JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Multiscale Modeling of Chemistry in Water: Are We There Yet?

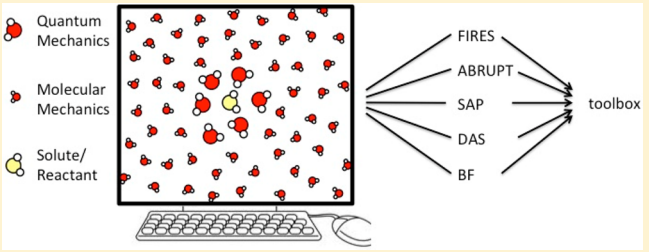
Rosa E. Bulo,*[,†,‡] Carine Michel,[†] Paul Fleurat-Lessard,[†] and Philippe Sautet[†]

[†]Université de Lyon, CNRS, Ecole Normale Supérieure de Lyon, Laboratoire de Chimie, 46, allée d'Italie, 69364 Lyon cedex 07, France

[‡]Department of Theoretical Chemistry, VU University Amsterdam, De Boelelaan 1083, 1081 HV Amsterdam, The Netherlands

Supporting Information

ABSTRACT: This paper critically evaluates the state of the art in combined quantum mechanical/molecular mechanical (QM/MM) approaches to the computational description of chemistry in water and supplies guidelines for the setup of customized multiscale simulations of aqueous processes. We differentiate between structural and dynamic performance, since some tasks, e.g., the reproduction of NMR or UV−vis spectra, require only structural accuracy, while others, i.e., reaction mechanisms, require accurate dynamic data as well. As a model system for aqueous solutions in general, the approaches were tested on a QM water cluster in an environment of MM water molecules. The key difficulty is the description of the possible diffusion of QM molecules into the MM region and vice versa. The flexible inner region ensemble separator (FIRES) approach constrains QM solvent molecules within an active (QM) region. Sorted adaptive partitioning (SAP), difference-based adaptive solvation (DAS), and buffered-force (BF) are all adaptive approaches that use a buffer zone in which solvent molecules gradually adapt from QM to MM (or vice versa). The costs of SAP and DAS are relatively high, while BF is fast but sacrifices conservation of both energy and momentum. Simulations in the limit of an infinitely small buffer zone, where DAS and SAP become equivalent, are discussed as well and referred to as ABRUPT. The best structural accuracy is obtained with DAS, BF, and ABRUPT, all three of similar quality. FIRES performs very well for dynamic properties localized deep within the QM region. By means of elimination DAS emerges as the best overall compromise between structural and dynamic performance. Eliminating the buffer zone (ABRUPT) improves efficiency and still leads to surprisingly good results. While none of the many new flavors are perfect, all together this new field already allows accurate description of a wide range of structural and dynamic properties of aqueous solutions.



1. INTRODUCTION

Water is a ubiquitous solvent. It is the main solvent for 'natural' chemical processes, such as those of a biological or hydrogeological nature. It is also increasingly used in organic synthesis and catalysis as a benign green solvent.[1−3] Its effect on chemical properties and reactivity most often goes beyond facilitating heat and mass transfer. In contrast to most organic solvents water often influences reactivity more directly and can considerably improve selectivity and yield. This was first illustrated in the 1980s by the work of Breslow on the rate enhancement and increased selectivity of Diels−Alder reactions conducted in water, as opposed to organic solvents such as isooctane or even methanol.[4] Following this pioneering work, exploring chemistry in water became a novel domain, expanding the toolbox of organic synthesis[1] and leading Sharpless et al. to define "on water" conditions when water is used as a solvent for insoluble reactants.[5−7] A number of reactions showed a dramatic increase in rate, from 18 h to completion in methanol to 10 min "on" water.[8] Replacing conventional organic solvents with water can also greatly improve enantioselectivities, as shown for instance by the group of Kobayashi for the asymmetric ring-opening of *meso*-epoxides with amines.[9] It can reduce or even eliminate side-reactions, and in some cases it is already applied on an industrial scale (i.e., by Novartis).[10] In the field of catalysis, water is increasingly used as a solvent as well. It is a natural choice for biomass conversion processes, being able to dissolve the targeted reactants (poly alcohols such as cellulose, sorbitol, glycerol, etc).[2] In addition, adding water as a co-solvent can dramatically increase the Pt-catalyzed oxidation rate of aliphatic alcohols.[3] The remarkable reactivity observed when exploring water as a solvent opens new perspectives for large-scale synthesis. However, experimental observations are not sufficient to provide a clear view of the underlying mechanisms. Theoretical chemistry—combined with experimental observations—can be a powerful tool to this effect. It can aid in the interpretation of spectra,[11] and it can provide detailed insights into reaction free energy profiles and mechanisms.[12−15]

However, predicting solvent effects has long been one of the great challenges in computational chemistry. This is particularly true for aqueous reactions, since water, in contrast to many organic solvents, is not an inert spectator. Its unique ability to form H-bonds ensures that water molecules several nanometers

Received: June 29, 2013
Published: October 17, 2013

   dx.doi.org/10.1021/ct4005596 | J. Chem. Theory Comput. 2013, 9, 5567−5577

away can influence the structure at the reactive site.[16] For the computational modeling of chemical reactions in solution, the ideal model system would thus include thousands upon thousands of solvent molecules. For such large and complex systems, the statistically relevant conformations can no longer be hand-picked. For a proper description, static calculations no longer suffice, and molecular dynamics (MD) or Monte Carlo (MC) simulations become a necessity. Despite the tremendous increase in computational power and the strong improvements in performance of quantum chemical methods,[17−19] quantum mechanical (QM) simulations today are still limited to hundreds of water molecules. To enhance the scale further we can take advantage of the fact that in solution chemistry the reactive site, while influenced by its environment, usually has a localized character. As a result, we can significantly reduce the costs of the simulations by using multiscale approaches, such as quantum mechanical/molecular mechanical (QM/MM)[20] methods. These methods have been used to simulate solution chemistry since the beginning of the 1980s. Initially, the solute molecules were treated QM, while the solvent molecules were treated MM.[21,22] As a consequence, direct participation of the solvent molecules in the reaction could not be accounted for. In fact, inclusion of solvent molecules in the QM region would not only allow the solvent molecules to react but would also improve the accuracy of the description of the solute−solvent interactions. A logical choice for the QM/MM study of a chemical reaction in solution is therefore a (set of) QM solute molecule(s) surrounded by a cluster of QM solvent molecules in an environment of a large number of MM solvent molecules. However, using traditional QM/MM[20] methods, the accurate QM water molecules would diffuse away from the reactive region with time, leaving the solute molecules surrounded by MM water.

To overcome this diffusivity problem, a new generation of advanced QM/MM approaches is emerging.[23−30] While this has already resulted in a variety of successful applications,[31,32] many of the methods were tested only for selected properties of simple model systems, often using an MM/MM description. In addition, they are not generally available in any of the standard MD codes. For potential users it is not only difficult to obtain the software, it is also hard to determine which of the approaches would best fit their needs. If the aim is to interpret a UV−vis or NMR spectrum, then the required simulation should properly describe the structure of a solution, so that a set of snapshots can be extracted from the QM/MM simulation and used to calculate the spectrum. If the goal is to decipher a reaction mechanism, then dynamical quantities, such as diffusion of the solute and solvent molecules, should be properly described by the approach as well. This paper will provide an overview of a selected set of the available advanced QM/MM approaches and assess their performances for each of these purposes.

