JCTC Journal of Chemical Theory and Computation

A New Empirical Correction to the AM1 Method for Macromolecular Complexes

Michael E. Foster and Karl Sohlberg*

Department of Chemistry, Drexel University, 3141 Chestnut Street, Philadelphia, Pennsylvania 19104

Received April 2, 2010

Abstract: Modeling systems that are governed by van der Waals (dispersion) interactions using empirically corrected DFT methods is becoming increasingly popular due to the promise of a CCSD(T) level accuracy at the computational cost of DFT. Although, DFT methods are computationally efficient in comparison to the CCSD(T) method, currently, structural optimizations using DFT methods are generally only feasible for systems of less than a few hundred atoms. We seek a method applicable to macromolecular complexes. In order to model such large systems, empirically corrected semiempirical methods appear to be an attractive alternative. As with most common DFT methods, the popular semiempirical methods (e.g., AM1) also do not model long-range dispersion (and therefore an empirical correction term is desirable), but this is not their only shortcoming. For weakly interacting systems, hydrogen bonding also poses a concern. A new empirically corrected AM1 method that uses two empirical correction terms, one for dispersion and one for hydrogen bonding interactions, is presented and termed AM1-FS1. This new empirically corrected AM1 method has been parametrized to a diverse training set of 66 complexes that includes nonequilibrium structures and yields sub-kilocalorie accuracy in the prediction of intermolecular interaction energies. More significantly, AM1-FS1 achieves this result with substantially less parametrization than existing empirically corrected semiempirical methods and *without modification of the original AM1 parameters* so that it retains both the computational efficiency and predictive power for thermo-chemical quantities of the original AM1 Hamiltonian. The performance of AM1-FS1 is also tested on several carbon nanostructure complexes and pseudorotaxanes and is found to produce results in very good agreement with the best first-principles calculations.

1. Introduction

Accurately and efficiently modeling nonbonding interactions, especially van der Waals interactions and hydrogen bonding, is a difficult task, but essential for the correct description of many systems of chemical and biological importance, such as the structure of molecular crystals,[1,2] the conformational preference[3−9] and folding of proteins, and the stability of two strands of DNA in a double helix.[3,10] These types of interactions are also important in macromolecular host/guest chemistry,[11−14] which motivates this work.

Density functional theory (DFT) is the most widely used quantum mechanical (QM) technique for chemical calculations due in part to its ability to accurately describe chemical and physical properties for a diversity of systems, often at modest computational expense. A major shortcoming with DFT is the inability of most popular XC functionals (exchange-correlation functionals) to accurately model long-range van der Waals (dispersion) interactions. Therefore, these methods predict systems like the benzene dimer to be unbound (Figure 1). Currently, an increasingly popular approach to overcome this hurdle is to add an empirical correction to the DFT total energy. Empirically corrected DFT methods for dispersion interactions, coined DFT-D, have become popular due to

* Corresponding author e-mail: kws24@drexel.edu.

10.1021/ct100177u
Published on Web 06/16/2010

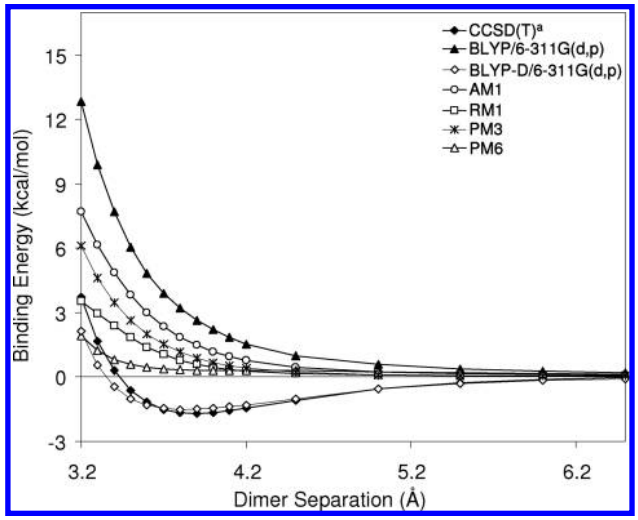

Figure 1. Potential energy curves for the parallel benzene dimer determined with various quantum mechanical methods. Superscript a refers to ref 23.

their success with essentially no added computational expense. Figure 1 shows the dramatic improvement that is achieved by adding an empirical correction term. Not only is the complex now predicted to be bound but excellent agreement with CCSD(T) results is achieved, the current "gold standard" in computational chemistry. For relatively small systems, DFT-D methods are computationally feasible and should provide quite accurate results, but modeling macromolecular host/guest complexes can be extremely computationally expensive. Therefore, alternative methods need to be explored.

Semiempirical (SE) techniques such as AM1,[15] PM3,[16] RM1,[17] and PM6[18] are sufficiently computationally efficient for modeling systems composed of hundreds or even thousands of atoms but typically perform poorly for dispersion and hydrogen bonding. These semiempirical methods are essentially incapable of modeling dispersion bound complexes because the form of the semiempirical wave function completely neglects electron correlation. Even qualitatively reliable modeling of dispersion-bound macromolecular systems, such as complexes of carbon nanostructures, is therefore out of the question; SE methods predict such complexes to be unbound. That is, the wrong sign of the interaction is predicted. Again, considering the parallel benzene dimer (Figure 1), which is a classic test case used for predicting the likely accuracy of a method for modeling complexes of carbon nanostructures, we see that the SE methods fail.

The accuracy of SE methods in modeling dispersion-bound systems can be dramatically improved by adding an empirical correction term. McNamara and Hillier[19] reported adding an empirical correction term to the AM1 and PM3 methods to incorporate dispersion interactions but found the overall results to be unsatisfactory. Therefore, to gain further improvements, in particular to improve the accuracy with which hydrogen-bonding is modeled, they reoptimized 18 of the original AM1 parameters, using the S22 database[20] as a training set. The resulting method, with *both* reoptimized semiempirical parameters *and* an empirical correction term, is referred to as AM1-D. (They have also produced an analogous PM3-D.) These empirically corrected methods show a substantial improvement in accuracy (over the corresponding original SE methods) for predicting intermolecular interaction energies, but at a significant cost. As detailed below, AM1-D is nearly 25-fold less accurate in the prediction of heats of formation than the original AM1 method. We seek a method that leads to good accuracy in the prediction of structures and interaction energies for macromolecular complexes, without sacrificing predictive power for the heat of formation.

More recently Řezáč and colleagues[21] published an empirically corrected PM6 method for modeling dispersion and hydrogen-bonding interactions, named PM6-DH. To incorporate dispersion interactions, the group used the empirical correction described by Jurecka et al.,[22] which is also discussed here in section 2.1. To improve the PM6 method for H-bonding, they included a second correction term involving three parameters. The correction term is applied to H-bonding situations, but not all types of H bonds are modeled with the same term. The group identified eight types of H bonds and used a different set of three parameters for each type, for a total of 24 H-bonding parameters. In their defense, it should be noted that they did use a relatively large training set to determine these H-bonding parameters. The major shortcoming of this method, as they acknowledge, is that knowledge of atom connectivity is required. One of the major benefits of QM techniques is that atom connectivity is not required, allowing bond formation/deformation to be modeled. Although, Řezáč and colleagues have obtained good results, for a method to be widely used, atom conductivity information should not be required. In addition to the limitations introduced, input of atom connectivity information is sufficiently burdensome to deter routine use, especially for macromolecular complexes where there may be thousands of atom−atom interactions that must be distinguished.

Herein, we present an empirical correction for the AM1 method that is suitable for modeling macromolecular complexes and avoids the above-noted shortcomings of existing techniques. We have chosen to apply separate empirical correction terms for dispersion and hydrogen bonding. Our method requires significantly less parametrization than the AM1-D and PM3-D methods of McNamara and Hillier[19] and also the PM6-DH method of Řezáč et al.[21] Additionally, it is important to note that we have not altered any of the original AM1 parameters. Such changes can have deleterious effects on predictions of properties not based strictly on the total energy or its derivatives, such as heats of formation, ionization potentials, and dipole moments, if these quantities are not taken into consideration during reparameterization. Our method also does not require knowledge of atom connectivity. We will henceforth refer to our new method as "AM1-FS1". AM1-FS1 achieves results that are comparable to (and in many cases better than) those of other empirically corrected SE methods, with *significantly* less parametrization and with no reparameterization of the AM1 method. The main objective of AM1-FS1 is to accurately model macromolecular host/guest systems that are currently

out of reach of DFT-D techniques. AM1-FS1 aims to not only accurately predict energies but also reasonable structures upon geometry optimization, since structural optimization is one of the main uses for such a technique. Herein, the accuracy of AM1-FS1 is tested by comparing interaction energies and distances to CCSD(T) and SAPT results; comparisons are also made with other empirically corrected semiempirical techniques.

2. Theory

2.1. Dispersion Correction. To correct the AM1 method for dispersion interactions, we have employed a method used by Grimme[24] with a slight modification suggested by Jurečka et al.[22] The resulting dispersion correction is of the form

$$E_{\mathrm{dis}} = -\frac{C_6^{ij}}{r^6} f_{\mathrm{damp}}(r_{ij}) \tag{1}$$

where r_{ij} is the atom−atom separation, C_6^{ij} is the dispersion coefficient, and f_{damp} is a damping function of the form

$$f_{\mathrm{damp}}(r_{ij}) = \frac{1}{1 + \exp(-d(\frac{r_{ij}}{S_R R_{\mathrm{vdw}}} - 1))} \tag{2}$$

This damping function depends on the equilibrium van der Waals separation (R_{vdw}) and the pairwise atom separation (r_{ij}). The damping function also depends on two unitless parameters, S_R and d, which have been optimized to a training set as discussed at length in section 2.3. The damping function operates as a switching function, turning off the dispersion term at short range. This is required because the SE wave function already models short-range repulsive interactions. Thus, the popular 6−12 Lennard-Jones (LJ) potential is not suitable for use as a dispersion correction since a repulsive term is involved (see ref 25 for a more detailed discussion and graphical representations).

