Further analysis and comparative study of intermolecular interactions using dimers from the S22 database

Laszlo Fusti Molnar, Xiao He, Bing Wang, and Kenneth M. Merz, Jr.^{a)}

Department of Chemistry and the Quantum Theory Project, 2328 New Physics Building, P.O. Box 118435, University of Florida, Gainesville, Florida 32611-8435, USA

(Received 19 March 2009; accepted 18 June 2009; published online 10 August 2009)

Accurate MP2 and CCSD(T) complete basis set (CBS) interaction energy curves (14 points for each curve) have been obtained for 20 of the dimers reported in the S22 set and analytical Morse curves have been fitted that can be used in developing updated density functional theory (DFT) and force field models. The magnitude and the effect of the basis set superposition error (BSSE) were carefully investigated. We found that going up to aug-cc-pVDZ and aug-cc-pVTZ basis sets is enough to obtain accurate CBS MP2 energies when BSSE corrected values are used but aug-cc-pVTZ and aug-cc-pVQZ basis sets are needed when the BSSE uncorrected total energies are used in CBS extrapolations. MP2 interaction energies with smaller basis sets such as 6-31G* are found to have very little dispersion energy and that the true source of dispersion attributed attractive interactions is almost entirely due to BSSE. MP2 and CCSD(T) CBS interaction energies are found to be very close to one another if aromatic systems are not involved. Comparative analyses have been performed with semiempirical and *ab initio* methods utilizing the moderate in size but affordable 6-31G* basis set both of which can be readily applied to macromolecular systems. The new M06-2X and M06-L DFT functionals were found to be more accurate than all methods tested herein. Interaction energy curves using the SG1 grid showed discontinuities for several of the dimer systems but this problem disappeared when finer DFT numerical grids were used. © 2009 American Institute of Physics. [DOI: [10.1063/1.3173809](https://doi.org/10.1063/1.3173809)]

I. INTRODUCTION

Obtaining accurate intermolecular interactions between small molecule inhibitors and binding sites in proteins is one of the keys for success in structure-based drug design.¹ The so-called “docking” process involving the search for the minimum of the interaction energy with respect to geometric parameters between a protein binding site and small molecule inhibitors is one of the most frequently used techniques in small molecule screening.² The overwhelming majority of these techniques use variations in molecular mechanics (MM) based scoring together with other empirical and statistical approaches due to the need for high computational efficiency.³ On the other hand, with constantly improving computational capabilities coupled with advances in algorithm and program development over the last 10–15 yr, quantum mechanically based and combined quantum mechanical-molecular mechanical based⁴ applications are getting more and more feasible.

In the *ab initio* field, density functional theory (DFT) based methods continue to make significant gains because of their ability to provide reasonable approximations for electron correlation coupled with better computational efficiency⁵ than traditional electron correlation methods such as MP2 and CCSD(T). In practice, DFT based computations can be significantly faster than Hartree–Fock (HF) calculations for large systems when the given DFT functional does

not require the evaluation of exact exchange. The most time consuming part of these so-called local DFT calculations is usually the evaluation of the Coulomb matrix elements. Modern multipole moment based techniques,⁶ the J-matrix engine technique,⁷ and the Fourier transform Coulomb (FTC) method⁸ can significantly speed up this computational step. The latter one is especially efficient when larger basis sets and/or diffuse basis functions are used; for instance, Baker *et al.*⁹ recently reported an approximately 100-fold speedup for parallel Coulomb energy force calculations based on the FTC technique. Since *ab initio* and especially DFT based calculations are getting more and more affordable for biological applications more attention has to be paid to how much gain in accuracy can be obtained by using *ab initio* methods relative to force field or semiempirical based techniques. One of the goals of this article is to compute intermolecular interactions using MP2, HF, and various DFT methods with basis sets that are affordable for biological applications and then to compare these techniques to semiempirical and force field techniques. In DFT, among the countless number of functional choices we have, we have chosen the widely used and cited B3LYP functional¹⁰ and two promising new functionals from the M06 family.¹¹ The latter ones are designed to provide accurate estimates for dispersion interactions that are of interest in the DFT field due to their importance in biological applications.^{11–13}

Our model systems were obtained from the S22 set that has an excellent balance between polar, dispersion, and hydrogen bonded (H-bond) interactions.¹⁴ For 20 out of the 22

^{a)} Author to whom correspondence should be addressed. Tel.: (352)392-6973. FAX: (352)392-8722. Electronic mail: merz@ctp.ufl.edu.

FIG. 1. MM interaction energy curve for the benzene-H₂O dimer.

dimers, 14 geometric points along the line of the center of masses of the monomers have been constructed in order to obtain more than one interaction energy value for each dimer. Energy curves provide additional information about the current methods and they are very useful for future functional and force field development as well. For the benzene-H₂O dimer, for instance, the CCSD(T) interaction energies extrapolated to the complete basis set (CBS) limit at the equilibrium geometry are about 3.28 kcal/mol. The interaction energy of a MM based calculation using the generalized AMBER force field (GAFF)¹⁵ and bond charge correction model (BCC) charges¹⁶ is 3.77 kcal/mol. Considering only the interaction energy at this geometric point the MM based

result looks to be excellent especially when one compares it to MP2 calculations. MP2 with aug-cc-pVDZ (Ref. 17) returns a 4.67 kcal/mol interaction energy when we do not correct for the effect of basis set superposition errors (BSSEs) while the results with aug-cc-pVTZ (Ref. 17) and aug-cc-pVQZ (Ref. 17) are 4.16 and 3.78 kcal/mol, respectively. This behavior of the MM interaction energy changes, however, when we consider the shapes of the potential curves that are shown in Fig. 1, for example.

