

# Hydrogen Bonding Described Using Dispersion-Corrected Density Functional Theory

J. Samuel Arey,<sup>\*,†</sup> Philippe C. Aeberhard,<sup>‡</sup> I-Chun Lin,<sup>§</sup> and Ursula Rothlisberger

Laboratory of Computational Chemistry and Biochemistry, Swiss Federal Institute of Technology at Lausanne (EPFL), Switzerland

Received: November 24, 2008; Revised Manuscript Received: January 28, 2009

In recent works, dispersion-corrected atom-centered potentials (DCACPs) were developed as a method to account for long-range dispersion forces between molecules in density functional theory calculations within the generalized gradient approximation (GGA). Here, we test the ability of DCACPs to improve the GGA treatment of hydrogen-bonded systems. We assessed both BLYP and dispersion-corrected BLYP with respect to benchmark calculations for the hydrogen bond lengths and binding energies of 20 complexes containing the elements C, H, N, O, and S. Benchmark data included geometries calculated using MP2 and CCSD(T) and binding energies using W2, W1, CBS-QB3, and other CCSD(T) extrapolation schemes. With respect to benchmark methods, dispersion-corrected BLYP exhibited a mean signed error of 0.010 Å in the hydrogen bond length and a mean relative error of 5.1% in the hydrogen bond binding energy. By comparison, uncorrected BLYP exhibited error statistics of 0.036 Å and 15.9%, respectively. We conclude that DCACPs robustly improve the BLYP description of hydrogen-bonded systems at small additional computational cost. New benchmark geometries (MP2/aug-cc-pVTZ) and new benchmark binding energies (W1) are presented for seven complexes, and the remaining benchmark data were taken from previous literature.

## Introduction

The phenomenon of hydrogen bonding attracts wide scientific interest. By its early definition, a hydrogen bond is a pairwise and predominantly ionic interaction that forms between an electronegative atom and a hydrogen covalently bound to another electronegative atom.<sup>1</sup> Although originally defined as noncovalent, hydrogen bonds are highly directional and are generally stronger than pure dispersion or dipole–dipole-based interactions, and it has been argued that hydrogen bonds include some degree of covalent character.<sup>2–5</sup> Hydrogen bonds exhibit Gibbs free energies of interaction typically in the range of –3 to –15 kcal/mol for neutral partners.<sup>6,7</sup> Consequently, they are central participants in determining the intermolecular order and forces that govern diverse natural and engineered systems, particularly for condensed phases. Hydrogen bonds are manifest in crystal and clathrate lattice networks and in liquid radial distribution functions,<sup>7</sup> and they affect bulk phase properties such as melting point and boiling point.<sup>8</sup> Moreover, hydrogen bonds deeply influence protein secondary structure, as well as the structures and self-assembly of DNA, membranes, and micelles.<sup>9</sup>

Weak interactions, including hydrogen bonds, pose challenges to density functional theory (DFT<sup>10</sup>) calculations using generalized gradient approximation (GGA<sup>10</sup>) models. Difficulties occur partly because GGAs lack the formalism needed to describe properly long-range electron correlation forces.<sup>11–13</sup> In Zhao and Truhlar's recent assessments<sup>14,15</sup> of DFT performances for weakly bound complexes, most GGA functionals systematically underpredicted dispersion forces between molecules. But GGAs

exhibited less bias and better relative accuracy for hydrogen bond binding energies.<sup>15,16</sup> Nonetheless, long-range electron correlation interactions, attributed to dispersion forces, can contribute significantly to hydrogen bond binding energies.<sup>17–21</sup> Hence, in the conventional GGA descriptions of hydrogen bonds, it is presumably the case that electrostatic, exchange, and induction effects compensate for the spurious or insufficient accounting of long-range electron correlation forces. Finally, it is worth mentioning that some meta-GGAs and hybrid-meta-GGAs handle long-range electron correlation effects better than GGAs, and meta-GGAs and hybrid-meta-GGAs are frequently more accurate on average than GGAs for a range of weak interactions.<sup>15,16,22</sup> But many applications in biochemistry, materials science, and other applied chemistry disciplines benefit from the computational efficiency<sup>23</sup> afforded by the GGA model.<sup>24–26</sup> The broad application of GGA functionals to condensed phase systems illustrates the need for accurate treatment of weak interactions at this level of theory.

Recent works have shown that application of an empirical nonlocal correction significantly improves GGA-level predictions of dispersion forces in weakly bound systems. In these studies, a “dispersion-corrected” atom-centered potential (DCACP) was added to the total electronic potential in order to correct for the absence of long-range electron correlation in the GGA expression.<sup>27–29</sup> The DCACP parameters were fitted to reproduce dispersion interactions for a small calibration set,<sup>27,28,30</sup> thereby accounting for a long-range interatomic force resembling dispersion.<sup>31</sup> Although DCACPs are tuned empirically, this multicenter functional correction can be laid on a rigorous theoretical foundation.<sup>32</sup> The resulting dispersion-corrected GGA showed significantly improved agreement with more rigorous benchmarks (MP2<sup>33</sup> or CCSD(T)<sup>34</sup>) for dispersion-dominated interactions. Moreover, improvements were found not only for systems bound by pure dispersion forces, but also for systems interacting via mixed electrostatic and dispersion forces, including  $\pi$ – $\pi$  stacked systems, mixed dipole–nonpolar complexes,

\* To whom correspondence should be addressed. E-mail: samuel.arey@epfl.ch. Phone: +41 21 693 80 31.

<sup>†</sup> Present affiliation: Environmental Chemistry Modeling Laboratory, Swiss Federal Institute of Technology at Lausanne (EPFL).

