JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

Noncovalent Interactions in SIESTA Using the vdW-DF Functional: S22 Benchmark and Macrocyclic Structures

Damien J. Carter[†,‡] and Andrew L. Rohl[*,†,‡]

[†]Nanochemistry Research Institute, Department of Chemistry, Curtin University of Technology, GPO Box U1987, Perth, WA, Australia, 6845

[‡]iVEC, 26 Dick Perry Avenue, Technology Park, Kensington, WA, Australia 6151

ABSTRACT: We investigate the performance of the vdW-DF functional of Dion et al. implemented in the SIESTA code. In particular, the S22 data set and several calixarene-based host−guest structures are examined to assess the performance of the functional. The binding energy error statistics for the S22 data set reveal that the vdW-DF functional performs very well when compared to a range of other methods of treating dispersion in density functional theory, and to vdW-DF implementations in other codes. For the calixarene host−guest structures, the structural properties and binding energies are compared to previous experimental and computational studies, and in most cases we find that vdW-DF provides superior results to other computational studies.

INTRODUCTION

Ab initio quantum mechanical methods, in particular density functional theory (DFT), arguably offer the most accurate methods for determining the stability and properties of structures; however, they are limited by the size of systems that can be examined. A number of DFT codes, such as SIESTA,[1] use localized basis sets, pseudopotentials, and other features such as linear scaling that facilitate faster calculations, enabling investigations of much larger systems.

Up until a few years ago, these relatively fast DFT calculations were not being commonly used for soft matter or biomolecular or molecular crystals. This is because standard DFT methods lacked a description of van der Waals (vdW), or dispersion, forces, which can be large and important contributions in these types of systems. The "gold standard" of chemical accuracy in quantum methods is arguably the perturbatively corrected coupled cluster CCSD(T) method;[2] however, the computational effort scales by a formal cost (where N is the system size) of $O(N^7)$, which limits its applicability for these soft matter applications. As a result, in the past few years, there has been an explosion of methods that have been developed for DFT to provide ever increasingly more accurate descriptions of the dispersion forces and have been generally termed DFT-D[3−5] methods.

A popular proposal for including dispersion forces has been to add an empirical correction using interatomic potentials of the form C_6R^{-6},[3,6−9] with parameters derived from fitting to quantum mechanical calculations. These types of corrections are added directly to standard exchange correction functionals, and a "-D" is appended to the name, such as BLYP-D[3] and B97-D[4] or to hybrid functionals such as B3LYP-D.[8] A number of methods have been published that include medium-range dispersion forces in conventional semilocal DFT, with hybrid meta-GGA methods such as X3LYP,[10] ωB97X-D,[11] M06,[12] and PW6B95.[13] Another popular option has been to account for dispersion by incorporating correlation components from wave function theory, sometimes called double hybrid density functionals, with methods such as B2PLYP[14] and XYG3.[15]

Another approach has been to generate dispersion coefficients based on the exchange-hole dipole method (XDM). In the XDM[16−18] approach, dispersion interactions are modeled by examining the instantaneous dipole that arises between an electron and its exchange hole. A more complex approach involves the development of explicitly nonlocal correlation functionals from first principles. Examples of this include the vdW-DF,[19] vdW-DF2,[20] and VV09[21] functionals. These methods are potentially more accurate than the parametrized methods mentioned above, particularly for vdW interactions that depend on their chemical environment.[22] The vdW-DF functional has recently been implemented[22] in the SIESTA code and successfully been used to examine binding energies in double-walled carbon nanotubes[22] and to calculate the properties of metal organic framework (MOF) materials.[23,24]

In this work, we performed calculations on the S22 test set of molecules developed by Jurecka et al.[25] for assessing vdW interactions to assess the accuracy of results of the vdW-DF implementation in the SIESTA code. In particular, we examined the effect of several basis sets and the effect of geometry relaxation and compare the binding energies to literature reports for the S22 test set using a variety of different density functional methods, including comparisons to results from vdW-DF implementations in other software codes. We also calculated the structures of two calixarene inclusion compounds, namely, *p-tert*-butylcalix[4]arene·CS_2 and *p-tert*-butylcalix[4]arene·toluene, and compare the results to previous theoretical and experimental studies.

METHODOLOGY

All DFT calculations were performed using the SIESTA[1] code. DFT-D calculations were performed using vdW-DF,[19] as described by Roman-Perez and Soler.[22] For comparison, standard DFT calculations were also performed using the PBE[26] functional.

Received: September 27, 2011
Published: November 16, 2011

Norm-conserving pseudopotentials of Troullier and Martins[27] were used with the valence electron configurations of hydrogen 1s, carbon $2s^2 2p^2$, nitrogen $2s^2 2p^3$, oxygen $2s^2 2p^4$, and sulfur $3s^2 3p^4$. Hartree and exchange correlation energies were evaluated on a uniform real-space grid of points with a defined maximum kinetic energy of 300 Ry. For basis set generation, we used soft confinement potentials[28] to generate both double-ζ and triple-ζ plus polarization basis sets. Standard basis sets were generated, where the numerical atomic orbitals were radially confined to an extent that induces an energy shift in each orbital of 0.001 Ry (we refer to these as the DZ, DZP, or TZP basis for double-ζ, double-ζ polarized, and triple-ζ polarized basis sets). The other basis set was of triple-ζ polarized quality (herein referred to as TZP-L) and used an 8 Bohr cutoff for all orbital types (s, p, d, and f) and an explicit polarization orbital defined as a single-ζ of $(l + 1)$ angular momentum, also with an 8 Bohr cutoff.

We use the S22 set of complexes of common molecules to examine the accuracy of our vdw-DF calculations. The S22 set can be grouped into three subgroups based on their noncovalent interactions: (i) hydrogen-bonded complexes, (ii) complexes with predominant dispersion contributions, and (iii) mixed complexes in which electrostatic and dispersion contributions are similar in magnitude. The S22 reference geometries are taken from Jurecka et al.,[25] where all geometries were optimized at either the CCSD(T) or MP2 level (and in several cases where hydrogen positions were not reported for reference structures, they optimized the hydrogen atom positions at the DFT B3LYP level). The binding energies of the S22 set have recently been revised (the reference geometries are unchanged) by Takatani et al.,[29] who used a larger and more complete basis set than the original work, so we will compare our binding energies to both the original S22 binding energies of Jurecka et al.[25] and the new S22A binding energies of Takatani et al.[29] Our calculations of the S22 set are performed using both the reference geometries and with full geometry relaxations.

