Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · SEPTEMBER 2007

Impact Factor: 2.69 · DOI: 10.1021/jp073358r · Source: PubMed

2 AUTHORS:

Kevin E Riley
Xavier University of Louisiana
41 PUBLICATIONS **3,320** CITATIONS
SEE PROFILE

Pavel Hobza
Academy of Sciences of the Czech Republic
318 PUBLICATIONS **18,109** CITATIONS
SEE PROFILE

Assessment of the MP2 Method, along with Several Basis Sets, for the Computation of Interaction Energies of Biologically Relevant Hydrogen Bonded and Dispersion Bound Complexes

Kevin E. Riley and Pavel Hobza*

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Flemingovonám. 2, 166 10 Prague 6, Czech Republic

Received: May 2, 2007; In Final Form: June 21, 2007

In the past several years the MP2 method has been used extensively in studies of noncovalent interactions within biological systems such as proteins, DNA/RNA, and protein−ligand complexes. In this work we assess the performance that can be expected of this method, when paired with several different medium and extended basis sets, for the accurate computation of binding energies of hydrogen bonded and dispersion bound biologically derived complexes. It is found that, overall, the MP2/cc-pVTZ method produces the best, most well balanced, description of noncovalent interactions. Another interesting observation made in this study is that generally the MP2 technique, when paired with any basis set, does not yield reliable results for cyclic hydrogen bonds such as those found in nucleic acid base pairs.

Introduction

Noncovalent interactions play a pivotal role in determining the structure, stability, and dynamics of biological systems such as proteins and DNA, and thus, the accurate theoretical description of these interactions is of critical importance. Relatively inexpensive, single determinant, methods such as Hartree−Fock (HF) and density functional theory (DFT) provide reliable results for intermolecular interactions that are largely of an electrostatic nature, such as hydrogen bonding,[1−3] but generally fail to yield accurate interaction energies (and geometries) for dispersion bound systems.[1,4−6] The failure of these methods to describe the interactions of dispersion bound systems is attributable to their limited ability (or inability) to describe electron correlation effects. On the other end of the computational spectrum, techniques such as the coupled cluster (CC) and configuration interaction (CI) methods, when used with large basis sets, produce accurate results for both interactions that are principally electrostatic and for those that are chiefly determined by dispersion forces. Unfortunately, because they scale very unfavorably with the number of basis sets, it is only feasible to use methods such as CC and CI for relatively small systems. The second-order Møller−Plesset perturbation theory (MP2), which describes electron correlation in a limited way and is much less computationally expensive than the CC and CI methods, can be described as being an intermediate level of theory. The success of the MP2 method in describing intermolecular interactions is largely attributable to a kind of "compensation of errors" involving the size of the basis set employed and the lack of higher order correlation energy contributions offered by this method. The purpose of this study is to characterize the accuracy that can be expected of the MP2 method, paired with several commonly used basis sets, for the computation of interaction energies of biologically relevant complexes.

Noncovalent interaction forces can be divided, for the most part, into two types, those of an electrostatic nature (such as in the case of hydrogen bonding) and those that arise as a result of the simultaneous electron correlation of separated subsystems (such as in the case of dispersion interactions). Dispersion forces are generally weaker than the forces associated with hydrogen bonding; nonetheless, because they can be very abundant, dispersion interactions play a large role in the stabilization of large biomolecules, such as proteins and DNA. To properly describe dispersion forces, which arise from the electrostatic interaction of fluctuating charge distributions, it is necessary that a system's dynamic correlation be treated accurately.

The MP2 method is used extensively in computational chemistry and has become a very popular tool for the treatment of intermolecular and intramolecular noncovalent interactions in biological systems. MP2 has been employed in studies concerning several different types of biomolecular noncovalent interactions associated with the stability of, for example, proteins,[5,7,8] DNA/RNA,[9,10] protein−ligand complexes,[11−13] protein−DNA/RNA complexes,[14−17] and protein−carbohydrate complexes.[18] Looking through the references listed above, one can see that the MP2 method is used with a wide variety of basis sets ranging in size from the, very small, STO-3G basis to the, much larger, aug-cc-pVDZ and cc-pVTZ bases; moreover, MP2 results are sometimes extrapolated to the complete basis set limit (CBS) to compute interaction energies.

The MP2/CBS (or MP2/extended basis set) stabilization energies are overestimated mainly due to the overestimation of dispersion energy. For complexes where stabilization energy originates in electrostatic interactions (H-bonded complexes), the MP2/CBS stabilization energy is relatively accurate, but for dispersion bound complexes (stacked structures), the MP2/CBS stabilization energy is too large. This overestimation is reduced when higher correlation energy contributions are considered, e.g., by performing the CCSD(T) calculations. The CCSD(T) correction term (defined as a difference between CCSD(T) and MP2 interaction energies) was determined for more than 100 stacked structures of DNA base pairs and amino acid pairs. This term was systematically repulsive by between 2 and 5 kcal/mol.[19] Surprisingly, accurate stabilization energies and complex geometries are sometimes determined with small or medium

10.1021/jp073358r CCC: $37.00
Published on Web 07/25/2007

basis sets. This is due to compensation of errors: the size of the basis set leads to overestimation of stabilization energy and neglect of higher correlation energy contributions leading to underestimation of stabilization energy. Evidently, the MP2 procedure combined with the extended basis set does not yield accurate stabilization energies and complex geometries whereas smaller basis sets can provide better results. Several modified (canonical and local) MP2 methods based on the scaling of correlation energy due to parallel and antiparallel electron pairs have been developed in the past several years.[20−23] These methods often yield stabilization energies and geometries superior to those of standard MP2. In terms of the treatment of biological systems, the "spin-component scaled for nucleobases" (SCSN) model of Hill and Platts,[23] which is based on spin-component scaled density fitting local MP2 (DF-SCS-LMP2) is the most promising among the modified MP2 methods.

