

# Lennard–Jones Parameters for the Combined QM/MM Method Using the B3LYP/6-31+G\*/AMBER Potential

MAREK FREINDORF,<sup>1</sup> YIHAN SHAO,<sup>2</sup> THOMAS R. FURLANI,<sup>1</sup> JING KONG<sup>2</sup>

<sup>1</sup>Center for Computational Research, The State University of New York at Buffalo,  
Buffalo, New York 14260

<sup>2</sup>Q-Chem, Inc., 5001 Baum Blvd, Suite 690, Pittsburgh, Pennsylvania 15213

Received 16 June 2004; Accepted 14 April 2005

DOI 10.1002/jcc.20264

Published online in Wiley InterScience (www.interscience.wiley.com).

**Abstract:** A combined DFT quantum mechanical and AMBER molecular mechanical potential (QM/MM) is presented for use in molecular modeling and molecular simulations of large biological systems. In our approach we evaluate Lennard–Jones parameters describing the interaction between the quantum mechanical (QM) part of a system, which is described at the B3LYP/6-31+G\* level of theory, and the molecular mechanical (MM) part of the system, described by the AMBER force field. The Lennard–Jones parameters for this potential are obtained by calculating hydrogen bond energies and hydrogen bond geometries for a large set of bimolecular systems, in which one hydrogen bond monomer is described quantum mechanically and the other is treated molecular mechanically. We have investigated more than 100 different bimolecular systems, finding very good agreement between hydrogen bond energies and geometries obtained from the combined QM/MM calculations and results obtained at the QM level of theory, especially with respect to geometry. Therefore, based on the Lennard–Jones parameters obtained in our study, we anticipate that the B3LYP/6-31+G\*/AMBER potential will be a precise tool to explore intermolecular interactions inside a protein environment.

© 2005 Wiley Periodicals, Inc. J Comput Chem 26: 1270–1278, 2005

## Introduction

Development of combined quantum mechanical and molecular mechanical (QM/MM) models have been the subject of extensive research during the past decade. In this approach, a small part of a large molecular system (an active site of a protein, or a solute immersed in a solvent) is treated quantum mechanically (QM) using DFT, Hartree–Fock, or the semiempirical method, while the remainder of the system is modeled using a molecular mechanical (MM) force field such as OPLS, CHARMM, AMBER, or Tripos. The combined QM/MM method, first introduced by Warshel and Levitt,<sup>1</sup> has been developed by a number of researchers and used to study of a large number of chemical processes in solution and in biological systems such as enzymes.<sup>2–28</sup> A key problem in the successful application of this approach is accurately describing the interaction between the QM and the MM regions, which consists of two terms: (1) the electrostatic energy between QM particles and MM atoms, which is usually represented by point charges on MM particles; and (2) the van der Waals energy between QM and MM atoms, which is usually described by a Lennard–Jones 6-12 potential. Details of the theoretical background for this method have been described previously.<sup>29</sup> There are also studies presenting more complex interactions between QM and MM atoms,<sup>22,30</sup> however, they are devoted to different models of a connection

between the QM and MM subunits. In the simplest combined QM/MM approach, used by us in the present study, external charges and van der Waals parameters associated with MM atoms are taken from the MM force field used in particular calculations. However, van der Waals parameters describing the QM atoms are usually evaluated separately, because they depend on the QM level of theory used in the combined calculations. The best technique to estimate the Lennard–Jones parameters of the QM atoms, is to evaluate hydrogen bonds between the QM and MM regions of a series of bimolecular systems. Energies and geometries of hydrogen-bonded systems obtained from combined QM/MM calculations can be compared with results obtained from computations where the entire system is treated quantum mechanically. Calculations of a large set of different systems yield a short list of the Lennard–Jones parameters for the QM atoms, which can be subsequently used in any combined QM/MM approach utilizing the desired QM level of theory. The present investigation is a contin-

**Correspondence to:** M. Freindorf; e-mail: mfrein@buffalo.edu

Contract/grant sponsor: NIH SBIR; contract/grant number: 1 R43 GM065617

This article includes Supplementary Material available from the authors upon request or via the Internet at <http://www.interscience.wiley.com/jpages/0192-8651/suppmat>

uation of a previous study in which Lennard–Jones parameters were determined for a combined semiempirical AM1/TIP3P model<sup>31</sup> and a combined *ab initio* HF/3-21G/TIP3P approach.<sup>29</sup> In both previous studies, the MM environment was represented by the TIP3P model of water molecules, and therefore the Lennard–Jones parameters were determined for combined QM/MM simulations only in water solution. Lennard–Jones parameters for QM/MM simulations in protein environments, have been evaluated by Murphy, Philipp and Friesner,<sup>22</sup> but their method and results differ from those employed in the present investigation. The Murphy study employed a frozen molecular orbitals approximation to describe the interface between QM and MM regions. Moreover, they do not include diffuse functions for electrons in the QM part of the system, which appears to be important to correctly describe hydrogen bonds at the B3LYP level of theory.

In the present study we have investigated the Lennard–Jones parameters for the combined B3LYP/6-31+G\*/AMBER potential, where the QM part of the system is described at the B3LYP/6-31+G\* level of theory, and the MM environment is described by the AMBER program. In recent years the B3LYP hybrid functional has been widely employed in molecular modeling studies because of its overall accuracy and computational efficiency in describing electron correlation, particularly for transition metals.<sup>32</sup> Furthermore, this functional represents one of the best tools for molecular modeling of an active site of biological systems,<sup>33</sup> especially in combination with a 6-31G\* basis set, where polarization functions are placed on nonhydrogen atoms. The reliability of DFT for describing hydrogen-bonded complexes has been discussed extensively in the literature.<sup>34</sup> Recently, Jorgensen and coworkers<sup>35</sup> have examined over 50 hydrogen-bonded molecular systems between water and well-known simple organic molecules as acids, alcohols, and ketons. In the calculations, they have compared the B3LYP level of theory with the HF and MP2 approaches using a variety of different basis sets. They have found that DFT approach utilizing the 6-31G\* basis set is not entirely adequate for geometry optimization of these systems; however, geometry optimization using a basis set beyond the 6-31+G\*\* level (having diffuse functions on nonhydrogen atoms, and polarization functions on all atoms) yields little improvement. The calculations presented in their article have been devoted to the simplest hydrogen-bonded molecular systems, for which very large basis sets can be feasibly utilized. Therefore, the 6-31+G\* basis set (having diffuse and polarization functions only on nonhydrogen atoms), seems to be in our opinion a compromise between a high-level of accuracy in calculations of hydrogen-bonded systems and computational feasibility. It has been our experience that the inclusion of diffuse functions on nonhydrogen atoms with the 6-31+G\* basis set generally decreases hydrogen bond energies and increases hydrogen bond distances, yielding better agreement with experimental values. Hydrogen-bonded molecular systems have also been examined using the combined DFT/MM approach,<sup>36</sup> utilizing both the local and nonlocal DFT methods. In that article it has been attributed that the nonlocal DFT method, which is also used in the B3LYP approach, is important for correct treatment of hydrogen bonds. Therefore, we use in our study the B3LYP hybrid functional, which is widely employed in molecular modeling because of its computational accuracy in describing electron correlation in particular for hydrogen-bonded systems.

