has the best performance of all current density functional methods in the calculation of absolute lattice energies (measured on the X23 set).

The performance of various cheap and multilevel methods was evaluated using three tests: absolute lattice energies (the X23 set), relative lattice energies (the EE14 set), and a practical CSP application (enantiopure and racemic forms of 1-aza[6]helicene, an organic semiconductor). The EE14 set is a new benchmark set for relative lattice energies derived from experimental measurements of enantiomeric excess of a solution in contact with the racemic and enantiopure crystals of the same chiral compound, comprising 14 relative lattice energies.

Our results show that absolute lattice energies are much more difficult to model than relative lattice energies, the latter being the important quantity in CSP. This is especially true for cheap methods. For instance, SCC-DFTB3-D3(BJ) gives a MAE of 12.8 kJ/mol on the X23, but only of 3.6 kJ/mol on the EE14. In comparison, PBEh-3c, whose corrections have been developed to improve binding and lattice energies accuracy relative to minimal-basis-set HF, has a relatively low MAE (5.7 kJ/mol) on the X23, but the MAE on the EE14 is higher than SCC-DFTB3-D3(BJ) (4.5 kJ/mol). This emphasizes the importance of targeting relative lattice energies in the development of cheap methods for

CSP.

Regarding the performance of multilevel methods (cheap methods with a final B86bPBE-XDM step), in general the performance on the X23 and EE14 of all multilevel methods improves relative to the corresponding cheap methods alone, except for those cheap methods that give equilibrium crystal geometries very different from B86bPBE-XDM. In particular, our results show that PBE-D2 with a double-zeta basis of numerical orbitals (PBE-D2/DZP) is particularly efficient at recovering B86bPBE-XDM geometries, despite its poor performance for absolute lattice energies. This results in an optimal multilevel method, with MAEs for the X23 and EE14 of 3.6 and 2.0 kJ/mol. In comparison, the MAEs of pure B86bPBE-XDM on the X23 and EE14 are 3.6 and 2.1 kJ/mol, respectively.

It should be noted, however, that the performance and computational cost of B86bPBE-XDM//PBE-D2/DZP depends critically on the choice of calculation parameters for the SIESTA method. In particular, the energy shift parameter, which controls the spatial extent of the orbitals, needs to be considerably lower than the default (0.001 Ry, as suggested previously in the literature^{52,77} is a good option). Likewise, choosing convergence criteria for the geometry optimization that strike a balance between performance and cost is essential. We have determined that the default convergence thresholds in SIESTA (0.04 eV/Å and 1.0 GPa maximum values for the components of force and stress, respectively) give an adequate compromise between the two, although they introduce a small error if the initial geometries are poor.

Because of its good performance, we tested B86bPBE-XDM//PBE-D2/DZP on the set of helicene candidate structures from our recent CSP study.⁶ It was found that, starting from the B86bPBE-XDM equilibrium geometries, the relative energies from B86bPBE-XDM//PBE-D2/DZP are essentially indistinguishable from pure B86bPBE-XDM. When we use relatively poor starting geometries (the equilibrium geometries using the W99 force field), then the performance is slightly worse, but the multilevel method is still able to recover

a reliable energy landscape. Thus, we conclude that multilevel approaches and, in particular, B86bPBE-XDM//PBE-D2/DZP are excellent candidates for energy ranking functions in molecular CSP.

Acknowledgement The authors thank Drs. J. Gerit Brandenburg and Stephen G. Dale, as well as Prof. Sally L. Price (University College London), for helpful discussions. We are grateful for financial support from the Natural Sciences and Engineering Research Council (NSERC), and for computational resources and support from ACEnet, Westgrid and Compute Canada/Calcul Canada. LML would also like to acknowledge the Walter C. Sumner Foundation for financial support.

Supporting Information Available

CPU timings for low-cost methods with selected crystal structures. Computed and benchmark lattice energies for the X23 set crystal structures, predicted and experimental enantiomeric excesses for the EE14 set crystal structures, and relative energy differences and densities of crystal structures from the CSP study of 1-aza[6]helicene.