It should be noted that we tried employing the global scaling factor, used by Grimme,[24,26] instead of scaling the equilibrium van der Waals separation (S_R); however, a smaller root-mean-square error (RMSE) was obtained on our training set using S_R (discussed in section 2.3). Scaling $R_{v\mathrm{dw}}$ seems theoretically well motivated, since this allows only the short-range interactions to be tailored and leaves untouched the long-range interactions for which the correct functional form of the interaction is known to follow r^{-6} (see ref 25 for a more detailed discussion and graphical representations).

Another decision concerns the choice of combination rules used for obtaining C_6^{ij} and R_{vdw}. We have chosen to employ the geometric mean and simple average combination rules for determining C_6^{ij} and R_{vdw}, respectively:

$$C_6^{ij} = \sqrt{C_6^i C_6^j}, R_{\mathrm{vdw}} = \frac{R_i + R_j}{2} \tag{3}$$

The dispersion coefficients (C_6^i and C_6^j) and van der Waals radii (R_i and R_j) for the different atoms were obtained from Grimme's 2006 publication.[24] The decision of using these particular combination rules was not made without considering other options. For C_6^{ij}, both the harmonic mean[26] and the combination rule suggested by Wu and Yang,[27] which uses the Slater−Kirkwood effective number of electrons, were considered. For R_{vdw}, the cubic mean suggested by Halgren[28] was also considered. We have also considered all possible combinations and found that the parameters (S_R and d) seemed to adjust to accommodate the different combination rules. The combination rules employed yielded the lowest RMSE for our training set. It should be noted that only Grimme's 2006 published dispersion coefficients and van der Waals radii values were considered. This dispersion correction scheme to the AM1 method has been implemented into a locally modified version of GAMESS.[29]

2.2. Hydrogen-Bonding Correction. Correcting the AM1 method for hydrogen bonding is a more difficult task than correcting for its neglect of dispersion since hydrogen-bonding interactions are already in part considered, given their partial electrostatic nature. It can be seen by looking at the H-bonded systems (1−7) in the S22 database (Table 1) that the AM1 method severely underbinds such complexes. The AM1 method does, however, produce more reasonable interaction energies for hydrogen-bonded systems upon geometry optimizations (Table 2); this is because AM1 generally predicts dispersion-bound complexes to be unbound, while for H-bonded complexes it predicts some binding, but generally with an unphysically large equilibrium separation (see Table 3). Thus, to improve the AM1 method for predicting H-bonding systems, the strength of these interactions needs to be increased at medium-to-short range. We have achieved this by adding a post-SCF pseudoelectrostatic term of the form

$$E_{\mathrm{HB}} = \alpha_1 \frac{Q_i Q_j}{r_{ij}} \cos^2(\theta) f_{\mathrm{damp2}}(r_{ij}) \tag{4}$$

where α_1 is a global scaling factor, Q_i and Q_j are the AM1 Coulson charges[30] (which are referred to as MOPAC charges in GAMESS[29]), r_{ij} is the H---Y separation, θ is the XH---Y angle, and f_{damp2} is a damping function of the form

$$f_{\mathrm{damp2}}(r_{ij}) = \exp[-(r_{ij} - \alpha_2 R_{\mathrm{vdw}})^2/\alpha_3^2(1 + \alpha_4(r_{ij} - \alpha_2 R_{\mathrm{vdw}}))^2] \tag{5}$$

where α_2, α_3, and α_4 are parameters and all other terms have the same meanings as in the dispersion correction. In this case, however, R_{vdw} is defined as the cubic mean

$$R_{\mathrm{vdw}} = \frac{R_i^3 + R_j^3}{R_i^2 + R_j^2} \tag{6}$$

The cubic mean is used in this case because it yields a slightly smaller RMSE for the F66 training set than using the simple average combination rule.

The damping function is an asymmetric distribution function (see Figure 2A) that turns the hydrogen-bonding function on or off over an appropriate range for correcting the AM1 method. To achieve an asymmetric distribution, three parameters (α_2, α_3, α_4) have been introduced, giving a total of four parameters in the H-bonding correction. We have optimized these four parameters to improve upon H

Table 1. Single-Point Interaction Energies (kcal/mol) at the S22 Geometries[a]

no.	molecule (symmetry)	ref values	AM1	PM3	AM1-D	PM3-D	PM6-DH	AM1-FS1
			Hydrogen-Bonded Complexes					
1	$(NH_3)2$ (C2h)	−3.17	−0.78	0.77	−3.43	−1.77	−3.74	−1.60
2	$(H_2O)2$ (Cs)	−5.02	−2.89	−2.79	−7.29	−5.14	−4.67	−5.53
3	formic acid dimer (C2h)	−18.61	1.54	−9.91	−15.45	−18.57	−17.39	−16.06
4	formamide dimer (C2h)	−15.96	−12.02	−8.08	−17.16	−15.37	−15.39	−15.75
5	uracil dimer (C2h)	−20.65	−5.79	−11.32	−20.15	−20.30	−18.84	−20.80
6	2-pyridoxine2-aminopyridine (C1)	−16.71	−4.45	−7.46	−16.50	−17.52	−17.35	−14.73
7	adenine thymine WC (C1)	−16.37	−4.28	−6.79	−16.58	−17.33	−17.83	−16.29
			Complexes with Predominant Dispersion Contribution					
8	$(CH_4)2$ (D3d)	−0.53	0.21	−0.25	−0.94	−1.24	−0.73	−0.61
9	$(C_2H_4)2$ (D2d)	−1.51	−0.13	−1.11	−3.31	−3.60	−1.52	−2.27
10	benzene CH_4 (C3)	−1.50	0.40	−0.19	−2.12	−2.42	−1.75	−1.79
11	benzene dimer (C2h)	−2.73	3.52	2.38	−2.90	−4.30	−3.62	−2.23
12	pyrazine dimer (Cs)	−4.42	2.49	3.90	−4.57	−4.20	−5.41	−3.81
13	uracil dimer (C2)	−10.12	0.12	5.80	−10.56	−6.78	−9.70	−8.47
14	indole benzene (C1)	−5.22	5.39	4.04	−4.04	−6.09	−5.20	−3.23
15	adenine thymine stack (C1)	−12.23	2.91	7.37	−12.20	−10.63	−12.78	−9.87
			Mixed Complexes					
16	ethene ethine (C2v)	−1.53	−0.35	−0.82	−1.61	−1.85	−1.11	−1.36
17	benzene H_2O (Cs)	−3.28	−0.69	−1.47	−3.43	−3.65	−3.41	−2.78
18	benzene NH_3 (Cs)	−2.35	−0.33	−0.59	−3.00	−2.96	−2.77	−2.65
19	benzene HCN (Cs)	−4.46	−0.81	−1.63	−4.44	−4.43	−3.20	−3.17
20	benzene dimer (C2v)	−2.74	0.37	−0.43	−3.85	−4.15	−2.84	−3.36
21	indole benzene T-shape (C1)	−5.73	−1.05	−1.25	−7.10	−6.65	−5.30	−4.63
22	phenol dimer (C1)	−7.05	−1.36	−1.37	−9.76	−7.52	−6.73	−6.91
	RMSE (hydrogen bonded)		11.64	7.77	1.56	0.76	1.07	1.37
	RMSE (dispersion bonded)		8.21	10.13	0.82	1.68	0.54	1.30
	RMSE (mixed bonded)		3.57	3.22	1.25	0.72	0.57	0.72
	RMSE		8.47	7.73	1.23	1.18	0.76	1.18
	MUE		6.54	5.94	0.85	0.90	0.59	0.88

[a] The AM1-D and PM3-D results have been taken from ref 19, and the PM6-DH results from ref 21.

bonding for the AM1 method. A detailed discussion is presented in the next section.

The H-bonding correction function also depends on the square of the cosine of the XH---Y angle. This is motivated

Table 2. Geometry Optimized Energies (kcal/mol) for the Complexes in the S22 Database[a]

no.	molecule (symmetry)	ref values	AM1	PM3	AM1-D	PM3-D	PM6-DH	AM1-FS1
			Hydrogen-Bonded Complexes					
1	$(NH_3)_2$ (C2h)	−3.17	−1.39	−0.71	−3.03	−1.99	−3.92	−2.82
2	$(H_2O)_2$ (Cs)	−5.02	−3.30	−3.55	−7.22	−6.53	−4.73	−5.59
3	formic acid dimer (C2h)	−18.61	−6.62	−9.58	−12.45	−16.16	−19.11	−17.76
4	formamide dimer (C2h)	−15.96	−2.06	−6.99	−14.64	−14.42	−15.01	−15.83
5	uracil dimer (C2h)	−20.65	−10.48	−10.70	−17.80	−18.83	−19.55	−25.06
6	2-pyridoxine2-aminopyridine (C1)	−16.71	−6.15	−7.06	−13.06	−18.32	−18.50	−15.16
7	adenine thymine WC (C1)	−16.37	−5.06	−6.90	−12.66	−18.66	−19.12	−21.10
			Complexes with Predominant Dispersion Contribution					
8	$(CH_4)_2$ (D3d)	−0.53	−0.21	−0.32	−4.10	−2.38	−0.73	−2.46
9	$(C_2H_4)2$ (D2d)	−1.51	−0.13	−1.08	−4.85	−4.11	−1.53	−4.09
10	benzene CH_4 (C3)	−1.50	0.35	−0.20	−2.93	−2.88	−1.88	−2.84
11	benzene dimer (C2h)	−2.73	0.01	−0.02	−3.10	−4.59	−3.59	−2.21
12	pyrazine dimer (Cs)	−4.42	−0.34	−0.26	−4.87	−4.45	−5.74	−4.73
13	uracil dimer (C2)	−10.12	−6.05	−4.26	−11.25	−7.59	−10.03	−9.99
14	indole benzene (C1)	−5.22	−1.33	−1.65	−8.16	−6.26	−5.99	−6.51
15	adenine thymine stack (C1)	−12.23	−5.15	−6.50	−15.13	−11.70	−13.61	−12.59
			Mixed Complexes					
16	ethene ethine (C2v)	−1.53	−0.57	−1.23	−2.47	−2.58	−1.17	−1.50
17	benzene H_2O (Cs)	−3.28	−1.03	−1.63	−3.90	−4.46	−3.95	−3.38
18	benzene NH_3 (Cs)	−2.35	−0.80	−0.93	−4.04	−3.99	−3.82	−4.70
19	benzene HCN (Cs)	−4.46	−0.92	−1.85	−4.28	−4.40	−3.21	−2.46
20	benzene dimer (C2v)	−2.74	−0.09	−0.52	−4.22	−4.39	−2.85	−2.15
21	indole benzene T-shape (C1)	−5.73	−1.24	−1.67	−7.74	−7.20	−5.22	−5.88
22	phenol dimer (C1)	−7.05	−3.39	−4.33	−11.55	−8.95	−7.46	−8.87
	RMSE (hydrogen bonded)		9.90	8.04	3.38	1.82	1.40	2.55
	RMSE (dispersion bonded)		3.73	3.65	2.36	1.71	0.80	1.34
	RMSE (mixed bonded)		2.96	2.40	2.09	1.40	0.82	1.37
	RMSE		6.25	5.22	2.65	1.65	1.04	1.82
	MUE		4.82	4.09	2.16	1.51	0.82	1.28