One way to deal with the diffusivity problem is to constrain the QM molecules to remain within a certain region (henceforth denoted 'active region').[33] The flexible inner region ensemble separator (FIRES) approach[26] is a recent and clever refinement of this approach, which uses a constraining potential that prevents the MM solvent molecules from entering the active region (Figure 1). This constrained method will be compared with adaptive QM/MM approaches that take the diffusivity of solvent molecules into account.[23−25,27−30] These approaches adapt the description of the solvent molecules from QM to MM and vice versa as they move in and out of the active region. The simplest adaptive approach abruptly changes the description as the solvent molecules cross a given cutoff radius. In this paper, this approach will be denoted by ABRUPT. To avoid the sudden changes in the energy and forces involved, other approaches introduce a buffer zone between the active region and the environment, which eases the QM to MM transition. The permuted adaptive partitioning (PAP),[27,28] the sorted adaptive partitioning (SAP) approach,[27,28] and the difference-based adaptive solvation (DAS) approach[29,34,35] enforce a gradual transition along the buffer zone. The former method does this in perhaps the most intuitive manner and has been shown to result in stable energy conserving simulations. This approach is, however, computationally extremely costly and for this reason will not be explored in this paper. The SAP and DAS approaches are more economic alternatives, and they differ in the way the gradual QM to MM transition is realized (see Section 2.2.2). The buffered-force (BF) approach[30] uses its buffer zone to resolve an intrinsic bias introduced by the QM/MM frontier. However, it does so at the expense of conservation of energy as well as conservation of momentum.

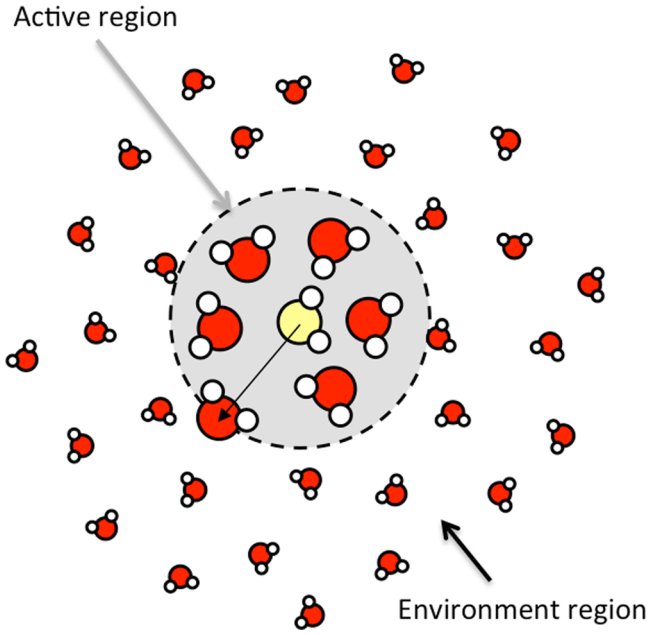


Figure 1. Schematic representation of the FIRES approach. The large water molecules are QM water molecules, while the small water molecules are treated MM. The yellow water molecule defines the center of the active (QM) region, and the thin arrow indicates the position of the spherical wall keeping the MM molecules out.

The aim of this paper is to assess the performance of the above approaches for the study of specific problems in solution chemistry. While our model system is water in water, the conclusions can be directly extrapolated to spectroscopic and/or mechanistic studies of small molecules in water. The paper is organized as follows: In Section 2 each of the five approaches is briefly reviewed. Section 3 presents the simulation setup and discusses the timings and availability of each approach. In Section 4 the performance of the approaches for structural and dynamic quantities is discussed. Finally, in Section 5, the general conclusions of this work will be briefly reviewed.

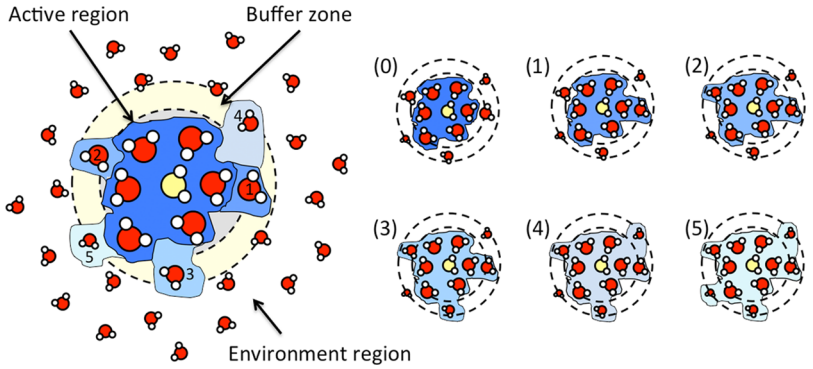


Figure 2. Schematic representation of the SAP and DAS schemes. The yellow water molecule defines the center of the active (QM) region, the large water molecules are QM, and the small water molecules are MM. The two dotted circles represent the inner and outer border of the buffer zone. The water molecules in the buffer zone are numbered based on their distance from the QM center (and therefore their QM character), and the size of these water molecules decreases with reduced QM character. There are five water molecules in the buffer zone, resulting in six ordered QM/MM partitionings. The first (dark blue) QM set only contains those water molecules that are inside the active region. The second set also contains water 1. The third set equals the second set plus water 2, etc. While SAP and DAS both construct the same set of partitionings, they each use them in a different manner.

2. ADVANCED QM/MM APPROACHES

In this paper a selection of representative QM/MM approaches are discussed and compared. In this section they will be briefly reviewed, and the relevant terms will be introduced. As mentioned in the Introduction, there are two main classes of approaches to solve the issue of diffusivity in the QM/MM study of solution chemistry. The first uses a constraint that keeps the QM molecules inside an active region and the MM molecules outside (in the environment region). The solvent molecules cannot diffuse from the active region to the environment region and vice versa. The FIRES[26] method is an elegant example of such an approach, and it is selected as representative in this paper. The ABRUPT approach, SAP,[27] DAS,[29] and BF[30] all belong to the second class: the adaptive approaches. In these cases the description of a solvent molecule (QM or MM) is modified based on its distance to the solute. In this manner solvent molecules can diffuse across the QM/MM frontier, changing description as they do so.

2.1. Flexible Inner Region Ensemble Separator (FIRES). The FIRES approach[26] uses a constraining potential that keeps the MM solvent molecules outside the active region (Figure 1). The constraining potential is located on a sphere around the active region, and the radius of the sphere is defined by the position of the QM solvent molecule farthest away from the QM center. Since the radius of the sphere is not constant throughout the simulation, the density of the QM region can adjust and equilibrate to the underlying free energy surface. While the constraint alters the dynamics with respect to an unconstrained system, the authors demonstrated that a structural quantity such as a radial distribution function (RDF) retains the correct values, at least for a test MM/MM simulation.

2.2. Adaptive Methods. The adaptive approaches selected here share a common methodological idea. In all of them, the description of the solvent molecules changes from QM to MM and vice versa, as they move in and out of the active region. This change can be abrupt (ABRUPT, BF) or smooth (SAP and DAS). The BF approach distinguishes itself from the other three by addressing the error introduced by short-range QM-MM interactions. All these adaptive approaches, in contrast to the constrained alternative (FIRES), take the diffusivity of solvent molecules fully into account.

2.2.1. ABRUPT. In this scheme, the description of the solvent molecules abruptly changes as they cross a predefined QM/MM boundary. The transition between the QM and MM description of a solvent molecule is discrete, which automatically means that the forces on the system change abruptly as solvent molecules cross the boundary. This simple scheme has previously been found to result in extreme heating of the system, even in the presence of a thermostat.[23]

2.2.2. SAP and DAS. To avoid abrupt changes in description and the subsequent strong discontinuities in the forces, the SAP and DAS approaches share a common strategy. They introduce a buffer zone between the active and the environment region where the description of a water molecule is partially QM and partially MM.[36] The smooth change in description requires a set of standard QM/MM calculations to be performed at each time-step. Each calculation corresponds to a different QM/MM partitioning, with the solvent molecules in the buffer zone either treated QM or treated MM. The SAP method defines an adaptive QM/MM potential $V^{\mathrm{ad}}(\mathbf{r})$ for the full system, as a weighted average of a set of standard QM/MM potentials $V_a(\mathbf{r})$ (eq 1). DAS directly defines a set of atomic forces $\boldsymbol{F}^{\mathrm{ad}}(\mathbf{r})$, as a weighted average of the gradients of the standard QM/MM potentials $V_a(\mathbf{r})$ (eq 2). It should be stressed that the force $\boldsymbol{F}^{\mathrm{ad}}(\mathbf{r})$ in eq 2 is not the gradient of the potential $V^{\mathrm{ad}}(\mathbf{r})$ defined in eq 1.