We would also like to note that the introduction of the approximate resolution of the identity MP2 method (RI-MP2),[24,25] which lowers the computational cost of MP2 calculations by about an order of magnitude with only marginal loss of accuracy,[26] increases the efficiency of the MP2 method for calculations of noncovalent interactions in biological systems. To our knowledge a systematic study assessing the quality of the MP2 interaction energies, computed using several different basis sets, has never been carried out. Given that MP2 will, presumably, continue to be used extensively in studies involving noncovalent interactions, we feel that it is important that its strengths and weaknesses for these types of computations be well characterized. A very important advantage of the MP2 procedure is the fact that it consistently describes all types of correlation energy, not only the intermolecular correlation term, which includes both the R^6 dispersion energy and all higher dispersion terms, but also the intramolecular correlation energy, which is mainly responsible for reducing the electrostatic term when passing from Hartree−Fock to a correlated level. The MP2 procedure can be said to be a truly ab initio method that requires no empirical parameters.

Recently the treatment of noncovalent interactions, especially those associated with dispersion, has been the subject of many theoretical investigations. As stated above, single determinant methods, such as HF and DFT, generally fail to produce reliable results for dispersion bound systems, it should be noted, however, that some DFT functionals do yield interaction energies that are qualitatively, or even semiquantitatively, accurate.[1,3,4,27] Among the functionals that have been noted to produce relatively good results for different types of noncovalent interactions, those recently developed by Truhlar and co-workers (M05,[28] M05-2X,[28] M06,[29] M06-2X[29]), seem to show the most promise for the treatment of dispersion interactions in biological systems. In recent years several groups have been able to treat dispersion interactions by augmenting single determinant methods with empirical terms meant to make up for the deficiencies of these methods in treating electronic correlation properly.[30−40] One of the most recently developed dispersion augmented density functional theory techniques (DFT-D), which was parametrized specifically for intermolecular interactions of biological interest, promises to be a useful tool for studying proteins, DNA, and protein−ligand complexes.[33,38]

Methods

In this work we compare the interactions energies obtained with the MP2 method along with ten different basis sets, namely 6-31G*, 6-31G*(0.25), 6-31+G*, TZVP, 6-311+G*, cc-pVDZ, cc-pVDZ+, aug-cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ, to those produced using higher level ab initio methods. Of the ten basis sets listed above, eight are very commonly used, the other two, 6-31G*(0.25) and cc-pVDZ+, are modified versions of the 6-31G* and cc-pVDZ bases, respectively. 6-31G*(0.25) is the same as 6-31G* except for the fact that the exponential coefficients of the first row atom polarization functions are changed from 0.80 to 0.25. cc-pVDZ+ is the usual cc-pVDZ with the addition of the s and p diffuse functions from aug-cc-pVDZ placed on the heavy (non-hydrogen) atoms.

The test set employed here is derived from the S22[19] set of Hobza and co-workers and contains noncovalent complexes of three types, those whose principle mode of interaction is through hydrogen bonding (hydrogen bonding), those which interact mainly through dispersion interactions (dispersion), and those whose interactions can be described as being a mixture of hydrogen bonding and dispersion (mixed). We have noted that cyclic hydrogen bonds, such as those found in the formic acid dimer, seem to behave quite differently than the more conventional, single, hydrogen bonds. As the S22 set contains several (five) examples of cyclic hydrogen bonds and only two examples of single hydrogen bonds, we have augmented this test set with four additional complexes whose interaction can be said to be of the single hydrogen bond type. The four added complexes are the methanol dimer, the methanol formaldehyde dimer, and two methyl amide dimers whose heavy atom structures are derived from the crystal structure of the protein rubredoxin. Concerning the rubredoxin derived hydrogen-bonding structures, the first of these, termed the methyl amide dimer (α), was obtained from an α helix and the second, termed the methyl amide dimer (β) comes from a β sheet within the protein. It should be noted that cyclic hydrogen bonds are very important to the structure of DNA/RNA but are not commonly found in proteins. The augmented S22 test set employed in this study will be referred to as the S26-07 set.

The geometries of the systems within the S22 database were determined using either MP2 or CCSD(T) with various basis sets. The reference interactions energies of these complexes were then determined by extrapolating MP2 results to the complete basis set limit and then adding a correction term corresponding to the difference between the CCSD(T) and MP2 interaction energies at a given basis set (please see ref 19 for further details). The optimized geometries for two of the four supplementary complexes included in this work, the methanol and methanol−formaldehyde dimers, were obtained at the MP2/cc-pVTZ level of theory on the counterpoise-corrected geometry hypersurface. The heavy atom (non-hydrogen) geometries of the two methyl amide dimers came from the crystal structure of the protein rubredoxin (PDB code 1RB9), and the hydrogen atom geometries were optimized at the DFT TPSS/TZVP level of theory. The reference interaction energies were then determined by extrapolating MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ energies to the complete basis set limit and adding CCSD(T) correction terms (determined with the aug-cc-pVDZ basis) to the MP2 derived interaction energies. We are aware that it is a bit suspect to use interaction energies that are determined, in part, using the MP2 method as reference data for our MP2 study; nonetheless, for most of the systems contained within our test set, these are the highest level results obtained to date. We would also note that the inclusion of the CCSD(T) correction terms (which can be quite large: up to ∼2−5 kcal/mol) seems to increase the accuracy of the reference interaction energies substantially.