[a] The AM1-D and PM3-D results have been taken from ref 19 and the PM6-DH results from ref 21.

Table 3. Interaction Distances (Ångstroms) for the Complexes in the S22 Database[a]

no.	molecule (symmetry)	ref values	AM1	PM3	AM1-D	PM3-D	AM1-FS1
			Hydrogen-Bonded Complexes				
1	$(NH_3)_2$ (C2h)	2.504	2.784	3.241	2.646	2.726	2.668
2	$(H_2O)2$ (Cs)	1.952	2.094	1.809	1.911	1.769	1.932
3	formic acid dimer (C2h)	1.670	2.101	1.776	1.925	1.737	1.567
4	formamide dimer (C2h)	1.841	2.072	1.807	1.981	1.763	1.916
5	uracil dimer (C2h)	1.775	2.044	1.787	1.946	1.744	1.563
6	2-pyridoxine-2-aminopyridine (C1)	1.859, 1.874	2.511, 2.107	1.798, 1.815	1.980, 1.981	1.722,1.768	1.760, 1.878
7	adenine thymine WC (C1)	1.819, 1.929	2.476, 2.101	1.780, 1.821	1.807,2.018	1.708, 1.769	1.597, 1.893
			Complexes with Predominant Dispersion Contribution				
8	$(CH_4)_2$ (D3d)	3.718	3.721	3.447	2.881	3.160	2.899
10	benzene CH_4 (C3)	3.716	3.746	3.718	3.315	3.450	3.457
11	benzene dimer (C2h)	3.765	6.952	6.096	3.643	3.499	3.753
12	pyrazine dimer (Cs)	3.479	4.848	4.760	3.695	3.437	3.681
13	uracil dimer (C2)	3.166	5.805	6.732	3.097	3.406	3.007
14	indole benzene (C1)	3.498	5.572	5.520	4.448	3.415	4.378
15	adenine thymine stack (C1)	3.172	6.202	5.788	4.320	3.280	3.099
			Mixed Complexes				
16	ethene ethine (C2v)	2.752	2.468	2.429	2.319	2.366	2.374
18	benzene NH3 (Cs)	3.592	4.092	4.025	2.995	3.069	3.014
19	benzene HCN (Cs)	3.387	3.472	3.694	3.228	3.343	3.303
20	benzene dimer (C2v)	3.513	5.225	3.606	3.253	3.370	3.351
21	indole benzene T-shape (C1)	3.210	3.811	3.807	3.010	3.233	3.208
22	phenol dimer (C1)	1.937, 4.921	2.174, 5.925	1.829, 5.712	2.001, 5.040	1.778,5.265	2.016, 4.937
	RMSE (hydrogen bonded)		0.387	0.257	0.137	0.134	0.129
	RMSE (dispersion bonded)		2.015	1.962	0.644	0.272	0.473
	RMSE (mixed bonded)		0.929	0.598	0.336	0.315	0.301
	RMSE		1.277	1.171	0.419	0.249	0.326
	MUE		0.853	0.691	0.301	0.199	0.222

[a] The AM1-D and PM3-D results have been taken from ref 19.

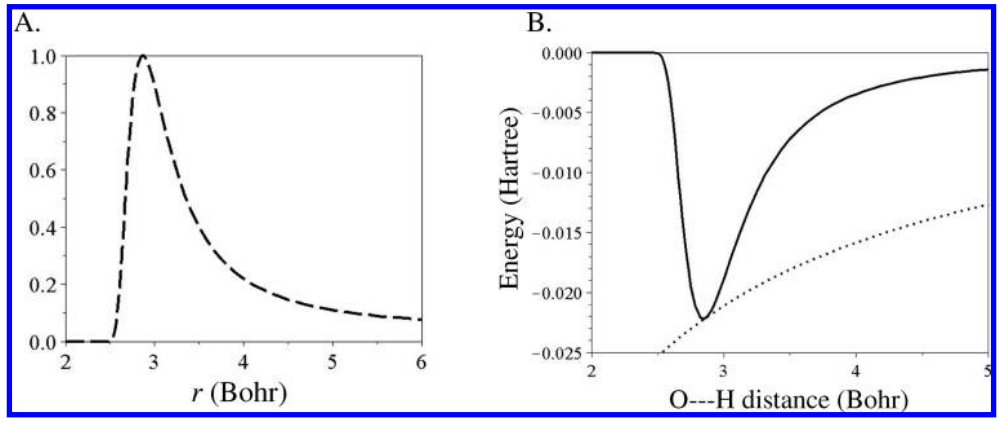

Figure 2. Graphical representation of the H-bonding damping function (A), the entire correction term (solid-line), and the electrostatic attractive portion (dotted-line) (B) used in the AM1-FS1 method. This model is for the case of the α-hydrogen atom (connected to the nitrogen atom) interacting with the parallel oxygen atom on the second monomer of the uracil dimer in the hydrogen bonding conformation. The MOPAC charges used correspond to the minimum energy structure (O---H, $R = 1.77$); this simplification has little effect on the functional form. This simplification has been used for graphical convenience.

by the observation that H-bonding interactions are directionally dependent.[31] The cosine squared function was used instead of the cosine function because it approaches zero smoothly. We have also chosen to make the function zero for all angles less than 90°. This helps exclude cases that are not H-bonding, such as an α-H atom in a carboxylic acid interacting with the adjacent carboxylate O atom. By using an appropriate summing scheme, we are able to identify highly likely H-bonding scenarios without knowledge of atom connectivity. This is done by first identifying H atoms for which the nearest nieghbor is an N, O, or F atom. These H atoms are then allowed to interact with other N, O, or F atoms.

The overall function (eq 4) is shown graphically in Figure 2B, where it can be seen that the function is only turned on over a short-range, peaking at approximately 2.8 bohr (1.5 Å; for the specific bonding scenario depicted). This is the behavior that is needed to improve the AM1 method for H-bonding, since these interactions only need to be increased over a short range and only at short distances. The nature of the charges (MOPAC) of the atoms involved in H-bonding, insures that eq 4 is negative, resulting in an attractive contribution. The charges are updated every optimization step. The optimization procedure also requires the gradient, which is determined by numerical differentiation. This correction

scheme to the AM1 method has been implemented into a locally modified version of GAMESS.[29]

The hydrogen bonding correction scheme as described above is continuous for proton transfer under most conditions. In most cases, the correction term effectively turns off (i.e., is essentially equal to zero) before the proton reaches the halfway point in a proton transfer. For example, when a proton transfers between two formic acid molecules, the intercomponent oxygen−oxygen separation is about 2.7 Å; therefore, when the proton reaches the halfway point, the H---Y distance is about 1.35 Å (2.55 bohr) and the H-bonding correction term is approximately zero (see Figure 2B). The function is also continuous when the molecules are separated by a greater distance even though there is a nonzero correction at the halfway point because at this point the function is identical in both directions (H---Y equals X---H). When the proton passes the halfway point, the H-bonding correction term corrects in the opposite direction. If a proton is transferring across an asymmetric system, however, a discontinuity can occur since the charges on the X and Y atoms may not be the same. This discontinuity can be eliminated by evaluating eq 4 in both directions at all times and taking the correction to be a weighted sum of the two. This correction has been implemented into AM1-FS1 and has no effect on any of the binding energies reported in this manuscript, since evaluating the function in the opposite direction (considering the X---H bond to be H-bonding) leads to no correction because the X---H distance is essentially always less than 1 Å (1.9 bohr). In summary, the switching transition from one H-bonding situation to another is effectively continuous during a proton transfer. We currently cannot recommend AM1-FS1 for modeling proton transfers, since it has not been tested and more importantly because our training set does not contain data to parametrize for such situations. Nevertheless, this correction scheme does not produce discontinuities. High-quality (CCSD(T) and DFT-SAPT) proton transfer potential energy curves are scarce, rendering such a parametrization difficult at this time. We plan to explore this avenue in the future.