$$V^{\mathrm{ad}}(\mathbf{r}) = \sum_a \sigma_a^{\mathrm{SAP}}(\mathbf{r}) V_a(\mathbf{r}) \tag{1}$$

$$F^{\mathrm{ad}}(\mathbf{r}) = \sum_a -\sigma_a^{\mathrm{DAS}}(\mathbf{r})\mathbf{grad}(V_a(\mathbf{r})) \tag{2}$$

As a solvent molecule m_i advances away from the QM center, the sum of the weights $\sigma_a(\mathbf{r})$ of those partitionings a in eqs 1 and 2 that treat m_i QM decreases, and it becomes zero as m_i exits the buffer zone and enters the environment. With M solvent molecules in the buffer zone, there are 2^M different QM/MM permutations a that can in principle contribute to $V^{\mathrm{ad}}(\mathbf{r})$ (or $\boldsymbol{F}^{\mathrm{ad}}(\mathbf{r})$). The value of M depends on many factors, such as the nature of the solute and the size of the active and buffer regions. For the model system used in this paper typical values of M are 10 or larger. As a result, the number of contributing partitionings can become very large, and the SAP

dx.doi.org/10.1021/ct4005596 | J. Chem. Theory Comput. 2013, 9, 5567−5577

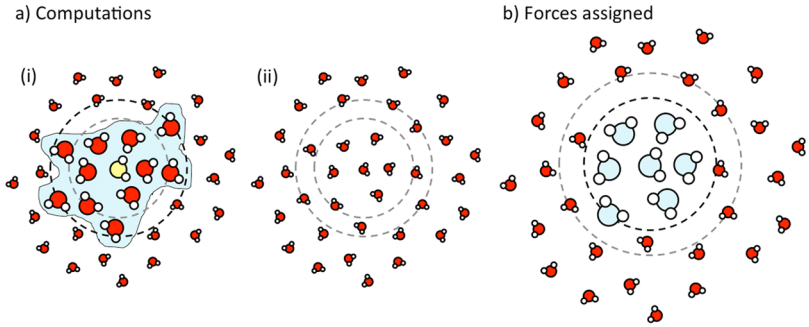


Figure 3. Schematic representation of the BF scheme. The yellow water molecule defines the center of the active (QM) region. The two dotted circles represent the inner and outer border of the buffer zone. (a) Two calculations are required: A QM/MM calculation with all molecules inside the outer sphere treated QM (i), and an MM calculation on the complete system (ii). (b) The water molecules in the inner sphere are assigned the accurate QM/MM forces from (i), while all other water molecules are assigned the MM forces from (ii).

and DAS methods are only affordable if this number is drastically reduced. Both methods construct the weight functions $\sigma_a(\mathbf{r})$ in such a way that only $M + 1$ 'ordered' partitionings contribute. The 'ordered' partitionings are those in which all QM solvent molecules are closer to the QM center than the MM molecules. In Figure 2 there are five water molecules in the buffer zone, which means that the six ordered (blue) sets of QM molecules, in addition to the water molecules in the active region, consecutively incorporate water molecules 1−5. The expressions for the weight functions $\sigma_a(\mathbf{r})$ are extensively discussed elsewhere.[27−29]

The SAP approach combines the QM and the MM potentials in a physically rigorous way and is in theory able to conserve the total energy of the system, so that simulations in a well-defined ensemble are possible. However, the adaptive potential $V^{ad}(\mathbf{r})$ often connects two very different potential energy surfaces. The resulting shape of $V^{ad}(\mathbf{r})$ can therefore be severely distorted from the original QM or MM shape, resulting in new erroneous minima.[29] The DAS approach, instead of defining an adaptive potential, defines an adaptive set of forces. Because no consistent adaptive potential is defined, a DAS simulation does not conserve energy, but its forces are completely continuous, and momentum is conserved. In addition, we have shown that DAS simulations reproduce the correct structural conformations.[29]

2.2.3. Buffered-Force (BF). There is a well-known problem associated with a QM/MM interface that none of the above approaches addresses. This problem stems from the fact that while the long-range interaction between QM and MM molecules is often quite well-described, the short-range interaction can deviate strongly from the ideal (QM) result. If the short-range forces between a QM and an MM solvent molecule differ very greatly from the interaction between two QM or two MM molecules, this can either lead to a preference for QM-MM coordination (in which case solvent molecules will collect on the QM/MM border) or an aversion to such interactions (which leads to a low solvent density at the QM/MM border). BF[37] addresses this problem by—yet again—introducing a buffer region between the active region and the environment. However, instead of using a weighted average of a set of standard QM/MM calculations, the BF method uses only two calculations: (i) a QM/MM calculation treating all solvent molecules in both the active region and the buffer zone QM and (ii) an MM calculation of the complete system (Figure 3a). The accurate forces from the QM/MM potential are then assigned only to the solvent molecules in the active region (Figure 3b). All molecules in the buffer zone and in the environment region get assigned forces from the fully MM potential. In this manner the solvent molecules at the edge of the active region see their direct neighbors on both sides as fully QM and do not feel a preference for either one of them. On the other hand, the solvent molecules in the buffer zone feel MM forces from all their neighboring molecules and do not feel a preference for either direction. While this will solve the problem of unwanted clustering or depletion at the QM/MM border (as described above), it intrinsically involves violation of Newton's third law (for every action there is an equal and opposite reaction). This will result in center of mass motion for the system as well as energy being lost or gained.

Our implementation of the BF method is a very basic one compared to the method presented in the work by Bernstein et al.[30] In their scheme, in addition to the force distribution as described above, very strong massive thermostats are introduced, allowing specific temperature regulation in the buffer zone. In addition, to avoid successive crossings by a solvent molecule at the borders surrounding the buffer zone, the borders are shifted based on the direction of the velocity of the solvent molecule. While these additional measures have certain advantages, they also lead to further violation of energy conservation and randomization of the dynamics. The extended scheme may affect dynamic quantities (temperatures and lifetimes) as reported in this paper, but the effect on structural properties (RDFs) is expected to be minor.

3. TECHNICAL DETAILS

In this section the simulation details are introduced, and some attention will be paid to the available implementations and the timings.

3.1. Simulation Details. We performed simulations with all combinations of two QM and two MM potentials. The QM potentials used are the semiempirical Hamiltonians PM6-DH+,[38,39] as implemented in the MOPAC package,[40] and DFTB,[41] as implemented in the ADF (2010) program package.[42] Periodic DFTB simulations for the full system were performed with the DFTB+ code.[58] Slater−Koster tables from the Dresden set were used.[43] As the two MM methods we employed the flexible TIP3P-Fs[49] force field, computed with the NAMD[44] program, and the reactive force field REAXFF,[50] as implemented in the ADF package. Periodic boundaries were computed using the MM code, using particle mesh Ewald in the case of the TIP3P-Fs simulations. All QM/MM potentials (with all four possible QM/MM combinations of the selected

dx.doi.org/10.1021/ct4005596 | J. Chem. Theory Comput. 2013, 9, 5567−5577

QM and MM methods) use mechanical embedding, which means that all interactions between QM and MM particles are computed with the MM potential. The advanced QM/MM schemes were all implemented in FlexMD, a python library for flexible multiscale MD simulations available in the ADF program package.[42] FlexMD uses the atomistic simulation environment (ASE)[45] python library for propagation of the MD simulations.

