

Chapter 2

Proton Transfer in Aqueous Solution: Exploring the Boundaries of Adaptive QM/MM

T. Jiang, J.M. Boereboom, C. Michel, P. Fleurat-Lessard
and R.E. Bulo

Abstract In this chapter, we review the current state-of-the-art in quantum mechanical/molecular mechanical (QM/MM) simulations of reactions in aqueous solutions, and we discuss how proton transfer poses new challenges for its successful application. In the QM/MM description of an aqueous reaction, solvent molecules in the QM region are diffusive and need to be either constrained within the region, or their description (QM versus MM) needs to be updated as they diffuse away. The latter approach is known as adaptive QM/MM. We review several constrained and adaptive QM/MM methods, and classify them in a consistent manner. Most of the adaptive methods employ a transition region, where every solvent molecule can continuously change character (from QM to MM, and vice versa), temporarily becoming partially QM and partially MM. Where a conventional QM/MM scheme partitions a system into a set of QM and a set of MM atoms, an adaptive method employs multiple QM/MM partitions, to describe the fractional QM character. We distinguish two classes of adaptive methods: Discontinuous and continuous. The former methods use at most two QM/MM partitions, and cannot completely avoid discontinuities in the energy and the forces. The more recent continuous adaptive methods employ a larger number of QM/MM partitions for a given configuration. Comparing the performance of the methods for the description of solution chemistry, we find that in certain cases the low-cost constrained methods are sufficiently accurate. For more demanding purposes, the continuous adaptive schemes provide a

T. Jiang · J.M. Boereboom · R.E. Bulo
Inorganic Chemistry and Catalysis Group, Debye Institute for Nanomaterials Science,
Utrecht University, Universiteitsweg 99, 3584CG Utrecht, The Netherlands
e-mail: R.E.Bulo@uu.nl

C. Michel · P. Fleurat-Lessard (✉)
Laboratoire de Chimie de L'ENS de Lyon, Université de Lyon,
46 Allée d'Italie, 69364 Lyon Cedex 7, France
e-mail: Paul.Fleurat-Lessard@ens-lyon.fr

P. Fleurat-Lessard
Université de Bourgogne, Institut de Chimie Moléculaire de L'Université de
Bourgogne (ICMUB), 9 Avenue Alain Savary, 21078 Dijon, France

good balance between dynamical and structural accuracy. Finally, we challenge the adaptive approach by applying it to the difficult topic of proton transfer and diffusion. We present new results, using a well-behaved continuous adaptive method (DAS) to describe an alkaline aqueous solution of methanol. Comparison with fully QM and fully MM simulations shows that the main discrepancies are rooted in the presence of a QM/MM boundary, and not in the adaptive scheme. An anomalous confinement of the hydroxide ion to the QM part of the system stems from the mismatch between QM and MM potentials, which affects the free diffusion of the ion. We also observe an increased water density inside the QM region, which originates from the different chemical potentials of the QM and MM water molecules. The high density results in locally enhanced proton transfer rates.

2.1 Introduction

Proton transfer is one of the unifying elements that (almost) all aqueous chemical reactions have in common. The computational study of aqueous reactions, already difficult due to the long distance effects in this highly structured solvent, [1, 2] is further complicated by this challenging phenomenon. Acid-base reactions involve proton transfer to or from a reactant compound, often across an extended chain of water molecules (Grotthuss mechanism) [3]. This means that in these processes many bonds break and form nearly simultaneously, causing the reaction to be less local than reactions in organic media [4]. Recent simulations on the solvated hydroxide ion have further shown that the proton transfer motion is strongly dependent on water wires linking the OH^- to the rest of the solution [5]. Global effects, like changes in the electric field or compression of the wires, can be the cause of proton transfer events [6]. If the wires compress, the proton can even hop several water molecules along, in a nearly concerted process.

Aqueous reactions such as those described above involve constant breaking and forming of bonds, and can therefore best be described using electronic structure methods, preferably including nuclear quantum effects [7, 8]. However, this approach is costly, due to the long time scales on which the reactions occur. A multi-scale quantum mechanical (QM)/molecular mechanical (MM) approach assumes a localized reactive region, and thereby makes a QM description of a reaction feasible. In most cases bond breaking or forming can indeed be described as local (with the possible exception of the proton transfer events discussed above), but even for local reactions a multi-scale description is not straightforward. In standard QM/MM approaches [9–15], the nature (QM or MM) of each particle is defined initially and remains fixed during the whole simulation. This can lead to problems in dealing with solvent molecules in solution chemistry, which are diffusive by nature. Let us consider a QM/MM molecular dynamics (MD) simulation of a solvated system. The initial QM particles comprise a solute (such as a sodium or a hydroxide ion) and its first solvation shell. However, after a few tens of picoseconds, the solvent molecules of the solvation shell are exchanged with bulk solvent molecules, leading to an incongruous simulation in

which the first solvation shell contains MM solvent molecules, while QM solvent molecule have diffused into the MM bulk. To overcome this diffusivity problem two classes of methods have emerged. The most direct approach constrains the QM molecules to remain within a certain region (henceforth denoted by ‘active region’), while the MM molecules are constrained to their respective part of space (‘environment region’). A more involved approach allows the nature of the solvent molecules to change from QM to MM and vice versa as they move in and out of the active region. This approach is called adaptive QM/MM [16–18]. Among the adaptive methods, some lead to discontinuities in the forces exerted on the atoms, while more complex ones manage to keep a continuous description. This book chapter aims at giving a comprehensive overview of the main constrained and adaptive methods currently available.

The adaptive QM/MM methods have already provided insights into several processes in water [19–21]. Local proton transfer events have been studied with conventional QM/MM [22], as well as with adaptive QM/MM [23]. However, none of the previous applications considered more global proton transfer and diffusion into the aqueous solution. A hydroxide ion in water can in principle exit the predefined QM region that is usually centered around a reactive solute. In this book chapter we address the issues involved with the description of such proton diffusion events. More specifically, we will consider the solvation of a methanol molecule in alkaline water, with the aim to understand the values and limitations of adaptive QM/MM.

