This is a repository copy of *Off-the-shelf DFT-DISPersion methods : Are they now "on-trend" for organic molecular crystals?*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/149207/

Version: Published Version

Article:

Geatches, Dawn, Rosbottom, Ian, Marchese Robinson, RIchard et al. (6 more authors) (2019) Off-the-shelf DFT-DISPersion methods : Are they now "on-trend" for organic molecular crystals? Journal of Chemical Physics. 044106. ISSN 1089-7690

https://doi.org/10.1063/1.5108829

Off-the-shelf DFT-DISPersion methods: Are they now "on-trend" for organic molecular crystals?

Dawn Geatches [iD], Ian Rosbottom [iD], Richard L. Marchese Robinson [iD], Peter Byrne [iD], Phil Hasnip [iD], Matt I. J. Probert [iD], Dominik Jochym [iD], Andrew Maloney [iD], and Kevin J. Roberts [iD]

View Online   Export Citation   CrossMark

ARTICLES YOU MAY BE INTERESTED IN

The Journal of Chemical Physics

Submit Today

The Emerging Investigators Special Collection and Awards
Recognizing the excellent work of early career researchers!

Off-the-shelf DFT-DISPersion methods: Are they now "on-trend" for organic molecular crystals?

Dawn Geatches,[1,a] (ID) Ian Rosbottom,[2,b] (ID) Richard L. Marchese Robinson,[2] (ID) Peter Byrne,[3] (ID) Phil Hasnip,[3] (ID) Matt I. J. Probert,[3] (ID) Dominik Jochym,[4] (ID) Andrew Maloney,[5] (ID) and Kevin J. Roberts[2] (ID)

AFFILIATIONS

[1] Science and Technologies Facilities Council, Daresbury Laboratory, Sci-Tech Daresbury, Warrington WA4 4AD, United Kingdom
[2] Centre for the Digital Design of Drug Products, School of Chemical and Process Engineering, University of Leeds, Leeds LS2 9JT, United Kingdom
[3] Department of Physics, University of York, Heslington YO10 5DD, United Kingdom
[4] Science and Technology Facilities Council, Rutherford Appleton Laboratory, Harwell Campus, Didcot OX11 0QX, United Kingdom
[5] The Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom

[a] Author to whom correspondence should be addressed: dawn.geatches@stfc.ac.uk
[b] Current address: Department of Chemical Engineering, Imperial College London, South Kensington Campus, South Kensington SW7 2AZ, United Kingdom.

ABSTRACT

Organic molecular crystals contain long-range dispersion interactions that can be challenging for solid-state methods such as density functional theory (DFT) to capture, and in some industrial sectors are overlooked in favor of classical methods to calculate atomistic properties. Hence, this publication addresses the critical question of whether dispersion corrected DFT calculations for organic crystals can reproduce the structural and energetic trends seen from experiment, i.e., whether the calculations can now be said to be truly "on-trend." In this work, we assess the performance of three of the latest dispersion-corrected DFT methods, in calculating the long-range, dispersion energy: the pairwise methods of D3(0) and D3(BJ) and the many-body dispersion method, MBD@rsSCS. We calculate the energetics and optimized structures of two homologous series of organic molecular crystals, namely, carboxylic acids and amino acids. We also use a classical force field method (using COMPASS II) and compare all results to experimental data where possible. The mean absolute error in lattice energies is 9.59 and 343.85 kJ/mol (COMPASS II), 10.17 and 16.23 kJ/mol (MBD@rsSCS), 10.57 and 18.76 kJ/mol [D3(0)], and 8.52 and 14.66 kJ/mol [D3(BJ)] for the carboxylic acids and amino acids, respectively. MBD@rsSCS produces structural and energetic trends that most closely match experimental trends, performing the most consistently across the two series and competing favorably with COMPASS II.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5108829

INTRODUCTION

In industry, scientific research and development needs to strike a balance between how fundamental the research is, and the time allowed for a product to progress from development through to market. While the most fundamental research might be the preserve of academia, industry, nevertheless, has a vested interest in advancing methods and methodology that could ultimately shorten the time spent in the product pipeline. In the pharmaceutical industry, the ADDoPT project (Advanced Digital Design of Pharmaceutical Therapeutics: https://www.addopt.org/about_addopt/) has combined expertise from industry, academia, and small-to-medium enterprises (SMEs) in a combined effort to digitalize the tablet-production pipeline, from the atomistic scale of single molecules to molecular crystals to macroscopic bulk-scale tablets. Both classical and first-principles atomistic simulations of organic molecular

crystals play a role in this pipeline, where their calculated lattice energies could inform structure, performance, properties, and processing,[1] such as thermodynamic solubility,[2–5] stability,[6–8] and crystallization[9–11] as well as crystal structure prediction.[12,13] Lattice energies can be calculated using classical, atomistic methods,[14–20] i.e., molecular mechanics based on force fields, quantum mechanical methods (see, for example, the work of Yang *et al.*[21] and the 2016 review and references within of Hoja, Reilly, and Tkatchenko[22]), and more rarely, hybrid quantum mechanics/molecular mechanics.[23,24] The choice of method depends on several factors, such as in-house expertise, computational resources, and time allotted for the research, and not least the modeling aims.

By definition, lattice energy corresponds to energy differences associated with the static lattice, ignoring zero-point vibrations, at 0 K. Our aim here was to evaluate the performance of lattice energy calculations. Rigorous derivation of experimental lattice energy would require experimental vibrational energy data and/or information regarding the zero point vibrational energy levels, coupled with accurate heat capacity data from the experimental temperature to 0 K.[25,26] Hence, in keeping with many literature studies, we employed a common approximation (further detailed in the section titled "Experimental lattice energies") to estimate the experimental lattice energy from the available experimental data at higher temperatures (typically 298 K). Although, in principle, lattice energy is a simple concept—the change in energy when infinitely separated, static gas-phase molecules in their lowest energy conformations, condense to form a static lattice[6]—calculating the lattice energies of organic molecular crystals is not straightforward. These systems prove challenging to model, thanks to their diverse inter- and intramolecular forces, such as the covalent forces binding the atoms within a molecule, and the different kinds of weaker interactions between molecules, i.e., intermolecular forces. The balance between the inter- and intramolecular forces results in polymorphism, where the same molecules pack in different orientations and potentially different conformations to produce different crystal structures.[27] The intermolecular forces depend on the electrostatic interactions, polarization, exchange-repulsion, and dispersion.[28] When viewed through the lens of electronic structure (i.e., the electron density), the foundations of these interactions are closely linked.[22] When viewed classically, these interactions are separable and encapsulated by the many terms required to build a typical force field.[29]

While dispersion forces are integral to force field modeling, albeit commonly captured imperfectly,[30] they have been largely absent from quantum mechanical-based electron density modeling until relatively recently. Nonetheless, following the introduction of a semiempirical method to predict intermolecular forces in 1975,[31] the application of dispersion-corrected density functional theory (DFT) (labeled within this work "DFT-DISP" to avoid confusion with "DFT-D" which usually refers solely to Grimme's methods) grew steadily throughout the 1990s and then exponentially through the 2000s,[32] and today, the development of accurate methods for condensed matter systems is thriving. Indeed, it can be said that progress in DFT-DISP is climbing a "Jacob's ladder of dispersion functionality" analogous to the more familiar Jacob's ladder of exchange-correlation functionals.[33] On the first step of the dispersion stairway is the most basic term required in dispersion—the r^{-6} term—that describes the asymptotic long-range interaction between particles (separated by distance r) in the gas-phase. The total energy is then $E_{tot} = E_{DFT} + E_{disp}$, where

$$E_{disp} = -\sum_{A,B} C_6^{A,B}/r_{A,B}^6, \qquad (1)$$

and the $C_6^{A,B}$ terms are the pairwise, additive, direction-independent, dispersion coefficients between elements A and B, obtained from a table of constants, which in the case of Grimme's original DFT-DISP method[34] is based on empirically derived C_6 terms.[35] The latter are obtained by least-squares fitting to molecular C_6 coefficients obtained from the dipole oscillator strength distribution method of Thomas and Meath (e.g., Ref. 36) although this is not the only means of obtaining the coefficients (see Ref. 32 and the references within for further examples). The inconsistency of deriving the dispersion coefficients led to the development of Grimme's DFT-DISP2 scheme,[37] where the ionization potentials are coupled with static polarizabilities of isolated atoms to generate a more first-principles scheme for deriving the C_6 coefficients.

The second step of the dispersion stairway was to include the local chemical environment that accounts for the number of neighbors and hence the hybridization state of elements, which means that the C_6 coefficients are no longer constant but instead decrease as the number of neighbor atoms increases. Grimme *et al.* developed this DFT-D3 scheme[38] using two- and three-body terms, where the former are dipole-dipole interactions and the latter triple-dipole. The dispersion C_6 coefficients are calculated using accurate, *ab initio* time dependent DFT (TDDFT) methods and are tabulated in a look-up table to be employed according to the geometrical environment of each atom.

Grimme *et al.*[38] extended DFT-D3 further by enabling the calculation of the C_6 coefficients on-the-fly and including the local electron density as opposed to purely geometrical information. As in DFT-D3, TDDFT polarizabilities are tabulated, but for the extended (DFT-D4) method, they are scaled according to a calculated charge (not simply the number and location of the nearest neighbors), which leads to charge- and coordination number-dependent C_6 coefficients.[39]

On the third step of the dispersion stairway sit the methods that calculate the dispersion interaction directly from the electron density, which, in principle, are an improvement over the methods below on steps one and two, where the C_6 coefficients are precalculated. The step three methods are described as "nonlocal correlation functionals" because they add long-range correlations to local or semilocal correlation functionals, perhaps the most well-known of which is the van der Waals density functional (vdW-DF).[40] Although they showed some improvements, further developments were needed.[41,42] Step three and below accommodates methods that are pairwise additive, which means that two atoms or molecules have the same dispersion interaction regardless of the other atoms or molecules with which they interact. Although this seems to be "good enough" for small molecules in the gas-phase, this is not representative of materials in the condensed phase.[43] Although organic molecular crystals are not as densely packed as metals, semiconductors, and ionic crystals, they contain many-body long-range correlation and dispersion, which explains the instances of failure of pairwise methods when applied to molecular crystals.[25,44,45] (The interested reader is directed to Ref. 32 for further details of the step three methods.) Many-body dispersion (MBD) energy includes

many-body effects such as the dispersion interactions between two atoms (i.e., the pairwise interactions), between three atoms described by the Axilrod-Teller-Muto term,[46] and between four, five, six, etc., up to an infinite number of atoms.[47,48] In addition, many-body dispersion inherently includes screened polarizability effects, which are prevalent in, for example, 2D nanomaterial systems[49,50] where long-range charge fluctuations are present. Considering the organic molecular crystals under focus in our research, it could be possible to find 2D planes of aligned molecules forming a metalliclike layer[51] and therefore enhanced, screened polarizability. Essentially, the many-body effects become significant in the absence of other interactions dominating, such as covalent and hydrogen-bonding and where a system is strongly anisotropic.
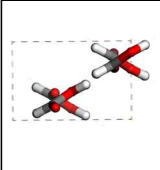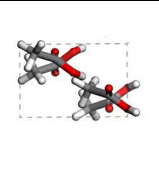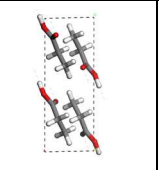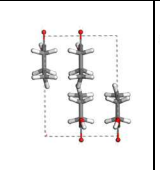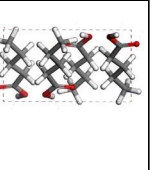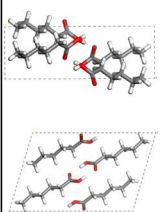
It is only beyond step three that we find the development of the many-body dispersion (MBD) method[52] that captures the aforementioned, long-range, many-body dispersion energy that has shown itself to be accurate for a range of molecular systems.[47,53] The MBD method combines the (step two) van der Waals method of Tkatchenko and Scheffler[54] (TS) with the self-consistent screening equation of classical electrodynamics. The TS method uses reference atomic polarizabilities and C_6 coefficients from which it is possible to calculate the C_6 coefficients for a pair of unlike atoms. The dispersion coefficients representing the chemical environment are obtained from effective atomic volumes by comparing the Hirshfeld-partitioned electron density of an atom-in-a-molecule with that of a free atom, which gives a C_6 scaling factor for a reference atom, thereby changing the value of the dispersion energy. The MBD method uses these environment-dependent C_6 values including effective long-range screening in the atomic polarizabilities, together with a scheme based on the coupled fluctuating dipole model[55] to account for many-body effects.

