JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Application of Adaptive QM/MM Methods to Molecular Dynamics Simulations of Aqueous Systems

Kyoyeon Park,[†] Andreas W. Götz,[‡] Ross C. Walker,[†,‡] and Francesco Paesani*[,†]

[†]Department of Chemistry and Biochemistry and [‡]San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, United States

ABSTRACT: The difference-based adaptive solvation quantum mechanics/molecular mechanics (adQM/MM) method (*J. Chem. Theory Comput.* **2009**, *5*, 2212) as implemented in the Amber software was applied to the study of several chemical processes in solution. The adQM/MM method is based on an efficient selection scheme that enables quantum-mechanical treatment of the active region of a molecular system in solution taking explicitly into account diffusion of solvent molecules between the QM and the MM regions. Specifically, adQM/MM molecular dynamics simulations are carried out to characterize (1) the free energy profiles of halide exchange S_N2 reactions in water, (2) the hydration structure of the Cl^- ion, and (3) the solvation structure of the zwitterionic form of glycine in water. A comparison is made with the results obtained using standard MM and QM/MM methods as well as with the available fully QM and experimental data. In all cases, it is shown that the adaptive QM/MM simulations provide a physically realistic description of the system of interest.

1. INTRODUCTION

Hybrid quantum mechanics/molecular mechanics (QM/MM) methods are among the most successful approaches used to model complex condensed-phase systems whose topology varies over time due to changes in the covalent bonding.[1−9] Within the QM/MM schemes only a small portion of the whole system, including the chemically most relevant region (e.g., the active site of an enzyme[1,6]), is treated quantum mechanically. The remaining part of the system, which effectively accounts for the environmental effects that exist in the condensed phase, is treated at the classical level using molecular mechanics force fields. In standard QM/MM methods, the atoms belonging to the QM and MM regions are determined at the beginning of the simulation from considerations based on the topology of the system and nature of the chemical process of interest. The composition of the two regions is then kept fixed throughout the simulation. As a result, standard QM/MM approaches can accurately describe chemical reactions in systems where the QM region can be uniquely identified as a relatively small part of a larger substrate or when the molecules initially included in the QM (MM) region do not diffuse into the MM (QM) region. By construction, therefore, standard QM/MM approaches suffer from severe limitations when applied to chemical processes in solution where, due to molecular diffusion, a continuous exchange of solvent molecules occurs between the QM and the MM regions. This becomes particularly relevant in studies of reactive and solvation phenomena in which explicit inclusion of purely quantum-mechanical electronic effects, such as polarization and charge transfer, is crucial for accurate representation of the overall process.

To overcome the limitations of standard QM/MM methods, several approaches have recently been developed which allow the molecules crossing the boundaries between the QM and the MM regions to dynamically change their "character" (from QM to MM and vice versa).[10−16] In these approaches, different adaptive selection schemes are used to determine which molecules belong to the QM and MM regions at any step of the simulation. This is generally accomplished either by adopting geometric criteria, such as a distance cutoff between the position of each molecule and the center of the QM region, or by predefining the total number of molecules that are allowed in the QM region at any step. However, since molecular diffusion across the QM/MM boundary is also accompanied by a change of the molecule "character", the corresponding dynamics becomes discontinuous. Several methods have been proposed for removing this discontinuity, including the difference-based adaptive solvation (DAS),[10,11] sorted adaptive partitioning (SAP),[14] and adaptive buffer methods.[15,16] All these methods rely on the definition of multiple QM/MM partitions that are then combined to determine the actual system dynamics using appropriate interpolation algorithms.

In this study, we describe the application of the adaptive QM/MM method with the DAS force interpolation (hereafter referred to as the adQM/MM method),[10] which we recently implemented[17] in the AMBER software,[18] to several representative molecular systems in solution and provide a detailed comparison with the results obtained using standard MM and QM/MM approaches. Specifically, we report on the results of adQM/MM simulations for (1) halide exchange S_N2 reactions in water, (2) Cl^- hydration, and (3) solvation structure of the zwitterionic form of glycine in water. This article is organized as follows. In section 2, a brief overview of the adQM/MM method is provided along with the description of the computational details specific to our studies. Results of the adQM/MM simulations are discussed in section 3, and conclusions are given in section 4.

Received: April 27, 2012
Published: June 21, 2012

2. METHODOLOGY

2.1. Difference-Based Adaptive Solvation Scheme. Within the DAS scheme the molecular system is divided into three regions, namely, the active (A), transition (T), and environment (E) regions,[10] which are schematically represented in Figure 1. The A region contains the part of the system

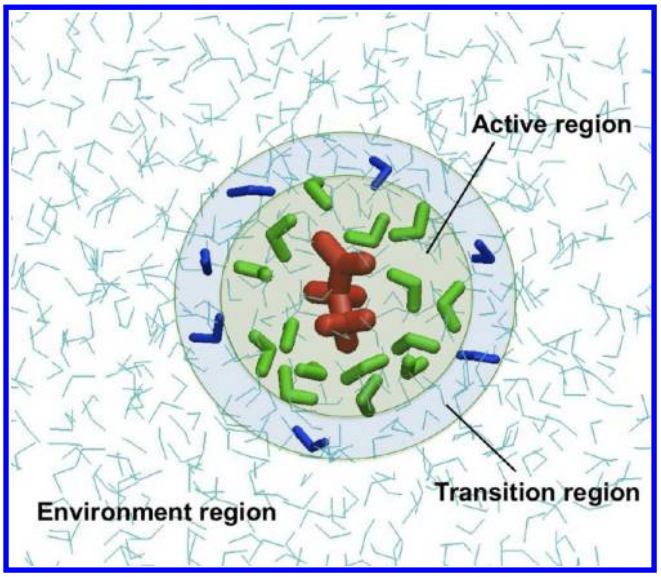

Figure 1. Schematic representation of the adQM/MM method. System is divided into 3 regions: active (A), transition (T), and environment (E) regions. Active region is treated quantum mechanically, while the environment region is treated at the MM level. Transition region T is located between the active and the environment regions. QM center is located at the center of the active region. See main text for details.

that is chemically relevant for the molecular process of interest and, therefore, is treated quantum mechanically throughout the simulation. The part of the system that does not directly affect the process of interest but is still important for accurate description of the overall environmental effects is instead defined as the E region. The E region is treated at the molecular-mechanics level throughout the simulation. The T region corresponds to the transition region between A and E in which the molecules are allowed to change their character from a purely QM to a purely MM description and vice versa. As a consequence, the molecules located in the T region effectively display both QM and MM "characters" that are smoothly combined according to an appropriate interpolation algorithm.[10]