In this work all MP2 calculations are made with either the geometries determined in ref 19 (for the S22 database) or those

TABLE 1: Interaction Energy Errors for the MP2 Method When Paired with the "Small" Basis Sets 6-31G*, 6-31G*(0.25), 6-31+G*, TZVP, and 6-311+G* [a]

		6-31G*		6-31G*(0.25)		6-31+G*		TZVP		6-311+G*	
	high level	CP	no-CP	CP	no-CP	CP	no-CP	CP	no-CP	CP	no-CP
Single Hydrogen Bonds											
$(NH_3)_2$ (C_{2h})	−3.17	−0.45	1.48	−0.90	1.95	−0.21	0.81	−0.40	0.38	−0.14	1.05
$(H_2O)_2$ (C_s)	−5.02	0.21	2.09	−1.06	2.59	−0.21	1.80	−0.35	0.70	0.19	2.55
methanol dimmer (C_1)	−5.70	−0.36	2.09	−1.08	3.12	−0.46	1.90	−0.47	1.12	−0.33	2.23
methanol−formaldehyde (C_1)	−5.31	−1.43	1.94	−1.41	2.53	−1.16	0.57	−1.48	0.03	−1.48	0.17
methyl amide dimer (α)	−6.69	−1.50	1.03	−1.35	2.73	−1.43	0.65	−1.47	0.01	−1.55	0.48
methyl amide dimer (β)	−7.65	−1.81	1.22	−1.58	3.94	−1.54	0.39	−1.51	−0.18	−1.56	0.18
Cyclic Hydrogen Bonds											
formic acid dimer (C_{2h})	−18.61	−4.28	1.20	−4.83	3.48	−5.28	−1.59	−4.28	−1.53	−6.07	−2.90
formamide dimer (C_{2h})	−15.96	−3.06	2.19	−3.33	3.71	−3.55	−0.91	−3.53	−1.31	−4.07	−1.72
uracil dimer (C_{2h})	−20.65	−3.70	1.20	−3.94	4.32	−3.74	−0.08	−3.66	−1.18	−4.28	−0.99
2-pyridoxine−2-aminopyridine (C_1)	−16.71	−2.45	2.77	−2.53	5.90	−2.93	1.02	−2.33	0.47	−2.94	0.49
adenine−thymine WC (C_1)	−16.37	−3.29	2.60	−3.08	6.59	−3.52	0.27	−2.92	−0.37	−3.51	−0.42
Dispersion Bound											
$(CH_4)_2$ (D_{3d})	−0.53	−0.62	−0.38	−0.40	−0.09	−0.59	−0.40	−0.44	−0.22	−0.53	−0.27
$(C_2H_4)_2$ (D_{2d})	−1.51	−1.43	−0.17	−0.83	0.62	−1.36	−0.40	−1.03	0.04	−1.13	0.00
benzene−CH_4 (C_3)	−1.50	−1.38	−0.21	−0.82	0.47	−1.16	0.38	−0.70	0.63	−0.85	0.81
benzene dimer (C_{2h})	−2.73	−2.74	0.44	−0.23	4.19	−1.07	3.16	−0.25	3.60	−0.17	4.26
pyrazine dimer (C_s)	−4.42	−1.88	1.44	0.47	5.64	−0.60	3.48	−0.39	3.52	−0.06	3.89
uracil dimer (C_2)	−10.12	−3.59	1.75	−1.75	7.06	−2.20	4.50	−1.66	4.90	−1.36	5.61
indole−benzene (C_1)	−5.22	−4.09	0.43	−0.46	6.37	−1.56	4.49	−0.56	5.26	−0.39	5.76
adenine−thymine stack (C_1)	−12.23	−4.31	2.97	−1.22	11.17	−1.90	7.06	−0.97	8.04	−0.58	8.44
Mixed Interactions											
ethene−ethyne (C_{2v})	−1.53	−0.38	0.63	−0.19	0.85	−0.52	0.19	−0.52	0.03	−0.43	0.28
benzene−H_2O (C_s)	−3.28	−0.96	0.76	−0.64	1.31	−1.03	1.02	−0.92	0.84	−0.85	1.33
benzene−NH_3 (C_s)	−2.35	−1.21	0.23	−0.70	0.90	−1.17	0.71	−0.87	0.71	−0.91	1.06
benzene−HCN (C_s)	−4.46	−1.09	0.90	−0.15	1.73	−1.15	1.31	−0.97	1.35	−0.73	1.56
benzene dimer (C_{2v})	−2.74	−1.49	0.58	−0.41	1.93	−0.98	2.02	−0.56	2.40	−0.53	2.57
indole−benzene T-shape (C_1)	−5.73	−1.66	1.61	−0.01	3.89	−1.33	2.74	−0.98	2.68	−0.69	3.43
phenol dimer (C_1)	−7.05	−1.36	2.29	−0.83	5.21	−1.18	3.07	−0.97	1.81	−0.89	3.52
Signed Average Errors											
hydrogen bonding		−2.01	1.80	−2.28	3.71	−2.18	0.44	−2.04	−0.17	−2.34	0.10
single H-bond		−0.89	1.64	−1.23	2.81	−0.84	1.02	−0.95	0.34	−0.81	1.11
cyclic H-bond		−3.36	1.99	−3.54	4.80	−3.80	−0.26	−3.34	−0.78	−4.17	−1.11
dispersion		−2.51	0.78	−0.66	4.43	−1.31	2.78	−0.75	3.22	−0.63	3.56
mixed		−1.16	1.00	−0.42	2.26	−1.05	1.58	−0.83	1.40	−0.72	1.96
total		−1.94	1.27	−1.28	3.54	−1.61	1.47	−1.32	1.30	−1.38	1.67
total (neglecting cyclic H-bonds)		−1.60	1.10	−0.74	3.24	−1.09	1.88	−0.83	1.79	−0.71	2.33
Unsigned Average Errors											
hydrogen bonding		2.05	1.80	2.28	3.71	2.18	0.91	2.04	0.66	2.37	1.20
single H-bonds		0.96	1.64	1.23	2.81	0.84	1.02	0.95	0.40	0.88	1.11
cyclic H-bond		3.36	1.99	3.54	4.80	3.80	0.77	3.34	0.97	4.17	1.30
dispersion		2.51	0.97	0.77	4.45	1.31	2.98	0.75	3.28	0.63	3.63
mixed		1.16	1.00	0.42	2.26	1.05	1.58	0.83	1.40	0.72	1.96
total		1.95	1.33	1.32	3.55	1.61	1.73	1.32	1.67	1.39	2.15
total (neglecting cyclic H-bonds)		1.62	1.17	0.79	3.25	1.09	1.95	0.83	1.83	0.73	2.35