For direct comparison of all our QM/MM results with full QM results, the simulations were performed on a relatively small cubic periodic box of 110 water molecules, with a diameter of 15 Å. We ran adaptive QM/MM simulations with two different sizes for the active region, corresponding to a radius of 4.0 and 5.5 Å. This corresponds approximately to a FIRES QM region of 14 and 27 H_2O molecules, respectively. In all SAP, DAS, and BF simulations a buffer region with a width of 0.9 Å is used. The time-step chosen for all simulations is 0.5 fs. For all SAP simulations (apart from DFTB/REAXFF) we reduced the time-step to 0.1 fs, because the simulations crashed almost instantaneously. With the smaller time-steps the simulations ran longer, but many of them crashed before reaching 10 ps. For the four reference simulations with PM6-DH+, DFTB, TIP3P-Fs, and REAXFF, the box was equilibrated at 300 K for 10 ps with each method, followed by a production run in the canonical ensemble (NVT) of again 10 ps, which was used to obtain the thermodynamic and kinetic results presented in Section 4. The final configuration was then used for a 10 ps simulation in the microcanonical ensemble (NVE). For all QM/MM simulations the equilibration was performed using the DAS method with an active region of 4 Å radius, using the PM6-DH+/REAXFF combination. The final coordinates and velocities of this run were used for all 10 ps QM/MM production runs in the canonical (NVT) ensemble. Then each of these runs was continued for 10 more ps in the microcanonical (NVE) ensemble. A small subset of the NVT QM/MM simulations was continued for another 50 ps, and it was confirmed that the main features of the RDFs did not change after the first 10 ps.

3.2. Performance and Availability. Of the five advanced QM/MM approaches discussed, FIRES, ABRUPT, and BF require only a single QM/MM calculation per time step, while SAP and DAS require $M + 1$ (roughly 10) QM/MM calculations (see Section 2). This suggests that the first three methods will be considerably faster than SAP and DAS. However, the SAP and DAS calculations can be straightforwardly parallelized and provided enough processors are available SAP and DAS can, in theory, be as fast as the other approaches. Among FIRES, ABRUPT, and BF, the latter is marginally slower, because the active region in the BF calculation is chosen larger than in FIRES and ABRUPT. Table 1 depicts real time timings for PM6-DH+/TIP3P-Fs simulations with all advanced QM/MM approaches compared to the full QM case for a periodic box of 110 water molecules (see Section 3). It shows that even with the small periodic box used the advantage over full QM is significant, and the simulation time is reduced by at least a factor 10 for all approaches but SAP. As expected ABRUPT and FIRES are the fastest methods, closely followed by BF. While DAS is somewhat slower, the disadvantage is small, due to efficient parallelization over the 12 cores of the machine (on average at each time step the number of computed partitionings is 7, while the maximum number of partitionings over the full 10 ps is 14). The SAP simulation is ∼5 times slower than the others, because the spiky potential energy surface imposes a propagation time step 5 times smaller than the other approaches (see Section 3).

Table 1. Real Time Timings (in h/ps) of PM6-DH+ and PM6-DH+/TIP3P-Fs Simulations of a Periodic Box Containing 110 Water Molecules on a Linux Cluster[a]

QM	FIRES	ABRUPT	SAP	DAS	BF
43.6	1.20	1.18	10.4	1.63	1.35

[a]All simulations are run in serial, except DAS and SAP, which are run in parallel on 12 cores. For all simulations except SAP, a time-step of 0.5 fs is used. The time-step for the SAP simulation is 0.1 fs.

All methods compared in this paper are available in the FlexMD python library for flexible multiscale MD simulations distributed with the ADF program package. The original BF approach is implemented in the QUIP/libAtoms software package,[46] and SAP is implemented in the QMMM package.[47] In addition, ABRUPT, DAS, and the original BF approach are under development in the widely used AMBER molecular dynamics package.[48,32]

4. RESULTS AND DISCUSSION

In the following we compare the performance of the five advanced QM/MM approaches for solution chemistry. The ideal approach is stable, fast, and correctly describes the structure as well as the dynamics of a solution. First the structural performance is addressed, using the radial distribution function (RDF) of the oxygen atoms around the central QM oxygen atom. In combination with stability and speed, a good description of the structure suffices for some purposes. Second the dynamical performance is discussed, evaluating temperature conservation and coordination lifetimes. In all cases the quality of the QM/MM results was assessed by comparison to the full QM values in the active region. To compare the intrinsic performances of each advanced QM/MM approach, we use different combinations of QM and MM potentials, and two different sizes for the active region (see Section 3).

4.1. Structural Data. Possibly the most basic and important data an MD simulation provides is the average structure of a system in equilibrium. The equilibrium structures are directly related to the free energy differences that accompany structure change, evaluation of which is often the ultimate goal in the study of chemical processes.

4.1.1. Reference Data. The performance of all five advanced QM/MM approaches (as described in Section 2) is strongly influenced by the difference between the QM and MM potential energy surfaces. We illustrate this by using two different QM (PM6-DH+ and DFTB) and two MM potentials (TIP3P-Fs[49] and REAXFF[50]). The chosen potentials are further described in the Section 3. As a quantifier for the difference between the potentials we report the RDFs for the O−O distance (Figure 4), for 10 ps NVT simulations with each of the four potentials, and we compare the resulting plots with an experimentally obtained curve (thick red line).[51] The REAXFF curve most strongly resembles the experimental result. PM6-DH+ overestimates the ordering of the solution and underestimates the shortest O−O distance, as a direct consequence of the fact that it underestimates the length of the H-bond. The two MM potentials are very similar to one another, but of the two, REAXFF is the most ordered (and therefore the most similar to PM6-DH+). The DFTB

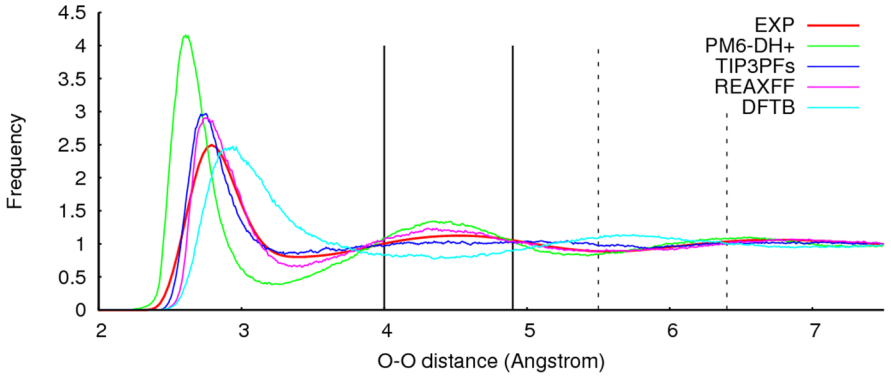


Figure 4. Experimental (thick red line), fully QM (PM6-DH+: green, DFTB: turquoise) and fully MM (TIP3P-Fs: blue, REAXFF: purple) O−O RDF values for water. The vertical lines represent the two different buffer zones that were used in the QM/MM simulations discussed later.

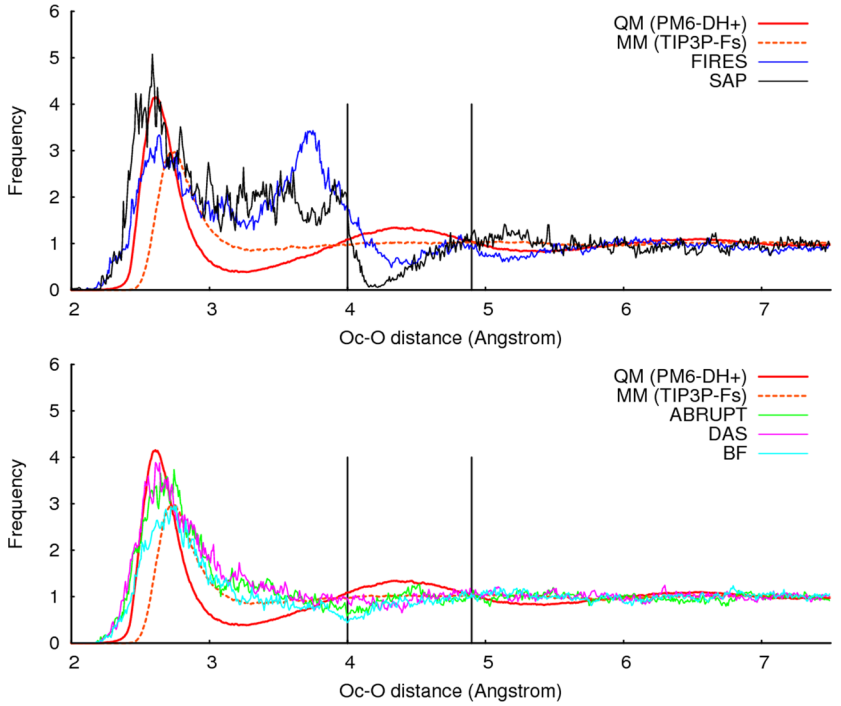


Figure 5. Oc−O RDF values (Oc is central oxygen) for water from QM/MM simulations (PM6-DH+/TIP3P-Fs) with an active region up to 4.0 Å (in FIRES simulation the active region is ∼4 Å). Top (FIRES, SAP) and bottom (ABRUPT, BF and DAS). The vertical lines represent the inner and outer borders of the buffer zone.