In the adQM/MM method, multiple partitions of the system are required to guarantee that the underlying molecular dynamics remains continuous throughout the simulation. Each partition is created by systematically assigning either the QM or the MM character to the molecules located in the transition region. Standard QM/MM calculations are then performed independently for all partitions. Finally, the actual forces required to propagate the system dynamics are obtained as weighted combinations of the forces calculated for each QM/MM partition. In principle, 2^N different partitions can be created for a system containing N molecules in the T region. However, it can be shown that, by construction, many of these partitions do not contribute.[10] Briefly, the adQM/MM scheme assigns a partial MM character, λ, to each molecule in the system. The specific value of λ depends on the distance (r) of the molecule from the center of the A region according to[10]

$$\lambda(r) = \begin{cases} 0 & \text{if } r < R_1 \\ \dfrac{(r - R_1)^2(3R_2 - R_1 - 2r)}{(R_2 - R_1)^3} & \text{if } R_1 \le r \le R_2 \\ 1 & \text{if } r > R_2 \end{cases} \quad (1)$$

where R_1 and R_2 are the inner and outer radii delimiting the T region. On the basis of this definition of λ, each QM/MM partition is then assigned a statistical weight, σ_i, given by

$$\sigma = \begin{cases} 0 & \text{if } \max(\{\lambda\}^{QM}) > \min(\{\lambda\}^{MM}) \\ \min(\{\lambda\}^{MM}) - \max(\{\lambda\}^{QM}) & \text{if } \max(\{\lambda\}^{QM}) \le \min(\{\lambda\}^{MM}) \end{cases} \quad (2)$$

where $\{\lambda\}^{QM}$ and $\{\lambda\}^{MM}$ indicate the sets of λ values assigned to the QM and MM regions, respectively. From eq 2 it follows that, given a specific molecular configuration, the statistical weight of a partition is zero when the partition contains one MM molecule closer to the center of the A region than any of the QM molecules. As a result, the total number of possible QM/MM partitions with nonzero weights in an adQM/MM simulation is $N + 1$. Importantly, the particular definition of σ in eq 2 also guarantees that the weight of each partition varies smoothly from 0 to 1. These particular features remove the discontinuities in the system dynamics that appear in standard QM/MM simulations anytime a molecule changes its character by diffusing in/out of the QM region.

Two different adQM/MM schemes have been proposed to define the boundaries between the A, the T, and the E regions in terms of distance-based[10] or number-based[11] criteria. According to the distance-based criterion, both the A and the T regions have fixed volumes but the number of molecules inside each region can vary during the simulation. By contrast, in adQM/MM simulations using the number-based criterion, the number of molecules in the A and T regions is fixed but the size of both regions can vary. An efficient parallel version of the adQM/MM method,[17] which employs both the distance-based and the number-based criteria, has been implemented by us in the Amber suite of codes for molecular dynamics simulations.[18] The specific details of our implementation are described in a separate publication.[17]

2.2. Computational Details. Depending on the system, different electronic structure methods were used to calculate the forces acting on the atoms located in the QM region. The general Amber force field (GAFF)[19] and the flexible SPC/Fw water model[20] were instead used to describe the MM water−solute and water−water interactions. The Lennard−Jones parameters for the halide ions were taken from ref 21. In all cases, classical molecular dynamics (MD) simulations in periodic boundary conditions were initially carried out in the isothermal−isobaric (NPT) ensemble to equilibrate each system at temperature T = 300 K and pressure P = 1 bar. The temperature was controlled using a Langevin thermostat[22] with a collision frequency of 5 ps^{-1}, while a Berendsen barostat[23] with a relaxation time of 5 ps was used to control the pressure. Starting from the equilibrated MD configurations, additional short NPT simulations (~50 ps) were carried out

with the adQM/MM method using the same parameters for the Langevin thermostat and Berendsen barostat as in the initial MD simulations. The cutoff distance for all nonbonded interactions, including the short-range electrostatic interactions between the QM and the MM atoms, was set to 9 Å. The center of the A region was chosen as the center of mass of the reactive complex $X\cdots CH_3\cdots X$ (with X = F, Cl, and Br) and the position of the Cl^- ion in the adQM/MM simulations of the halide exchange S_N2 reactions and the chloride ion hydration, respectively. In the adQM/MM simulations of glycine in water the center of the A region was chosen as the middle point between the alpha and the carboxyl carbon atoms to ensure that both the $-NH_3^+$ and the $-COO^-$ groups were solvated by QM water molecules. The relative distance between the center of the A region and the position of the oxygen atoms of the water molecules was used to determine the value of λ in eq 1 and, consequently, to define the A, T, and E regions.

QM calculations in the adQM/MM simulations were carried out at both the semiempirical and the density functional theory (DFT) levels of theory using, respectively, the Sander/SQM[24] and Sander/TeraChem interfaces[25−29] available in Amber. The electrostatic embedding (EE) scheme, which is based on a QM/MM compatible version of the particle mesh Ewald (PME) approach,[24] was adopted in simulations with the Sander/SQM interface to describe the electrostatic interactions between the QM and the MM regions with periodic boundary conditions. The mechanical embedding (ME) scheme based on 2-level ONIOM calculations[30] was instead used in the simulations with the Sander/TeraChem interface since the latter currently does not support the PME approach.

3. RESULTS

3.1. Halide Exchange S_N2 Reactions. As a first application of the adQM/MM method, we studied the halide exchange S_N2 reactions. Specifically, the free energy profiles for the reaction $X'^- + CH_3X \rightarrow CH_3X' + X^-$ with X = F, Cl, and Br were calculated in both gas and aqueous phases and compared with the results obtained using standard QM/MM simulations. In all cases, the PM3-PDDG semiempirical Hamiltonian[31,32] was used for the QM region since it gives accurate potential energy barriers for halide exchange S_N2 reactions in the gas phase with values that are comparable with high-level ab initio calculations.[33] For each S_N2 reaction, the potential of mean force (PMF) was calculated in the (a) gas phase at the full QM level, (b) aqueous phase using the standard QM/MM method with all water molecules being described by the SPC/Fw model, and (c) aqueous phase using the adQM/MM method. In the latter case, 10 (for X = F) and 15 (for X = Cl and Br) water molecules were included in the A region while 7 water molecules were assigned to the T region. All simulations in the aqueous phase were performed in the NPT ensemble at temperature T = 300 K and pressure P = 1 bar as described in section 2.2. Periodic cubic boxes containing 1690, 1419, and 1803 water molecules were used for simulations with X = F, Cl, and Br, respectively. The Lennard−Jones parameters for the neutral and ionic forms of the X atoms are listed in Table 1. Gas-phase simulations were carried out at T = 300 K. In all cases, the PMF curves were calculated using the umbrella sampling method along a reaction coordinate (RC) defined as the difference of the distances (d) between the carbon and the two halide atoms, i.e., RC = d(C−X) − d(C−X′). In total, 31 evenly spaced umbrella windows were used for each PMF calculation with 0.0 Å ≤ RC ≤ 3.0 Å.

Table 1. Lennard−Jones Parameters Employed in the adQM/MM Simulations of the S_N2 Reactions Described in Section 2[a]

X	$R_{min}/2$ (Å)	ε (kcal/mol)
F	1.750	0.061
F^-	2.257	0.0074005
Cl	1.948	0.265
Cl^-	2.711	0.0127859
Br	2.020	0.420
Br^-	2.751	0.0269586

[a]Parameters for the halogen atoms are from ref 19, while those for the halide atoms are from ref 21.

