7820 *J. Phys. Chem. A* **2002,** *106,* 7820−7827

An ab Initio Quantum Mechanical Study of Hydrogen-Bonded Complexes of Biological Interest

Niu Huang and Alexander D. MacKerell, Jr.*

Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland,
Baltimore, Maryland 21201

Received: January 10, 2002; In Final Form: May 24, 2002

A series of hydrogen-bonded dimers are examined via several quantum mechanics (QM) methods, including the Hartree−Fock (HF), second-order Moller−Plesset perturbation (MP2), and local MP2 (LMP2) theories, with different basis sets. The effects of electron correlation, basis set size, and basis set superposition error (BSSE) are systematically analyzed and results are compared with available experimental data. Results from the lower levels of theories examined depend on “error cancelation” and in some cases are not in satisfactory agreement with experiment or high level QM calculations. Higher level methods yield improved results, with the LMP2/cc-pVQZ single point energy evaluation for geometries optimized at the MP2/6-31G* level indicated to be a reliable and economical procedure for accurately determining both the hydrogen-bonding geometries and energies. The importance of the inclusion of electron correlation during geometry optimization is discussed.

I. Introduction

Hydrogen bonds play a critical role in many chemical and biochemical processes. They not only dominate the influence of the aqueous environment on molecular structure, but also play a key role in determining the structure and function of biomolecules, such as proteins and nucleic acids.[1,2] It is therefore critical to accurately determine the geometries and energies of hydrogen-bonded systems in order to reliably quantitate their impact on biological systems.

Due to the difficulties in measuring hydrogen-bond geometries and strengths by experimental methods, quantum mechanical (QM) calculations have been widely used to determine these quantities. Until recently, the majority of QM studies of hydrogen bonding have been low-level electronic structure calculations at the Hartree−Fock (HF) level with relatively small basis sets. However, these levels of theory are not adequate for the reliable prediction of hydrogen-bond geometries and strengths.[3,4] These failures are due to the errors concerning basis set incompleteness, basis set superposition error (BSSE), and the neglect of correlation energy.[5]

Recently, QM studies of hydrogen bonding have been extended to include treatment of electron correlation. Density functional theory (DFT), especially the popular B3LYP approach,[6] has been shown to yield reliable predictions of the geometries and stabilization energies of some hydrogen-bonded and ionic complexes.[7−13] However, it is known that DFT is not able to reproduce dispersion interactions and other weak forces.[14] This is problematic for weak hydrogen-bonded systems, where the dispersion effects become more important as has been shown for a number of weak interacting systems.[15] Thus, the use of DFT theory for a wide range of intermolecular interactions is not appropriate.

On the contrary, MP2[16] theory accounts for the full range of intermolecular interactions: electrostatic, induction and dispersion effects.[17] Accordingly, it is believed to be reliable for calculations of intermolecular interactions when a sufficiently large basis set is used. However, recent studies indicate that a very large basis set, such as cc-pVQZ or cc-pV5Z,[18] is essential to study interactions of hydrogen-bonded complexes in conjunction with MP2 level calculations.[4] Such large basis sets, however, are computationally intractable for calculations on biological molecules, such as nucleic acid base pairs, and are complicated by the presence of basis set superposition errors (BSSE).[19]

The most widely used method to correct for BSSE is the counterpoise (CP) method proposed by Boys and Bernardi.[20] This approach eliminates the BSSE by treating the monomers in the basis set of the supermolecular complex instead of their individual basis sets. Introducing the CP correction for BSSE proved to be crucial to the present success of the supermolecular approach to study weak-bonded complexes.[5,21,22] However, in practice the CP correction is computationally expensive and a tedious procedure due to the two extra energy evaluations required for each dimer calculation. In addition, there is evidence that the BSSE values obtained using the CP method are only approximate and that this method overestimates the size of the BSSE with small basis sets.[23]

The local MP2 (LMP2) method is virtually free from BSSE on the correlated level via the use of a subset of the virtual orbitals for the perturbation calculation.[24] In addition, LMP2 offers significant computational savings over MP2.[25] With a large basis set, LMP2 yields good agreement with experimental data for the conformational energy differences of small organic molecules.[26] The LMP2 method has been applied to study hydrogen bonding between amino acid backbone and side chain analogues.[27] However, rigorous comparison of LMP2 calculated hydrogen-bonding interactions against available experimental data and high level QM calculations has not been performed.

In the present study, a systematic study of hydrogen-bond interactions at selected QM levels of theory is performed. Comparison is made with available experimental data and previously reported high level QM calculations. On the basis of these comparisons, levels of theory are suggested that are

---

* To whom correspondence should be addressed. E-mail: alex@outerbanks.umaryland.edu. Fax: 410-706-0346.

10.1021/jp020059n CCC: $22.00

appropriate for accurately determining hydrogen-bonding geometries and energies.

The remainder of this paper is organized as follows. In section II, we discuss the computational methodology. Section III presents the results and compares them with experimental data and previous high level QM calculations. The effect of different basis sets, BSSE and correlation energies is analyzed in detail. Finally, the conclusions are presented in section IV.

II. Method

All HF and MP2 calculations were performed using the program Gaussian 98.[28] LMP2 single-point energy calculations were performed with the program Jaguar 4.1.[29] Full geometry optimizations were carried out at the HF or MP2 level with the 6-31G* basis set[30] and with an extended triple-ζ-type 6-311++G** basis set.[31] The uncorrected hydrogen-bonding energies were obtained by subtracting the energy of two fully optimized monomers from the energy of the dimer. For MP2 interaction energy calculations, BSSE was estimated by the standard CP method.[20] For the LMP2 calculations with the 6-311++G** and cc-pVQZ(-g) basis sets,[18] the interaction energies were corrected for BSSE at the HF level. Throughout the manuscript cc-pVQZ indicates the cc-pVQZ(-g) basis set. The BSSE-corrected interaction energy, $E_{\text{corrected}}$, is determined as follows:

$$E_{\text{corrected}} = \{E(\text{AB}) - [E(\text{A})_{\text{mon}} + E(\text{B})_{\text{mon}}]\} + E_{\text{BSSE}} \quad (1)$$

$$E_{\text{BSSE}} = \{E(\text{A})_{\text{dim}} - E(\text{A})_{\text{dim}}^{*} + E(\text{B})_{\text{dim}} - E(\text{B})_{\text{dim}}^{*}\} \quad (2)$$

where E(AB) represents the energy of the dimer, and E_{mon} represent the energy of the individually optimized monomer. E_{BSSE} is the correction for the BSSE, where E_{dim} and E_{dim}* are the energies of the monomers in their supermolecular (i.e., dimer) optimized geometries for the basis set of the monomer and the dimer, respectively.

