JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

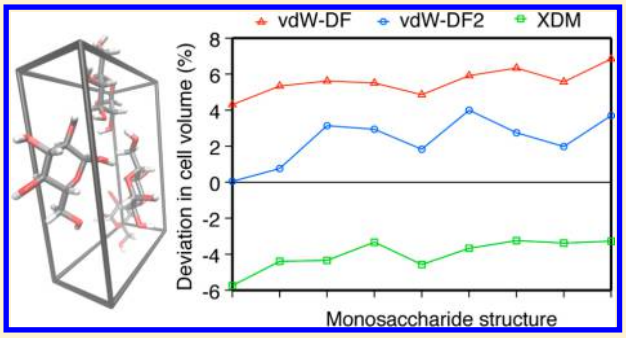
Benchmarking Calculated Lattice Parameters and Energies of Molecular Crystals Using van der Waals Density Functionals

Damien J. Carter* and Andrew L. Rohl

Nanochemistry Research Institute and Department of Chemistry, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia

Supporting Information

ABSTRACT: The development of new functionals and methods to accurately describe van der Waals forces in density functional theory (DFT) has become popular in recent years, with the vast majority of studies assessing the accuracy of the energetics of collections of molecules, and to a lesser extent molecular crystalline systems. As the energies are a function of the atom positions, we assess the accuracy of DFT calculations from both a geometric and energetics point of view for the C21 reference data set of Otero-de-la-Roza and Johnson for molecular crystals, and a set of monosaccharide molecular crystals. In particular, we examine the performance of exchange-correlation functionals designed to handle van der Waals forces, including the vdW-DF, vdW-DF2, and XDM methods. We also assess the effect of using small and large basis sets, the choice of basis functions (local atomic orbitals using the SIESTA code versus planewaves using the Quantum ESPRESSO code), and the effect of corrections for basis set superposition errors. Finally, we examine the geometries and energies of the S22 reference set of molecular complexes. Overall, the most accurate geometries for both choices of basis functions are obtained with the vdW-DF2 functional, while the most accurate lattice energies are obtained using vdW-DF2 with local atomic orbitals and XDM with planewaves with mean absolute errors of less than 4 kJ/mol.



INTRODUCTION

The absence of van der Waals forces in standard density functional theory (DFT) calculations is well-known, which restricted the use of DFT for soft matter, biomolecular or molecular crystalline systems for many years. In particular, van der Waals forces in molecular crystals are extremely important from both an energetic and geometric perspective and influence a range of aspects including crystal packing and polymorphism.

In recent years, many methods have been developed to provide ever-increasingly more accurate descriptions of the dispersion forces. Grimme[1] proposed adding an empirical correction using interatomic potentials (commonly referred to as DFT-D) of the form C_6R^{-6} (where R is the pairwise atom cutoff distance and C_6 is the dispersion coefficient), which lead to the many exchange-correlation functionals[1−5] incorporating this correction. Grimme et al.[6] later published a refined method for computing coefficients and cutoff radii, referred to as DFT-D3. Another approach has been to include medium-range dispersion forces in conventional semilocal DFT, by using hybrid meta-GGA methods such as X3LYP,[7] ωB97X-D,[8] M06,[9] and PW6B95.[10]

Methods to include van der Waals forces have also been developed that incorporate correlation components from wave function theory such as XYG3[11] and B2PLYP[12] functionals. A drawback of the simplistic C_6R^{-6} approach is that that the coefficients do depend on the molecular environment and the partial charge of an atom,[13] which is particularly important when examining solids. Several methods[14,15] have been developed that deal with this problem. An alternative approach involves calculating dispersion coefficients based on the exchange-hole dipole method (XDM).[16−18] In this approach, dispersion interactions are modeled by examining the instantaneous dipole that arises between an electron and its exchange hole. Other functionals have been developed that explicitly include nonlocal correlation, for example, the various implementations of vdW-DF,[19−21] vdW-DF2,[22] VV09,[23] and VV10[24] functionals.

Assessments of the accuracy of DFT methods for van der Waals forces have typically been carried out on molecular systems, where the energetics are calculated for various molecular data sets containing dimers or complexes. The most common data set is arguably the S22 set of complexes,[25] where there have been a vast array of computational studies comparing the accuracy of particular DFT methods with the binding energies in this data set.[21,22,26−31] There are also many other data sets of molecular compounds such as S66,[32] JSCH-2005,[33] and GMTKN30,[34,35] or from the work of Kannemann and Becke.[29] The reference energies in many of these systems are calculated using accurate quantum methods based on the

Received: April 18, 2014
Published: June 20, 2014

   dx.doi.org/10.1021/ct500335b | J. Chem. Theory Comput. 2014, 10, 3423−3437

perturbatively corrected coupled cluster CCSD(T) technique;[36] however, the computational effort scales by a formal cost of $O(N^7)$ (where N is the system size), which greatly limits its applicability, from both energy and optimized geometry perspectives, in molecular crystal applications.

van der Waals corrected DFT calculations of molecular crystals have focused on systems ranging from amino acids,[37] calixarene inclusion compounds,[28] metal organic framework materials,[38,39] zeolites,[40] and many others.[5,41−52] These studies have reported a range of properties from various geometric parameters to cohesive energies or lattice energies. A range of different methods have been used to describe the van der Waals forces in these systems including various flavours of vdW-DF,[28,37−39,41,46] vdW-DF2,[41,46] XDM,[46,47] and empirical DFT-D.[51,52,2,43−47] Other schemes that have shown promising results include methods that have incorporated many-body interactions coupled with either second-order Møller−Plesset (MP2) perturbation theory[44] or with standard or hybrid DFT functionals,[52,48,49] while another approach uses Gaussian and/or planewave approaches coupled with MP2.[40,42,50,53]

Van de Streek and Neumann[54] used DFT-D to examine a large collection of molecular crystals published experimentally during 2008, to examine the correctness of the proposed experimental structures, in particular relating to the location of protons. They highlighted how well DFT-D methods can reproduce experimental crystal structures and show how calculations combined with experimental data can help solve ambiguities or incorrect structures. Neumann and Perrin[45] used DFT-D to examine a collection of 31 small molecular crystals, highlighting how much smaller cell deformations (3.2%) are observed when compared to results from standard DFT methods (21.5%). B3LYP-D has also shown the importance of including dispersion corrections for superior performance over standard DFT methods[2,43] and also highlight[2] the importance of the dispersion coefficients and scaling factors when using an empirical C_6R^{-6} approach. In our recent investigation[28] of crystals of *p-tert*-butylcalix[4]arene containing CS_2 or toluene molecules, we highlighted the superior performance of vdW-DF to standard DFT from a geometric and binding energy perspective. Similarly, Sabatini et al.[37] examined crystal structures of glycine and alanine and also showed that vdW-DF performs much better than standard DFT for geometries and phase stability. Berland et al.[41] assessed the performance of three different implementations of vdW-DF, namely vdW-DF(revPBE_x),[20] vdW-DF(C09_x),[19] and vdW-DF(optPBE_x)[21] with vdW-DF2[22] for a small set of molecular crystals such as hexamine and dodecahedrane. They found vdW-DF2 outperforms the various vdW-DF implementations based on lattice parameters and cohesive energies of the systems they examined.

In summary, there is a wealth of reference data sets for molecular systems, and very limited reference sets for crystalline systems. Motivated by this, Otero-de-la-Roza and Johnson[46] published a benchmark for noncovalent interactions in solids (C21). The reference set contains a range of intermolecular interactions with reference values obtained from sublimation energies and X-ray crystal structures, with corrections for calculated thermal and zero-point energy effects, to allow DFT functionals to be assessed in a straightforward way. This detailed study largely focused on the energetics relating to lattice energies and found that XDM methods give the best results compared to other methods that such as DFT-D, vdW-DF, and vdW-DF2. Reilly and Tkatchenko[48] recently expanded this set of 21 molecular crystals to 23 and examined the performance of their many-body dispersion (MBD) method, which can be coupled with standard or hybrid DFT functionals. The X23 reference set has also been recently investigated using DFT-D3[51,52] methods.

In this work, we compare the accuracy, from both a geometric and an energetics perspective, of van der Waals DFT functionals including vdW-DF, vdW-DF2, and XDM for molecular crystals. This study is motivated by the need to find van der Waals DFT methods that accurately predict both geometries and energies of large molecular crystalline systems such as calixarenes, or for use in crystal structure prediction calculations to locate stable polymorphs. We compare the effect of basis sets quality (double-ζ plus polarization quality versus a larger triple-ζ plus polarization quality), and the choice of basis function, by comparing approaches using numerical atomic orbitals and planewave representations. Next, we extend our investigations to examine a set of monosaccharide molecular crystals. Sameera and Pantazis[55] recently reported a study of the performance of various van der Waals DFT methods for calculating the geometries and energetics of monosaccharide gas phase isomers, so a logical extension of this work was extend this to monosaccharide molecular crystals. Finally, we revisit the S22 reference set of complexes to assess whether or not the conclusions drawn for crystalline systems also apply to molecular systems.

METHODOLOGY

DFT calculations were performed using the SIESTA[56] and Quantum ESPRESSO[57] codes. In SIESTA, Troullier and Martins[58] norm-conserving pseudopotentials were used. A small basis set, referred to as double-ζ polarized (DZP), was generated using numerical atomic orbitals that were radially confined to an extent that induces an energy shift in each orbital of 0.001 Ry. A large optimized basis set, referred to as triple-ζ polarized (TZP), was generated based on the work of Louwerse and Rothenberg.[59] As they only published optimized basis sets for N, O, and H atoms, we generated an optimized basis set for C atoms, based on fitting to structures from the S22 reference set of Jurecka et al.[25] The exact cutoff radii for the C basis set are listed in the Supporting Information. Hartree and exchange correlation energies were evaluated on a uniform real-space grid of points with a defined maximum kinetic energy of 300 Ry. In Quantum ESPRESSO, we employ the projector-augmented wave (PAW) method[60] using a planewave basis set, with pseudopotentials adapted from the atompaw library.[61] Small (S) and large (L) basis sets were generated using planewave kinetic energy cutoffs of 65 and 80 Ry, respectively.

DFT calculations were carried out using the PBE,[62] vdW-DF,[20] and vdW-DF2[22] exchange-correlation functionals with SIESTA. Quantum ESPRESSO calculations utilized the vdW-DF and vdW-DF2 functionals, and the exchange-hole dipole moment (XDM) method.[47,63] In particular, we used the B86b-XDM functional, which combines B86b[64] exchange with PBE[62] correlation, where the XDM parameters were A1 = 0.337 and A2 = 2.488. Atomic relaxations were performed in SIESTA using a conjugate gradients method with a maximum tolerance of 0.01 eV/Å, while calculations in Quantum ESPRESSO use a BFGS quasi-newton method with a maximum tolerance of 0.001 Ry/Bohr. To minimize Pulay stresses in Quantum ESPRESSO calculations, geometry optimizations were repeatedly restarted until the change in volume between the initial and final structures in an optimization was less than 0.01%.