This chapter is organized as follows. In Sect. 2.2 the most dominant diffusive QM/MM methods aimed at aqueous reactions are reviewed, first the constrained, and then the adaptive methods. Their performances are compared, based on published results, in Sect. 2.3. In Sect. 2.4, one of the methods is applied to describe proton transfer in solution, and the performances of the method are discussed based on structural and dynamical data. Finally, in Sect. 2.5, the general conclusions of this work are briefly summarized.

2.2 Review of QM/MM Methods for the Simulation of Chemistry in Solution

In this section the available QM/MM methods designed for solution chemistry are reviewed, and the relevant terms are introduced. As we focus on describing a reactive process in solution at the atomistic level, multi-scale methods resorting to coarse-graining will not be described and we refer the interested readers to recent reviews [24–26]. Similarly, methods targeting gas phase reactions only are omitted [27, 28].

All the QM/MM methods dedicated to treat solution chemistry are based on the definition of two zones: (i) an active region that is generally defined as a sphere around the reactive center and (ii) an environment region that embeds the active region. This is illustrated in Fig. 2.1. The active region generally includes the reactive center (the orange disk in Fig. 2.1) and at least the first coordination sphere and is described at the QM level. QM molecules will always be shown in

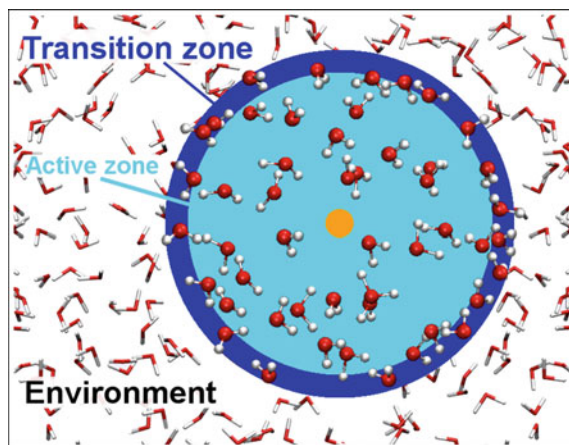


Fig. 2.1 Partitioning of a simulation box into an active zone (with light *blue* background), an environment zone (*white* background) and for some methods a transition zone (with *blue* background). The active center is symbolized with an *orange* disk. QM molecules are shown in *Ball&Stick* while MM water molecules are shown with *lines*

Ball&Stick in the Figures. The environment region includes the bulk solvent molecules described at the MM level, and will be depicted with lines. Each solvent molecule is either in the active zone, or in the transition region or in the environment zone. The QM/MM border does not cut through solvent bonds, even though methods that allow it have been developed [29].

As mentioned in the introduction, there are two main classes of methods. The first class constrains the QM molecules to remain inside the active region, while simultaneously keeping the MM molecules in the environment region (outside the active region). The solvent molecules cannot diffuse across the boundary between the active region and the environment. The most prominent examples of this approach will be discussed in Sect. 2.2.1. The second class of methods discussed does allow the solvent molecules to diffuse across the QM/MM boundary, changing the nature (QM vs. MM) of those molecules on the fly. This change can be instantaneous [16, 30–32] (Sect. 2.2.2: Discontinuous adaptive QM/MM), or smooth [33–35] (Sect. 2.2.3: Continuous adaptive QM/MM).

Including the continuity referred to above, there are three defining characteristics of the adaptive QM/MM methods that reveal much about their performance:

1. Continuity of the energy and/or the forces;
2. Energy conservation;
3. Momentum conservation.

The methods are classified according to these definitions in Table 2.1. By definition none of the discontinuous methods are energy conserving, but even among the continuous methods only some fall into this class. All categories have advantages as well as disadvantages, and these will be addressed in the following subsections.

Table 2.1 Classification of the various methods

	Continuous	Energy conserving	Momentum conserving
Abrupt [30]	✗	✗	✓
ONIOM-XS [30]	✗	✗	✓
LOTF [31]	✗	✗	✓
Hot-spot [16]	✗	✗	✗
BF [32]	✗	✗	✗
PAP [33]	✓	✓	✓
SAP [33]	✓	✓	✓
DAS [34]	✓	✗	✓
SCMP [35]	✓	✗	✓

They are classified according to three features: Continuity of the energy and/or forces, energy conservation, and momentum conservation

It is worth emphasizing that all constrained and adaptive QM/MM methods described in this chapter are fundamentally QM/MM simulations and will thus experience the same problems, all stemming from the fact that two different potentials are applied to describe the same system. The bulk of the problem originates in the interaction between QM and MM particles, and as a result, artifacts are found mainly at the QM/MM boundary. Several different schemes exist for the description of the interaction between QM and MM particles in conventional QM/MM [18, 36–38]: (i) mechanical embedding, (ii) electrostatic embedding, (iii) polarizable embedding, and (iv) flexible embedding [39, 40]. The most straightforward choice is mechanical embedding, in which one simply uses the MM (pair)-potential to describe the interaction between the QM and the MM molecules. Electrostatic embedding includes the partial charges of the MM atoms in the QM Hamiltonian, thereby describing part of the interaction using the QM description and allowing the MM environment to polarize the electrons in the QM region. Polarizable embedding is a less used option that employs a polarizable force-field in combination with electrostatic embedding, thereby allowing the QM region to in turn polarize the MM region. The mutual polarization then needs to be solved in an iterative manner. The last scheme, flexible embedding, is very recent, and allows the total charges of the active and environment regions to adapt during the course of the simulation, effectively allowing charge transfer between the two regions. In this chapter we will not go into the advantages and disadvantages of these options, but they have been recently reviewed and compared [18]. Suffice it to say that all of them have certain advantages and disadvantages, and none of them are exempt from structural and dynamical artifacts at the boundary. Nonetheless, all QM/MM methods discussed in this chapter can be straight-forwardly combined with either mechanical, electrostatic or polarizable embedding.