The experimental interaction energies reported in the present work for the CH3COO$^-$—imidazole and imidazole—imidazolium dimers were derived from experimental enthalpies of binding by correcting for the E_{ZPE} using unscaled MP2/6-31G* frequencies,[8] in which E_{ZPE} is the change in zero-point energy between the individual monomers and the dimer.

III. Results and Discussion

A common practice in determining interaction energies has been to obtain optimized structures at a relatively low level of theory and evaluate the energetic properties at a higher level theory. In situations where full geometry optimization at the high level is prohibitively expensive, the use of a lower level theory may offer a valuable compromise. However, the use of low level approximations has to be carefully investigated on a wide range of interacting pairs, including comparison with experimental data and high-level QM data in order to determine its accuracy.

The pseudospectral local MP2 (LMP2) methods in the program Jaguar[29] provide a feasible methodology for investigating significantly larger systems with the treatment of electron correlation and large basis sets. LMP2 reduces the computational scaling with system size to $N^{2.5}$, in contrast to N^5 for canonical MP2 approaches.[25] Moreover, it has been suggested that the LMP2 methodology is particularly useful in examining hydrogen-bonding energies.[32] As presented above, electron correlation is necessary for calculating accurate structural properties of hydrogen-bonded complexes. However, previous LMP2 studies of interaction energies have all been based on geometry optimization at the HF level followed by single point LMP2 energy evaluations.[27,32−34] Unfortunately, calculations performed in our laboratory as part of the present study (not shown) and by Hobza et al.[35] have indicated that the use of the LMP2 method for full optimization of molecular complexes can be problematic. Therefore, for supermolecular complexes, LMP2 calculations are currently most useful for single-point energy evaluations.

Hydrogen-bonded dimers were chosen to encompass a wide range of hydrogen-bond strengths and geometries and on the availability of experimental or high level QM data. They include (I) H_2O-H_2O, (II) NH_3-NH_3, (III) NH_3-H_2O, (IV$_{a,b}$) CH_3OH-H_2O (water as both the hydrogen-bond donor and acceptor), (V) CH_3OH-CH_3OH, (VI) Me_2O-H_2O, (VII) H_2CO-H_2O, (VIII) HCOOH−HCOOH, (IX) CH_3COO^-−imidazole (Aac$^-$−Im), and (X$_{a,b}$) imidazole−imidazolium (Im−Im$^+$) (both parallel and antiparallel conformation) and are shown in Figure 1. Hydrogen-bonding energies, intermolecular distances, and angles from the full geometry optimizations using the HF/6-31G*, MP2/6-31G*, and MP2/6-311++G** level of theories are reported in Table 1. Comparing the geometries from the three methods shows the heteroatom to heteroatom distances to decrease from between 0.05 and 0.1 Å upon going from the HF/6-31G* to the MP2/6-31G* method. Increasing the basis set size from 6-31G* to 6-311++G** at the correlated level further decreases the intermolecular distances by 0.03 Å, except for the charged species where decreases of up to 0.08 Å occur. The only exception to these trends is the NH_3-NH_3 dimer. As may be seen, significant differences in the interaction distances and donor angle as a function of the level of theory occur, indicating this dimer to be especially sensitive to computational protocol.

It is essential to compare the calculated hydrogen-bond geometries with available experimental data. For the water dimer, the experimental O···O distance is 2.95 Å after correction for the anharmonicity of the dimer vibrations.[36,37] The HF calculation gives a longer distance (2.97 Å), while the MP2 values of 2.92 Å are in good agreement with the most recent spectroscopic determination of 2.92 Å.[38] For the $(HCOOH)_2$ dimer, the reported experimental O···O distance is 2.70 Å.[39,40] As with the water dimer, the HF calculated distance (2.80 Å) is too long with the MP2/6-31G* value of 2.75 Å being shorter, although still longer than the experimental estimate. For the $(HCOOH)_2$ dimer, the MP2/6-311++G** calculated distance of 2.72 Å is in the best agreement with experiment. For the imidazole-imidazolium dimer, no direct experimental evidence is available. However, the crystal structure of 9-ethylguanine hemihydrochloride[41] contains a neutral and a protonated imidazole moiety. These moieties interact directly, having an intermolecular hydrogen-bond N···N distance of 2.64 Å. Both the HF/6-31G* (2.84 Å) and MP2/6-31G* (2.73 Å) values are longer than this while the MP2/6-311++G** value (2.65 Å) is in satisfactory agreement. Similarly, the short intermolecular hydrogen-bond N···O distance of 2.65 Å between the carboxylate and imidazole moieties in the crystal structure of *cyclo*-(L-histidyl-L-aspartyl) trihydrate indicates the MP2/6-311++G** value of 2.64 Å for the Aac$^-$−imidazole dimer to be reasonable. Both the HF/6-31G* (2.84 Å) and MP2/6-31G* (2.73 Å) are significantly longer than the experimental value. Therefore, MP2/6-31+G* optimizations were performed to test if the addition of diffuse function can improve the geometries without significantly increasing the computational requirements. Surprisingly, for the four largest dimers (HCOOH−HCOOH, Aac$^-$−Im, Im−Im$^+$ (anti), and Im−Im$^+$ (para)), inclusion of the diffuse functions actually increased the interaction distances, yielding