Our results for the CCSD(T) CBS interaction energy curves for 20 dimers similar to the already available values at the equilibrium geometries¹⁴ will serve as a useful tool for future development of efficient functionals and force field

TABLE I. BSSE errors and differences between CCSD(T) and MP2 computed interaction energies in kcal/mol unit.

Molecule	BSSE ^a (MP2,DZ)	BSSE ^a (MP2,TZ)	BSSE ^a [CCSD(T),DZ]	$E[\text{CCSD(T),DZ}] - E(\text{MP2,DZ})^b$	Delta BSSE ^c
Ammonia dimer	0.70(21%)	0.25(8%)	0.75(22%)	0.05	0.05
Water dimer	0.84(16%)	0.47(9%)	0.99(19%)	0.03	0.15
Methane dimer	0.53(58%)	0.14(23%)	0.56(57%)	-0.04	0.03
Ethane dimer	0.92(44%)	0.41(22%)	0.98(48%)	0.10	0.06
Ethane-ethine	1.14(45%)	0.41(21%)	1.35(52%)	0.16	0.21
Formic acid dimer	2.58(14%)	1.52(8%)	3.01(16%)	0.02	0.43
Formamide dimer	2.20(14%)	1.25(8%)	2.52(15%)	-0.08	0.31
Benzene water	1.69(36%)	0.80(19%)	2.00(42%)	0.26	0.31
Benzene ammonia	1.76(44%)	0.73(22%)	2.03(52%)	0.33	0.27
Benzene methane	1.81(55%)	0.71(29%)	2.03(64%)	0.35	0.22
Benzene dimer t-shape	3.40(52%)	1.39(28%)	3.54(62%)	0.89	0.14
Benzene dimer stack	3.87(48%)	1.59(25%)	4.04(67%)	2.23	0.17
Indole benzene t-shape	4.33(41%)	2.13(23%)	4.39(48%)	1.42	0.06
Indole benzene stack	5.72(44%)	2.56(24%)	5.93(62%)	3.48	0.21
Pyrazine dimer	3.88(39%)	1.53(19%)	4.02(54%)	2.57	0.14
2-pyridoxine 2-aminopyridine	3.42(18%)	2.11(11%)	3.41(19%)	0.64	-0.01
Phenol dimer	3.30(33%)	1.72(19%)	3.43(36%)	0.73	0.12
Uracil dimer stack	5.77(37%)	2.54(19%)	6.10(42%)	1.26	0.33
Uracil dimer hb	3.32(15%)	2.07(9%)	3.45(16%)	-0.06	0.13
Benzene hcn	2.57(37%)	1.11(18%)	3.07(45%)	0.63	0.50

^aThe BSSE values were computed at the equilibrium geometries. The BSSE contributions to the non-BSSE corrected interaction energies are tabulated in parentheses. DZ: aug-cc-pVDZ; TZ: aug-cc-pVTZ.

^bBSSE corrected values are used.

^cBSSE differences between CCSD(T) and MP2 energies computed at the equilibrium geometry using aug-cc-pVDZ basis set.

FIG. 2. CCSD(T) and MP2 interaction energies with and without BSSE corrections using different basis sets: (a) water dimer, (b) methane dimer, and (c) benzene dimer.

TABLE II. Quality analysis of the MP2 CBS extrapolations.

Molecule	$E(\text{MP2}, \text{TE})^a$ DZ,TZ	$E(\text{MP2}, \text{CE})^b$ DZ,TZ	$\Delta(\text{TE}-\text{CE})$ DZ,TZ	$E(\text{MP2}, \text{TE})^a$ TZ,QZ	$E(\text{MP2}, \text{CE})^b$ TZ,QZ	$\Delta(\text{TE}-\text{CE})$ TZ,QZ
Ammonia dimer	-3.28	-3.13	-0.15	-3.19	-3.16	-0.04
Water dimer	-5.22	-4.91	-0.31	-5.05	-4.95	-0.10
Methane dimer	-0.53	-0.49	-0.05	-0.49	-0.49	0.00
Ethane dimer	-1.86	-1.56	-0.29	-1.60	-1.58	-0.02
Ethane-ethine	-1.88	-1.65	-0.23	-1.67	-1.66	0.00
Formic acid dimer	-19.24	-18.30	-0.95	-18.89	-18.45	-0.44
Formamide dimer	-16.53	-15.70	-0.83	-16.04	-15.82	-0.22
Benzene water	-4.16	-3.53	-0.62	-3.58	-3.54	-0.04
Benzene ammonia	-3.17	-2.65	-0.53	-2.65	-2.67	0.02
Benzene methane	-2.27	-1.80	-0.47	-1.80	-1.82	0.03
Benzene hcn	-5.93	-5.13	-0.80	-5.23	-5.18	-0.05

^aTE stands for total energies in kcal/mol unit. The tabulated values are CBS extrapolated interaction energies where the extrapolations are performed by using the total energies and not the interaction energies and without any BSSE corrections.

^bCE stands for corrected energies in kcal/mol unit. The tabulated values are CBS extrapolated interaction energies where the extrapolations are performed by using the BSSE corrected interaction energies.

models that yield accurate intermolecular interactions. The total of 280 CCSD(T) CBS interaction energy values provides a large number of parameters to fit inspiring us to also provide fitted analytic Morse functions for all 20 CCSD(T) CBS interaction curves. The effect of BSSE in traditional electron correlation methods, the expected outcome of the CBS extrapolations with and without BSSE corrections, and the difference between MP2 and CCSD(T) results are also analyzed in great detail. Only a few selected cases and figures are presented herein due to space limitations. However, all the numerical results for all 20 dimers including their geometries and all figures are provided in the supplementary materials.¹⁸

II. COMPUTATIONAL DETAILS

The QCHEM 3.1 program¹⁹ was used for all DFT and MP2 calculations. The single processor version of ACESII (Ref. 20)

was used for the smaller CCSD(T) calculations, while parallel GAUSSIAN 03 (Ref. 21) was used for the larger CCSD(T) calculations. The latter ones ran on eight core processors and used the high performance I/O system provided by the High Performance Computer Center at University of Florida. The AMBER 9.0 program²² with BCC charges¹⁶ and the GAFF force field¹⁵ were used for molecular mechanical calculations and to obtain the Lennard-Jones corrections used to approximate dispersion interactions. A recent version of the DCQTP program²³ was used to obtain the semiempirical results. JAGUAR 7.5 (Ref. 24) was used for all local MP2 calculations.