The original MBD method has since been revised as the range-separated self-consistent screened MBD, i.e., MBD@rsSCS,[56] that separates short- and long-range correlation, whereby the former Coulomb interaction is calculated using semilocal or hybrid DFT, and the latter, long-range correlation using the localized, coupled atomic response functions of MBD. Determination of the range separation is achieved via a single parameter that is fitted to accurate quantum chemistry benchmarks (see Ref. 56 for full details). MBD@rsSCS has been shown to be a highly efficient and accurate method for calculating the long-range correlation energy in finite-gap systems including molecular crystals–see, for example, results of Ambrosetti et al.[56] for the X23 set of structures.[45,57] With MBD@rsSCS, we have reached the current state-of-the-art modeling method for the inclusion of dispersion forces in the condensed phase, for systems of any size that DFT already models "comfortably," where "comfortably" means using whatever hardware and resources are currently used for nondispersion DFT calculations.

Even once the quantum chemical or classical method for computing the energies of the solid state and gas phase structures is selected, lattice energies can be calculated in a number of different ways, depending on whether, for example, the unit cell parameters are allowed to relax or are kept fixed at their experimental values, the extent to which the gas phase molecules extracted from the crystal structure are allowed to relax.[58] The pharmaceutical industry partners involved in the ADDoPT project mainly employ classical force field methods for lattice energy calculations. One outcome of the ADDoPT project was to determine an optimum protocol
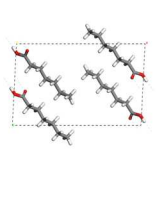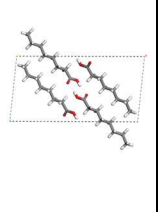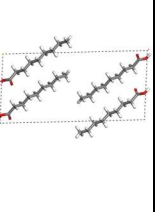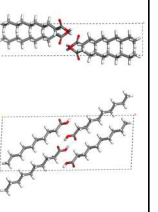
for calculating lattice energies of pharmaceutically relevant organic molecular crystals, and this protocol involves the classical force field COMPASS II.[59] This is an extension of COMPASS[60] (condensed-phase optimized molecular potentials for atomistic simulation studies) that is itself based on the polymer consistent force field (PCFF) that forms part of a second generation of closely related force fields that were designed specifically for polymers and organic materials.[61] COMPASS was parameterized using ab initio calculations and empirical data, which entailed adding new molecular classes to PCFF. In addition, nonbond parameters were reparameterized, whereby the electrostatic and van der Waals terms combine quantum mechanical calculations and fitting to experimental condensed phase properties of liquids and crystals. COMPASS II extends this to include parameters specific to polymers and druglike molecules, hence its wide adoption in the pharmaceutical industry. However, given the current availability of DFD-DISP options, is it time to rethink this default position of using classical methods to calculate lattice energies? Attempting to answer this question generated the research we present in this paper, in which we assess whether state-of-the-art DFT-DISP methods are comparably predictive to, or better than classical methods. We use two different simulation methods to calculate the lattice energies of the organic molecular crystals, namely, classical molecular mechanics using a force field and ab initio density functional theory (DFT) plus dispersion (DFT-DISP). The molecular mechanics was performed using the Forcite module of Materials Studio,[62] with the aforementioned COMPASS II force field, and the DFT-DISP calculations were carried out using the pseudopotential, plane wave code CASTEP[63] employing D3(0)[38] and D3(BJ) of Grimme et al.[64] and MBD@rsSCS of Ambrosetti et al.[56] (The improvement of CASTEP to implement state-of-the-art DFT-DISP methods efficiently was catalyzed by the ADDoPT project: https://www.addopt.org/news/latest_news/castep_archer/.) These methods are described as "Forcite," "D3," "D4," and "MBD*" (and "MBD" in the figures), respectively, from this point onwards. [At the time of developing CASTEP, implementing Grimme's extended DFT-D3 method, i.e., DFT-D4,[39] was out-of-scope; for the purposes of simplifying labeling, the label "D4" is used to distinguish DFT-D3(BJ) from DFT-D3(0) only.] We also include synthon (i.e., intermolecular) analysis of the dispersion contributions within the calculated structures, using the software HABIT98.[65] (For an overview of synthonic analysis, see, for example, the work of Roberts et al.[66,67]) We have chosen two series of homologous crystals—carboxylic acids and amino acids—wherein both the dispersion forces are expected to increase with an increasing number of carbon atoms, and where the carboxylic acids are expected to have a higher ratio of van der Waals to hydrogen bonding, and the reverse is true for the amino acids. The amino acid series also potentially exhibit charge-charge interactions, due to being zwitterionic in the solid state.

In the Results and Discussion section, we explore trends in the calculated lattice energies, compare lattice and molecular structures, and bonding, as well as the contribution of dispersion energy, with comparison to the experimental structures and experimental data where possible. Our aim is to compare the performance of the different methods throughout the two series and to draw conclusions based on the analysis of their trends as to whether current, "off-the-shelf" DFT-DISP methods are now worthy of becoming part of industry's modeling pipeline toolkit.

Formic C = 1 Pna21 FORMAC01	Acetic C = 2 Pna2$_1$ ACETAC07	Propanoic C = 3 P2$_1$/c PRONAC	Butyric C = 4 C2/m BUTRAC	Valeric C = 5 P2$_1$/c VALRAC
Hexanoic C = 6 P2$_1$/c ISENID	Heptanoic C = 7 P2$_1$/c ISENOJ	Octanoic C = 8 P2$_1$/c ISENUP	Nonanoic C = 9 P2$_1$/c ISEPAX	Decanoic C = 10 P2$_1$/c ISEPEB

FIG. 1. Carboxylic acid series—experimental structures (after the addition of hydrogen atoms where applicable) of the first 10 carboxylic acids, the number of carbon (C) atoms, their symmetry groups, and CSD REFCODES.

CRYSTAL STRUCTURES, MODELS, AND METHODS

Two homologous series of organic molecular crystals were explored. Series 1 comprises the set of carboxylic acids from 1 to 10 carbon atoms in the hydrocarbon backbone, i.e., formic, acetic propanoic, butyric, valeric, hexanoic, heptanoic, ocatanoic, nonanoic, and decanoic acids (see Fig. 1). Series 2 comprises the amino acids with one to four carbon atoms in the hydrocarbon backbone: alpha-glycine, L-alanine, L-valine, and L-isoleucine (see Fig. 2).
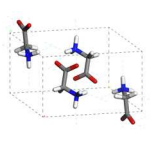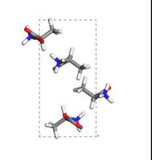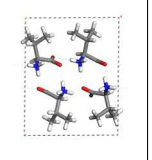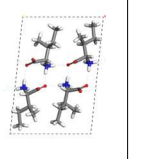
All CIF files were obtained from the Cambridge Structural Database (5.38)[68] and uploaded into Materials Studio[62] where they were visualized, inspected, and supplemented with missing hydrogen atoms by-hand. (All REFCODES can be found in the supplementary material as well as in Figs. 1 and 2.) The input parameters were set (as described below) to optimize the geometry of the crystal structures, where geometry optimization involved relaxing

the lattice lengths and angles according to the structures' symmetry constraints, including all molecules and their constituent atoms. For the gas-phase calculations for both force field and DFT methods, a single molecule from each of the unrelaxed crystals was also geometry-optimized until the respective force field and DFT convergence criteria were met. Lattice energies (LEs) were calculated using the following equation: $LE = \frac{E_c}{Z} - E_g$, where E_c is the total, calculated energy of the unit cell, Z is the number of molecules per unit cell, and E_g is the total energy of a single gas-phase molecule. From this point onwards, "experimental structures" and "pre-optimized structures" are used interchangeably.

FORCE FIELD CALCULATIONS

All force field calculations were performed using the Forcite module of Materials Studio. Detailed instructions for how to repeat these calculations are provided under the section titled "Forcite calculations" in the supplementary material. Following structure preprocessing, including bond assignment required for force field calculations, geometry optimizations then proceeded as described above. The force field used was COMPASS II;[59] the total energy was converged to 8.37×10^{-5} kJ/mol, forces to 0.0042 (kJ/mol)/Å, stress (i.e., for bulk crystals) to 0.001 GPa, and displacement to 1.0×10^{-5} Å. Ewald summation[69,70] was employed throughout for all electrostatic interactions.

All Forcite calculations were run using a single core on a desktop computer [Intel(R) Core(TM) i5-6300U CPU at 2.40 GHz, 7.41 GB usable RAM] running Windows 7 Enterprise (Service Pack 1, 64-bit). Running each optimization operation took under 1 min, including Forcite module start-up time, based on timing the operations for L-valine (starting from the CIF file prepared from CSD refcode LVALIN05).

α-Glycine P2$_1$/n GLYCIN86	L-Alanine P2$_1$2$_1$2$_1$ LALNIN24	L-Valine P2$_1$ LVALIN05	L-Isoleucine P2$_1$ LISLEU02

FIG. 2. Amino acid series—experimental structures of the first four amino acids, their symmetry groups, and CSD REFCODES.

DFT CALCULATIONS

All DFT calculations were performed using the plane wave, pseudopotential code, CASTEP v19.0[63] and visualized in Materials Studio.[62] For the gas phase calculations, a single molecule was copied from the unrelaxed bulk phase and placed in a cube of vacuum, 20 Å³. The plane wave basis set representing the valence electron wave functions was converged to an accuracy of greater than 0.001 meV/atom (i.e., 0.1 J mol⁻¹), corresponding to a kinetic energy cutoff of 700 eV. The exchange-correlation interactions were described by the generalized gradient approximation density functional of Perdew, Burke, and Ernzerhof, i.e., PBE-GGA,[71] and the electron-ion interactions were described by corresponding ultrasoft pseudopotentials generated on-the-fly. Three types of long-range dispersion forces were applied using the semiempirical dispersion correction (SEDC) module[72] including the many-body MBD@rsSCS interactions of Ambrosetti *et al.*[56] and the newly implemented D3(0)[38] and D3(BJ) of Grimme *et al.*[64] We also used PW91-GGA[73] without any dispersion corrections to serve as a control.