Table 1. Deviations (%) of the Cell Volumes for the C21 Reference Data Set Using SIESTA and Quantum ESPRESSO, as a Function of Basis Set Size and vdW Functional[a]

	SIESTA				Quantum ESPRESSO						
	vdW-DF		vdW-DF2		vdW-DF		vdW-DF2		XDM		
	DZP	TZP	DZP	TZP	S	L	S	L	S	L	COMPASS
14-cyclohexanedione	2.96	7.79	−0.39	3.23	4.49	8.25	−2.24	0.74	−5.02	−1.40	0.17
acetic acid	3.30	8.80	−0.43	2.99	5.10	9.13	−1.07	2.05	−3.46	0.03	−3.46
adamantane	2.56	7.66	−0.06	2.50	2.88	7.18	−3.74	−0.53	−8.35	−4.32	−3.62
ammonia	3.46	5.16	−2.24	−1.96	6.21	7.96	−2.92	0.08	−7.04	−6.82	−4.55
anthracene	6.34	9.15	3.88	6.15	6.36	10.87	0.68	3.70	−5.83	−1.76	−1.61
benzene	4.74	8.90	0.86	4.18	5.10	10.13	−1.93	1.58	−7.83	−3.86	−2.58
carbon dioxide	−1.73	9.08	−6.20	2.47	5.34	9.60	−4.79	−1.40	−0.23	4.68	−6.90
cyanamide	4.57	9.21	2.24	5.37	7.25	9.72	1.14	3.09	−2.21	−0.02	2.76
cytosine	3.71	6.89	2.73	4.67	5.17	8.09	0.51	2.87	−4.45	−1.83	0.47
ethylcarbamate	−0.21	3.57	−3.31	−0.41	1.25	4.99	−4.67	−1.70	−6.33	−2.85	−8.38
formamide	3.70	8.39	0.25	3.71	5.78	9.43	−1.18	1.95	−3.48	−0.23	−5.54
imidazole	3.52	7.50	1.73	4.42	5.25	9.36	−0.92	2.63	−5.65	−2.11	−5.27
naphthalene	7.15	10.61	4.30	7.30	7.39	12.05	1.11	4.29	−5.34	−1.48	−0.65
α-oxalic acid	2.14	7.31	1.29	5.20	5.39	8.31	−0.04	2.67	−2.35	0.34	5.35
β-oxalic acid	1.55	6.04	−0.43	3.14	3.91	7.29	−1.89	0.91	−2.70	0.69	−0.70
pyrazine	3.04	6.24	0.03	2.56	3.87	7.88	−2.45	0.58	−7.36	−4.03	−8.23
pyrazole	6.94	11.22	4.14	7.47	7.73	12.02	1.59	4.40	−3.30	0.27	−3.53
triazine	−1.03	2.91	−4.28	−1.07	1.00	4.31	−5.35	−2.94	−8.73	−5.88	−11.03
trioxane	1.29	5.53	−2.33	−0.32	3.73	7.30	−3.13	−0.31	−4.25	−0.85	−4.20
uracil	2.12	5.50	0.76	3.37	3.84	6.88	−0.82	1.83	−4.49	−1.74	−2.41
urea	6.41	8.30	3.92	4.96	6.38	8.69	1.75	3.92	−2.49	−0.41	−1.64
vol %MD	3.17	7.42	0.31	3.33	4.92	8.54	−1.45	1.45	−4.80	−1.60	−3.12
vol %MAD	3.45	7.42	2.25	3.69	4.92	8.54	2.09	2.10	4.80	2.17	3.95
cell length MAD	0.11	0.18	0.08	0.11	0.12	0.19	0.07	0.07	0.12	0.08	0.36
cell angle MAD	0.23	0.14	0.38	0.30	0.62	0.82	0.85	0.80	0.54	0.64	1.86
cell length MA%D	1.40	2.50	1.16	1.51	1.73	2.72	0.96	0.90	1.70	1.06	5.10
cell angle MA%D	0.22	0.12	0.35	0.27	0.55	0.73	0.78	0.73	0.51	0.61	3.64

[a]For comparison, we also report the results for molecular mechanics calculations using the COMPASS force field. The mean, absolute, and percentage deviations of the cell volumes, lengths, and angles are also reported.

Molecular mechanics calculations were carried out using the Materials Studio software package.[65] Electrostatic interactions were computed with an Ewald summation to an accuracy of 1×10^{-5} kcal/mol and atomic charges were assigned by the force field. van der Waals interactions were evaluated using an atom-based summation with a cutoff of 18.5 Å. Structural optimization was performed with the COMPASS[66,67] force field within the Forcite module in Materials Studio.

The C21 reference data set was taken from Otero-de-la-Rosa and Johnson[46] and consists of 21 small molecular crystals (at temperatures ranging from 10 to 298 K) that span several types of intermolecular interactions (for example π-stacking, electrostatic or hydrogen-bonding dominated) and a range of interaction energies. We created a room temperature version of the reference C21 data set by removing the gases (ammonia and carbon dioxide) and using room temperature (278−298 K) crystal structures from the literature for the remaining molecular crystals. We have labeled this set $C19_{RT}$. The monosaccharide data set consists of 9 molecular crystals at temperatures ranging from 95 to 298 K. Specific details of the crystal structures, including the Cambridge Structural Database[68] (CSD) reference numbers and lattice parameters, of the room temperature $C19_{RT}$ data set, and the monosaccharide data set are described in the Supporting Information. The S22 reference data set was taken from Jurecka et al.[25] and consists of 22 complexes of common molecules dominated by a variety different intermolecular forces.

Localized orbital approaches (such as SIESTA) can suffer from the problem of basis set superposition errors (BSSE). To correct binding energies for BSSE in dimers and molecular species, the standard counterpoise correction (CP) method[69,70] was used. To correct lattice energies for BSSE, we developed a CP method suitable for molecular crystals. Here, we define $E_{bsse} = E_{mol+ghosts}^{supercell} - E_{mol}^{supercell}$, where $E_{mol+ghosts}^{supercell}$ is the energy of a supercell of the molecular crystal (using the fully relaxed lattice parameters and atomic coordinates) with all atoms described by ghost orbitals except those in a single molecule, and $E_{mol}^{supercell}$ is the energy of that single molecule of the molecular crystal within this same supercell. This procedure needs to be repeated for each symmetry unique molecule in the unit cell; for all structures studied here there was only one (i.e., Z′=1). To determine an appropriate supercell size, we measured the minimum separation distance that was required to keep the orbitals of a molecule and its periodic images from overlapping (approximately 9.5 Å for the largest optimized TZP basis set); we increased the supercell dimensions until this minimum criterion was met. The exact sizes of the supercells used for the CP corrections are listed in the Supporting Information. We have made an online data set available[71] that contains all input, output, and pseudopotential files from this work.

dx.doi.org/10.1021/ct500335b | J. Chem. Theory Comput. 2014, 10, 3423−3437

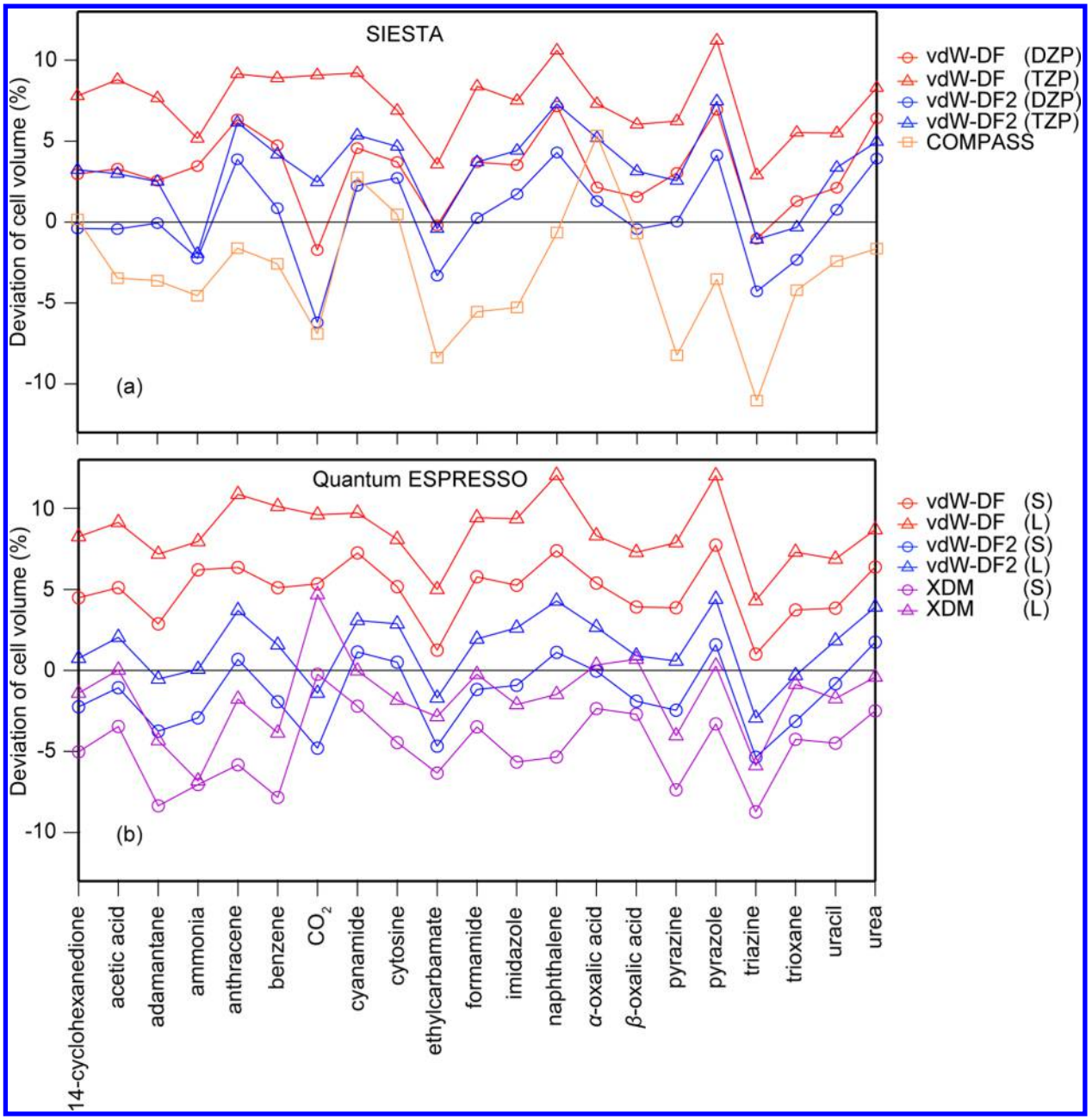


Figure 1. Deviations (%) of the cell volumes compared to the C21 experimental reference structures using (a) SIESTA and (b) Quantum ESPRESSO. Results are shown for a small basis set (DZP for SIESTA and S for Quantum ESPRESSO) and large basis set (TZP for SIESTA, and L for Quantum ESPRESSO), using the vdW-DF, vdW-DF2 and XDM functionals. For comparison, we have also included the results from the COMPASS force field in panel a.

RESULTS AND DISCUSSION

1. C21 Reference Data Set. The C21 reference data set of Otero-de-la-Rosa and Johnson[46] consists of 21 small molecular crystals with structures determined at temperatures ranging from 10 to 298 K. In their original work, Otero-de-la-Rosa and Johnson calculated thermal and zero point effects and optimized the geometries under an external pressure equal to the thermal pressure (which is calculated from the vibrations of the crystals). For small crystals, this procedure is quite feasible, but for large molecular crystal systems involving hundreds of atoms, this is still largely difficult and not currently computationally feasible. Hence, in our study here, we will directly compare the optimized geometries from 0 K static DFT calculations with the lattice parameters from the experimental crystal structures.