The two point extrapolation scheme²⁵

$$E_{\text{MP2,CBS}} = E_{\text{MP2},x} + \text{constant} \times x^{-3} \quad (1)$$

was used to obtain the CBS extrapolated values of the MP2 correlation energies (E_{CBS}) from two separate energy calculations using a smaller and a larger basis set. The variable x

TABLE III. Accuracy study of BSSE corrected and noncorrected MP2 6-31G* interaction energies at published equilibrium geometries.

Molecule	Percentage of MP2 CBS energy, no BSSE corrections	Percentage of CCSD(T) CBS energy, no BSSE corrections	Percentage of MP2 CBS energy with BSSE corrections	Percentage of CCSD(T) CBS energy with BSSE corrections
Ammonia dimer	146	149	86	87
Water dimer	144	145	106	107
Methane dimer	30	28	-18	-17
Ethane dimer	83	89	5	5
Ethane-ethine	130	144	69	77
Formic acid dimer	108	108	78	78
Formamide dimer	115	115	82	81
Benzene water	114	123	65	70
Benzene ammonia	97	110	43	50
Benzene methane	71	88	7	9
Benzene dimer t-shape	88	116	33	43
Benzene dimer stack	63	114	-1	-2
Indole benzene t-shape	97	120	54	66
Indole benzene stack	66	114	13	22
Pyrazine dimer	84	135	36	58
2-pyridoxine 2-aminopyridine	111	116	82	85
Phenol dimer	119	131	72	80
Uracil dimer stack	99	111	52	59
Uracil dimer hb	108	107	84	83
Benzene hcn	101	116	64	73

FIG. 3. BSSE corrected CCSD(T) and MP2 interaction energy curves together with the CBS extrapolated results: (a) water dimer, (b) methane dimer, and (c) benzene dimer.

FIG. 4. Effect of the BSSE at MP2 6-31G* level of theory: (a) methane dimer, (b) benzene-methane dimer, and (c) benzene dimer.

FIG. 5. Fitted Morse and Rydberg curves to the CCSD(T) and MP2 CBS results: (a) water dimer, (b) methane dimer, and (c) benzene dimer.

FIG. 6. HF, dispersion corrected HF, and B3LYP DFT energy curves compared to the CCSD(T) CBS results: (a) ammonia dimer, (b) methane dimer, and (c) benzene-methane dimer.

TABLE IV. Fitted Morse parameters for the BSSE corrected CCSD(T) CBS interaction energy curves and zero point energy differences at equilibrium geometries.

Molecule	D_e (kcal/mol)	α (Å)	ΔRe^a (Å)	Standard deviation	ΔZPE^b (kcal/mol)
Ammonia dimer	3.10	1.38	0.05	0.08	0.93
Water dimer	4.91	1.42	0.05	0.16	1.23
Methane dimer	0.53	1.57	-0.07	0.01	0.07
Ethane dimer	1.48	1.48	0.00	0.01	0.65
Ethane ethane	1.49	1.29	0.02	0.03	0.56
Formic acid dimer	18.49	1.54	0.05	0.43	1.81
Formamide dimer	15.85	1.34	0.05	0.48	2.25
Benzene water	3.25	1.18	-0.02	0.08	0.28
Benzene ammonia	2.32	1.23	0.00	0.05	0.25
Benzene methane	1.48	1.36	0.05	0.02	0.20
Benzene dimer t-shape	2.81	1.25	0.08	0.05	0.13
Benzene dimer stack	2.84	1.33	0.15	0.03	0.36
Indole benzene t-shape	5.98	1.13	0.07	0.14	0.48
Indole benzene stack	4.99	1.36	0.12	0.09	0.15
Pyrazine dimer	4.32	1.36	0.11	0.06	0.44
2-pyridoxine 2-aminopyridine	16.72	1.34	0.03	0.38	1.08
Phenol dimer	6.99	1.23	0.05	0.32	0.89
Uracil dimer stack	9.69	1.32	0.01	0.23	0.46
Uracil dimer hb	20.12	1.31	0.04	0.58	0.91
Benzene hcn	4.50	1.13	0.03	0.11	1.82

^a ΔRe is the difference of the Morse curve minima versus the published equilibrium geometry along our chosen coordinate.

^bThe zero point energies are obtained from B3LYP DFT 6-31G** vibrational energies using harmonic approximation.

in Eq. (1) represents the largest angular momentum of the given basis set, i.e., $x=2$ for aug-cc-pVDZ, $x=3$ for aug-cc-pVTZ, and $x=4$ for aug-cc-pVQZ. The HF energies were not extrapolated and were simply taken from the results of the largest basis set calculation used in the given extrapolation

TABLE V. Overall accuracy of different methods.