For the bulk phase molecular crystals, the Brillouin zone was sampled on various Monkhorst-Pack grids,[74] as listed in Table S1. For the gas phase molecules, a single sampling point corresponding to the gamma point was used. The geometry of both the molecular crystals and the gas-phase molecules was optimized using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm,[75] and the electronic method used in the self-consistent field (SCF) calculations was density mixing. The following convergence criteria were also applied: electronic energy for the SCF cycles: 1×10^{-12} eV; total energy following geometry optimization: 5×10^{-8} eV; maximum force component: 0.001 eV Å⁻¹; and geometrical displacement 5×10^{-4} Å.

All DFT calculations were run using between 24 and 96 core processors (Intel Xeon processors—Ivy Bridge E5-2697v2 2.7 GHz) optimized according to the number of atoms, volume of the unit cell, and type of calculation. The calculations took from less than 1 h (formic acid crystal) to approximately 70 h (decanoic acid crystal), before the aforementioned convergence criteria were met.

EXPERIMENTAL LATTICE ENERGIES

Sublimation enthalpy data at a defined temperature were retrieved, or—in the case of valeric acid—estimated (see below), for all compounds studied, save for the following entries in the carboxylic acid series: hexanoic (C = 6), heptanoic (C = 7), and nonanoic (C = 9) acids. Synonyms were retrieved via consulting the corresponding CSD entry and used to search the US National Institute for Standards and Technology (NIST) Chemistry Webbook: https://webbook.nist.gov/chemistry/name-ser/ and the literature[76–80] for standard enthalpy of sublimation data at 298–298.15 K, or, if no such data were available, enthalpy of sublimation data at the closest temperature to 298 K. Efforts were made to trace experimental sublimation data points back to their primary reference although these could not always be accessed.

In one case (valeric acid), no experimental sublimation enthalpy could be found. Hence, the approach of Lifson *et al.*[81]

was applied: the sublimation enthalpy at 298.15 K was estimated via adding the melting enthalpy[81] to the enthalpy of vaporization[81,82] at 298.15 K. In another case (L-valine), the sublimation enthalpy estimate recommended by Dorofeeva and Ryzhova,[78] obtained via the difference in enthalpy of formation in the gaseous and solid states, was used.

For only one datapoint—one of the datapoints[80] retrieved for decanoic acid linked to CSD refcode ISEPEB—was experimental metadata identified that allowed confirmation that the sublimation data corresponded to the polymorph for which the crystal structure was retrieved from the CSD. However, pairwise differences in calculated lattice energies and the related[2] experimental sublimation enthalpies between polymorphs are reported to typically be less than 2 kJ/mol (greater than 7.2 kJ/mol in only 5% of cases)[83] and a few percent,[84] respectively. Indeed, these are comparable to some estimates of typical experimental error in sublimation enthalpies,[57] albeit these may scale with molecular size,[85] and variation between literature values[86] for the same compound. Moreover, an analysis, documented in the supplementary material, of known polymorphism for all dataset entries found that only formic acid and alpha-glycine were associated with known polymorphism in the CSD (version 5.38).

In the case of decanoic acid and propanoic acid, multiple sublimation enthalpy data points were retrieved according to the noted selection criteria. Hence, the arithmetic mean enthalpy value was used.

After all crystal structures were associated with a single sublimation enthalpy datapoint, the experimental lattice energy (*LE*) was estimated from the (average) sublimation enthalpy (*deltaHsub*) using the following approximate relationship,[2,25] where *R* is the molar gas constant and *T* is the temperature in Kelvin, which ranged from 223 to 455 K, across different dataset entries: $LE = -deltaHsub - 2RT$. N.B. The assumptions made in obtaining this relationship are discussed in the cited references. These approximate experimental lattice energies were used to benchmark the lattice energy calculations obtained in the current work.

An Excel workbook linking all CSD refcodes, compound names, sublimation enthalpy data/experimental estimates, and the sublimation datapoint specific reference is provided in the supplementary material. In addition, the experimental lattice energies derived from these data, along with a Python script for performing the derivation, is also provided in the supplementary material.

SYNTHON ANALYSIS

Synthon analysis was carried out on all of the geometry-optimized structures with the aim of rationalizing the contribution of dispersion forces to the overall lattice energy. The dispersion synthons were approximated by measuring the distance between the centers of the molecules in the relaxed crystals and comparing this distance to the distance measured between the two same molecules in the pre-optimized structure. Although this is not a direct comparison with the analysis of the DFT dispersion energy contributions of the relaxed MBD*, D3, and D4 structures to the total energy, both types of analyses (classical and DFT) are expected to produce similar trends.

RESULTS AND DISCUSSION

Carboxylic acid series: Series 1

Crystal structure and lattice energies

Table I gives the details of the experimental lattice energies, lattice lengths, and angles for the pre-optimized CIF (labeled "Expt.") and all structures (except PW91) following relaxation using Forcite, MBD*, D3, and D4, together with the mean absolute errors (MAEs) of the calculated lattice energies, volumes, and β-angles. These and further details of carbonyl and hydroxyl torsions and hydrogen bond lengths are shown in the graphs of Fig. 3. Figures of the crystal structures obtained following relaxation under each method (except PW91) can be found in Fig. S17 of the supplementary material.

The *lattice energies* for the dispersion corrected calculations generally show monotonic increases in magnitude (i.e., they become increasingly negative) with an increasing number of carbon atoms (see Fig. 3). This is also the general trend for the nondispersion corrected calculations of PW91 although the increase in magnitude is much less significant. Consequently, we do not include any further analysis of the PW91-relaxed structures. The experimental values for hexanoic, heptanoic, and nonanoic acids have been interpolated by drawing straight lines between the lattice energies of valeric, octanoic, and decanoic acids. The Forcite lattice energies underestimate the magnitudes of the experimental lattice energies by 6–12 kJ/mol, whereas the MBD* values typically overestimate the experimental values by between 5 and 15 kJ/mol, with the largest difference occurring for decanoic acid. This larger difference occurs because the MBD* trend is slightly positively quadratic compared to the experimental values that show a more linear trend. For the carboxylic acids with up to three carbon atoms, MBD*, D3, and D4 have similar lattice energies and agree with the experimental value from between 10% and 15%; this trend is broken by butyric acid, whose D3 and D4 lattice energies are smaller (in magnitude) than that calculated using MBD*, and for valeric acid, the reverse is true. For butyric acid, the MBD*, D3, and D4 lattice energies are 16%, 10%, and 12% larger than the experimental value, and for valeric acid, they are 11%, 25%, and 26%, respectively. The lattice energy MAE follows the trend D3 > MBD* > Forcite > D4 within the range 8.52–10.57 kJ/mol, although within each method, the variation in values is largest in D4 where the lattice energy differs by approximately 5–23 kJ/mol from the experimental value. The MAEs were calculated using the experimental values available and hence do not reflect the off-trend values seen in D3 and D4 for hexanoic acid. To check the sensitivity of the optimization parameters on the lattice energy, two additional lattice energy values were calculated for hexanoic acid by using a different k-point set (see the brown "pluses" on the top graph of Fig. 3 and Table S1 for details) and these are lower by about 9 kJ/mol than those already discussed. Although the lattice energy decreased further, the structure of hexanoic acid did not distort further, and so future reference to hexanoic's lattice energy concerns its highest lattice energy value and corresponding structure.

Considering the *lattice volumes*, there is a monotonic increase in lattice volumes with an increasing number of carbon atoms, except for D4 from five to six carbon atoms. In every case except formic acid, the MBD* volume decreased from the experimental value, whereas the D3 and D4 volumes increased for formic, acetic, and valeric acids, decreased for butyric, hexanoic, and nonanoic acids, and both increased (D3) and decreased (D4) for heptanoic, octanoic, and decanoic acids. Nonetheless, the D3 and D4 calculated volumes are always larger than the MBD* calculated volumes. Throughout Series 1, the experimental volumes are typically the largest or close to the largest, which is not surprising given that the X-ray crystallography experiments are carried out at temperatures higher than the 0 K, DFT calculation temperature, and therefore will include a crystal expansion. Even so, the experimental volumes are closely matched by the D3 and D4 volumes where these have increased. The Forcite and MBD* volumes are also close to one another and are consistently smaller than D3, D4, and experimental volumes. The difference between the two groups (Forcite and MBD* vs D3 and D4) increases with the number of carbon atoms.

In keeping with the observation regarding the change in optimized volumes, the lattice lengths of the D3 and D4 structures are generally larger than the experimental, Forcite, and MBD* calculated lattice lengths, and the latter two methods produce similar lattice lengths. The MAE of the lattice volumes decreases in the order MBD* > Forcite > D4 > D3 with the range being from 60.75 to 18.77 Å^3.

The *beta angles* of the molecular crystals that most closely match those of the pre-optimized CIFs are calculated by Forcite, and in general, there is close agreement among all methods, except for the hexanoic acid structures relaxed using D3 and D4. Although they closely match one another, they deviate significantly (~30°) from the experimental, Forcite, and MBD* beta angles. This difference clearly indicates that the crystal structures of the two sets are different. An examination of Fig. 4 shows that, in the D3 and D4 optimized structures, the molecules appear to have "stretched" although in reality, they have rotated along the x-axis and relaxed to become more parallel to the x-axis. The crystal structure of hexanoic acid is unique among the carboxylic acids insofar as the dimer-bonded molecules are aligned directly along the x-z axis (see Fig. S1 for comparison). The MAE of the beta-angles decreases in the order D3 > D4 > MBD* > Forcite within the range 4.50°–1.33°.

The intermolecular *hydrogen bonds* (see Fig. 3) were measured within Materials Studio and were indicated by a distance between a hydrogen donor and oxygen acceptor less than or equal to 2.5 Å, with an angle of at least 90°. The hydrogen bond lengths remain fairly consistent within each of the five methods with the largest variation in length seen in the experimental structures, where the hydrogen bond lengths found in propanoic acid are approximately 0.2 Å shorter than those found in all other experimental structures. The average lengths of the hydrogen bonds are approximately: experimental: 1.78 Å, Forcite: 1.68 Å, D3 and D4: 1.57 Å, and MBD*: 1.53 Å. In formic and acetic acids, the two hydrogen bonds (H-bonds) per molecule are between two other molecules forming strongly hydrogen-bonded sheets of molecules in the y-z plane that are weakly bound in the x-direction. Generally, the H-bonds in the experimental crystal structures are longer by an approximate average of 14% than those of the relaxed Forcite, MBD*, D3, and D4 structures, which is to be expected given that their positions are difficult to obtain experimentally. Also, the experimental structures reflect the effects of thermal expansion at nonzero Kelvin as well as zero-point motion. For each of the remaining crystals of Series 1,

TABLE I. Lattice energies and structures of the carboxylic acid series. *a*, *b*, and *c* are lattice lengths (per row); v is the unit cell volume (per row); and α, β, γ are the lattice angles (per row). The experimental lattice energy values are calculated from enthalpies of sublimation using the approximation: $\Delta H_{sub} = -$ lattice energy $- 2RT$, where *R* is the gas constant and *T* is the experimental temperature. Experimental values for hexanoic, heptanoic, and nonanoic acids could not be found in the literature. MAE: mean absolute error of lattice energies, volumes, and *β*-angles.