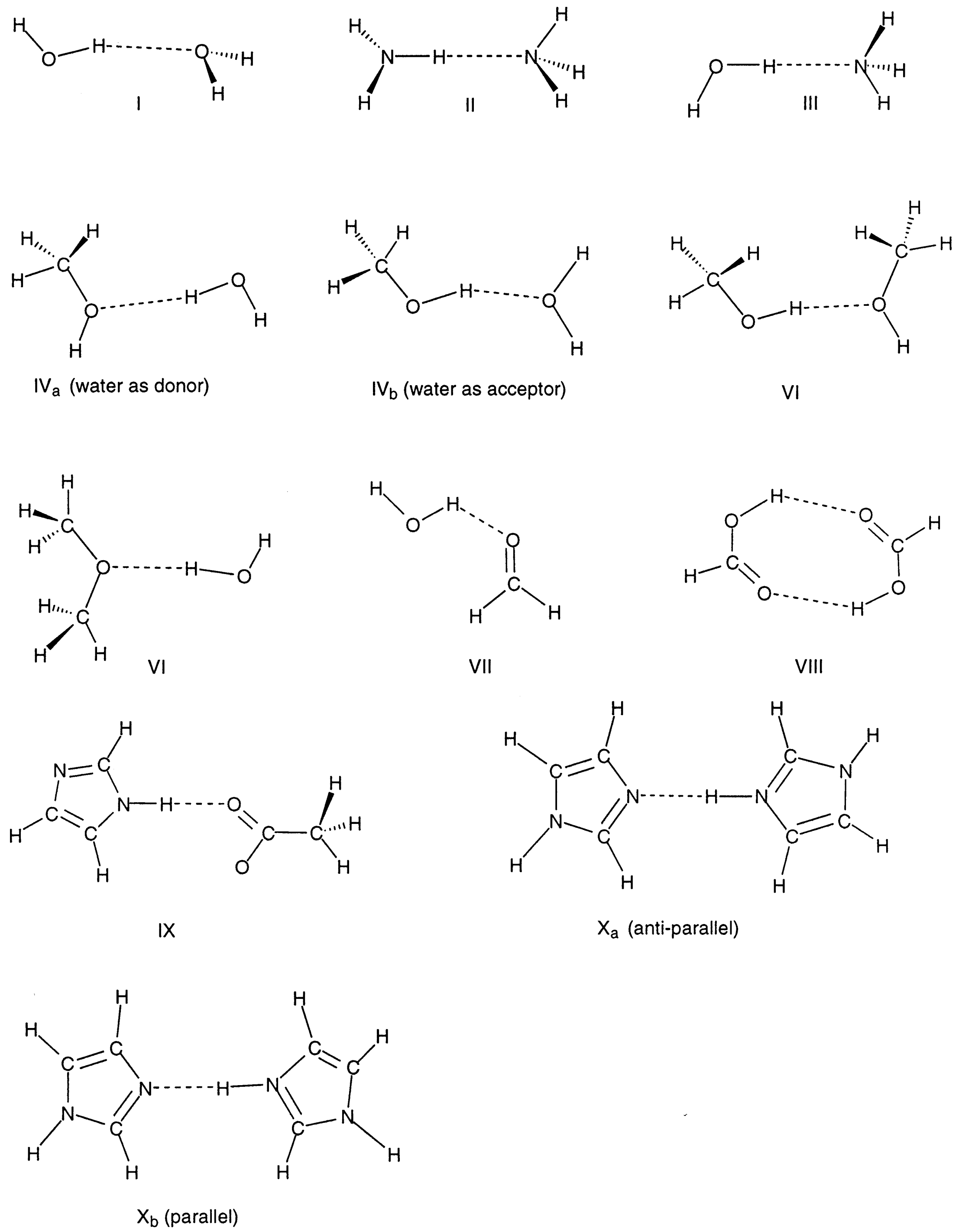

Figure 1. Structures of the studied hydrogen-bonded dimers. (I) H_2O-H_2O, (II) NH_3-NH_3, (III) NH_3-H_2O, ($IV_{a,b}$) CH_3OH-H_2O (water as both the hydrogen-bond donor and acceptor), (V) CH_3OH-CH_3OH, (VI) Me_2O-H_2O, (VII) H_2CO-H_2O, (VIII) HCOOH−HCOOH, (IX) $CH3COO^--$imidazole (Aac^--Im), and ($X_{a,b}$) imidazole−imidazolium ($Im-Im^+$) (both parallel and antiparallel conformations).

poorer agreement with the available experimental data and MP2/6-311++G** values (Table 1, Supporting Information). In previous studies, HF calculations generally gave longer intermolecular distances than the corrected experimental values[42−44] while MP2/6-31G* optimized geometries were shown to yield noticeable improvements as compared to the HF optimized geometries.[26,34] MP2/6-311++G** calculated geometries of some hydrogen-bonded complexes, including charged species, have been shown to be in good agreement with experimental data.[45,46] Thus, the present results, along with previous studies, indicate that it is necessary to explicitly treat electron correlation in order to obtain accurate geometries in hydrogen-bonded systems. The MP2/6-31G* method appears to be adequate for the optimization of neutral systems, though the larger basis sets may be required for treating charged species, while HF/6-31G* calculations yield results that are consistently too long.