Method	Average std ^a (kcal/mol)	Average normalized std ^b (kcal/mol)
MP2(aug-cc-pVDZ)	2.426	0.635
MP2(6-31G*)	1.129	0.232
M06-2X (6-31G*, SG1)	1.197	0.223
B3LYP (6-31G*, SG1)	2.420	0.622
HF+VDW6 (6-31G*)	1.954	0.277
HF (6-31G*)	2.950	0.755
M06-L [6-31G++(df,pd), 75 302]	0.430	0.110
M06-L(6-31G*, 75 302)	0.638	0.134
M06-L (6-31G*, SG1)	0.701	0.150
PDDG+LJ6	2.653	0.609
PDDG	5.351	0.819
AM1+LJ6	2.755	0.375
PM3+LJ6	1.959	0.424
AM1	6.168	1.027
PM3	5.325	0.868
MM(BCC)	1.327	0.249

^aThe standard deviations were obtained by using six points around the minimum. The average is over all 20 systems.

^bSame as in (a) but the energies are normalized by dividing them with the dissociation energies of the CCSD(T) CBS at equilibrium distance. This normalization provides the same weights for every system regardless of the strengths of the interactions while (a) is bias toward the stronger interactions.

scheme. The CBS correlation energies for CCSD(T) were obtained using

$$E_{\text{CCSD(T),CBS}} = E_{\text{MP2,CBS}} + (E_{\text{CCSD(T),aug-cc-pVDZ}} - E_{\text{MP2,aug-cc-pVDZ}}), \quad (2)$$

which is based on the observation that the difference between the MP2 and CCSD(T) correlation energies converges faster in basis set size than the correlation energies themselves.²⁶

III. RESULTS AND DISCUSSIONS

The BSSE corrected and uncorrected energy curves for both MP2 and CCSD(T) are plotted in Fig. 2 and the numerical values at the equilibrium points are tabulated in Table I. Clearly, the amount of BSSE is very significant especially for nonhydrogen bonding interactions and for both the MP2 and CCSD(T) methods using the aug-cc-pVDZ basis set. The BSSE drops when the larger aug-cc-pVTZ is used but still remains far from negligible and is larger than our target accuracy, which is at most a few tenths of a kcal/mol. Thus, extrapolation to the CBS limit is essential in order to study the computational accuracy and we discuss this topic in more detail later in this section. It is very important to recognize that at least at the aug-cc-pVDZ level, the CCSD(T) and MP2 interaction energies are almost identical and the differences between the two are less than our target accuracy for all dimers that do not involve aromatic rings. The differences are also much less than the BSSE error at the DZ level for either the MP2 or the CCSD(T) results. To investigate how general and transferable this statement is to other systems

FIG. 7. Semiempirical energy curves compared to the CCSD(T) CBS results: (a) ammonia dimer, (b) methane dimer, and (c) benzene dimer.

certainly requires further investigation but the picture is very clear for the dimers of the S22 set. Indeed, this information is very useful in practical applications for larger systems that have biological or drug design relevance. It is very rare that one can correct for BSSE in these applications and, if it is possible, the increase in computational costs makes this an impractical solution. Indeed, if aromatic rings do not play a

major role in a particular system of interest then it is much more economical to expend computational effort carrying out MP2 calculations with larger basis set where the BSSE is reduced or can be eliminated via extrapolation versus carrying out smaller basis CCSD(T) calculations.

This point is best illustrated by an example. We have completed a project to create the best possible conforma-

FIG. 8. M06-L DFT and molecular mechanical (BCC) energy curves compared to the CCSD(T) CBS results: (a) ammonia dimer, (b) methane dimer, (c) formamide dimer, (d) benzene dimer, (e) benzene-water dimer, (f) benzene-HCN dimer, and (g) indole-benzene (stack) dimer.

FIG. 8. (Continued).

FIG. 8. (Continued).

tional profiles for glycine₃ and alanine₃ using high-level electron correlation methods in order to refine biomolecular force fields. In this instance, it is impractical to estimate the BSSE for these intramolecular interactions. Based on the observations delineated above it was deemed best to perform MP2 calculations with large basis sets and then perform extrapolations for the correlation energy, thereby avoiding extremely costly CCSD(T) calculations. Unfortunately, things are not so simple for aromatic compounds or for cases involving interactions between aromatic compounds. In these instances CCSD(T) calculations appear to be unavoidable.

Our results for a more detailed quality analysis of the CBS extrapolation are tabulated in Table II. We performed this analysis only for the 11 smaller dimers which included at most one benzene ring. There were four CBS extrapolations performed, two used the total energies without any BSSE correction and two used the BSSE corrected interaction energies for the extrapolations. One extrapolation used aug-cc-pVDZ and aug-cc-pVTZ while the other used aug-cc-pVTZ and aug-cc-pVQZ energies in both cases. By definition, there should be no BSSE at the CBS limit. Thus if our extrapolation is accurate and our extrapolated interaction energies truly reflect the complete basis MP2 limit then the interaction energies obtained using the total energies without any BSSE corrections should be similar to those obtained using BSSE corrected relative energies. The importance of this analysis is that one can usually perform only the former one in practical applications. On the other hand the formula for CBS extrapolations is based on the convergence behavior of the physical correlation energies and it is not clear how it is applicable to BSSE.