Each umbrella window was equilibrated independently for 100 ps using standard QM/MM simulations. For each umbrella window, biased MD simulations were then performed for 100 ps using restraining harmonic potentials applied to RC. The harmonic force constant was set to k = 400, 200, and 150 kcal/mol for X = F, Cl, and Br, respectively. For all systems, the WHAM algorithm was used to reconstruct the actual PMF along RC from the biased simulation data and standard block averaging techniques were used to estimate the statistical uncertainty.[34] The results are summarized in Table 2 and Figure 2.

Table 2. Potential and Free Energy Barrier Heights for $X'^- + CH_3X \rightarrow X'CH_3 + X^-$ S_N2 Reactions in the Gas and Aqueous Phase[a]

	potential energy barrier		free energy barrier		
	gas phase		gas phase	in water	
X	wo/ZPE corr.	w/ZPE corr.		standard QM/MM[c]	adQM/MM[d]
F	43.8	43.3	57.6 ± 0.2	62.5 ± 0.5	61.6 ± 0.9
Cl	15.7	14.5[e]	19.3 ± 0.9	26.4 ± 0.2	27.8 ± 0.9
Br	9.8	8.2	9.6 ± 0.2	17.7 ± 0.5	20.0 ± 0.1

[a]The experimental value for the reaction with X = Cl is 26.6 kcal/mol.[72] Energies are in kcal/mol. Barrier heights were obtained from the difference of the highest and lowest energies along the PMF curves. [b]ZPE was calculated from the harmonic frequencies using the PM3-PDDG semiempirical Hamiltonian. [c]Only the reactive system $X'^- + CH_3X$ is treated at the QM level. [d]Number of water molecules in the A region was 10 and 15 for X = F and X = Cl and Br, while in all cases, 7 water molecules were included in the T region. [e]Barrier height is 13.09 kcal/mol at the CBS-QB3(+) level of theory including zero-point energy correction.[33]

Our QM/MM simulations reproduce closely the free-energy profiles obtained in previous studies of halide exchange S_N2 reactions in both gas and aqueous phases.[35,36] In the gas phase, the backside attack of the nucleophilic halide ion leads to pronounced free-energy minima between 0.9 (X = Br) and 1.4 Å (X = F) due to strong ion−dipole interactions. In solution, this effect is compensated by the energy required for partial desolvation of the halide ions. Importantly, due to more diffuse charge distributions, the transition states in solution are characterized by weaker ion−dipole interactions with the surrounding water molecules than the corresponding pre-reaction complexes. This leads to relatively higher free-energy barriers in the aqueous phase than in the gas phase.

The comparison between the standard QM/MM and the adQM/MM results for the smaller and almost nonpolarizable

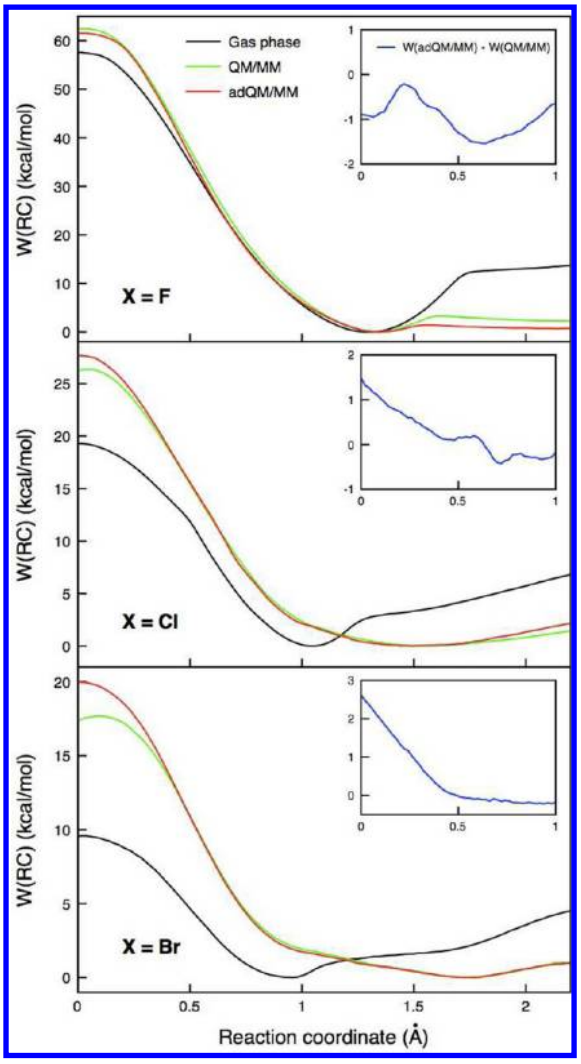

Figure 2. PMF curves calculated along the reaction coordinate (RC) of the S_N2 reactions $X'^- + CH_3X \rightarrow X'CH_3 + X^-$. Reaction coordinate is defined as the difference between the C–X and the C–X′ distances. Transition state is located at RC = 0 Å. Number of QM solvent molecules included in the adQM/MM calculations is 10 (A region) and 7 (T region) for X = F and 15 (A region) and 7 (T region) for X = Cl and Br.

F^- ion indicates that both methods predict essentially identical barrier heights. However, appreciable differences between the two methods exist for S_N2 reactions with X = Cl and Br. In these cases, the adQM/MM simulations predict barrier heights that are 1.3 (for X = Cl) and 2.3 kcal/mol (for X = Br) higher than the corresponding results obtained from standard QM/MM simulations. Interestingly, contrary to the results obtained with the adQM/MM method, the PMF curves from standard QM/MM simulations do not reach their maxima at the transition state (RC = 0). The difference between the QM/MM and the adQM/MM results (blue curves in Figure 2) can be explained by considering the artificial solvation effects that appear near the transition state in standard QM/MM simulations. In this case, the QM region includes only the reactive system, $X'^- + CH_3X$, and consequently, quantum-mechanical electronic effects such as charge transfer to the surrounding water molecules and solvent polarization are not allowed. As a result, near the transition state the negative charge remains artificially localized on the reactive complex, which leads to stronger ion–dipole interactions with the surrounding water molecules. This unphysical effect is thus partially responsible for the relatively lower free energy calculated at the transition state using standard QM/MM simulations. By contrast, the adQM/MM method, which treats the water molecules close to the reactive complex at the quantum-mechanical level, correctly takes into account charge transfer processes to the solvent. This leads to a reduction of charge localization on the reaction complex and, consequently, to weaker ion–dipole interactions with the surrounding water molecules.