References

- (1) Cruz-Cabeza, A. J. Facts and Fictions About Polymorphism. *Chem. Soc. Rev.* **2015**, *44*, 8619–8635.
- (2) Bučar, D.; Lancaster, R. W.; Bernstein, J. Disappearing Polymorphs Revisited. *Angew. Chem. Intl. Ed.* **2015**, *54*, 6972–6993.
- (3) Thakur, T. S.; Dubey, R.; Desiraju, G. R. Crystal Structure and Prediction. *Annu. Rev. Phys. Chem.* **2015**, *66*, 21–42.
- (4) Zykova-Timan, T.; Raiteri, P.; Parrinello, M. Investigating the Polymorphism in PR179: A Combined Crystal Structure Prediction and Metadynamics Study. *J. Phys. Chem. B* **2008**, *112*, 13231–13237.
- (5) C. Wei, H. H.; Duan, X.; Pei, C. Structures and Properties Prediction of HMX/TATB Co-Crystal. *Propellants Explos. Pyrotech.* **2011**, *36*, 416–423.
- (6) Yang, Y.; Rice, B.; Shi, X.; Brandt, J. R.; Correa da Costa, R.; Hedley, G. J.; Smilgies, D.; Frost, J. M.; Samuel, I.; Otero-de-la-Roza, A.; Johnson, E. R.; Jelfs, K. E.; Nelson, J.; Campbell, A. J.; Fuchter, M. J. Emergent Properties of an Organic Semiconductor Driven by its Molecular Chirality. *ACS Nano* **2017**, *11*, 8329–8338.
- (7) Ravva, M. K.; Risko, C.; Brédas, J. L. In *Non-covalent Interactions in Quantum Chemistry and Physics*; Otero-de-la-Roza, A., DiLabio, G., Eds.; Elsevier: Amsterdam, Netherlands, 2017; Chapter 9, pp 277–298.
- (8) Lommerse, J.; Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Gavezzotti, A.; Hofmann, D.; Leusen, F.; Mooij, W.; Price, S. L.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. A Test of Crystal Structure Prediction of Small Organic Molecules. *Acta Crystallogr., Sect. B: Struct. Sci.* **2000**, *56*, 697–714.
- (9) Motherwell, W.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D.; Leusen, F.; Lommerse, J.; Mooij, W.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. Crystal Structure Prediction of Small Organic Molecules: A Second Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 647–661.
- (10) Day, G. M.; Motherwell, W.; Ammon, H. L.; Boerrigter, S.; Della Valle, R. G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J. D.; Schweizer, B.; van Eijck, B. P.; Erk, P.; Facelli, J. C.;

- Bazterra, V. E.; Ferraro, M. B.; Hofmann, D.; Leusen, F. J. J.; Liang, C.; Pantelides, C. C.; Karamertzanis, P. G.; Price, S. L.; Lewis, T. C.; Nowell, H.; Torrisi, A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Verwer, P. A. Third Blind Test of Crystal Structure Prediction. *Acta Crystallogr., Sect. B: Struct. Sci.* **2005**, *61*, 511–527.
- (11) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, R. S.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D.; Neumann, M. A.; Leusen, F.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A. K.; Schweizer, B. Significant Progress in Predicting the Crystal Structures of Small Organic Molecules – A Report on the Fourth Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci.* **2009**, *65*, 107–125.
- (12) Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F.; ad A. J. Misquitta, A. V. M.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K. Towards Crystal Structure Prediction of Complex Organic Compounds – A Report On the Fifth Blind Test. *Acta Crystallogr., Sect. B: Struct. Sci.* **2011**, *67*, 535–551.
- (13) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fustimolnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Janckiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J. Z.; Zhu, Q.; Groom, C. R. Report on the Sixth Blind Test of Organic Crystal-Structure Prediction Methods. *Acta Crystallogr., Sect. B: Struct. Sci.* **2016**, *72*, 439–459.
- (14) Price, S. L. Predicting Crystal Structures of Organic Compounds. *Chem. Soc. Rev.* **2014**, *43*, 2098–2111.
- (15) Cruz-Cabeza, A. J. Crystal Structure Prediction: Are We There Yet? *Acta Crystallogr., Sect. B: Struct. Sci.* **2016**, *72*, 437–438.
- (16) Beran, G. Modelling Polymorphic Molec-