Interaction energies from the HF and MP2 optimizations are included in Table 1. Comparing the uncorrected HF and MP2 energies shows that, although the HF calculations do capture a large portion of the interaction energies, the inclusion of electron correlation makes the hydrogen-bonding energies significantly more favorable, especially with the charged species, Aac^--Im and Im^+-Im. However, upon going to the 6-311++G** basis set, the uncorrected MP2 energies become less favorable. This

TABLE 1: Minimum Interaction Geometries and Energies at HF/6-31G*, MP2/6-31G*, and MP2/6-311++G Levels of Theory**

		distance (Å)		angle(degree)	energy (kcal/mol)		
complex[a]	level of theory	X···Y	H···Y	X−H···Y	$E_{\text{uncorrected}}$	E_{BSSE}	$E_{\text{corrected}}$
H_2O-H_2O	HF/6-31G*[b]	2.97	2.03	172.3	−5.63		
	MP2/6-31G*	2.92	1.96	166.5	−7.32	2.23	−5.09
	MP2/6-311++G**	2.92	1.95	176.8	−6.08	1.62	−4.46
NH_3-NH_3	HF/6-31G*[b]	3.42	2.42	179.0	−3.07		
	MP2/6-31G*	3.11	2.29	136.7	−4.56	2.00	−2.56
	MP2/6-311++G**	3.26	2.26	166.3	−3.76	1.09	−2.67
NH_3-H_2O	HF/6-31G*[b]	3.04	2.09	176.4	−6.56		
	MP2/6-31G*	2.97	1.99	175.3	−8.40	2.41	−5.99
	MP2/6-311++G**	2.94	1.98	171.4	−7.32	1.56	−5.76
CH_3OH-H_2O	HF/6-31G*[b]	2.96	2.01	169.0	−5.55		
(water as donor)	MP2/6-31G*	2.87	1.93	160.2	−7.72	3.16	−4.56
	MP2/6-311++G**	2.85	1.89	170.3	−6.67	1.77	−4.90
CH_3OH-H_2O	HF/6-31G*[b]	2.96	2.01	174.9	−5.60		
(water as acceptor)	MP2/6-31G*	2.90	1.93	170.8	−7.47	2.26	−5.21
	MP2/6-311++G**	2.91	1.94	179.1	−6.15	1.82	−4.33
CH_3OH-CH_3OH	HF/6-31G*[b]	2.95	2.00	174.6	−5.54		
	MP2/6-31G*	2.85	1.91	160.4	−8.04	3.40	−4.64
	MP2/6-311++G**	2.85	1.89	171.9	−6.81	1.96	−4.85
H_2CO-H_2O	HF/6-31G*[b]	2.96	2.11	147.6	−5.28		
	MP2/6-31G*	2.91	2.03	148.4	−6.94	1.54	−5.40
	MP2/6-311++G**	2.89	2.01	150.2	−5.22	1.07	−4.15
Me_2O-H_2O	HF/6-31G*[b]	2.95	2.02	163.9	−5.31		
	MP2/6-31G*	2.86	1.93	159.8	−7.47	3.12	−4.35
	MP2/6-311++G**	2.83	1.86	173.4	−6.89	1.93	−4.96
HCOOH−HCOOH	HF/6-31G*[c]	2.80	1.84	173.5	−15.28		
	MP2/6-31G*	2.75	1.75	178.8	−18.26	5.11	−13.15
	MP2/6-311++G**	2.72	1.73	178.0	−14.53	2.82	−11.71
$Aac^- -Im$	HF/6-31G*[c]	2.79	1.77	179.7	−29.51		
	MP2/6-31G*	2.72	1.66	180.0	−36.01	7.10	−28.91
	MP2/6-311++G**	2.64	1.56	178.0	−30.82	3.28	−27.54
$Im-Im^+$ (anti)	HF/6-31G*[c]	2.84	1.81	177.9	−23.60		
	MP2/6-31G*	2.73	1.64	178.9	−29.54	2.70	−26.84
	MP2/6-311++G**	2.65	1.54	179.1	−28.71	2.27	−26.43
$Im-Im^+$(para)	HF/6-31G*[c]	2.84	1.81	176.8	−23.46		
	MP2/6-31G*	2.73	1.64	176.8	−29.50	2.70	−26.80
	MP2/6-311++G**	2.65	1.55	176.8	−28.66	2.26	−26.40

[a] CH_3COO^-−imidazole (Aac^--Im); imidazole−imidazolium ($Im-Im^+$). [b] Reference 59. [c] From the present study.

indicates that uncorrected MP2 calculations with the smaller basis set overestimate the hydrogen-bonding energies due to the BSSE, which vary from 1.3 to 7.1 kcal/mol with the 6-31G* basis set. Though the BSSE is smaller with the 6-311++G** basis set (1.1−3.3 kcal/mol), it still has a significant influence on the interaction energies. In general, the uncorrected MP2/6-311++G** energies are more favorable than the experimental or high level QM estimates while the corresponding CP corrected energies are underestimated. For example, the uncorrected and corrected MP2/6-311++G** energy for the water dimer are −6.08 and −4.46 kcal/mol, respectively, as compared to the most often quoted experimental value of −5.4 ± 0.7 kcal/mol.[47] Similarly, the uncorrected and corrected MP2/6-311++G** energy for the $(HCOOH)_2$ dimer are −14.53 and −11.71 kcal/mol, respectively, versus the experimental value of −13.2 kcal/mol.[4,48,49] Thus, our present results are in good agreement with a previous study indicating that, due to BSSE, a very large basis set, such as the cc-pVQZ and cc-pV5Z sets, is essential to study hydrogen-bonding interactions in MP2 level calculations.[4,22,50]

Use of LMP2 to treat electron correlation, as discussed in the Introduction, has been indicated to be an effective method to remove BSSE from the correlation portion of the interaction energy. The remaining BSSE at the HF level may be eliminated by standard CP corrections or by basis set saturation.[24] Table 2 presents interaction energy data calculated by three different methods: MP2/6-311++G**//MP2/6-31G*, LMP2/6-311++G**//MP2/6-31G*, and LMP2/6-311++G**//MP2/6-311++G**. Table 3 presents LMP2/cc-pVQZ energies on geometries optimized by HF/6-31G*, MP2/6-31G*, and MP2/6-311++G**. Available experimental data and published high level ab initio results are also included in Table 3 for comparison. Analysis of the HF associated BSSE estimated by the CP method (in parentheses in Table 2) shows that LMP2 with the 6-311++G** basis set is still not adequate to eliminate the BSSE, which ranges from 0.25 to 1.13 kcal/mol (average = 0.64 kcal/mol) for the different dimers. Upon going to Dunning's cc-pVQZ correlation consistent basis set, however, the BSSE becomes much smaller, ranging from 0.02 to 0.99 kcal/mol (average = 0.25 kcal/mol) for the different dimers. Moreover, LMP2/6-311++G** does not always accurately estimate the interaction energies. For example, the corrected LMP2/6-311++G**//MP2/6-31G* energy for the water dimer is −4.05 kcal/mol. Even in conjunction with MP2/6-311++G** optimized geometry, the LMP2/6-311++G** calculated energy (−4.47 kcal/mol) is still not able to accurately reproduce the experimental value of −5.4 ± 0.7 kcal/mol.[47] However, the corrected LMP2/cc-pVQZ//MP2/6-311++G** value of −4.84 kcal/mol is in good agreement with experiment, especially as compared to the most recent spectroscopic determination of −4.91 kcal/mol.[38] For the methanol dimer, the corrected LMP2/6-311++G** energies calculated from MP2/6-31G* and MP2/6-311++G** optimized geometries are −3.70 and −4.37 kcal/mol individually, as compared to the experimental range of −4.6 to −5.9 kcal/mol.[4,51,52] The corrected LMP2/cc-pVQZ//MP2/6-311++G** value of −4.82 kcal/mol is in good agreement