We begin by first examining the how well our DFT methods can reproduce the experimental geometries of the molecular crystals. In Table 1, we report the deviation of the cell volumes (in percent) from the experimental reference structures in the C21 data set, for SIESTA and Quantum ESPRESSO. Here, we use the vdW-DF, vdW-DF2, and XDM functionals with the small and large basis sets options described in the Methodology section. We also report the mean deviation (MD) and mean absolute deviation (MAD) of the volumes (in percent), the MAD and MA%D of the cell lengths (in Å and %, respectively), and the MAD and MA%D of the cell angles (in degrees and %, respectively). For comparison we also include the results for

dx.doi.org/10.1021/ct500335b | J. Chem. Theory Comput. 2014, 10, 3423−3437

Table 2. Lattice Energies (kJ/mol) for the C21 Data Set Using SIESTA and Quantum ESPRESSO, as a Function of Basis Set Size and vdW Functional[a]

	SIESTA					Quantum ESPRESSO						
	vdW-DF		vdW-DF2			vdW-DF		vdW-DF2		XDM		
	DZP	TZP	DZP	TZP	hybrid[b]	S	L	S	L	S	L	ref
14-cyclohexanedione	126.51	108.25	125.41	103.68	102.60	102.09	103.29	102.11	102.67	86.36	87.14	86.53
acetic acid	93.02	80.44	94.56	79.70	79.58	75.94	76.40	76.67	76.97	72.57	72.95	71.80
adamantane	93.61	95.34	87.65	87.89	86.88	86.79	89.20	83.19	84.17	72.08	73.06	62.43
ammonia	43.29	40.64	47.29	43.67	43.79	37.08	37.17	40.20	40.25	39.59	39.61	37.57
anthracene	124.93	131.66	113.80	115.79	114.07	113.63	116.39	107.06	108.43	100.30	101.71	100.58
benzene	65.60	67.37	60.99	60.20	59.29	58.38	59.84	55.36	55.88	51.08	51.71	50.41
carbon dioxide	46.72	35.92	46.26	32.03	31.59	35.06	35.24	33.39	33.48	22.32	22.53	27.80
cyanamide	98.88	89.56	102.16	91.31	91.50	84.24	84.45	88.34	88.53	88.48	88.68	79.16
cytosine	179.74	163.44	180.78	162.24	161.71	152.46	153.21	156.49	157.06	152.91	153.57	168.81
ethylcarbamate	114.21	98.80	114.79	96.99	96.39	93.79	94.56	93.69	94.15	84.73	85.23	84.17
formamide	94.81	82.71	98.35	84.75	84.47	78.73	79.05	82.29	82.54	79.15	79.44	78.74
imidazole	103.53	94.79	102.28	92.01	91.70	88.77	89.52	88.76	89.18	88.35	88.85	85.75
naphthalene	94.49	98.00	86.74	87.11	86.24	85.62	87.87	80.49	81.41	75.20	76.21	76.32
α-oxalic acid	127.57	107.28	132.47	109.37	109.14	102.86	103.25	107.58	108.42	90.93	91.56	96.02
β-oxalic acid	126.44	107.53	129.32	107.81	107.18	103.33	103.73	107.35	107.66	93.24	94.00	95.85
pyrazine	84.02	75.47	83.37	72.33	72.01	67.65	68.56	66.98	67.43	60.18	60.77	62.16
pyrazole	92.70	85.35	91.52	82.61	82.09	78.18	78.86	78.04	78.39	76.73	77.10	76.83
triazine	86.51	71.75	86.00	68.58	68.85	66.91	67.58	65.60	66.00	55.40	55.96	60.51
trioxane	94.14	76.55	96.88	76.84	76.66	74.31	75.02	74.61	74.95	58.34	58.90	62.54
uracil	164.21	143.98	165.12	140.50	140.20	135.59	136.25	138.60	139.10	131.51	132.16	132.90
urea	120.00	105.37	126.03	109.94	110.41	100.82	101.23	107.55	107.88	103.54	103.85	99.43
MAE (kJ/mol)	22.79	13.08	22.64	10.58	10.20	7.93	8.41	7.75	8.18	3.58	3.49	

[a]A hybrid strategy is also included for the SIESTA code, combining the functional/basis set that gives the best geometries and best lattice energies. The mean absolute errors (MAE) are reported with reference to the corrected experimental values reported by Otero-de-la-Rosa and Johnson.[46] CP corrections to account for BSSE errors in SIESTA calculations are not included in these results. [b]Hybrid strategy in SIESTA uses vdW-DF2(DZP) geometries with vdW-DF2(TZP) single point energy calculations.

molecular mechanics calculations with the COMPASS[66,67] force field (which has been optimized for use with solids). In Figure 1, we illustrate the deviations of the cell volumes for the SIESTA, Quantum ESPRESSO, and COMPASS calculations.

The cell volume deviations in Figure 1 show that the vdW-DF functional almost always overestimates the cell volumes, while the XDM functional underestimates most of them. The vdW-DF2 results generally lie between the XDM and vdW-DF results; that is, the vdW-DF2 functional appears to give the best results for both the small and large basis sets within the local (SIESTA) and PAW (Quantum ESPRESSO) approaches. Overall, all functionals and basis set qualities appear to follow similar trends, with the one exception being XDM calculations of the CO_2 structure.

Turning now to Table 1, we examine in more detail the deviations in the cell volumes, cell lengths, and cell angles, to compare how well each method performs compared to the reference geometries. In particular, we focus on the MA%D of the cell lengths to ascertain the best performance. Using SIESTA, the best result of 1.16% (smallest deviation) is obtained using the vdW-DF2 functional with the smaller DZP basis set, followed by vdW-DF (DZP), vdW-DF2(TZP), and vdW-DF(TZP). Using Quantum ESPRESSO, we also find the vdW-DF2 functional performs best, with only very small differences between the L (0.90%) and S (0.96%) basis sets for the MA%D of the cell lengths. The corresponding %MAD of the cell volumes are also almost identical, and vary by only 0.01%. Using the MA%D of the cell lengths, the best results are obtained by vdW-DF2(L), very closely followed by vdW-DF2(S), then XDM(L). There is a noticeable gap to the next best result from XDM(S) and vdW-DF(S), and finally, vdW-DF(L) performs worst. In summary for both codes, the vdW-DF2 functional gives the best geometries, and the vdW-DF functional, in particular with the larger basis set, performs worst.

In contrast to our finding that vdW-DF2 gives the best results for both SIESTA and Quantum ESPRESSO, Otero-de-la-Rosa and Johnson[46] report that XDM(B86b) gives geometries that are slightly better than vdW-DF2 and much better than vdW-DF. This difference is primarily because Otero-de-la-Rosa and Johnson employ a lattice expansion technique (using a negative pressure) to account for thermal affects in solids, whereas we only use the 0 K optimized geometries. The XDM cell volumes shown in Figure 1 and reported in Table 1 are almost all underestimated (negative) at 0 K compared to the experimental values. Surprisingly, when the lattice expansion technique was applied by Otero-de-la-Rosa and Johnson,[46] the XDM cell volumes became almost all overestimated according to their Supporting Information. At 0 K, our best results for the MA%D of the cell lengths from Quantum ESPRESSO is 0.90% with vdW-DF2(L), and the values for vdW-DF2(S) and XDM(L) are 0.96% and 1.06%, respectively. Significantly, Otero-de-la-Rosa and Johnson[46] report the MA%D cell length values after applying the lattice expansion technique are 1.88% and 1.76%, for vdW-DF2(L) and XDM(L), respectively. Thus, the lattice expansion technique appears to overcorrect the lattice parameters and volumes. Our finding that vdW-DF2 outperforms vdW-DF has also been reported elsewhere for molecular crystals.[41,46]

In Table 1, we also include the C21 results using the COMPASS force field, which has been optimized for use with solids. Examining the average MA%D of the cell lengths shows the best results from SIESTA and Quantum ESPRESSO are 1.16% and 0.90%, respectively, while the COMPASS results show a 5.1% deviation. Note that the error in the cell volume % MAD for COMPASS is somewhat smaller; however, this is unsurprising given the Lennard-Jones potentials describing the intermolecular interactions in COMPASS are fitted to experimental densities and consequently will always calculate a reasonable value for the densities, while generating large deviations in the lattice parameters. This illustrates that DFT with an appropriate van der Waals functional and basis set greatly outperforms (from a geometric perspective) a general force field, albeit at much greater computational cost.

We now examine how well our DFT methods can reproduce the experimental lattice energies reported for the C21 reference set.[46] In Table 2, we report the lattice energies for each structure in the C21 reference set, using the various choices of exchange-correlation functionals, basis sets functions and basis set sizes, as described previously. The inclusion of corrections for BSSE for local orbital approaches can be a nontrival exercise, and in very large and complicated molecular and crystalline systems, CP corrections simply may not be feasible due to the complexity. We discuss the effects of including CP corrections later in this section (see Table 3); consequently, the lattice energies for the SIESTA calculations reported in Table 2 contain no CP correction. In Table 2, we report the mean absolute errors (MAE) of the lattice energies, and we illustrate the deviations in the lattice energies from the reference values in Figure 2.

Table 3. Lattice Energies (kJ/mol) for the C21 Data Set Using SIESTA as a Function of Basis Set Size for the vdW-DF and vdW-DF2 Exchange-Correlation Functionals[a]

	SIESTA				
	vdW-DF		vdW-DF2		
	DZP	TZP	DZP	TZP	ref
14-cyclohexanedione	95.18	95.46	93.57	91.22	86.53
acetic acid	70.96	71.28	71.33	70.30	71.80
adamantane	76.21	79.21	72.96	72.70	62.43
ammonia	35.43	36.37	39.15	39.15	37.57
anthracene	106.31	103.72	97.50	92.85	100.58
benzene	55.84	54.01	52.31	48.84	50.41
carbon dioxide	28.69	31.48	26.19	27.46	27.80
cyanamide	79.46	81.76	82.86	83.87	79.16
cytosine	146.09	146.46	147.85	146.78	168.81
ethylcarbamate	86.43	88.42	86.42	86.39	84.17
formamide	75.38	75.74	78.27	77.82	78.74
imidazole	85.47	84.45	85.06	82.68	85.75
naphthalene	80.21	77.50	74.21	70.02	76.32
α-oxalic acid	91.36	94.84	94.43	97.10	96.02
β-oxalic acid	92.20	95.08	92.91	95.29	95.85
pyrazine	65.64	63.56	64.90	61.63	62.16
pyrazole	76.58	74.64	76.10	72.94	76.83
triazine	63.67	61.34	62.50	59.02	60.51
trioxane	70.38	69.57	72.00	69.49	62.54
uracil	128.46	128.28	128.97	126.18	132.90
urea	94.74	97.23	100.41	101.95	99.43
MAE (kJ/mol)	4.88	4.42	3.85	4.32	

[a]Lattice energies are CP corrected to account for BSSE errors. The mean absolute errors (MAE) are reported with reference to the corrected experimental values reported by Otero-de-la-Rosa and Johnson.[46]

We begin by first examining the general trends for the lattice energy deviations in Figure 2. Overall, there is broad agreement in the behavior (whether the deviation increases or decreases) as we move through each crystal structure, with the main difference being the magnitude of the deviation. For example, there are downward spikes in the deviations for cytosine. The deviations in lattice energies with SIESTA are almost all larger than for Quantum ESPRESSO, with SIESTA almost always overestimating the lattice energy (with the exception of cytosine).

Examining the mean absolute errors in the lattice energies in Table 2, there are a number of interesting conclusions that can be made. Overall, the XDM(L) functional in Quantum ESPRESSO gives the most accurate lattice energies (lowest MAE) with an MAE of 3.49 kJ/mol. The XDM(S) method is next best by a very small margin, with an MAE of 3.58 kJ/mol. Using Quantum ESPRESSO, the vdW-DF and vdW-DF2 results are very similar regardless of functional or basis set option, with MAE values varying between 7.75 and 8.41 kJ/mol. For lattice energies using a given basis set size, the choice of vdW-DF or vdW-DF2 functionals appears to make little difference, as was also reported by Otero-de-la-Roza and Johnson.[46] This observation is true for SIESTA as well, although there is a little more variation than with Quantum ESPRESSO.