2.3. Parameter Optimization. To improve the AM1 method for dispersion and hydrogen-bonding interactions, two empirical correction terms have been added as discussed above. These two correction terms involve a total of six parameters: two for the dispersion term (eq 1) and four for the H-bonding term (eq 4). These six parameters have been mathematically optimized to the RMSE of the interaction energies of 66 complexes (the F66 training set, see Table S1, Supporting Information). All of the interaction energies in the training set are CCSD(T) or SAPT quality. The training set consists of complexes not only at their minimum energy structures but also at greater and lesser separation than the potential minimum. Inclusion of these nonequilibrium structures is intended to increase the reliability of geometry optimization with AM1-FS1.

Our F66 training set includes the complexes in the S22 database,[20] which has been used by others for similar parametrization purposes.[19,22,32] We have also included the four additional H-bonded complexes[33] that were later introduced to the S22 database, now termed the S26 database. The additional interaction energies are also CCSD(T) quality. In our F66 training set, the water dimer, T-shaped benzene dimer, and both uracil dimer structures from the S22 database have been replaced by five points on their respective interaction potential energy curves. Five-point sampling of the potential energy curves has also been added for the nitromethane dimer,[34] parallel[23] and M1[35] benzene dimer, and three different benzene−acetylene[36] conformations. For a detailed list of complexes in the training set, refer to Table S1 in the Supporting Information. It would be desirable to have more potential energy curves in the training set, but there is limited high quality data available. For training set proposes, we have restricted ourselves to using only CCSD(T) or SAPT results, and only at or near the complete basis set limit.

Upon optimization of the parameters, the damping coefficient (d) in eq 2 optimized to infinity. This is because the AM1 method, as well as other semiempirical methods, inaccurately models repulsive interactions at close range for dispersion bound complexes. This can be observed by comparing DFT and semiempirical (AM1, PM3, RM1, and PM6) potential energy curves for the parallel benzene dimer, as shown in Figure 1. The figure clearly shows that at close separation the semiempirical methods (AM1, PM3, RM1, and PM6) differ significantly from the DFT (BLYP/6-311G(d,p)) results, severely underestimating the repulsion at close separations. The inaccurate repulsive inner wall of the potential is a consequence of the minimal basis set and parametrization of the SE methods.[25] This inaccuracy is the origin of d optimizing to infinity. As d becomes larger, the dispersion correction is turned off more rapidly; however, the function cannot become positive as needed to correct for underestimation of the repulsion at short range by the SE method. This problem could potentially be improved if a 6-12 LJ potential was used and only intercomponent atom pairs were considered; however, this introduces the requirement of atom connectivity information. It would also introduce a discontinuity in the potential and/or its derivative during bond breaking and formation processes (not to mention that intracomponent dispersion interactions would be neglected completely, thereby rendering the method ineffective for modeling conformational preference in macromolecules). We have therefore chosen to set d equal to 1000 and fully optimize the other five parameters. The damping coefficient was chosen to be 1000, because this is at the computational limit for evaluating the derivative of eq 1 within double precision. (Derivative information is needed for structural optimizations.) The other five parameters optimized to the following values: $S_6 = 1.1059$, $\alpha_1 = 0.4882$, $\alpha_2 = 0.6211$, $\alpha_3 = 0.3344$, and $\alpha_4 = 1.5451$.

2.4. Why Begin with AM1? AM1 has long been accepted as one of the most robust semiempirical methods. This method has been used many times with success for modeling large systems, but this is not the only reason for choosing AM1. We applied the same correction scheme described above to the RM1 method, which is a reparameterized version of AM1. The "corrected" RM1 method was actually less successful, on the basis of the RMSE for the F66 training set. Upon further investigation, we found that the RM1

method (as well as the PM3 and PM6 methods) performs worse than the AM1 method for the benzene dimer when compared to the DFT results that neglect dispersion interactions, as discussed above. Thus, if the same dispersion correction is applied to these mentioned SE methods, the AM1 method will produce the best result, even though other uncorrected SE methods produce potential energy curves closer to the CCSD(T) results. This is because the AM1 method has the strongest repulsive wall, therefore, producing a potential energy curve closest to the DFT result (see Figure 1). The functional form of the dispersion correction (eq 1) does not allow the term to become positive as is needed in some cases. This can be easily seen for the PM6 results in Figure 1. At close range, the CCSD(T) results are more repulsive then the PM6 results; thus to make the PM6 curve identical to the CCSD(T) curve, a repulsive correction would be needed. This problem is less severe for the AM1 method, rendering it more suitable for modeling dispersion interactions at close range. Note that these findings again might lead one to believe that using a function like the LJ potential would be beneficial, since a repulsive term is included. In fact, the LJ potential was among our many attempts to improve the AM1 method, but without success. This was due to the fact that we were/are unwilling to add the burden of requiring atom connectivity information. We believe such a burden outweighs the potential added benefit.

3. Validation Studies

3.1. Single-Point Energies. In Table 1, the single-point interaction energies for the structures in the S22 database[20] are compared for various corrected and uncorrected SE methods. First, note that the interaction energies for the uncorrected AM1 and PM3 methods deviate significantly from the CCSD(T) reference values. The root-mean-square error (RMSE) for the AM1 and PM3 methods are 8.47 and 7.73 kcal/mol, respectively. Not only are these errors very large, but in many cases the sign of the interaction is predicted incorrectly. That is, the interactions are predicted to be repulsive not attractive. The addition of an empirical correction term(s) can drastically improve these methods. Our AM1-FS1 method reduces the RMSE to 1.18 kcal/mol, with the correct sign being predicted in all cases.

The results from McNamara and Hillier's[19] AM1-D and PM3-D methods and the PM6-DH method of Řezáč et al.[21] are also reported in Table 1. AM1-FS1 shows a slight improvement over the AM1-D method in two of the three subcategories and overall has a lower RMSE. AM1-FS1 achieves comparable accuracy to the PM3-D method for intermolecular interaction energies; the RMSEs are both 1.18 kcal/mol, with AM1-FS1 achieving a slightly lower MUE. While the overall improvement achieved by AM1-FS1 in the accuracy with which intermolecular binding energies are predicted is minor, we note that this has been achieved with significantly less parametrization and *no* modification of the original AM1 parameters.

The recently published PM6-DH method[21] slightly outperforms AM1-FS1 on the basis of the single-point energies for the S22 database. Looking at the hydrogen bonded complexes, the RMSEs are 1.07 and 1.37 kcal/mol for the PM6-DH and AM1-FS1 methods, respectively. Given that PM6-DH requires *different* parameters for each type of hydrogen bond, the 0.3 kcal/mol improvement in RMSE shown by PM6-DH is not especially significant. The group has identified eight H-bonding scenarios resulting in a total of 24 parameters for their H-bond correction term (three parameters for each H-bonding type). AM1-FS1 only uses four parameters; AM1-FS1 also does not introduce the requirement of knowing atom conductivity. The PM6-DH method does show significant improvement for many dispersion bonded cases, but it performs poorly for modeling the potential energy surface of the benzene dimer. This is discussed below and shown graphically in section 3.4. It should be noted that Řezáč et al.[21] only used complexes 8−22 of the S22 database for determining the dispersion parameters for PM6-DH. Thus, it is not unexpected that good agreement was achieved for the eight dispersion bound complexes.

3.2. F66 Results. Many of the other empirically corrected SE methods discussed have been parametrized to the S22 database, thus they should achieve accurate results for those complexes. AM1-FS1 has been parametrized to a larger training set consisting of 66 complexes. Parameterizing to this larger training set has led to an increase in RMSE for the S22 database. This is not unexpected and, in our opinion, is a worthwhile sacrifice that should make AM1-FS1 more versatile. (In fact, we tried optimizing AM1-FS1 solely to the S22 database and achieved near DFT-D level accuracy, but when the method was subsequently tested on the F66 training set, a larger RMSE resulted.) AM1-FS1 is parametrized to the F66 training set and achieves a sub-kilocalorie RMSE (0.99 kcal/mol) and MUE (0.69 kcal/mol). The individual results are reported in Table S1 in the Supporting Information. Both AM1-D and PM3-D were parametrized solely to the S22 database, so high accuracy is not surprising when the S22 database is used as the "test set". We have performed calculations on the 66 complexes of the F66 set using McNamara and Hillier's AM1-D method for comparison. This provides for a much more comprehensive test of the method than the S22 database because it contains a wider variety of structures *and* nonequilibrium structures. AM1-D produces a RMSE and MUE of 1.49 and 1.02 kcal/mol respectively, approximately 50% less accurate than AM1-FS1. Upon close inspection of Table S1, it can be seen that AM1-FS1 significantly outperforms AM1-D on the repulsive wall, an issue we will look at more closely in section 3.4.