In our approach we are interested in the evaluation of Lennard–Jones parameters for use in combined B3LYP/6-31+G\*/AMBER calculations of the active site of proteins. Accordingly, we have chosen seven typical small organic molecules, namely methane, methanol, dimethyl ether, dimethyl ketone, acetic acid, methyl amine, and imidazol as a first reference set of molecules because they include some of the most common functional groups found at an active site. The molecules of the first reference set have been treated on a QM level of theory in our combined QM/MM computations. We have also chosen a second set of seven molecules representing a MM protein environment. The second set contains five amino acids namely asparagine, aspartic acid, histidine, lysine, serine, N-methylacetamide (NMA)—representing the protein backbone, and a single water molecule—characterizing protein waters. The molecules in the second set have been treated at an MM level of theory in our combined QM/MM calculations. From these two sets of molecules, we created all possible bimolecular systems, where functional groups of the molecules of the first set were hydrogen bonded to functional groups of the molecules of the second set, yielding a total of 116 different bimolecular systems.

## Computational Details

In the current study, combined QM/MM calculations at the B3LYP/6-31+G\*/AMBER level of theory were carried out for the 116 hydrogen bonded systems discussed above. First, we performed geometry optimizations for all hydrogen bonded molecular systems using the B3LYP/6-31+G\* level of theory. In these calculations, the molecules from the MM set, as well as those from the QM set, were treated quantum mechanically. Then for each bimolecular system, we performed combined QM/MM calculations at the B3LYP/6-31+G\*/AMBER level of theory, where the geometry of the QM monomer was fully optimized in the presence of external point charges and van der Waals potentials, representing the MM monomer. In these calculations, the MM atoms were fixed at their positions obtained from the previous calculations of the bimolecular complex performed at the QM level of theory. The MM parameters of these atoms including values of point charges and Lennard–Jones parameters represented by  $\sigma$  and  $\epsilon$  constants, were taken from the AMBER<sup>37</sup> program using the Cornell et al. (PARM94) force field,<sup>38</sup> for asparagine, aspartic acid, histidine, lysine, and serine monomers. For NMA and water, the MM parameters were taken from OPLS<sup>39</sup> and TIP3P<sup>40</sup> models, respectively. The MM parameters remain unchanged in our combined QM/MM calculations; therefore, our approach does not take into consideration the polarization of the MM environment by the QM system during its geometry optimization. The  $\sigma$  and  $\epsilon$  Lennard–Jones parameters for the QM atoms were then adjusted to obtain the best agreement between hydrogen bond energies and geometries, obtained from the combined QM/MM method and those from the full QM calculations. All calculations have been carried out in our investigation using the Q-Chem programming suite<sup>41</sup> and computational resources of the University at Buffalo's Center for Computational Research ([www.ccr.buffalo.edu](http://www.ccr.buffalo.edu)).

**Table 1.** Optimized Lennard–Jones Parameters of the Combined B3LYP/6-31+G\*/AMBER Potential Compared with the HF/3-21G/TIP3P Method, the AM1/TIP3P Model, and the AMBER'94 Force Field.

Element	B3LYP/6-31+G*/AMBER		HF/3-21G/TIP3P		AM1/TIP3P <sup>c</sup>		AMBER'94 <sup>c</sup>	
	$\epsilon$	$\sigma$	$\epsilon$	$\sigma$	$\epsilon$	$\sigma$	$\epsilon$	$\sigma$
C	0.06	4.02	0.08	3.80	0.08	3.50	0.10	3.40
H <sup>a</sup>	0.03	2.22	0.01	2.60	0.01	2.00	0.02	2.53
O	0.19	3.33	0.15	3.60	0.17	2.52	0.19	3.04
N	0.13	3.86	0.20	3.90	0.15	2.80	0.17	3.25
H <sup>b</sup>	0.03	1.06	0.10	1.30	0.10	0.80	0.01	0.36

<sup>a</sup>H is a hydrogen atom connected to carbon.

<sup>b</sup>H is a hydrogen atom connected to heteroatoms (O, N).

<sup>c</sup>For the AM1/TIP3P model and the AMBER'94 force field, there are different values of Lennard–Jones parameters for different types of the same atoms; therefore, for comparison we present arithmetic average values of them.

The  $\epsilon$  parameter is expressed in kcal/mol, and the  $\sigma$  parameter is expressed in Å.

## Results and Discussion

In evaluation of the Lennard–Jones parameters for the combined B3LYP/6-31+G\*/AMBER potential, we started with the values already used in the AMBER force field, and the values obtained previously from combined semiempirical AM1/TIP3P and *ab initio* HF/3-21G/TIP3P approaches. Then, we modified these parameters to reproduce hydrogen bond energies and hydrogen bond geometries of bimolecular systems, which were calculated at the QM level of theory (as described in Section II). The final set of the  $\sigma$  and  $\epsilon$  Lennard–Jones parameters for the combined B3LYP/6-31+G\*/AMBER method is presented in Table 1. A comparison

between our results and other relevant results reported in the literature is also included in Table 1. To show how the presented QM/MM method is sensitive on the QM level of theory used in the calculations, we present computed hydrogen bond energies and geometries of a short list of bimolecular systems investigated in our study, including only bimolecular complexes between water and acetic acid in Tables 2, 3, and 4. Table 2 presents computed hydrogen bond energies and geometries obtained using the HF and HF/AMBER models, Table 3 using the B3LYP and B3LYP/AMBER method, whereas Table 4 shows results of calculations using the BLYP and BLYP/AMBER approach. Table 5 presents a comparison of hydrogen bond energies and geometries of the short

**Table 2.** Computed Hydrogen Bond Energies and Geometries of the Acetic Acid and Water Bimolecular Systems Using the HF Level of Theory and the HF/AMBER Model with the Lennard–Jones Parameters Evaluated from the B3LYP/6-31+G\*/AMBER Potential.