[a] Errors are calculated by subtracting the calculated values from the high level ones and are given in kcal/mol. CP denotes counterpoise-corrected results, and no-CP indicates results obtained with no counterpoise correction.

determined at the MP2/cc-pVTZ level as described above (for the supplementary hydrogen-bonding complexes). Interaction energies are computed both with and without the counterpoise (CP)[41] correction scheme of Boys and Bernardi, which compensates for the basis set superposition error (BSSE). For each of the basis sets included in this study, with the exception of aug-cc-pVTZ, MP2 calculations were made using the Gaussian suite of molecular structure programs.[42] MP2/aug-cc-pVTZ calculations were carried out using the RI-MP2 method as implemented in the Turbomole 5.8 molecular structure program package.[43]

Discussion

Following is a discussion concerning the results obtained with the counterpoise correction for the basis set superposition error, below the results for non-counterpoise-corrected calculations, which are generally very poor, will be discussed.

Table 1 gives the MP2 interaction energy errors, along with the average signed and unsigned (absolute) errors, associated with the Pople type (6-31G*, 6-31G*(0.25), 6-31+G*, and 6-311+G*) and TZVP basis sets (which we will refer to as the "small" basis sets). Here it can be seen that 6-31G*(0.25) and TZVP yield overall average unsigned interaction energy errors of 1.32 kcal/mol, the lowest value among these basis sets. Not surprisingly, the highest unsigned error of 1.95 kcal/mol is obtained with 6-31G*, the only basis here that does not contain diffuse functions. Another observation that can be made from these data is that the MP2 method, when paired with each of these basis sets, tends to underestimate the stabilization energies for all types of interactions considered here. Indeed, with only

three exceptions ($(H_2O)_2$ with 6-31G* and 6-311+G*, and pyrazine dimer with 6-31G*(0.25)), the stabilization energies of all of the systems, along with each of these basis sets, is underestimated. When cyclic hydrogen bonds, which are treated very poorly by these methods, are ignored, the MP2/6-311+G* method gives the lowest overall average unsigned interaction energy error with a value of 0.73 kcal/mol. It is also worth pointing out that the, very small, 6-31G*(0.25) basis performs relatively well, giving an average unsigned error of 0.79 kcal/mol.

In terms of hydrogen-bonding, TZVP yields the lowest average unsigned interaction energy error with a value of 2.04 kcal/mol; the highest unsigned error of 2.37 kcal/mol is produced by the 6-311+G* basis. It should also be noted that the 6-31G* basis set gives a, relatively low, unsigned error of 2.05 kcal/mol. Stabilization energies for the cyclic hydrogen bonds are underestimated to a much greater extent than those of the single hydrogen bonding complexes, with average signed errors for cyclic hydrogen bonding complexes generally being about 4−5 times greater than those of single hydrogen bonding species. The best result for single hydrogen bonds is obtained with the 6-31+G* basis with an average unsigned error of 0.84 kcal/mol whereas, for cyclic hydrogen bonds, the lowest unsigned error of 3.34 kcal/mol is produced with TZVP. Single hydrogen bonds involving a carbonyl oxygen as the hydrogen bond acceptor are generally not as well described by the MP2 method (with the small basis sets) as those whose hydrogen bond acceptor is an amino nitrogen or hydroxyl oxygen (including the water dimer).

Considering the results for dispersion bound complexes, the MP2/6-311+G* method, with an average unsigned interaction energy error of 0.63 kcal/mol, is the best performer among the methods being considered here. The TZVP and 6-31G*(0.25) basis sets give unsigned errors of 0.75 and 0.77 kcal/mol, respectively, and the 6-31G* and 6-31+G* bases yield poor results with average unsigned errors of 2.51 and 1.31 kcal/mol, respectively. Among the dispersion bound complexes treated here, the stacked uracil dimer proved to be particularly problematic for MP2 methods, the best result for this system, corresponding to an error of 1.36 kcal/mol, was obtained with the 6-311+G* basis.