Structure	Lattice energy magnitude (kJ/mol)					Lattice lengths (Å),					Lattice angles (deg)				
	Expt.	Forcite	MBD*	D3	D4	Expt.	Forcite	MBD*	D3	D4	Expt.	Forcite	MBD*	D3	D4
Formic acid	65.07	55.84	70.89	72.44	71.34	*a*: 10.24	10.46	10.42	10.72	10.53	α: 90.00	90.00	90.00	90.00	90.00
						b: 3.54	3.48	3.66	3.63	3.66	β: 90.00	90.00	90.00	90.00	90.00
						c: 5.36	5.38	5.33	5.47	5.38	γ: 90.00	90.00	90.00	90.00	90.00
						v: 194.4	195.9	203.4	213.1	207.5					
Acetic acid	70.71	63.22	79.83	80.78	77.56	*a*: 13.15	14.28	12.93	13.23	13.14	α: 90.00	90.00	90.00	90.00	90.00
						b: 3.92	3.81	3.94	3.93	3.95	β: 90.00	90.00	90.00	90.00	90.00
						c: 5.76	5.26	5.73	5.86	5.83	γ: 90.00	90.00	90.00	90.00	90.00
						v: 297.3	286.5	291.6	305.1	302.4					
Propanoic acid	77.37	67.36	85.34	85.02	82.43	*a*: 4.04	3.88	3.91	3.96	3.95	α: 90.00	90.00	90.00	90.00	90.00
						b: 9.06	8.96	8.77	9.23	9.02	β: 91.25	90.76	92.48	85.00	89.26
						c: 11.00	10.81	11.03	11.17	11.08	γ: 90.00	90.00	90.00	90.00	90.00
						v: 402.5	375.9	377.7	406.8	395.3					
Butyric acid	80.12	74.62	92.59	87.69	89.36	*a*: 8.01	7.75	7.74	8.09	8.00	α: 90.00	95.49	96.78	96.85	96.49
						b: 6.82	6.68	6.58	6.60	6.67	β: 111.45	114.05	112.5	109.8	110.9
						c: 10.14	9.92	10.09	10.31	10.13	γ: 90.00	96.04	91.07	91.22	91.16
						v: 515.6	460.9	470.9	513.1	500.6					
Valeric acid	88.46	82.37	98.56	110.97	111.17	*a*: 5.55	5.51	5.52	5.56	5.50	α: 90.00	90.00	90.00	90.00	90.00
						b: 9.66	9.19	9.57	10.33	11.05	β: 101.82	101.65	101.3	100.3	99.21
						c: 11.34	11.14	10.93	11.27	10.96	γ: 90.00	90.00	90.00	90.00	90.00
						v: 595.4	552.8	566.1	637.3	657.8					
Hexanoic acid	. . .	86.10	107.85	117.36	120.18	*a*: 15.02	14.61	14.8	18.76	18.59	α: 90.00	90.00	90.00	90.00	90.00
						b: 5.02	4.83	4.89	4.03	4.02	β: 106.55	106.7	109.2	78.04	78.98
						c: 9.94	9.95	9.36	9.24	9.08	γ: 90.00	90.00	90.00	90.00	90.00
						v: 719.0	673.4	638.9	683.8	666.5					
Heptanoic acid	. . .	95.29	115.49	110.72	110.00	*a*: 16.01	14.81	14.42	15.61	15.22	α: 90.00	90.00	90.00	90.00	90.00
						b: 5.08	4.87	4.89	5.17	5.14	β: 92.86	92.14	91.70	89.66	89.99
						c: 10.16	10.28	10.29	10.24	10.23	γ: 90.00	90.00	90.00	90.00	90.00
						v: 824.6	741.5	725.3	827.3	801.0					
Octanoic acid	118.26	101.12	125.66	118.05	118.55	*a*: 18.66	18.35	18.17	19.11	19.02	α: 90.00	90.00	90.00	90.00	90.00
						b: 4.98	4.93	4.85	5.04	4.90	β: 95.77	93.33	93.07	94.46	96.34
						c: 9.57	9.24	9.21	9.47	9.52	γ: 90.00	90.00	90.00	90.00	90.00
						v: 884.6	835.3	810.8	909.1	881.2					
Nonanoic acid	. . .	110.51	137.20	134.19	129.00	*a*: 21.11	18.99	18.45	21.15	21.63	α: 90.00	90.00	90.00	90.00	90.00
						b: 4.92	4.71	4.78	4.93	4.82	β: 93.78	99.05	99.60	95.54	94.61
						c: 10.12	10.41	10.33	9.76	9.53	γ: 90.00	90.00	90.00	90.00	90.00
						v: 1048	919.9	897.7	1012	991					
Decanoic acid	129.16	117.51	144.99	145.80	138.36	*a*: 22.84	23.18	22.48	23.44	23.45	α: 90.00	90.00	90.00	90.00	90.00
						b: 4.96	4.86	4.81	4.98	4.92	β: 93.56	94.97	96.07	92.70	92.52
						c: 9.40	8.88	9.05	9.24	9.12	γ: 90.00	90.00	90.00	90.00	90.00
						v: 1063	997.2	972.4	1077	1050					
MAE	. . .	9.59	10.17	10.57	8.52	v: . . .	50.81	60.75	18.77	25.24	β: . . .	1.33	1.76	4.50	3.80
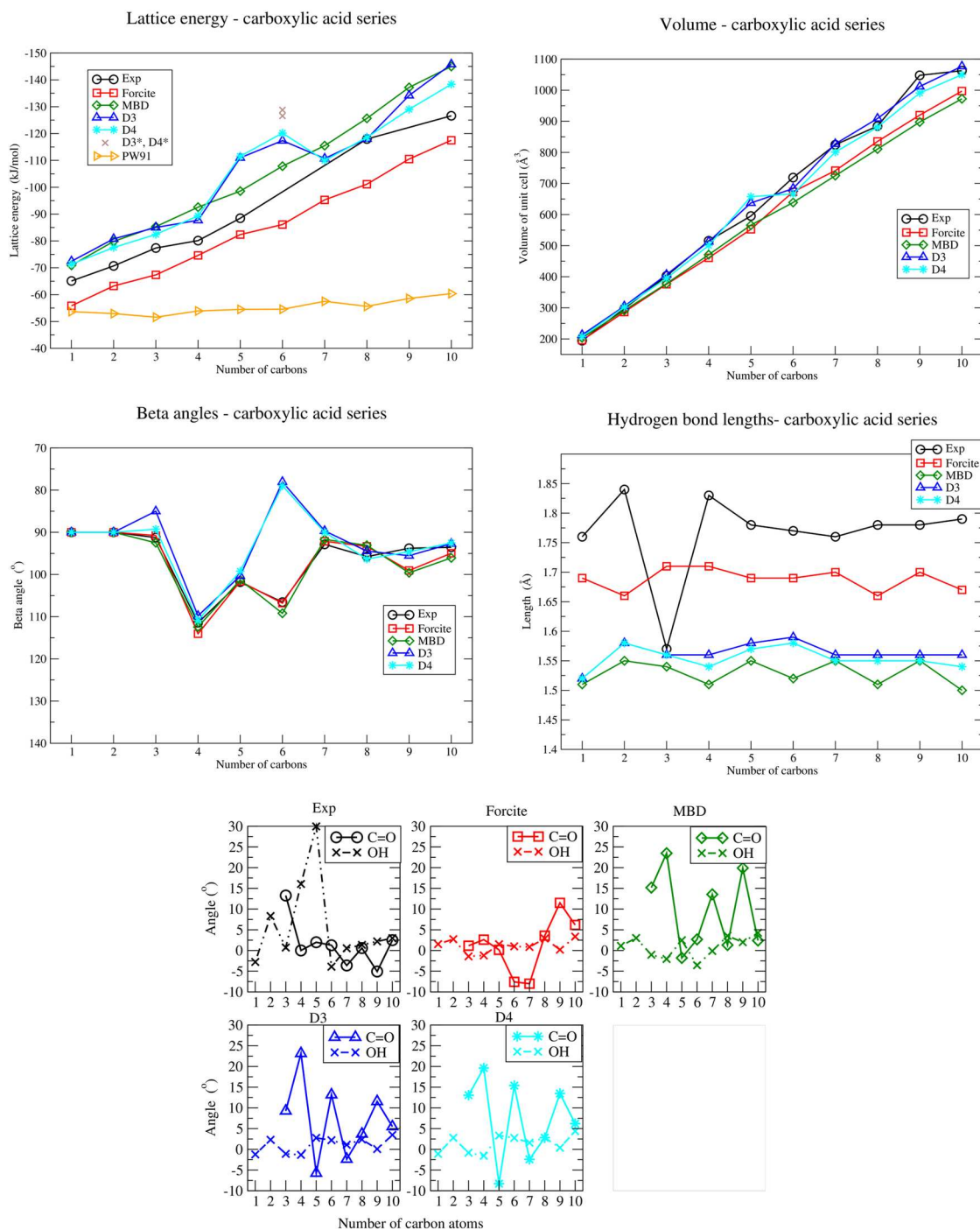
FIG. 3. Analysis of carboxylic acid series, exploring lattice energies, volumes, beta angles, length of hydrogen bonds, and torsion angles of the carbonyl (C=O) bond relative to the hydrocarbon backbone, and of the hydroxyl (OH) bond relative to the C—C=O group. The lattice energy graph shows two extra data points (D3* and D4*) for hexanoic acid (C = 6) that were calculated using a different set of k-points (see the section titled "Results and discussion" for Series 1). The lattice parameters of hexanoic acid's D3, D3*, D4, and D4* relaxed structures were not significantly different; the analysis presented in the text is for the D3 and D4 hexanoic acid structures unless otherwise stated.
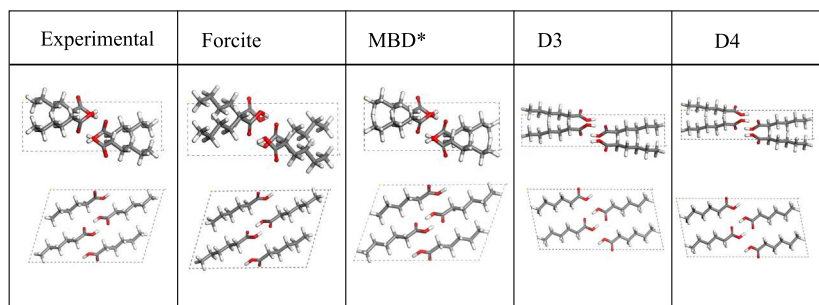
Experimental	Forcite	MBD*	D3	D4

FIG. 4. The original .cif structure (far left hand side) and relaxed structures of hexanoic acid (C = 6).

the hydrogen bonds form dimer-pairs. The average H-bond lengths increase in the order MBD* < D4/D3 < Forcite.