<sup>‡</sup> Present affiliation: Inorganic Chemistry Laboratory, University of Oxford.

<sup>§</sup> Present affiliation: Department of Chemistry, New York University.

TABLE 1: Hydrogen-Bonded Complexes Evaluated in This Study

hydrogen bond complex <sup>a</sup>	approximate classifications, following Gilli et al. <sup>4,5</sup>	
	strength	type
	single hydrogen bond	
ammonia–ammonia	weak	ordinary, isolated
methanimine–water	weak	ordinary, isolated
ammonia–formaldehyde	weak	ordinary, isolated
water–hydrogen cyanide	weak	ordinary, isolated
water–water	weak	ordinary, isolated
methanol–water	weak	ordinary, isolated
water–formaldehyde	weak	ordinary, isolated
methanol–dimethylsulfide	weak	ordinary, isolated
water–ammonia	moderate	$\sigma$ -bond cooperative, polarization-assisted
water–methanimine	moderate	$\sigma$ -bond cooperative, polarization-assisted
	charged, single hydrogen bond	
water–cyanide	strong	negative charge-assisted
water–acetylide	strong	negative charge-assisted
ammonium–water	strong	positive charge-assisted
water–hydroxide	strong	negative charge-assisted
hydronium–water	strong	positive charge-assisted
	multiple hydrogen bonds	
water trimer	moderate	$\sigma$ -bond cooperative, polarization-assisted
formamide–formamide	strong	$\pi$ -bond cooperative, resonance-assisted
formic acid–formic acid	strong	$\pi$ -bond cooperative, resonance-assisted
adenine–thymine	strong	$\pi$ -bond cooperative, resonance-assisted
guanine–cytosine	strong	$\pi$ -bond cooperative, resonance-assisted

<sup>a</sup> The monomer components of the complex are shown in the order “donor–acceptor”. The vertical ordering approximately reflects increasing hydrogen bond strength within each category.

and biomolecular complexes.<sup>27,28,30,35–38</sup> In other words, the DCACP description of weak interactions was found to be robustly transferable to bond hybridization states that differed significantly from the electronic environments in which the DCACPs were calibrated. The DCACP treatment has been tested for systems containing the elements C, H, O, N, S, and noble gases He, Ne, Ar, and Kr.<sup>27,28,30</sup> Beyond these successes, preliminary evidence suggests that DCACPs may also improve the GGA description of hydrogen-bonded systems. For example, recent work shows that dispersion-corrected BLYP predicts both the static and dynamical properties of liquid water more accurately than BLYP alone,<sup>39</sup> and DCACPs significantly improve the BLYP description of Watson–Crick conformation DNA base pair interactions.<sup>38</sup>

In the present work, we systematically evaluate DCACPs for their ability to improve a GGA description of hydrogen bond binding energies. We test both the GGA functional BLYP<sup>40,41</sup> and dispersion-corrected BLYP predictions of hydrogen bond binding energies and hydrogen bond geometry parameters for 20 complexes containing the elements C, H, O, N, and S. The BLYP model provides an interesting case study GGA for our purposes; although BLYP performs poorly among the GGAs for pure dispersion interactions,<sup>14,15</sup> a recent assessment found BLYP to perform better than both PBE and BP86, on average, over a test set of 18 hydrogen-bonded complexes.<sup>16</sup> We use the results of the current investigation to draw conclusions about the general applicability of DCACPs for weak interactions.

## Methods

**Selection of Molecular Systems.** We aimed to assess the performance of dispersion-corrected BLYP for hydrogen bonds that span a meaningful range of realistic chemical environments. We therefore assembled a data set of 20 hydrogen-bonded complexes which explore different electronic hybridization and charge states (Table 1), using elements for which we have previously parametrized DCACPs, that is, C, H, N, O, and S.

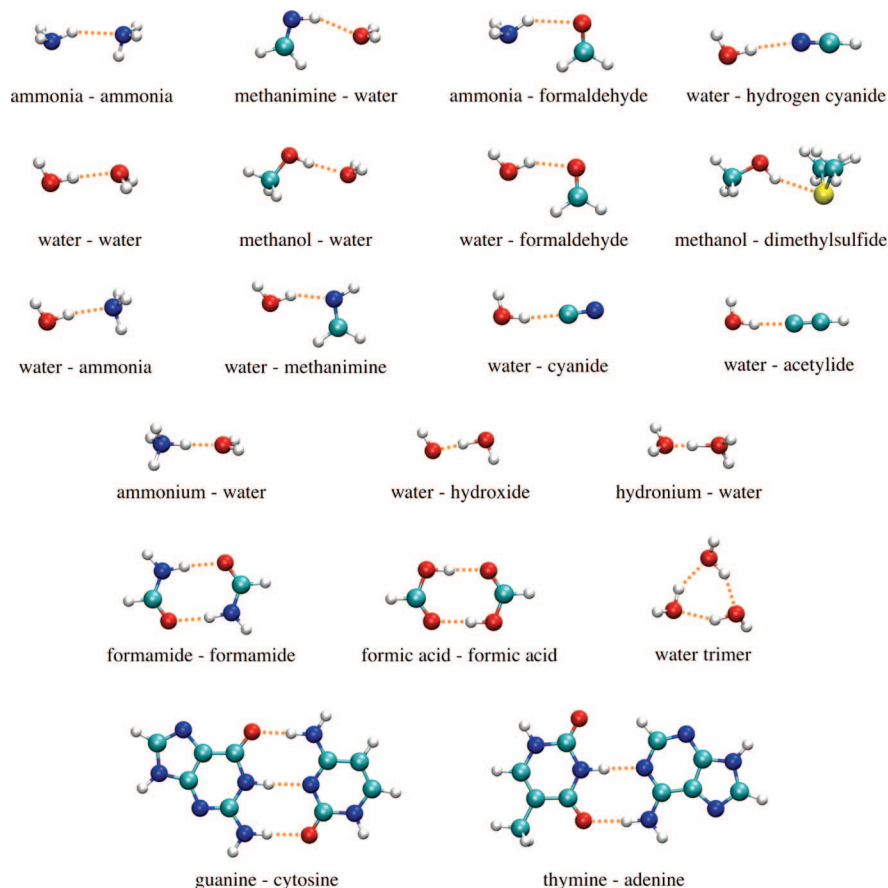
In the terminology of Gilli and co-workers,<sup>4,5</sup> this set includes representatives from several major categories of hydrogen bonds: ordinary, weak; polarization-assisted or  $\sigma$ -bond cooperative; resonance-assisted or  $\pi$ -bond cooperative; positive charge-assisted; and negative charge-assisted (Table 1). The hydrogen bond electronic binding energy for each system was defined as