Molecule	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$
		HF/3-21G			HF/3-21G/AMBER	
acd_wat_a1	17.82	1.69	106	7.73	1.82	105
acd_wat_a2	20.15	1.63	180	9.75	1.79	180
acd_wat_d1	20.15	1.94	108	9.75	2.17	108
acd_wat_d2	11.03	1.91	111	6.04	1.97	117
acd_wat_d3	8.25	1.92	180	4.81	1.83	180
		HF/6-31G			HF/6-31G/AMBER	
acd_wat_a1	11.52	1.79	120	8.29	1.91	123
acd_wat_a2	14.48	1.76	180	10.51	1.88	180
acd_wat_d1	14.48	2.09	109	10.51	2.26	110
acd_wat_d2	7.65	1.96	116	6.97	2.00	120
acd_wat_d3	5.53	2.01	180	5.44	1.89	180
		HF/6-31+G*			HF/6-31+G*/AMBER	
acd_wat_a1	7.67	1.91	106	8.21	1.97	107
acd_wat_a2	9.06	1.92	180	10.39	1.96	180
acd_wat_d1	9.06	2.18	112	10.39	2.35	109
acd_wat_d2	5.46	2.04	120	6.54	2.12	119
acd_wat_d3	3.19	2.16	180	4.16	2.11	180

The hydrogen bond energy  $E_{\text{HB}}$  is expressed in kcal/mol, the hydrogen bond distance  $r_{\text{HB}}$  is expressed in Å, and the hydrogen bond angle  $\phi_{\text{HB}}$  is expressed in degrees.

**Table 3.** Computed Hydrogen Bond Energies and Geometries of the Acetic Acid and Water Bimolecular Systems Using the B3LYP Level of Theory and the B3LYP/AMBER Model with the Lennard–Jones Parameters Evaluated from the B3LYP/6-31+G\*/AMBER Potential.

Molecule	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$	
		B3LYP/3-21G			B3LYP/3-21G/AMBER		
acd_wat_a1	25.12	1.58	98	7.48	1.75	97	
acd_wat_a2	27.34	1.47	180	9.32	1.67	180	
acd_wat_d1	27.34	1.70	105	9.32	1.99	105	
acd_wat_d2	14.76	1.83	108	5.14	1.95	113	
acd_wat_d3	11.19	1.82	180	4.17	1.85	180	
		B3LYP/6-31G			B3LYP/6-31G/AMBER		
acd_wat_a1	15.37	1.70	106	7.87	1.83	106	
acd_wat_a2	18.74	1.61	180	9.98	1.76	180	
acd_wat_d1	18.74	1.88	108	9.98	2.12	107	
acd_wat_d2	9.40	1.87	111	6.00	1.95	117	
acd_wat_d3	6.95	1.92	180	4.84	1.83	180	
		B3LYP/6-31+G*			B3LYP/6-31+G*/AMBER		
acd_wat_a1	9.09	1.80	101	8.09	1.89	102	
acd_wat_a2	10.63	1.82	180	10.19	1.85	180	
acd_wat_d1	10.63	1.97	111	10.19	2.21	107	
acd_wat_d2	6.23	1.92	114	6.02	2.04	116	
acd_wat_d3	3.88	2.02	180	4.03	2.06	180	

The hydrogen bond energy  $E_{\text{HB}}$  is expressed in kcal/mol, the hydrogen bond distance  $r_{\text{HB}}$  is expressed in Å, and the hydrogen bond angle  $\phi_{\text{HB}}$  is expressed in degrees.

list of bimolecular systems, obtained from the QM/MM calculations using the Lennard–Jones parameters evaluated in our study, with hydrogen bond energies and geometries obtained from the similar QM/MM calculations but using the original AMBER mo-

lecular mechanical parameters. The full list of computed hydrogen bond energies and geometries is included in Table 1 of supplementary materials. Figure 1 shows the 2D sketches of the simple organic molecules used in our calculations, and Figure 2 presents

**Table 4.** Computed Hydrogen Bond Energies and Geometries of the Acetic Acid and Water Bimolecular Systems Using the BLYP Level of Theory and the BLYP/AMBER Model with the Lennard–Jones Parameters Evaluated from the B3LYP/6-31+G\*/AMBER Potential.

Molecule	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$	
		BLYP/3-21G			BLYP/3-21G/AMBER		
acd_wat_a1	26.09	1.58	96	7.30	1.76	94	
acd_wat_a2	27.46	1.44	180	9.09	1.65	180	
acd_wat_d1	27.46	1.67	105	9.09	1.97	105	
acd_wat_d2	14.54	1.85	107	4.81	1.98	112	
acd_wat_d3	10.97	1.84	180	3.93	1.86	180	
		BLYP/6-31G			BLYP/6-31G/AMBER		
acd_wat_a1	18.09	1.70	103	7.64	1.83	102	
acd_wat_a2	20.90	1.60	180	9.69	1.74	180	
acd_wat_d1	20.90	1.86	107	9.69	2.11	107	
acd_wat_d2	11.50	1.89	110	5.64	1.97	116	
acd_wat_d3	9.10	1.94	180	4.53	1.90	180	
		BLYP/6-31+G*			BLYP/6-31+G*/AMBER		
acd_wat_a1	8.37	1.81	101	8.02	1.89	102	
acd_wat_a2	9.69	1.83	180	10.08	1.85	180	
acd_wat_d1	9.69	1.97	111	10.08	2.22	107	
acd_wat_d2	5.61	1.95	115	5.79	2.10	114	
acd_wat_d3	3.40	2.03	180	3.97	2.10	180	