TABLE 2: Comparison of Hydrogen-Bond Energies Calculated by Three Different Basis Sets at the MP2 and LMP2 Levels of Theory[a]

level of theory/complex[b]	MP2/6-311++G**//MP2/6-31G*			LMP2/6-311++G**//MP2/6-31G*			LMP2/6-311++G**//MP2/6-311++G**		
H_2O-H_2O	−5.39	(1.30)	−4.09	−4.56	(0.50)	−4.05	−5.11	(0.64)	−4.47
NH_3-NH_3	−3.40	(0.90)	−2.50	−2.49	(0.24)	−2.25	−3.06	(0.42)	−2.64
NH_3-H_2O	−7.33	(2.03)	−5.30	−6.25	(1.13)	−5.12	−6.23	(0.66)	−5.57
CH_3OH-H_2O (water as donor)	−5.70	(1.49)	−4.21	−4.59	(0.57)	−4.02	−5.35	(0.64)	−4.71
CH_3OH-H_2O (water as acceptor)	−5.60	(1.57)	−4.03	−4.50	(0.65)	−3.84	−5.00	(0.75)	−4.25
H_2CO-H_2O	−4.79	(1.24)	−3.55	−3.67	(0.24)	−3.43	−4.24	(0.22)	−4.02
Me_2O-H_2O	−6.26	(1.68)	−4.58	−4.81	(0.54)	−4.27	−5.32	(0.64)	−4.68
HCOOH−HCOOH	−14.47	(2.70)	−11.77	−11.79	(0.80)	−10.99	−11.96	(0.83)	−11.13
Aac^-−Im	−30.51	(3.03)	−27.48	−26.95	(0.67)	−26.28	−29.80	(0.70)	−29.10
Im−Im^+ (anti)	−28.52	(2.03)	−26.49	−22.69	(0.60)	−22.09	−22.42	(0.70)	−21.72
Im−Im^+ (para)	−28.50	(2.03)	−26.47	−22.58	(0.63)	−21.95	−22.24	(0.64)	−21.60

[a] Energies in kcal/mol. For each level of theory, column one is the uncorrected interaction energy, column two (in parentheses) is the BSSE correction energy and column three is the BSSE corrected interaction energy. [b] CH_3COO^-−imidazole (Aac^--Im); imidazole−imidazolium (Im−Im^+).

TABLE 3: Comparison of Hydrogen-Bond Energies Calculated at the LMP2/cc-pVQZ Level of Theory on the Geometries Optimized by Different Basis Sets[a]

level of theory/ complex[m]	LMP2/ cc-pVQZ//HF/6-31G*			LMP2/ cc-pVQZ//MP2/6-31G*			LMP2/ cc-pVQZ//MP2/6-311++G**			exptl	other QM data
H_2O-H_2O	−4.99	(0.18)	−4.81	−4.95	(0.25)	−4.70	−4.95	(0.11)	−4.84	−5.4 ± 0.7[b] −4.91[c]	−5.0 ± 0.1[d]
NH_3-NH_3	−2.81	(0.28)	−2.53	−2.86	(0.31)	−2.55	−3.05	(0.42)	−2.63		−2.67[e]
NH_3-H_2O	−6.01	(0.14)	−5.87	−6.27	(0.99)	−5.28	−6.32	(0.39)	−5.93		−5.5[f]
CH_3OH-H_2O (water as donor)	−5.44	(0.21)	−5.23	−5.02	(0.36)	−4.66	−5.20	(0.23)	−4.97		−5.2[g]
CH_3OH-H_2O (water as acceptor)	−4.98	(0.16)	−4.82	−4.62	(0.14)	−4.48	−4.69	(0.02)	−4.67		−4.6[g] −4.9[h]
CH_3OH-CH_3OH	−5.61	(0.45)	−5.16	−4.81	(0.23)	−4.58	−4.95	(0.13)	−4.82	−4.6 to −5.9[i]	−4.98[j]
H_2CO-H_2O	−4.85	(0.12)	−4.73	−4.97	(0.07)	−4.90	−4.99	(0.06)	−4.93		−5.15[c] −5.17[h]
Me_2O-H_2O	−5.28	(0.18)	−5.10	−5.47	(0.44)	−5.03	−5.63	(0.17)	−5.46		−5.51[h]
HCOOH−HCOOH	−13.40	(0.30)	−13.10	−14.37	(0.43)	−13.94	−14.39	(0.42)	−13.97	−13.2[k]	−13.93[h]
Aac^-−Im	−28.56	(0.58)	−27.98	−29.26	(0.66)	−28.60	−29.36	(0.64)	−28.72	−28.8[l]	
Im−Im^+ (anti)	−25.82	(0.09)	−25.73	−26.68	(0.09)	−26.59	−26.32	(0.09)	−26.23	−24.0[l]	
Im−Im^+ (para)	−25.90	(0.12)	−25.78	−26.78	(0.11)	−26.67	−26.43	(0.13)	−26.30		