The CBS extrapolated interaction energies using BSSE corrected interaction energies (third and sixth columns in Table II) are very close to each other in all cases. The largest difference (0.15 kcal/mol) is for the formic acid dimer while the remaining examples have smaller differences. Thus the CBS extrapolations for the physical interaction part of the correlation energy appear accurate even when aug-cc-pVDZ and aug-cc-pVTZ basis sets are used. On the other hand, the

extrapolated interaction energies using the total energies still contain some errors due to BSSE contaminations, which are not eliminated completely upon extrapolation. This error can be greatly reduced most of the time through the use of aug-cc-pVTZ and aug-cc-pVQZ uncorrected energies except for very strong hydrogen bonding systems where the bond distances are relatively short. Based on this analysis we recommend the use of at least triple and quadruple zeta quality basis sets in applications where the main objective is to obtain accurate CBS MP2 energies and where BSSE correction is not possible or not an attractive option. If one uses a Pople-type basis, or perhaps especially designed basis sets, we recommend that test calculations be performed in order to obtain accuracy estimates for that particular basis set. Our geometries and energy curves may offer some practical help to do such test calculations. Figure 3 shows the BSSE corrected interaction energy curves together with the MP2 and CCSD(T) CBS curves for some selected systems and similar figures for all dimers are included in the supplementary material.¹⁸

We need to highlight, however, that MP2 calculations with higher quality basis sets such as aug-cc-pVTZ are computationally very expensive for medium and large systems that have biological relevance even if fragment based or linear-scaling based approximations²⁷ are used. The computational cost of a MP2 calculation (on a modern 2.4 GHz OPTERON processor) for a small benzene dimer is about 2.5 min when a relatively small 6-31G* basis set is used but it increases to about an hour using the aug-cc-pVDZ basis set and to over 300 h using the aug-cc-pVTZ basis set (direct MP2 calculations with QCHEM 3.0). Importantly, the size of the aug-cc-pVDZ and aug-cc-pVTZ basis sets is still not enough to eliminate all of the BSSE error by extrapolation.

It is widely appreciated that MP2 is the most affordable choice among traditional electron correlation methods that are able to capture dispersion interactions, but it is sometimes forgotten that this important feature strongly depends on the quality of the basis set used in the actual calculations. Table III and Fig. 4 show our results for interaction energies

TABLE VI. Comparison of the interaction energies and BSSE for MP2, local-MP2 (LMP2), and DFT-M06L methods at the published equilibrium geometries and by using 6-31G* basis set (values are in kcal/mol unit).

Molecule	$E(\text{MP2})$	BSSE(MP2)	$E(\text{LMP2})$	BSSE(LMP2)	$E(\text{M06L})$	BSSE(M06L)	$E[\text{CCSD(T), CBS}]$
Ammonia dimer	-4.63	1.91	-4.39	1.68	-5.45	1.11	-3.10
Water dimer	-7.15	1.90	-6.92	1.62	-8.31	1.21	-4.92
Methane dimer	-0.15	0.24	-0.04	0.17	-1.10	0.10	-0.53
Ethane dimer	-1.32	1.24	-0.87	0.93	-2.35	0.47	-1.49
Ethane-ethine	-2.16	1.01	-1.87	0.70	-2.24	0.39	-1.50
Formic acid dimer	-19.92	5.57	-16.34	3.83	-22.99	3.05	-18.43
Formamide dimer	-18.24	5.30	-16.23	3.68	-20.19	2.90	-15.90
Benzene water	-4.03	1.72	-3.17	1.03	-5.07	0.79	-3.28
Benzene ammonia	-2.58	1.42	-1.86	0.72	-3.41	0.51	-2.34
Benzene methane	-1.30	1.17	-0.58	0.54	-2.33	0.37	-1.48
Benzene dimer t-shape	-3.27	2.05	NA ^a	NA ^a	-4.07	0.58	-2.81
Benzene dimer stack	-3.11	3.15	-1.26	2.03	-4.30	1.27	-2.72
Indole benzene t-shape	-7.20	3.24	-4.69	1.40	-7.31	0.98	-6.00
Indole benzene stack	-5.54	4.47	-3.06	2.95	-6.30	1.64	-4.87
Pyrazine dimer	-5.72	3.25	-3.80	2.04	-6.55	1.15	-4.24
2-pyridoxine 2-aminopyridine	-19.41	5.19	-12.34	3.12	-20.80	2.19	-16.76
Phenol dimer	-9.28	3.62	-7.37	2.31	-10.28	1.71	-7.09
Uracil dimer stack	-10.86	5.15	-7.03	2.75	-13.75	2.14	-9.76
Uracil dimer hb	-21.75	4.87	-15.78	2.74	-24.54	2.26	-20.28
Benzene hcn	-5.25	1.94	-3.94	0.68	-6.04	0.53	-4.54

^aCalculation finished unsuccessfully with error.

using the 6-31G* basis set with MP2 calculations. This level of theory is very affordable for application to much larger systems especially if fragment based linear-scaling techniques combined with localization and/or density fitting approximations are used. It is clear from the results of the methane, ethane, and benzene-methane dimers that MP2 with the 6-31G* basis captures very little or no contribution from true dispersion interactions. The negative numbers in Table III show that there is no attractive dispersion interaction between the monomers at published equilibrium geometries. The BSSE corrected interaction curve has a very shallow minimum at longer separations of the monomers when compared to our best estimate of the true potential energy curves. Those interested in MP2 computations for macromolecules should pay close attention to the choice of basis set to ensure that dispersion is correctly modeled and is not replaced by nonphysical BSSE. Unfortunately, there is no reason to believe that the BSSE even approximates true dispersion energy curves.

Given that we have 14 CCSD(T) CBS interaction energies along our approximate interaction coordinate we can fit the analytical Morse curves for all 20 dimers. The parameters for the Morse curves together with the standard deviation of the fitting are listed in Table IV and the fitted curves are shown in Fig. 5 for some selected dimers. Having analytical curves can be useful in practice since one can create any number of points along the interaction coordinate which can aid in parameter fitting of simple potentials, for example.