Using SIESTA, the vdW-DF and vdW-DF2 results overall are slightly worse compared to the equivalent Quantum ESPRESSO calculation, and the choice of basis set has a large effect on the MAE results. The smaller DZP basis set in SIESTA leads to MAE lattice energies of about 23 kJ/mol for both functionals, substantially worse than the values around 8 kJ/mol with Quantum ESPRESSO. Using the TZP basis set, the errors using vdW-DF and vdW-DF2 improve to 13.08 and 10.58 kJ/mol, respectively. In the original C21 paper, Otero-de-la-Roza and Johnson examined a number of difference exchange-correlation functionals using a larger (L) basis set, and found MAE lattice energies of 4.81 (XDM), 10.22 (vdW-DF), 10.11 (vdW-DF2), 16.97 (PBE-TS), and 35.83 (PBE) kJ/mol. As noted before, Otero-de-la-Roza and Johnson[46] use a negative pressure to account for thermal expansion of crystals in their calculations, and we use the most stable oxalic acid molecule rotamer reported by Blair and Thakker,[73] which explains why their values are slightly different to ours. The best SIESTA result for the lattice energies has an MAE of 10.58 kJ/mol and, while not as accurate as the XDM results, is still very good and comparable to similar Quantum ESPRESSO results and represents a substantial improvement over the standard DFT functionals.

The large basis set effect in SIESTA is perhaps not surprising when the basis functions used in the construction of the basis set are considered. The local orbital approach for basis functions reduces the computational cost compared to using planewaves or other basis functions, but this compromise comes at the expense of the accuracy of the energies. In bulk crystals, the electron densities of neighboring atoms can overlap strongly, so relatively smaller confinement radii can possibly be used without losing too much accuracy.[59] However, for molecules (or surfaces), the local basis orbitals must also describe the electron density reaching into a vacuum. Although the DZP basis set gives the most accurate geometries, it gives

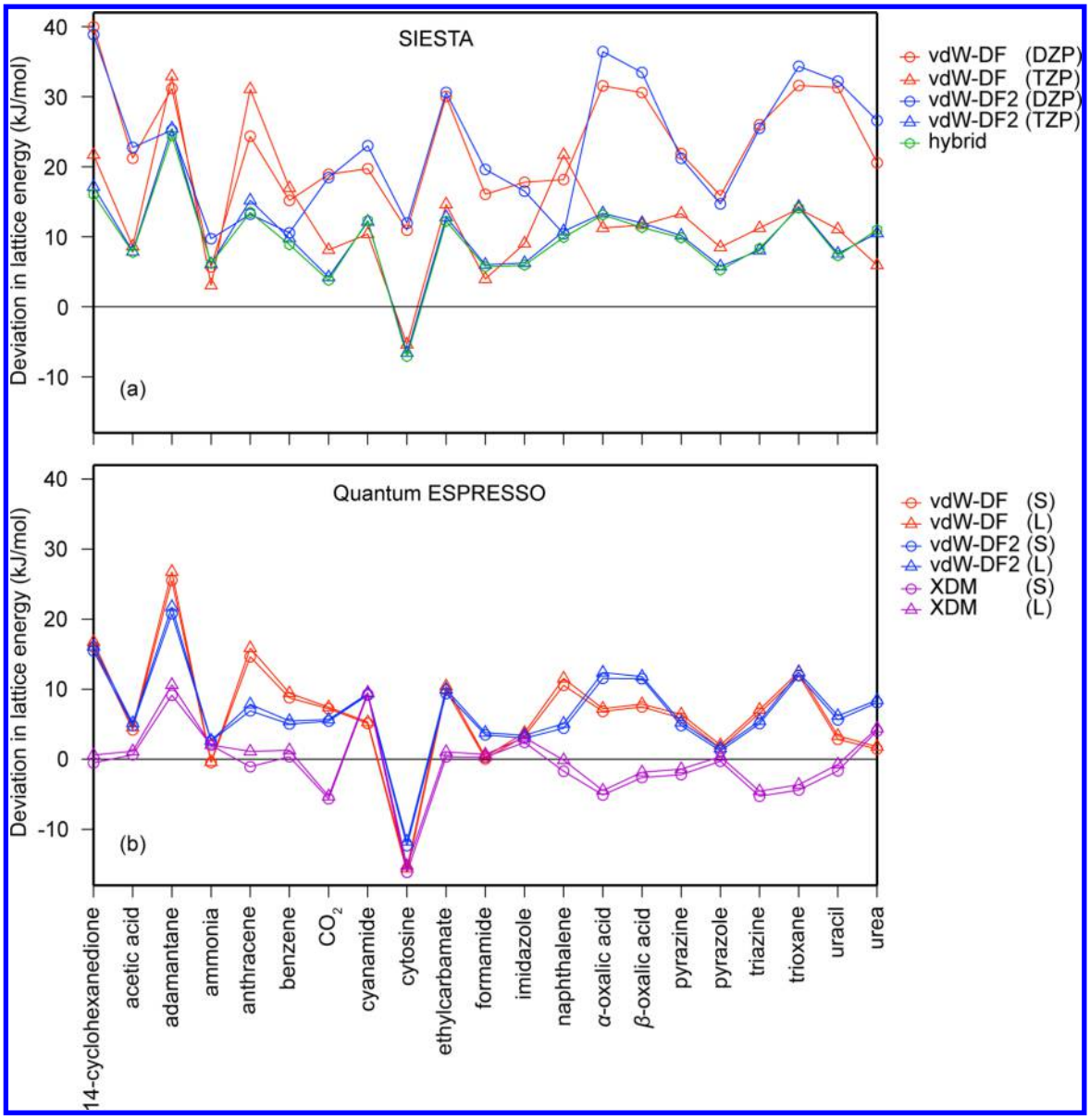


Figure 2. Deviations (kJ/mol) of the lattice energies compared to the C21 reference lattice energies using (a) SIESTA and (b) Quantum ESPRESSO. Results are shown for a small basis set (DZP for SIESTA and S for Quantum ESPRESSO) and large basis set (TZP for SIESTA and L for Quantum ESPRESSO), using the vdW-DF, vdW-DF2 and XDM functionals. The hybrid strategy in SIESTA uses vdW-DF2(DZP) geometries with vdW-DF2(TZP) single point energy calculations. CP corrections to account for BSSE errors in SIESTA calculations are not included in these results.

the worst lattice energies, in particular because of the accuracy of the energy of the isolated molecule from each of the C21 crystal structures. The optimized TZP basis from Louwerse[59] has much longer cutoff radii in comparison to the standard DZP basis set. The larger basis adds a small computational cost but gives a more accurate energy of the isolated molecules and thus more accurate lattice energies. This effect is seen in Figure 2, where the smaller DZP basis sets leads to a greater overestimation of the lattice energies, compared with the TZP basis.

To summarize the SIESTA results so far, we find vdW-DF2 with a smaller DZP basis gives the best geometric match, but vdW-DF2 with a TZP basis gives the most accurate lattice energies. This lead us to formulate a hybrid approach, whereby we optimize the geometries with vdW-DF2(DZP), but perform a single point calculation of the energies using vdW-DF2(TZP). This type of approach is not new to quantum chemistry and is often used for computationally intensive methods, for example MP2, where the geometries are obtained at a lower level of theory and a single point energy calculation is performed at the higher level of theory.[72]

The lattice energies for the hybrid strategy are listed in Table 2, and the lattice energy differences are also illustrated in Figure 2. The hybrid strategy gives the most accurate lattice energies for any of the SIESTA calculations, with an MAE of 10.20 kJ/mol. This result highlights the usefulness of the hybrid strategy, since using vdW-DF2(DZP) for the geometries and energies has an MAE of the lattice energy of 22.64 kJ/mol, but using the hybrid strategy, the MAE reduces by approximately 12 kJ/mol, to a value of 10.20 kJ/mol.

We now examine the effect of BSSE corrections on the lattice energies from SIESTA calculations. In Table 3, we report the

dx.doi.org/10.1021/ct500335b | J. Chem. Theory Comput. 2014, 10, 3423−3437

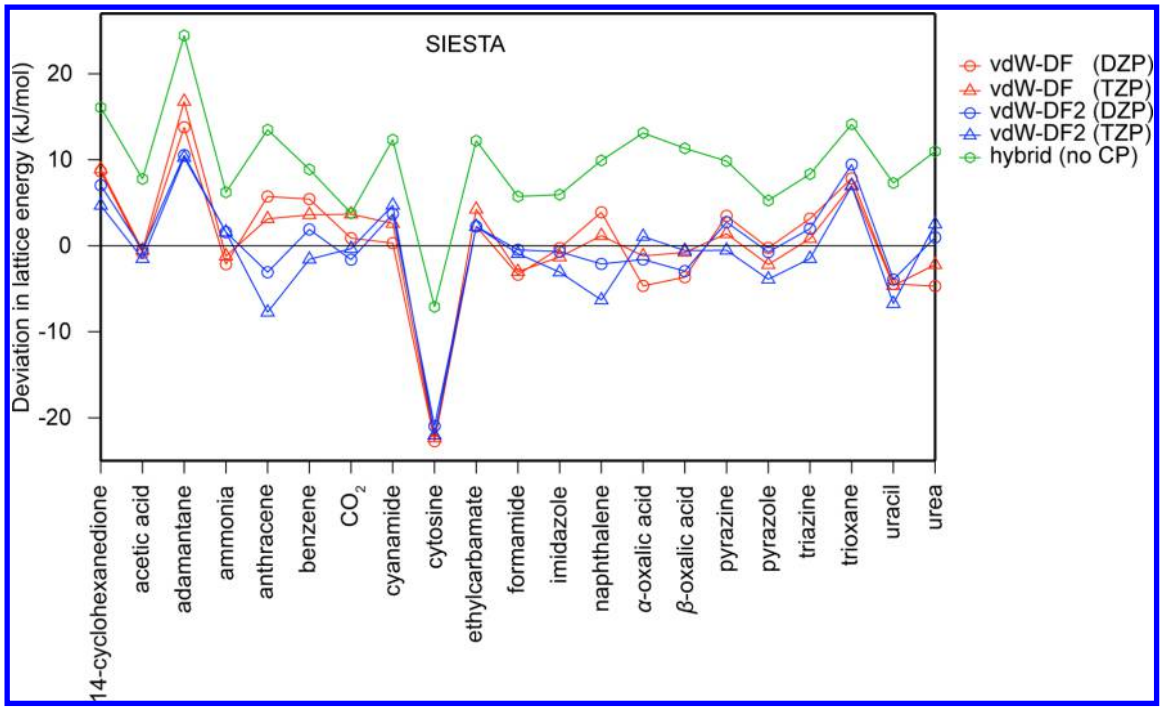


Figure 3. Deviations (kJ/mol) of the CP corrected lattice energies compared to the C21 reference lattice energies using SIESTA. Results are shown for a small (DZP) and larger (TZP) basis set, using the vdW-DF and vdW-DF2 functionals. Also shown are the results for the hybrid strategy in SIESTA, which uses vdW-DF2(DZP) geometries combined with vdW-DF2(TZP) single point energy calculations and has no CP correction for BSSE errors.