$$D_e = E_{\text{AH-B}} - E_{\text{AH}} - E_{\text{B}} \quad (1)$$

where AH is the hydrogen-bond donor, B is the hydrogen-bond acceptor, and AH–B is the complex, and  $D_e$  excludes zero point vibrational or thermal energy terms.<sup>42</sup> An exception was the case of water trimer, where the three-body interaction energy was defined as

$$D_{\text{e,H}_2\text{O-H}_2\text{O-H}_2\text{O}} = E_{\text{H}_2\text{O-H}_2\text{O-H}_2\text{O}} - 3E_{\text{H}_2\text{O}} \quad (2)$$

**Benchmark Calculations.** Highly reliable (“benchmark”) estimates of hydrogen bond geometries and  $D_e$  values were sought, so that DFT predictions could be evaluated against these benchmark results. We aimed to obtain benchmark geometries having a minimum quality of MP2/aug-cc-pVTZ; previous assessments suggest that this level of theory is appropriate for accurate hydrogen bond geometries.<sup>16,43,44</sup> Benchmark geometries of 13 complexes (Figure 1) were taken from the literature,<sup>14,16,45,46</sup> 11 of which had geometries optimized using MP2 or CCSD(T) with a polarized triple- $\zeta$  or larger basis set (Table 2). Two complexes, formic acid dimer and formamide dimer, had been optimized using the empirically parametrized composite method MC-QCISD/3.<sup>47</sup> For further confirmation of the MC-QCISD/3 benchmark geometries, we compared these two cases with the CCSD(T)/cc-pVTZ geometry optimizations reported by Jurecka and co-workers.<sup>46</sup> Benchmark  $D_e$  values of these 13 complexes were taken from the same studies as were



**Figure 1.** Benchmark geometries of 20 hydrogen-bonded complexes.

the geometries.<sup>14,16,45,46</sup> We accepted only literature  $D_e$  benchmarks that relied on methods which attempt to extrapolate CCSD(T) results to the basis set limit. It is worth remarking that our selected benchmark set partly overlaps with complexes studied in the recent S22 benchmark database of Jurecka et al.,<sup>46</sup> but we used none of the S22 values. We intentionally excluded  $D_e$  values reported in the S22 benchmark database because those calculations did not account for geometric deformation (relaxation) energies of the monomers.<sup>48</sup> Note that we did use the Jurecka et al.<sup>46</sup> JSCH-2005 benchmark set for two systems (adenine–thymine complex and guanine–cytosine complex). Unlike the S22 set, the JSCH-2005 benchmark data accounts for deformation energies.

To supplement the literature benchmark set, we performed additional calculations for seven new complexes. For this new set, both complexes and monomers were geometry optimized using MP2/aug-cc-pVTZ. Refer to Supporting Information for these benchmark geometries. All complexes and monomers were fully relaxed in the conformations shown in Figure 1, with the exception of water–methanimine complex. A stationary point was difficult to locate for this system in the absence of any geometric constraints, hence this complex was constrained to  $C_s$  symmetry during the optimization. Using these geometries, we computed  $D_e$  values using the W1 composite method<sup>49</sup> for six complexes. The W1 method is a high quality extrapolation to the CCSD(T) basis set limit,<sup>49</sup> and it has been recently assessed for some hydrogen-bonded complexes.<sup>50</sup> The remaining (seventh) complex, methanol–dimethylsulfide, was too large to evaluate using W1, and we applied CBS-QB3<sup>51</sup> for this case. A recent assessment found that CBS-QB3 was in good agreement (mean signed error (MSE) = 0.17 kcal/mol) with W1 and W2 predictions for hydrogen bond energies,<sup>52</sup> given a common

set of geometries. Note that we used MP2 geometries in place of the default B3LYP geometry optimizations normally prescribed in both the W1 and CBS-QB3 protocols. MP2 has been found to give slightly better geometric parameters than B3LYP for hydrogen bond complexes on average.<sup>16,43,44,53</sup> All benchmark calculations were performed using Gaussian 03, revision D02 software.<sup>54</sup>

**DFT Calculations.** BLYP and dispersion-corrected BLYP were used to predict both geometries and  $D_e$  values of the 20 hydrogen-bonded complexes. All monomer and complex geometries were fully relaxed, with the exception of water–methanimine complex, which was constrained to  $C_s$  symmetry. DFT predictions were then assessed against the benchmark calculations. For DFT calculations, the plane-wave quantum chemical code CPMD<sup>55</sup> version 3.12 was employed. The contribution of core electrons was treated using the pseudopotentials of Troullier and Martins;<sup>56</sup> this approximation enables plane-wave (i.e., basis set) convergence at a plane-wave energy cutoff of 70 Ry. The dimensions of the orthorhombic plane-wave cell were set to ensure a minimum distance of 5 Å between any nucleus and the nearest cell boundary. An isolated (gas phase) system was imposed; thus a condition of negligible electronic density was assumed at cell boundaries, using the Poisson solver of Martyna and Tuckerman.<sup>57</sup> Default criteria were used to determine convergence of both the electronic density (largest element of the gradient  $<10^{-5}$  au) and molecular geometry ( $10^{-4}$  au).