simulations strongly underestimate the order and result in the longest H-bond of all methods. The first solvation shell is significantly shifted to the right compared to the other methods, and the second solvation shell is shifted even further. For all potentials except DFTB the second solvation peak lies in the buffer zone corresponding to the smallest active region (black vertical lines), illustrating once more how much DFTB differs from all other methods.

4.1.2. Comparison of the Radial Distributions. RDFs for the distance between the central QM oxygen and all other oxygens were obtained from 10 ps simulations in the NVT ensemble. For each advanced QM/MM approach, eight simulations were performed using different combinations of QM and MM potentials (see Section 3). Figure 5 contains the RDF plots for a few selected simulations with PM6-DH+/TIP3P-Fs and an active region of 4.0 Å. The RDFs obtained with the FIRES and SAP simulations have a very strong peak at the inner edge of the buffer zone, and, in the case of SAP, also a depletion inside the buffer zone. This behavior is representative of the majority of the QM/MM combinations (see Supporting Information). The distorted structures in the SAP simulations are caused by the fact that the shape of the adaptive potential is very different from the original QM or MM shape. The result is in agreement with previously published data for MM/MM simulations, which were analyzed in detail.[29] On the contrary, the RDFs produced with DAS, BF, and ABRUPT do not show any large local deviations from the QM and MM curves. The QM first solvation peak is well reproduced by all three methods. Beyond the first solvation shell the QM/MM water density is higher than in the fully QM simulation (in the QM/MM simulation the QM density has some freedom to adjust to the underlying potential, whereas in the fully QM simulation the density is enforced by the NVT conditions). The fact that the QM H-bonds are significantly shorter than the MM H-bonds (first peak in Figure 4) is the cause of this redistribution of water densities over the different regions.[52,53]

dx.doi.org/10.1021/ct4005596 | J. Chem. Theory Comput. 2013, 9, 5567−5577

Table 2. Average (over all QM/MM combinations) RMSD of the QM/MM RDF (between the central QM oxygen and all others) from the QM RDF Inside the Active Region[a]

	FIRES	ABRUPT	SAP	DAS	BF
4.0 Å	0.51	0.26	0.48	0.27	0.28
5.5 Å	0.56	0.38	0.55	0.40	0.33

[a]Two different active region sizes (4.0 and 5.5 Å) are tabulated.

For each advanced QM/MM approach with all QM/MM pairs and both buffer zone sizes the root-mean-square deviation (RMSD) from the QM RDF inside the active region was computed. The average deviation over the four QM/MM pairs was obtained for each approach, and the results are given in Table 2. In agreement with Figure 5 the FIRES and SAP errors in the active region are larger than the errors produced by the three other approaches. Out of the three approaches that perform best (DAS, BF, and ABRUPT), no consistently superior performance can be discerned for the simulations with a small active region. For those simulations with a larger active region the deviations are generally larger. This is because the surface of the active region has increased as well, and more molecules are present at the boundary. For these simulations BF clearly has the advantage over the others.

4.1.3. On the BF Results. Analysis shows that the advantage of BF is not related to any pronounced local differences in the shapes of the curves, but from the fact that BF consistently produces a lower water density than the other methods throughout the active region. The fact that BF does not appear to solve any local issues at the boundary can be considered surprising, as the BF method was developed specifically for this purpose. The original BF paper[30] demonstrated that for water in water a simulation without a buffer zone (comparable to our ABRUPT simulations) results in a spurious peak in the RDF at the QM/MM boundary, which is corrected for by the introduction of the BF buffer zone. Our ABRUPT simulations, however, display no such artifact. We assign this effect to our use of mechanical embedding to describe the interaction between the QM and the MM regions, whereas the authors of the BF method use electronic embedding. Electronic embedding is known to introduce difficulties in the description of the interaction at the QM/MM boundary.[54,55] It is an invaluable approach in two important situations: (1) If one needs spectroscopic data for which the influence of the environment on the QM electron density is required; and (2) if, during the simulation, the electron density of the active region undergoes very large changes that strongly affect the interaction with the environment. Mechanical embedding, on the other hand, circumvents difficulties at the interface and uses thoroughly tested force field parameters to describe the interactions between the QM and the MM region. Therefore, when the simulation objectives allow it, mechanical embedding may be the safer choice. Our results imply that, even though BF produces slightly better structural properties than ABRUPT and DAS, the use of mechanical embedding greatly reduces the advantage of BF over the other QM/MM approaches for simulations in water.

4.1.4. Analysis of the FIRES Results. We further investigated FIRES, to understand why it causes such large deviations from the QM structures. We ran a TIP3P-Fs/TIP3P-Fs FIRES simulation with 14 water molecules in the active zone. In agreement with the SWM4-NDP/SWM4-NDP[56] simulations in the original paper,[26] the RDF closely resembles the reference MM curve, and the effect of the wall appears to be minimal (Figure 6, top). Using the same TIP3P-Fs potential, we also ran a simulation using a simpler approach, constraining only the

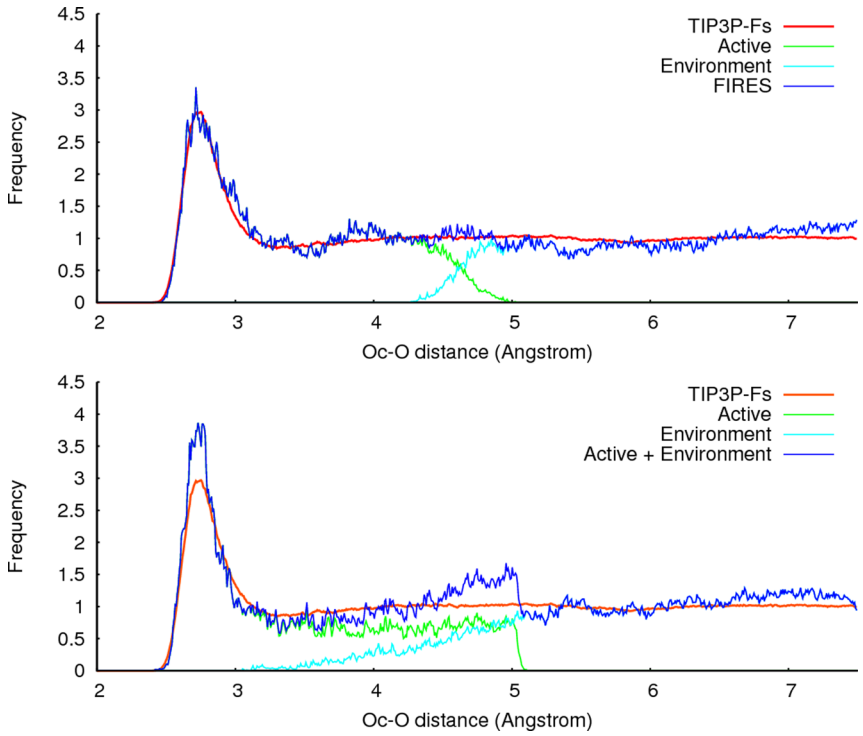


Figure 6. Top: RDF values for FIRES MM/MM simulation (TIP3P-Fs/TIP3P-Fs) with an active region of 14 water molecules. Bottom: RDF values for a 10 ps TIP3P-Fs/TIP3P-Fs simulation with an active region of 14 water molecules, using a fixed quadratic wall at a distance of 5.0 Å from the central water molecule (Oc). This wall constrains the active region water that has a maximum distance from the QM center.