While many literature studies of test sets of compounds include an estimate for basis sets superposition errors (BSSE), usually via the counterpoise correction (CP) method, there are just as many literature studies that do not report any BSSE corrections with their binding energies. The counterpoise method can be problematic because, although it can improve the accuracy for very small basis sets (smaller than 6-31+G(d,p)), it can lead to less accurate results for moderate and large basis sets.[30] In more complex systems such as biopolymers, trimers, and other soft materials, the CP method can be impractical or ambiguous.[31] Grimme and co-workers[3,4,32] argue that with good quality basis sets, such as polarized triple-ζ basis sets, BSSE effects are small and are not required, and this is the approach that we will follow. We calculated a BSSE correction for several S22 examples and found it was of similar magnitude to that reported by Antony and Grimme,[32] supporting our decision not to report BSSE calculations in this paper.

For calculations of the structures of *p-tert*-butylcalix[4]arene·CS_2 and *p-tert*-butylcalix[4]arene·toluene, we use the TZP-L basis set and the other computational parameters listed above. We optimized the structures in both the gas phase and solid state (fixed at the experimental lattice parameters and fully relaxed), using the *p-tert*-butylcalix[4]arene·CS_2 crystal structure of Schatz et al.[33] and the *p-tert*-butylcalix[4]arene·toluene crystal structure of Arduini et al.[34]

RESULTS AND DISCUSSION

A. S22 Data Set. In Table 1, we report the binding energies for calculations of the S22 data set for a range of basis sets and the GGA functionals with (vdW-DF) and without (PBE) dispersion, at the reference and fully relaxed geometries. The reference energies in Table 1 are the S22A binding energies from Takatani et al.[29]

There are some clear trends in the binding energies in Table 1 with, for example, PBE overestimating the binding energies for hydrogen bonded complexes. As a visual aid to clearly show the trends in Table 1, we have plotted the difference between the binding energies from our calculations (ΔE) and the binding energies of the reference S22A data set (ΔE_{ref}), as shown in Figure 1. For the PBE functional, we see a clear trend of overestimation of binding energies for hydrogen bonded complexes and an underestimation for complexes with predominantly dispersion interactions or for mixed interaction compounds. This is the case for both the reference geometries and fully relaxed structures. Due to some fortuitous error cancellation, PBE still performs reasonably well for some compounds, such as the ammonia dimer or the ethene−ethyne dimer. Jurecka et al.[35] also reported similar behavior in their calculations of the S22 data set. GGA functionals like PBE are known to typically overestimate the strength of hydrogen bonds,[36] and we found similar behavior in our previous SIESTA calculations of the strongly hydrogen bound potassium dihydrogen phosphate system.[37,38]

For the vdW-DF functional using TZP and TZP-L basis sets, we find the opposite behavior of that of the PBE functional, in general underestimating the binding energy for hydrogen bonded complexes and overestimating for predominantly dispersion and mixed complexes, although the deviation is much less than for the PBE functional. Using the slightly smaller DZP basis sets, the vdW-DF generally overestimates (although some are still underbound) the binding of all complexes. Using the DZ basis set, the smallest basis set we examined, the results for the dispersion and mixed complexes are similar in magnitude and direction to the larger basis set results; however for the hydrogen bonded complexes, there is a large problem of overbinding. The fully relaxed geometries exhibit similar trends to those with the geometries fixed at the S22 reference geometry.

To quantify the trends in the binding energy results for the S22 data set as a whole, we compute three quantities—namely, the mean deviation (MD), mean absolute deviation (MAD), and root-mean-square deviation (RMSD). These are defined in the following way:

$$\mathrm{MD} = \frac{1}{n}\sum_{\mathrm{sys}}^{n}(\Delta E - \Delta E_{\mathrm{ref}}),$$

$$\mathrm{MAD} = \frac{1}{n}\sum_{\mathrm{sys}}^{n}|\Delta E - \Delta E_{\mathrm{ref}}|,$$

$$\mathrm{RMSD} = \sqrt{\frac{1}{n}\sum_{\mathrm{sys}}^{n}(\Delta E - \Delta E_{\mathrm{ref}})^2}$$

In Table 2, we compare the MD, MAD, and RMSD for our calculated binding energies based on the S22A reference energies and compare the results to a recent study by Burns et al.,[40] who reported a comprehensive investigation of the S22 data set with a wide range of exchange-correlation (XC) functionals. In Table 2, we report only a selection of different XC functional results from

Table 1. Binding Energies (kcal/mol) for the S22 Data Set[a]