A second factor that contributes to the differences between the standard QM/MM and the adQM/MM methods is directly related to the nature of the interactions between the reactive complex and the surrounding water molecules. At the transition state, the two X atoms are located at the same distance from the carbon atom of CH_3 forming two equivalent mixed covalent–ionic bonds. This symmetric configuration of the reactive complex implies that the hydration structure of the two X atoms should be, on average, identical. However, this is not the case in standard QM/MM simulations as illustrated in Figure 3. This figure shows the radial distribution functions (RDFs) describing the spatial correlation between the two Br atoms and the oxygen (O_w) and hydrogen (H_w) atoms of the water molecules calculated at the transition state using adQM/MM (left panels) and standard QM/MM (right panels) simulations. Similar results (not reported) were also obtained for X = F and Cl. The two sets of results clearly indicate that while the adQM/MM method correctly reproduces the symmetric hydration structure around the reactive complex at the transition state, the RDFs obtained from standard QM/MM simulations for the two Br atoms are significantly different. This difference can be explained by considering that the QM and MM atoms also interact through pairwise classical Lennard–Jones potentials whose parameters depend on the chemical nature of each atom pair. Therefore, the X and X′ atoms are characterized by different Lennard–Jones parameters that represent their different chemical nature at the beginning of the simulation (representing a covalently bound bromine atom and a bromide ion, respectively). For example, the Lennard–Jones parameters used for simulations where X = Br were $R_{min}/2 = 2.020$ Å and $\varepsilon = 0.420$ kcal/mol for the covalently bound Br atom[19] and $R_{min}/2 = 2.751$ Å and $\varepsilon = 0.027$ kcal/mol for the Br^- ion.[21] The difference in the LJ parameters becomes problematic for standard QM/MM representations of the $X'^- + CH_3X$ systems in which the X atoms are at the interface between QM and MM regions.[8] Although this description is valid at relatively large separations, it becomes increasingly less accurate near the transition state where both the X and X′ atoms have mixed covalent–ionic characters. This mixed character cannot be correctly reproduced in standard QM/MM simulations that, by construction, use the same Lennard–Jones parameters for the X and X′ atoms independently of the location along the reaction coordinate. By contrast, since the interactions between the reactive complex and the first surrounding water molecules are treated at the quantum-mechanical level, the adQM/MM method is capable of reproducing the correct physics along the entire reaction coordinate.

3.2. Cl^- Hydration. Analysis of the hydration properties of the Cl^- ion was carried out using different adQM/MM simulation setups to determine the dependence of the results on (1) the size of the A and T regions, (2) the choice of the

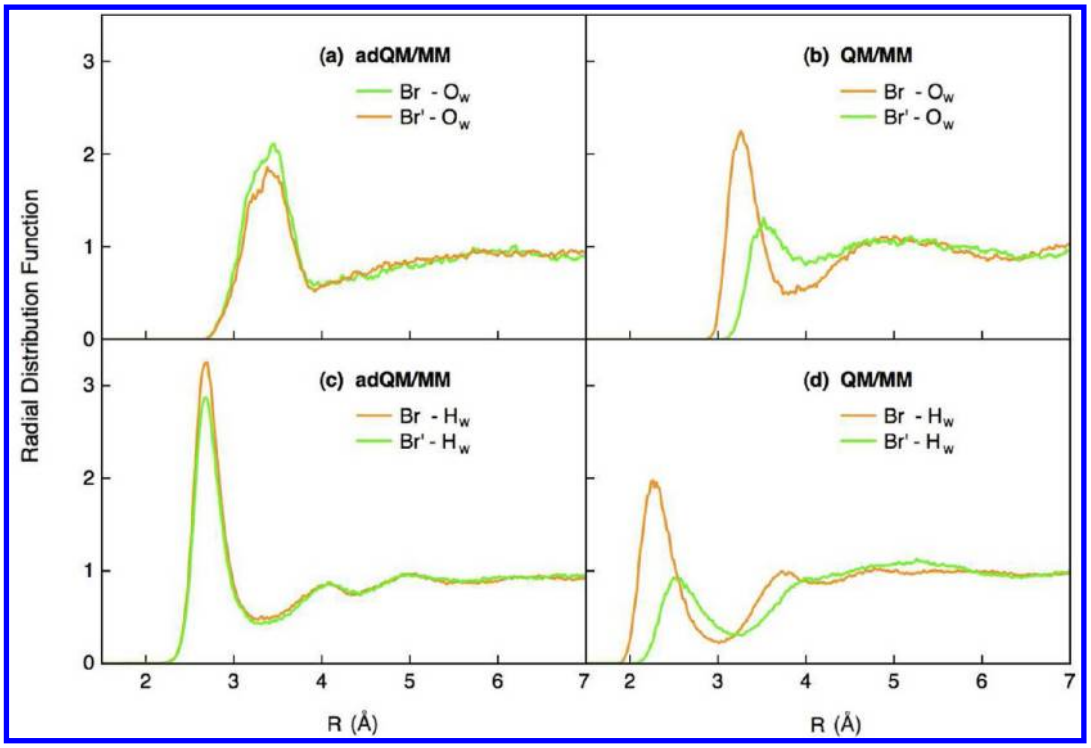

Figure 3. Radial distribution functions of the water molecules around the two Br atoms at the transition state of the reaction $CH_3Br + Br'^- \rightarrow Br^- + CH_3Br'$. RDFs are calculated from the simulation of the umbrella window at RC = 0 Å.

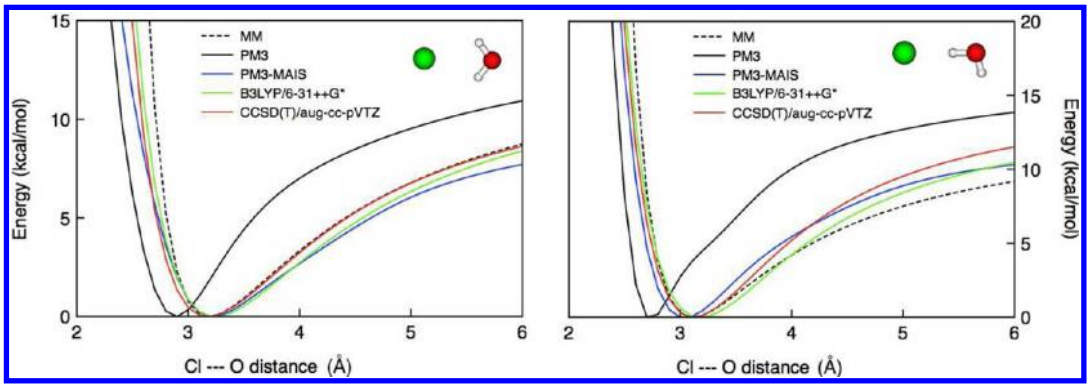

Figure 4. Potential energy scans for two Cl^-−H_2O orientations calculated using different QM levels of theory. Geometry of the water molecule is fixed at its MP2/aug-cc-pVTZ-optimized structure. MM force field parameters for the Cl^- ion were taken from ref 21, and the SPC/Fw model was used to describe the water interactions.[20]

embedding scheme for the description of the QM/MM interactions, and (3) the QM level of theory. All simulations were performed in the NPT ensemble (T = 300 K and P = 1 bar) for a system containing a single Cl^- ion and 1343 water molecules in a periodic cubic box of side length ≈ 34 Å. The semiempirical PM3-MAIS method[37,38] and the B3LYP density functional[39] with the 6-31++G* basis set[40−44] were used in the adQM/MM simulations. Both methods provide an accurate description of the Cl^-−H_2O interactions when compared to the corresponding results obtained at the CCSD(T) level of theory (Figure 4). The RDFs describing the spatial correlation between the Cl^- ion and the oxygen (left panels) and hydrogen (right panels) atoms of the water molecules are shown in Figure 5 along with the corresponding curves derived from neutron diffraction measurements of dilute KCl solutions.[45] The experimentally derived RDFs indicate that, on average, 6 and 10 water molecules are located in the first and second hydration shells, respectively.