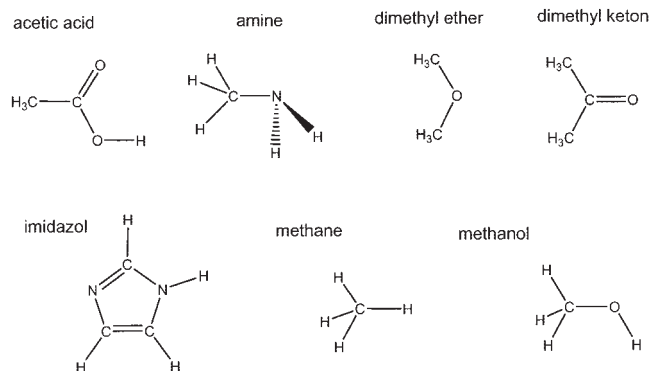
The hydrogen bond energy  $E_{\text{HB}}$  is expressed in kcal/mol, the hydrogen bond distance  $r_{\text{HB}}$  is expressed in Å, and the hydrogen bond angle  $\phi_{\text{HB}}$  is expressed in degrees.

**Table 5.** Comparison of the Hydrogen Bond Energies and Geometries of the Acetic Acid and Water Bimolecular Systems Obtained Using the B3LYP/6-31+G\* Level of Theory and the B3LYP/6-31+G\*/AMBER Model with the Lennard–Jones Parameters Evaluated from the QM/MM Potential and the MM Potential.

Molecule	B3LYP/6-31+G*/AMBER								
	B3LYP/6-31+G*			QM/MM parameters			MM parameters		
	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$	$E_{\text{HB}}$	$r_{\text{HB}}$	$\phi_{\text{HB}}$
acd_wat_a1	9.09	1.80	101	8.09	1.89	102	10.68	1.67	103
acd_wat_a2	10.63	1.82	180	10.19	1.85	180	13.64	1.63	180
acd_wat_d1	10.63	1.97	111	10.19	2.21	107	13.64	2.00	106
acd_wat_d2	6.23	1.92	114	6.02	2.04	116	7.68	1.78	116
acd_wat_d3	3.88	2.02	180	4.03	2.06	180	4.95	1.89	180

The hydrogen bond energy  $E_{\text{HB}}$  is expressed in kcal/mol, the hydrogen bond distance  $r_{\text{HB}}$  is expressed in Å, and the hydrogen bond angle  $\phi_{\text{HB}}$  is expressed in degrees.

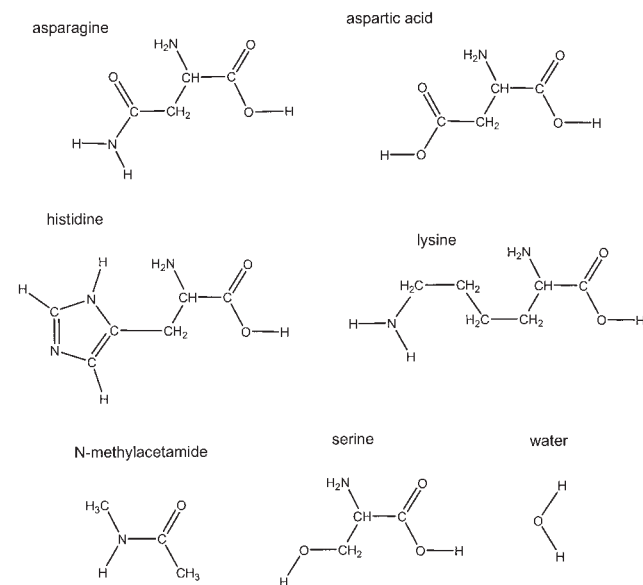
the 2D sketches of amino acids under our investigation. 2D sketches of all structures of bimolecular systems that were investigated in our study are shown in Figure 1 of supplementary materials. In tables and figures the abbreviations mth, mol, crb, eth, acd, amn, and imd refer to the QM hydrogen bond monomers namely: methane, methanol, dimethyl ketone, dimethyl ether, acetic acid, methyl amine, and imidazol, respectively. The abbreviations: asn, asp, his, lys, ser, nma, and wat refer to the MM hydrogen bond monomers, namely asparagine, aspartic acid, histidine, lysine, serine, N-methylacetamide, and water, respectively. For example, abbreviations (1) acd-asn-a1 and (5) acd-asn-d1 stand for bimolecular systems including acetic acid and asparagine, where the asparagine molecule is a hydrogen bond acceptor (denoted by a) and a hydrogen bond donor (denoted by d), respectively. Numbers 1 and 5 in brackets refer to a number of the structure from the total list of all 116 bimolecular systems. Figures 3, 4, and 5 show correlation diagrams between the results of the B3LYP/6-31+G\* method and the combined B3LYP/6-31+G\*/AMBER potential for the hydrogen bond energies, hydrogen bond distances, and hydrogen bond angles, respectively, using the Lennard–Jones parameters evaluated in our study. Figures 6, 7, and 8



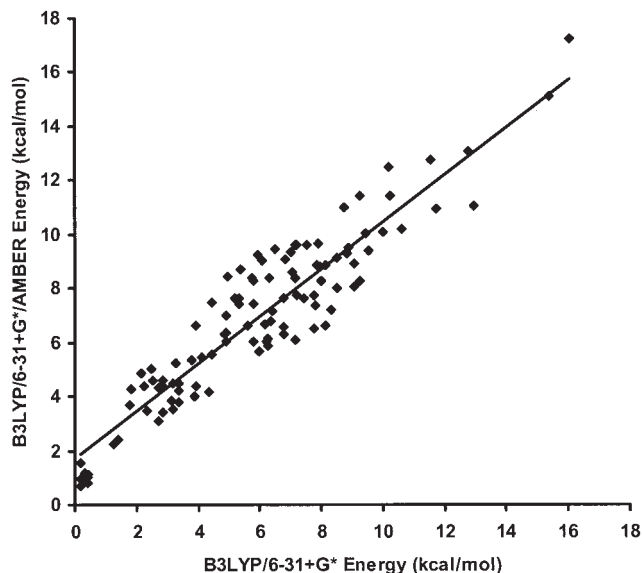
**Figure 1.** Simple organic molecules investigated in the present study using the combined B3LYP/6-31+G\*/AMBER method.

show similar correlation diagrams between the results of the B3LYP/6-31+G\* method and the combined B3LYP/6-31+G\*/AMBER potential, using the original molecular mechanical AMBER Lennard–Jones parameters.