CP corrected lattice energies for vdW-DF and vdW-DF2 using the small and large basis sets. The CP corrections have quite a dramatic effect on the lattice energies, significantly improving them by around 6−20 kJ/mol, depending on the choice of functional and basis set size. The MAE of the vdW-DF2(DZP) lattice energies improves from 22.64 to 3.85 kJ/mol by the inclusion of CP corrections, improving it from equal worst to our most accurate SIESTA result, with an MAE only slightly larger than the XDM results. The MAE results for vdW-DF(DZP), vdW-DF(TZP), and vdW-DF2(TZP) are almost identical to each other and are slightly higher with values around 4.5 kJ/mol. The spread of the SIESTA results are now consistent with the planewave Quantum ESPRESSO results, which showed that the MAE values for vdW-DF and vdW-DF2 with the small and large basis were all very similar and within 1 kJ/mol of each other. The MAE values are now also about 3.5 kJ/mol better than the equivalent planewave results in Table 2.

In Figure 3 we illustrate the deviations of the C21 lattice energies from the reference values for the CP corrected lattice energies in Table 3. For visual comparison, we also show the results from the hybrid method (which has no CP correction). Figure 3 clearly shows that the inclusion of a CP correction for the BSSE errors improves the accuracy of the lattice energies for all functionals and basis set sizes, compared to the best results when no CP corrections are applied (hybrid method). In all cases except for cytosine, the deviations in lattice energies are noticeably reduced by the inclusion of the CP correction.

In summary, we have shown that when using the planewave Quantum ESPRESSO approach, the XDM(L) method gives the most accurate lattice energies with an MAE of 3.49 kJ/mol. Using the local orbital SIESTA approach, when CP corrections for BSSE are not feasible, a hybrid strategy produces the most accurate lattice energies with an MAE of 10.20 kJ/mol. When CP corrections are included, the vdW-DF2(DZP) method gives the most accurate lattice energies, with an MAE of 3.85 kJ/mol. The most accurate lattice energies from SIESTA and Quantum ESPRESSO are within the coveted "chemical accuracy" of 4.2 kJ/mol,[74] indicating the high quality of these results. Brandenburg et al.[75] recently devised a CP correction method, called gCP, that can be used while optimizing cells and atom positions, and they applied it to the X23[49] set of molecular crystals (adapted from the C21 reference set). They report an empirically derived 50% gCP contribution applied to B3LYP-D3 can give MAE lattice energies of 7.1 kJ/mol. However, we must note that they optimized the atomic coordinates with the lattice parameters fixed at the experimental values. The gCP method combined with D3 dispersion has been recently applied[51] to Hartree−Fock calculations of the X23 set, giving an MAE of 6.28 kJ/mol for the lattice energies. Other recent D3 dispersion calculations[52] of the X23 set report MAE values of 4.60 (PBE-D3), 3.77 (TPSS-D3), 5.02 (HSE06-D3), and 5.02 (PBE0-D3) kJ/mol.

The recent work of Reilly and Tkatchenko[49] showed that including many-body effects with their PBE0-MPD method gives an MAE of the lattice energies of their X23 data set of 3.92 kJ/mol. The lattice parameters and cell coordinates were optimized using the PBE-TS functional before applying the many-body single-point energy corrections. It would be very interesting to see if optimizing the structures with the PBE0-MPD method would further improve the lattice energy MAE, although this would require considerable computational cost.

2. Room Temperature C21 Reference Data Set Geometries. When assessing the accuracy of 0 K DFT optimized geometries against experimental crystal structures, the temperature that the experimental structure was recorded is an important factor. As noted above, the C21 reference data set has temperatures ranging from 10 to 298 K; therefore, to further investigate the accuracy of our DFT calculations, we

Table 4. Mean, Absolute, and Percentage Deviations of the Cell Volumes, Lengths, and Angles for the $C19_{RT}$ Data Set

	SIESTA				Quantum ESPRESSO						
	vdW-DF		vdW-DF2		vdW-DF		vdW-DF2		XDM		
	DZP	TZP	DZP	TZP	S	L	S	L	S	L	COMPASS
vol %MD	1.75	5.72	−0.83	1.99	3.14	6.77	−2.79	0.03	−6.48	−3.26	−4.39
vol %MAD	2.97	5.72	3.44	3.35	3.43	6.77	3.66	2.98	6.48	3.54	5.39
cell length MAD	0.13	0.17	0.12	0.13	0.12	0.18	0.11	0.11	0.16	0.10	0.40
cell angle MAD	0.85	0.93	0.95	1.00	1.04	0.96	1.41	1.32	1.32	1.10	1.96
cell length MA%D	1.79	2.26	1.75	1.76	1.71	2.39	1.62	1.54	2.16	1.40	5.66
cell angle MA%D	0.80	0.88	0.88	0.93	1.01	0.94	1.36	1.28	1.25	1.07	1.79

Table 5. Deviations (%) of the Cell Volumes Compared to the Reference Experimental Monosaccharide Data Set Using SIESTA and Quantum ESPRESSO, as a Function of Basis Set Size and vdW Functional[a]

	SIESTA				Quantum ESPRESSO					
	vdW-DF		vdW-DF2		vdW-DF		vdW-DF2		XDM	
	DZP	TZP	DZP	TZP	S	L	S	L	S	L
α-D-talose	1.96	4.31	0.05	1.79	2.47	5.00	−1.56	1.07	−5.74	−2.82
β-D-galactose	2.79	5.34	0.76	3.14	3.63	6.36	−0.87	2.76	−4.39	−1.56
β-D-allose	3.81	5.62	3.14	4.05	4.00	6.74	0.61	2.72	−4.34	−2.43
β-D-altropyranose	3.93	5.51	2.94	4.10	4.01	6.37	0.51	3.13	−3.33	−2.88
β-D-fructose	3.21	4.86	1.83	3.01	3.42	6.29	−0.08	2.28	−4.57	−2.45
α-D-glucose	4.54	5.92	4.00	4.76	4.51	7.04	1.20	3.78	−3.67	−1.59
β-D-glucose	4.36	6.34	2.75	3.47	4.73	7.64	0.75	3.25	−3.24	−0.65
β-D-psicopyranose	3.78	5.57	1.98	3.28	3.63	6.77	0.08	2.36	−3.38	−1.35
α-D-tagatopyranose	4.78	6.85	3.70	4.81	5.49	7.70	1.62	3.75	−3.27	−1.41
vol %MD	3.69	5.59	2.35	3.60	3.99	6.66	0.25	2.79	−3.99	−1.90
vol %MAD	3.69	5.59	2.35	3.60	3.99	6.66	0.81	2.79	3.99	1.90
cell length MAD	0.13	0.17	0.15	0.13	0.13	0.21	0.09	0.10	0.12	0.08
cell angle MAD	0.01	0.02	0.06	0.03	0.00	0.00	0.00	0.00	0.00	0.00
cell length MA%D	1.39	1.78	1.46	1.33	1.29	2.12	0.86	0.98	1.38	0.93
cell angle MA%D	0.03	0.03	0.13	0.07	0.00	0.00	0.01	0.00	0.00	0.00

[a]The mean, absolute, and percentage deviations of the cell volumes, lengths, and angles are also reported.

created a room temperature version of the data set. This was motivated by our interest in using DFT to assist in determining the structures of polymorphs where often the only experimental information regarding the structure is room temperature XRD. The original C21 data set has only 6 of the 21 experimental crystal structures measured at room temperature. However, with the exception of the gases ammonia and carbon dioxide, room temperature structures exist in the literature, yielding the $C19_{RT}$ data set. In Table 4 we report the mean deviation (MD) of the volumes (%) for each structure in the room temperature version of the C21 data set. As with Table 1, we report the MD and MAD of the volumes (in percent), the MAD and MA%D of the cell lengths (in Ångstrom and percent, respectively), and the MAD and MA%D of the cell angles (in degrees and percent, respectively). For comparative purposes we again include the results for molecular mechanics calculations with the COMPASS force field.

The most interesting trends from Table 4 relate to the % MAD of the cell volumes and the MA%D of the cell lengths. For SIESTA, vdW-DF(DZP) gives the best result for %MAD of the cell volume, in contrast to the result of the original C21 data set (Table 1), where vdW-DF2(DZP) gives the most accurate result. For Quantum ESPRESSO vdW-DF2(L) still gives the best %MAD of the cell volume, but vdW-DF(S) is much improved and is only about 0.5% larger, compared to 2.8% larger for the C21 results. The fact that the vdW-DF results improve when compared to a room temperature C21 data set is expected. As can be seen by the mean deviations of the cell volumes in Figure 1, vdW-DF largely overestimates volumes compared to the largely low temperature experimental crystal structures in C21, so when compared to the room temperature C21 data set, where the cell volumes are almost all larger compared to their values in the original C21 data set, the results are improved.

For the MA%D of the cell lengths in SIESTA, we find that vdW-DF2(DZP) combination still gives the best result, but we also find vdW-DF(DZP), vdW-DF2(TZP) and vdW-DF(DZP) results are almost identical, with values of 1.75, 1.76 and 1.79%, respectively. This is in contrast to the original C21 data set, where these same quantities were 1.16, 1.51 and 1.40%, respectively. The best MA%D of the cell lengths in Quantum ESPRESSO is again observed for the vdW-DF2(L) combination, but with vdW-DF2(L), vdW-DF2(S), and vdW-DF(S) having all very similar to values of 1.54, 1.62 and 1.71%, respectively. This again contrasts with the results for the original C21 reference set, where these same quantities were 0.90, 0.96, and 1.73%, respectively. Additionally, in the original comparison XDM with a large basis gave a MA%D of the cell lengths of 1.06%, making it the third best option, whereas here XDM(L) gives the best result with a value of 1.40%.

In summary, the methods that gave the best MA%D of the cell lengths for the original C21 set still give the best (or equal best) results for the room temperature $C19_{RT}$ set. For SIESTA, it is vdW-DF2(DZP), and for Quantum ESPRESSO, it is vdW-