The comparison between the RDFs calculated with the PM3-MAIS method for different sizes of the A region using the EE scheme is reported in Figure 5a and 5b. On the basis of the experimental values mentioned above, in the adQM/MM simulations with 6 water molecules in the A region and 3 water molecules in the transition region (6A/3T) all molecules belonging to the first solvation shell were described at the PM3-MAIS level. In the 16A/15T simulations, the water molecules in both the first and the second solvation shells were treated quantum mechanically, while 15 water molecules were included in the transition region. Analysis of the Cl^-−O and Cl^-−H RDFs obtained from the 6A/3T, 8A/3T, and 16A/15T simulations shows that both the position and the shape of the first peak are essentially insensitive to the number of water molecules included in the A region. In all cases, the first peak of

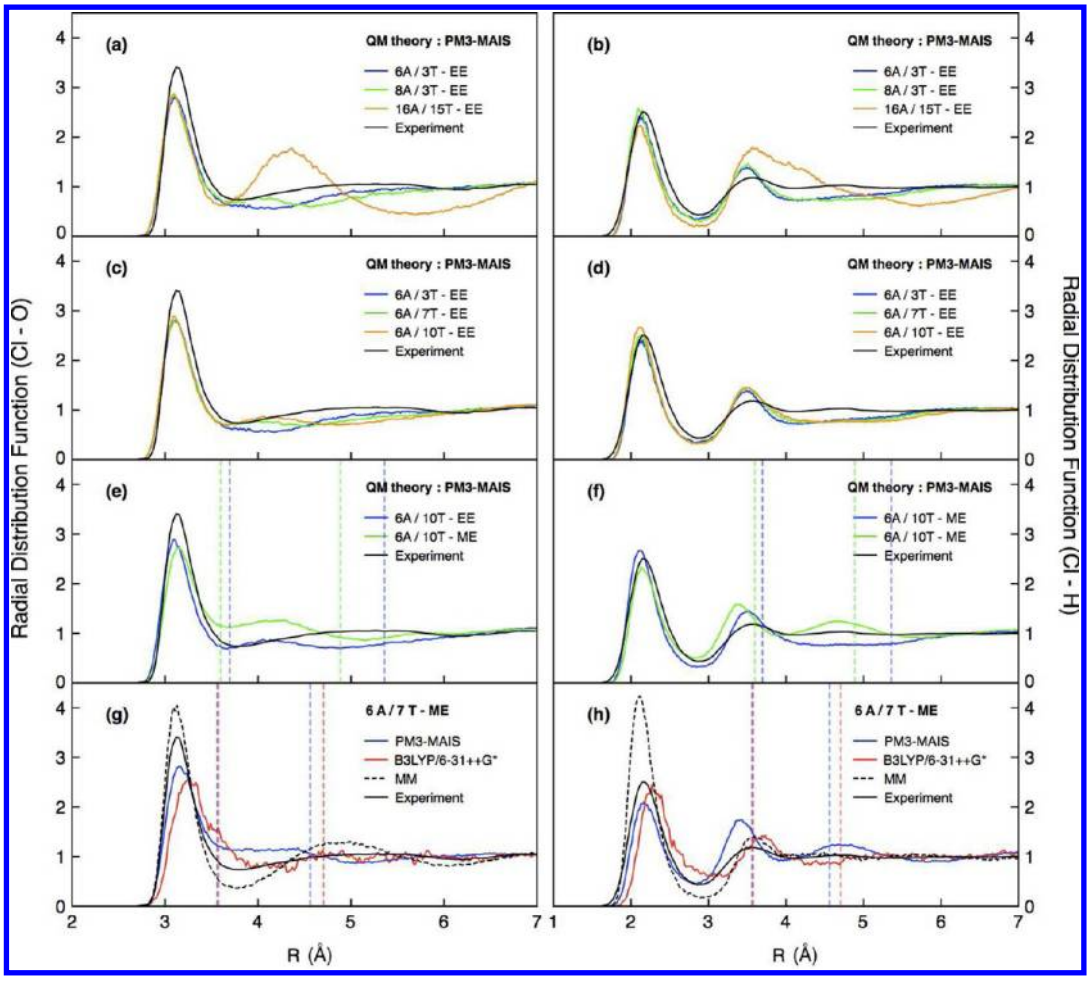

Figure 5. Cl^-−O and Cl^-−H radial distrubution functions. Experimentally derived RDFs are from ref 45. Graphs from top to bottom describe the comparisons for (1) different A region sizes, (2) different T region sizes with a fixed A region size, (3) mechanical (ME) versus electrostatic (EE) embedding, and (4) different QM levels of theory. Boundaries of the T regions are indicated with colored dashed lines. Green and blue dashed lines in panels e and f correspond to the inner and outer boundaries of the T regions for ME and EE, respectively. Blue and red dashed lines in panels g and h correspond to the inner and outer boundaries of the T regions for PM3-MAIS and B3LYP, respectively. Force field parameters for the Cl^- ion were taken from ref 21, and the SPC/Fw model was used to describe the water interactions.[20]

the calculated RDFs is in good agreement with the available experimental data. By contrast, noticeable differences exist at larger separations. In particular, a second peak appears in the Cl−O RDF for simulations with 8 water molecules in the A region, which becomes significantly more pronounced in the 16A/15T simulations. These results suggest that although the PM3-MAIS method provides an accurate representation of the water−Cl^- interactions, the water−water interactions in solution appear not to be well reproduced. This leads to an unphysically overstructured second solvation shell in adQM/MM simulations where a larger number of H_2O molecules is treated at the quantum-mechanical level.

Panels c and d of Figure 5 show the comparison between the RDFs obtained from simulations carried out with an identical number of water molecules located in the A region (6) and a varying number of molecules in the T region. Also in this case, the position and shape of the first peak are independent of the specific partitioning scheme used in the simulations. As the size of the T region increases, a small peak at ∼4 Å appears in the Cl^-−O RDF. Since a larger T region implies that an increasing number of water molecules acquire a partial QM character, the appearance of the second peak can be attributed again to inaccuracies in the PM3-MAIS description of the water−water interactions.

The RDFs obtained from adQM/MM simulations carried out using the same partitioning (6A/10T) of water molecules between the A and the T regions but different embedding schemes are shown in Figure 5e and 5f. Despite the similarity of the first peaks of both Cl^-−O and Cl^-−H RDFs, noticeable differences between the two sets of results exist at larger separations. In particular, the ME scheme features a narrower T region (delimited by vertical dashed lines) as a result of the increased density as well as relatively more structured RDFs.