- ular Crystals with Electronic Structure Theory. *Chem. Rev.* **2016**, *116*, 5567–5613.
- (17) Price, S. L.; Braun, D. E.; Reutzeldens, S. M. Can Computed Crystal Energy Landscapes Help Understand Pharmaceutical Solids? *Chem. Commun.* **2016**, *52*, 7065–7077.
- (18) Hoja, J.; Reilly, A. M.; Tkatchenko, A. First-Principles Modeling of Molecular Crystals: Structures and Stabilities, Temperature and Pressure. *WIREs Comput. Mol. Sci.* **2017**, *7*, e1294.
- (19) Price, S. L. Why Don't We Find More Polymorphs? *Acta Crystallogr., Sect. B: Struct. Sci.* **2013**, *69*, 313–328.
- (20) Marqués, M.; Morales, A.; Menéndez, J. M. In *An Introduction to High-Pressure Science and Technology*; Recio, J. M., Menendez, J. M., de la Roza, A. O., Eds.; CRC Press, 2015; Chapter 4, pp 105–130.
- (21) Nyman, J.; Day, G. M. Static and Lattice Vibrational Energy Differences Between Polymorphs. *CrystEngComm* **2015**, *17*, 5154–5165.
- (22) Becke, A. D. On the Large-Gradient Behavior of the Density Functional Exchange Energy. *J. Chem. Phys.* **1986**, *85*, 7184–7187.
- (23) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- (24) Becke, A. D.; Johnson, E. R. Exchange-Hole Dipole Moment and the Dispersion Interaction Revisited. *J. Chem. Phys.* **2007**, *127*, 154108.
- (25) Otero-de-la-Roza, A.; Johnson, E. R. Van der Waals Interactions in Solids Using the Exchange-Hole Dipole Moment Model. *J. Chem. Phys.* **2012**, *136*, 174109.
- (26) Otero-de-la-Roza, A.; Johnson, E. R. A Benchmark for Non-Covalent Interactions in Solids. *J. Chem. Phys.* **2012**, *137*, 054103.
- (27) Johnson, E. R. In *Non-Covalent Interactions in Quantum Chemistry and Physics*; Otero-de-la-Roza, A., DiLabio, G., Eds.; Elsevier: Amsterdam, Netherlands, 2017; Chapter 5, pp 169–192.
- (28) Chickos, J. S. Enthalpies of Sublimation after a Century of Measurement: A View as Seen Through the Eyes of a Collector. *Netsu Sokutei* **2003**, *3*, 116–124.
- (29) Perdew, J. P. Some Fundamental Issues in Ground-State Density Functional Theory: A Guide for the Perplexed. *J. Chem. Theory Comput.* **2009**, *5*, 902–908.
- (30) Whittleton, S. R.; Otero-de-la-Roza, A.; Johnson, E. R. Exchange-Hole Dipole Dispersion Model for Accurate Energy Ranking in Molecular Crystal Structure Prediction. *J. Chem. Theory Comput.* **2017**, *13*, 441–450.
- (31) Whittleton, S. R.; Otero-de-la-Roza, A.; Johnson, E. R. Exchange-Hole Dipole Dispersion Model for Accurate Energy Ranking in Molecular Crystal Structure Prediction II: Non-Planar Molecules. *J. Chem. Theory Comput.* **2017**, *13*, 5332–5342.
- (32) Grimme, S. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105–5154.
- (33) Marom, N.; DiStasio, Jr., R. A.; Atalla, V.; Levchenko, S.; Reilly, A. M.; Chelikowsky, J. R.; Leiserowitz, L.; Tkatchenko, A. Many-Body Dispersion Interactions in Molecular Crystal Polymorphism. *Angew. Chem. Int. Ed.* **2013**, *52*, 6629–6632.
- (34) Kronik, L.; Tkatchenko, A. Understanding Molecular Crystals with Dispersion-Inclusive Density Functional Theory: Pairwise Corrections and Beyond. *Acc. Chem. Res.* **2014**, *47*, 3208–3216.