3.3. Optimized Energies and Structures. This section considers the effect of geometry optimization on interaction energy and structural distortion for systems in the S22 database. The ability of an empirically corrected SE method to perform accurately in this role is crucial because one of the principal uses of SE methodology is structural optimization of systems that are too large for optimization with first-principles methods. The ability of a method to reproduce interaction energies at *reference* geometries is not very useful, because if we know the CCSD(T) geometry, and therefore its energy, there is little value in knowing the SE energy for that structure. In Table 2, the interaction energies for the

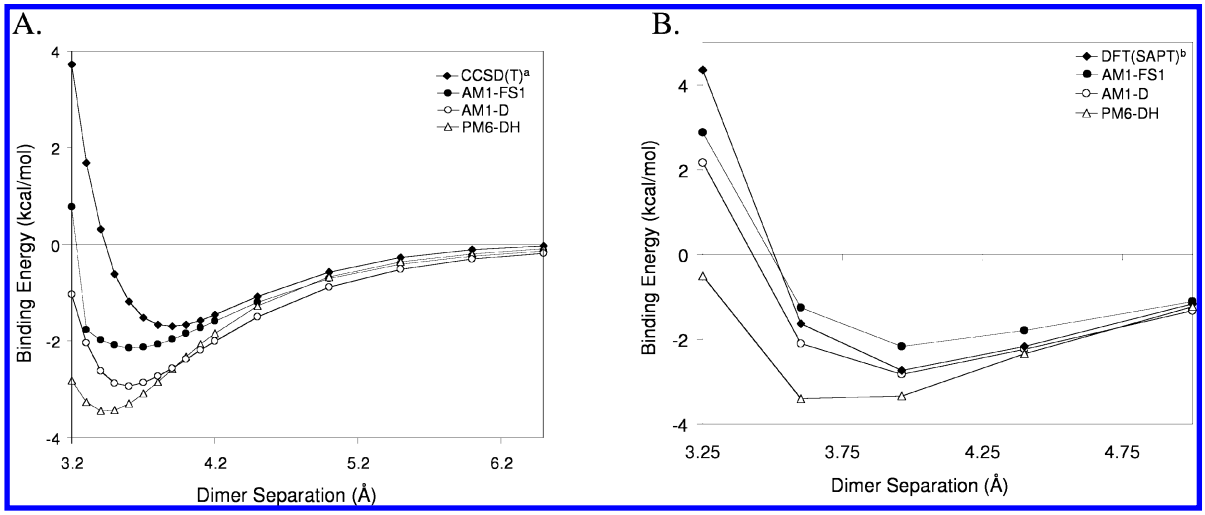

Figure 3. Parallel and M1 benzene dimer potential energy curves determined with various computational methods. Superscript a refers to ref 23 and b to ref 35.

geometry optimized complexes are reported for a variety of corrected and uncorrected SE methods. Again, the uncorrected AM1 and PM3 methods perform poorly. Upon applying our correction scheme to the AM1 method, the RMSE is lowered from 6.25 to 1.80 kcal/mol. AM1-FS1 is also an improvement over AM1-D; AM1-FS1 outperforms AM1-D in all subcategories by about 1 kcal/mol. This increase in performance for optimization, compared to the AM1-D method, presumably results from our use of a substantially larger training set that includes nonequilibrium structures.

The performance of McNamara and Hillier's PM3-D method is comparable to our AM1-FS1 method. Depending of the statistical metric selected, either may be said to outperform the other for predicting interaction energies upon geometry optimization of the structure in the S22 database. AM1-FS1 does perform better in two of the three categories, the dispersion and mixed bounded complexes. The PM6-DH method outperforms all the other methods. However, structural distortion should also be considered, but unfortunately, data for such a comparison are not available. Again, this aspect of a SE method is especially important since such a method will likely be used for optimization purposes.

To gauge the degree of structural distortion upon geometry optimization, select interaction distances are compared and are shown in Table 3. The interaction distances are defined as the center-of-mass separation and/or atom−atom distance(s) between the two monomers depending on the system (see Figure S1 of ref 19 for the specific interaction distances). Comparing the different empirically corrected SE methods, we find that AM1-FS1 outperforms AM1-D in every category. Our method is generally comparable to PM3-D on the basis of interaction distance. AM1-FS1 performs better in two of the three categories. This time, AM1-FS1 outperforms PM3-D for the H-bonded complexes on the basis of interaction distances, but not for the dispersion bound complexes. Based solely on the total RMSE for the S22 database would be difficult to choose which method, AM1-FS1 or PM3-D, is better; however, AM1-FS1 does not require reoptimization of the AM1 parameters thereby preserving the predictive power of AM1 for calculation of heats of formation, discussed below. As noted above, interaction distances and/or structural geometries were not made available for the PM6-DH method preventing structural comparisons upon optimization of the S22 complexes.

To further test the ability of AM1-FS1 to model H-bonding complexes, 16 additional hydrogen bonded DNA base pairs have been considered. The 16 additional complexes were chosen from ref 20 since these are the only complexes from the H-bonding subsection that have CCSD(T) quality binding energies. The geometries of these complexes, however, are from MP2 optimizations or experimental data. Therefore, these structures do not correspond to the CCSD(T) potential minimum; this is also the case for most of the S22 database structures. We have computed the binding energies for these complexes on the basis of the reference geometries and also AM1-FS1 optimized geometries. The RMSEs for the binding energies are 1.78 and 2.18 kcal/mol, respectively. This error is consistent with the error associated with the hydrogen bonding complexes in the S22 database, which were used for parametrization. The 16 complexes as well as the reference CCSD(T), single point, and optimized AM1-FS1 binding energies are reported in Table S2 of the Supporting Information.

3.4. Potential Energy Curves. The value of a computational method is significantly enhanced if it is able to accurately describe the potential energy surface apart from the minimum. A given method could accurately predict the interaction energy at a specific molecular geometry yet yield a very inaccurate picture of the remainder of the potential energy surface. (See ref 25 for a detailed discussion.) In this section, potential energy curves will be compared for various empirically corrected SE methods.

In Figure 3, potential energy curves for two different benzene dimer conformations are shown. Figure 3A shows the interaction energy for the parallel dimer as a function of monomer separation. The parallel dimer is not the lowest energy conformation, but it is important to be able to model a variety of geometries correctly for the correct description of $\pi-\pi$ interactions involved in large systems, and the

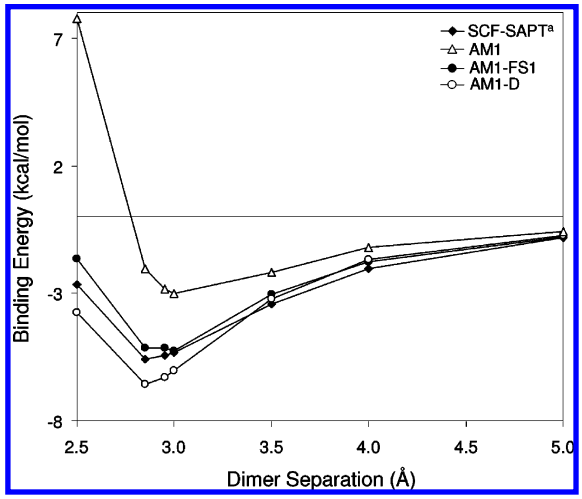

Figure 4. Water dimer potential energy curves as a function of O---O separation, as determined with various computational methods. Superscript a refers to ref 37.

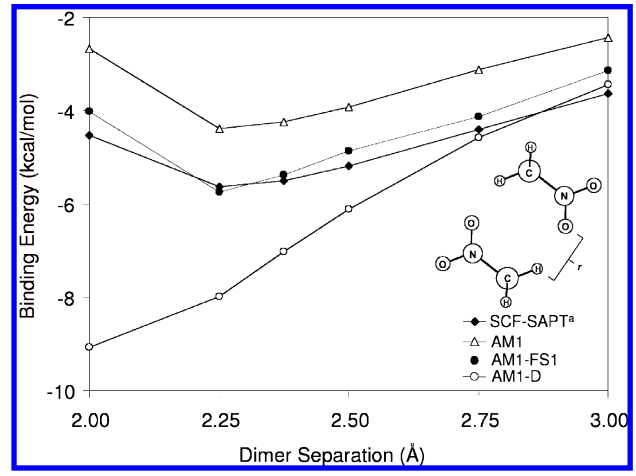

Figure 5. Nitromethane dimer potential energy curve in the "double hydrogen bond" configuration, as determined with various computational methods. Superscript a refers to ref 34.

parallel dimer represents a widely used test case, probably owing to the simplicity of its construction. The M1 benzene dimer, according to ref 35, is the lowest energy structure known. Compared to the CCSD(T) and DFT(SAPT) reference values, among the empirically corrected SE methods PM6-DH performs the worst for these systems. The PM6-DH method seems to overbind $\pi-\pi$ interactions. This is due to the fact that the PM6 method performs poorly for dispersion bound complexes compared to DFT-BLYP results as discussed earlier and shown in Figure 1 (further discussion is presented in ref 25). McNamara and Hillier's AM1-D method performs very well for the M1 dimer, however, not so well for the parallel dimer. On average, AM1-FS1 performs the best for these two systems. AM1-FS1 has a very steep potential wall at close separation (Figure 3A); this is an artifact of using a large damping parameter (d) in the dispersion correction term (eq 2). Again, the large term is required because of the inability of the AM1 method to properly capture short-range repulsive interactions.

In Figures 4 and 5, potential energy curves for the water dimer and the nitromethane dimer are shown, respectively. The water dimer is a classic hydrogen bonding system. The potential energy curves in Figure 4 are shown as a function of O---O separation. The figure shows that AM1-FS1 dramatically improves upon the AM1 method and outperforms McNamara and Hillier's AM1-D method. The correlation to SCF-SAPT results[37] again shows that the hydrogen bonding correction term (eq 4) is a worthwhile addition to the AM1 method. The AM1-D method also performs relatively well for the water dimer. This means that the changes they have made to the AM1 parameters improve the results for this particular system; however, the same is not observed if we consider the nitromethane dimer. In Figure 5, we see that AM1-D performs poorly for the nitromethane dimer. The SCF-SAPT curve[34] is for the so-called "double hydrogen bond" configuration; however, nitromethane is not a classical H-bonding system. It lacks a hydrogen atom attached to a highly electronegative atom (N, O, or F); nevertheless, this system is said to form weak H bonds.[34] As shown (Figure 5), the AM1 method performs relatively well for this system, whereas McNamara and Hiller's modification of the AM1 parameters has caused the AM1-D method to inadequately model this system. AM1-FS1, on the other hand, does not consider this a H-bonding case. Therefore, the H-bonding correction term is not turned on for this system. Consequently, AM1-FS1 performs well for this system by applying only the dispersion correction. This potential energy curve demonstrates that the AM1 parameters should not be changed in all cases. It should be noted that the AM1-D training set does not contain this system, whereas the training set for AM1-FS1 does. (We have not compared the PM6-DH method of Řezáč et al.[21] in H-bonding cases because we do not have code for their elaborate H-bonding correction scheme.)