The *torsion angles* are measured with respect to the carbonyl (C=O) angle to the hydrocarbon backbone and the subsequent deviation of the hydroxyl group (HO) from that (O=C—C—C) plane. The experimental and Forcite structures show the least variation in the C=O torsion throughout the series, and the Forcite structures show the least variation in the OH torsion, meaning that the carboxylic group of the Forcite structures is close to planar throughout the series. D3, D4, and MBD* show very similar variations in the range of C=O torsions across the series with a notable exception at C = 5, where D3 and D4 show a less planar carboxylic group than that found by MBD*. These D3 and D4 results correspond with a nonplanar carboxylic group in the experimental CIF although in the latter, the nonplanarity is much larger. In general, where the pre-optimized CIF (i.e., experimental) carboxylic group is nonplanar, all modeling methods produce some degrees of nonplanarity although this is least pronounced in the Forcite models, especially for C = 3, 4, and 5. Given that the crystals of butyric acid (C = 4) and valeric acid (C = 5) from the CSD contained no hydrogen atoms and that these were added via the Materials Studio interface, the variation across all model results in the torsions of the carboxylic group components implies that no bias was introduced by the manually applied, and orientation of, hydrogen atoms that represent those within the physical crystal structures.

In formic acid, the hydroxyl torsions and H-bonds decrease on relaxation causing a slight change in the orientation of the formic acid molecules in MBD*, D3, and D4, mostly causing a slight expansion of the lattice parameters. In acetic acid, the hydroxyl torsions and H-bonds also decrease on relaxation with all modeling methods, with an increase in lattice lengths for the D3 and a mixed increase and decrease for Forcite, MBD*, and D4. The main difference between formic and acetic acid molecules is one methyl group, which suggests that the presence of van der Waals interactions in acetic acid crystals counterbalances the strength of the H-bonding, which is the sole interaction in formic acid crystals, hence the reduced expansion of crystals of acetic acid compared to formic acid.

Figure 5 shows the *dispersion energy* contributions to the total energy for the DFT, semiempirical dispersion contribution (SEDC) methods MBD*, D3, and D4, as well as for the classical method using synthon analysis. For MBD*, the SEDC contribution increases monotonically with an increasing number of carbon atoms both in the bulk crystal and for the gas-phase molecules where the

dispersion is purely intramolecular. Similarly, D3 and D4 show non-monotonic increases in SEDC for the bulk state, except for valeric (C = 5) and hexanoic (C = 6) acids, where there are larger-than-would-be-interpolated increases in the SEDC energies. This contrasts with the gas-phase, where the D3 and D4 SEDC values mirror MBD* in that they show a monotonic increase in intramolecular SEDC energy, with the trend in gradients, of the trendline between SEDC and number of carbons, being D4 > MBD* > D3. The percentage SEDC contributions to the total energy reveal that, for both the gas-phase molecules and the bulk crystals, the dispersion energies are leveling-off for all three methods with the proportion of D3 contributions to the total energy decreasing the fastest. The overall rates of decrease are faster for the gas-phase molecules than for the bulk crystals, which is somewhat reassuring because a gas-phase molecule should only have intramolecular interactions, and, given the "linear" chainlike nature of carboxylic acids, such self-interactions are minimal.

The absolute and percentage SEDC contributions to the total lattice energy (Fig. 5 middle row, right) show linear, monotonic trends for MBD* and similar trends for D3 and D4 with lower absolute and percentage SEDC contributions above C = 6. All trendlines show that the dispersion energy contributions increase with an increasing number of carbon atoms, from 30% (formic acid) to between 65% and 80% (decanoic acid). The synthon analysis (Fig. 5 bottom row) of the percentage change in the van der Waals distance between two molecules (i.e., carbon-to-carbon) in the relaxed crystal compared to the pre-optimized structure shows that up to C = 7, the relaxed molecules mostly moved closer together and less often, further apart than they were prior to relaxation under all methods. Beyond C = 7, the molecules move closer together on relaxation under all methods. Furthermore, following relaxation, molecules with more than seven carbon atoms will be found 10%–15% closer together than their pre-optimized parent structures. This implies that the attractive part of the dispersion energy dominates as the molecules increase in length, and indeed, this is supported by the increasing contribution of the dispersion energy to the overall lattice energy.

As previously noted, the D3 and D4 calculated lattice energies for valeric and hexanoic acids lie outside the general trends identified for these methods. These larger-than-would-be interpolated lattice energies both have larger-than-would-be-interpolated SEDC energies in the bulk crystal. In the relaxed D3 and D4 crystal structures of valeric acid, the outstanding feature compared to the Forcite and MBD* crystals is the larger torsion of the carboxylic group. This
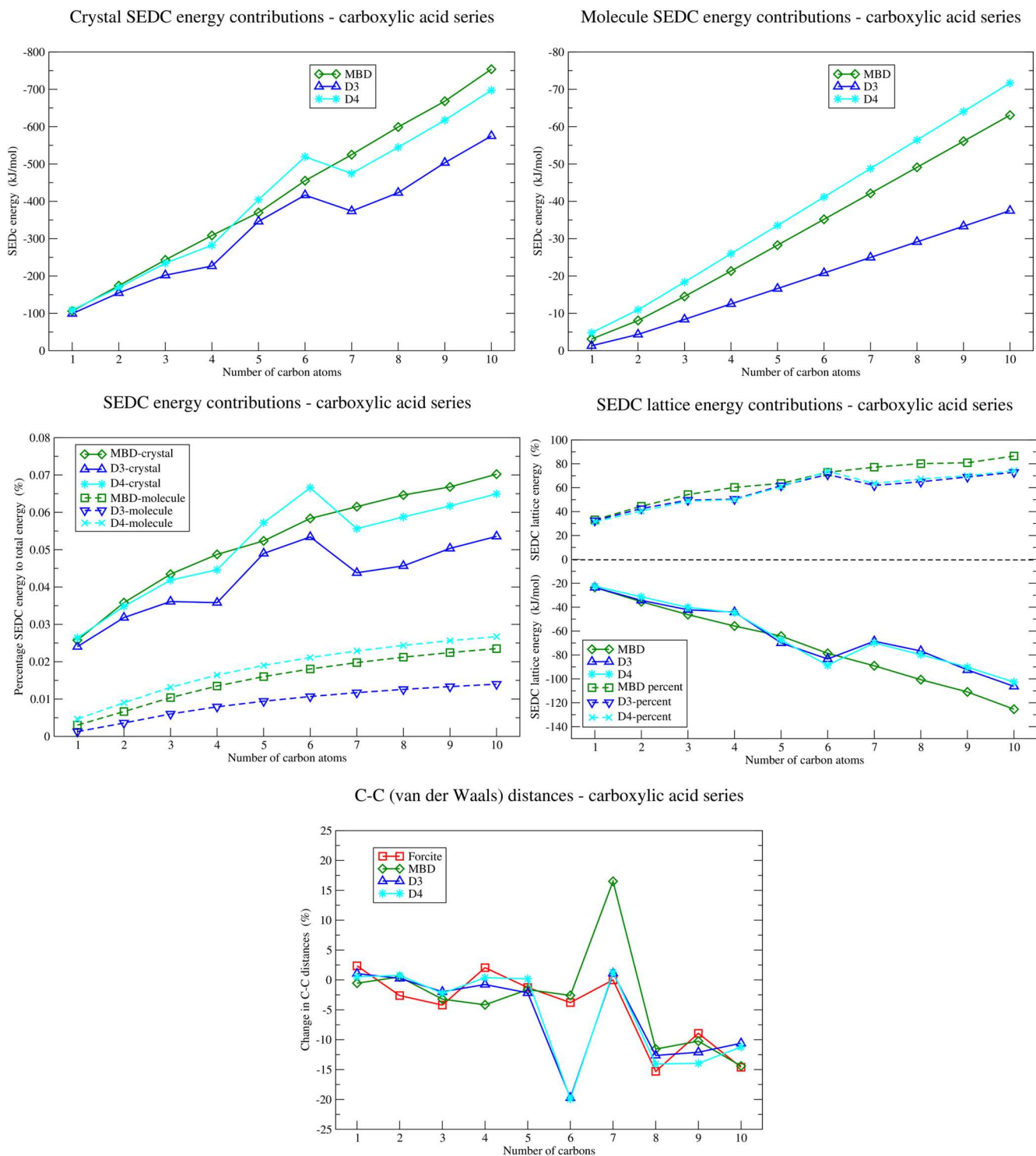
FIG. 5. Top row, semiempirical dispersion correction energies of the three DFT-DISP methods applied to the carboxylic acid series. Middle row left, the percentage of the SEDC energy contributions to the overall total energy; middle row right, the absolute and percentage contributions of the SEDC energy to the total lattice energy. Bottom row, average change in distances after optimization between centers of the molecules (i.e., van der Waals distances) relative to the pre-optimized structures.
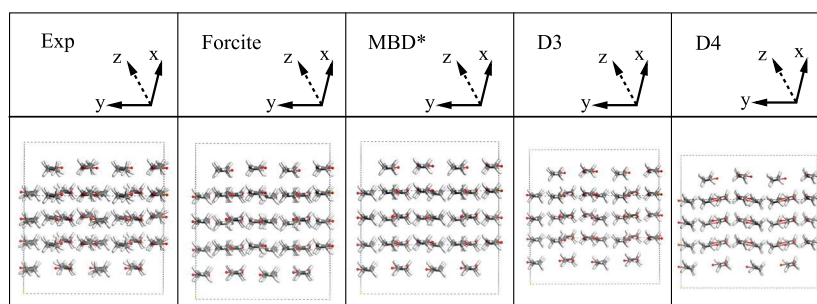
FIG. 6. Supercells of relaxed valeric acid (C = 5) structures showing difference in orientation parallel to the y-axis. The dotted line of the z-axis is to indicate that it is pointing out-of-the-page.

larger torsion enables adjacent molecules to lie slightly staggered parallel to the y-axis, whereas in the Forcite and MBD* structures, adjacent molecules lie in a plane parallel to the y-axis (see Fig. 6). This would affect the dispersion energy between adjacent molecules in the bulk crystal without an accompanying dramatic change in the crystal structure.

Contrasting with the subtle change in the valeric acid crystal are the D3 and D4 hexanoic acid crystal structures where, as previously described, their beta-angles are approximately 30° smaller than those seen in the pre-optimized, Forcite, and MBD* crystals. Like valeric acid, the D3 and D4 SEDC energy contributions in the bulk crystal are larger-than-would-be-interpolated, and the corresponding carboxylic torsions are this time, only slightly larger than those found in the Forcite and MBD* crystals.

A closer examination of the overlap between adjacent molecules along the c-length reveals that in the Forcite and MBD* crystals, the carboxyl groups of adjacent (nonbonded) dimer pairs are well-aligned with and slightly nonparallel to the c-length (see Fig. 7), whereas those of D3 and D4 lie at a more acute angle to the c-length. In the D3 and D4 structures, adjacent molecules are rotated by about 120°, whereas in Forcite, they are rotated by 75° and in MBD* about 90°. The rotations and alignments in the D3 and D4 structures have decreased the ß-angle and elongated the a-length, with a consequent increase in the magnitude of the SEDC energy. Hexanoic acid was also sensitive to the k-point spacing during relaxation under D3 and D4 (see D3* and D4* lattice energies in Fig. 3), whereas MBD* was not. Increasing the fineness of the k-point grid (from 0.07 to a maximum of 0.04 Å^{-1}) led to an increase in the magnitude of the D3* and D4* lattice energies by about 9 kJ/mol,

whereas for MBD*, the corresponding increase in lattice energy was less than 0.02 kJ/mol.