## Results and Discussion

We found that both BLYP and dispersion-corrected BLYP complex geometries had similar conformations to those of

**TABLE 2: Hydrogen Bond Distances Predicted by Benchmark Methods, BLYP, and Dispersion-Corrected BLYP<sup>a</sup>**

hydrogen bond complex <sup>b</sup>	benchmark Geometry			BLYP		disp. corrected-BLYP	
	distance	reference	method	distance	error	distance	error
single hydrogen bond							
ammonia–ammonia	3.232	16	CCSD(T)/aug-cc-pVQZ <sup>d</sup>	3.348	0.116	3.270	0.038
methanimine–water	3.174	this work <sup>c</sup>	MP2/aug-cc-pVTZ	3.275	0.101	3.191	0.017
ammonia–formaldehyde	3.124	this work <sup>c</sup>	MP2/aug-cc-pVTZ	3.229	0.105	3.163	0.039
water–hydrogen cyanide	3.050	this work <sup>c</sup>	MP2/aug-cc-pVTZ	3.109	0.059	3.078	0.028
water–water	2.912	16	CCSD(T)/aug-cc-pVQZ <sup>d</sup>	2.918	0.006	2.908	−0.004
methanol–water	2.905	this work <sup>c</sup>	MP2/aug-cc-pVTZ	2.964	0.060	2.929	0.025
water–formaldehyde	2.844	this work <sup>c</sup>	MP2/aug-cc-pVTZ	2.936	0.092	2.899	0.055
methanol–dimethylsulfide	3.222	this work <sup>c</sup>	MP2/aug-cc-pVTZ	3.349	0.127	3.260	0.038
water–ammonia	2.939	16	CCSD(T)/aug-cc-pVQZ <sup>d</sup>	2.957	0.018	2.935	−0.004
water–methanimine	2.851	this work <sup>c</sup>	MP2/aug-cc-pVTZ	2.946	0.095	2.911	0.060
charged, single hydrogen bond							
water–cyanide	2.913	16	CCSD(T)/aug-cc-pVQZ <sup>d</sup>	2.889	−0.024	2.882	−0.031
water–acetylide	2.866	16	CCSD(T)/aug-cc-pVQZ <sup>d</sup>	2.825	−0.041	2.818	−0.048
ammonium–water	2.703	16	CCSD(T)/aug-cc-pVQZ <sup>d</sup>	2.714	0.011	2.708	0.005
water–hydroxide	2.500	16	CCSD(T)/aug-cc-pVQZ <sup>d</sup>	2.528	0.028	2.535	0.035
hydronium–water	2.386	16	CCSD(T)/aug-cc-pVQZ <sup>d</sup>	2.434	0.048	2.433	0.047
multiple hydrogen bonds							
water trimer (O <sup>1</sup> –O <sup>2</sup> )	2.790	45	MP2/aug-cc-pVQZ	2.846	0.056	2.805	0.015
water trimer (O <sup>2</sup> –O <sup>3</sup> )	2.787	45	MP2/aug-cc-pVQZ	2.825	0.038	2.804	0.017
water trimer (O <sup>3</sup> –O <sup>1</sup> )	2.785	45	MP2/aug-cc-pVQZ	2.813	0.028	2.806	0.021
formamide–formamide	2.881	14	MC-QCISD/3	2.923	0.042	2.892	0.011
formic acid–formic acid	2.701	14	MC-QCISD/3	2.657	−0.044	2.642	−0.059
thymine–adenine (N <sup>3</sup> H–N <sup>1</sup> )	2.861	46	MP2/TZVPP	2.857	−0.004	2.854	−0.007
adenine–thymine (N <sup>6</sup> H–O <sup>4</sup> )	2.941	46	MP2/TZVPP	2.901	−0.040	2.901	−0.040
cytosine–guanine (N <sup>6</sup> H–O <sup>6</sup> )	2.790	46	MP2/TZVPP	2.810	0.020	2.787	−0.003
guanine–cytosine (N <sup>1</sup> H–N <sup>3</sup> )	2.934	46	MP2/TZVPP	2.956	0.022	2.929	−0.005
guanine–cytosine (N <sup>2</sup> H–O <sup>2</sup> )	2.937	46	MP2/TZVPP	2.952	0.015	2.927	−0.010
mean unsigned error (MUE)					0.050		0.026
mean signed error (MSE)					0.036		0.010
root mean squared error (RMSE)					0.061		0.032

<sup>a</sup> Defined as the distance between the two non-hydrogen atoms forming the hydrogen bond. <sup>b</sup> The monomer components of the complex are shown in the order “donor–acceptor”. The vertical ordering reflects increasing hydrogen bond strength. <sup>c</sup> Refer to the Supporting Information for the benchmark geometries of these complexes and monomers. <sup>d</sup> The aug-cc-pVQZ basis set was applied to all elements except hydrogen, for which the cc-pVQZ basis set was used.

benchmark results. This enabled meaningful quantitative comparisons of geometric parameters and binding energies.