Another important decision in conventional QM/MM lies in the choice of the size of the QM region (the further away the boundary is from the reactive site, the better). The same holds for adaptive QM/MM simulations, and the possible options as to

where to locate the QM/MM boundary (distance to the QM center, number of active solvent molecules, or density based) will be discussed at the end of this section.

2.2.1 Constrained QM/MM

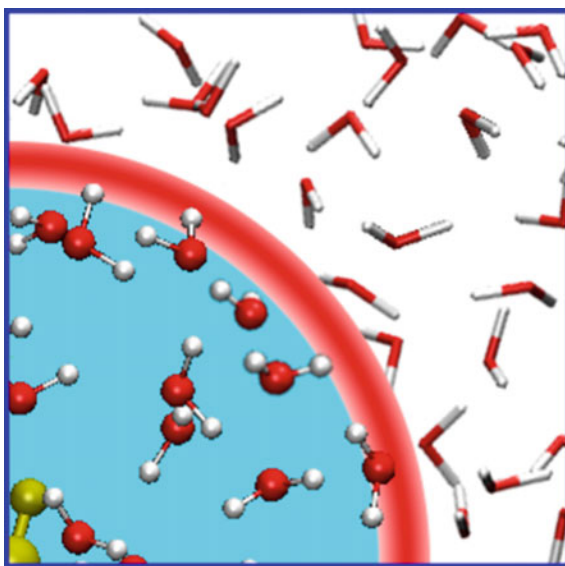
In this section we will discuss two constrained QM/MM methods: the Flexible Inner Region Ensemble Separator (FIRES) method, [41] and the recent Boundary based on Exchange Symmetry Theory (BEST) method [42].

Flexible Inner Region Ensemble Separator (FIRES)

The FIRES method is an elegant example of the constrained approach. It prevents the solvent molecules from crossing the QM/MM boundary, but does not require the volume of the QM region to be determined in advance. 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. This is illustrated in Fig. 2.2. 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. Such a simulation can be viewed as having the active region placed in a balloon immersed in the environment. FIRES calculations are as efficient as conventional QM/MM since the overhead of the computation of the forces keeping the two sets of molecules apart is minimal.

While the imposed constraint alters the dynamics with respect to an unconstrained system, the authors argue that if the constraint is infinitely steep, average values of a property $X(\mathbf{r})$ (i.e. the distance between two reactants, the number of

Fig. 2.2 FIRES method. The constraining potential is schematized as a *red* ring



hydrogen bonds to a molecule, etc.) remain unaltered by the constraint. If we define the partition function Z for a system consisting of a central reactant a and N_s equivalent solvent molecules,

$$Z = \frac{1}{N_s!} \int d\mathbf{r}_a \int e^{-V(\mathbf{r})/k_B T} d\mathbf{r}^{N_s}, \quad (2.1)$$

with $V(\mathbf{r})$ the potential energy of the system. We can define the configurational average of $X(\mathbf{r})$,

$$\langle X \rangle = \frac{\int d\mathbf{r}_a \int X(\mathbf{r}) e^{-V(\mathbf{r})/k_B T} d\mathbf{r}^{N_s}}{\int d\mathbf{r}_a \int e^{-V(\mathbf{r})/k_B T} d\mathbf{r}^{N_s}}, \quad (2.2)$$

In the following, we will prove that providing the constrained particles are identical and $V(\mathbf{r})$ and $X(\mathbf{r})$ are invariant under exchange of two particles, constraining two sets of particles to different regions of space will not affect the value of $\langle X \rangle$. If we define a central molecule (the reactant a), and then partition our system into a set S_n of n molecules and a set of S_m of ($m = N_s - n$) molecules, we can rewrite the partition function as,

$$Z = \frac{1}{N_s!} \int d\mathbf{r}_a \int d\mathbf{r}^n \int d\mathbf{r}^m e^{-V(\mathbf{r})/k_B T}. \quad (2.3)$$

We can select a subset of configurations contributing to the integral, where the molecules in S_n are the n solvent molecules closest to the reactant a (which we would like to assign QM character). They occupy a flexible region $A(\mathbf{r})$ generally defined as the region of space containing the n solvent molecules closest to a . Then, the MM molecules in the set S_m occupy a further part of space ($E(\mathbf{r})$).

$$Z' = \int d\mathbf{r}_a \frac{1}{n!} \int^{A(\mathbf{r})} d\mathbf{r}^n \frac{1}{m!} \int^{E(\mathbf{r})} d\mathbf{r}^m e^{-V(\mathbf{r})/k_B T}. \quad (2.4)$$

The contributions to the integral Z that we discarded in Z' are simply those where solvent molecules in S_n and in S_m exchanged position. Since all solvent molecules are identical, the discarded contributions to the integral Z are equivalent to the ones in Z' (Eq. 2.4), and the two expressions differ only by a multiplicative constant factor C : $Z' = Z/C$.

This separation is exact, and implementing the integral of Eq. (2.4) into Eq. (2.2) has the factors C cancel in the numerator and denominator, yielding an unchanged thermodynamic average $\langle X \rangle$, provided that the molecules in the active and environment region do not interchange [41].

To ensure that QM and MM molecules do not exit their regions, an infinitely steep constraint can be used. However, in practice, such a constraint would lead to divergent molecular dynamics. Therefore, a soft constraint is used, typically a simple half-harmonic potential that exerts a repulsive force on any MM molecules

as they approach the active region $A(\mathbf{r})$. As this constraint is not infinitely steep, it might happen that a MM molecule sneaks into the active region $A(\mathbf{r})$ and thus becomes closer to the QM center than a QM molecule. As a result, the averages computed with Eq. (2.4) are no longer exact. The authors have shown that an aqueous system retains the correct structure in a test MM/MM simulation [41]. However, it has been shown that when two different potentials are used for the active zone and the environment, even equilibrium properties can be wrong at the boundary such as the radial distribution of oxygen atoms around the core QM water oxygen [17, 42].