[a] Energies in kcal/mol. For each level of theory, column one is the uncorrected interaction energy, column two (in parentheses) is the BSSE correction energy, and column three is the BSSE corrected interaction energy. [b] Reference 47. [c] Reference 38. [d] Reference 43. [e] Reference 13. [f] Reference 32. [g] Reference 56. [h] Reference 4. [i] References 4 and 51−52. [j] Reference 61. [k] References 4 and 48−49. [l] Reference 62, derived from experimental binding enthalpies by back-correcting the zero-point energy contribution using unscaled MP2/6-31G* frequency calculations. [m] CH_3COO^-−imidazole (Aac^--Im); imidazole−imidazolium (Im−Im^+).

with the experimental data. Similarly, the corrected LMP2/cc-pVQZ//MP2/6-311++G** energy for the $(HCOOH)_2$ dimer is −13.97 kcal/mol in satisfactory agreement with the experimental value of −13.2 kcal/mol[4,48,49] versus the LMP2/6-311++G** energy of −11.13 kcal/mol. Accordingly, the LMP2/cc-pVQZ method yields improved agreement with experiment and is able to eliminate the majority of the BSSE associated with the HF aspect of the calculations.

The purpose of this study, as discussed above, is to identify a feasible computational method for the determination of hydrogen-bonding geometries and energies, that is accessible for calculations on relatively large systems (e.g. nucleic acid base pairs). As shown above, the LMP2/cc-pVQZ//MP2/6-311++G** calculated hydrogen-bond geometries and energies are in good agreement with experimental and published high level ab initio data. However, the use of MP2/6-311++G** for geometry optimization is very expensive and only accessible for the optimization of small systems. Accordingly, the use of lower levels of theory for geometry optimization is necessary. Presented in Figure 2 are scatter plots of the uncorrected LMP2/cc-pVQZ//HF/6-31G* (Figure 2A) or the uncorrected LMP2/cc-pVQZ//MP2/6-31G* energies (Figure 2B) verses the corrected LMP2/cc-pVQZ//MP2/6-311++G** values. In both cases the correlation coefficients are close to one as are the slopes, while the *Y* intercept is somewhat larger with the HF/6-31G* optimized structures (Figure 2A) as compared the MP2/6-31G* optimized structures (Figure 2B). These results suggest that the use of the smaller basis set for geometry optimization with MP2 to treat the correlation energy may not adversely impact the results.

Another important question that needs to be addressed is the validity of using lower level theories for determination of the interaction energies, such as HF/6-31G* and MP2/6-31G*, in conjunction with scaling methods. Scatter plots of the uncorrected HF/6-31G* or the corrected MP2/6-31G* energies versus the corrected LMP2/cc-pVQZ//MP2/6-311++G** values are included in Figure 2. It is striking that both the uncorrected HF energies (Figure 2C) and corrected MP2/6-31G* energies (Figure 2D) have high correlation with the corrected LMP2/cc-pVQZ//MP2/6-311++G** data. In addition, the slopes are