			reference geometry					relaxed geometry				
	functional		PBE	vdW-DF				PBE	vdW-DF			
	basis set	ref	TZP	DZ	DZP	TZP	TZP-L	TZP	DZ	DZP	TZP	TZP-L
						hydrogen bonded complexes						
1	$(NH_3)_2$	−3.17	−3.46	−3.96	−3.10	−3.04	−3.01	−3.52	−3.91	−3.26	−3.14	−3.12
2	$(H_2O)_2$	−5.02	−6.49	−7.25	−5.57	−5.49	−5.36	−6.86	−8.22	−5.95	−5.76	−5.56
3	formic acid dimer ($C_{2\,h}$)	−18.61	−22.16	−23.04	−20.32	−18.53	−18.82	−22.46	−23.57	−18.58	−16.90	−17.68
4	formamide dimer ($C_{2\,h}$)	−15.96	−17.44	−18.60	−16.68	−15.16	−15.37	−17.78	−19.83	−16.36	−14.75	−15.46
5	uracil dimer ($C_{2\,h}$)	−20.65	−21.37	−24.17	−21.17	−19.64	−20.05	−21.22	−24.54	−19.99	−18.35	−19.32
6	2-pyridoxine-2-aminopyridine	−16.71	−18.57	−22.18	−17.85	−17.02	−17.36	−19.25	−23.56	−17.37	−16.47	−16.98
7	adenine−thymine WC	−16.37	−17.91	−23.61	−17.46	−16.42	−16.97	−18.28	−22.56	−17.09	−15.95	−16.58
						complexes with predominant dispersion contributions						
8	$(CH_4)_2$ (D_{3d})	−0.53	−0.48	−1.33	−1.13	−1.22	−0.96	−0.44	−3.91	−1.13	−1.21	−0.96
9	$(C_2H_4)_2$ (D_{2d})	−1.51	−0.77	−1.43	−1.50	−1.65	−1.59	−0.84	−8.22	−1.67	−1.73	−1.76
10	benzene−CH_4 (C_3)	−1.50	−0.31	−1.26	−1.37	−1.44	−1.78	−0.76	−23.57	−1.78	−1.85	−2.04
11	benzene dimer (C_{2h})	−2.73	0.02	−3.24	−3.23	−4.01	−3.59	−0.63	−19.83	−3.60	−4.18	−4.00
12	pyrazine dimer (C_s)	−4.42	−1.43	−5.45	−4.86	−5.31	−5.18	−1.93	−24.54	−5.31	−5.61	−5.57
13	uracil dimer (C_s)	−10.12	−6.04	−9.55	−11.87	−11.97	−12.09	−5.63	−23.56	−11.18	−10.82	−11.37
14	indole benzene	−5.22	−0.47	−4.97	−5.07	−6.00	−5.61	−2.16	−22.56	−5.57	−6.29	−6.01
15	adenine thymine (stack)	−12.23	−5.75	−12.58	−12.81	−13.38	−13.40	−6.06	−3.91	−13.19	−13.42	−13.43
						mixed complexes						
16	ethene−ethyne (C_{2v})	−1.53	−1.64	−1.33	−1.13	−1.94	−2.14	−1.80	−1.32	−1.68	−2.00	−2.21
17	benzene H_2O (C_s)	−3.28	−3.46	−1.43	−1.67	−4.24	−3.91	−3.72	−1.70	−4.50	−4.34	−3.95
18	benzene NH_3 (C_s)	−2.35	−1.75	−1.26	−1.78	−2.69	−2.80	−1.86	−1.63	−2.65	−2.82	−2.88
19	benzene HCN (C_s)	−4.46	−3.62	−3.24	−3.60	−4.37	−4.76	−3.71	−3.68	−4.18	−4.44	−4.84
20	benzene dimer (C_{2v})	−2.74	−1.24	−5.45	−5.31	−3.19	−3.44	−1.36	−5.91	−2.76	−3.22	−4.00
21	indole benzene T-shape	−5.73	−3.84	−9.55	−11.18	−5.87	−6.47	−3.82	−10.40	−5.64	−5.88	−6.39
22	phenol dimer	−7.05	−6.14	−4.97	−5.57	−7.29	−7.91	−6.60	−5.60	−7.19	−7.45	−8.26

[a] The reference binding energies (Ref.) are the S22A binding energies from Takatani *et al.*[29]

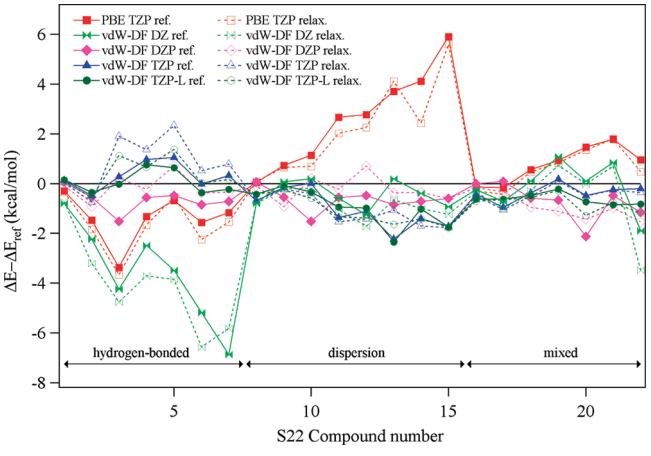


Figure 1. Binding energies differences ($\Delta E - \Delta E_{ref}$) for the S22 data set with respect to the S22A reference binding energies of Takatani et al.[29] The geometries of molecules are either fixed at the reference (ref.) S22 geometries of Jureka et al.[25] or fully relaxed (relax.).

Burns et al.,[40] in particular, only choosing results using a basis set (aug-cc-pVTZ) similar in size to the largest basis sets used in this study and choosing results without BSSE corrections.

The results in Table 2 highlight why using several error statistics can give a better gauge of the performance of a particular choice of XC functional and basis set, rather than just one. For example, if we only examined the MD of results, our vdW-DF TZP result for the full relaxed geometry gives the smallest deviation (−0.26 kcal/mol); however, when examining the MAD and RMSD results, we find it actually has slightly worse results than the DZP and TZP-L results for both the reference and fully relaxed geometries. Comparing the reference geometry results to the fully relaxed results we find that the errors are smaller for the results at the ideal reference geometries (by approximately 0.2 kcal/mol) when using the DZ, TZP, and TZP-L basis sets. Using the DZP basis sets, the results are very similar at the reference and fully relaxed geometries.

The overall results for the MAD and RMSD show that, when the vdW-DF functional is used, the DZP basis sets gives the best results, slightly outperforming the TZP-L basis set, which is slightly better than the TZP basis set results. Using the smallest DZ basis sets, the vdW-DF performs overall as badly, or worse, than the PBE functional, which was not designed to include van der Waals forces. Delving into the MAD and RMSD values in more detail, we can examine which basis set performs best for particular types of van der Waals complexes. For example, for the mixed complexes at the reference and fully relaxed geometries,

Table 2. Mean Deviation (MD), Mean Absolute Deviation (MAD), and Root Mean Square Deviation (RMSD) for S22 Data Set Binding Energies Using the S22A Reference Binding Energy Results of Takatani et al.[29] [a]