Amino acid series: Series 2

Table II gives the details of the lattice energies, lattice lengths, and angles for the pre-optimized CIF (labeled "Expt.") and all structures (except PW91) following relaxation using Forcite, MBD*, D3, and D4. These and further details of hydrogen bond lengths (as defined for Series 1) are shown in the graphs of Fig. 8. Images of the crystal structures obtained following relaxation under each method (except PW91) can be found in the supplementary material in Fig. S18. In the bulk crystal, the amino acid molecules are zwitterions, whereas in the gas-phase, they become neutral in the DFT geometry optimizations and are built as neutral molecules in the Forcite calculations (see the section titled "Zwitterion workflow" in the supplementary material for further details). In the following, the number of carbon atoms represents those along the linear chain of the amino acids and does not include the carbon atoms that form branches.

The experimental *lattice energies* increase monotonically from α-glycine (C = 1) to L-alanine (C = 2) and then decrease in magnitude for both L-valine (C = 3) and further again for L-isoleucine (C = 4). The Forcite lattice energies are the largest in magnitude with the most negative value calculated for α-glycine. The calculated lattice energies for MBD*, D3, and D4 are very similar and show an increase in magnitude with an increasing number of carbon atoms. The general trend for the nondispersion corrected calculations of PW91 is similar to the experimental trend although with much

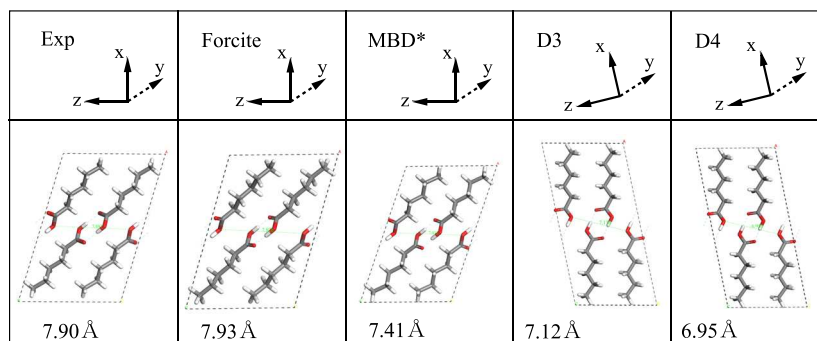FIG. 7. Hexanoic acid (C = 6) showing differences in orientation of the carboxyl groups with respect to the z-axis. The dotted line of the y-axis is to indicate that it is pointing into-the-page. The green dotted line is drawn between like-rotated molecules connecting their hydroxyl groups through the intermediate dimer pair. These distances are given in Angstroms.

TABLE II. Lattice energies and structures of the amino acids. a, b, and c are lattice lengths (per row); v is the unit cell volume (per row); and α, β, and γ are the lattice angles (per row). The experimental lattice energy values are calculated from enthalpies of sublimation using the approximation: $\Delta H_{sub} = -lattice\ energy - 2RT$, where R is the gas constant and T is the experimental temperature. All experimental and calculated lattice energies were negative. MAE: mean absolute error of lattice energies, volumes, and β-angles.

Structure	Lattice energy magnitude (kJ/mol)					Lattice lengths (Å)					Lattice angles (deg)				
	Expt.	Forc	MBD*	D3	D4		Expt.	Forc	MBD*	D3	D4	Expt.	Forc	MBD*	D3	D4
α-glycine	143.16	511.16	151.20	154.24	150.89	a:	5.09	4.50	5.08	5.15	5.14	α: 90.00	90.00	90.00	90.00	90.00
						b:	11.77	12.93	11.86	11.98	11.90	β: 111.99	114.59	111.42	110.99	110.77
						c:	5.46	5.37	5.44	5.46	5.45	γ: 90.00	90.00	90.00	90.00	90.00
						v:	303.2	283.8	305.2	314.3	311.8					
L-alanine	149.76	454.22	153.97	155.89	152.74	a:	5.93	6.01	5.89	6.67	6.10	α: 90.00	90.00	90.00	90.00	90.00
						b:	12.26	12.58	12.18	11.46	12.02	β: 90.00	90.00	90.00	90.00	90.00
						c:	5.79	5.15	5.82	5.83	5.83	γ: 90.00	90.00	90.00	90.00	90.00
						v:	421.1	389.7	416.8	445.2	427.4					
L-valine	143.86	489.63	160.67	163.58	157.66	a:	9.67	9.56	9.66	9.67	9.67	α: 90.00	90.00	90.00	90.00	90.00
						b:	5.27	4.86	5.17	5.27	5.24	β: 90.80	99.16	90.47	91.30	91.16
						c:	12.06	11.83	11.54	11.88	11.74	γ: 90.00	90.00	90.00	90.00	90.00
						v:	615.2	544.3	576.6	605.2	594.7					
L-isoleucine	127.67	484.83	163.51	165.78	161.80	a:	9.68	9.55	9.58	9.68	9.67	α: 90.00	90.00	90.00	90.00	90.00
						b:	5.30	4.93	5.23	5.30	5.27	β: 96.16	87.88	93.02	95.59	95.74
						c:	13.96	13.98	13.45	14.02	13.88	γ: 90.00	90.00	90.00	90.00	90.00
						v:	712.1	658.0	673.1	716.04	703.44					
MAE	. . .	343.85	16.23	18.76	14.66	v:	. . .	43.95	20.98	12.29	11.02	β: . . .	4.81	1.01	0.52	0.50

smaller (in magnitude) lattice energies. Consequently, we do not include any further analysis of the PW91-relaxed structures. Of this group, the trend in magnitude of the lattice energies is D3 > MBD* > D4, the D3-MBD*-D4 variation is from between 4 to approximately 7 kJ/mol, and they are larger than the experimental values by between 2% to 30%. The MAE of the lattice energies decreases from Forcite > D3 > MBD* > D4 within the range 14.66–18.76 kJ/mol for D3, MBD*, and D4.

The *lattice volumes* of the D3, D4, MBD*, and experimental structures agree to within approximately 1%–6%, whereas the Forcite volumes are consistently smaller than the pre-optimized volumes by between approximately 7% and 12%. All volumes increase monotonically with an increasing number of carbon atoms, and the MAE of the lattice volumes decreases from Forcite > MBD* > D3 > D4 within the range 43.95–11.02 Å³.

The *beta angles* of D3, D4, and MBD* agree with the experimental beta angles to within approximately 3%, as do those of Forcite for α-glycine and L-alanine. For L-valine and L-isoleucine, the Forcite beta angles are larger and smaller, respectively, than the experimental values by approximately 10%. The MAE of the beta-angles decreases from Forcite > MBD* > D3 > D4 within the range 4.81°–0.50°.

The *hydrogen bond lengths* are longest in the pre-optimized structures (as expected and discussed previously with reference to Series 1) and shortest in the Forcite structures. D3 and D4 produce very similar hydrogen bond lengths as exemplified by the trends seen in Fig. 8 and the average values in Table III. In the MBD* structures, the average minimum H-bond is shorter than D3/D4, and the average maximum H-bond longer, with the consequence that the range of H-bond lengths seen in the MBD* structures is wider than those seen in D3 and D4. The average H-bond length increases Forcite < D3/D4 < MBD*.

The *dispersion energy* contributions (Fig. 9) for D3, D4, and MBD* show similar trends in both the bulk crystal and gas-phase, and there is near-perfect agreement between the SEDC energies of D4 and MBD* in the bulk crystal and agreement to within approximately 10 kJ/mol in the gas-phase. D3 calculates the smallest SEDC energies in both bulk and gas-phase and as the amino acids increase in size, the D3 SEDC energy increasingly diverges from that of D4 and MBD*. The percentage SEDC contributions to the overall total energy (Fig. 9, middle row, left) show a widening gap between the D3 contributions and those of MBD* and D4 for both the gas-phase molecules and the crystal. Interestingly, the percentage D3 contribution decreases slightly rather than increasing for the crystal of L-alanine, which could be due to the slightly larger distances (than those of the D4 structures) between the methyl groups of the amino acids in the D3 L-alanine structure.

The absolute and percentage SEDC contributions to the total lattice energies (Fig. 9, middle row right) show near-identical trends between the three DFT methods, with contributions to the lattice
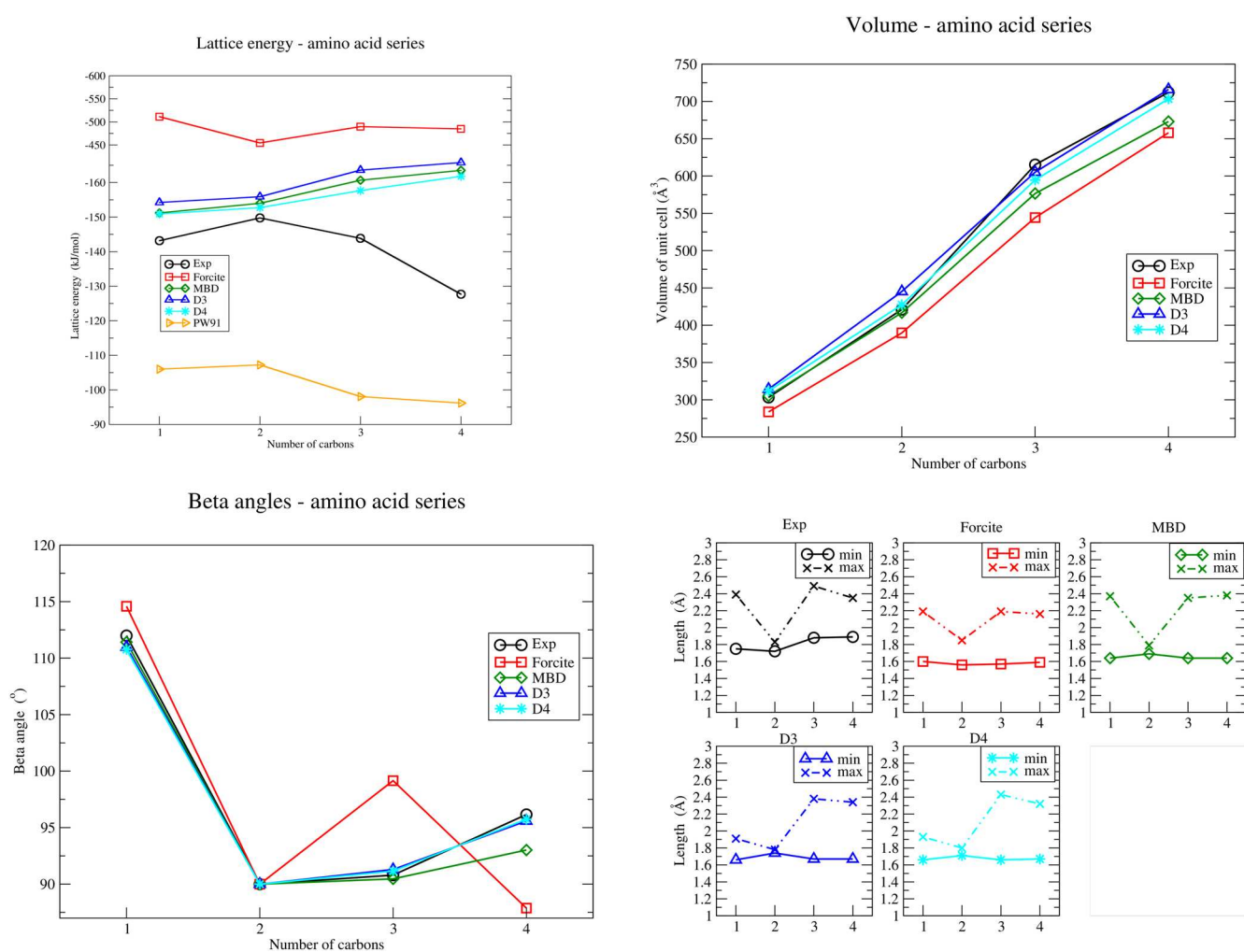
FIG. 8. Analysis of amino acid series, exploring lattice energies, volumes, beta angles, and maximum and minimum lengths of hydrogen bonds (bottom right).

energy increasing with the number of carbon atoms, and the percentage contribution increases from about 35% to 50% across Series 2. The synthon analysis (Fig. 9, bottom row) shows that under MBD* and D4, the molecules (beyond C = 1) move closer together, under D3, they move slightly apart (including C = 1), and under Forcite, the molecules move closer together and closer than is seen under the DFT methods, except for C = 2, where under both Forcite and D3, the molecules move further apart by approximately 2%. Forcite shows the largest variation in changes in van der Waals distances (−15% to +2%) and MBD*, D3 and D4 produce changes in distances between −2 and +2%.