Somewhat surprisingly, the best result for the mixed interactions is obtained with the 6-31G*(0.25) basis, with an average unsigned interaction energy error of 0.42 kcal/mol. The next best result, corresponding to an average error of 0.72 kcal/mol, is obtained with the 6-311+G* basis set. The 6-31G* and 6-31+G* basis sets both yield relatively high unsigned errors of 1.05 and 1.16 kcal/mol, respectively. The stabilization energy of the phenol dimer is consistently underestimated with all of these methods by about 1 kcal/mol; the best result for this complex is obtained with 6-31G*(0.25), which is in error by −0.83 kcal/mol.

Table 2 gives the MP2 interaction energy errors, along with the signed and unsigned average errors, associated with the Dunning type basis sets (cc-pVDZ, cc-pVDZ+, aug-cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ). Here it can be seen that the aug-cc-pVTZ basis set produces the lowest average unsigned interaction energy error with a value of 0.67 kcal/mol. The largest average error of 1.93 kcal/mol is obtained with the cc-pVDZ basis set. The smaller basis sets considered here, cc-pVDZ and cc-pVDZ+, both have a strong tendency to underestimate stabilization energies of all types, with the former underestimating the interaction energies of all the complexes in the test set and the latter underestimating all but one of them (benzene dimer (C_{2h})). The largest basis set employed in this study, aug-cc-pVTZ, tends to overestimate the stabilization energies of the dispersion bound and mixed complexes while underestimating the binding energies of most of the hydrogen-bonding complexes. When cyclic hydrogen bonds, for which all MP2 methods investigated in this study tend to yield high errors, are neglected, the MP2/cc-pVTZ method yields the lowest overall error of 0.51 kcal/mol.

In terms of hydrogen bonding there is clearly a tendency for the accuracy of the interaction energies to increase as the size of the basis set is increased, this tendency can also be seen for both single hydrogen bonds and cyclic ones. The aug-cc-pVTZ basis set gives the most accurate hydrogen bonding interaction energies, with an overall average unsigned error of 0.46 kcal/mol, and can be said to be the only basis set considered in this work that yields relatively good results for cyclic hydrogen bonds (unsigned error of 0.76 kcal/mol). It is interesting to note that for each basis set considered here, with the exception of aug-cc-pVTZ, all of the hydrogen bonding interaction energies are underestimated. The aug-cc-pVTZ basis underestimates the binding energies of all hydrogen-bonding complexes except the methyl amide dimers (both α and β). As in the case of the small basis sets, most of the Dunning bases yield lower errors for single hydrogen bonds in which the hydrogen bond acceptor is an amino nitrogen or a hydroxyl oxygen rather than a carbonyl oxygen.

Considering dispersion interactions, the MP2/cc-pVTZ method yields the lowest unsigned interaction energy error with a value of 0.61 kcal/mol. It is also notable that the, relatively small, cc-pVDZ+ basis gives the second lowest unsigned error of 0.76 kcal/mol. The aug-cc-pVTZ basis set overestimates the stabilization energies of all but two of the dispersion bound complexes (methane dimer and ethene dimer) and overestimates the stabilization energies of the pyrazine dimer, indole−benzene complex, and stacked adenine−thymine complex by more than 2.00 kcal/mol.

For the mixed interactions, the best results are obtained with the aug-cc-pVDZ and cc-pVTZ basis sets with average unsigned errors of 0.25 and 0.26 kcal/mol, respectively. It is interesting to note that the MP2/aug-cc-pVTZ method overestimates the stabilization energies of all these interactions.

Extrapolated MP2 complete basis set interaction energies for the systems in the S26-07 test set are given in Table 3. Here it can be seen that the MP2/CBS results for hydrogen-bonding complexes are generally quite good, with errors for all but two of the systems being below 0.1 kcal/mol. Errors for the 2-pyridoxine−2-aminopyridine and the adenine−thymine (WC) complexes (cyclic hydrogen bonding) are larger, with values of 0.66 and 0.17 kcal/mol, respectively. As would be expected, the MP2/CBS method generally overbinds dispersion bound and mixed complexes significantly, with interaction energies that are in error by up to 66% (for the parallel displaced benzene dimer).

Here we will briefly describe the interaction energy data obtained without the use of the counterpoise correction method. These results are generally much worse than those produced using the counterpoise correction and would not be recommended for studies on interactions of noncovalent complexes in biological systems. As might be expected, non-counterpoise calculations generally underestimate stabilization energies because of their tendency to underestimate the energies of monomers relative to those of their corresponding complexes. This overbinding tendency can be seen for each of the basis sets employed in this work. For two of the basis sets, 6-31G*

TABLE 2: Interaction Energy Errors for the MP2 Method When Paired with the Dunning Type Basis Sets cc-pVDZ, cc-pVDZ+, aug-cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ[a]