3.5. Heat of Formation. As mentioned earlier, modifying the original semiempirical parameters can have deleterious effects, especially for thermodynamic properties. For example, the experimental heat of formation of benzene is 19.8 kcal/mol[38] and is predicted to be 22.0 kcal/mol by the AM1 method.[15] The AM1-D method, however, predicts a value of −12.9 kcal/mol. (PM3-D performs even more poorly, yielding −21.8 kcal/mol.) Reparameterization has rendered AM1-D (and PM3-D) unreliable for predicting thermodynamic properties. On the other hand, AM1-FS1 does not change any of the original AM1 parameters and predicts the heat of formation of benzene to be 20.0 kcal/mol, in good agreement with experimental results and, serendipitously, even a slight improvement over AM1. The AM1-FS1 empirical correction is designed to have little effect on quantities that are already predicted relatively well by the AM1 method. Table 4 collects results for calculations of heat of formation on 53 test molecules. Note that the RMSE in predictions of heat of formation with AM1-FS1 is comparable to that of the original AM1 method, but AM1-D is 24 times (2400%) less accurate. Reparameterization of the original PM3 method in the development of PM3-D has also seriously degraded its predictive power for heats of formation (see Table 4). This clearly shows the negative consequences of changing the original semiempirical parameters without

Table 4. Heat of Formation (kcal/mol)[a]

	heat of formation (kcal/mol)				
molecule	expt	AM1	AM1-D	PM3-D	AM1-FS1
methane	−17.8	−8.8	−78.4	−6.6	−8.8
ethane	−20.0	−17.4	−114.1	−12.2	−18.3
ethylene	12.5	16.5	−39.4	−11.0	16.1
acetylene	54.5	54.8	38.6	−9.8	54.8
propane	−25.0	−24.3	−148.4	−17.7	−27.1
propene	4.8	6.6	−75.6	−16.5	5.3
propyne	44.2	43.4	3.4	−15.3	43.0
allene	45.5	46.1	6.6	−15.3	45.7
n-butane	−30.0	−31.1	−182.7	−23.2	−36.1
isobutane	−32.0	−29.4	−181.1	−23.2	−35.3
but-1-ene	−0.1	0.4	−109.2	−22.0	−2.6
trans-2-butene	−2.8	−3.3	−111.8	−22.0	−5.6
cis-2-butene	−1.7	−2.2	−110.9	−22.0	−4.8
isobutene	−4.0	−1.2	−109.8	−22.0	−4.3
1,2-butadiene	38.8	37.1	−28.5	−20.8	35.9
trans-1,3-butadiene	26.3	29.9	−38.2	−20.8	28.3
1-butyne	39.5	37.5	−29.8	−20.8	35.6
2-butyne	34.8	32.0	−31.8	−20.8	31.1
vinylacetylene	72.8	67.9	42.1	−19.6	66.9
diacetylene	113.0	106.1	122.4	−18.4	105.8
n-pentane	−35.1	−37.9	−216.9	−28.7	−45.1
neopentane	−40.2	−32.8	−212.3	−28.7	−42.8
benzene	19.8	22.0	−12.9	−29.6	20.0
toluene	12.0	14.5	−46.6	−35.1	10.7
ammonia	−11.0	−7.3	−154.0	−7.7	−7.3
methylamine	−5.5	−7.4	−165.2	−13.3	−8.2
dimethylamine	−4.4	−5.6	−175.1	−18.8	−7.5
trimethylamine	−5.7	−1.7	−183.3	−24.3	−5.1
ethylamine	−11.3	−15.1	−200.4	−18.8	−17.5
n-propylamine	−16.8	−22.1	−234.7	−24.3	−26.5
isopropylamine	−20.0	−19.2	−231.9	−24.3	−23.9
tert-butylamine	−28.9	−21.2	−261.6	−29.8	−29.3
pyrrole	25.9	39.9	−56.0	−26.3	38.5
pyridine	34.6	32.1	−29.5	−30.7	30.6
pyridazine	66.5	55.3	−33.6	−31.8	54.2
water	−57.8	−59.2	−200.8	−12.0	−59.2
methanol	−48.2	−57.0	−193.7	−17.5	−57.5
ethanol	−56.2	−62.7	−225.8	−23.0	−64.4
1-propanol	−61.0	−70.6	−261.9	−28.5	−74.7
2-propanol	−65.2	−67.7	−258.2	−28.5	−71.8
t-butyl_alcohol	−74.7	−71.6	−288.7	−34.0	−78.6
dimethyl_ether	−44.0	−53.2	−185.7	−23.0	−53.8
diethyl_ether	−60.3	−64.4	−249.6	−34.0	−67.8
oxirane	−12.6	−8.9	−95.3	−21.8	−9.2
furan	−8.3	3.0	−52.7	−30.5	2.1
phenol	−23.0	−22.2	−120.0	−40.4	−24.9
anisole	−16.2	−15.8	−110.1	−45.9	−19.8
benzaldehyde	−8.8	−8.9	−58.5	−44.8	−12.1
formic_acid	−90.5	−97.4	−222.7	−27.2	−97.4
acetic_acid	−103.4	−103.0	−252.9	−32.7	−103.7
propionic_acid	−108.4	−108.0	−285.5	−38.2	−111.1
oxalic_acid	−173.0	−172.4	−370.9	−53.2	−172.8
benzoic_acid	−70.3	−68.0	−181.4	−55.6	−71.0
RMSE		4.9	132.5	46.7	5.6
MUE		3.6	118.4	36.1	4.3

Binding Energies (kcal/mol) of Carbon Nanostructure Complexes[c]

	AM1-FS1	AM1-D	M06-2X/6-31+G(d,p)// M06-L/MIDI
HMB@6CPPA	−16.6	−17.6	−14.7
C60@6CPPA	−26.9	−30.1	−28.0
C70@6CPPA	−36.3	−41.0	−31.1
3,3@6CPPA	−17.7	−22.1	−5.4
4,4@6CPPA	−32.7	−42.0	−24.0
5,5@6CPPA	−43.2	−46.4	−43.3
C60@BuckyCatcher[b]	−29.3	−36.8	−26.4
C60@Coroanulene[b]	−13.4	−16.7	−12.4

[a] The experimental and AM1 results were obtained from ref 15. Note: the AM1-D and PM3-D results were obtained by coding the method outlined in ref 19; however, slight disagreements in the binding energies were observed with PM3-D for compounds containing oxygen, suggesting a misprint in the published PM3-D oxygen parameters. [b] Results were obtained from ref 45. [c] The M06 results were obtained from ref 44.

the consideration of such quantities during optimization of the parameters.

While the original AM1 parameters have not been altered in AM1-FS1, the FS1 correction terms do influence the predicted heat of formation. This occurs because the heat of formation is in part determined from the total energy of the complex, which now contains the empirical correction energy, but also in part from the energies of the isolated atoms. The isolated atom energies do not include any empirical correction energy since; by design, the correction terms are not implemented for a single atom. Therefore, the difference between the heat of formation as computed with AM1 and AM1-FS1 will generally become larger as the correction term(s) contribution increases. This will also be the case for DFT-D methods when the total energy is used in the determination of the heat of formation. The influence of the FS1 correction on the predicted heat of formation can have both undesirable and desirable consequences. For example, in Table 4, it can be seen that as the number of methylene units in the aliphatic hydrocarbons increases (methane → ethane → propane → etc.), the error in the predicted heat of formation increases. Fortunately, since the original AM1 parameters have not been altered, the AM1 heat of formation can be easily obtained by subtracting out the empirical correction energy from the AM1-FS1 heat of formation. This approach of subtracting the correction energy will also be effective for DFT-D methods when the total energy is used in the determination of the heat of formation. On the other hand, the correction to the total energy sometimes has a beneficial impact on the predicted heat of formation. For example, the experimental heat of formation of the benzene dimer is 30.4 kcal/mol[39] and is predicted to be 37.9 and 44.1 kcal/mol on the basis of structures optimized with the AM1-FS1 and AM1 methods, respectively. Note that these structures are significantly different upon optimization because the AM1 method does not consider dispersion interactions. If we determine the heat of formation of the AM1-FS1 optimized geometry with the AM1 method, the error is even larger; the heat of formation is predicted to be 46.7 kcal/mol. Since the original AM1 parameters were optimized to give reliable predictions of the heat of formation at AM1-optimized geometries, it seems reasonable to conclude that, in general, if the heat of formation of some large multicomponent carbon structure is desired, the AM1-FS1 method will likely produce a more accurate result at the AM1-FS1 geometry. Certainly, the AM1-FS1 structure will be more accurate since the dominant intercomponent interaction will be incorporated. This situation is much less complicated when a heat of reaction is of interest since the correction term(s) are applied to both the reactants and products.