Finally, to determine the sensitivity of the adQM/MM results on the specific QM level of theory, the RDFs obtained from simulations with the B3LYP/6-31++G* method are compared in Figure 5g and 5h with the corresponding results obtained with the PM3-MAIS method. Both simulations were performed using the ME scheme. Also shown for reference are the RDFs obtained from MD simulations in which classical force fields were used for the Cl^-−H_2O and H_2O−H_2O interactions as described in section 2.2. Overall, the B3LYP/6-31++G* method provides the best agreement with the experimentally derived data. However, both the Cl^-−O and the Cl^-−H RDFs are slightly shifted to larger distances, and the

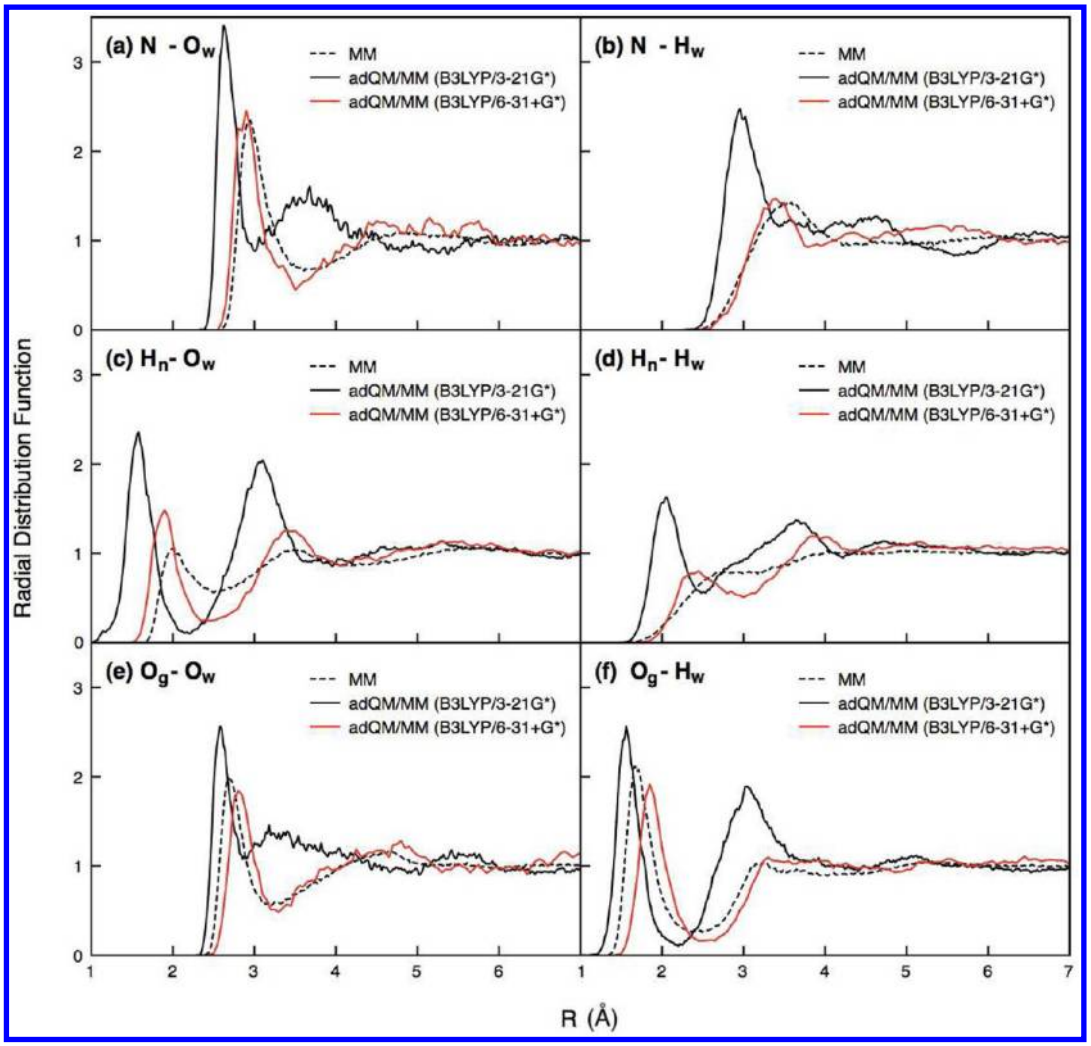

Figure 6. Radial distrubution functions of the water molecules around the hydrophilic group of the glycine zwitterion, $-NH_3^+$ and $-COO^-$. O_w and H_w labels the oxygen and hydrogen atoms in water, respectively. O_g labels both oxygen atoms of the $-COO^-$ group, and H_n labels all hydrogen atoms of the $-NH_3^+$ group.

first peak of the Cl^-−O RDF appears to be somewhat too low and broad. Since analysis of the PM3-MAIS RDFs shows that the shape of the first peak is essentially independent of the specific embedding scheme used in the calculations, the differences between the B3LYP/6-31++G* and the experimental RDFs are likely due to inaccuracies of the B3LYP method in the description of the Cl^-−H_2O interactions. It is important to note that standard classical simulations with MM potentials predict Cl^-−O and Cl^-−H RDFs that are significantly overstructured.

3.3. Solvation Structure of the Zwitterionic Form of Glycine in Water. It has been shown that protein motion is intimately linked to the motion of the surrounding aqueous environment.[46−52] For example, it is well known that in the absence of water the reactivity of enzymes stops or is severely slowed down.[53] Molecular-level characterization of hydration effects is therefore crucial for development of a comprehensive understanding of the relationship between protein structure and function under physiologically relevant conditions.[54] The interactions between biomolecules and water have been extensively studied using different experimental techniques[55−59] and theoretical methodologies.[60−64] In particular, classical MD simulations with empirical force fields have largely contributed to the current understanding of the structural, thermodynamic, and dynamical properties of protein−water interfaces. However, it has been shown that the results of these simulations strongly depend on the specific potential functions used to describe the water−protein interactions.[65,66] Importantly, most of the empirical force fields used in the MD simulations are not capable of accurately reproducing quantum-mechanical electronic effects associated with charge transfer and polarization, which can play a significant role in molecular systems with strong hydrogen bonds. In this regard, full ab initio molecular dynamics simulations at the Hartree−Fock level have recently been reported for an entire protein in water.[67]

In this study, the adQM/MM method is used to characterize the hydration structure of the zwitterionic form of glycine for which both experimental[68] and ab initio MD[69] results are available. The adQM/MM simulations were carried out in the NPT ensemble for a system containing 1 glycine molecule and 779 water molecules in a periodic cubic box of side length ∼28 Å. In all simulations, 18 water molecules were included in the A region while 7 water molecules were assigned to the T region. Three sets of simulations were performed: (1) classical MD simulations with empirical force fields, (2) adQM/MM simulations using the B3LYP density functional with the 3-

21G* basis set[70] for the QM calculations, and (3) adQM/MM simulations using the B3LYP density functional with the 6-31+G* basis set[40−44] for the QM calculations. The hydration structure around the glycine molecule was characterized through analysis of the RDFs describing the spatial correlation between the water molecules and the two hydrophilic functional groups $-NH_3^+$ and $-COO^-$. The results are summarized in Figure 6.