**BLYP and Dispersion-Corrected BLYP Predictions of Hydrogen Bond Lengths.** Dispersion-corrected BLYP was significantly more accurate than BLYP for predicting hydrogen bond lengths. In most cases, BLYP predicted hydrogen-bonded lengths that were too long compared to benchmark predictions (Table 2). The BLYP hydrogen bond length, defined here as the distance between the two non-hydrogen atoms of the hydrogen bond, exhibited a mean unsigned error (MUE) of 0.050 Å with respect to benchmark geometries. This was comparable to the observed mean signed error (MSE) of 0.036 Å, indicating that BLYP was biased toward overly long hydrogen bonds on average. For 20 out of the 25 individual hydrogen bonds considered, BLYP predicted a longer hydrogen bond than the benchmark, and in 4 cases this error exceeded 0.1 Å. By comparison, dispersion-corrected BLYP showed better accuracy for hydrogen bond lengths (MUE = 0.026 Å) and also exhibited less bias (MSE = 0.010 Å). Dispersion-corrected BLYP predicted shorter hydrogen bond lengths than BLYP for all complexes studied, consistent with the notion that the DCACP acts primarily to increase BLYP’s attractive interaction of weakly bound systems. By any statistical metric (MUE, root mean squared error, or largest error), dispersion-corrected BLYP displayed about half the error of BLYP for hydrogen bond lengths. This finding has broad practical relevance; the length of a hydrogen bond offers diagnostic insights into its category,

strength, and other properties.<sup>4</sup> Finally, the reader should be aware that many of the benchmark data presented here are probably accurate to within  $\pm 0.02$  Å, at best, in the hydrogen bond length. Boese and co-workers<sup>16</sup> recently found that MP2/aug-cc-pVTZ exhibited a root mean squared error (RMSE) of 0.026 Å for hydrogen bond lengths when compared to CCSD(T)/aug-cc-pVQZ results for 18 hydrogen bond complexes. Moreover, we found that Zhao and Truhlar’s<sup>14</sup> MC-QCISD/3 hydrogen bond lengths for formamide dimer and formic acid dimer were 0.021 and 0.036 Å longer,<sup>58</sup> respectively, than the CCSD(T)/cc-pVTZ predictions reported by Jurecka et al.<sup>46</sup> From these results, we conclude that the error exhibited by dispersion-corrected BLYP (RMSE = 0.034 Å) is similar to the uncertainty of most of the benchmark hydrogen bond length data (i.e., the MP2/aug-cc-pVTZ and MC-QCISD/3 geometries) reported here.

A closer examination of particular cases yielded further insights. BLYP performed worst for the hydrogen bond lengths of the weakest complexes, consistent with BLYP’s known failure to treat dispersion-dominated interactions. Dispersion-corrected BLYP gave significantly improved results for BLYP’s four worst cases: ammonia dimer, methanimine–water, ammonia–formaldehyde, and methanol–dimethylsulfide. BLYP suffered error of  $>0.1$  Å in the predicted hydrogen bond length of all of these cases, whereas dispersion-corrected BLYP improved this accuracy to better than 0.05 Å for all four complexes. These four cases all constitute weak hydrogen bonds; BLYP’s failure to predict accurately these geometries is



TABLE 3: Hydrogen Bond Binding Energies Predicted by Benchmark Methods, BLYP, and Dispersion-Corrected BLYP<sup>a</sup>

hydrogen bond complex <sup>b</sup>	benchmark			BLYP		disp. corrected-BLYP	
	<i>D<sub>e</sub></i>	reference	method	<i>D<sub>e</sub></i>	error/h-bond	<i>D<sub>e</sub></i>	error/h-bond
single hydrogen bond							
ammonia–ammonia	−3.15	16	W2	−2.27	0.88	−3.34	−0.19
methanimine–water	−3.20	this work	W1 <sup>c</sup>	−2.25	0.95	−3.31	−0.11
ammonia–formaldehyde	−3.89	this work	W1 <sup>c</sup>	−0.92	2.97	−3.55	0.34
water–hydrogen cyanide	−3.95	this work	W1 <sup>c</sup>	−3.20	0.75	−4.01	−0.06
water–water	−4.97	16	W2	−4.09	0.88	−5.19	−0.22
methanol–water	−5.03	this work	W1 <sup>c</sup>	−4.30	0.73	−5.53	−0.50
water–formaldehyde	−5.43	this work	W1 <sup>c</sup>	−3.97	1.46	−5.24	0.19
methanol–dimethylsulfide	−6.04	this work	CBS-QB3 <sup>c</sup>	−2.64	3.40	−6.29	−0.25
water–ammonia	−6.41	16	W2	−6.07	0.34	−7.17	−0.76
water–methanimine	−6.59	this work	W1 <sup>c</sup>	−5.48	1.11	−6.84	−0.25
charged, single hydrogen bond							
water–cyanide	−15.43	16	W2	−15.90	−0.47	−16.85	−1.42
water–acetylide	−18.25	16	W2	−19.84	−1.59	−19.84	−1.59
ammonium–water	−20.27	16	W2	−20.16	0.11	−21.49	−1.22
water–hydroxide	−26.37	16	W2	−25.86	0.51	−27.14	−0.77
hydronium - water	−33.74	16	W2	−35.28	−1.54	−36.63	−2.89
multiple hydrogen bonds							
water trimer	−15.76	45	ΔCCSD(T)/CBS <sup>e</sup>	−13.11	0.88	−16.61	−0.28
formamide–formamide	−14.94	14	W1 <sup>d</sup>	−12.21	1.37	−15.11	−0.09
formic acid–formic acid	−16.15	14	W1 <sup>d</sup>	−14.45	0.85	−17.54	−0.70
adenine–thymine	−15.43	46	ΔCCSD(T)/CBS <sup>f</sup>	−10.92	2.26	−15.35	0.04
guanine–cytosine	−28.80	46	ΔCCSD(T)/CBS <sup>f</sup>	−22.63	2.06	−27.80	0.33
mean unsigned error/h-bond (MUE)					1.26		0.61
mean signed error/h-bond (MSE)					0.90		−0.52
root mean squared error/h-bond (RMSE)					1.51		0.92
mean relative error/h-bond (MRE)					15.9%		5.1%