As is evident in Table 1, the values of the  $\sigma$  Lennard–Jones parameter obtained in our investigation are slightly larger than corresponding values of this parameter found in the previous combined QM/MM studies, especially using the AM1/TIP3P approach. On the other hand, the  $\epsilon$  parameters obtained in the current study have similar values to those found previously. The similar relation is observed between the Lennard–Jones parameters evaluated in our present study at the DFT/AMBER level of theory, and the parameters obtained using the AM1/TIP3P and PM3/TIP3P

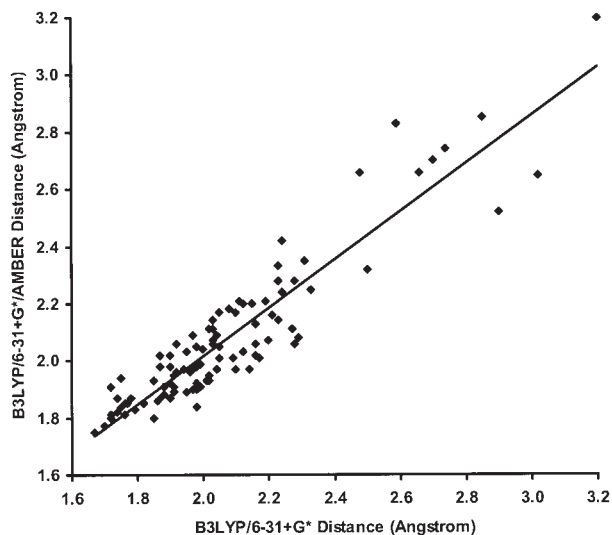


**Figure 2.** Amino acids investigated in the present study using the combined B3LYP/6-31+G\*/AMBER method.

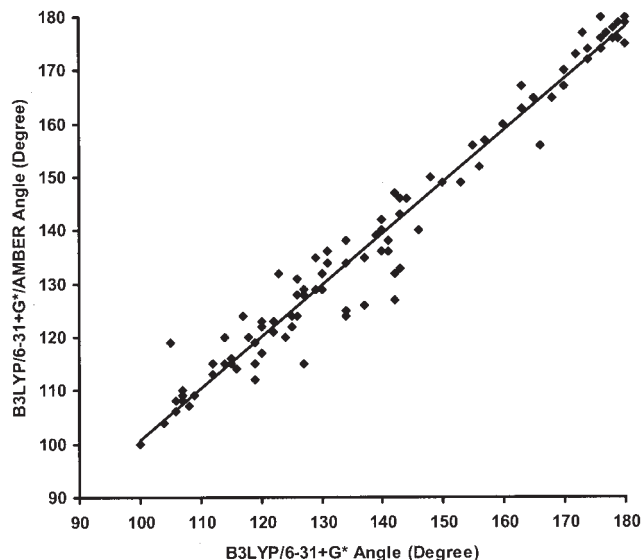


**Figure 3.** Comparison of the hydrogen bond energies obtained from the B3LYP/6-31+G\* method, and the combined B3LYP/6-31+G\*/AMBER method with the QM/MM Lennard–Jones parameters.

methods by others,<sup>42</sup> showing that the parameters that have been evaluated at the combined DFT/MM level of theory, are slightly larger the parameters obtained from the semiempirical QM/MM methods. We also observe systematic larger values of the Lennard–Jones parameters obtained at the B3LYP/AMBER level of theory in our present study than the parameters used in the CHARMM program utilizing the SCC-DFTB/MM approach.<sup>43</sup> It is also interesting to compare the values of  $\sigma$  and  $\epsilon$  determined in our study with the values of these parameters used in the molecular

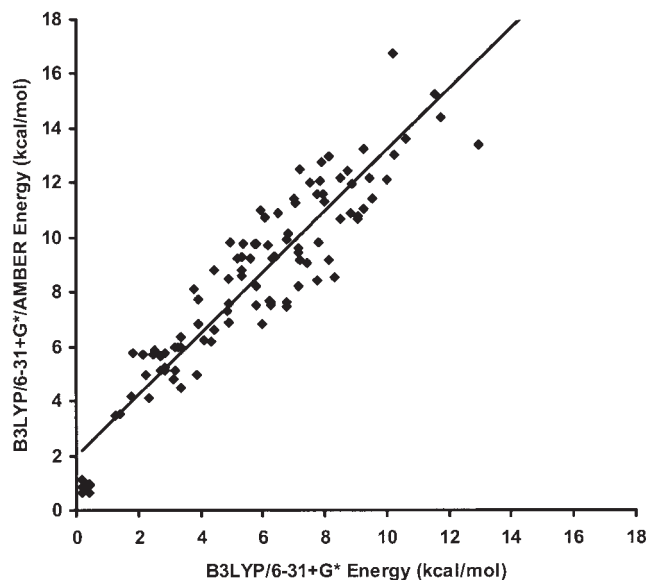


**Figure 4.** Comparison of the hydrogen bond distances obtained from the B3LYP/6-31+G\* method, and the combined B3LYP/6-31+G\*/AMBER method with the QM/MM Lennard–Jones parameters.

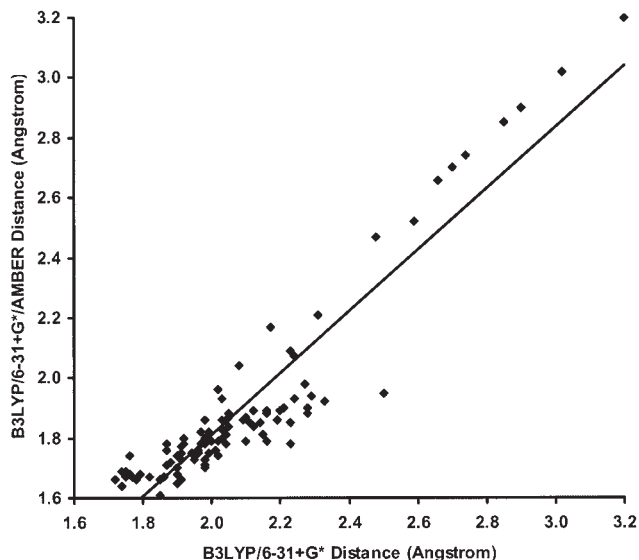


**Figure 5.** Comparison of the hydrogen bond angles obtained from the B3LYP/6-31+G\* method, and the combined B3LYP/6-31+G\*/AMBER method with the QM/MM Lennard–Jones parameters.