dx.doi.org/10.1021/ct4005596 | J. Chem. Theory Comput. 2013, 9, 5567−5577

Table 3. The Average Temperature Drift (in K/ps) in NVE Simulations[a]

	DFTB		PM6-DH+		REAXFF		TIP3P-Fs	
	−0.4		3.3		−0.6		−0.9	
	DFTB/REAXFF		DFTB/TIP3P-Fs		PM6-DH+/REAXFF		PM6-DH+/TIP3P-Fs	
	active region radius (Å)							
	4.0	5.5	4.0	5.5	4.0	5.5	4.0	5.5
FIRES	0.7	−0.3	0.7	0.4	0.3	3.2	3.2	0.2
ABRUPT	41.9	115.4	140.0	437.4	38.6	218.7	95.6	**919.4**
SAP	−4.5	42.7	**737.9**	**5.35 × 10^5**	**Nan**	154.3	**249.7**	**655.1**
DAS	14.6	42.8	80.3	221.8	13.4	34.9	28.6	90.6
BF	184.2	533.1	**579.3**	1040.1	3189.1	**758.3**	**627.0**	**1708.1**

[a]If the numbers are bold, then the simulation crashed before reaching 10 ps.

active water molecules within a 5.0 Å radius of the central water molecule. We know from the FIRES simulation (Figure 6, top) that this constraint is far enough away from the center to allow the active region water molecules to reach equilibrium density. Figure 6 (bottom) depicts the RDF for the active and environment water molecules. The data show that the unconstrained (environment) water enters into the active region, while the constrained (active region) water clusters at the wall in response, creating a false minimum there. We now carefully speculate that this tendency for the water molecules to cluster at a wall is also present in the FIRES simulation, but much less visible. Because the FIRES wall is mobile, the position of the wall is delocalized over a certain region, and the false minimum disappears into the noise.

4.1.5. Conclusions. To conclude the section on structural properties, this comparative study shows that neither SAP nor FIRES can be relied upon to consistently return representative RDFs. DAS, BF, and ABRUPT all perform well, with only a slight preference for the BF method. Since the costs of BF and ABRUPT simulations are smaller than the costs of a DAS simulation, BF and ABRUPT emerge as the best choices if only structural data are required.

4.2. Dynamics. While the RDFs provide a measure for the quality of structural averages obtained from the QM/MM schemes, the advantage of MD over stochastic methods like MC is that it is able to produce dynamic quantities as well. In fact, if a dynamic quantity like diffusion is incorrect in a simulation (i.e., underestimated) a system might never explore all the possible stable states. This, in turn, will result in improper sampling, and unless special care is taken to enhance the sampling along the relevant degrees of freedom, this will result in inaccurate free energy profiles.

4.2.1. Temperature Conservation. A first measure for the quality of the dynamics of a simulation is energy or temperature conservation. If the dynamics of a simulation is wrong, this generally means that there is unphysical local heating or cooling of the system. A method that defines a continuous potential energy surface with a continuous gradient, and numerically integrates with an appropriate time-step, will generally conserve energy, and for an equilibrated system it will conserve the temperature as well. Of all the approaches discussed here, SAP is the only one that meets this criterion. Energy conservation is more sensitive to the quality of the dynamics than temperature conservation, but it cannot be tested for DAS, ABRUPT and especially BF, which do not provide a unique potential energy for the entire system. To compare all five approaches on the same ground, we used temperature conservation as the decisive measure. Table 3 shows the average temperature drift during the QM/MM simulations with active regions of 4.0 Å and 5.5 Å. The temperature drift is defined as the average slope of the temperature in time.

The temperature drift is <1 K/ps for the fully QM and MM simulations. Only the PM6-DH+ simulation has a slightly larger drift, which can be attributed to insufficient precision of the gradient calculations in MOPAC. Since in the QM/MM simulations only a small part of the forces are derived from a PM6-DH+ potential, the drift caused by the MOPAC error is expected to remain small. Only the FIRES simulations match the fully QM and fully MM simulations in temperature conservation. The sharp peaks on the SAP potential energy surface lead to large temperature drifts, even with a small time-step of 0.1 fs. In many cases the steep forces involved caused the SAP simulations to crash. Further reduction of the SAP time-step would eventually have resulted in energy and temperature conservation, but at such a large cost in simulation time that it is not worth the effort. Contrary to the SAP simulations, where the temperature increase occurs in short bursts alternated by temperature conserving stretches of time, the temperature increase in the three remaining simulations (DAS, BF, and ABRUPT) is gradual and exponential. Out of these three methods, BF results in the largest temperature drift. The smallest temperature drift is found for the DAS simulations, although the values are far from negligible. Both BF and ABRUPT simulations regularly crash due to large temperature jumps, while DAS appears to be the most stable approach of the three. Because the only real difference between BF and ABRUPT is the lack of momentum conservation of the former, comparison between the two suggests that the majority of the large temperature drift in the BF simulation is caused by this effect.

Because we need to compare our QM/MM results to full QM reference simulations the total size of our system is rather small. In general, the main advantage of a QM/MM simulation is that the MM region is comparatively free of cost, so that very large periodic cells can be used. In a much larger water box the local temperature fluctuations in the buffer zone are dampened by the large MM environment, so that overall temperature drifts remain smaller. To illustrate this we performed a DAS simulation (DFTB/TIP3P-Fs) on a larger water box, more representative of a practical application of QM/MM. A 10 ps simulation of a periodic water box of 30 Å (915 water molecules) with DAS (DFTB/TIP3P-Fs), and a buffer zone between 4.0 Å and 4.9 Å, has a temperature drift of 7.2 K/ps, which is significantly smaller than the 80.3 K/ps drift in Table 3 for the 15 Å system.

dx.doi.org/10.1021/ct4005596 | J. Chem. Theory Comput. 2013, 9, 5567−5577

Table 4. Lifetimes (ps) of the First Coordination Shell of the Central QM Water Molecule[a]

	DFTB/REAXFF		DFTB/TIP3P-Fs		PM6-DH+/REAXFF		PM6-DH+/TIP3P-Fs			
τ_{QM}	1.65		1.65		2.42		2.42			
τ_{MM}	2.22		2.05		2.22		2.05			
	DFTB/REAXFF		DFTB/TIP3P-Fs		PM6-DH+/REAXFF		PM6-DH+/TIP3P-Fs		ε_{RMSD}	
	active region radius (Å)									
	4.0	5.5	4.0	5.5	4.0	5.5	4.0	5.5	4.0	5.5
FIRES	2.29	1.90	2.52	2.48	**2.48**	**2.37**	**2.53**	**2.43**	0.42	0.29
ABRUPT	**1.74**	**1.61**	1.28	**1.24**	1.99	2.25	1.49	2.05	0.45	**0.25**
SAP	2.19	2.06	2.44							
DAS	1.87	1.40	**1.64**	**1.24**	2.34	2.20	2.10	1.97	**0.16**	0.33
BF	1.82	1.11	1.47	1.03	1.72	2.04	1.35	2.14	0.53	0.45

[a]The values in bold have the smallest deviation from the QM result. At the top the fully QM (τ_{QM}) and fully MM (τ_{MM}) values are given for comparison. Only limited data is available for the SAP simulations, because most of the simulations crashed before reaching 10 ps. On the right the average root mean square deviations from the QM values (ε_{RMSD}) are listed.

In conclusion, the temperature data suggest FIRES as the most rigorous method for dynamic data. However, it should be kept in mind that FIRES is able to correctly reproduce the dynamics for an artificial system with limited diffusion, which may differ considerably from a real solution. The second best performing method in terms of temperature conservation, and (with FIRES) the most stable method, is DAS, with the important advantage that it allows diffusion of solvent molecules throughout the system.