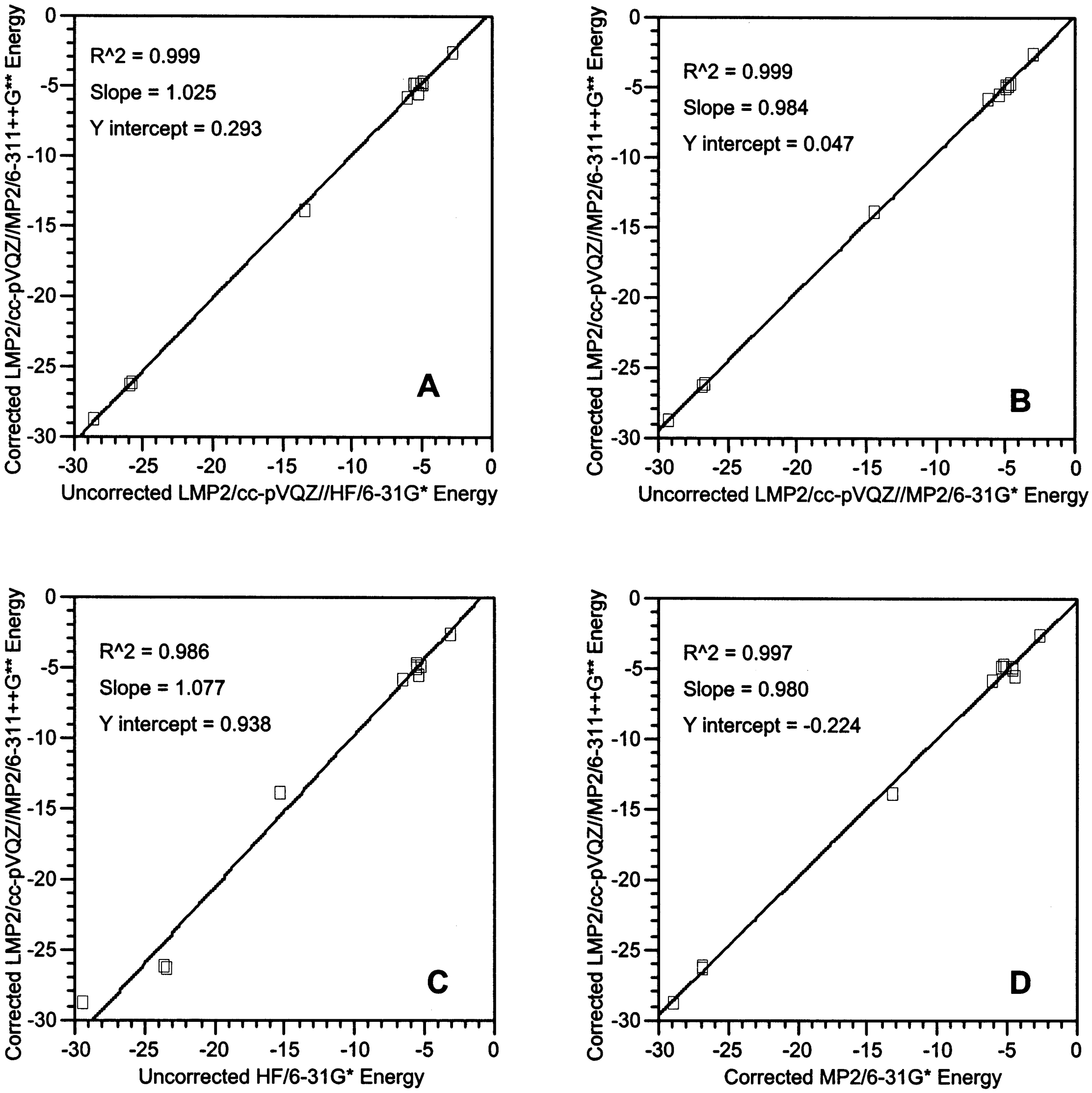

Figure 2. Relationship between the corrected LMP2/cc-pVQZ//MP2/6-311++G** energies and (A) the uncorrected LMP2/cc-pVQZ//HF/6-31G* energies, (B) the uncorrected LMP2/cc-pVQZ//MP2/6-31G* energies, (C) the uncorrected HF/6-31G* energies, and (D) the corrected MP2/6-31G* energies. Energies in kcal/mol.

close to one and the Y intercepts are relatively small. This appears to be due to "error cancelation" in both low level of theories (see above). Such high correlation may also be why most force fields based on these low level theories are able to achieve good results in biomolecular simulations.[53−55] Therefore, the question of whether it is necessary to employ the high level theory proposed here versus a scaled, lower level of theory still remains.

The above question may be answered by analyzing results on the methanol−water dimer. In the methanol−water heterodimer each monomer may behave either as a hydrogen-bond donor or as a hydrogen-bond acceptor. The LMP2/cc-pVQZ energies values (Table 3) show that the most stable methanol-water heterodimer conformer corresponds to that in which water behaves as a hydrogen-bond donor, being favored by 0.4 kcal/mol. This result is in good agreement with high level ab initio G2 theory studies[56] and with the most recent experimental data on methanol−water heterodimers.[57,58] However, both the uncorrected HF/6-31G* and corrected MP2/6-31G* calculations predict the dimer where water is an acceptor to be more favorable by 0.06 and 0.57 kcal/mol, respectively. This discrepancy indicates that the low level theories are not able to reproduce experimental results in some cases.

Halgren did extensive comparisons of different empirical force fields with scaled HF/6-31G* data for hydrogen-bonded complexes.[59] His results show that the HF/6-31G* method as well as many of the force fields predict that water and methanol have comparable strengths as hydrogen-bond donors and acceptors, with some force fields predicting methanol to be a stronger hydrogen-bond donor than acceptor. This error is suggested to be due to the use of lower level ab initio data in their respective parametrizations.[53−55] Therefore, this scenario emphasizes the importance of using high level QM calculations to reproduce the subtle aspects of hydrogen-bond interactions.

In all the cases studied in the present work, the uncorrected LMP2/cc-pVQZ energies for geometries optimized by MP2/6-31G* are able to combine the advantages of both MP2 and LMP2 methods to successfully reproduce available experimental and high level QM data. However, this procedure might not achieve good results for much bigger hydrogen-bonded systems, such as nucleic acid base pairs. Therefore, the GC−WC base pair was selected to test the LMP2/cc-pVQZ//MP2/6-31G* method. The calculated hydrogen-bonding energy of −21.63 kcal/mol is in excellent agreement with the corrected experimental value of −22.2 kcal/mol.[34,60] Notably, the uncorrected LMP2/cc-pVQZ energy value is −19.05 kcal/mol for G:C pair from the HF/6-31G* optimized geometry, confirming our observation (see above) that the HF optimized geometries are problematic for accurately calculating hydrogen-bond properties.

IV. Conclusion

Progress in computational chemistry now allows for the use of sufficiently large basis sets and correlated methods in QM calculations on model compounds of biological relevance. In the present work, it was shown that use of the LMP2 method with cc-pVQZ basis set greatly reduces the computational cost and diminishes the BSSE at both the correlated and HF levels. In contrast, when MP2 is used to treat election correlation, even triple-ζ-type basis sets are not sufficient to adequately reduce the BSSE and the CP correction appears to overestimate the BSSE. Moreover, such calculations are still prohibitively expensive for large hydrogen-bonded systems, such as nucleic acid base pairs.

Comparison of the tested levels of theory with available experimental and high level QM data indicates that the uncorrected LMP2/cc-pVQZ//MP2/6-31G* is an accurate, yet computationally tractable approach to calculate hydrogen-bond geometries and energies. However, limitations in the use of MP2/6-31G* optimized geometries for interactions involving charged species should be noted. Tests of lower levels of theory indicate that, although good correlation with higher level QM data is achieved, in specific cases these methods fail.

Accurate hydrogen-bond-interaction geometries and energies will greatly aid in parametrizing molecular mechanics force fields for modeling and simulation studies of biological systems. However, the use of QM results as target data for the optimization of force fields must include checks against available experimental data whenever possible to ensure that the force field adequately reproduces experimental data.

Acknowledgment. This work has been financially supported by NIH grant GM51501. We also thank the National Partnership for Advanced Computational Infrastructure and the DOD ASC Major Shared Resource Computing and High Performance Computing for providing computational resources.

Nomenclature

Abbreviations

HF: Hartree-Fock
MP2: second-order Moller-Plesset
LMP2: local MP2
BSSE: basis set superposition error
CP: counterpoise
QM: quantum mechanics
DFT: density functional theory

Supporting Information Available: A table containing the MP2/6-31+G* optimized geometric and energetic data and the total energies in hartrees and full geometric data for all ab initio optimized structures (including monomers and dimers) in *XYZ* format. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Jeffrey, G. A. *An introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997.

(2) Scheiner, S. *Hydrogen Bonding*; Oxford University Press: New York, 1997.

(3) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 520.

(4) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *J. Chem. Phys.* **1999**, *110*, 11906.