mechanical AMBER force field. In general, according to Table 1, the  $\sigma$  Lennard–Jones parameter obtained in our investigation, are about 10–20% larger then the corresponding values of this parameter used in the AMBER force field. The  $\epsilon$  parameters are similar, however. We also observe the analogous relation between parameters evaluated in our present study, and molecular mechanical parameters Tripos force field,<sup>44</sup> which have been used in the QM/MM calculations of palladium complexes.<sup>45</sup> The  $\sigma$  Lennard–



**Figure 6.** Comparison of the hydrogen bond energies obtained from the B3LYP/6-31+G\* method, and the combined B3LYP/6-31+G\*/AMBER method with the MM Lennard–Jones parameters.



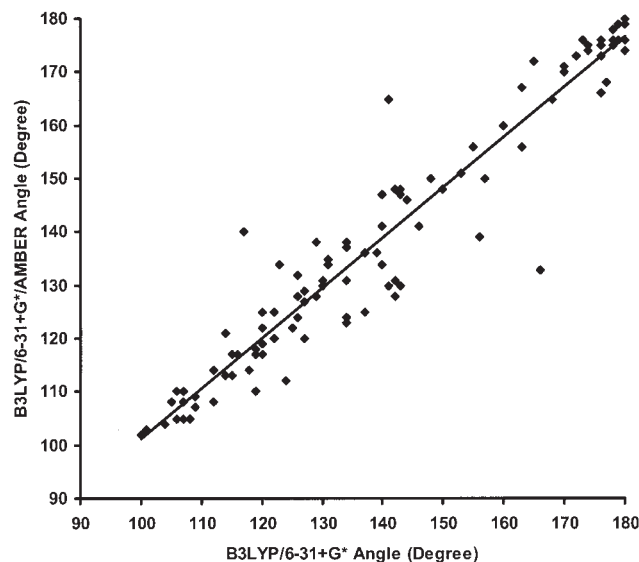
**Figure 7.** Comparison of the hydrogen bond distances obtained from the B3LYP/6-31+G\* method, and the combined B3LYP/6-31+G\*/AMBER method with the MM Lennard–Jones parameters.

Jones parameters obtained in our study are slightly larger than the corresponding values of the Tripos force-field parameters, keeping the similar value for the  $\epsilon$  parameters.

According to Figure 3, there is a good agreement between the hydrogen bond energies obtained at the combined B3LYP/6-31+G\*/AMBER level of theory and the results from the B3LYP/6-31+G\* method. In terms of the root-mean-square (RMS) error, there is a small deviation (1.5 kcal/mol) between the energy obtained from the combined QM/MM approach presented in this study and those from the full QM calculations. However, the combined QM/MM approach used in our investigation, tends to predict slightly higher hydrogen bond energies than the QM model. According to Figure 4, there is also a similar small deviation (0.11 Å) observed for hydrogen bond distances between the combined B3LYP/6-31+G\*/AMBER model and the QM calculations. However, the largest discrepancies between these two computational models are observed for a range of a hydrogen bond longer than 2.5 Å, where the interaction between two molecules has rather a long-range dispersion character than a hydrogen-bonding character. Figure 5 shows a very good agreement between hydrogen bond angles computed using the combined B3LYP/6-31+G\*/AMBER approach and the QM model. The deviation between hydrogen bond angles calculated using these two computational methods is only 4°. Generally, the B3LYP/6-31+G\*/AMBER approach predicts hydrogen bond parameters much closer to the QM results than the AM1/TIP3P and HF/3-21G/TIP3P models. However, the most significant improvement is observed for the hydrogen bond geometries (in particular, for the hydrogen bond angles). Because the goal of this work is to utilize the B3LYP/6-31+G\*/AMBER potential to study chemical processes occurring in a protein environment, the correct prediction of molecular geometries, which depends on a good-quality set of the Lennard–Jones parameters, is more important than accurate calcu-

lations of hydrogen bond energies. Hence, we anticipate that the B3LYP/6-31+G\*/AMBER potential will be a precise tool to explore intermolecular interactions inside a protein environment. The small deviations in hydrogen bond energies and hydrogen bond distances observed in our study, especially for a longer distance range, can be improved in our B3LYP/6-31+G\*/AMBER model by inclusion of electron polarization and diffuse functions on the QM hydrogen atoms. This subject has already been examined by Jorgensen et al., at the QM level of theory using the HF, MP2, and B3LYP approach and utilizing a variety of different basis sets for a large set of simple hydrogen bonds. Comparing the B3LYP method with the other approaches (HF and MP2), they have found a much better agreement with experiment for the calculations using a basis set including diffuse functions, especially using a basis set of a triple-zeta quality with polarization functions that is augmented by more than one diffuse function. They have also found the importance of the nonstandard augmentations,<sup>46</sup> where a basis set is augmented by an additional set of diffuse d polarization functions having an exponent one-fourth as large as that used in standard basis sets. This issue is a subject of our current investigations. The other possible source of the small deviations in hydrogen bond energies observed in our study can be related to the basis set superposition error (BSSE). However, according to Jorgensen et al., the BSSE calculated using the well-known counterpoise method is much smaller at the B3LYP level of theory than at the MP2 level for the same basis set. Therefore, we do not include the BSSE correction in our final reported binding energies.