Boundary Based on Exchange Symmetry Theory (BEST)

The soft harmonic constraint used in the FIRES method changes the partition function, resulting in slightly distorted equilibrium structures at the QM/MM boundary. The BEST method aims to solve this problem [42]. In this section, we illustrate the BEST method for the simplest example of only two particles i and j (belonging to S_n and S_m respectively) that violate the requirement in Eq. (2.4): the molecules in S_m cannot be closer to reactant a than any of the molecules in S_n . The partition function Z' in Eq. (2.4) can be rewritten as,

$$Z' = \frac{1}{n!m!} \int d\mathbf{r}_a \int^{A(\mathbf{r})} d\mathbf{r}_i \int^{E(\mathbf{r})} d\mathbf{r}_j \int^{A(\mathbf{r})} d\mathbf{r}_{rest}^n \int^{E(\mathbf{r})} d\mathbf{r}_{rest}^m e^{-V(\mathbf{r})/k_B T}. \quad (2.5)$$

Because FIRES uses a soft wall, additional configurations are explored, where particles i and j have swapped positions. Because the soft wall prevents these configurations from occurring too often, they have a relatively low probability. So they appear with a fractional weight $f(\mathbf{r}_i, \mathbf{r}_j)$ in the computed FIRES partition function Z'' :

$$\begin{aligned} Z'' &= \frac{1}{n!m!} \int d\mathbf{r}_a \int^{A(\mathbf{r})} d\mathbf{r}_i \int^{E(\mathbf{r})} \mathbf{r}_j \int^{A(\mathbf{r})} d\mathbf{r}_{rest}^n \int^{E(\mathbf{r})} d\mathbf{r}_{rest}^m e^{-V(\mathbf{r})/k_B T} \\ &+ \frac{1}{n!m!} \int d\mathbf{r}_a \int^{E(\mathbf{r})} d\mathbf{r}_i \int^{A(\mathbf{r})} d\mathbf{r}_j f(\mathbf{r}_i, \mathbf{r}_j) \int^{A(\mathbf{r})} d\mathbf{r}_{rest}^n \int^{E(\mathbf{r})} d\mathbf{r}_{rest}^m e^{-V(\mathbf{r})/k_B T} \end{aligned} \quad (2.6)$$

The partition function Z'' thus differs from the exact one (Z) by something other than a constant. This leads to inexact average values (Eq. 2.1). The authors of BEST recognized that the partition function Z' in Eq. (2.5) can be recovered by additional weighting of the first term of Eq. (2.6).

$$f(\mathbf{r}_i, \mathbf{r}_j) + f'(\mathbf{r}_i, \mathbf{r}_j) = 1 \quad (2.7)$$

$$\begin{aligned}
Z' = & \frac{1}{n!m!} \int d\mathbf{r}_a \int^{A(\mathbf{r})} d\mathbf{r}_i \int^{E(\mathbf{r})} d\mathbf{r}_j f'(\mathbf{r}_i, \mathbf{r}_j) \int^{A(\mathbf{r})} d\mathbf{r}_{rest}^n \int^{E(\mathbf{r})} d\mathbf{r}_{rest}^m e^{-V(\mathbf{r})/k_B T} \\
& + \frac{1}{n!m!} \int d\mathbf{r}_a \int^{E(\mathbf{r})} d\mathbf{r}_i \int^{A(\mathbf{r})} d\mathbf{r}_j f(\mathbf{r}_i, \mathbf{r}_j) \int^{A(\mathbf{r})} d\mathbf{r}_{rest}^n \int^{E(\mathbf{r})} d\mathbf{r}_{rest}^m e^{-V(\mathbf{r})/k_B T}.
\end{aligned}
\tag{2.8}$$

In BEST, this formulation was extended to include all possible exchanges of solvent molecules between the QM and the MM sets. This approach does, however, become very laborious, and the recommended implementation assigns weights $f(\mathbf{r}_i, \mathbf{r}_j)$ to only the most frequent exchange contributions and neglects the rest.

2.2.2 Discontinuous Adaptive QM/MM

The list of discontinuous adaptive QM/MM methods discussed here counts five members: (1) Abrupt, [30] (2) ONIOM-XS, [30] (3) Hot Spot, [16] (4) Learn On The Fly (LOTF), [31] and (5) Buffered Force (BF) [32]. In QM/MM simulations for chemistry in solution, the molecular system is generally partitioned into a subset of QM solvent molecules S_{QM} and a subset of MM solvent molecules S_{MM} (see also Sect. 2.2.1). In order to account for the diffusivity of the solvent molecules, the QM/MM partition needs to be adjusted during the course of the simulation to always describe the solvent molecules in the active region as QM. As pointed out in the introduction to this section, none of the discontinuous methods are energy conserving. In addition, Hot Spot and BF do not conserve momentum, thereby breaking Newton's third law, which states that the force a molecule exerts on its neighbor is equal and opposite to the force the other molecule exerts in return. At first glance this seems an odd setup, and to clarify this, we address the different reasons why this option can be advantageous in the paragraphs below.

2.2.2.1 Abrupt

The simplest and most intuitive adaptive QM/MM simulation involves an abrupt switch of description as a solvent molecule crosses a given cut-off radius. This corresponds to a sudden change in the QM/MM partition, as the molecule is reassigned to a different set (QM or MM). For each configuration, an energy $V^{(a)}(\mathbf{r})$ is defined, which corresponds to the energy of the desired QM/MM partition with label a . In this chapter, this approach will be denoted by Abrupt.