	MD	MAD	RMSD
	this study		
PBE TZP − ref.	0.76	1.68	2.23
vdW-DF DZ − ref.	−1.34	1.58	2.39
vdW-DF DZP − ref.	−0.53	0.58	0.78
vdW-DF TZP − ref.	−0.41	0.67	0.89
vdW-DF TZP-L − ref.	−0.53	0.67	0.85
PBE TZP − relax.	0.47	1.61	2.11
vdW-DF DZ relax.	−1.78	1.91	2.70
vdW-DF DZP relax.	−0.49	0.59	0.75
vdW-DF TZP relax.	−0.26	0.90	1.10
vdW-DF TZP-L relax.	−0.52	0.83	0.98
	GGA-D		
B97-D3[b]	0.14	0.36	0.44
BP86-D3[b]	−0.73	0.77	0.96
	hybrid functionals		
M05-2X[b]	0.42	0.64	0.85
M05-2X-D3[b]	−0.28	0.39	0.49
PBE0-D3[b]	−0.40	0.55	0.70
B3LYPD-D3 cc-pVTZ[b]	−0.32	0.37	0.49
wB97X-D cc-pVTZ[b]	−0.79	0.79	0.92
	other		
SCS-MP2[c]	−0.72	0.80	0.96
vdW-DF(revPBE)[d]	0.83	0.94	1.38
vdW-DF2(PW86)[d]	0.48	0.52	0.71

[a] Binding energy error statistics are in kcal/mol and do not include BSSE corrections. Our results use either the reference (ref.) or fully relaxed (relax.) geometries. [b] Results from Burns et al.[40] using an aug-cc-pVTZ basis set. [c] Results from Takatani et al.[29] using an aug-cc-pVTZ basis set. [d] Results from Lee et al.[20]

Table 3. Mean Deviation (MD), Mean Absolute Deviation (MAD), and Root Mean Square Deviation (MRSD) for S22 Data Set Binding Energies Using the S22 Reference Binding Energy Results of Jurecka et al.[25] [a]

	MD	MAD	RMSD
	this study		
PBE TZP − ref.	0.80	1.82	2.44
vdW-DF DZ − ref.	−1.30	1.60	2.48
vdW-DF DZP − ref.	−0.49	0.55	0.75
vdW-DF TZP − ref.	−0.36	0.56	0.73
vdW-DF TZP-L − ref.	−0.49	0.61	0.72
PBE TZP − relax.	0.51	1.74	2.32
vdW-DF DZ relax.	−1.74	1.87	2.77
vdW-DF DZP relax.	−0.44	0.53	0.64
vdW-DF TZP relax.	−0.21	0.75	0.94
vdW-DF TZP-L relax.	−0.48	0.73	0.83
	GGA-D		
BLYP-D[b]	−0.39	0.53	0.64
B3LYP-D[c]	−0.28	0.48	
B97-D[c]	0.44	0.50	
	hybrid functionals		
mPW2PLYP-D[d]	0.64	0.71	0.87
B2PLYP-D[d]	0.57	0.58	0.72
M05-2X[e]	0.29	0.57	0.72
M06-2X[f]	−0.14	0.44	0.56
	XDM		
PW86PBE-XDM[g]		0.46	
	other		
vdW-DF(revPBE)[h]	0.88	0.95	1.32
vdW-DF2(PW86)[h]	0.52	0.55	0.72
vdW-DF(revPBE)[i]	1.36	1.39	1.96
vdW-DF(PBE)[i]	−1.15	1.19	1.39
vdW-DF(revPBE)[j]		1.50	
vdW-DF(B86)[j]		0.53	
optB88-vdW[j]		0.23	

[a] Binding energy error statistics are in kcal/mol and do not include BSSE corrections. Our results use either the S22 reference (ref.) or fully relaxed (relax.) geometries. [b] Results from Antony and Grimme[32] using a TZV(2df,2pd) basis set. [c] Results from Chai and Head-Gordon[11] using a 6-311++G(3df,3pd) basis set. [d] Results from Scwabe and Grimme[14] using a TZV(2df,2pd) basis set. [e] Results from Zhao and Truhlar[30] using a 6-31+G(d,p) basis set. [f] Results from Zhao and Truhlar[41] using a 6-311+G(2df,2p) basis set. [g] Results from Kannemann and Becke[39] using the basis-set-free Numol method.[42] [h] Results from Lee et al.[20] [i] Results from Gulans et al.[43] using a TZP basis set. [j] Results from Klimes et al.[44] using a aug-cc-pVTZ basis set.

the TZP basis performs best. For hydrogen-bonded complexes, the TZP-L basis performs best at the reference geometry, and the DZP basis performs best at the fully relaxed geometry. Burns et al.[40] found that a DZP type basis performed slightly better for hydrogen bonded complexes, while a TZP type basis performed better for the dispersion complexes. These results illustrate how the choice of basis set and the type of intermolecular interaction affects the binding energy results.

Comparing the S22A error statics in Table 2, we find that the vdW-DF in SIESTA performs well when compared to the subset chosen here of the many XC functional results reported by Burns et al.,[40] SCS-MP2 results of Takatani et al.,[29] and the vdW-DF/2 results of Lee et al.[20] In particular the MD, MAD, and RMSD for our vdW-DF calculations of the reference structures are similar in magnitude to the M05-2X and wB97X-D cc-pVTZ hybrid XC functionals and are slightly better than the SCS-MP2 and BP86-D3 results. The vdW-DF error statistics are slightly worse than the B97-D3 GGA functional results and the B3LYPD-D3 and M05-2X-D3s hybrid functional results. Lee et al.[20] report MAD and RMSD values of 0.94 and 1.38 kcal/mol, respectively, for their vdW-DF (revPBE) implementation. Our MAD and RMSD results from SIESTA are slightly better using the DZP, TZP, and TZP-L basis sets, both at the reference and the fully relaxed geometries, indicating that the performance of the vdW-DF in SIESTA is very good. Although the form of the vdW-DF (revPBE) functional is the same, these differences may arise due to the implementation in the respective codes, the type of basis set (e.g., planewave or localized orbitals), the quality of basis set, or the type of pseudopotentials used or other factors. Compared to the vdW-DF2 results of Lee et al.,[20] our results are slightly worse.