<sup>a</sup> Given as *D<sub>e</sub>* values;<sup>42</sup> in other words, zero-point vibrational energy and thermal effects are excluded. <sup>b</sup> The monomer components of the complex are shown in the order “donor–acceptor”. <sup>c</sup> These composite methods were applied to MP2/aug-cc-pVTZ geometries, rather than the B3LYP geometries prescribed by the method protocols. <sup>d</sup> Zhao and Truhlar applied the W1 method to MC-QCISD/3 geometries, rather than B3LYP geometries prescribed by the W1 protocol. <sup>e</sup> The ΔCCSD(T)/CBS extrapolation method is described in Anderson et al.<sup>45</sup> <sup>f</sup> The ΔCCSD(T)/CBS extrapolation method for the “JSCH-2005” benchmark set is described in Jurecka et al.<sup>46</sup>

consistent with the fact that dispersion forces contribute importantly to weak hydrogen bond complexes. For the strongest interactions, including resonance-assisted, positive charge-assisted, and negative charge-assisted hydrogen bonds, both BLYP and dispersion-corrected BLYP predicted reasonable hydrogen bond lengths. Finally, it is worth mentioning that in 19 out of the 20 complexes, both BLYP and dispersion-corrected BLYP found a conformational minimum that was qualitatively similar to the reference geometry. The arguable exception was ammonia dimer. In this case, the benchmark geometry (CCSD(T)/aug-cc-pVQZ) exhibited a so-called bent hydrogen bond (NHN  $\angle = 159.3^\circ$ ).<sup>50</sup> Both DFT methods predicted a significantly less bent hydrogen bonding angle. BLYP led to an energetic minimum at NHN  $\angle = 170.0^\circ$ , and dispersion-corrected BLYP improved upon this slightly, finding NHN  $\angle = 167.5^\circ$ . In previous work, detailed theoretical study of the ammonia dimer has revealed that this potential energy surface is very flat, and DFT methods have consistently overpredicted the NHN angle of this bent hydrogen bond.<sup>50</sup>

**BLYP and Dispersion-Corrected BLYP Predictions of Binding Energies.** Dispersion-corrected BLYP predicted hydrogen bond binding energies significantly more accurately on average than BLYP (Table 3). BLYP was biased toward underbinding, exhibiting a MSE = 0.90 kcal/mol for the 20 complex set. In most cases (17 out of 20), BLYP predicted a binding energy that was too weak compared to the benchmark, consistent with the previously discussed bias toward overly long hydrogen bond lengths. By comparison, dispersion-corrected BLYP was biased toward overbinding (MSE = −0.52 kcal/mol). However, dispersion-corrected BLYP was much more

accurate than BLYP overall. Dispersion-corrected exhibited an MUE of 0.61 kcal/mol (compared to MUE = 1.26 kcal/mol for BLYP) and mean relative error, MRE, of 5.1% (compared to MRE = 15.9% for BLYP). BLYP gave an unsigned error per h-bond of >1 kcal/mol for nearly half of the test set (nine cases), whereas dispersion-corrected BLYP exhibited an error of this magnitude for only four cases. The observed improvements in both relative error and absolute error indicated that dispersion-corrected BLYP performed better than BLYP, on average, for both weak and strong hydrogen bonds.

Going beyond the statistics of average DFT predictions for *D<sub>e</sub>* values, it is worth inspecting particular cases and subsets. BLYP made the largest relative errors for predictions of weak hydrogen bond energies, exhibiting relative errors ranging from 27 to 76% in the *D<sub>e</sub>* for the ammonia–ammonia, methanimine–water, ammonia–formaldehyde, water–formaldehyde, and methanol–dimethylsulfide complexes. Notably, BLYP also suffered errors of approximately 0.1 Å or more in the hydrogen bond lengths for all five of these cases (Table 2). By comparison, dispersion-corrected BLYP made systematic improvements for these complexes, reducing the error to less than 10% in the *D<sub>e</sub>* and less than 0.06 Å in the hydrogen bond length, in every case. Although BLYP and dispersion-corrected BLYP both overpredicted the NHN hydrogen bond angle for the ammonia dimer, these methods both showed reasonable agreement (−2.27 and −3.34 kcal/mol, respectively) with the weak interaction energy predicted by the W2 benchmark (−3.15 kcal/mol). Although perhaps fortuitous, this result is consistent with the previous observation that the potential energy surface of the ammonia dimer is very flat in the region of the hydrogen-bonded

conformation.<sup>50</sup> Dispersion-corrected BLYP's significantly improved treatment of weak hydrogen bonds is consistent with the fact that dispersion forces play an important role in these intermolecular interactions. In addition to this, DCACPs improved BLYP performance for stronger hydrogen bonds, including the resonance-assisted hydrogen bonds formed by formamide dimer, formic acid dimer, and the Watson–Crick conformation base pairs adenine–thymine and guanine–cytosine. These are nontrivial achievements for DFT; in a recent assessment, Zhao and Truhlar<sup>14</sup> found that formamide dimer and formic acid dimer were difficult cases for all the DFT methods that they tested.