It is also interesting to compare the results of the combined calculations using the B3LYP/6-31+G\*/AMBER potential with the Lennard–Jones parameters evaluated in our present study, with the results of similar combined calculations but using a different level of theory for the QM system. Table 2 shows the results of this comparison for a small set of bimolecular systems including acetic



**Figure 8.** Comparison of the hydrogen bond angles obtained from the B3LYP/6-31+G\* method, and the combined B3LYP/6-31+G\*/AMBER method with the MM Lennard–Jones parameters.

acid and water, using the HF level of theory and three basis sets: 3-21G, 6-31G, and 6-31+G\*. In Table 2 we report hydrogen bond energies, and geometries of these molecular complexes, and we compare the results of the QM/MM calculations with the results of the full QM calculations, which have been performed at the same QM level of theory. The QM/MM calculations have been performed using Lennard–Jones parameters evaluated in our study, which are presented in Table 1. Tables 3 and 4 present the similar comparison, where the B3LYP and BLYP level of the QM theory have been used with the same three basis sets, for both the full QM calculations and the combined QM/MM calculations. The best agreement between the results of the full QM calculations and the results obtained from the combined QM/MM calculations is observed for the B3LYP/6-31+G\* level of theory as it is presented in Table 3, which is not surprising because the Lennard–Jones parameters used in the combined calculations have been evaluated for this level of the QM theory. According to Table 3, we observe that hydrogen bond energies become slightly smaller, and hydrogen bond distances become slightly shorter, when we are using a smaller basis set in the combined B3LYP/AMBER calculations, which indicates sensitivity of the presented QM/MM method on the basis set used in the calculations. It is interesting to mention, that according to Table 3, the results of hydrogen bond energies and geometries obtained using the combined B3LYP/AMBER method with a smaller basis set such as the basis 3-21G or 6-31G are even better than the results of the calculation performed at the full QM level of theory using the same smaller basis set. This conclusion indicates that the combined B3LYP/AMBER method with a relatively small basis set reproduces better long-range interactions between monomers of hydrogen bonds, than the full QM level of theory using the same basis set. The similar trend is observed for the HF and BLYP methods, as it is shown in Tables 2 and 3, respectively. Comparing the results of the B3LYP/AMBER calculations using the 6-31+G\* basis set shown in Table 3, with the results of the HF/AMBER and BLYP/AMBER methods using with the same basis set shown in Tables 2 and 3, respectively, we observe small changes in the hydrogen bond energies and geometries. These differences are even smaller when comparing the results of the QM/AMBER calculations between two methods using different levels of DFT theory (B3LYP/AMBER shown in Table 3, and BLYP/AMBER shown in Table 4). This indicates that the presented QM/AMBER approach is less sensitive on the correlation method used in the calculations such as B3LYP or BLYP, and can be accurately used in combined QM/AMBER calculations with the 6-31+G\* basis set, utilizing a different level of theory than B3LYP. However, the presented combined B3LYP/AMBER method is slightly more sensitive on the choice of the basis set used in the calculations; therefore, a new evaluation of Lennard–Jones parameters is needed for precise calculations with a basis set different than the basis of the 6-31+G\* quality. It is also interesting to compare the results of the combined QM/MM calculations using the B3LYP/6-31+G\*/AMBER level of theory and the Lennard–Jones parameters obtained in our present study, with the results of the calculations using the same level of the QM/MM theory but utilizing the original molecular mechanical (MM) parameters. Table 5 presents this comparison, for a small set of the bimolecular systems, including acetic acid and water, similar as in our previous test calculations. According to Table 5, hydrogen

bond energies obtained from the combined calculations using the original AMBER Lennard–Jones parameters are much bigger, and hydrogen bond distances are much smaller, than energies and distances obtained from the calculations using the QM/MM Lennard–Jones parameters. This observation is consistent with Table 1, which presents smaller values of the MM parameters than the corresponding QM/MM parameters. The full list of the results obtained from the combined calculations using the MM parameters and the QM/MM parameters is included in Table 1 of supplementary materials. Figures 6, 7, and 8 demonstrate a correlation between hydrogen bond energies, hydrogen bond distances, and hydrogen bond angles obtained from the B3LYP/6-31+G\* method and from the B3LYP/6-31+G\*/AMBER potential, where MM Lennard–Jones parameters have been used. According to Figures 6 and 7, hydrogen bond energies obtained from the QM/MM calculations are systematically too big and hydrogen bond distances are systematically too short in comparison with the results from the full QM calculations, which confirms the general conclusion that MM parameters are too small in use of the combined QM/MM method. Moreover, according to Figure 8, the QM/MM calculations using the original MM parameters reproduce hydrogen bond angles much less accurate (with an RMS error of almost 7°) than the QM/MM method with the QM/MM Lennard–Jones parameters evaluated in our present study (with an RMS error within 4°).

## Conclusion

A combined B3LYP/6-31+G\*/AMBER potential for use in combined QM/MM molecular modeling of chemical processes in a protein environment is presented. In this approach, the active site of a large biological system is treated at the B3LYP/6-31+G\* level of theory, whereas the remaining protein environment is represented by the AMBER force field. The interaction between the QM and the MM regions consists of an electrostatic perturbation of the QM wave function by external MM point charges, and a van der Waals interaction between QM and MM atoms, which is expressed in terms of the 6-12 Lennard–Jones potential. The Lennard–Jones potential depends on two empirical parameters,  $\sigma$  and  $\epsilon$ , the values of which depend on the level of the QM theory employed in the combined approach. In our current investigation, we have evaluated the  $\sigma$  and  $\epsilon$  parameters for the B3LYP/6-31+G\*/AMBER potential to reproduce hydrogen bond energies and geometries of a large set of bimolecular systems, representing hypothetical interactions between the QM active site of a protein and a MM protein environment. Hydrogen bond energies and geometries obtained in our combined B3LYP/6-31+G\*/AMBER approach have been compared with results from the calculations using the B3LYP/6-31+G\* level of theory, and we have found a very good agreement between these two computational methods. The RMS error in hydrogen bond energies was 1.5 kcal/mol, whereas hydrogen bond distances have been reproduced within an rms error 0.11 Å, and hydrogen bond angles within an RMS error of 4°. The agreement between the combined QM/MM potential and the full QM calculations presented in our study is better than that for QM/MM potentials published previously. Therefore, we anticipate that the B3LYP/6-31+G\*/AMBER potential will be-

come a useful tool for better describing chemical processes occurring at the active sites of proteins.

## Acknowledgments

We would like to thank Prof. Harry F. King for critical reading this manuscript and also for his helpful comments related to the subject investigated in this work. We also would like to thank Shawn T. Brown and Prakashan P. Korambath for their help using the Q-Chem program. This work was performed in part at the Center for Computational Research at the University at Buffalo, the State University of New York.