The newly revised S22A binding energies for the S22 data set were recently published by Sherrill and co-workers.[29,40] S22

Table 4. Comparison of the Structural Parameters for *p-tert*-Butylcalix[4]arene·CS_2 from Theory and Experiment[a]

	gas phase			solid state			
				ref.	relax.	ref.	
	vdW-DF	PBE[b]	HF[c]	vdW-DF	vdW-DF	PBE[b]	exptl[d]
a (Å)					12.911		12.770
b (Å)					12.811		12.770
c (Å)					13.185		13.314
Ar−O_4 θ_1 (deg)	120	126	125	120, 120	121, 121	125, 125	122
Ar−O_4 θ_2 (deg)	123	126	125	123, 123	125, 124	125, 125	122
Ar−O_4 θ_3 (deg)	121	126	125	124, 123	123, 124	124, 125	122
Ar−O_4 θ_4 (deg)	123	126	125	124, 124	124, 124	124, 124	122
$C(CS_2)$−O_4 (Å)	5.75	5.93	6.51	5.64, 5.66	5.58, 5.64	5.75, 5.80	5.45, 5.45
C−S (Å)	1.59, 1.61	1.59, 1.60	1.55	1.59, 1.61	1.59, 1.61	1.56, 1.58	1.50,1.61
O_H−O_H (Å)	2.70, 2.71		2.73	2.69, 2.71	2.70, 2.71		2.68
O_4-CS_2 (deg)	4.6	1.2	0.0	2.4, 2.5	2.9, 3.2	0.9, 1.2	0.0
I^b (%)	89		57	89, 89	89, 90		92
d_A (Å)	8.20, 8.43		8.48	8.39, 8.49	8.40, 8.58		8.34

[a] Theoretical structures are either fixed at the reference experimental lattice parameters (ref.) or have fully optimized lattice parameters and atomic coordinates (relax.). Ar−O4 angles are the angles between the four aromatic rings (Ar) and the O4 plane, which is defined by the four phenol oxygen atoms (O_H) in the calixarene. $C(CS_2)$−O_4 is the height of the CS_2 molecule above the O_4 plane. C−S is the CS_2 molecule bond lengths, with S_1 being the inner sulphur atom and S_2 being the outer atom. O_4−CS_2 is the tilt between the long axis of the CS_2 molecule and the O_4 plane. The distance d_A is the separation between the top-most carbon atoms of the aromatic rings (Ar) on opposite sides of the calixarene cage. The inclusion percentage (I^b) is calculated using $I^b = (\text{dist}(S_1-S_2) - \text{dist}(S2-\text{plane}_b))/(\text{dist}(S_1-S_2))$, where plane_b represents the top of the calixarene cavity. [b] Results from Ogden et al.[46] [c] Results from Schatz et al.[45] [d] Crystal structure from Schatz et al.[33]

literature studies before these S22A results were published used the reference binding energies from Jurecka et al.[25] To enable direct comparison of our results to these earlier published studies, in Table 3 we compare the MD, MAD, and RMSD for our calculated binding energies based on the S22 binding energies reported by Jurecka et al.[25]

The literature S22 binding error statistics in Table 3 are by no means an attempt to collate all studies in the computational literature; we have simply made a selection of the wide ranging literature with examples of different types of dispersion implementations such as XDM or hybrid methods for which to compare our results. Many of the studies[14,32,35,40] we referenced in Table 3 reported results for other density functionals and also reported results for a variety of basis set sizes or other computational parameters.

Examining the error statistics in Table 3, in particular, the MAD and RMSD values, we find that at the reference geometries, our vdW-DF results compares well to other DFT-D methods. The MAD values of approximately 0.6 kcal/mol for the DZP, TZP, and TZP-L basis sets are similar to, or better than, the majority of literature results reported in Table 3. As mentioned before, using the fully relaxed geometries, the results are slightly worse for the DZ, TZP, and TZP-L basis sets. Comparing our results to the vdW-DF results of Lee et al.,[20] this time for the S22 binding energies, we see similar behavior to that for the S22A binding energies in Table 2 with our MAD and RMSD values being slightly better, both at the reference and the fully relaxed geometries. Again, our results are slightly worse compared to the vdW-DF2 results. Gulans et al.[43] reported their own implementations of vdW-DF(revPBE) and vdW-DF(PBE) in the SIESTA code, and their results appear to be much less accurate than our results at both the reference and fully relaxed geometries, although these values used a CP correction, which will alter these values, particularly depending on their basis set choices etc. Klimes et al.[44] examined the vdW-DF based on six different GGA functionals, finding the MAD values varied from 1.50 kcal/mol for the vdwDF(revPBE) functional to 0.53 kcal/mol for the vdwDF-(B86) functional. At the reference geometries, our MAD values of 0.56 and 0.61 kcal/mol for the two basis set options indicate that the vdw-DF(revPBE) implementation in SIESTA performs extremely well compared to these results. Klimes et al.[44] also report MAD values for several new optimized functionals, with the optB88-vdW functional performing best with a MAD of 0.23 kcal/mol.

A particular advantage of SIESTA is that the localized numerical basis set enables us to examine (with full relaxation of unit cells and atomic coordinates) systems such as large biomolecules and soft matter systems, which would not be accessible to a planewave basis set or using high-order correlated based methods. The compromise for this speed and ability to examine large systems is accuracy; however, our results for the S22A and S22 binding energies in Tables 1−3 show that the vdW-DF implementation in SIESTA overall performs extremely well when using good quality DZP or TZP basis sets. In particular, when compared to vdW-DF implementations in other software codes reported by Lee et al.[20] and Klimes et al.,[44] the SIESTA MAD and RMSD values were better than or at least equally as good as the other implementations. A revised version of the vdW-DF has recently been published and is termed vdW-DF2[20] and reports improved accuracy for the S22 data set as reported above. In the future, if this is implemented in SIESTA, this could lead to even better performance for the S22 data set in SIESTA. There is little difference between the binding energies for the DZP, TZP, and TZP-L basis sets at the reference S22 geometries; however, when using fully relaxed geometries, the DZP basis set performs slightly better than the TZP-L and TZP basis sets. On the basis

of these results, we then examined the structure of two calixarene inclusion compounds.