In terms of absolute error, dispersion-corrected BLYP displayed the worst performance for the five charged hydrogen bond systems. For these strongly interacting complexes, dispersion-corrected BLYP exhibited a systematic bias toward overbinding, whereas BLYP handled these systems better. The observed bias for dispersion-corrected BLYP could not be clearly attributed to poor geometries; predictions of hydrogen bond lengths were within 0.05 Å of the benchmark in all cases, for both BLYP and dispersion-corrected BLYP. Hence, the DCACP appears to slightly overcorrect the strength of the charge-assisted hydrogen bonds. Nonetheless, using either model (BLYP or dispersion-corrected BLYP), the relative error did not exceed 10% for any of these cases.

For hydrogen bond binding energies, dispersion-corrected BLYP exhibits reasonable accuracy compared to the M06, M06-2X, and M06-L methods recently developed by Zhao and Truhlar.<sup>22</sup> In a recent assessment, this “M06 suite” of meta-GGAs and hybrid-meta-GGAs displays superior or competitive performance compared to other DFT methods for a broad range of chemical problems.<sup>22</sup> We do not provide a complete side-by-side assessment of dispersion-corrected BLYP and the M06 suite, but some comparisons can be made. Across Zhao and Truhlar's “HB6/04”<sup>14</sup> database set, consisting of NH<sub>3</sub> dimer, HF dimer, H<sub>2</sub>O dimer, H<sub>2</sub>O–NH<sub>3</sub>, formamide dimer, and formic acid dimer, Zhao and Truhlar reported that M06, M06-2X, and M06-L exhibit MUEs of 0.28, 0.35, and 0.36 kcal/mol, respectively.<sup>22</sup> Our test set includes all of these complexes except the HF dimer, because a DCACP has not yet been developed for the element F. Across the remaining five out of these six complexes, dispersion-corrected BLYP gives MUE = 0.55 kcal/mol. However, it should be noted that all three M06 suite functionals were calibrated using the HB6/04 database as part of their training sets,<sup>22,59</sup> whereas DCACPs were calibrated using no hydrogen-bonded systems at all. Additionally, Zhao and Truhlar used a set of common geometries for both their benchmark and DFT calculations, whereas we have employed DFT geometries for DFT binding energies and wave function-based geometries for benchmark binding energies. Hence the current comparisons are preliminary.

## Conclusions

Having low computational overhead and useful accuracy, GGA functionals are used widely to treat condensed phase systems.<sup>24–26</sup> In these studies, condensed phases are simulated explicitly precisely because weak interactions, including dispersion forces, may play a key role in the chemical structures and kinetics of interest. However, it has been long understood that the GGAs' formal lack of nonlocal electron correlation will often give rise to errors where dispersion forces are important. In this study, we focused attention on hydrogen bonds, where dispersion forces can be significant.<sup>17–21</sup> Here, we amended the BLYP Hamiltonian with a nonlocal interaction potential (DCACPs) that has been designed to account for dispersion forces. In the

case of hydrogen bonding, it is too simple to interpret the DCACP as a correction to dispersion forces only. As has been remarked previously, DCACPs correct for errors in the GGA description of both Pauli repulsion and London dispersion forces,<sup>28,29</sup> and both forces contribute importantly to hydrogen bond binding energies. Nonetheless, the resulting predictions of hydrogen bond geometries and binding energies are substantially improved. This is a significant landmark of the transferability of DCACPs, which were originally calibrated to reproduce a benchmark interaction potential for pure-dispersion van der Waals clusters. These results suggest that DCACPs are highly transferable to weak interactions that explore a wide variety of molecular hybridization states and electrostatic environments. Hence DCACPs may inform the practical development of future functional designs for DFT. The present findings confirm growing evidence that DCACPs improve GGA performance for weak interactions that extend significantly beyond the chemical space of the original calibration systems.

**Acknowledgment.** The authors thank Geoffrey P. F. Wood for his comments on the manuscript. Jan Cornelius Brauer, Thomas Schoettl, Jean Soncini, Sylvain Ruetti, and Lucien Ruffino are thanked for their contributions to the formative stages of the project. The U.S. National Science Foundation (MPS-DRF Award 0502600) and the Swiss National Supercomputing Center (CSCS) are gratefully acknowledged for their support. I.-C. L. acknowledges the SNF postdoctoral research grant No. PBELP-123062.

**Supporting Information Available:** Benchmark geometries (MP2/aug-cc-pVTZ) for seven complexes and for the corresponding monomers, as indicated in Table 2. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References and Notes