The preceding analysis shows that Forcite has not performed as well as the DFT methods in the energetics and (for two of the crystals) the structure. Examining the output files of the Forcite calculations carried out within Materials Studio (not shown here) reveals relative positive, nonbonded (i.e., van der Waals and electrostatic) intramolecular energies of the gas-phase (neutral) molecules.

In the crystal, the nonbonded energies (the sum of intra- and intermolecular) are positive for α-glycine and L-alanine and negative for L-valine and L-isoleucine although except for α-glycine, they are proportionately smaller in magnitude than the nonbonded energies found in the gas-phase. In α-glycine, the crystal contains pairs of hydrogen-bonded amino acids that are "stacked" along the b-length

TABLE III. Average maximum and average minimum lengths of hydrogen bonds found across the whole of the amino acid series (each structure contains a different number of various lengths of hydrogen bonds).

Method	Expt.	Forcite	MBD*	D3	D4
Minimum (Å)	1.81	1.58	1.65	1.69	1.68
Maximum (Å)	2.27	2.10	2.22	2.10	2.12
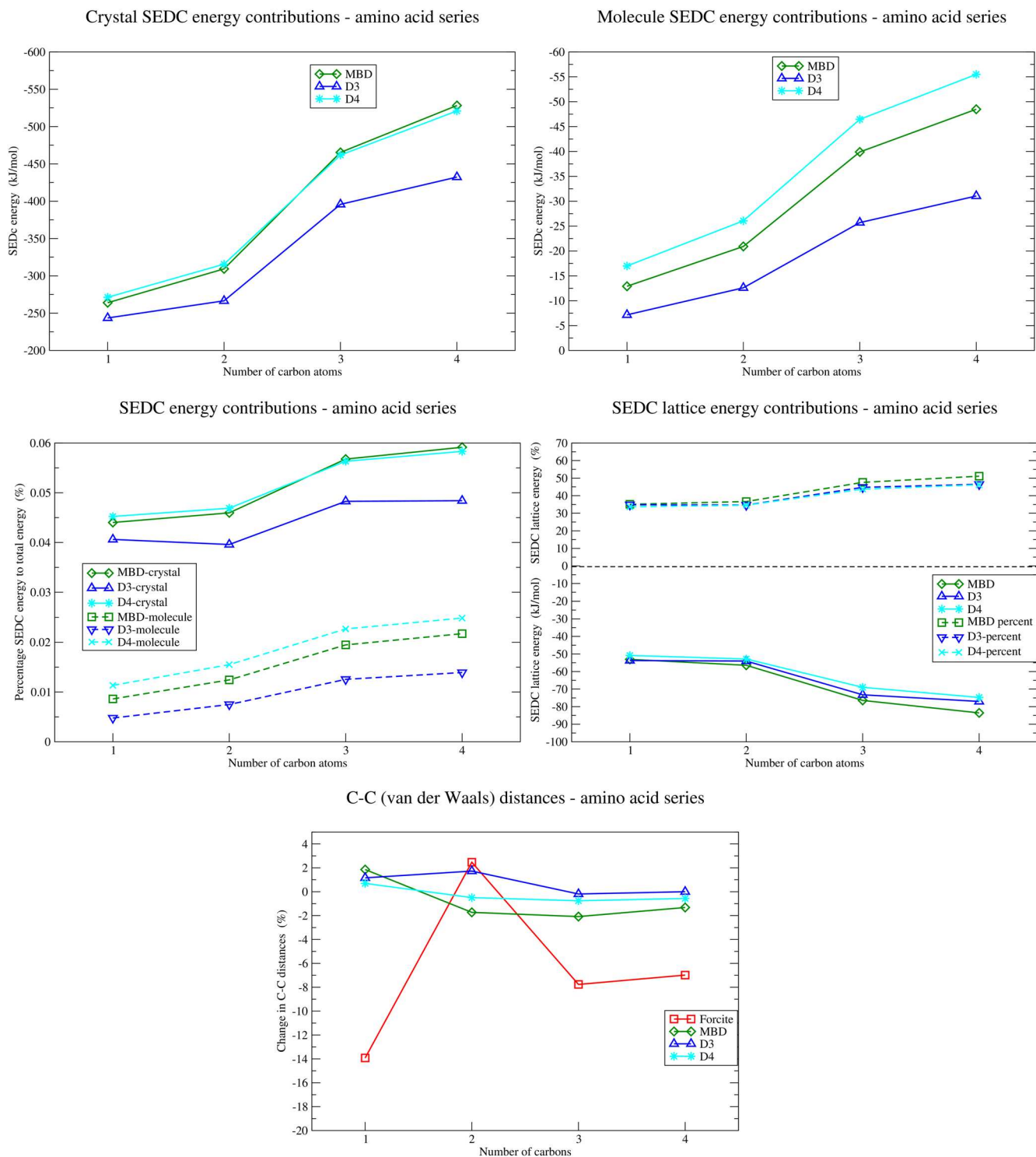
FIG. 9. Top row, semiempirical dispersion correction energies (SEDC) of the three DFT-DISP methods for the amino acid series. Middle row left, the percentage of the SEDC energy contributions to the overall total energy; middle row right, the absolute and percentage contributions of the SEDC energy to the total lattice energy. Bottom row, average change in distances after optimization between centers of the molecules (i.e., van der Waals distances) relative to the pre-optimized structures.

and weakly bound by van der Waals forces, which would explain the positive, nonbonded energy found in this crystal. In L-valine and L-isoleucine, the molecules are hydrogen-bonded in bimolecular layers, leaving the main, nonbonded interactions to be weak van der Waals between opposing hydrocarbon backbones (within the bilayers) and methyl groups (between opposing bilayers). In the smaller L-alanine crystal, the molecules are hydrogen-bonded throughout the crystal with the nonbonded interactions confined to staggered methyl-methyl interactions. The changes in the torsion angles between the gas-phase $NH_2\cdots COH$ and the bulk, crystal phase $NH_3^+\cdots COO$-groups are larger in all of the Forcite models (see Table S2) than in any of the DFT models, which means that there is greater molecular distortion in the Forcite crystals. This appears to be due to overshortening of the hydrogen bonds and is particularly visible in L-valine and L-isoleucine to the extent that the foreshortening has drawn the molecules within the crystal closer together, thereby affecting the beta angles of the two crystals. In general, modeling hydrogen bond-dominated systems might be improved by using a hybrid exchange-correlation functional (plus MBD*) such as PBE0,[87] where in similar studies, it outperformed PBE (plus MBD*).[45,56] Unfortunately, PBE0 is computationally more expensive than PBE and therefore unsuitable for inclusion in our exploration of trends across increasingly large systems.

SUMMARY

By including an exchange-correlation functional without any dispersion corrections (i.e., PW91-GGA), the results show that across the two series—the van der Waals-dominated carboxylic acids and the polarized- and hydrogen bond-dominated amino acids—dispersion corrections are essential to reproduce experimental trends in lattice energies. Of the dispersion corrections, MBD@rsSCS has performed the most reliably producing consistent trend lines in lattice energetics and structure. Grimme D3(0) and D3(BJ) performed similarly well with D3(BJ) returning the smallest MAE for lattice energies (8.52 and 14.66 kJ/mol for Series 1 and 2, respectively) although this is excluding hexanoic acid from the energetics, which is sensitive to the number of sampling points under D3(0) and D3(BJ). The many-body nature of the MBD@rsSCS dispersion interactions has kept the structure and energetics of hexanoic acid "on-trend" during optimization. The smallest lattice volume MAEs were produced by D3(0) (18.77 $Å^3$) for Series 1 and D3(BJ) (11.02 $Å^3$) for Series 2. The smallest MAEs in the beta-angles were given by Forcite (1.33°) for Series 1 and D3(BJ) (0.5°) for Series 2. The average length of H-bonds shows the shortest calculated by MBD@rsSCS and the longest by Forcite for Series 1, and this order is reversed for Series 2. D3(BJ) was developed to remove the artificial repulsion originating from the damping function of D3(0), and herein, its overall slightly better performance than D3(0) encourages its recommendation as the preferred DFT-D3 functional, in agreement with the review of Grimme et al.[88] However, within the dispersion corrected DFT methods, MBD@rsSCS captures more of the different chemical environments than does D3(0) and D3(BJ), which is attributed to the inclusion of the many-body terms in MBD@rsSCS.

The performance of MBD@rsSCS competes favorably with the COMPASS II force field method used in Forcite, in the calculated energetics as well as the structural features and outperforms it for the polarized- and hydrogen bond-dominated series of amino acids. The MBD@rsSCS performs sufficiently consistently throughout Series 1 that the addition of a size-independent, constant shift in the lattice energy would produce an almost exact match with the experimental lattice energies for these types of organic molecular crystals.

CONCLUSION

In terms of functionality, the latest solid state DFT-DISP methods compete well with classical methods although computationally they are more demanding of hardware and computational time. In particular, the many body dispersion term, MBD@rsSCS, has almost consistently reproduced experimental energetic and structural trends across two homologous series separately dominated by (1) van der Waals and (2) polar and hydrogen bond interactions. The pairwise DFT-DISP methods [D3(0) and D3(BJ)] perform similarly well across both series with a couple of energetic and structural exceptions that undermine their reliability, and supervision is recommended when performing batches of calculations. The classical method employing the COMPASS II force field reproduces the experimental and structural trends well for the van der Waals dominated carboxylic acids but not for the polar and hydrogen bond dominated amino acids.

The advantage of probing two series rather than a set of diverse structures is that physically explicable trends can be identified, as exemplified by the predicted and simulated increasing contribution of the dispersion energy to the lattice energy, with increasing number of carbon atoms in the backbone chain. In turn, deviations from those trends allowed anomalies, such as those produced by the pairwise methods (for hexanoic acid) to be identified, which could then be explored further to determine where the sources of error lie (currently beyond the scope of this work).