B. *p-tert*-Butylcalix[4]arene·CS_2 and *p-tert*-Butylcalix[4]arene·toluene. We have performed vdW-DF calculations on the *p-tert*-butylcalix[4]arene structure with toluene and carbon disulfide (CS_2) guest molecules to examine the performance of the vdW-DF for calixarene host−guest structures. In Table 4, we report the structural parameters of the *p-tert*-butylcalix[4]arene·CS_2 compound in both the gas phase and the solid state. We examined the solid state calixarene structure fixed at the experimental lattice parameters of Schatz et al.[33] (allowing atomic coordinates to relax), and with fully optimized lattice parameters and atomic coordinates, to allow for comparison with other theoretical investigations.[45,46]

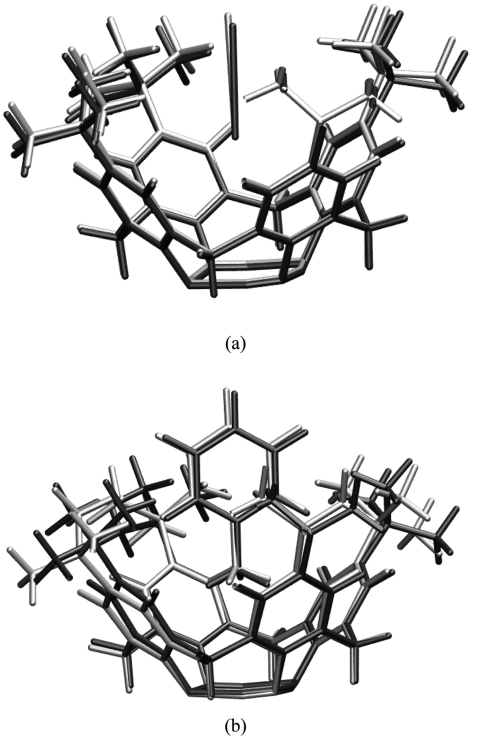
The fully optimized *p-tert*-butylcalix[4]arene·CS_2 structure in Table 4 shows good agreement with the room temperature experimental crystal structure of Schatz et al.[33] For the fully optimized crystal structure, there is a small distortion from a tetragonal to orthorhombic unit cell, although the lattice parameters are within 1% of the experimental values, and the rest of the structural properties are in good agreement with those calculated at the experimental lattice parameters.

Comparing the values in Table 4, the main differences that appear in both the gas phase and solid state are the $C(CS_2)-O_4$ distance and the inclusion percentage (I^b), both of which give a measure of how much the CS_2 guest molecule is included within the calixarene cage (a higher inclusion percentage will have a smaller $C(CS_2)-O_4$ distance and vice versa). For both I^b and $C(CS_2)-O_4$ values, the vdW-DF results in the gas phase and solid state phase are much closer to the experimental values than the other theoretical calculations.

The inclusion percentage for the vdW-DF calculations is approximately 90% in both the gas phase and solid state, very close to the experimental value of 92% and noticeably higher than the HF value of 57%. Correspondingly, the $C(CS_2)-O_4$ distances are shorter in the vdW-DF calculations with 5.75 Å in the gas phase to 5.58−5.66 Å in the solid state (reference and relaxed geometries, respectively) for the vdW-DF. The PBE results for the $C(CS_2)-O_4$ distances are 5.93 and 5.75−5.80 Å for the gas phase and solid state, respectively, and 6.51 Å for the gas phase HF results calculations. In theoretical calculations using the vdW-DF and PBE functionals, there is a small tilt of the CS_2 molecule of several degrees with respect to the O4 plane, which is not seen in the experimental structure or HF calculations. In Figure 2a, we show the gas phase (light gray) and fully relaxed solid state structures (dark gray) of *p-tert*-butylcalix[4]arene·CS_2, showing a single calixarene unit overlaid for each. In this figure, the CS_2 molecule is clearly further into the cage in the solid state (dark gray) structure than the gas phase structure (light gray), as quantified earlier by the $C(CS_2)-O_4$ distances. There are also some small differences in the orientations of the methyl side chains of the *tert*-butyl groups.

In Table 5, we report the structural parameters of the *p-tert*-butylcalix[4]arene·toluene compound in both the gas phase and the solid state. We examined the solid state calixarene structure fixed at the experimental $P112/a$ crystal structure reported by Arduini et al.[34] (allowing atomic coordinates to relax) and with fully optimized lattice parameters and atomic coordinates.

The experimental crystal structure of *p-tert*-butylcalix[4]arene·toluene has been reported in several different crystal structures at a range of temperatures. The high temperature structures were reported as being in the tetragonal $P4/n$ space



(a)

(b)

Figure 2. Superposition of the gas phase and solid state (fully relaxed) structures of (a) *p-tert*-butylcalix[4]arene·CS_2 and (b) *p-tert*-butylcalix[4]arene·toluene. The gas phase structures and solid state crystal structures are colored light gray and dark gray, respectively.

group;[47] then later structural refinements reported monoclinic structures.[34,48] In Table 5, we report the two monoclinic structures for comparison to our computational results. In the monoclinic structures, the toluene guest molecules induce a distortion in the host calixarene molecules, and there is a correlation of guest molecules in adjacent calixarene molecules.[48] If no distortion occurred, the four $Ar-O_4$ angles in Table 5 would be the same value. The actual position of the toluene molecule in the experimental structures is dynamically disordered and typically averages to give either a 4-fold[47] or 2-fold[34,48] symmetry. In this study, we perform static electronic structure energy minimizations. Molecular dynamics simulations would be required to investigate the dynamical nature of the guest molecule, but this was not the focus of this investigation.

Examining the gas phase structures in Table 5, we find that the vdW-DF structure is slightly closer to the experimental structures than the PBE results of Ogden et al.,[46] in particular for the $Ar-O_4$ angles and the CH_3-O_4 distance. The CH_3-O_4 distance is a measure of how much the toluene molecule is included within the host molecule, so the PBE value of 3.93 Å, compared to the vdW-DF value of 3.77 Å, indicates that the toluene is more included in the vdW-DF calculations.