Due to the difference between the QM and the MM potential, an instantaneous change of the partition is accompanied by a sudden change in the potential energy $V^{(a)}(\mathbf{r})$, and in the forces on the atoms. As a result, the total energy of a simulation is not conserved, which is demonstrated by an acceleration of the molecules at the

QM/MM boundary [17, 30]. Strong thermostats are required to correct for this acceleration. Although the energy is not conserved, at each configuration the forces on all atoms are defined as the negative gradient of $V^{(a)}(\mathbf{r})$. As a result, the total force on the system equals zero, and the momentum is conserved.

2.2.2.2 ONIOM-XS

This method aims to introduce a certain degree of continuity into the switch from one QM/MM partition to the next (which accompanies the change in the QM character of the molecules). ONIOM-XS is still classified as discontinuous, as some degree of discontinuity remains. The sudden changes occurring in Abrupt are diminished through introduction of a transition region between the active region and the environment, where the solvent molecules can semi-continuously switch between QM and MM character. As a consequence, the molecules in that region have partial QM and partial MM character. We define a transition region variable T as a set of two distances R_s and R_e ($T = \{R_s, R_e\}$, with $R_s < R_e$), denoting the start and the end of the transition region. The partial QM character of the solvent molecules in the transition region is reflected in a quantity $\lambda(\mathbf{r}_i)$, which is a switching function that depends on the position of the molecule i in the transition region. In ONIOM-XS, the switching function $\lambda(\mathbf{r}_i)$ is a polynomial function that equals 1 inside the QM zone (for $r_i < R_s$), 0 inside the environment (for $R_e < r_i$) and smoothly switch between these two values inside the transition zone:

$$\lambda(\mathbf{r}_i : T) = 6\left(x_i - \frac{1}{2}\right)^5 - 5\left(x_i - \frac{1}{2}\right)^3 + \frac{15}{8}\left(x_i - \frac{1}{2}\right) + \frac{1}{2}, \quad (2.9)$$

with $x_i = (r_i - R_s)/(R_e - R_s)$.

Another, more common choice for a continuous switching function in adaptive methods is,

$$\lambda(\mathbf{r}_i : T) = \begin{cases} 1 & \text{if } r_i < R_s, \\ \frac{(R_e - r_i)^2(2r_i + R_e - 3R_s)}{(R_e - R_s)^3} & \text{if } R_s \leq r_i \leq R_e, \\ 0 & \text{if } R_e < r_i. \end{cases} \quad (2.10)$$

The switching function $\lambda(\mathbf{r}_i : T)$ is a function of the distance r_i between the center of the QM core and the position of (a representative atom in) the solvent molecule i . The value of $\lambda(\mathbf{r}_i : T)$ decreases smoothly from 1 to 0 if the molecule diffuses from the center of the QM core across the transition region (see Fig. 2.3).

As illustrated in Fig 2.4, at each time-step, two QM/MM partitions a and b are computed. In partition a , only molecules in the active region are included in the set $S_{QM}^{(a)}$. In partition b , the set $S_{QM}^{(b)}$ also contains the molecules in the transition region.

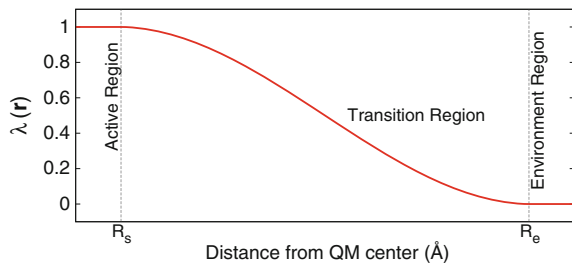


Fig. 2.3 Value of the switching function $\lambda(\mathbf{r})$ from Eq. (2.10) over the transition region

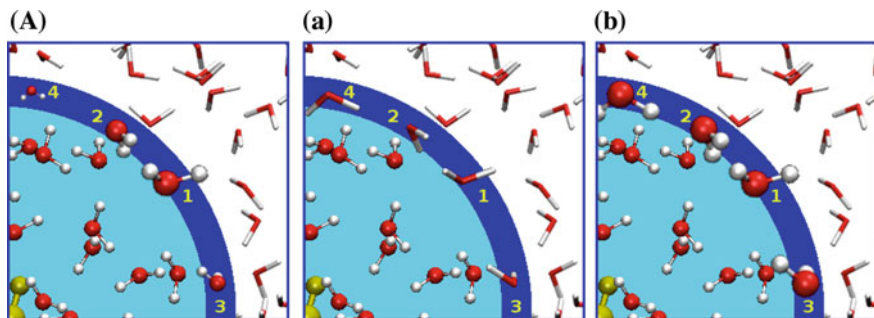


Fig. 2.4 Schematic illustration of ONIOM-XS method. (A) ONIOM-XS energy with (partially) QM solvent molecules in *Ball&Stick*, while molecules with *lines* are MM. Size of the *Ball&Stick* QM molecules in the transition zone illustrate their percentage of QM character; (a) Small QM/MM partition: $S_{QM}^{(a)}$ contains only solvent molecule in the active zone; (b) Large QM/MM partition: $S_{QM}^{(b)}$ contains solvent molecules in the active and transition regions

The energy is then defined as a weighted average of the two computed partition energies,

$$V^{XS} = \sigma^{(a)}(\mathbf{r})V^{(a)}(\mathbf{r}) + \sigma^{(b)}(\mathbf{r})V^{(b)}(\mathbf{r}). \quad (2.11)$$

The functions $\sigma^{(a)}(\mathbf{r})$ and $\sigma^{(b)}(\mathbf{r})$ are weights of the two computed partition energies, and $\sigma^{(a)}(\mathbf{r}) + \sigma^{(b)}(\mathbf{r}) = 1$. If the M molecules in the transition region define a set S_T , the value of $\sigma^{(b)}(\mathbf{r})$ is defined as the average over the QM character ($\lambda(\mathbf{r}_i : T)$ values) of all M molecules in S_T ,

$$\sigma^{(b)}(\mathbf{r}) = \frac{1}{M} \sum_{i \in S_T}^M \lambda(\mathbf{r}_i : T). \quad (2.12)$$

When only one molecule is in the transition region, the change from QM to MM (and vice versa) is continuous. However, when the transition region contains more

than one molecule, the method still involves small jumps when solvent molecules exit or enter the transition region. At that point, one of the partitions (a or b) is replaced, and the set of M solvent molecules in Eq. (2.12) changes. As a result of the discontinuity this method does not conserve energy, but because the forces on all atoms are the negative gradient of the energy defined in Eq. (2.11), the total momentum is conserved.