4. Application to Macromolecular Complexes

The ultimate goal of AM1-FS1 is to be able to efficiently and accurately model large weakly bound systems, such as complexes of carbon nanostructures and molecular devices. Such large systems are currently out of reach for CCSD(T), and extremely computationally demanding for DFT methods. Furthermore, most DFT functionals are incapable of model-

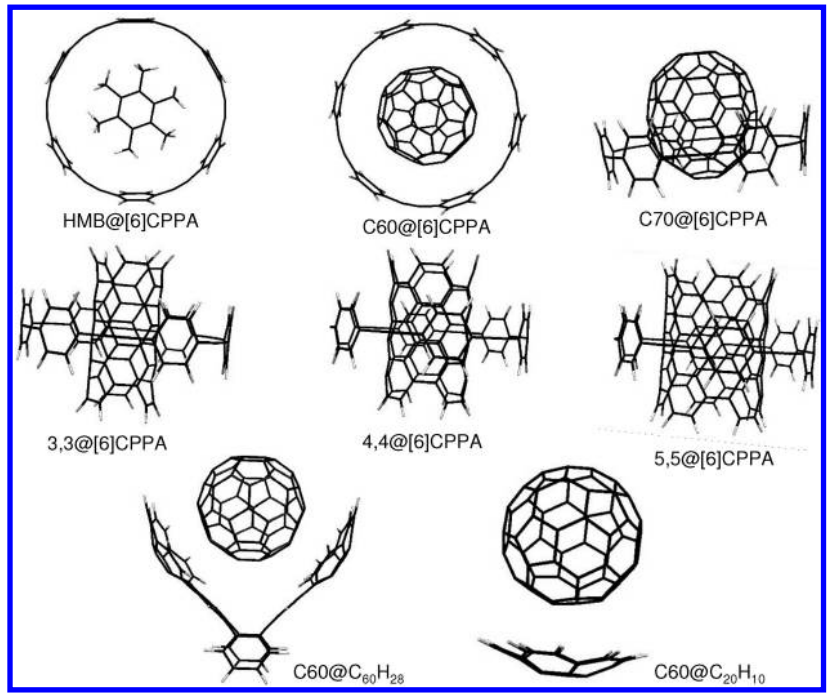

Figure 6. AM1-FS1 geometry optimized carbon nanostructure complexes.

ing carbon nanostructure complexes due to the fact that these systems are governed by van der Waals interactions. Presumably, the current most accurate methods capable of modeling such systems are DFT-D methods and the M05,[40] M06,[41] and M08[42] family of functionals developed by Zhao and Truhlar. Performing geometry optimizations with DFT-based methods on systems larger than 100 atoms is currently extremely computationally expensive, but such optimizations can be routinely performed with semiempirical-based techniques even on "PC"-class computers. This great computational efficiency of semiempirical methods provides the central motivation for the present work.

4.1. Carbon Nanostructures. To test the performance of AM1-FS1 on complexes of carbon nanostructures, we have performed geometry optimizations and determined the binding energy of several inclusion complexes. The hosts considered are corannulene ($C_{20}H_{10}$), a double-concave hydrocarbon buckycatcher[43] ($C_{60}H_{28}$), and cyclic[6]paraphenylacetylene (6CPPA). The AM1-FS1 optimized structures are shown in Figure 6. (It is important to note that these complexes would be predicted to be unbound if the standard AM1 and most DFT methods were used.) To date, the best binding energy values for these complexes are from DFT calculations reported by Zhao and Truhlar,[44,45] using the M06-2X functional. The binding energies along with the AM1-FS1 and AM1-D results are reported in Table 4. The AM1-FS1 results are very comparable to the M06 values; however, AM1-FS1 overestimates the binding energy for 3,3@[6]CPPA and 4,4@[6]CPPA in comparison to DFT-M06-2X. This may be a result of the M06 functional underestimating dispersion interactions at long range. This hypothesis is supported by the potential energy curve for the parallel benzene dimer shown in Figure 7, which clearly shows that the M06 functional fails to accurately model dispersion interactions in the long-range regime, where the predicted interaction energy even becomes slightly positive.

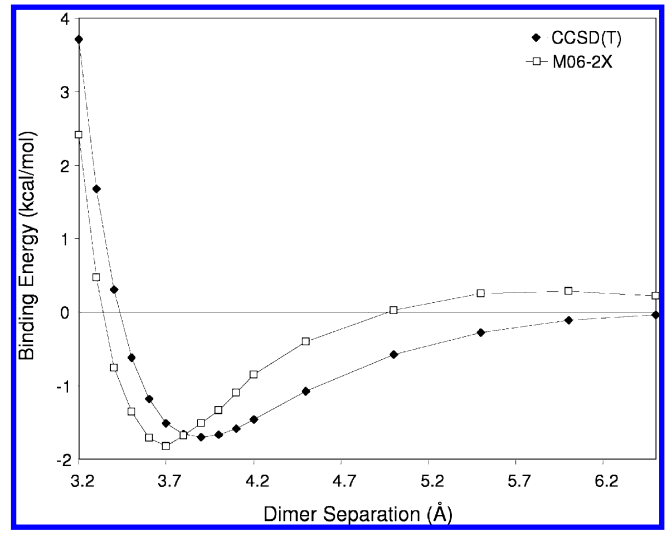

Figure 7. Parallel benzene dimer potential energy curve calculated with the M06-2X functional, using the 6-311G(d,p) basis. CCSD(T) results were obtained from ref 23.

This behavior might be easily overlooked since upon optimization of the parallel benzene dimer, a reasonable energy and structure will be produced. To show that this is the case for 3,3@[6]CPPA, the binding energy was determined using the BLYP-D functional. The resulting binding energy of 18.9 kcal/mol is in very good agreement with the AM1-FS1 result. The M06-2X functional underestimates binding for 3,3@[6]CPPA because the nearest intermolecular interaction is 4.5 Å, a distance at which M06 underestimates the interaction energy as exhibited by the benzene dimer potential energy curve.

The AM1-FS1 method significantly outperforms AM1-D in every case on the basis of the current benchmark M06 values. AM1-FS1 achieves this correlation with only two added parameters (due to the nature of the systems only the

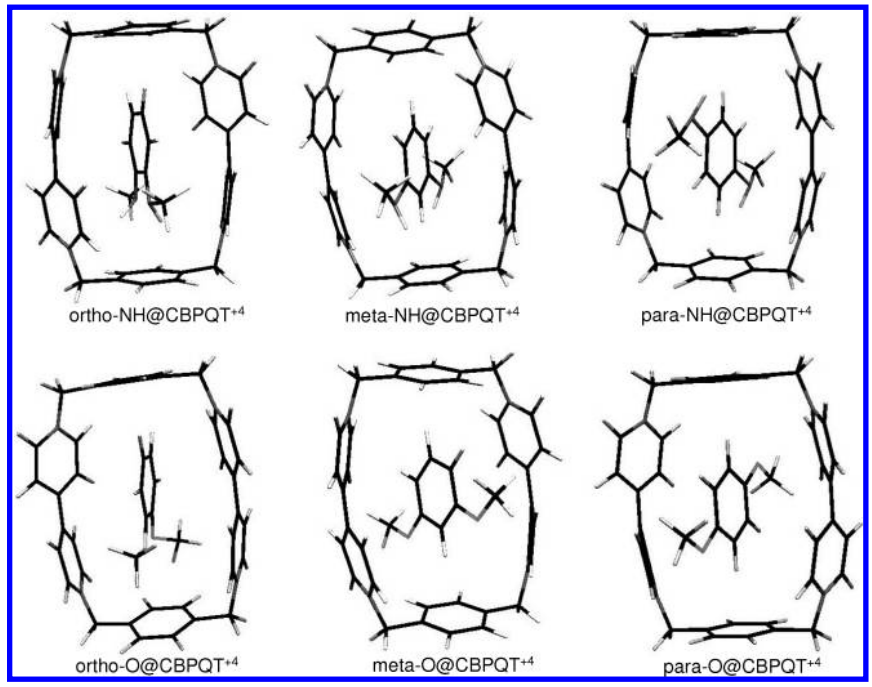

Figure 8. AM1-FS1 geometry optimized pseudorotaxane complexes.

Table 5. Binding Energies (kcal/mol) of Pseudorotaxane Complexes[a]

	Binding Energy (kcal/mol)		
	AM1-FS1	M06−2X/6-311G(d,p)//M06-L/MIDI	LMP2/6-311+G(d,p)//BHandHLYP/6-31G(d)
ortho-O@CBPQT^{+4}	−32.1	−34.7	−21.0
meta-O@CBPQT^{+4}	−31.7	−35.0	−16.1
para-O@CBPQT^{+4}	−33.4	−34.7	−21.3
ortho-NH@CBPQT^{+4}	−37.7	−41.7	−22.3
meta-NH@CBPQT^{+4}	−38.2	−36.9	−22.5
para-NH@CBPQT^{+4}	−38.5	−40.1	−23.9

[a] The LMP2 results were obtained from ref 46.

dispersion correction term is "turned on" during the AM1-FS1 calculations), whereas AM1-D utilizes 10 parameters. We credit the success of AM1-FS1 to parametrizing to a larger training set containing nonequilibrium complexes.

4.2. Pseudorotaxanes. We also tested the performance of AM1-FS1 on six different pseudorotaxanes, since these types of complexes are of central interest to our research group. All of the systems considered incorporate cyclobis-(paraquat-p-phenylene) (CBPQT^{+4}), a tetracationic ring structure. Six inclusion complexes with this ring have been formed with dimethoxybenzene and benzenedimethanamine in the *ortho*, *meta*, and *para* conformations. (AM1-FS1 optimized structures are shown in Figure 8.) We have performed geometry optimizations and determined the binding energies of these complexes and compared them to previously reported LMP2/6-311+G(d,p)//BHandHLYP/6-31G(d) results.[46] We also computed these binding energies at the M06-2X/6-311G(d,p)//M06-L/MIDI level of theory for additional comparisons. (All results are reported in Table 5.) On the basis of the results, the LMP2/6-311+G(d,p)//BHandHLYP/6-31G(d) results appear to underestimate the binding energy. This is likely a result of the geometry produced by the BHandHLYP functional and not the LMP2 method. This conclusion is based on the binding energies determined at the M06-2X/6-311G(d,p)//M06-L/MIDI level. The differences in binding energies between the isomers are sufficiently small that they may be taken to be insignificant given the level of theory and the large conformational space associated with these complexes. On the basis of these results, we believe AM1-FS1 is a valuable tool for modeling this class of macromolecular complexes.