dx.doi.org/10.1021/ct500335b | J. Chem. Theory Comput. 2014, 10, 3423−3437

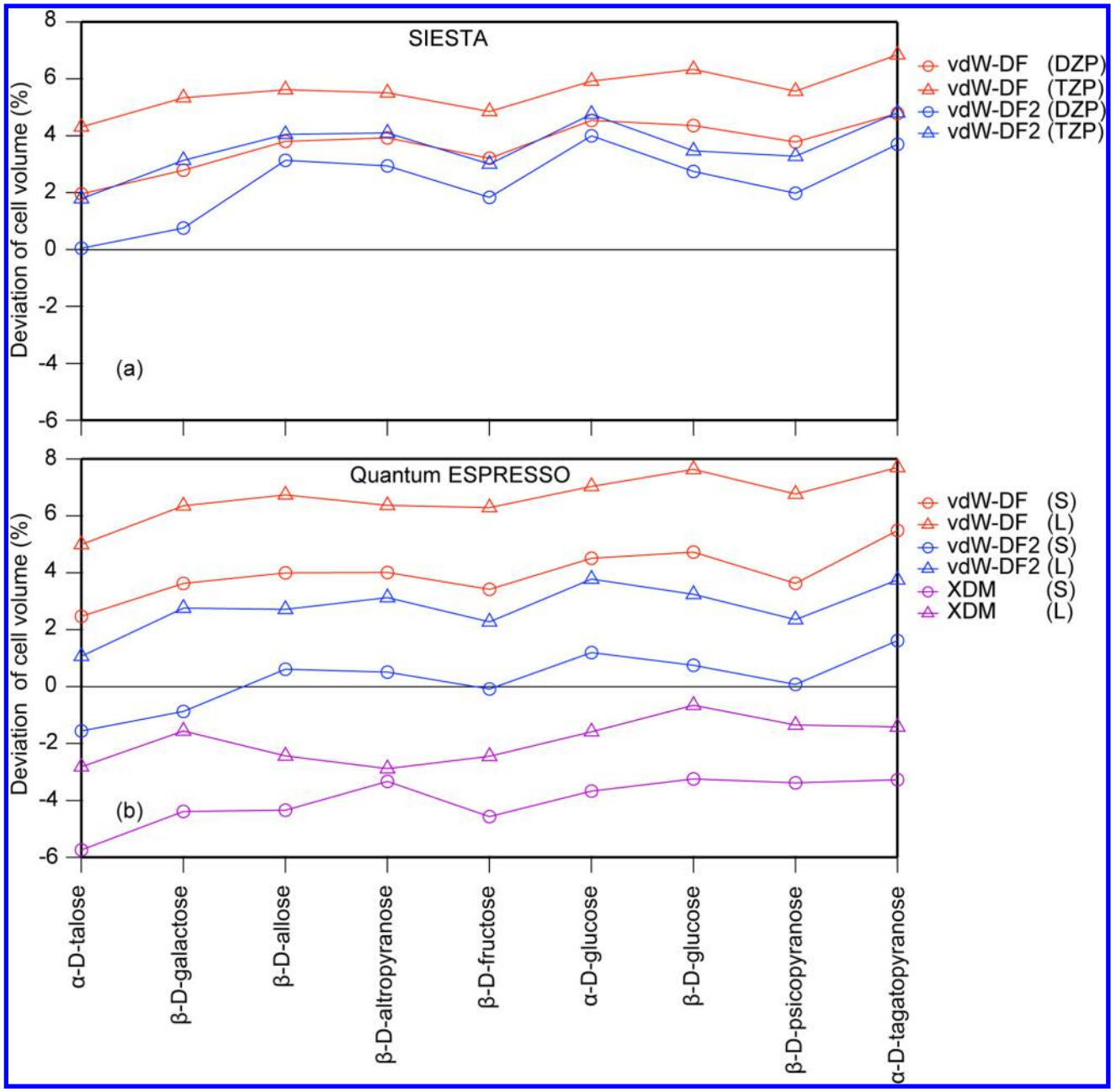


Figure 4. Deviations of the cell volume (%) for the monosaccharide data set using (a) SIESTA and (b) Quantum ESPRESSO. Results are shown for a small basis set (DZP for SIESTA and S for Quantum ESPRESSO) and large basis set (TZP for SIESTA and L for Quantum ESPRESSO), using the vdW-DF, vdW-DF2, and XDM functionals.

DF2(L). Using the $C19_{RT}$ set, where the cell volumes are almost all larger than in the C21 reference set, several basis set and vdW functional options perform relatively better than their results for the C21 set. There are also small changes in the % MAD of the cell volumes compared to trends in the original set.

3. Monosaccharide Data Set. To assess the performance of our vdW methods, and to see if the general trends and observations from the C21 reference set of molecular crystals were continued, we undertook simulations on a series of monosaccharide structures taken from the CSD[68] database determined at temperatures ranging from 95 to 298 K. In Table 5, we report the deviations of the cell volumes for SIESTA and Quantum ESPRESSO as a function of the choice of basis set and exchange-correlation functional. We also report the mean, absolute, and percentage deviations of the cell volumes, lengths, and angles compared to the reference monosaccharide data set. In Figure 4 we illustrate the deviations of the cell volumes to highlight the trends.

Figure 4 shows similar trends overall for both SIESTA and Quantum ESPRESSO and for both basis set sizes. Each vdW functional/basis set combination follows a similar general deviation trend across all nine monosaccharide structures. As observed for the C21 reference set, the vdW-DF functional overestimates cell volumes, XDM underestimates cell volumes, and vdW-DF2 falls somewhere in the middle of these two. The results in Figure 4 suggest that vdW-DF2 with a smaller basis set gives the smallest deviations for the monosaccharides for both SIESTA and Quantum ESPRESSO. Examining the detailed results in Table 5, this is indeed the case for both SIESTA and Quantum ESPRESSO, with vdW-DF2(DZP) and

dx.doi.org/10.1021/ct500335b | J. Chem. Theory Comput. 2014, 10, 3423−3437

Table 6. Lattice Energies (kJ/mol) for the Monosaccharide Data Set Using SIESTA and Quantum ESPRESSO[a]

	SIESTA					Quantum ESPRESSO					
	vdW-DF		vdW-DF2			vdW-DF		vdW-DF2		XDM	
	DZP	TZP	DZP	TZP	hybrid[b] (no CP)	S	L	S	L	S	L
α-D-talose	180.43	181.79	190.30	189.09	209.24	−193.27	−193.92	−203.53	−204.07	−196.84	−191.17
β-D-galactose	161.01	162.69	172.66	173.25	193.57	−173.82	−174.37	−184.46	−184.96	−178.90	−175.70
β-D-allose	161.25	162.51	172.15	171.19	192.60	−173.78	−174.49	−184.25	−184.91	−174.13	−171.07
β-D-altropyranose	164.20	164.36	177.38	174.88	196.35	−179.32	−179.89	−191.18	−191.93	−187.83	−182.96
β-D-fructose	154.79	156.99	166.94	167.02	187.77	−168.41	−169.20	−178.28	−179.05	−169.89	−164.68
α-D-glucose	183.05	185.57	193.83	194.23	215.30	−197.42	−198.06	−207.69	−208.29	−202.88	−199.58
β-D-glucose	157.99	161.60	171.13	172.69	194.57	−173.57	−174.18	−183.74	−184.38	−179.61	−173.69
β-D-psicopyranose	156.83	158.52	167.24	166.58	186.27	−169.08	−170.06	−178.34	−178.97	−169.22	−165.53
α-D-tagatopyranose	163.57	163.99	176.53	174.94	196.48	−175.33	−175.88	−186.25	−187.00	−175.85	−171.48

[a]SIESTA lattice energies are CP corrected to account for BSSE errors, except for the hybrid strategy. [b]Hybrid strategy in SIESTA uses vdW-DF2(DZP) geometries with vdW-DF2(TZP) single point energy calculations.

vdW-DF2(S) giving the smallest deviations (%MAD) of the cell volumes of 2.35 and 0.81%, respectively. This is consistent with the trend observed for the C21 reference data set comparisons, with the vdW-DF2 functional giving the best geometric matches with the experiment crystal structures.

Examining the MA%D of the cell lengths we find some slight differences compared to observations from the C21 reference set. For Quantum ESPRESSO, vdW-DF2(S) is the most accurate, followed by XDM(L) and vdW-DF2(L) with values of 0.86, 0.93, and 0.98%, respectively. As reported in Table 1 for the C21 reference set, vdW-DF2(L) gave the most accurate geometries, although the differences between the S and L basis set results was 0.01% for the %MAD of the cell volumes, and 0.06% for the MA%D of the cell lengths; that is, essentially they perform equally as well as each other. For the monosaccharide data set, the vdW-DF2(S) cell volumes are about 2% better but the MA%D of the cell lengths only varies by 0.12%.

SIESTA results for the %MAD of the cell volumes are consistent with those from the C21 reference set, with the vdW-DF2(DZP) combination giving the most accurate result of 2.35% (the second best result is vdW-DF2(TZP) with 3.60%). There is a slightly different trend in the MA%D of the cell lengths with vdW-DF2(TZP) giving the most accurate value of 1.33%, closely followed by vdW-DF(DZP) and vdW-DF2(DZP) with values of 1.39 and 1.46%, respectively. In this case the vdW-DF2(DZP) combination does not give the best result, but the difference in MA%D of the cell lengths is only 0.13% between first and third best, which indicates that they have a very similar level of accuracy.

The lattice energies for the monosaccharide data set are listed in Table 6. For SIESTA calculations we include CP corrections for BSSE errors, and we also include calculations with the hybrid method (with no CP correction) for comparison. Overall, the lattice energies in many cases are quite similar to each other, not unsurprising given the similarities of many of the structures. The CP corrected lattice energies in SIESTA are virtually all smaller than the equivalent Quantum ESPRESSO values (by roughly 10−15 kJ/mol). Similar behavior is seen for the C21 results as can be seen by comparing the SIESTA CP corrected lattice energies in Table 3, with the Quantum ESPRESSO lattice energies in Table 2. The effect of basis set size is very small for both codes for vdW-DF and vdW-DF2 and only results in differences of typically 1−2 kJ/mol. This is also consistent with the C21 reference set observations. For XDM, the basis size effect is also very small for the C21 reference set, but for the monosaccharides it is a little larger with differences close to 5 kJ/mol. When the basis set size is kept fixed, and the functional is changed from the vdW-DF to vdW-DF2, both codes show an increase in the lattice energy of around 10 kJ/mol.

In Table 6 we also include the results from the hybrid strategy with no CP corrections. Compared to the vdW-DF2(DZP) results, which gave the most accurate CP-corrected lattice energies for the C21 reference set, the hybrid results are approximately 20 kJ/mol higher in all cases. This is consistent with the behavior we observed for the C21 reference set, where the lattice energies are also approximately 20 kJ/mol higher using the hybrid method with no CP corrections.

Despite the wealth of structural information for the monosaccharides there is a distinct lack of experimental sublimation enthalpies to compare with our calculated results. This is because measurements of vapor pressures of low-volatility materials, such as sugars, require high temperatures to obtain conveniently measurable pressures. However, high temperatures cannot be used because of thermal decomposition.[76] The only two experimental lattice energies we have found in the literature are for α-D-glucose, and the values are quite different, ranging from 138.4 ± 7.9[77] to 194.4 ± 5.0 kJ/mol.[76]

Given the mean absolute error of the lattice energies with Quantum ESPRESSO was 3.49 kJ/mol using XDM(L) and with SIESTA was 3.85 kJ/mol using vdW-DF2(DZP), then our calculated lattice energies for α-D-glucose (199.58 and 194.23 kJ/mol, respectively) match much more closely the value reported by Oja and Suuberg.[76] The consistency of the results from both SIESTA and Quantum ESPRESSO for both geometries and energetics give us confidence that they are performing at a similar level as they did for the C21 reference set results.