		cc-pVDZ		cc-pVDZ+		aug-cc-pVDZ		cc-pVTZ		aug-cc-pVTZ	
	high level	CP	no-CP	CP	no-CP	CP	no-CP	CP	no-CP	CP	no-CP
Single Hydrogen Bonds											
$(NH_3)_2$ (C_{2h})	−3.17	−1.27	1.72	−0.67	0.09	−0.49	0.20	−0.40	0.88	−0.18	0.07
$(H_2O)_2$ (C_s)	−5.02	−1.10	2.36	−0.87	0.25	−0.66	0.18	−0.57	1.10	−0.33	0.14
methanol dimer (C_1)	−5.70	−1.21	2.76	−0.88	0.77	−0.47	0.72	−0.49	1.47	−0.07	0.56
methanol−formaldehyde (C_1)	−5.31	−2.14	2.54	−1.62	−0.19	−0.69	0.46	−0.80	1.23	−0.20	0.39
methyl amide dimer (α)	−6.69	−1.92	1.10	−1.65	0.14	−0.68	1.06	−0.72	0.71	0.22	0.70
methyl amide dimer (β)	−7.65	−2.25	1.67	−1.67	−0.16	−0.88	0.82	−0.93	0.71	0.27	0.71
Cyclic Hydrogen Bonds											
formic acid dimer (C_{2h})	−18.61	−4.55	2.73	−4.91	−1.71	−2.62	−0.05	−1.72	1.09	−1.06	0.46
formamide dimer (C_{2h})	−15.96	−4.30	2.20	−4.02	−1.70	−2.01	0.20	−1.71	0.77	−0.93	0.32
uracil dimer (C_{2h})	−20.65	−4.52	1.09	−4.03	−0.53	−2.24	1.07	−1.91	0.42	−1.10	0.84
2-pyridoxine−2-aminopyridine (C_1)	−16.71	−2.89	2.91	−2.69	0.71	−1.15	2.26	−0.80	1.71	−0.12	1.71
adenine−thymine WC (C_1)	−16.37	−3.71	2.81	−3.30	0.17	−1.66	2.01	−1.45	1.18	−0.58	1.50
Dispersion Bound											
$(CH_4)_2$ (D_{3d})	−0.53	−0.45	−0.22	−0.42	−0.23	−0.14	0.39	−0.21	−0.11	−0.07	0.07
$(C_2H_4)_2$ (D_{2d})	−1.51	−1.10	−0.04	−0.99	−0.12	−0.33	0.59	−0.36	0.08	−0.05	0.36
benzene−CH_4 (C_3)	−1.50	−1.02	0.05	−0.78	0.44	−0.14	1.67	−0.07	0.39	0.20	0.91
benzene dimer (C_{2h})	−2.73	−1.08	1.44	0.09	3.26	1.51	5.38	1.01	2.44	1.97	3.55
pyrazine dimer (C_s)	−4.42	−1.17	2.03	−0.09	3.18	1.59	5.46	1.02	2.59	2.14	3.67
uracil dimer (C_2)	−10.12	−3.27	2.57	−2.20	3.97	0.11	5.88	−0.09	2.97	0.50	3.01
indole−benzene (C_1)	−5.22	−1.74	2.13	−0.07	4.73	1.84	7.53	1.20	3.43	2.52	4.84
adenine−thymine stack (C_1)	−12.23	−3.19	4.78	−1.45	6.38	1.35	9.71	0.91	5.06	2.04	5.51
Mixed Interactions											
ethene−ethyne (C_{2v})	−1.53	−0.45	0.32	−0.49	0.14	−0.14	1.00	−0.10	0.20	0.05	0.46
benzene−H_2O (C_s)	−3.28	−1.11	0.53	−0.78	0.62	−0.26	1.43	−0.34	0.96	0.07	0.88
benzene−NH_3 (C_s)	−2.35	−0.97	0.28	−0.72	0.64	−0.17	1.59	−0.21	0.62	0.17	0.90
benzene−HCN (C_s)	−4.46	−1.05	0.49	−0.85	0.94	−0.01	2.56	0.14	0.82	0.46	1.58
benzene dimer (C_{2v})	−2.74	−0.84	1.00	−0.39	1.96	0.41	3.81	0.26	1.14	0.72	2.09
indole−benzene T-shape (C_1)	−5.73	−1.04	1.56	−0.65	2.53	0.47	4.74	0.45	1.87	0.98	2.87
phenol dimmer (C_1)	−7.05	−1.81	2.62	−1.35	2.07	−0.32	2.96	−0.35	1.66	0.31	1.95
Signed Average Errors											
hydrogen bonding		−2.71	2.17	−2.39	−0.20	−1.23	0.81	−1.05	1.02	−0.37	0.67
single H-bond		−1.65	2.03	−1.23	0.15	−0.65	0.57	−0.65	1.02	−0.05	0.43
cyclic H-bond		−3.99	2.35	−3.79	−0.61	−1.94	1.10	−1.52	1.03	−0.76	0.97
dispersion		−1.63	1.59	−0.74	2.70	0.72	4.58	0.43	2.11	1.16	2.74
mixed		−1.04	0.97	−0.75	1.27	0.00	2.58	−0.02	1.04	0.39	1.53
total		−1.93	1.67	−1.44	1.09	−0.30	2.45	−0.32	1.36	0.31	1.54
total (neglecting cyclic H-bonds)		−1.44	1.51	−0.88	1.50	0.09	2.77	−0.03	1.44	0.56	1.68
Unsigned Average Errors											
hydrogen bonding		2.71	2.17	2.39	0.58	1.23	0.82	1.05	1.02	0.46	0.67
single H-bonds		1.65	2.03	1.23	0.27	0.65	0.57	0.65	1.02	0.21	0.43
cyclic H-bond		3.99	2.35	3.79	0.96	1.94	1.12	1.52	1.03	0.76	0.97
dispersion		1.63	1.66	0.76	2.79	0.88	4.58	0.61	2.13	1.19	2.74
mixed		1.04	0.97	0.75	1.27	0.25	2.58	0.26	1.04	0.39	1.53
total		1.93	1.69	1.45	1.45	0.86	2.45	0.70	1.37	0.67	1.54
total (neglecting cyclic H-bonds)		1.44	1.53	0.89	1.56	0.60	2.77	0.51	1.45	0.64	1.68

[a] Errors are calculated by subtracting the calculated values from the high level ones and are given in kcal/mol. CP denotes counterpoise-corrected results, and no-CP indicates results obtained with no counterpoise correction.

and cc-pVDZ, the non-counterpoise-corrected interaction energies are predicted to be more accurate than their counterpoise-corrected counterparts. This may be indicative of a tendency for the counterpoise correction to overestimate the basis set superposition error. This phenomenon is especially evident in cases where relatively small basis sets are employed. The lowest overall average unsigned binding energy error is obtained by the smallest basis set considered here, 6-31G*, with a value of 1.33 kcal/mol.