- (35) Moellmann, J.; Grimme, S. DFT-D3 Study of Some Molecular Crystals. *J. Phys. Chem. C* **2014**, *118*, 7615–7621.
- (36) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
- (37) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465.
- (38) Tkatchenko, A.; DiStasio, Jr., R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Phys. Rev. Lett.* **2012**, *108*, 236402.
- (39) Ambrosetti, A.; Reilly, A. M.; DiStasio, Jr., R. A.; Tkatchenko, A. Long-Range Correlation Energy Calculated From Coupled Atomic Response Functions. *J. Chem. Phys.* **2014**, *140*, 18A508.
- (40) Reilly, A. M.; Tkatchenko, A. Understanding the Role of Vibrations, Exact Exchange, and Many-Body van der Waals Interactions in the Cohesive Properties of Molecular Crystals. *J. Chem. Phys.* **2013**, *139*, 024705.
- (41) Brandenburg, J. G.; Grimme, S. Dispersion Corrected Hartree-Fock and Density Functional Theory for Organic Crystal Structure Prediction. *Top. Curr. Chem.* **2014**, *345*, 1–24.
- (42) Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions. *J. Phys. Chem. Lett.* **2014**, *5*, 4275–4284.
- (43) Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301–5337.
- (44) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. Self-Consistent-Charge Density-Functional Tight-Binding Method for Simulations of Complex Materials Properties. *Phys. Rev. B* **1998**, *58*, 7260–7268.
- (45) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *J. Chem. Theory Comput.* **2011**, *7*, 931–948.
- (46) Gaus, M.; Cui, Q.; Elstner, M. Density Functional Tight Binding: Application to Organic and Biological Molecules. *WIREs Comput. Mol. Sci.* **2014**, *4*, 49–61.
- (47) Brandenburg, J. G.; Grimme, S. Accurate Modeling of Organic Molecular Crystals by Dispersion-Corrected Density Functional Tight Binding (DFTB). *J. Phys. Chem. Lett.* **2014**, *5*, 1785–1789.
- (48) Grimme, S.; Brandenburg, J. G.; Banwarth, C.; Hansen, A. Consistent Structures and Interactions by Density Functional Theory with Small Atomic Orbital Basis Sets. *J. Chem. Phys.* **2015**, *143*, 054107.
- (49) Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems. *J. Comput. Chem.* **2013**, *34*, 1672–1685.
- (50) Sure, R.; Brandenburg, J. G.; Grimme, S. Small Atomic Orbital Basis Set First-Principles Quantum Chemical Methods for Large Molecular and Periodic Systems: A Critical Analysis of Error Sources. *ChemistryOpen* **2016**, *5*, 94–109.
- (51) Cutini, M.; Civalleri, B.; Corno, M.; Orlando, R.; Brandenburg, J. G.; Maschio, L.; Ugliengo, P. Assessment of Different Quantum Mechanical Methods for the Prediction of Structure and Cohesive

- Energy of Molecular Crystals. *J. Chem. Theory Comput.* **2016**, *12*, 3340–3352.
- (52) Carter, D. J.; Rohl, A. L. Benchmarking Calculated Lattice Parameters and Energies of Molecular Crystals Using van der Waals Density Functional. *J. Chem. Theory Comput.* **2014**, *10*, 3423–3437.
- (53) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.
- (54) Lee, K.; Murray, E. D.; Kong, L.; and D. C. Langreth, B. I. L. Higher-Accuracy van der Waals Density Functional. *Phys. Rev. B* **2010**, *82*, 081101.
- (55) Soler, J. M.; Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. The SIESTA Method for *Ab Initio* Order-*N* Materials Simulation. *J. Phys.: Condens. Matter* **2002**, *14*, 2745–2779.
- (56) Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Martin, R. M.; Ordejón, P.; Pruneda, J. M.; Sánchez-Portal, D.; Soler, J. M. The SIESTA Method; Developments and Applicability. *J. Phys.: Condens. Matter* **2008**, *20*, 064208.
- (57) Giannozzi, P.; Andreussi, O.; Brumme, T.; Bunau, O.; Buongiorno Nardelli, M.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Cococcioni, M.; Colonna, N.; Carnimeo, I.; Dal Corso, A.; de Gironcoli, S.; Delugas, P.; DiStasio, R.; Ferretti, A.; Floris, A.; Fratesi, G.; Fugallo, G.; Gebauer, R.; Gerstmann, U.; Giustino, F.; Gorni, T.; Jia, J.; Kawamura, M.; Ko, H.-Y.; Kokalj, A.; Küçükbenli, E.; Lazzeri, M.; Marsili, M.; Marzari, N.; Mauri, F.; Nguyen, N. L.; Nguyen, H.-V.; Otero-de-la-Roza, A.; Paulatto, L.; Poncé, S.; Rocca, D.; Sabatini, R.; Santra, B.; Schlipf, M.; Seitsonen, A.; Smogunov, A.; Timrov, I.; Thonhauser, T.; Umari, P.; Vast, N.; Baroni, S. Advanced Capabilities for Materials Modelling with Quantum ESPRESSO. *J. Phys.: Condens. Matter* **2017**, *29*, 465901.
- (58) Louwerson, M. J.; Rothenberg, G. Transferable Basis Sets of Numerical Atomic Orbitals. *Phys. Rev. B* **2012**, *85*, 035108.
- (59) Otero-de-la-Roza, A.; Cao, B. H.; Price, I. K.; Hein, J. E.; Johnson, E. R. Density-Functional Theory Predicts the Relative Solubilities of Racemic and Enantiopure Crystals. *Angew. Chem. Int. Ed.* **2014**, *53*, 7879–7882.
- (60) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Non-covalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements ($Z = 1-86$). *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009.
- (61) Pople, J. A.; Head-Gordon, M.; Fox, D. J.; Raghavachari, K.; Curtiss, L. A. Gaussian-1 Theory: A General Procedure for Prediction of Molecular Energies. *J. Chem. Phys.* **1989**, *90*, 5622–5629.
- (62) Curtiss, L. A.; Jones, C.; Trucks, G. W.; Raghavachari, K.; Pople, J. A. Gaussian-1 Theory of Molecular Energies for Second-Row Compounds. *J. Chem. Phys.* **1990**, *93*, 2537–2545.
- (63) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 Theory for Molecular Energies of First- and Second-Row Compounds. *J. Chem. Phys.* **1991**, *94*, 7221–7230.
- (64) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) Theory for Molecules Containing First and Second-Row Atoms. *J. Chem. Phys.* **1998**, *109*, 7764–7776.
- (65) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Gaussian-4 Theory. *J. Chem. Phys.* **2007**, *126*, 084108.