Depending on the output required—e.g., structure, energetics, and general trends—all three "off-the-shelf" DFT methods are sufficiently "on-trend" to be used as alternatives or complementary to classical, force field methods. We intend to test their functionality further with our next goal being to use the above methods to calculate the lattice energies and elastic constants of a diverse set of organic molecular crystals.

SUPPLEMENTARY MATERIAL

Supplementary material includes the following: a ZIP file containing the Python scripts, an Excel file with experimental sublimation data, CSV files with the experimental lattice energies, and a document capturing the information referred to throughout this paper as "supplementary material."

REFERENCES

[1]C. Sun, J. Pharm. Sci. **98**, 1744 (2009).

[2]J. L. McDonagh, D. S. Palmer, T. van Mourik, and J. B. O. Mitchell, J. Chem. Inf. Model. **56**, 2162 (2016).

[3]R. Docherty, K. Pencheva, and Y. A. Abramov, J. Pharm. Pharmacol. **67**, 847 (2015).

[4]R. E. Skyner, J. L. McDonagh, C. R. Groom, T. van Mourik, and J. B. O. Mitchell, Phys. Chem. Chem. Phys. **17**, 6174 (2015).

[5]R. L. M. Robinson, K. J. Roberts, and E. B. Martin, J. Cheminf. **10**, 44 (2018).

[6]S. L. Price, Faraday Discuss. **211**, 9 (2018).

[7]C. Červinka and M. Fulem, J. Chem. Theory Comput. **13**, 2840 (2017).

[8]J. G. Brandenburg and S. Grimme, Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. **72**, 502 (2016).

[9]T. T. H. Nguyen, I. Rosbottom, I. Marziano, R. B. Hammond, and K. J. Roberts, Cryst. Growth Des. **17**, 3088 (2017).

[10]I. Rosbottom, K. J. Roberts, and R. Docherty, CrystEngComm **17**, 5768 (2015).

[11]*Engineering Crystallography: From Molecule to Crystal to Functional Form*, edited by K. J. Roberts, R. Docherty, and R. Tamura (Springer, The Netherlands, 2017), pp. 155–176.

[12]S. L. Price, Chem. Soc. Rev. **43**, 2098 (2014).

[13]G. M. Day, Crystallogr. Rev. **17**, 3 (2011).

[14]A. Pertsin and A. I. Kitaigorodsky, *The Atom-Atom Potential Method: Applications to Organic Molecular Solids* (Springer-Verlag, Berlin, Heidelberg, 1987).

[15]D. E. Williams, J. Mol. Struct. **485-486**, 321 (1999).

[16]D. E. Williams, J. Chem. Phys. **45**, 3770 (1966).

[17]A. Warshel and S. Lifson, J. Chem. Phys. **53**, 582 (1970).

[18]A. Gavezzotti, *Molecular Aggregation: Structure Analysis and Molecular Simulation of Crystals and Liquids* (Oxford University Press, Oxford, New York, 2013).

[19]J. C. Osborn and P. York, J. Mol. Struct. **474**, 43 (1999).

[20]D. E. Williams, Science **147**, 605 (1965).

[21]J. Yang, W. Hu, D. Usvyat, D. Matthews, M. Schütz, and G. K.-L. Chan, Science **345**, 640 (2014).

[22]J. Hoja, A. Reilly, and A. Tkatchenko, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **7**, e1294 (2017).

[23]S. Wen and G. J. O. Beran, J. Chem. Theory Comput. **7**, 3733 (2011).

[24]T. Teuteberg, M. Eckhoff, and R. A. Mata, J. Chem. Phys. **150**, 154118 (2019).

[25]A. Otera-de-la-Roza and E. R. Johnson, J. Chem. Phys. **137**, 054103-1 (2012).

[26]J. S. Chickos, S. Hosseini, D. G. Hesse, and J. F. Liebman, Struct. Chem. **4**, 271 (1993).

[27]A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein, Chem. Soc. Rev. **44**, 8619 (2015).

[28]A. Stone, *The Theory of Intermolecular Forces*, 2nd ed. (Oxford Press, Oxford, 2013).

[29]A. R. Leach, *Molecular Modelling: Principles and Applications* (Addison Wesley Longman Limited, 1996), pp. 196–204.

[30]J. Nyman, O. S. Pundyke, and G. M. Day, Phys. Chem. Chem. Phys. **18**, 15828 (2016).

[31]J. Hepburn, G. Scoles, and R. Penco, Chem. Phys. Lett. **36**, 451 (1975).

[32]J. Klimeš and A. Michaelides, J. Chem. Phys. **137**, 120901 (2012).

[33]J. P. Perdew and K. Schmidt, AIP Conf. Proc. **577**, 1 (2001).

[34]S. Grimme, J. Comput. Chem. **25**, 1463 (2004).

[35]Q. Wu and W. T. Yang, J. Chem. Phys. **116**, 515 (2002).

[36]G. F. Thomas and W. J. Meath, Mol. Phys. **34**, 113 (1977).

[37]S. Grimme, J. Comput. Chem. **27**, 1787 (2006).

[38]S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).

[39]E. Caldeweyher, C. Bannwarth, and S. Grimme, J. Chem. Phys. **147**, 034112 (2017).

[40]Y. Zhang and W. Yang, Phys. Rev. Lett. **80**, 890 (1998).

[41]O. A. Vydrov and T. Van Voorhis, J. Chem. Phys. **130**, 104105 (2009).

[42]O. A. Vydrov and T. Van Voorhis, Phys. Rev. Lett. **103**, 063004 (2009).

[43]G.-X. Zhang, A. Tkatchenko, J. Paier, H. Appel, and M. Scheffler, Phys. Rev. Lett. **107**, 245501 (2011).

[44]N. Marom, R. A. DiStasio, V. Atalla, S. Levchenko, A. M. Reilly, J. R. Chelikowsky, L. Leiserowitz, and A. Tkatchenko, Angew. Chem., Int. Ed. **52**, 6629 (2013).

[45]A. M. Reilly and A. Tkatchenko, J. Phys. Chem. Lett. **4**, 1028 (2013).

[46]B. M. Axilrod and E. Teller, J. Chem. Phys. **11**, 299 (1943).

[47]A. Tkatchenko, D. Alfè, and K. S. Kim, J. Chem. Theory Comput. **8**, 4317 (2012).

[48]A. Ambrosetti, D. Alfè, R. A. DiStasio, and A. Tkatchenko, J. Phys. Chem. Lett. **5**, 849 (2014).

[49]A. Ambrosetti and P. L. Silvestrelli, J. Phys. Chem. Lett. **10**, 2044 (2019).

[50]A. Ambrosetti and P. L. Silvestrelli, Carbon **139**, 486 (2018).

[51]P. R. Tulip, S. P. Bates, and S. J. Clark, J. Chem. Phys. **137**, 024701 (2012).

[52]A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler, Phys. Rev. Lett. **108**, 236402 (2012).

[53]R. A. DiStasio, O. A. von Lilienfeld, and A. Tkatchenko, Proc. Natl. Acad. Sci. U. S. A. **109**, 14791 (2012).

[54]A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).

[55]M. W. Cole, D. Velegol, H.-Y. Kim, and A. A. Lucas, Mol. Simul. **35**, 849 (2009).

[56]A. Ambrosetti, A. M. Reilly, R. A. DiStasio, Jr., and A. Tkatchenko, J. Chem. Phys. **140**, 18A508 (2014).

[57]A. M. Reilly and A. Tkatchenko, J. Chem. Phys. **139**, 024705 (2013).

[58]H. P. G. Thompson and G. M. Day, Chem. Sci. **5**, 3173 (2014).

[59]H. Sun, Z. Jin, C. Yang, R. L. C. Akkermans, S. H. Robertson, N. A. Spenley, S. Miller, and S. M. Todd, J. Mol. Model. **22**, 47 (2016).

[60]H. Sun, J. Phys. Chem. B **102**, 7338 (1998).

[61]H. Sun, S. J. Mumby, J. R. Maple, and A. T. Hagler, J. Am. Chem. Soc. **116**, 2978 (1994).

[62]Dassault Systèmes, Dassault Systèmes BIOVIA, Materials Studio v2016, Dassault Systèmes 2016, San Diego, 2016.

[63]S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, Z. Kristallogr.: Cryst. Mater. **220**, 567 (2005).

[64]S. Grimme, S. Ehrlich, and L. Goerigk, J. Comput. Chem. **32**, 1456 (2011).

[65]G. Clydesdale, R. Docherty, and K. J. Roberts, Comput. Phys. Commun. **64**, 311 (1991).

[66]K. J. Roberts, R. B. Hammond, V. Ramachandran, and R. Docherty, *Computational Approaches in Pharmaceutical Solid State Chemistry* (Wiley, NJ, 2016).

[67]J. Pickering, R. B. Hammond, V. Ramachandran, M. Soufian, and K. J. Roberts, *Engineering Crystallography: From Molecule to Crystal to Functional Form* (Springer, Dordrecht, The Netherlands, 2017), pp. 155–176.

[68]C. R. Groom, I. J. Bruno, M. P. Lightfoot, and S. C. Ward, Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. **B72**, 171–179 (2016).

[69]P. P. Ewald, Ann. Phys. **369**, 253 (1921).

[70]M. P. Tosi, *Solid State Physics* (Academic Press, 1964), pp. 1–120.

[71]J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[72]E. R. McNellis, J. Meyer, and K. Reuter, Phys. Rev. B **80**, 205414 (2009).

[73] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).

[74] H. Monkhorst and J. Pack, Phys. Rev. B **13**, 5188 (1976).

[75] B. G. Pfrommer, M. Côté, S. Louie, and M. L. Cohen, J. Comput. Phys. **131**, 233 (1997).

[76] S. N. Ngauv, R. Sabbah, and M. Laffitie, Thermochim. Acta **20**, 371 (1977).

[77] H. J. Svec and D. D. Clyde, J. Chem. Eng. Data **10**, 151 (1965).

[78] O. V. Dorofeeva and O. N. Ryzhova, J. Phys. Chem. A **118**, 3490 (2014).

[79] C. D. Cappa, E. R. Lovejoy, and A. R. Ravishankara, J. Phys. Chem. A **112**, 3959 (2008).

[80] D. P. Baccanari, J. A. Novinski, Y.-C. Pan, M. M. Yevitz, and H. A. Swain, Trans. Faraday Soc. **64**, 1201 (1968).

[81] S. Lifson, A. T. Hagler, and P. Dauber, J. Am. Chem. Soc. **101**, 5111 (1979).

[82] J. D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds* (Academic Press, Inc., London, 1970).

[83] J. Nyman and G. M. Day, CrystEngComm **17**, 5154 (2015).

[84] G. Filippini and A. Gavezzotti, Acta Crystallogr., Sect. B: Struct. Sci. **49**, 868 (1993).

[85] M. H. Charlton, R. Docherty, and M. G. Hutchings, J. Chem. Soc., Perkin Trans. 2 **2**, 2023 (1995).

[86] L. Maschio, B. Civalleri, P. Ugliengo, and A. Gavezzotti, J. Phys. Chem. A **115**, 11179 (2011).

[87] J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).

[88] S. Grimme, A. Hansen, J. G. Brandenburg, and C. Bannwarth, Chem. Rev. **116**, 5105 (2016).