It is interesting to mention that counterpoise and non-counterpoise-corrected MP2 interaction energies for hydrogen bonding and stacked nucleic acids have been shown to converge when a very large basis set (aug-cc-pV5Z) is used.[44] This is an expected result, as the basis set superposition error should become smaller as larger basis sets are employed.

Conclusions

In assessing the performance of MP2 with the various basis sets studied here it is important to consider not only the overall average performance for a particular basis but also the quality of the results obtained for each type of interaction. For instance, if a method were to produce very good hydrogen bonding and mixed interaction results but describe dispersion interactions fairly poorly, it might yield a fairly good overall average interaction energy error, but it could not be said to be a good general method for treating noncovalent interactions. It is very important that any method used to describe biologically relevant intermolecular interactions be well balanced, in terms of its description of electrostatic and dispersion forces, because in practice these forces can rarely be separated, and generally there will always be some electrostatic and dispersion component

TABLE 3: Complete Basis Set Limit Counterpoise-Corrected MP2 Interaction Energies and CCSD(T)-Corrected Interaction Energies for Model Complexes in the S26-07 Test Set (Interaction Energies in kcal/mol)

	ΔE_{CBS}^{MP2}	$\Delta E_{CCSD(T)/CBS}$
Single Hydrogen Bonds		
$(NH_3)_2$ (C_{2h})	−3.20	−3.17
$(H_2O)_2$ (C_s)	−5.03	−5.02
methanol dimer (C_1)	−5.79	−5.70
methanol− formaldehyde (C_1)	−5.32	−5.31
methyl amide dimer (α)	−6.66	−6.69
methyl amide dimer (β)	−7.63	−7.65
Cyclic Hydrogen Bonds		
formic acid dimer (C_{2h})	−18.60	−18.61
formamide dimer (C_{2h})	−15.86	−15.96
uracil dimer (C_{2h})	−20.61	−20.65
2-pyridoxine−2-aminopyridine (C_1)	−17.37	−16.71
adenine−thymine WC (C_1)	−16.54	−16.37
Dispersion Bound		
$(CH_4)_2$ (D_{3d})	−0.51	−0.53
$(C_2H_4)_2$ (D_{2d})	−1.62	−1.51
benzene−CH_4 (C_3)	−1.86	−1.50
benzene dimer (C_{2h})	−4.95	−2.73
pyrazine dimer (C_s)	−6.90	−4.42
uracil dimer (C_2)	−11.39	−10.12
indole−benzene (C_1)	−8.12	−5.22
adenine−thymine stack (C_1)	−14.93	−12.23
Mixed Interactions		
ethene−ethyne (C_{2v})	−1.69	−1.53
benzene−H_2O (C_s)	−3.61	−3.28
benzene−NH_3 (C_s)	−2.72	−2.35
benzene−HCN (C_s)	−5.16	−4.46
benzene dimer (C_{2v})	−3.62	−2.74
indole−benzene T-shape (C_1)	−7.03	−5.73
phenol dimer (C_1)	−7.76	−7.05

within any noncovalent interaction. The following discussion will focus on counterpoise-corrected results and we will only note here that, in general, MP2 does not yield accurate interaction energies when the counterpoise correction method is not employed.

Overall, the aug-cc-pVTZ and cc-pVTZ basis sets yield the lowest average unsigned interaction energy errors with values of 0.67 and 0.70 kcal/mol, respectively. The aug-cc-pVTZ basis set produces very good results for hydrogen bonding and mixed interactions but tends to strongly overestimate the stabilization energies of dispersion bound systems whereas cc-pVTZ generally gives fairly good results for each of the interaction types and describes mixed interactions particularly well. The conclusion that we would draw from these observations is that, among the basis sets studied here, cc-pVTZ produces the most accurate, and well balanced, interaction energies when used with the MP2 method. It must, however, be stressed again that this is due to fortuitous compensation of errors.

Among the small basis sets, 6-31G*(0.25) and TZVP produce the best results, both yielding an average error of 1.32 kcal/mol. Both of these basis sets produce fairly high errors for hydrogen bonding interactions and relatively low ones for dispersion interactions; 6-31G*(0.25), however, gives substantially lower errors for mixed systems than TZVP. In cases where the use of large basis sets with the MP2 method is not possible, because of large system size, and only qualitative results are required, the 6-31G*(0.25) basis seems to be the best choice, among basis sets studied here, for the computation of noncovalent interaction energies.

One of the main conclusions that can be drawn from this study is that MP2 methods do not generally describe cyclic hydrogen bonds very well, the aug-cc-pVTZ basis set, the largest one considered here, produces reasonably low errors for these types of interactions and cc-pVTZ can perhaps be said to describe cyclic hydrogen bonds to a qualitative level. This result has deep implications for the study of nucleic acids, such as those found in DNA, which form base pairs through cyclic hydrogen bonds. It should be noted that cyclic hydrogen bonds are not particularly common within proteins or most protein−ligand complexes, and so this deficiency of MP2 is not critically important for studies of these types of systems.