- (1) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960.
- (2) Coulson, C. A.; Danielsson, U. *Ark. Fys.* **1954**, *8*, 239.
- (3) Coulson, C. A.; Danielsson, U. *Ark. Fys.* **1954**, *8*, 245.
- (4) Gilli, G.; Gilli, P. *J. Mol. Struct.* **2000**, *552*, 1.
- (5) Gilli, P.; Bertolasi, V.; Pretto, L.; Gilli, G. *J. Mol. Struct.* **2006**, *790*, 40.
- (6) Joesten, M. D.; Schaad, L. J. *Hydrogen Bonding*; Marcel Dekker, Inc.: New York, NY, 1974.
- (7) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, 1997.
- (8) Vinogradov, S. N.; Linnell, R. H. *Hydrogen Bonding*; Van Nostrand Reinhold: New York, 1971.
- (9) Jeffrey, G. A.; Saenger, W. *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, 1991.
- (10) Parr, R. G.; Yang, W. *Annu. Rev. Phys. Chem.* **1995**, *46*, 701.
- (11) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.
- (12) Meijer, E. J.; Sprik, M. *J. Chem. Phys.* **1996**, *105*, 8684.
- (13) Kohn, W.; Mier, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.
- (14) Zhao, Y.; Truhlar, D. G. *J. Chem. Theor. Comput.* **2005**, *1*, 415.
- (15) Zhao, Y.; Truhlar, D. G. *J. Chem. Theor. Comput.* **2007**, *3*, 289.
- (16) Boese, A. D.; Martin, J. M. L.; Klopper, W. *J. Phys. Chem. A* **2007**, *111*, 11122.
- (17) Cybulski, S. M.; Chalasinski, G.; Moszynski, R. *J. Chem. Phys.* **1990**, *92*, 4357.
- (18) Cybulski, S. M.; Scheiner, S. *Chem. Phys. Lett.* **1990**, *166*, 57.
- (19) Scheiner, S. *Annu. Rev. Phys. Chem.* **1994**, *45*, 23.
- (20) Langlet, J.; Caillet, J.; Caffarel, M. *J. Chem. Phys.* **1995**, *103*, 8043.
- (21) Zierkiewicz, W.; Jurecka, P.; Hobza, P. *ChemPhysChem* **2005**, *6*, 609.
- (22) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
- (23) Guidon, M.; Schiffmann, F.; Hutter, J.; VandeVondele, J. *J. Chem. Phys.* **2008**, *128*, 214104.
- (24) Parrinello, M. *Solid State Comm.* **1997**, *102*, 107.
- (25) Colombo, M. C.; Guidoni, L.; Laio, A.; Magistrato, A.; Maurer, P.; Piana, S.; Rohrig, U.; Spiegel, K.; Sulpizi, M.; VandeVondele, J.; Zumstein, M.; Rothlisberger, U. *Chimia* **2002**, *56*, 13.

- (26) Meng, S.; Wang, E. G.; Gao, S. *Phys. Rev. B* **2004**, *69*, 195404.
- (27) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. Lett.* **2004**, *93*, 153004.
- (28) Lin, I.-C.; Coutinho-Neto, M. D.; Felsenheimer, C.; von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. B* **2007**, *75*, 205131.
- (29) Lin, I.-C.; Rothlisberger, U. *Chimia* **2008**, *62*, 231.
- (30) Aeberhard, P. C.; Arey, J. S.; Lin, I.-C.; Rothlisberger, U. *J. Chem. Theor. Comput.* **2009**, *5*, 23.
- (31) Note that the DCACP treatment differs from pairwise nuclear potentials and other dispersion-correction methods that have been devised for DFT. The reader is referred to von Lilienfeld et al. (ref 27) for further discussion of the distinctions between these methods.
- (32) Tavernelli, I.; Lin, I.-C.; Rothlisberger, U. *Phys. Rev. B* **2009**, *79*, 045106.
- (33) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. *Chem. Phys. Lett.* **1988**, *153*, 503.
- (34) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. *J. Chem. Phys.* **1987**, *87*, 5968.
- (35) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. B* **2005**, *71*, 195119.
- (36) Lin, I.-C.; von Lilienfeld, O. A.; Coutinho-Neto, M. D.; Tavernelli, I.; Rothlisberger, U. *J. Phys. Chem. B* **2007**, *111*, 14346.
- (37) Tapavicza, E.; Lin, I.-C.; von Lilienfeld, O. A.; Tavernelli, I.; Coutinho-Neto, M. D.; Rothlisberger, U. *J. Chem. Theor. Comput.* **2007**, *3*, 1673.
- (38) Lin, I.-C.; Rothlisberger, U. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2730.
- (39) Lin, I.-C.; Seitsonen, A. P.; Tavernelli, I.; Coutinho-Neto, M. D.; Rothlisberger, U. *J. Phys. Chem. B* **2009**, *113*, 1127.
- (40) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.
- (41) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (42) Herzberg, G. *Molecular Spectra and Molecular Structure. I. Spectra of Diatomic Molecules*, 2nd ed.; Van Nostrand Reinhold Co.: Princeton, NJ, 1950.
- (43) Del Bene, J. E.; Person, W. B.; Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705.
- (44) Tuma, C.; Boese, A. D.; Handy, N. C. *Phys. Chem. Chem. Phys.* **1999**, *1*, 3939.
- (45) Anderson, J. A.; Crager, K.; Fedoroff, L.; Tschumper, G. S. *J. Chem. Phys.* **2004**, *121*, 11023.
- (46) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
- (47) Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2003**, *107*, 3898.
- (48) In separate calculations, we ascertained that these deformation energies are significant for some cases.
- (49) Martin, J. M. L.; de Oliveira, G. *J. Chem. Phys.* **1999**, *111*, 1843.
- (50) Boese, A. D.; Chandra, A.; Martin, J. M. L.; Marx, D. *J. Chem. Phys.* **2003**, *119*, 5965.
- (51) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822.
- (52) Zhao, Y.; Truhlar, D. G. *J. Chem. Theor. Comput.* **2006**, *2*, 1009.
- (53) Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.
- (54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (55) CPMD, version 3.12; IBM Corp. and MPI Stuttgart: Stuttgart, Germany, 2006.
- (56) Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993.
- (57) Martyna, G. J.; Tuckerman, M. E. *J. Chem. Phys.* **1999**, *110*, 2810.
- (58) We could not assess objectively whether Zhao and Truhlar's (ref 14) MC-QCISD/3 geometries were superior to those reported by Jurecka et al. (ref 46), who used CCSD(T)/cc-pVTZ. We selected the MC-QCISD/3 geometries as part of our benchmark set because Zhao and Truhlar calculated W1 binding energies using these geometries, which we report in Table 3. The hydrogen bond lengths predicted by BLYP and dispersion-corrected BLYP agree very slightly better with the CCSD(T)/cc-pVTZ geometries.
- (59) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.

JP810323M