2.2.2.3 Hot Spot

The Hot Spot method, like ONIOM-XS, is based on the introduction of a transition region between the active region and the environment region. The “Hot Spot” corresponds to the combined active and transition region. Again, like ONIOM-XS, Hot Spot considers two partitions for each conformation, as illustrated in Fig. 2.5. In partition θ , the whole system is MM, leading to the energy $V^{(0)}$. In partition a , the solvent molecules of the Hot Spot are QM. Originally, partition a contained

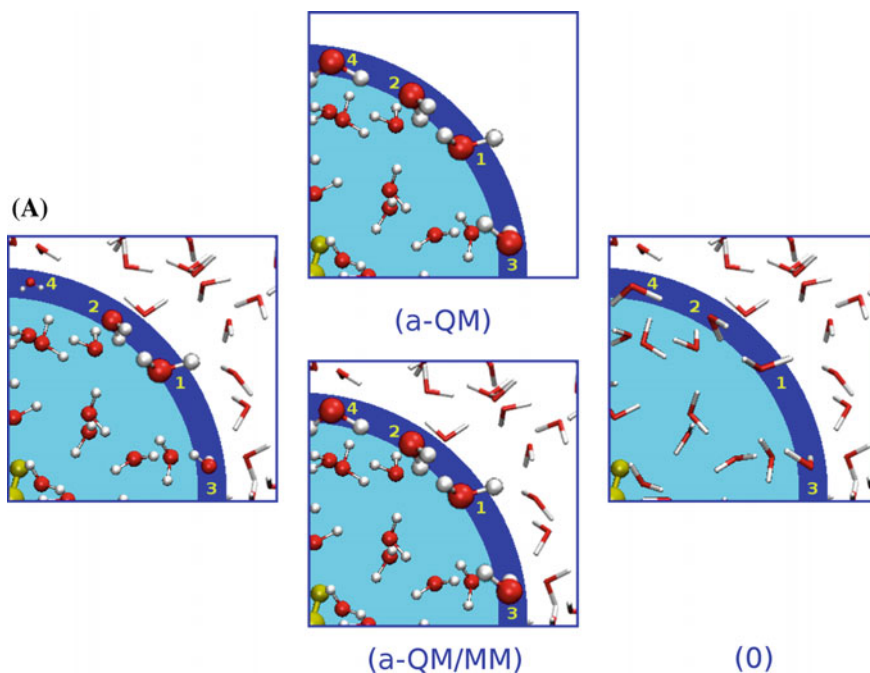


Fig. 2.5 Illustration of Hot Spot method. (A) Forces applied to each molecule. *Ball&Stick* molecules are assigned forces from the QM/MM calculations, while molecules with *lines* are assigned MM forces. Sizes of the *Ball&Stick* QM molecules in the transition zone illustrate their percentage of QM character. (a) QM or QM/MM simulation to get the force for the molecules in the active and transition zones; (0) MM calculation to get the forces for the molecules in the transition zone and in the environment

only these molecules, [16] but recent applications used a QM/MM energy for this partition: molecules in active region and the transition change are QM while the environment is MM [43]. The molecules in the active region feel the forces computed for partition a at the QM or QM/MM level. The molecules belonging to the environment region feel the forces obtained from the MM calculation. And last, the molecules in the transition region feel a weighted average of the two forces. The force exerted on a molecule i in Hot Spot writes:

$$\mathbf{F}_i^{HS}(\mathbf{r}) = -\lambda(\mathbf{r}_i : T) \frac{\partial V^a(\mathbf{r})}{\partial \mathbf{r}_i} - (1 - \lambda(\mathbf{r}_i : T)) \frac{\partial V^{(0)}(\mathbf{r})}{\partial \mathbf{r}_i}. \quad (2.13)$$

In this approach, no unique energy can be defined that is consistent with the force on all atoms, and Newton's third law is violated. This happens because the force that a molecule exerts on a neighboring molecule is no longer equal to the negative force the second molecule exerts in return, since it comes from a different calculation. This type of approach has a continuity similar to an ONIOM-XS type simulation, at a reduced cost, since only one QM/MM calculation per conformation is required. Like ONIOM-XS, the forces in this method are still discontinuous, which results in an energy gain (or loss) during the simulation. As a consequence of the violation of Newton's third law, the total force acting on the system is not zero. This results in an additional energy flux into the system, so that the simulations require even stronger thermostats than Abrupt to maintain the desired temperature.

A very large part of the accelerations of the atoms in the Hot Spot simulations comes from the lack of momentum conservation. A way to reduce this effect to the level of ONIOM-XS, without requiring an additional QM/MM computation, is to fit a simple force field to the Hot Spot forces for the entire system on the fly (for each conformation). This results in a unique energy expression $V^{FF}(\mathbf{r})$, which has approximately the Hot Spot forces as its gradient. Then, the forces used for the molecular dynamics propagation of the system are taken as the negative gradient of the fitted potential $V^{FF}(\mathbf{r})$.

A method such as this has been developed under the name Learn On The Fly (LOTF) [31]. As $V^{FF}(\mathbf{r})$ is fitted at each time step, it is a discontinuous method that does not conserve the total energy. It behaves similarly to a Hot Spot simulation, but with the added advantage that it conserves momentum. Indeed, the forces on all atoms come from a unique potential $V^{FF}(\mathbf{r})$ so that the total force is zero.