The sixth column in Table IV shows the zero point energy differences obtained using B3LYP harmonic frequencies with the 6-31G** basis set. Zero point energy differences obtained using MP2 are published as supplementary material.¹⁸ This energy is coming from the nonclassical quantum nature of the vibrations and it can be an important

contributor to the interaction enthalpy. For instance, if the aim is to develop new and more accurate scoring functions for enzyme-ligand docking applications, then the high computational efficiency requirements make dealing with the vibrational frequencies directly a less satisfactory approach. On the other hand, one can obtain new interaction energy curves through a consideration of our fitted CCSD(T) CBS Morse curves and deduct zero point energies from Morse parameters of the dissociation energies. The resulting effective Morse curve will implicitly include zero point energy corrections and new scoring function parameters for docking applications can be fitted against these modified curves. This is an approximate solution since the vibrations are coupled and strongly depend on atomic masses as well as the potential curves and certainly need further study, but perhaps it is a good starting point for developing a useful approximate approach instead of completely neglecting zero point energy contributions to the interaction enthalpy.

Table V presents the comparative results of several different and widely used methods demonstrating their relative abilities to describe intermolecular interactions. Six geometric points around the published minimum (the minimum point, two at shorter and three at longer distances) were used to evaluate the standard deviations versus CCSD(T) CBS values. The methods are ordered based on their approximate computational expense for macromolecules from the top to the bottom. The most expensive among those methods is MP2 and the least expensive *ab initio* method is DFT calculations using the M06-L functional that only uses functional exchange. The overall accuracy of MP2 interaction energy curves with the 6-31G* basis set is very disappointing. It is just a bit better than very fast molecular mechanical calculations. In addition, most of the dispersion interaction is coming from nonphysical BSSE as discussed previously. The

overall performance of MP2 with a larger aug-cc-pVDZ basis set is poorer because most of the interactions are overestimated due to the large BSSE still present with this basis set. HF and B3LYP perform decently for H bonds but, as is widely known, dispersion interactions are not modeled at these levels of theory. Adding empirical Lennard-Jones attractive terms to HF without changing the BCC force field parameters improves the accuracy for the dispersion interactions but unfortunately negatively affects the accuracy for hydrogen bonded systems. The overall performance increases due to this correction but it is still a bit less accurate than the results from the MM calculations. Figure 6 shows these curves for some selected dimers. The accuracy of semi-empirical methods, not unexpectedly, is poor for all interaction types, with Fig. 7 presenting some representative examples.

The accuracy of DFT calculations using the M06 family of functionals outperforms all methods in our comparison and the M06-L functional is particularly attractive since it can be implemented very efficiently due to its local nature. Among the methods that were included in this comparison, this is the only quantum mechanical based method that clearly outperforms the molecular mechanical results for intermolecular interactions and for which the extra computational expense can be justified in biological or drug design applications. Similarly accurate results have been published recently using other dispersion corrected DFT functionals.¹³ Figure 8 shows some representative examples comparing the DFT M06-L curves with MM and exact curves. One should note that unlike in the MP2 case where the BSSE corrections lead to less accurate results, the DFT M06-L method with the 6-31G* basis set overestimates the interaction energies at equilibrium distances and if the BSSE error is corrected then the results improve for all our dimers. Table VI shows our comparison of the dissociation energies and the BSSE for MP2, local-MP2, and DFT M06-L methods using the 6-31G* basis set. Local MP2 and its derivative techniques are another alternative to canonical MP2 calculations since its BSSE is less and it converges to the CBS limit faster than canonical MP2.²⁸ Through the use of a relatively small and computationally efficient basis set such as 6-31G*, however, the BSSE in local-MP2 is still significant and the physical interactions after BSSE correction (sum of the fourth and fifth columns in Table VI) are similarly inaccurate and even worse than in the canonical MP2 cases. The local-MP2 based approach may be the method of choice when one can afford a significantly larger basis set where the BSSE free interaction energies are adequately described while the interactions are overestimated due to BSSE in the corresponding canonical MP2 calculations. For the DFT calculations one important technical aspect should be noted. The efficient and popular SG1 numerical grid²⁹ caused discontinuities on the energy curves for several systems. This cannot be recognized from individual energy calculations since they all converged and terminated normally. These discontinuities diminish through the use of larger numerical grids where the realistic performance of the M06-L functional is shown.

Based on a consideration of both the computational expense and accuracy of the MP2 method, one must conclude

that this method is not the method of choice for intermolecular interactions even when fragment or linear-scaling based approximations increase the computational performance. The M06 functional family or other explicit dispersion corrected local DFT methods seem to be much better choices in this regard. On the other hand, a lot of credit should be given to developers of electron correlation methods over the last few decades since obtaining the “right” results using large basis MP2 and coupled-cluster methods is a key and a vital ingredient for the future development of DFT and force field methodologies.

ACKNOWLEDGMENTS

We gratefully acknowledge the computing support from the University of Florida High Performance Computing Center. We thank the NIH (Grant No. GM066859) for the financial support of this research.