4.2.2. Lifetimes of the First Coordination Shell. A dynamic quantity of importance in solution chemistry is the lifetime of coordination of first and second solvation shells. We defined the first solvation shell of the central water molecule as consisting of those water molecules with an oxygen−oxygen distance to the QM water between 2.0 and 4.0 Å. The lifetime of this shell is expected to be larger than the lifetime of a single H-bond (τ_{HB}), which is 1.4 ps for TIP3P-Fs.[57] Because in many cases the NVE simulations resulted in very large temperature increases, the NVT simulations were used to compute the time-correlation functions $C(t)$ for the water residence in the first solvation shell (eq 3). The summation is over all water molecules, and h is a function that equals 1 if the water molecule is in the first solvation shell and 0 otherwise.

$$C(t) = \frac{\sum_w \langle h_w(0) h_w(t) \rangle}{\sum_w \langle h_w(0) h_w(0) \rangle} \tag{3}$$

In those cases where the temperature in the NVE run was stable we found that the autocorrelation functions obtained from the NVT simulations were only minimally affected by the presence of the thermostat (because the system was equilibrated). The lifetime τ of the water molecules in the first solvation shell was computed according to eq 4 over a period of 5 ps. For the purposes of this paper we do not need the converged value of the lifetime τ, because we use it solely as a quantity that describes the curve of the time-correlation function $C(t)$. The lifetimes we compare are therefore relative measures only. Decorrelation is slow with respect to our simulation time, and longer time correlation windows would lead to larger values of τ.

$$\tau = \frac{\int t C(t) \mathrm{d}t}{\int C(t) \mathrm{d}t} \tag{4}$$

Table 4 lists the lifetime values of the first solvation shell of water obtained from the fully QM and fully MM simulations, with DFTB producing the shortest lifetimes and PM6-DH+ the longest. This is in agreement with the findings in Figure 4, which indicate that DFTB results in the weakest water−water interaction, while PM6-DH+ has the strongest. Out of the QM/MM simulations with a small active region (4.0 Å radius), DAS shows the smallest deviation from the QM result as shown by the average RMSD value. The first solvation shell is located within the active region, directly at the border to the buffer zone (or MM environment). FIRES (and SAP, although the SAP data is limited) result in the longest lifetimes, due to the steep potential energy that the solvent experiences at the edge of the active region. This potential prevents the molecules in the first shell from diffusing away. The shortest lifetimes are reported for the BF simulations. The lack of energy conservation results in an increase in kinetic energy of the water molecules near the buffer zone, where the laws of physics are violated.

It should be noted that the massive thermostats recommended by the developers of BF may improve the results by relaxing the water molecules near the buffer zone. However, the BF method itself introduces a degree of randomization not only of the size but also of the direction of the motions. A thermostat can only correct the former error, leaving (or even further randomizing) the effect of the latter. The majority of the unphysical heating in the buffer zone stems from the intrinsic violation of Newtons third law in the interaction between water molecules in the active region and water molecule in the buffer zone. The use of different potentials to propagate two interacting water molecules can result in the transformation of vibrational motion across a H-bond to translational motion of the water pair, which would cause water molecules at the QM/MM boundary to diffuse away from their original position inside the first solvation shell. Even if a thermostat can improve results by reducing the translational motion, an unphysical effect on the lifetimes of the first solvation shell coordination will remain.

For those simulations with a larger active region (5.5 Å radius) the FIRES simulations result in somewhat shorter lifetimes (more diffusion) than reported with the smaller active region and a smaller RMSD error. This is related to the fact that the first solvation shell water molecules are now somewhat farther away from the potential energy wall at the QM/MM border, and more diffusion is allowed. Still, the values are generally larger than the results from the fully QM simulations. The ABRUPT simulations also show surprisingly good results.

dx.doi.org/10.1021/ct4005596 | J. Chem. Theory Comput. 2013, 9, 5567−5577

The DAS simulations result in a larger error, despite the fact that the solvation shell lies further from the buffer zone. This is due to the fact that the larger active region results in more water molecules crossing the buffer zone, thereby introducing heat into the system.

In conclusion, the results indicate that FIRES can give good results even for dynamic quantities involving diffusion, as long as the effect studied is deep enough inside the QM region. However, a note of caution should be placed with this observation, as it is difficult to a priori predict the effect of the artificial constraint at the QM/MM border. For dynamic effects related to molecules closer to the QM/MM border DAS may be the best choice.

5. CONCLUSION

We have comparatively evaluated five different advanced QM/MM methods for the description of solute−solvent systems. We used water in water as our model and compared structural parameters (RDFs) as well as dynamic quantities (temperature conservation and the lifetimes in the first solvation shell of the solute molecule). Of these three quantities we feel the RDFs are the most important, since for any other property (free energy differences for reactive processes, spectroscopic data, etc.) the structure of the system needs to be correctly represented. The best RDFs were produced with DAS, BF, and ABRUPT, with BF and ABRUPT requiring less computational time than DAS without apparent loss of accuracy in the structure. SAP produces very strong structural deviations at the QM/MM border. The quality of the RDFs produced with FIRES very strongly depends on the form of the QM and MM potential energy surfaces. This makes the performance of FIRES very hard to predict. In terms of temperature conservation FIRES is the best method, followed with some distance by DAS. If the intended result requires a proper description of diffusion in the system, especially if that diffusion happens close to the QM/MM interface, then DAS gives the best results, while the worst performance is given by BF. In reproducing dynamic quantities ABRUPT generally performs worse than DAS, but, considering its simple nature and the low expense of the simulations, this buffer zone-free method performs remarkably well.

Recent years have hence provided some very important advances in the QM/MM description of chemistry in solution. This paper demonstrates that we are currently able to properly describe the structure of complex solute−solvent systems. When it comes to describing dynamical quantities, which are related to phase-space sampling and thus can be important for the computation of reaction mechanisms and free energy differences, all methods introduce some artifacts near the QM/MM boundary. DAS minimizes the spurious motions, can lead to satisfactory results, and is infinitely more efficient than a full QM simulation. However, room for improvement remains in the QM/MM description of chemical reactivity in a diffusive environment.

■ ASSOCIATED CONTENT

Supporting Information

RMSD values for the RDFs with each QM/MM combination, the RDF graphs themselves, and the time-correlation functions for the residence time of water molecules in the first solvation shell. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: r.e.bulo@vu.nl

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The research has been supported by the European Community under a Marie Curie EIF (contract no. MEIF-CT-2004−011109). Calculations were performed using the local HPC resources of PSMN at ENS-Lyon.

■ ABBREVIATIONS

RDF, radial distribution function; QM/MM, quantum mechanical/molecular mechanical; RMSD, root-mean-square deviation