2.2.2.4 Buffered Force (BF)

BF, like Hot Spot, obtains the forces assigned to different atoms from different calculations, as schematized in Fig. 2.6. As a consequence, no unique potential energy is defined, and Newton's third law is violated. The total momentum of the simulations will therefore not be conserved and the method requires an efficient thermostat to be used. There is, however, a major distinction between Hot Spot and

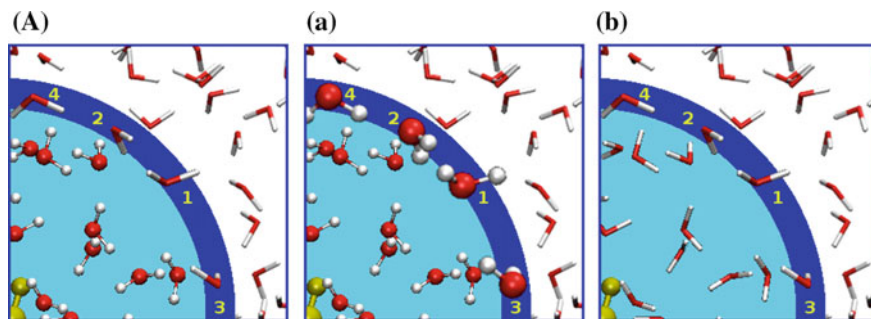


Fig. 2.6 Schematic illustration of BF method. (A) Forces applied to each molecule. *Ball&Stick* molecules are assigned forces from the QM/MM calculations, while molecules with *lines* are assigned MM forces; (a) Nature of the solvent molecules in the QM/MM simulation; (b) Nature of the solvent molecules in the MM calculation

BF, in that they are designed to solve different problems. One general problem with QM/MM, whether conventional or designed for diffusive systems, is that particles at the boundary interact with other particles of both QM and MM character, and may prefer one over the other. This may result in either a density increase, or a density depletion at the boundary, which would not be present in a fully QM simulation. This artifact may be reduced somewhat in the above-mentioned ONIOM-XS, Hot Spot, and LOTF, if the transition region is chosen sufficiently large, but in the BF method this issue is tackled in a more thorough manner.

Like Hot Spot and LOTF, BF computes one QM/MM energy $V^{(b)}(\mathbf{r})$ for a partition b with a large set of QM molecules ($S_{QM}^{(b)}$ in ONIOM-XS) and a fully MM energy $V^{(0)}(\mathbf{r})$ for the entire system. In a manner analogue to the Hot Spot method, the QM/MM forces are then assigned to all molecules in a small predefined active region ($S_{QM}^{(a)}$ in ONIOM-XS), while the MM forces are assigned to all the rest.

In BF, there are no molecules that have partial QM and partial MM character, exactly as in the Abrupt approach. In that sense, there is no transition region, but we can define a buffer region as $S_{QM}^{(b)} - S_{QM}^{(a)}$. The choice of transition region thickness is carefully selected based on the error in the forces on the QM reactive center, and amounts to values below 1 Å, similar to the sizes used in other methods. The considerable advantage over the Abrupt approach is that in BF the QM molecules at the boundary feel forces from a calculation that describes all their neighbors QM, while the MM molecules feel forces from a calculation that describes all their neighbors MM. As a result, unwanted clustering or depletion is avoided.

During a molecular dynamics simulation most molecules exhibit Brownian motion. Therefore, close to the boundary, a wandering molecule could switch between QM and MM character very often. To avoid this effect, the authors of BF introduced hysteresis in the QM \leftrightarrow MM change: Four boundaries are defined in total. A molecule going out of the active region will be assigned MM forces beyond the R_s^{out} radius whereas a molecule coming towards the active region will be

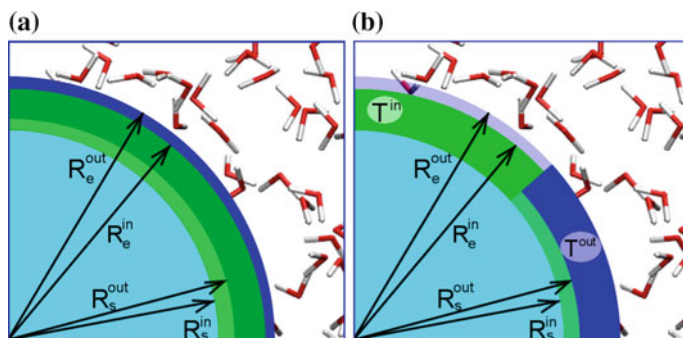


Fig. 2.7 Schematic hysteresis approach in the BF method. (a) Definition of the four boundaries; (b) Four boundaries seen as two buffer regions. T^{in} is shown as a *green* ring, T^{out} is shown in *blue*. For the sake of clarity, they are only partially shown

assigned QM forces below the R_s^{in} radius, with $R_s^{in} \leq R_s^{out}$. Similarly, R_e^{in} and R_e^{out} are introduced as the outer boundaries of the buffer region. These four boundaries can be grouped into two buffer regions defined by $T^{in} = \{R_s^{in}, R_e^{in}\}$ and $T^{out} = \{R_s^{out}, R_e^{out}\}$. This is schematized on Fig. 2.7.