4. S22 Reference Data Set Geometries. We finish our investigation by shifting our focus from calculations of molecular crystals to calculations of molecules. In our earlier study[28] of the S22 reference set of complexes using the SIESTA code, we predominately focused on the binding energy trends for the S22 set, even comparing binding energy results using fixed and fully relaxed geometries. However, we did not analyze the relaxed geometries (although we made the coordinates freely available[78]) and given the importance of a good description of the geometries for the C21 and monosaccharide reference sets determined here, it is timely to revisit this now.

dx.doi.org/10.1021/ct500335b | J. Chem. Theory Comput. 2014, 10, 3423−3437

Table 7. Root Mean Square Deviations (RMSD) for the S22 Data Set of Structures That Have Been Optimized in SIESTA Using the PBE, vdW-DF, and vdW-DF2 Functionals[a]

	PBE	vdW-DF				vdW-DF2	
S22 complex	TZP	DZ	DZP	TZP	TZP-L	DZP	TZP
adenine–thymine stack	0.172	0.147	0.112	0.100	0.110	0.094	0.103
adenine–thymine WC	0.057	0.066	0.054	0.053	0.061	0.077	0.079
ammonia dimer	0.021	0.051	0.028	0.029	0.017	0.039	0.040
benzene–ammonia	0.064	0.113	0.072	0.065	0.059	0.042	0.052
benzene dimer C_{2h}	0.183	0.102	0.100	0.081	0.134	0.109	0.117
benzene dimer C_{2v}	0.109	0.121	0.119	0.109	0.101	0.102	0.099
benzene–HCN	0.038	0.134	0.071	0.058	0.061	0.068	0.064
benzene–methane	0.120	0.116	0.115	0.120	0.093	0.051	0.083
benzene–water	0.140	0.177	0.094	0.091	0.070	0.034	0.036
ethene dimer	0.089	0.114	0.111	0.088	0.105	0.018	0.008
ethene–ethylene	0.072	0.042	0.026	0.023	0.021	0.016	0.018
formamide dimer	0.047	0.063	0.041	0.035	0.040	0.059	0.054
formic acid dimer	0.101	0.070	0.029	0.031	0.031	0.037	0.047
indole–benzene stack	0.771	0.142	0.114	0.100	0.121	0.138	0.133
indole–benzene t-shape	0.096	0.145	0.117	0.107	0.101	0.110	0.114
methane dimer	0.022	0.021	0.020	0.019	0.013	0.010	0.000
phenol dimer	0.365	0.168	0.083	0.072	0.084	0.053	0.071
2-pyridoxine-2-aminopyridine	0.161	0.209	0.083	0.078	0.063	0.067	0.078
pyrazine dimer	0.159	0.112	0.107	0.097	0.116	0.114	0.141
uracil dimer hb	0.043	0.074	0.060	0.046	0.049	0.087	0.075
uracil dimer stack	0.219	0.124	0.085	0.064	0.068	0.061	0.072
water dimer	0.082	0.128	0.165	0.141	0.065	0.043	0.024
avg.	0.142	0.111	0.082	0.073	0.072	0.065	0.069

[a]The basis sets used for the PBE and vdW-DF calculations were double-zeta (DZ), double-zeta polarized (DZP), triple-zeta polarized (TZP) and triple-zeta polarized long (TZP-L), as described by Carter and Rohl.[28] The basis sets used for the vdW-DF2 calculations were the DZP and TZP basis sets described in the methodology section. The RMSD values are in Ångstrom.

Also, at the time of our earlier study the vdW-DF2 method was not available in the SIESTA code. In Table 7 we report the root-mean-square deviations (RMSD) of the optimized S22 structures using the vdW-DF, vdW-DF2 and PBE functionals for a range of basis set sizes.

Table 7 shows that the vdW-DF2 functional overall gives the smallest average RMSD errors, with the DZP basis producing an average RMSD error of 0.065 Å, and the larger TZP basis producing the second smallest RMSD error of 0.069 Å. The vdW-DF functional produces the next best results, and in contrast with the vdW-DF2 behavior, it has the trend that as the basis set size increases, the RMSD error decreases. The largest TZP-L basis using the vdW-DF functional has the smallest RMSD of 0.072 Å, followed by the TZP and DZP basis sets, and the smallest DZ basis has an RMSD of 0.111 Å. The binding energies for the vdW-DF functional, however, do not follow this trend, with the smaller DZP basis giving the smallest error in the binding energies, followed by the largest TZP-L basis. The PBE average RMSD error of 0.142 Å is the worst result, highlighting the importance of accounting for van der Waals from not only an energetics perspective, but from a geometric perspective as well.

In our earlier study,[28] the best (smallest) mean absolute deviation (MAD) of the binding energies was obtained using vdW-DF with a DZP basis set with a value of 0.59 kcal/mol. The next best result was vdW-DF with a large TZP-L basis set, producing a MAD value of 0.83 kcal/mol. The PBE functional gave the worst results. Using a DZP basis set, we find vdW-DF2 improves over the vdW-DF results, with a MAD value of 0.53 kcal/mol. This is very close to the MAD of 0.52 kcal/mol reported in the original vdW-DF2 paper by Lee et al.[22] Using the optimized TZP basis from this current study, we find an MAD of 0.72 kcal/mol with vdW-DF2. In summary, the vdW-DF2 functional with the DZP basis set produces both the best match in geometries and best match in binding energies for the S22 reference data set of complexes.

In our calculations of the C21 reference set, the hybrid computation with SIESTA was used a model to examine situations when CP corrections for BSSE were not feasible. Using the hybrid method, the MAD of the binding energy is 0.73 kcal/mol, essentially identical to the results for the optimized TZP basis (which uses TZP geometries, rather than DZP geometries). This would make it equal second best, behind the vdW-DF2(DZP) results.

When using local basis functions, in particular for molecular systems, basis set superposition errors (BSSE) can be an important consideration for binding energy calculations, due the artificial lowering of the energies of the complex. In our earlier study,[28] we did not include any account of the BSSE. Given the lattice energy results for the C21 and monosaccharide reference sets and the large changes we observed when including CP corrections, we have calculated the CP corrected binding energies for S22 using only the new vdW-DF2 results (DZP, TZP and hybrid). The CP correction[69,70] was applied using the standard methods. The MAD values for the binding energies using vdW-DF2 with DZP, TZP basis sets and the hybrid method change from 0.52, 0.72, and 0.73 kcal/mol to 0.49, 0.86, and 0.65 kcal/mol, respectively, after applying the CP correction. The BSSE corrections do not significantly change any of the observations and have little effect on the binding energies, also supporting our decision not to include CP corrections in our earlier calculations.[28]

CONCLUSIONS

We have performed DFT calculations on the C21 reference data set and a new monosaccharide reference set to assess the accuracy of both geometries and energies using the vdW-DF, vdW-DF2, and XDM functionals in the SIESTA and Quantum ESPRESSO codes. These codes provided a contrast between the basis function representations of numerical atomic orbitals and planewaves, and we also investigated the effects of a smaller and a larger basis set for each.

For Quantum ESPRESSO the most accurate geometries are obtained using the vdW-DF2 functional, particularly when assessing deviations in the cell volume and cell lengths. There is little effect from the choice of basis set, with the larger L basis slightly preferred for C21, and the S basis is slightly preferred for the monosaccharides. The most accurate lattice energies are obtained using the XDM functional, where the XDM(L) combination produced an MAE of 3.49 kJ/mol for the C21 reference set. Using the SIESTA code the most accurate geometries for the C21 and monosaccharide data sets are obtained using vdW-DF2(DZP). Including CP corrections for BSSE in SIESTA calculations significantly improves the accuracy of lattice energies, with vdW-DF2(DZP) also producing the most accurate lattice energies with an MAE of 3.85 kJ/mol for the C21 reference set. For situations in which CP corrections for BSSE are not feasible, such as for very large systems, we have shown that a hybrid approach, combining vdW-DF2(DZP) geometries with vdW-DF2(TZP) single point calculations for the energies, reduces the MAE of the lattice energies from 22.64 kJ/mol for vdW-DF2(DZP) to 10.20 kJ/mol. Using the $C19_{RT}$ room-temperature reference set, the trends for the %MAD of the cell volumes slightly changes in terms of which performs best, but the MA%D in cell length trend stays essentially the same. Finally, we also re-examined the S22 set of molecular complexes with SIESTA and found the most accurate geometries and binding energies were again obtained using vdW-DF2(DZP).

We have shown in this paper that through a judicious choice of vdW-corrected functional and basis set, excellent geometries and lattice energies can be calculated using both Quantum ESPRESSO and SIESTA. This knowledge is particularly important for the polymorph prediction community, where accurate energies are required for the ranking of candidate structures, and accurate structures are needed for comparison with unknown polymorphs identified from powder XRD.

ASSOCIATED CONTENT

Supporting Information

Details of the optimized basis set for use in SIESTA, the supercells used for CP corrections, and the lattice parameters and CSD codes for structures in the $C19_{RT}$ and monosaccharide reference sets. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Email: d.carter@curtin.edu.au.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Alberto Otero-de-la-Roza and Erin Johnson for helpful discussions and advice relating to the PAW pseudopotentials, C21 data sets, and the XDM method. They also thank J. D. Gale for his advice relating to BSSE corrections. The authors also thank Jan Gerit Brandenburg for his suggestions for the manuscript. The project used computational resources provided by the iVEC facility at Murdoch University.

REFERENCES

(1) Grimme, S. Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

(2) Civalleri, B.; Zicovich-Wilson, C. M.; Valenzano, L.; Ugliengo, P. B3LYP Augmented with an Empirical Dispersion Term (B3LYP-D*) as Applied to Molecular Crystals. *CrystEngComm* **2008**, *10*, 405−410.

(3) Neumann, M. A.; Leusen, F. J. J.; Kendrick, J. A Major Advance in Crystal Structure Prediction. *Angew. Chem., Int. Ed.* **2008**, *47*, 2427−2430.

(4) Riley, K. E.; Vondrasek, J.; Hobza, P. Performance of the DFT-D Method, Paired with the PCM Implicit Solvation Model, for the Computation of Interaction Energies of Solvated Complexes of Biological Interest. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5555−5560.

(5) Williams, R. W.; Malhotra, D. van der Waals Corrections to Density Functional Theory Calculations: Methane, Ethane, Ethylene, Benzene, Formaldehyde, Ammonia, Water, PBE, and CPMD. *Chem. Phys.* **2006**, *327*, 54−62.

(6) 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.

(7) Xu, X.; Goddard, W. A., III The X3LYP Extended Density Functional for Accurate Descriptions of Nonbond Interactions, Spin States, and Thermochemical Properties. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673−2677.

(8) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom−Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.

(9) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2006**, *120*, 215−241.

(10) Zhao, Y.; Truhlar, D. G. Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions. *J. Phys. Chem. A* **2005**, *109*, 5656−5667.

(11) Zhang, Y.; Xu, X.; Goddard, W. A., III Doubly Hybrid Density Functional for Accurate Descriptions of Nonbond Interactions, Thermochemistry, and Thermochemical Kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 4963−4968.

(12) Schwabe, T.; Grimme, S. Double-Hybrid Density Functionals with Long-Range Dispersion Corrections: Higher Accuracy and Extended Applicability. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397−3406.

(13) Johnson, E. R.; Becke, A. D. A Post-Hartree−Fock Model of Intermolecular Interactions. *J. Chem. Phys.* **2005**, *123*, 024101.

(14) Sato, T.; Nakai, H. Density Functional Method Including Weak Interactions: Dispersion Coefficients Based on the Local Response Approximation. *J. Chem. Phys.* **2009**, *131*, 224104.

(15) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Phys. Rev. Lett.* **2009**, *102*, 073005.

(16) Becke, A. D.; Johnson, E. R. Exchange-Hole Dipole Moment and the Dispersion Interaction. *J. Chem. Phys.* **2005**, *122*, 154104.

(17) Becke, A. D.; Johnson, E. R. Exchange-Hole Dipole Moment and the Dispersion Interaction: High-Order Dispersion Coefficients. *J. Chem. Phys.* **2006**, *124*, 014104.

(18) Hebelmann, A. Derivation of the Dispersion Energy As an Explicit Density- and Exchange-Hole Functional. *J. Chem. Phys.* **2009**, *130*, 084104.