Looking now at the solid state structures there is quite reasonable agreement between the vdW-DF and PBE results[46] with the experimental structures. For the vdW-DF results at the experimental (ref.) geometry, the CH_3-O_4 distances are 3.61 and 3.66 Å, very similar to the PBE results of 3.64 and 3.66 Å. The experimental monoclinic structures report a similar lower value in the range of 3.65−3.70 Å, but the second distance is about 0.1 Å longer than the vdW-DF and PBE results. The interplanar tilt of the toluene molecules is also slightly larger in the vdW-DF

Table 5. Comparison of the Structural Parameters for *p-tert*-Butylcalix[4]arene·toluene from Theory and Experiment[a]

	gas phase		solid state				
			ref.	relaxed	ref.		
	vdW-DF	PBE[b]	vdW-DF	vdW-DF	PBE[b]	exptl.[c]	exptl.[d]
a (Å)				18.239	12.756	17.889	17.814
b (Å)				18.127	12.756	17.899	17.806
c (Å)				13.647	13.793	13.827	13.890
Ar−O_4 θ_1 (deg)	118	123	117, 119	119, 120	123, 121	119	120
Ar−O_4 θ_2 (deg)	124	132	124, 122	125, 125	127, 128	124	125
Ar−O_4 θ_3 (deg)	118	124	120, 120	121, 121	122, 123	120	120
Ar−O_4 θ_4 (deg)	125	127	126, 126	125, 124	126, 126	125	125
CH_3−O_4 (Å)	3.77	3.93	3.61, 3.66	3.63, 3.67	3.64, 3.66	3.70,3.84	3.64, 3.83
toluene tilt (deg)	0.3	3.5	3.7, 4.1	4.2, 5.5	1.1, 3.8	0	4.2, 7.7
interplanar angle	19.1		23.7, 26.6	23.3, 26.3	19.0, 24.8	19.6, 23.1	18.0, 21.9

[a] Theoretical structures are either fixed at the reference experimental lattice parameters (ref.) or have fully optimized lattice parameters and atomic coordinates (relax.). Ar−O4 angles are the angles between the four aromatic rings (Ar) and the O4 plane, which is defined by the four phenol oxygen atoms (O_H) in the calixarene. CH_3−O_4 is the height of the toluene molecule above the O_4 plane. The toluene tile is the angle between the long axis of the toluene molecule and the O_4 plane. The interplanar angle is the angle between the pseudo-mirror plane of two of the methylene C atoms and the plane of the aromatic ring in the toluene molecule. [b] Results from Ogden et al.[46] using the $P4/n$ crystal structure of Andreeti et al.,[47] where $\alpha = \beta = \gamma = 90°$. [c] Crystal structure from Arduini et al.[34] using the $P112/a$ space group. [d] Crystal structure from Enright et al.[48] using the $P2/c$ space group, where $\alpha = \beta = 90$ and $\gamma = 89.91°$.

Table 6. Binding Energy (kcal/mol) per Guest Molecule of *p-tert*-Butylcalix[4]arene·toluene (Toluene) and *p-tert*-Butylcalix[4]arene·CS_2 (CS_2) for the Gas Phase and Solid State Structures[a]

	gas phase	
	vdW-DF	PBE DZP[b]
CS_2	−13.27	−0.76
toluene	−21.78	

	solid state		
	vdW-DF		PBE DZP[b]
	fixed at exptl unit cell	relaxed unitcell	fixed at exptl unit cell
CS_2	−83.42	−83.59	−1.84
toluene	−92.27	−92.67	

[a] The solid state structures are those reported in Tables 4 and 5. [b] Results from Ogden et al.[46] using the $P4/n$ crystal structure of Andreeti et al.,[47] where $\alpha = \beta = \gamma = 90°$.

calculation, with angles of 23.7 and 26.6°, compared to experimental values ranging from 18 to 23°. When we fully relaxed the solid state structure, we found that the lattice parameter changes were less than 2% different compared to the experimental crystal structure. The CH_3−O_4 distances of 3.63 and 3.67 Å are similar to the result fixed at the experimental lattice parameters and are again slightly less than the experimental results. The interplanar angles of 23.3 and 26.3° are slightly larger than for the fixed lattice parameters (values of 23.7 and 26.1°) but are slightly overestimated compared with the experimental values. Overall, the PBE and vdW-DF functionals appear to perform quite similarly for the structural parameters of the solid state structures, while the vdW-DF appears to perform slightly better for the gas phase structure. In Figure 2b, we show the gas phase (light gray) and fully relaxed solid state structures (dark gray) of *p-tert*-butylcalix[4]arene·toluene, showing a single calixarene unit overlaid for each. In this figure, the toluene molecule sits further into the cage in the solid state (dark gray) structure than the gas phase structure (light gray), as quantified earlier by the CH_3−O_4 distances. There are also some small differences in the orientations of the methyl side chains of the *tert*-butyl groups.

After examining the structural properties of the *p-tert*-butylcalix[4]arene·CS_2 and *p-tert*-butylcalix[4]arene·toluene compounds, we now examine the binding energies. In Table 6, we compare the binding energies of the *p-tert*-butylcalix[4]arene·toluene and *p-tert*-butylcalix[4]arene·CS_2 compounds in both the gas phase and solid state, defining the binding energy as per guest molecule, with reference to the isolated gas phase species. As reported earlier with the S22 data set, we do not include any BSSE corrections in these binding energies.

The binding energy results in Table 6 clearly show that the binding of toluene and CS2 molecules is weaker in the gas phase than the solid state. This is not unsurprising as the cavity in the solid state is well-defined and the guest molecules are also influenced by binding with neighboring calixarene host molecules. For the *p-tert*-butylcalix[4]arene·CS_2 structure, the binding energy in the gas phase is −13.27 kcal/mol, while in the solid state, the values are −83.42 and −83.59 kcal/mol, respectively, for the fixed and fully relaxed structures, respectively. For the *p-tert*-butylcalix[4]arene·toluene in the gas phase, the binding energy is −21.78 kcal/mol, compared to −92.27 and −92.67 kcal/mol for the fixed and fully relaxed solid state structures, respectively. Previous PBE calculations of Ogden et al.[46] of the *p-tert*-butylcalix[4]arene·CS_2 structure also predicted this behavior with values of −0.76 and −1.84 kcal/mol for the gas phase and solid state structures, respectively, although they acknowledge the magnitude of the binding energies will be underestimated due to a lack of suitable description of the dispersion forces. Overall, the vdW-DF in SIESTA appears to provide good

descriptions of both the gas phase and solid state structures of *p-tert*-butylcalix[4]arene·CS_2 and *p-tert*-butylcalix[4]arene·toluene, based on structural properties reported in Tables 4 and 5 and the binding energies reported in Table 6 and in most cases provide superior results to other theoretical studies.