- (66) Brandenburg, J. G.; Grimme, S. Organic Crystal Polymorphism: A Benchmark for Dispersion Corrected Mean Field Electronic Structure Methods. *Acta Crystallogr. Sect. B: Struct. Sci.* **2016**, *72*, 502–513.
- (67) Hylton, R. K.; Tizzard, G. J.; Threlfall, T. L.; Ellis, A. L.; Coles, S. J.; Seaton, C. C.; Schulze, E.; Lorenz, H.; Seidel-Morgenstern, A.; Stein, M.; Price, S. L. Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids Determined by Kinetics or Thermodynamics? *J. Am. Chem. Soc.* **2015**, *137*, 11095–11104.
- (68) Artacho, E.; Sánchez-Portal, D.; Ordejón, P.; García, A.; Soler, J. M. Linear-Scaling ab-initio Calculations for Large and Complex Systems. *Phys. Stat. Sol.* **1999**, *215*, 809–817.
- (69) Junquera, J.; Paz, O.; Sánchez-Portal, D.; Artacho, E. Numerical Atomic Orbitals for Linear-Scaling Calculations. *Phys. Rev. B* **2001**, *64*, 235111.
- (70) Sánchez-Portal, D.; Ordejón, P.; Artacho, E.; Soler, J. M. Density-Functional Method for Very Large Systems with LCAO Basis Sets. *Int. J. Quantum Chem.* **1997**, *65*, 453–461.
- (71) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- (72) Peverati, R.; Baldrige, K. K. Implementation and Performance of DFT-D with Respect to Basis Set and Functional for Study of Dispersion Interactions in Nanoscale Aromatic Hydrocarbons. *J. Chem. Theory Comput.* **2008**, *4*, 2030–2048.
- (73) Hamann, D. R.; Schlüter, M.; Chiang, C. Norm-Conserving Pseudopotentials. *Phys. Rev. Lett.* **1979**, *43*, 1494–1497.
- (74) Bachelet, G. B.; Hamann, D. R.; Schlüter, M. Pseudopotentials That Work: From H to Pu. *Phys. Rev. B* **1982**, *26*, 4199–4228.
- (75) Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. *Phys. Rev. B* **1991**, *43*, 1993–2006.
- (76) 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>.
- (77) Chapman, C.; Ting, E. C.; Kereszti, A.; Paci, I. Self-Assembly of Cysteine Dimers at the Gold Surface: A Computational Study of Competing Interactions. *J. Phys. Chem. C* **2013**, *117*, 19426–19435.
- (78) Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalieri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D’Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B. CRYSTAL14: A Program for the Ab Initio Investigation of Crystalline Solids. *Int. J. Quantum Chem.* **2014**, *114*, 1287–1317.
- (79) Dovesi, R.; Erba, A.; Orlando, R.; Zicovich-Wilson, C. M.; Civalieri, B.; Maschio, L.; Rérat, M.; Casassa, S.; Baima, J.; Salustro, S.; Kirtman, B. CRYSTAL17. **2017**,
- (80) Kruse, H.; Grimme, S. A Geometrical Correction for the Inter- and Intra-Molecular Basis Set Superposition Error in Hartree-Fock and Density Functional Theory Calculations for Large Systems. *J. Chem. Phys.* **2012**, *136*, 154101.
- (81) Brandenburg, J. G.; Alessio, M.; Civalieri, B.; Peintinger, M. F.; Bredow, T.; Grimme, S. Geometrical Correction for the Inter- and Intramolecular Basis Set Superposition Error in Periodic Density Functional Theory Calculations. *J. Phys. Chem. A* **2013**, *117*, 9282–9292.