- ¹M. B. Peters, K. Raha, and K. M. Merz, *Curr. Opin. Drug Discovery Dev.* **9**, 370 (2006); K. Raha, M. B. Peters, B. Wang, N. Yu, A. M. Wollacott, L. M. Westerhoff, and K. M. Merz, *Drug Discovery Today* **12**, 725 (2007).
- ²A. R. Leach, B. K. Shoichet, and C. E. Peishoff, *J. Med. Chem.* **49**, 5851 (2006); A. N. Jain, *Curr. Protein Pept. Sci.* **7**, 407 (2006); R. T. Kroemer, *ibid.* **8**, 312 (2007).
- ³R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, and D. T. Mainz, *J. Med. Chem.* **49**, 6177 (2006); E. Perola, W. P. Walters, and P. S. Charifson, *Proteins: Struct., Funct., Bioinf.* **56**, 235 (2004); M. Kontoyianni, L. M. McClellan, and G. S. Sokol, *J. Med. Chem.* **47**, 558 (2004); E. Kellenberger, J. Rodrigo, P. Muller, and D. Rognan, *Proteins: Struct., Funct., Bioinf.* **57**, 225 (2004); T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard, and J. L. Banks, *J. Med. Chem.* **47**, 1750 (2004); R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin, *ibid.* **47**, 1739 (2004); R. X. Wang, Y. P. Lu, and S. M. Wang, *ibid.* **46**, 2287 (2003); Z. Zsoldos, D. Reid, A. Simon, S. B. Sadjad, and A. P. Johnson, *J. Mol. Graphics Modell.* **26**, 198 (2007); M. Zeragrat, H. Steuber, C. Koch, C. La Motta, S. Sartini, C. A. Sotriffer, and G. Klebe, *Angew. Chem., Int. Ed.* **46**, 3575 (2007); S. Tietze and J. Apostolakis, *J. Chem. Inf. Model.* **47**, 1657 (2007); M. R. Lee and Y. X. Sun, *J. Chem. Theory Comput.* **3**, 1106 (2007); P. Khodade, R. Prabhu, N. Chandra, S. Raha, and R. Govindarajan, *J. Appl. Crystallogr.* **40**, 598 (2007); A. Gupta, A. Gandhimathi, P. Sharma, and B. Jayaram, *Protein Pept. Lett.* **14**, 632 (2007); A. Grosdidier, V. Zoete, and O. Michielin, *Proteins: Struct., Funct., Bioinf.* **67**, 1010 (2007); H. M. Chen, B. F. Liu, H. L. Huang, S. F. Hwang, and S. Y. Ho, *J. Comput. Chem.* **28**, 612 (2007); R. Thomsen and M. H. Christensen, *J. Med. Chem.* **49**, 3315 (2006); S. Y. Huang and X. Q. Zou, *J. Comput. Chem.* **27**, 1876 (2006).
- ⁴U. C. Singh and P. A. Kollman, *J. Comput. Chem.* **7**, 718 (1986); M. J. Field, P. A. Bash, and M. Karplus, *ibid.* **11**, 700 (1990); M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, *J. Phys. Chem.* **100**, 19357 (1996); T. Z. Mordasini and W. Thiel, *Chimia* **52**, 288 (1998); I. Antes and W. Thiel, *J. Phys. Chem. A* **103**, 9290 (1999); Y. K. Zhang, T. S. Lee, and W. T. Yang, *J. Chem. Phys.* **110**, 46 (1999); D. Das, K. P. Eurenium, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodoseck, and B. R. Brooks, *ibid.* **117**, 10534 (2002); G. Csanyi, T. Albaret, M. C. Payne, and A. De Vita, *Phys. Rev. Lett.* **93**, (2004); I. Berente and G. Naray-Szabo, *Int. J. Quantum Chem.* **104**, 328 (2005); T. Hori, H. Takahashi, and T. Nitta, *J. Theor. Comput. Chem.* **4**, 867 (2005); H. Lin and D. G. Truhlar, *J. Phys. Chem. A* **109**, 3991 (2005); Y. Nakamura, N. Takahashi, M. Okamoto, T. Uda, and T. Ohno, *J. Comput. Phys.* **225**, 1985 (2007); Y. Zhang and H. Lin, *J. Chem. Theory Comput.* **4**, 414 (2008).
- ⁵W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory* (Wiley, New York, 2000); J. Baker and P. Pulay, *J. Chem. Phys.* **117**, 1441 (2002); *J. Comput. Chem.* **24**, 1184 (2003); V. N. Staroverov, G. E. Scuseria, J. M. Tao, and J. P. Perdew, *J. Chem. Phys.* **119**, 12129