Large differences exist between the three sets of results, which are particularly evident in the hydration of the $-NH_3^+$ group (Figure 6a−d). In this case, the adQM/MM results are sensitive to the choice of the basis set, with the RDFs obtained from simulations with the 3-21G* predicting a more compact water structure. The first peak of the $N-O_w$ RDF calculated from the B3LYP/3-21G* and B3LYP/6-31+G* adQM/MM and the fully classical MM simulations is located at 2.62, 2.86, and 2.95 Å, respectively. The B3LYP/6-31+G* adQM/MM result is in excellent agreement with the corresponding experimental value of 2.85 ± 0.05 Å.[68] An important feature that is evident from analysis of the hydration structure around the $-NH_3^+$ group is the presence of two well-separated regions in the H_n-O_w RDF (Figure 6c) calculated at the adQM/MM level. This specific feature, which also appears in the RDFs calculated using ab initio MD simulations, is completely missing in the RDFs obtained from MM simulations. This difference can be explained by considering formation of strong hydrogen bonds between the oxygen atoms of the water molecules and the hydrogen atoms of the $-NH_3^+$ group involving a significant amount of charge transfer that is neglected when MM potentials are used. Qualitatively similar differences were obtained for the RDFs describing the spatial correlation between the water molecules and the COO^- group, with the B3LYP/3-21G* adQM/MM simulations again predicting a more compact hydration structure. All three O_g-H_w RDFs display a well-defined first peak between 1.8 and 2.1 Å, which was also found in the ab initio MD simulations of ref 69.

As expected from analysis of the radial distribution functions, significant differences are found in the local structure of water around the $-NH_3^+$ group obtained from the adQM/MM simulations with the B3LYP functional depending on the basis set. As shown in Figure 7, the more compact structure obtained with the 3-21G* basis set leads to establishment of an extended hydrogen bond network around the glycine zwitterion (left panel). Due to this particular arrangement, proton transfer involving the $-NH_3^+$ group and surrounding water molecules was observed to occur with relatively high frequency (>1 ps^{-1}). By contrast, no proton transfer events were observed in the adQM/MM simulations with the 6-31+G* basis set, which also provides a more accurate description of the overall hydration properties. This analysis clearly indicates that besides the density functional, the choice of a proper basis set is critical for the physically correct representation of molecular processes in solutions using adQM/MM simulations.

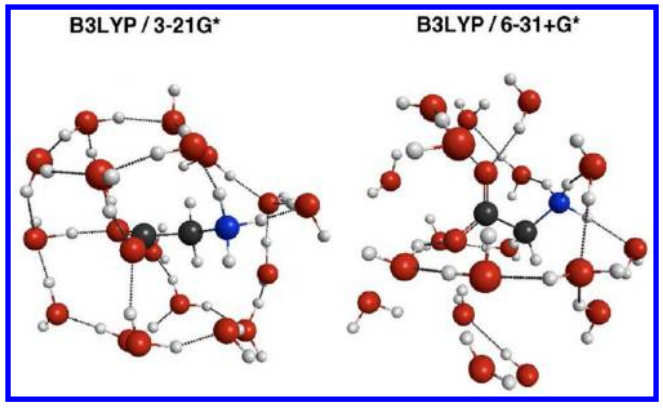

Figure 7. Representative snapshots of the hydration structure around the glycine zwitterion obtained from adQM/MM simulations with the B3LYP/3-21G* (left) and B3LYP/6-31+G* (right) methods. Only the 18 closest water molecules are shown.

4. CONCLUSION

The adQM/MM method was applied to the study of (1) halide exchange S_N2 reactions in aqueous phase, (2) Cl^- hydration, and (3) solvation structure of the zwitterionic form of glycine in water. In all cases, a direct comparison with the results obtained from classical MD simulations with MM potentials and standard QM/MM simulations and, when available, with data from full QM simulations and experiments was made. It was shown that compared to the standard QM/MM results, the adQM/MM simulations provide a more physically realistic representation of chemical reactions in solution by correctly describing the interactions between the reactive complex and the water molecules along the entire reaction coordinate. Although the applications discussed in this study are focused on thermodynamic and structural properties, the adQM/MM method also represents a promising approach for investigating the dynamics of condensed-phase systems. In particular, the force continuity guaranteed by the adQM/MM scheme enables calculation of time-correlation functions that can be Fourier transformed to obtain dynamical and spectroscopic quantities, which are not accessible using currently available adaptive QM/MM approaches.[16,71] Work along these lines is ongoing in our lab.

Analysis of the Cl^- hydration structure indicates that the adQM/MM method represents an effective approach for studying solvation processes. In particular, it was shown that the structure of the first solvation shell is essentially independent of both the size and the location of the transition region between the active and the environment regions. Thus, the quality of the results obtained is mainly a function of the QM method chosen. Since the properties of the active region quickly converge with its size, solvation processes can be studied at higher QM levels of theory than those used in standard ab initio MD simulations at effectively the same computational cost.

Finally, analysis of the solvation structure of the zwitterionic form of glycine in water emphasized the importance of purely quantum-mechanical electronic effects in the description of strong hydrogen bonds in solution. Results that are comparable to full ab initio MD simulations and experiment were obtained with the present adQM/MM simulations. It was shown that besides the QM Hamiltonian an adequate basis set is critical for a physically correct representation of the molecular process of interest.

Although this study shows that the properties of the active region are effectively independent of the location of the transition region, the results appear to be sensitive to the specific scheme used to describe the electrostatic interactions between the QM and the MM regions. Future work will focus on development of improved embedding schemes for a more accurate representation of the boundary properties between the QM and the MM regions as well as application of the adQM/

MM method using more physically correct descriptions of the MM region including polarization effects.

AUTHOR INFORMATION

Corresponding Author

*E-mail: fpaesani@ucsd.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation through grant CHE-1111364 to F.P. and the University of California through UC Lab award 09-LR-06-117792 to R.C.W. and a TRO award to A.W.G. and F.P. Funding provided by the Department of Energy, Office of Biological and Environmental Research and Office of Advanced Computing Research, SciDAC award DE-AC36-99G0-10337, is acknowledged by A.W.G. We are grateful to the National Science Foundation for a generous allocation of computing time on XSEDE resources via award TG-CHE110009 to F.P. and award TG-MCB090110 to R.C.W. and A.W.G. as well as to the San Diego Supercomputer Center for a computing time allocation on the Triton Computing Cluster through the TRO program. The authors thank Prof. Todd Martinez for his help with the Amber/TeraChem interface and Dr. Rosa Bulo for many helpful discussions about the adQM/MM method.

REFERENCES

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.

(2) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718−730.

(3) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.

(4) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170−1179.

(5) Sherwood, P. In *Modern methods and algorithms of quantum chemistry proceedings*; Grotendorst, J., Ed.; John von Neumann Institute for Computing: Jülich, Germany, 2000; Vol. 3, pp 257−277.

(6) Gao, J. L.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.

(7) Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H. B.; Ghosh, N.; Prat-Resina, X.; Konig, P.; Li, G. H.; Xu, D. G.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458−6469.