## References

1. Warshel, A.; Levitt, M. *J Mol Biol* 1976, 103, 227.
2. Singh, U. C.; Kollman, P. A. *J Comput Chem* 1986, 7, 718.
3. Bash, P. A.; Field, M. J.; Karplus, M. *J Am Chem Soc* 1987, 109, 8092.
4. Gao, J.; Xia, X. *Science* 1992, 258, 631.
5. Luzhkov, V.; Warshel, A. *J Comput Chem* 1992, 13, 199.
6. Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J Phys Chem* 1993, 97, 11868.
7. Maseras, F.; Morokuma, K. *J Comput Chem* 1995, 16, 1170.
8. Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH: New York, 1995, p. 119, vol. 7.
9. Stanton, R. V.; Hartsough, D. S.; Merz, K. M., Jr. *J Phys Chem* 1993, 97, 11868; Hartsough, D. S.; Merz, K. M., Jr. *J Phys Chem* 1995, 99, 384; Hartsough, D. S.; Merz, K. M., Jr. *J Phys Chem* 1995, 99, 11266.
10. Chatfield, D. C.; Brooks, B. R. *J Am Chem Soc* 1995, 117, 5561.
11. Tunon, I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-Lopez, M. F.; Rivail, J. L. *J Comp Chem* 1996, 17, 19.
12. Bakowies, D.; Thiel, W. *J Phys Chem* 1996, 100, 10580; Bakowies, D.; Thiel, W. *J Comput Chem* 1996, 17, 87.
13. Gao, J. *Acc Chem Res* 1996, 29, 298.
14. Assfeld, X.; Rivail, J. L. *Chem Phys Lett* 1996, 263, 100.
15. Liu, H.; Muller-Plathe, F.; van Gunsteren, W. F. *J Mol Biol* 1996, 261, 454.
16. Gao, J.; Thomson, M., Eds. *ACS Symposium: Combined Quantum Mechanical and Molecular Mechanical Methods*; American Chemical Society: Washington, DC, 1998, p. 712.
17. Alhambra, C.; Wu, L.; Zhang, Z.-Y.; Gao, J. *J Am Chem Soc* 1998, 120, 3858.
18. Assfeld, X.; Ferre, N.; Rivail, J. L. *ACS Symposium: Combined Quantum Mechanical and Molecular Mechanical Methods*; American Chemical Society: Washington, DC, 1998, p. 234.
19. Philipp, D. M.; Friesner, R. A. *J Comp Chem* 1999, 20, 1468.
20. Amara, P.; Field, M. J. *Computational Molecular Biology*; Leszczynski, J., Ed.; Elsevier: Amsterdam, 1999, p. 1.
21. Alhambra, C.; Gao, J.; Corchado, J. C.; Villa, J.; Truhlar, D. G. *J Am Chem Soc* 1999, 121, 2253.
22. Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J Comput Chem* 2000, 21, 1442.
23. Mo, Y.; Gao, J. *J Comput Chem* 2000, 21, 1458.
24. Schoneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Oglaro, F.; Shaik, S. *J Am Chem Soc* 2002, 124, 8142.
25. Rajamani, R.; Gao, J. *J Comput Chem* 2002, 23, 96.
26. Guallar, V.; Baik, M. H.; Lippard, S. J.; Friesner, R. A. *Proc Natl Acad Sci USA* 2003, 100, 6998.
27. Freindorf, M.; Kozlowski, P. M. *J Am Chem Soc* 2004, 126, 1928.
28. Schoneboom, J. C.; Cohen, S.; Lin, H.; Shaik, S.; Thiel, W. *J Am Chem Soc* 2004, 126, 4017.
29. Freindorf, M.; Gao, J. *J Comp Chem* 1996, 17, 386.
30. Ferre, N.; Assfeld, X.; Rivail, J. L. *J Comput Chem* 2002, 23, 610.
31. Gao, J. In *Modeling the Hydrogen Bond*; ACS Symp Ser; American Chemical Society, Washington, DC, 1994, p. 8.
32. Becke, A. D. *J Chem Phys* 1993, 98, 1372.
33. Kozlowski, P. M. *Curr Opin Chem Biol* 2001, 5, 736.
34. Sim, F.; St. Amant, A.; Papai, I.; Salahub, D. R. *J Am Chem Soc* 1992, 114, 4391.
35. Rablen, P. R.; Lockman, J. W.; Jorgensen, W. L. *J Phys Chem A* 1998, 102, 3782.
36. Long, X. P.; Nicholas, J. B.; Guest, M. F.; Ornstein, R. L. *J Mol Struct* 1997, 412, 121.
37. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 6*; University of California: San Francisco, 1999.
38. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J Am Chem Soc* 1995, 117, 5179.
39. Jorgensen, W. L.; Tirado-Rives, J. *J Am Chem Soc* 1988, 110, 1657.
40. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J Chem Phys* 1983, 79, 926.
41. Kong, J.; White, C. A.; Krylov, A. I.; Sherrill, C. D.; Adamson, R. D.; Furlani, T. R.; Lee, M. S.; Lee, A. M.; Gwaltney, S. R.; Adams, T. R.; Ochsenfeld, C.; Gilbert, A. T. B.; Kedziora, G. S.; Rassolov, V. A.; Maurice, D. R.; Nair, N.; Shao, Y.; Besley, N. A.; Maslen, P. E.; Dombroski, J. P.; Daschel, H.; Zhang, W.; Korambath, P. P.; Baker, J.; Byrd, E. F. C.; Van Voorhis, T.; Oumi, M.; Hirata, S.; Hsu, C.-P.; Ishikawa, N.; Florian, J.; Warshel, A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M.; Pople, J. A. *J Comput Chem* 2000, 21, 1532.
42. Luque, F. J.; Reuter, N.; Cartier, A.; Ruiz-Lopez, M. F. *J Phys Chem A* 2000, 104, 10923.
43. Riccardi, D.; Li, G.; Cui, Q. *J Phys Chem B* 2004, 108, 6467.
44. Clark, M.; Cramer, R. D., III; van Opdenbosch, N. *J Comput Chem* 1989, 10, 982.
45. Magistrato, A.; Pregosin, P. S.; Albinati, A.; Rothlisberger, U. *Organometallics* 2001, 20, 4178.
46. Ochterski, J. W. Ph.D. Thesis, Wesleyan University; 1994.