(5) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **2000**, *100*, 4227.

(6) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.

(7) Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J. Phys. Chem. A* **1998**, *102*, 3782.

(8) Del Bene, J. E.; Person, W. B.; Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 9, 10705.

(9) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.

(10) Tuma, C.; Boese, A. D.; Handy, N. C. *Phys. Chem. Chem. Phys.* **1999**,*1*, 3939.

(11) Kieninger, M.; Suhai, S. *J. Comput. Chem.* **1996**, *17*, 1508.

(12) Lozynski, M.; Rusinska-Roszak, D. *J. Phys. Chem. A* **1998**, *102*, 2899.

(13) Novoa, J. J.; Sosa, C. *J. Phys. Chem.* **1995**, *99*, 15837.

(14) Zaremba, E.; Kohn, W. *Phys. Rev.* **1976**, *B13*, 2278.

(15) Rappe, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.

(16) Moller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.

(17) Krishnan, R.; Pople, J. A. *Int. J. Quantum Chem.* **1978**, *14*, 91.

(18) Dunning, Jr., T. H. *J. Chem. Phys.* **1989**, *90*, 1007.

(19) Kestner, N. R. *J. Chem. Phys.* **1968**, *48*, 252.

(20) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(21) Feller, D. *J. Chem. Phys.* **1992**, *96*, 6104.

(22) Novoa, J. J.; Plannas, M.; Whangbo, M. H. *Chem. Phys. Lett.* **1994**, *225*, 240.

(23) Gutowski, M.; Chalasinski, G. *J. Chem. Phys.* **1993**, *98*, 5540.

(24) Saebo, S.; Tong, W.; Pulay, P. *J. Chem. Phys.* **1993**, *98*, 2170.

(25) Friesner, R. A.; Dunietz, B. D. *Acc. Chem. Res.* **2001**, *34*, 351.

(26) Murphy, R. B.; Beachy, M. D.; Friesner, R. A. *J. Chem. Phys.* **1995**, *103*, 1481.

(27) Kim, K.; Friesner, R. A. *J. Am. Chem. Soc.* **1997**, *119*, 12952.

(28) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millan, 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.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortic, 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.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98, revision A.9; Gaussian, Inc.: Pittsburgh, PA, 1998.

(29) Jaguar, 4.0 ed.; Schrodinger, Inc: Portland, OR, 2000.

(30) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213.

(31) Frisch, M. J.; Pople, J. A.; Binkley, J. S. *J. Chem. Phys.* **1984**, *80*, 3265.

(32) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 11775.

(33) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.

(34) Brameld, K.; Dasgupta, S.; Goddard, III, W. A. *J. Phys. Chem. B* **1997**, *101*, 4851.

(35) Hobza, P.; Havlas, Z. *Chem. Rev.* **2000**, *100*, 4253.

(36) Dyke, T. R.; Mack, K. M.; Muenter, J. S. *J. Chem. Phys.* **1977**, *66*, 498.

(37) Odutola, J. A.; Dyke, T. R. *J. Chem. Phys.* **1980**, *72*, 5062.

(38) Fellers, R. S.; Leforestier, C.; Braly, L. B.; Brown, M. G.; Saykally, R. J. *Science* **1999**, *284*, 945.

(39) Harmony, M. D.; Laurie, V. W.; Kiczkowski, R. L.; Schwendeman, R. H.; Ramsay, D. A.; Lovas, F. J.; Lafferty, W. J.; Maki, A. G. *J. Phys. Chem.* **1979**, *8*, 619.

(40) Turi, L. *J. Phys. Chem.* **1996**, *100*, 11285.

(41) Mandel, G. S.; Marsh, R. E. *Acta Crystallogr.* **1975**, *B31*, 2862.

(42) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Phys. Chem. A* **1999**, *103*, 1640.

(43) Feyereisen, M. W.; Feller, D.; Dixon, D. A. *J. Phys. Chem.* **1996**, *100*, 2993.

(44) Gould, I. R.; Kollman, P. A. *J. Am. Chem. Soc.* **1994**, *116*, 2493.

(45) Alkorta, I.; Elguero, J. *J. Phys. Chem. A* **1999**, *103*, 272.

(46) Araujo, R. C. M. U.; Ramos, M. N. *J. Braz. Chem. Soc.* **1998**, *9*, 499.

(47) Curtiss, L. A.; Frurip, D. J.; Blander, M. *J. Chem. Phys.* **1979**, *71*.

(48) Henderson, G. *J. Chem. Educ.* **1987**, *64*, 88.

(49) Colominas, C.; Teixido, J.; Cemeli, J.; Luque, F. J.; Orozco, M. *J. Phys. Chem. B* **1998**, *102*, 2269.

(50) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873.

(51) Curtiss, L. A.; Blander, M. *Chem. Rev.* **1988**, *88*, 827.

(52) Bizzarri, A.; Stolte, S.; Reuss, J.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Duijneveldt, F. B. *Chem. Phys.* **1990**, *143*, 423.

(53) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(54) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, J., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(55) MacKerell, Jr., A. D.; Bashford, D.; Bellott, M.; Dunbrack Jr., R. L.; Evanseck, J.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I., W. E.; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.

(56) Gonzalez, L.; Mo, O.; Yanez, M. *J. Chem. Phys.* **1998**, *109*, 139.

(57) Huisken, F.; Stemmler, M. *Chem. Phys. Lett.* **1991**, *180*, 332.

(58) Stockman, P. A.; Blake, G. A.; Lovas, F. J.; Suenran, R. D. *J. Chem. Phys.* **1997**, *107*, 3782.

(59) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730.

(60) Yanson, I. K.; Teplitsky, A. B.; Sukhodub, L. F. *Biopolymers.* **1979**, *18*, 1149.

(61) Mo, O.; Yanez, M.; Elguero, J. *J. Chem. Phys.* **1997**, *107*, 3592.

(62) Meot-Ner Mautner, M. *J. Am. Chem. Soc.* **1988**, *110*, 3075−3080.