(8) Lin, H.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *117*, 185−199.

(9) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(10) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. *J. Chem. Theory Comput.* **2009**, *5*, 2212−2221.

(11) Takenaka, N.; Kitamura, Y.; Koyano, Y.; Nagaoka, M. *Chem. Phys. Lett.* **2012**, *524*, 56−61.

(12) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chem. Phys.* **1996**, *211*, 313−323.

(13) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* **2002**, *355*, 257−262.

(14) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 2231−2241.

(15) Bernstein, N.; Kermode, J. R.; Csanyi, G. *Rep. Prog. Phys.* **2009**, 72.

(16) Bernstein, N.; Varnai, C.; Solt, I.; Winfield, S. A.; Payne, M. C.; Simon, I.; Fuxreiter, M.; Csanyi, G. *Phys. Chem. Chem. Phys.* **2012**, *14*, 646−656.

(17) Götz, A. W.; Park, K.; Bulo, R. E.; Paesani, F.; Walker, R. C. Manuscript in preparation.

(18) Case, D. A.; Darden, T. A.; Cheatham, T. E., III ; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Götz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vaníček, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *Amber 12*; University of California-San Francisco: San Francisco, CA, 2012.

(19) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(20) Wu, Y. J.; Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2006**, *124*, 024503.

(21) Joung, I. S.; Cheatham, T. E. *J. Phys. Chem. B* **2008**, *112*, 9020−9041.

(22) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. *Biopolymers* **1992**, *32*, 523−535.

(23) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.

(24) Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019−1031.

(25) Götz, A. W.; Isborn, C. M.; Clark, M.; Martínez, T. M.; Walker, R. C. Manuscript in preparation.

(26) *TeraChem*; PetaChem: Los Altos, CA, 2012.

(27) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2008**, *4*, 222−231.

(28) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2009**, *5*, 1004−1015.

(29) Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2009**, *5*, 2619−2628.

(30) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815−826.

(31) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622.

(32) Tubert-Brohman, I.; Guimarães, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 138−150.

(33) Parthiban, S.; de Oliveira, G.; Martin, J. M. L. *J. Phys. Chem. A* **2001**, *105*, 895−904.

(34) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(35) Yang, S. Y.; Fleurat-Lessard, P.; Hristov, I.; Ziegler, T. *J. Phys. Chem. A* **2004**, *108*, 9461−9468.

(36) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154−163.

(37) Bernal-Uruchurtu, M. I.; Ruiz-López, M. F. *Chem. Phys. Lett.* **2000**, *330*, 118−124.

(38) Arillo-Flores, O. I.; Ruiz-López, M. F.; Bernal-Uruchurtu, M. I. *Theor. Chem. Acc.* **2007**, *118*, 425−435.

(39) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(40) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257.

(41) Dill, J. D.; Pople, J. A. *J. Chem. Phys.* **1975**, *62*, 2921−2923.

(42) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; Defrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, *77*, 3654−3665.

(43) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(44) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. *J. Comput. Chem.* **1983**, *4*, 294−301.

(45) Mancinelli, R.; Botti, A.; Bruni, F.; Ricci, M. A.; Soper, A. K. *J. Phys. Chem. B* **2007**, *111*, 13570−13577.

(46) Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Parak, F. G. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 16047−16051.

(47) Frauenfelder, H.; Fenimore, P. W.; McMahon, B. H. *Biophys. Chem.* **2002**, *98*, 35−48.

(48) Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Young, R. D. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14408−14413.

(49) Frauenfelder, H.; Fenimore, P. W.; Chen, G.; McMahon, B. H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 15469−15472.

(50) Hermans, J.; Berendsen, H. J. C.; Van Gunsteren, W. F.; Postma, J. P. M. *Biopolymers* **1984**, *23*, 1513−1518..

(51) Rain, J. C.; Selig, L.; De Reuse, H.; Battaglia, V.; Reverdy, C.; Simon, S.; Lenzen, G.; Petel, F.; Wojcik, J.; Schächter, V. *Nature* **2001**, *409*, 211−215.
(52) Hunt, N. T.; Kattner, L.; Shanks, R. P.; Wynne, K. *J. Am. Chem. Soc.* **2007**, *129*, 3168−3172.
(53) Maréchal, Y. *The hydrogen bond and the water molecule: The physics and chemistry of water, aqueous and bio-media*; Elsevier: Amsterdam, 2006.
(54) Ball, P. *Chem. Rev.* **2008**, *108*, 74−108.
(55) Grossman, M.; Born, B.; Heyden, M.; Tworowski, D.; Fields, G. B.; Sagi, I.; Havenith, M. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1102−1108.
(56) Bryant, R. G. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 29−53.
(57) Halle, B. *Philos. Trans. R. Soc. London, Ser. B* **2004**, *359*, 1207−1224.
(58) Qiu, W.; Kao, Y.-T.; Zhang, L.; Yang, Y.; Wang, L.; Stites, W. E.; Zhong, D.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 13979−13984.
(59) Russo, D.; Hura, G.; Head-Gordon, T. *Biophys. J.* **2004**, *86*, 1852−1862.
(60) Tarek, M.; Tobias, D. J. *Phys. Rev. Lett.* **2002**, *88*, 138101.
(61) Li, T.; Hassanali, A. A.; Kao, Y.-T.; Zhong, D.; Singer, S. J. *J. Am. Chem. Soc.* **2007**, *129*, 3376−3382.
(62) Cheng, Y.-K.; Rossky, P. J. *Nature* **1998**, *392*, 696−699.
(63) Pizzitutti, F.; Marchi, M.; Sterpone, F.; Rossky, P. J. *J. Chem. Phys. B* **2007**, *111*, 7584−7590.
(64) Marchi, M.; Sterpone, F.; Ceccarelli, M. *J. Am. Chem. Soc.* **2002**, *124*, 6787−6791.
(65) Daura, X.; Oliva, B.; Querol, E.; Aviles, F. X.; Tapia, O. *Proteins: Struct., Funct., Genet.* **1996**, *25*, 89−103.
(66) Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. *J. Chem. Theory Comput.* **2012**, *8*, 1409−1414.
(67) Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. *J. Phys. Chem. Lett.* **2011**, *2*, 1789−1793.
(68) Kameda, Y.; Ebata, H.; Usuki, T.; Uemura, O.; Misawa, M. *Bull. Chem. Soc. Jpn.* **1994**, *67*, 3159−3164.
(69) Sun, J.; Bousquet, D.; Forbert, H.; Marx, D. *J. Chem. Phys.* **2010**, 133.
(70) Mitin, A. V.; Hirsch, G.; Buenker, R. J. *Chem. Phys. Lett.* **1996**, *259*, 151−158.
(71) Rowley, C. N.; Roux, B. *J. Chem. Theory Comput.* **2012**, ASAP, DOI: 10.1021/ct300091w.
(72) Albery, W. J.; Kreevoy, M. M. In *Advances in Physical Organic Chemistry*; Gold, V., Bethell, D., Eds.; Academic Press: New York, 1978; Vol. *16*, pp 87−157.