■ REFERENCES

(1) Simon, M.-O.; Li, C. J. *Chem. Soc. Rev.* **2012**, *41*, 1415−1427.
(2) Ruppert, A. M.; Weinberg, K.; Palkovits, R. *Angew. Chem. Int Ed.* **2012**, *51*, 2564−2601.
(3) Frassoldati, A.; Pinel, C.; Besson, M. *Catal. Today* **2011**, *173*, 81.
(4) Rideout, D. C.; Breslow, R. *J. Am. Chem. Soc.* **1980**, *102*, 7816.
(5) Narayan, S.; Muldoon, J.; Finn, M. G.; Fokin, V. V.; Kolb, H. C.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2005**, *44*, 3275.
(6) Arani Chanda, A.; Fokin, V. V. *Chem. Rev.* **2009**, *109*, 725−74.
(7) Butler, R. N.; Coyne, A. G. *Chem. Rev.* **2010**, *110*, 6302−6337.
(8) Narayan, S.; Muldoon, J.; Finn, M. G.; Fokin, V. V.; Kolb, H. C.; Sharpless, K. B. *Angew. Chem., Int. Ed.* **2005**, *44*, 3275.
(9) Azoulay, S.; Manabe, K.; Kobayashi, S. *Org. Lett.* **2005**, *7*, 4593.
(10) Portmann, R. World Patent, WO 9802423, 1998.
(11) Bulo, R. E.; Jacob, C. R.; Visscher, L. *J Phys. Chem. B* **2008**, *112*, 2640−2647.
(12) Trinh, T. T.; Rozanska, X.; Delbecq, F.; Sautet, P. *Phys. Chem. Chem. Phys.* **2012**, *14*, 3369−3380.
(13) Chibani, S.; Michel, C.; Delbecq, F.; Pinel, C.; Besson, M. *Catal. Sci. Technol.* **2013**, *3*, 339−350.
(14) Park, J. M.; Laio, A.; Ianuzzi, M.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 11318−11319.
(15) Michel, C.; Auneau, F.; Delbecq, F.; Sautet, P. *ACS Catalysis* **2011**, *1*, 1430.
(16) Shelton, D. P. *Chem. Phys. Lett.* **2000**, *325*, 513−516.
(17) Kresse, G.; Furtmuller, J. *Phys. Rev. B* **1996**, *54*, 11169.
(18) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. *Comput. Phys. Commun.* **2005**, *167*, 103−128.
(19) I.S. Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2009**, *5*, 2619.
(20) (a) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249. (b) Thole, B. T.; Van Duijnen, P. Th. *Theor. Chim. Acta* **1980**, *55*, 307−318. (c) Field, M. J.; Bash, P. A. *J. Comput. Chem.* **1990**, *11*, 700−733. (d) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VHC: New York, 1995; Vol. 7, pp 119−185; (e) Sherwood, P. In *Modern Methods and Algorithms of Quantum Computing*; Grotendorst, J., Eds.; John von Neumann Institute for Computing: Jülich, Germany, 2000; pp 257−277; Carloni, P.; (f) Rothlisberger, U.; Parrinello, M. *Acc. Chem. Res.* **2002**, *35*, 455. (g) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2008**, *4*, 2067−2084. (h) Magistrato, A.; DeGrado, W. F.; Laio, A.; Rothlisberger, U.; VandeVondele, J.; Klein, M. L. *J. Phys. Chem. B* **2003**, *107*, 4182−4188.
(21) Gao, J. L.; Xia, X. F. *J. Am. Chem. Soc.* **1993**, *115*, 9667−9675.
(22) Gao, J. L. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(23) Kerdcharoen, T.; Morokuma, K. *Chem. Phys. Lett.* **2002**, *355*, 257−262.
(24) (a) Kerdcharoen, T.; Liedl, K. R.; Rode, B. M. *Chem. Phys.* **1996**, *211*, 313−323. (b) Hofer, T. S.; Pribil, A. B.; Randolf, B. R.; Rode, B.

M. *J. Am. Chem. Soc.* **2005**, *127*, 14231−14238. (c) Schwenk, C. F.; Loeffler, H. H.; Rode, B. M. *J. Am. Chem. Soc.* **2003**, *125*, 1618−1624.
(25) Csanyi, G.; Albaret, T.; Payne, M. C.; De Vita, A. *Phys. Rev. Lett.* **2004**, *93* (17), 175503.
(26) Rowley, C. N.; Roux, B. *J. Chem. Theory Comp.* **2012**, *8*, 3526−3535.
(27) Heyden, A.; Lin, H.; Truhlar, D. G. *J. Phys. Chem. B* **2007**, *111*, 2231−3341.
(28) Pezeshki, S.; Lin, H. *J. Chem. Theory Comput.* **2011**, *7*, 3625−3634.
(29) Bulo, R. E.; Ensing, B.; Sikkema, J.; Visscher, L. *J. Chem. Theory Comput.* **2009**, *5*, 2212−2221.
(30) 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.
(31) (a) Rode, B. M.; Schwenk, C. F.; Tongraar, A. *J. Mol. Liq.* **2004**, *110*, 105−122. (b) Rode, B. M.; Hofer, T. S. *Pure Appl. Chem.* **2006**, *78*, 525−539. (c) Rode, B. M.; Hofer, T. S.; Randolf, B. R.; Schwenk, C. F.; Xenides, D.; Vchirawongk- win, V. *Theor. Chim. Acc.* **2006**, *115*, 77−85.
(32) Park, K.; Goetz, A. W.; Walker, R. C.; Paesani, F. *J. Chem. Theory Comput.* **2012**, *8*, 2868−2877.
(33) Caratzoulas, S.; Courtney, T.; Vlachos, D. G. *J. Phys. Chem. A* **2011**, *115*, 8816−8821.
(34) Fleurat-Lessard, P.; Michel, C.; Bulo, R. E. *J. Chem. Phys.* **2012**, *137*, 074111.
(35) Nielsen, S. O.; Bulo, R. E; Moore, P. E.; Ensing, B. *Phys. Chem. Chem. Phys.* **2010**, *12*, 12401−12414.
(36) In the limit of an infinitely thin buffer region, SAP and DAS are equivalent and give the ABRUPT scheme.
(37) 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.
(38) Korth, M. *J. Chem. Theory Comput.* **2010**, *6*, 3808−3816.
(39) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173−213.
(40) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1.
(41) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.
(42) Velde, G.; Bickelhaupt, F.; Baerends, E.; Fonseca-Guerra, C.; Van Gisbergen, S.; Snijders, J.; Ziegler, T. *J. Comput. Chem.* **2001**, *22* (22), 931.
(43) (a) Frenzel, J.; Oliveira A. F.; Jardillier, N.; Heine, T.; Seifert, G. Semi-relativistic, self-consistent charge Slater-Koster tables for density-functional based tight-binding (DFTB) for materials science simulations. *DFTB Method Web site*; TU-Dresden: Dresden, Germany, 2004−2009. (b) Frenzel, J.; Oliveira, A. F.; Duarte, H. A.; Heine, T.; Seifert, G. *Z. Anorg. Allg. Chem.* **2005**, *631*, 1267. (c) Guimarães, L.; Enyashin, A. N.; Frenzel, F.; Heine, T.; Duarte, H. A.; Seifert, G. *Nano* **2007**, *1*, 362. (d) Luschtinetz, R.; Oliveira, A. F.; Frenzel, J.; Joswig, J.; Seifert, G.; Duarte, H. A. *Surf. Sci.* **2008**, *602*, 1347. (e) Luschtinetz, R.; Frenzel, J.; Milek, T.; Seifert, G. *J. Phys. Chem. C* **2009**, *113*, 5730.
(44) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.
(45) Bahn, S. R.; Jacobson, K. W. *Comput. Sci. Eng.* **2002**, *4*, 56−66.
(46) http://libatoms.org/.
(47) Lin, H.; Zhang, Y.; Pezeshki, S.; Truhlar, D. G. *QMMM*; version 1.4.0.CO; University of Minnesota: Minneapolis, MN, 2012.
(48) 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.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, 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, CA, 2012.
(49) Wu, Y.; Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2006**, *124*, 024503.
(50) Rahaman, O.; van Duin, A. C. T.; Goddard, W. A., III; Doren, D. J. *J. Phys. Chem. B* **2011**, *115*, 249−261.
(51) Soper, A. K. *ISRN Phys. Chem.* **2012**, 279463.
(52) Potestio, R.; Fritsch, S.; Espanol, P.; Delgado-Buscalioni, R.; Kremer, K.; Everaers, R.; Donadio, D.; *Phys. Rev. Lett. in press.*
(53) A PM6-DH+ simulation of a droplet (no predefined volume) with the same number of water molecules (110) as the periodic simulations produced an RDF around the central oxygen closely resembling the QM/MM result.
(54) Laio, A.; VandeVondele, J.; Rothlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941−6947.
(55) Lin, H.; Truhlar, D. *Theor. Chem. Acc.* **2007**, *117*, 185−199.
(56) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. *Chem. Phys. Lett.* **2006**, *418*, 245−249.
(57) Wu, Y.; Tepper, H. L.; Voth, G. A. *J. Chem. Phys.* **2006**, *124*, 024503.
(58) Aradi, B.; Hourahine, B.; Frauenheim, Th. *J. Phys. Chem. A* **2007**, *111* (26), 5678.