- (2003); R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, England, 2004); Y. Zhao and D. G. Truhlar, *J. Chem. Phys.* **124**, (2006); K. E. Riley and K. M. Merz, *J. Phys. Chem. A* **111**, 6044 (2007); K. E. Riley, B. T. Op't Holt, and K. M. Merz, *J. Chem. Theory Comput.* **3**, 407 (2007).
- ⁶H. G. Petersen, D. Soelvason, J. W. Perram, and E. R. Smith, *J. Chem. Phys.* **101**, 8870 (1994); C. A. White and M. Headgordon, *ibid.* **101**, 6593 (1994); C. A. White, B. G. Johnson, P. M. W. Gill, and M. Headgordon, *Chem. Phys. Lett.* **230**, 8 (1994); J. C. Burant, M. C. Strain, G. E. Scuseria, and M. J. Frisch, *ibid.* **258**, 45 (1996); J. C. Burant, M. C. Strain, G. E. Scuseria, and M. J. Frisch, *ibid.* **248**, 43 (1996); M. Challacombe, E. Schwegler, and J. Almlof, *J. Chem. Phys.* **104**, 4685 (1996); M. C. Strain, G. E. Scuseria, and M. J. Frisch, *Science* **271**, 51 (1996); C. A. White, B. G. Johnson, P. M. W. Gill, and M. Headgordon, *Chem. Phys. Lett.* **253**, 268 (1996); K. E. Schmidt and M. A. Lee, *J. Stat. Phys.* **87**, 955 (1997); G. E. Scuseria, *J. Phys. Chem. A* **103**, 4782 (1999); L. F. Greengard and J. F. Huang, *J. Comput. Phys.* **180**, 642 (2002); E. Rudberg and P. Salek, *J. Chem. Phys.* **125**, 084106 (2006).
- ⁷Y. Shao and M. Head-Gordon, *Chem. Phys. Lett.* **323**, 425 (2000); C. A. White and M. HeadGordon, *J. Chem. Phys.* **104**, 2620 (1996).
- ⁸L. Fusti-Molnar and P. Pulay, *J. Chem. Phys.* **117**, 7827 (2002); **116**, 7795 (2002); *J. Mol. Struct.: THEOCHEM* **666**, 25 (2003); L. Fusti-Molnar, *J. Chem. Phys.* **119**, 074108 (2003); L. Fusti-Molnar and J. Kong, *ibid.* **122**, (2005).
- ⁹J. Baker, K. Wolinski, and P. Pulay, *J. Comput. Chem.* **28**, 2581 (2007).
- ¹⁰A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988); C. T. Lee, W. T. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- ¹¹Y. Zhao and D. G. Truhlar, *J. Chem. Phys.* **125**, (2006); *Theor. Chem. Acc.* **120**, 215 (2008).
- ¹²A. J. Du and S. C. Smith, *Nanotechnology* **16**, 2118 (2005); A. D. Becke and E. R. Johnson, *J. Chem. Phys.* **124**, 174104 (2006); R. W. Williams and D. Malhotra, *Chem. Phys.* **327**, 54 (2006); Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A* **110**, 5121 (2006); J. A. Alonso and A. Mananes, *Theor. Chem. Acc.* **117**, 467 (2007); Q. S. Du, P. J. Liu, and J. Deng, *J. Chem. Theory Comput.* **3**, 1665 (2007); S. Grimme, J. Antony, T. Schwabe, and C. Muck-Lichtenfeld, *Org. Biomol. Chem.* **5**, 741 (2007); P. Jurecka, J. Cerny, P. Hobza, and D. R. Salahub, *J. Comput. Chem.* **28**, 555 (2007); C. A. Morgado, J. P. McNamara, I. H. Hillier, and N. A. Burton, *J. Chem. Theory Comput.* **3**, 1656 (2007); T. Schwabe and S. Grimme, *Phys. Chem. Chem. Phys.* **9**, 3397 (2007); Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.* **3**, 289 (2007); V. R. Cooper, T. Thonhauser, and D. C. Langreth, *J. Chem. Phys.* **128**, 204102 (2008); P. L. Silvestrelli, *Phys. Rev. Lett.* **100**, 053002 (2008); Y. Y. Sun, Y. H. Kim, K. Lee, and S. B. Zhang, *J. Chem. Phys.* **129**, 154102 (2008).
- ¹³J. Antony and S. Grimme, *Phys. Chem. Chem. Phys.* **8**, 5287 (2006).
- ¹⁴P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).
- ¹⁵J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, *J. Comput. Chem.* **25**, 1157 (2004).
- ¹⁶A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, *J. Comput. Chem.* **21**, 132 (2000).
- ¹⁷R. A. Kendall, T. H. Dunning, and R. J. Harrison, *J. Chem. Phys.* **96**, 6796 (1992); K. A. Peterson, R. A. Kendall, and T. H. Dunning, *ibid.* **99**, 1930 (1993).
- ¹⁸See EPAPS supplementary material at <http://dx.doi.org/10.1063/1.3173809> for attachment files for geometries and interaction energy curves of the dimers.
- ¹⁹Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio, R. C. Lochan, T. Wang, G. J. O. Beran, N. A. Besley, J. M. Herbert, C. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Dachsel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, C. P. Hsu, G. Kedziora, R. Z. Khalliulin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer, J. Kong, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon, *Phys. Chem. Chem. Phys.* **8**, 3172 (2006).
- ²⁰J. F. Stanton, J. Gauss, J. D. Watts, M. Nooijen, N. Oliphant, S. A. Perera, P. G. Szalay, W. J. Lauderdale, S. A. Kucharski, S. R. Gwaltney, S. Beck, A. Balková, D. E. Bernholdt, K. K. Baeck, P. Rozyczko, H. Sekino, C. Hober, and R. J. Bartlett, *ACES II* program package (Quantum Theory Project, University of Florida, Gainesville, FL).
- ²¹M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN03, Revision D.01, Gaussian, Inc., Wallingford, CT, 2004.
- ²²D. A. Case, T. A. Darden, T. E. Cheatham III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, B. Wang, S. Hayik, A. Roitberg, G. Seabra, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, J. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, and P. A. Kollman, *AMBER*, 2006.
- ²³B. Wang, K. Raha, N. Liao, M. B. Peters, H. Kim, L. M. Westerhoff, A. M. Wollacott, A. van de Vaart, V. Gogonea, D. Suarez, S. L. Dixon, J. J. Vincent, E. N. Brothers, and K. M. Merz, *DCQTP*, 2005.
- ²⁴L. Schrodinger, *Jaguar 7.5* (Schrodinger, LLC, New York, 2008).
- ²⁵A. Halkier, T. Helgaker, P. Jorgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, *Chem. Phys. Lett.* **286**, 243 (1998).
- ²⁶P. Hobza and J. Sponer, *J. Am. Chem. Soc.* **124**, 11802 (2002); P. Jurecka and P. Hobza, *Chem. Phys. Lett.* **365**, 89 (2002); I. Daabkowska, P. Jurecka, and P. Hobza, *J. Chem. Phys.* **122**, 204322 (2005).
- ²⁷D. G. Fedorov and K. Kitaura, *J. Phys. Chem. A* **111**, 6904 (2007); T. E. Exner and P. G. Mezey, *ibid.* **106**, 11791 (2002); X. He and J. Z. H. Zhang, *J. Chem. Phys.* **122**, 031103 (2005).
- ²⁸L. F. Holroyd and T. van Mourik, *Chem. Phys. Lett.* **442**, 42 (2007); T. B. Adler, H. J. Werner, and F. R. Manby, *J. Chem. Phys.* **130**, (2009); J. Kaminsky, R. A. Mata, H. J. Werner, and F. Jensen, *Mol. Phys.* **106**, 1899 (2008).
- ²⁹P. M. W. Gill, B. G. Johnson, and J. A. Pople, *Chem. Phys. Lett.* **209**, 506 (1993).