5. Conclusions

AM1-FS1 is a new empirically corrected semiempirical method suitable for performing geometry optimizations on macromolecular complexes. AM1-FS1 displays considerable improvement over the traditional AM1 method for nonbonding interactions, yet it retains the computational efficiency and predictive power for thermochemical quantities of the original AM1 Hamiltonian. Validation testing shows that the method reduces the RMSE for the popular S22 database from 8.47 to 1.18 kcal/mol. More impressively, this new method has achieved kilocalorie accuracy on a training set of 66 complexes. This was accomplished with just six empirical parameters (two for dispersion and four for hydrogen-bonding) and *no* reparameterization of AM1 (which we show here has led to serious consequences in existing empirically corrected SE methods). This is a dramatic reduction in the total number of adjustable parameters compared to other previously published empirically corrected SE methods. Validation testing shows that, while the existing PM6-DH method does outperform AM1-FS1 on the basis of the S22 database, PM6-DH is shown to be inaccurate for reproducing potential energy curves for the benzene dimer, a classic test case used for predicting the likely accuracy of a method for modeling complexes of carbon nanostructures. Moreover, unlike PM6-DH, AM1-FS1 does *not* require knowledge of atom connectivity. On the basis of the examples reported,

on average, AM1-FS1 is also the most reliable empirically corrected SE method for reproducing the potential energy curve away from the global minimum. We credit this success to using a training set that contains nonequilibrium complexes.

This new AM1-FS1 method has been shown to yield results comparable in accuracy to the best available calculations on complexes of carbon nanostructures and carbohydrate pseudorotaxanes. We believe AM1-FS1 is a useful computational tool for obtaining reliable results for such systems at limited computational expense. It should prove to be a valuable asset for routine modeling of macromolecular complexes that are currently at (or beyond) the limit of DFT based techniques, or out of reach of higher levels of theory.

Acknowledgment. This work was supported by National Science Foundation grant CHE0449595 and E. I. du Pont de Nemours & Co., Inc.

Supporting Information Available: The complete F66 training set is reported along with the interaction energies and appropriate references. The interaction energies for the F66 complexes are reported for AM1-FS1 and McNamara and Hillier's[19] AM1-D method. This information is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Vallee, R.; Damman, P.; Dosiere, M.; Toussaere, E.; Zyss, J. Nonlinear Optical Properties and Crystalline Orientation of 2-Methyl-4-nitroaniline Layers Grown on Nanostructured Poly(tetrafluoroethylene) Substrates. *J. Am. Chem. Soc.* **2000**, *122*, 6701–6709.

(2) Thalladi, V. R.; Brasselet, S.; Weiss, H.-C.; Blaser, D.; Katz, A. K.; Carrell, H. L.; Boese, R.; Zyss, J.; Nangia, A.; Desiraju, G. R. Crystal Engineering of Some 2,4,6-Triaryloxy-1,3,5-triazines: Octupolar Nonlinear Materials. *J. Am. Chem. Soc.* **1998**, *120*, 2563–2577.

(3) Hunter, C. A.; Sanders, J. K. M. The nature of pi-pi interactions. *J. Am. Chem. Soc.* **1990**, *112*, 5525–5534.

(4) Xiao, Y.; Chen, C.; He, Y. Folding Mechanism of Beta-Hairpin Trpzip2: Heterogeneity, Transition State and Folding Pathways. *Int. J. Mol. Sci.* **2009**, *10*, 2838–2848.

(5) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of Docking: An Overviewof Search Algorithms and a Guide to Scoring Functions. *Proteins: Str. Funct. Genet.* **2002**, *47*, 409–443.

(6) Jonikas, M. C.; Collins, S. R.; Denic, V.; Oh, E.; Quan, E. M.; Schmid, V.; Weibezahn, J.; Schwappach, B.; Walter, P.; Weissman, J. S.; Schuldiner, M. Comprehensive Characterization of Genes Required for Protein Folding in the Endoplasmic Reticulum. *Science* **2009**, *323*, 1693–1697.

(7) Dobson, C. M. Protein folding and misfolding. *Nature* **2003**, *426*, 884–890.

(8) Griffiths-Jones, S. R.; Searle, M. S. Structure, Folding, and Energetics of Cooperative Interactions between the beta-Strands of a *de Novo* Designed Three-Stranded Antiparallel beta-Sheet Peptide. *J. Am. Chem. Soc.* **2000**, *122*, 8350−8356.

(9) Burley, S. K.; Petsko, G. A. Aromatic-Aromatic Interaction: A Mechanism of Protein Structure Stabilization. *Science* **1985**, *229*, 23–28.

(10) Guckian, K. M.; Krugh, T. R.; Kool, E. T. Solution Structure of a Nonpolar, Non-Hydrogen-Bonded Base Pair Surrogate in DNA. *J. Am. Chem. Soc.* **2000**, *122*, 6841–6847.

(11) Foster, M. E.; Sohlberg, K. Theoretical Study of Binding Site Preference in ^{2}Rotaxanes. *J. Chem. Theory Comput.* **2007**, *3*, 2221–2233.

(12) Zheng, X.; Sohlberg, K. Modeling bistability and switching in a ^{2}catenane. *Phys. Chem. Chem. Phys.* **2004**, *6*, 809–815.

(13) Nepogodiev, S. A.; Stoddart, J. F. Cyclodextrin-Based Catenanes and Rotaxanes. *Chem. Rev.* **1998**, *98*, 1959–1976.

(14) Zheng, X.; Sohlberg, K. Modeling of a Rotaxane-based Molecular Device. *J. Phys. Chem.* **2003**, *107*, 1207–1215.

(15) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

(16) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209–220.

(17) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101–1111.

(18) Stewart, J. J. P. Optimization fo parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model* **2007**, *13*, 1173–1213.

(19) McNamara, J. P.; Hillier, I. H. Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362–2370.

(20) Jurečka, 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.

(21) Řezáč, J.; Fanfrlik, J.; Salahub, D.; Hobza, P. Semiempirical Quantum Chemical PM6Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* **2009**, *5*, 1749–1760.

(22) Jurečka, P.; Cerny, J.; Hobza, P.; Salahub, D. Density Functional Theory Augmented with an Empirical Dispersion Term. Interaction Energies and Geometries of 80 Noncovalent Complexes Compared with Ab Initio Quantum Mechanics Calculations. *J. Comput. Chem.* **2007**, *28*, 555–569.

(23) Sinnokrot, M. O.; Sherrill, C. D. Highly Accurate Coupled Cluster Potential Energy Curves for the Benzene Dimer: Sandwich, T-Shaped, and Parallel-Displaced Configurations. *J. Phys. Chem. A* **2004**, *108*, 10200–10207.

(24) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27*, 1787–1799.

(25) Foster, M. E.; Sohlberg, K. Empirically corrected DFT and semi-empirical methods for non-bonding interactions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 307–322.

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

(27) Wu, Q.; Yang, W. Empirical correction to density functional theory for van der Waals interactions. *J. Chem. Phys.* **2002**, *116*, 515–524.

(28) Halgren, T. A. Representation of van der Waals (vdW) Interaction in Molecular Mechanics Force Fields: Potential Form, Combination Rules, and vdW Parameters. *J. Am. Chem. Soc.* **1992**, *114*, 7827–7843.

(29) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(30) Chirgwin, H.; Coulson, C. A. The Electronic Structure of Conjugated Systems. VI. *Proc. R. Soc. London, Ser. A* **1950**, *201*, 196–209.

(31) Buckingham, A. D.; Bene, J. E. D.; McDowell, S. A. C. The hydrogen bond. *Chem. Phys. Lett.* **2008**, *463*, 1–10.

(32) 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.

(33) Riley, K. E.; Hobza, P. Assessment of the MP2Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes. *J. Phys. Chem. A* **2007**, *111*, 8257–8263.

(34) Cole, S. J.; Szalewicz, K., III; Bartlett, R. J. Correlated calculation of the interaction in the nitromethane dimer. *J. Chem. Phys.* **1986**, *84*, 6833–6836.

(35) Podeszwa, R.; Bukowski, R.; Szalewicz, K. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of pi-pi Interactions. *J. Phys. Chem. A* **2006**, *110*, 10345–10354.

(36) Tekin, A.; Jansen, G. How accurate is the density functional theory combined with symmetry-adapted perturbation theory approach for CH-π and π-π interactions? A comparison to supermolecular calculations for the acetylene-benzene dimer. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1680–1687.

(37) Rybak, S.; Jeziorski, B.; Szalewicz, K. Many-body symmetry-adapted perturbation theory of intermolecular interactions. H20 and HF dimers. *J. Chem. Phys.* **1991**, *95*, 6576–6601.

(38) Pedley, J. B.; Rylance, J. *Sussex-N.P.L. Computer Analysed Thermochemical Data: Organic and Organometallic Compounds*; University of Sussex: East Sussex, U.K., 1977.

(39) Grover, J. R. Dissociation Energies of the Benzene Dimer and Dimer Cation. *J. Phys. Chem.* **1987**, *91*, 3233–3237.

(40) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. Design of Density Functionals by Combining the Method of Constraint Satisfaction with Parametrization for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.

(41) 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.* **2008**, *120*, 215–241.

(42) Zhao, Y.; Truhlar, D. G. Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2008**, *4*, 1849–1868.

(43) Sygula, A.; Fronczek, F. R.; Sygula, R.; Rabideau, P. W.; Olmstead, M. M. A Double Concave Hydrocarbon Buckycatcher. *J. Am. Chem. Soc.* **2007**, *129*, 3842–3843.

(44) Zhao, Y.; Truhlar, D. G. Size-selective supramolecular chemistry in a hydorcarbon nanoring. *J. Am. Chem. Soc.* **2007**, *129*, 8440–8442.

(45) Zhao, Y.; Truhlar, D. G. Computational Characterization and modeling of buckyball tweezers: density functional study of concave-convex pi---pi interactions. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2813–2818.

(46) Romero, C.; Fomina, L.; Fomone, S. How important is the dispersion interaction for cyclobis(paraquat-p-phenylene)-based molecular "shuttles"? A theoretical study. *Int. J. Quantum Chem.* **2005**, *102*, 200–208.

(47) Pitonak, M.; Riley, K. E.; Neogrady, P.; Hobza, P. Highly Accurate CCSD(T) and DFT-SAPT Stabilization Energies of H-Bonded and Stacked Structures of the Uracil Dimer. *ChemPhysChem* **2008**, *9*, 1636–1644.

CT100177U