CONCLUSIONS

We have investigated the performance of the vdW-DF implemented in the SIESTA code for the S22 data set, examining the effect of basis set choice and atomic relaxation. Using the MD, MAD, and RMSD error statistics, we have quantified the results against both the S22[25] and S22A[39] reference binding energies and find that SIESTA performs very well compared to a range of results from other studies. We find that at the reference geometries there is little difference between the DZP, TZP, and TZP-L basis sets; however, when full atomic relaxations are carried out, the DZP basis overall gives the best results by approximately 0.2 kcal/mol. Dividing the S22 compounds into their dominant van der Waals interactions (hydrogen bonded, dispersion dominated, or mixed), we also demonstrated how different basis sets performed better for particular types of van der Waals interactions.

We then examined the performance of vdW-DF for two calixarene host−guest inclusion compounds, namely, *p-tert*-butylcalix[4]arene·CS_2 and *p-tert*-butylcalix[4]arene·toluene. We examined both the gas phase and solid state structures and compared the results against other theoretical investigations and experimental data. Overall, for both inclusion compounds, vdW-DF performs extremely well and outperforms other Hartree−Fock and DFT PBE results where published. In particular, for the *p-tert*-butylcalix[4]arene·CS_2 compound, the vdW-DF calculated structures show that the CS_2 molecule is included further into the host compound than other theoretical calculations and much better matches the experimental structures. Binding energy calculations show that the guest molecules are much more strongly bound in the solid state than the gas phase, as expected.

AUTHOR INFORMATION

Corresponding Author

*E-mail: a.rohl@curtin.edu.au.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

The authors thank iVEC and the NCI for the provision of computational resources. A complete archive of the SIESTA calculations of the S22 data set is available at http://dx.doi.org/10.4225/06/4ED6B979EBEC4.

REFERENCES

(1) Soler, J. M.; Artacho, E.; Gale, J. D.; Garcia, A.; Junquera, J.; Ordejon, P.; Sanchez-Portal, D. *J. Phys.: Condens. Matter* **2002**, *14*, 2745–2779.

(2) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phy. Lett.* **1989**, *157*, 479–483.

(3) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(4) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(5) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(6) Williams, R. W.; Malhotra, D. *Chem. Phys.* **2006**, *327*, 54–62.

(7) Riley, K. E.; Vondrasek, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5555–5560.

(8) Civalleri, B.; Zicovich-Wilson, C. M.; Valenzano, L.; Ugliengo, P. *CrystEngComm* **2008**, *10*, 405–410.

(9) Neumann, M. A.; Leusen, F. J. J.; Kendrick, J. *Angew. Chem., Int. Ed.* **2008**, *47*, 2427–2430.

(10) Xu, X.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673–2677.

(11) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620.

(12) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2006**, *120*, 215–241.

(13) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 5656–5667.

(14) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406.

(15) Zhang, Y.; Xu, X.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S. A.* **2009**, *106*, 4963–4968.

(16) Hebelmann, A. *J. Chem. Phys.* **2009**, *130*, 084104.

(17) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 154104.

(18) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2006**, *124*, 014104.

(19) Dion, M.; Rydberg, H.; Schroder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.

(20) Lee, K.; Murray, E. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.

(21) Vydrov, O. A.; Van Voorhis, T. *Phys. Rev. Lett.* **2009**, *103*, 063004.

(22) Roman-Perez, G.; Soler, J. M. *Phys. Rev. Lett.* **2009**, *103*, 096102.

(23) Kong, L.; Roman-Perez, G.; Soler, J. M.; Langreth, D. C. *Phys. Rev. Lett.* **2009**, *103*, 096103.

(24) 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. *Angew. Chem., Int. Ed.* **2010**, *122*, 7663–7665.

(25) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.

(26) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(27) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993–2006.

(28) Junquera, J.; Paz, O.; Sanchez-Portal, D.; Artacho, E. *Phys. Rev. B* **2001**, *64*, 235111.

(29) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 144104.

(30) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289–300.

(31) van Mourik, T.; Karamertzanis, P. G.; Price, S. L. *J. Phys. Chem. A* **2006**, *110*, 8–12.

(32) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293.

(33) Schatz, J.; Scholdbach, F.; Lentz, A.; Rastatter, S.; Schilling, J.; Dormann, J.; Ruoff, A.; Debaerdemaeker, T. *Z. Naturforsch., B: J. Chem. Sci.* **2000**, *55*, 213–221.

(34) Arduini, A.; Caciuffo, R.; Geremia, S.; Ferrero, C.; Ugozzoli, F.; Zontone, F. *Supramol. Chem.* **1998**, *10*, 125–132.

(35) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555–569.

(36) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*, 2nd ed.; Wiley-VCH Verlag GmbH: Weinheim, Germany, 2001.

(37) Carter, D. J.; Rohl, A. L.; Gale, J. D. *J. Chem. Theory Comput.* **2006**, *2*, 797–800.

(38) Carter, D. J.; Rohl, A. L. *J. Chem. Theory Comput.* **2011**, *7*, 1604–1609.

(39) Kannemann, F. O.; Becke, A. D. *J. Chem. Theory Comput.* **2010**, *6*, 1081–1088.

(40) Burns, L. A.; Vazquez-Mayagoitia, A.; Sumpter, B. G.; Sherrill, C. D. *J. Chem. Phys.* **2011**, *134*, 084107.

(41) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. C* **2008**, *112*, 4061–4067.

(42) Becke, A. D.; Dickson, R. M. *J. Chem. Phys.* **1998**, *89*, 2993–2997.

(43) Gulans, A.; Puska, M. J.; Nieminen, R. M. *Phys. Rev. B* **2009**, *79*, 201105.

(44) Klimes, J.; Bowler, D. R.; Michaelides, A. *J. Phys.: Condens. Matter* **2010**, *22*, 022201.
(45) Schatz, J.; Backes, A. C.; Siehl, H. *J. Chem. Soc., Perkin Trans. 2* **2000**, *2*, 609–610.
(46) Ogden, M. I.; Rohl, A. L.; Gale, J. D. *Chem. Commun.* **2001**, 1626–1627.
(47) Andreeti, G. D.; Ungaro, R.; Pochini, A. *J. Chem. Soc., Chem. Commun.* **1979**, 1005–1007.
(48) Enright, G. D.; Brouwer, E. B.; Udachin, K. A.; Ratcliffe, C. I.; Ripmeester, J. A. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 1032–1035.