(19) Cooper, V. Van der Waals Density Functional: An Appropriate Exchange Functional. *Phys. Rev. B* **2010**, *81*, 161104.
(20) Dion, M.; Rydberg, H.; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.* **2004**, *92*, 246401.
(21) Klimes, J.; Bowler, D. R.; Michaelides, A. Chemical Accuracy for the van der Waals Density Functional. *J. Phys.: Condens. Matter* **2010**, *22*, 022201.
(22) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. Higher-Accuracy van der Waals Density Functional. *Phys. Rev. B* **2010**, *82*, 081101.
(23) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals Density Functional Made Simple. *Phys. Rev. Lett.* **2009**, *103*, 063004.
(24) Vydrov, O. A.; Van Voorhis, T. Nonlocal van der Waals Density Functional: The Simpler the Better. *J. Chem. Phys.* **2010**, *133*, 244103.
(25) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. Benchmark Database of Accurate (MP2 and CCSD(T) Complete Basis Set Limit) Interaction Energies of Small Model Complexes, DNA Base Pairs, and Amino Acid Pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(26) Antony, J.; Grimme, S. Density Functional Theory Including Dispersion Corrections for Intermolecular Interactions in a Large Benchmark Set of Biologically Relevant Molecules. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287−5293.
(27) Burns, L. A.; Vazquez-Mayagoitia, A.; Sumpter, B. G.; Sherrill, C. D. Density-Functional Approaches to Noncovalent Interactions: A Comparison of Dispersion Corrections (DFT-D), Exchange-Hole Dipole Moment (XDM) Theory, and Specialized Functionals. *J. Chem. Phys.* **2011**, *134*, 084107.
(28) Carter, D. J.; Rohl, A. L. Noncovalent Interactions in SIESTA Using the vdW-DF Functional: S22 Benchmark and Macrocyclic Structures. *J. Chem. Theory Comput.* **2012**, *8*, 281−289.
(29) Kannemann, F. O.; Becke, A. D. van der Waals Interactions in Density-Functional Theory: Intermolecular Complexes. *J. Chem. Theory Comput.* **2010**, *6*, 1081−1088.
(30) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. Basis Set Consistent Revision of the S22 Test Set of Noncovalent Interaction Energies. *J. Chem. Phys.* **2010**, *132*, 144104.
(31) Zhao, Y.; Truhlar, D. G. A Prototype for Graphene Material Simulation: Structures and Interaction Potentials of Coronene Dimers. *J. Phys. Chem. C* **2008**, *112*, 4061−4067.
(32) Rezác, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7* (8), 2427−2438.
(33) Hohenstein, E. G.; Chill, S. T.; Sherill, C. D. Assessment of the Performance of the M05-2X and M06-2X Exchange-Correlation Functionals for Noncovalent Interactions in Biomolecules. *J. Chem. Theory Comput.* **2008**, *4*, 1996−2000.
(34) Goerigk, L.; Grimme, S. A Thorough Benchmark of Density Functional Methods for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.
(35) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2011**, *7*, 291−309.
(36) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479−483.
(37) Sabatini, R.; Küçükbenli, E.; Kolb, B.; Thonhauser, T.; de Gironcoli, S. Structural Evolution of Amino Acid Crystals under Stress from a Nonempirical Density Functional. *J. Phys.: Condens. Matter* **2012**, *24* (42), 424209.
(38) Kong, L.; Roman-Perez, G.; Soler, J. M.; Langreth, D. C. Energetics and Dynamics of H_2 Adsorbed in a Nanoporous Material at Low Temperature. *Phys. Rev. Lett.* **2009**, *103*, 096103.
(39) Walker, A. M.; Civalleri, B.; Slater, B.; Mellot-Draznieks, C.; Cora, F.; Zicovich-Wilson, C. M.; Roman-Perez, G.; Soler, J. M.; Gale, J. D. Flexibility in a Metal−Organic Framework Material Controlled by Weak Dispersion Forces: The Bistability of MIL-53(Al). *Angew. Chem., Int. Ed.* **2010**, *122*, 7663−7665.
(40) Göltl, F.; Grüneis, A.; Hafner, J. Van der Waals Interactions between Hydrocarbon Molecules and Zeolites: Periodic Calculations at Different Levels of Theory, from Density Functional Theory to the Random Phase Approximation and Møller−Plesset Perturbation Theory. *J. Chem. Phys.* **2012**, *137*, 114111.
(41) Berland, K.; Borck, O.; Hyldgaard, P. Van der Waals Density Functional Calculations of Binding in Molecular Crystals. *Comput. Phys. Commun.* **2011**, *182* (9), 1800−1804.
(42) DelBen, M.; Hutter, J.; VandeVondele, J. Second-Order Møller−Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach. *J. Chem. Theory Comput.* **2012**, *8* (11), 4177−4188.
(43) Feng, S.; Li, T. Predicting Lattice Energy of Organic Crystals by Density Functional Theory with Empirically Corrected Dispersion Energy. *J. Chem. Theory Comput.* **2006**, *2*, 149−156.
(44) Nanda, K. D.; Beran, G. O. Prediction of Organic Molecular Crystal Geometries from MP2-Level Fragment Quantum Mechanical/Molecular Mechanical Calculations. *J. Chem. Phys.* **2012**, *137* (17), 174106.
(45) Neumann, M. A.; Perrin, M.-A. Energy Ranking of Molecular Crystals Using Density Functional Theory Calculations and an Empirical van der Waals Correction. *J. Phys. Chem. B* **2005**, *109* (32), 15531−15541.
(46) Otero-de-la-Roza, A.; Johnson, E. R. A Benchmark for Noncovalent Interactions in solids. *J. Chem. Phys.* **2012**, *137*, 054103.
(47) 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.
(48) Reilly, A. M.; Tkatchenko, A. Seamless and Accurate Modeling of Organic Molecular Materials. *J. Phys. Chem. Lett.* **2013**, *4* (6), 1028−1033.
(49) 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.
(50) Ayala, P. Y.; Scuseria, G. Electron Correlation in Large Molecular Systems Using the Atomic Orbital Formalism. The Case of Intermolecular Interactions in Crystalline Urea As an Example. *J. Comput. Chem.* **2000**, *21*, 1524−1531.
(51) Brandenburg, J. G.; Grimme, S. Dispersion Corrected Hartree−Fock and Density Functional Theory for Organic Crystal Structure Prediction . In *Prediction and Calculation of Crystal Structures. Methods and Applications*; Atahan-Evrenk, S., Aspuru-Guzik, A., Eds.; Springer: New York, 2014123
(52) Moellmann, J.; Grimme, S. DFT-D3 Study of Some Molecular Crystals. *J. Phys. Chem C* **2014**, *118*, 7615−7621.
(53) Maschio, L.; Usvyat, D.; Civalleri, B. Ab Initio Study of van der Waals and Hydrogen-Bonded Molecular Crystals with a Periodic Local-MP2 Method. *CrystEngComm* **2010**, *12*, 2429−2435.
(54) van de Streek, J.; Neumann, M. A. Validation of Experimental Molecular Crystal Structures with Dispersion-Corrected Density Functional Theory Calculations. *Acta Crystallogr., Sect. B: Struct. Sci.* **2010**, *B66*, 544−558.
(55) Sameera, W. M. C.; Pantazis, D. A. A Hierarchy of Methods for the Energetically Accurate Modeling of Isomerism in Monosaccharides. *J. Chem. Theory Comput.* **2012**, *8*, 2630−2645.
(56) Soler, J. M.; Artacho, E.; Gale, J. D.; Garcia, A.; Junquera, J.; Ordejon, P.; Sanchez-Portal, D. The SIESTA Method for Ab Initio Order-*N* Materials Simulation. *J. Phys.: Condens. Matter* **2002**, *14*, 2745−2779.
(57) Giannozzi, P.; Baroni, S.; Bonini, N.; Calandra, M.; Car, R.; Cavazzoni, C.; Ceresoli, D.; Chiarotti, G. L.; Cococcioni, M.; Dabo, I.; Dal, C. A.; de, G. S.; Fabris, S.; Fratesi, G.; Gebauer, R.; Gerstmann, U.; Gougoussis, C.; Kokalj, A.; Lazzeri, M.; Martin-Samos, L.; Marzari, N.; Mauri, F.; Mazzarello, R.; Paolini, S.; Pasquarello, A.; Paulatto, L.; Sbraccia, C.; Scandolo, S.; Sclauzero, G.; Seitsonen, A. P.; Smogunov, A.; Umari, P.; Wentzcovitch, R. M. QUANTUM ESPRESSO: a

Modular and Open-Source Software Project for Quantum Simulations of Materials. *J. Phys.: Condens. Matter* **2009**, *21* (39), 395502.
(58) Troullier, N.; Martins, J. L. Efficient Pseudopotentials for Plane-Wave Calculations. *Phys. Rev. B* **1991**, *43*, 1993−2006.
(59) Louwerse, M. J.; Rothenberg, G. Transferable Basis Sets of Numerical Atomic Orbitals. *Phys. Rev. B* **2012**, *85*, 035108.
(60) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50*, 17953−17979.
(61) Holzwarth, N. A. W.; Tackett, A. R.; Matthews, G. E. A Projector Augmented Wave (PAW) Code for Electronic Structure Calculations, Part I: Atompaw for Generating Atom-Centered Functions. *Comput. Phys. Commun.* **2001**, *135*, 329−347.
(62) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.
(63) Becke, A. D.; Johnson, E. R. A Unified Density-Functional Treatment of Dynamical, Nondynamical, and Dispersion Correlations. *J. Chem. Phys.* **2007**, *127*, 124108.
(64) Becke, A. D. On the Large-Gradient Behavior of the Density Functional Exchange Energy. *J. Chem. Phys.* **1986**, *85*, 7184−7187.
(65) *Materials Studio Release Notes*, Release 6.0; Accelrys Software Inc.: San Diego, 2013.
(66) Sun, H.; Ren, P.; Fried, J. R. The COMPASS Forcefield: Parameterization and Validation for Polyphosphazenes. *Comp. Theor. Polym. Sci.* **1988**, *8*, 229−246.
(67) Sun, H. COMPASS: An Ab Initio Forcefield Optimized for Condensed-Phase Applications—Overview with Details on Alkane and Benzene Compounds. *J. Phys. Chem. B* **1988**, *102*, 7338−7364.
(68) Allen, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.
(69) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553−566.
(70) Van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. State of the art in counterpoise theory. *Chem. Rev.* **1994**, *94*, 1873−1885.
(71) Carter, D. J.; Rohl, A. L. *Calculated Structures and Energies of Molecular Crystals Using Dispersion Corrected Density Functionals*; Curtin University: Bentley, WA, Australia. http://dx.doi.org/10.4225/06/4ED6B979EBFD5 (accessed June 18, 2014).
(72) Leach, A. R. *Molecular Modelling: Principles and Applications*. 2nd ed. ed.; Prentice Hall: New York, 2001.
(73) Blair, S. A.; Thakkar, A. J. How many Intramolecular Hydrogen Bonds Does the Oxalic Acid Dimer Have? *Chem. Phys. Lett.* **2010**, *495*, 198−202.
(74) Perdew, J. P.; Ruzsinszky, A.; Constantin, L. A.; Sun, J.; Csonka, G. I. Some Fundamental Issues in Ground-State Density Functional Theory: A Guide for the Perplexed. *J. Chem. Theory Comput.* **2009**, *5*, 902−908.
(75) Brandenburg, J. G.; Alessio, M.; Civalleri, 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.
(76) Oja, V.; Suuberg, E. M. Vapor Pressures and Enthalpies of Sublimation of D-Glucose, D-Xylose, Cellobiose, and Levoglucosan. *J. Chem. Eng. Data* **1999**, *44*, 26−29.
(77) Kabo, G. J.; Voitkevich, O. V.; Blokhin, A. V.; Kohut, S. V.; Stepurko, E. N.; Paulechka, Y. U. Thermodynamic Properties of Starch and Glucose. *J. Chem. Thermodyn.* **2013**, *59*, 87−93.
(78) Carter, D. J.; Rohl, A. L. *SIESTA Input and Output Files for Calculations on the S22 Data Set*; Curtin University: Bentley, WA, Australia. http://dx.doi.org/10.4225/06/4ED6B979EBEC4 (accessed June 2, 2014).