When cyclic hydrogen bonding interactions are neglected, the cc-pVTZ and aug-cc-pVDZ basis sets produce the most accurate interaction energies when used with MP2, with average unsigned errors of 0.51 and 0.60 kcal/mol, respectively. These bases give essentially identical results for both hydrogen bonding and mixed interactions, but cc-pVTZ produces an error that is about 0.3 kcal/mol lower than that of aug-cc-pVDZ for interactions dominated by dispersion forces. These results lead us to believe that the cc-pVTZ basis set is the best basis set to use, along with MP2, for the computation of noncovalent interactions in proteins and protein−ligand systems. It should also be noted that, with the omission of the cyclic hydrogen bonds, the 6-311+G* basis set produces relatively good results, with an average unsigned interaction energy error of 0.73 kcal/mol. This basis set gives a very balanced description of intermolecular interactions, giving its largest average error of 0.88 kcal/mol for dispersion interactions.

In this work it has been shown that the MP2 method can be used to obtain interaction energies, for biologically relevant noncovalent complexes, that are at least semiquantitative if the basis set is chosen carefully. For general purposes, we feel that the cc-pVTZ basis offers the best balance in terms of its description of noncovalent interactions arising from both electrostatic and dispersion forces. One interesting question that has not been addressed in this study is the quality of the local and global geometry minima obtained by the MP2 technique when paired with various basis sets. We believe that this would be a very interesting avenue for future investigation.

Acknowledgment. This work has been part of the research project Z4 055 905 and was supported by grants from the Grant Agency of the Czech Republic (203/05/0009), the Grant Agency of the Academy of Sciences of the Czech Republic (A400550510), and the Ministry of Education of the Czech Republic (Center for Biomolecules and Complex Molecular Systems, LC512).

References and Notes

(1) Johnson, E. R.; Dilabio, G. A. *Chem. Phys. Lett.* **2006**, *419*, 333.
(2) Riley, K. E.; Merz, K. M., Jr. *J. Chem. Theory Comput.* **2007**, *3*, 407.
(3) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 1009.
(4) Johnson, E. R.; Wolkow, R. A.; DiLabio, G. A. *Chem. Phys. Lett.* **2004**, *394*, 334.
(5) Vondrasek, J.; Bendova, L.; Klusak, V.; Hobza, P. *J. Am. Chem. Soc.* **2005**, *127*, 2615.
(6) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 415.
(7) Riley, K. E.; Merz, K. M., Jr. *J. Phys. Chem. B* **2006**, *110*, 15650.
(8) Guo, H. B.; Beahm, R. F.; Guo, H. *J. Phys. Chem. B* **2004**, *108*, 18065.
(9) Vanommeslaeghe, K.; Mignon, P.; Loverix, S.; Tourwe, D.; Geerlings, P. *J. Chem. Theory Comput.* **2006**, *2*, 1444.
(10) Daabkowska, I.; Jurecka, P.; Hobza, P. *J. Chem. Phys.* **2005**, *122*, 204322.
(11) Fukuzawa, K.; Mochizuki, Y.; Tanaka, S.; Kitaura, K.; Nakano, T. *J. Phys. Chem. B* **2006**, *110*, 16102.
(12) Dobes, P.; Otyepka, M.; Strnad, M.; Hobza, P. *Chem.-a Eur. J.* **2006**, *12*, 4297.

(13) Riley, K. E.; Cui, G.; Merz, K. M., Jr. *J. Phys. Chem. B* **2007**, *111*, 5700.
(14) Aylward, N.; Bofinger, N. *J. Theor. Biol.* **2007**, *244*, 258.
(15) Hunter, K. C.; Millen, A. L.; Wetmore, S. D. *J. Phys. Chem. B* **2007**, *111*, 1858.
(16) Rutledge, L. R.; Campbell-Verduyn, L. S.; Hunter, K. C.; Wetmore, S. D. *J. Phys. Chem. B* **2006**, *110*, 19652.
(17) Fukuzawa, K.; Komeiji, Y.; Mochizuki, Y.; Kato, A.; Nakano, T.; Tanaka, S. *J. Comput. Chem.* **2006**, *27*, 948.
(18) Spiwok, V.; Lipovova, P.; Skalova, T.; Buchtelova, E.; Hasek, J.; Kralova, B. *Carbohydr. Res.* **2004**, *339*, 2275.
(19) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
(20) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
(21) Werner, H. J.; Manby, F. R.; Knowles, P. J. *J. Chem. Phys.* **2003**, *118*, 8149.
(22) Jung, Y. S.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793.
(23) Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80.
(24) Weigend, F.; Haser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.
(25) Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143.
(26) Jurecka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
(27) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289.
(28) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.
(29) Zhao, Y.; Truhlar, D. G. *Theo. Chem. Acc.*, in press.
(30) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287.
(31) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.
(32) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2006**, *124*, 014104.
(33) Cerny, J.; Jurecka, P.; Hobza, P.; Valdes, H. *J. Phys. Chem. A* **2007**, *111*, 1146.
(34) Gonzalez, C.; Allison, T. C.; Lim, E. C. *J. Phys. Chem. A* **2001**, *105*, 10583.
(35) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
(36) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.
(37) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2006**, *124*, 174104.
(38) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555.
(39) Williams, R. W.; Malhotra, D. *Chem. Phys.* **2006**, *327*, 54.
(40) Johnson, E. R.; Becke, A. D. *Chem. Phys. Lett.* **2006**, *432*, 600.
(41) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*.
(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Chesseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Forseman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; AlLoham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 03*; Gaussian Inc.: Wallingford, CT, 2003.
(43) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.
(44) Dabkowska, I.; Jurecka, P.; Hobza, P. *J. Chem. Phys.* **2005**, *122*, 204322.