- (82) Johnson, E. R.; Becke, A. D. A Post-Hartree-Fock Model of Intermolecular Interactions: Inclusion of Higher-Order Corrections. *J. Chem. Phys.* **2006**, *124*, 174104.
- (83) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *J. Phys. Chem. A* **2007**, *111*, 5678–5684.
- (84) Gaus, M.; Goez, A.; Elstner, M. Parametrization and Benchmark of DFTB3 for Organic Molecules. *J. Chem. Theory Comput.* **2013**, *9*, 338–354.
- (85) Kubillus, M.; Kubař, T.; Gaus, M.; Řezáč, J.; Elstner, M. Parametrization of the DFTB3 Method for Br, Ca, Cl, F, I, K, and Na in Organic and Biological Systems. *J. Chem. Theory Comput.* **2015**, *11*, 332–342.
- (86) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953–17979.
- (87) Otero-de-la-Roza, A.; Johnson, E. R.; Luaña, V. CRITIC2: A Program for Real-Space Analysis of Quantum Chemical Interactions in Solids. *Comput. Phys. Commun.* **2014**, *185*, 1007–1018.
- (88) de Gelder, R.; Wehrens, R.; Hageman, J. A. A Generalized Expression for the Similarity of Spectra: Application to Powder Diffraction Pattern Classification. *J. Comput. Chem.* **2001**, *22*, 273–289.
- (89) Buchholz, H. K.; Hylton, R. K.; Brandenburg, J. G.; Seidl-Morgenstern, A.; Lorenz, H.; Stein, M.; Price, S. L. Thermochemistry of Racemic and Enantiopure Organic Crystals for Predicting Enantiomer Separation. *Cryst. Growth Des.* **2017**, *17*, 4676–4686.
- (90) Rougeot, C.; Situ, H.; Cao, B. H.; Vlachos, V.; Hein, J. E. Automated Reaction Progress Monitoring of Heterogeneous Reactions: Crystallization-Induced Stereoselectivity in Amine-Catalyzed Aldol Reactions. *React. Chem. Eng.* **2017**, *2*, 226–231.
- (91) Hein, J. E.; Tse, E.; Blackmond, D. G. A Route to Enantiopure RNA Precursors From Nearly Racemic Starting Materials. *Nat. Chem.* **2011**, *3*, 704–705.
- (92) Price, I. K. Investigation and Application of Crystal Phase Behavior to Obtain Enantiopure Chemicals. Ph.D. thesis, University of California, Merced. Available from <http://eprints.cdlib.org/uc/item/24r9v8mc>., 2014.
- (93) Williams, D. E. Improved Intermolecular Force Field for Molecules Containing H, C, N, and O Atoms, with Application to Nucleoside and Peptide Crystals. *J. Comput. Chem.* **2001**, *22*, 1154–1166.
- (94) Stone, A. J. Distributed Multipole Analysis, or How to Describe a Molecular Charge Distribution. *Chem. Phys. Lett.* **1981**, *83*, 233–239.

Graphical TOC Entry

B86bPBE-XDM HF-3c/MINI(x) HF-D2/MINI(s)
PBE-D2/DZP CRYSTAL
SIESTA PBEh-3c/def2mSVP
vdW-DF2/DZP DFTB+
SCC-DFTB3-D3(BJ)