2.2.3 Continuous Adaptive QM/MM

In the previous subsection, we presented several methods that aim to provide some continuity in the QM/MM transition of solvent molecules (ONIOM-XS, Hot Spot, LOTF). In this subsection, we discuss a set of more recent QM/MM methods that ensure a completely continuous transition: (1) Permuted Adaptive Partitioning (PAP), [33] (2) Sorted Adaptive Partitioning (SAP), [33] (3) Difference-based Adaptive Solvation (DAS), [34] and (4) Size-Consistent Multi-Partitioning (SCMP) [35]. Similar to the discontinuous methods from Sect. 2.2.2, a transition region is again introduced, in which the solvent molecules are assigned fractional QM (and MM) character based on their position. This is illustrated in Fig. 2.8a. As opposed to the discontinuous schemes that require only two different QM/MM partitions to be computed each time step, the continuous methods require calculation of a large set of partitions of the solvent molecules. This can be rationalized as follows: to assign a partial QM (and MM) character to a single molecule i , one should launch two calculations: one in which i is QM and one in which it is MM. This is illustrated in Fig. 2.8b, c for the second solvent molecule in the transition region. Then, an adaptive potential energy expression can be defined that is a weighted average of the energies for the two partitions (equivalent to Eq. (2.11) for ONIOM-XS). As all molecules in the transition region can have different QM characters, an adaptive potential energy expression that assigns the desired fractional QM characters to all

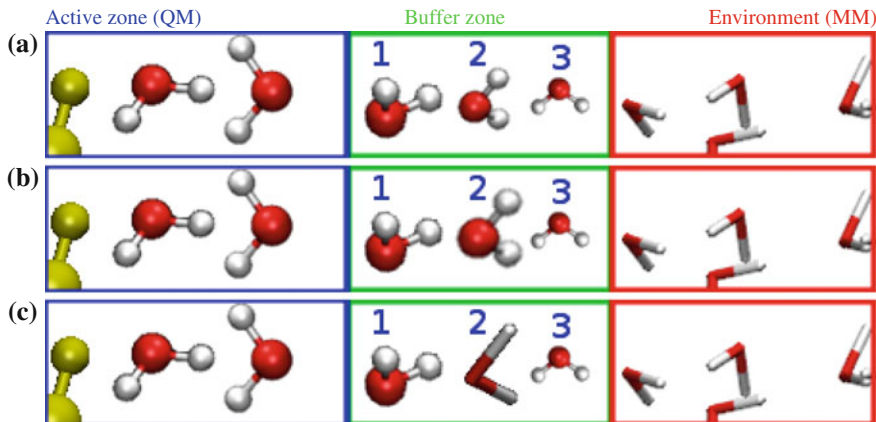


Fig. 2.8 Schematic illustration of adaptive methods. (a) Continuous evolution of the nature of solvent molecule from QM (shown as *Ball&Stick*) to MM (shown as *lines*). The percentage of QM nature is schematized by the size of the *Ball&Stick* drawing. (b) and (c) Two contributing partitions that determine the QM/MM character of water 2: (b) Water 2 is QM, (c) Water 2 is MM

these molecules needs to include contributions from many QM/MM partitions. All of the methods define this adaptive potential energy $V^{ad}(\mathbf{r})$ as,

$$V^{ad}(\mathbf{r}) = \sum_n \sigma^{(n)}(\mathbf{r}) V^{(n)}(\mathbf{r}). \quad (2.14)$$

For a system with M solvent molecules in the transition region, the maximum number of partitions that can meaningfully contribute to this partial character corresponds to the 2^M possible partitions of the M molecules into two sets. The behavior of the weight functions $\sigma^{(n)}(\mathbf{r})$ is such that when a solvent molecule m_i moves away from the QM center, the weight $\sigma_{(n)}(\mathbf{r})$ of a partition n that describes m_i QM decreases, becoming zero as m_i exits the transition region and enters the environment region.

As mentioned, all continuous methods have the same basic definition for the potential energy (Eq. 2.14). The most important difference between the continuous methods lies in the form of the weight-functions $\sigma^{(n)}(\mathbf{r})$, where the PAP method assigns them in perhaps the most intuitive manner. However, the PAP method also requires computation of a large number of computationally expensive QM/MM partitions for each conformation, since it includes all 2^M possible partitions, with M the number of solvent molecules in the transition region. SAP, DAS and SCMP are considerably more economic alternatives that include only the most important contributions, typically $M + 1$.

There is, however, a less defining choice in how to perform a continuous adaptive molecular dynamics simulation. Molecular dynamics works on the premise that the forces (the negative gradient of the energy) are a good estimate for $\Delta V^{ad}(\mathbf{r})/\Delta \mathbf{r}$ over a time-step Δt . Only then can the correct kinetic energy for the

next time step be predicted, and will the total energy be conserved. The forces are only a good estimate for $\Delta V^{ad}(\mathbf{r})/\Delta \mathbf{r}$ if the potential energy is continuous, which it is in all continuous methods. In adaptive approaches, however, there are two ways in which the forces on the atoms can be obtained from $V^{ad}(\mathbf{r})$, and the choice determines whether or not the simulation conserves energy.

- (i) Energy conserving: The forces on the atoms are the negative gradients of the energy in Eq. (2.14) according to,

$$\mathbf{F}_i^{ad}(\mathbf{r}) = - \sum_n \sigma^{(n)}(\mathbf{r}) \frac{\partial V^{(n)}(\mathbf{r})}{\partial \mathbf{r}_i} - \sum_n \frac{\partial \sigma^{(n)}(\mathbf{r})}{\partial \mathbf{r}_i} V^{(n)}(\mathbf{r}). \quad (2.15)$$

- (ii) non-Energy conserving: The forces on the atoms are again the negative gradient of the energy in Eq. (2.14), but now the terms that describe the change of $\sigma^{(n)}(\mathbf{r})$ are neglected, resulting in the definition,

$$\mathbf{F}_i^{ad'}(\mathbf{r}) = - \sum_n \sigma_{\mathbf{r}}^{(n)} \frac{\partial V^{(n)}(\mathbf{r})}{\partial \mathbf{r}_i}. \quad (2.16)$$

The forces in Eq. (2.16) provide an inexact prediction for $\Delta V^{ad}(\mathbf{r})$, resulting in simulations that do not conserve energy.

Of the methods discussed below, PAP and SAP historically used the weighted energy from Eq. (2.15), and thus are energy conserving. DAS and SCMP are generally used in a non-energy conserving approach, directly computing the forces from Eq. (2.16).

As stated above, the four methods below differ mainly in the form of the weight functions, and as long as the weight functions have continuous derivatives (PAP, SAP, DAS), they can equally well be used in an energy conserving as in a non-energy conserving manner. While the former has the advantage of being formally more sound, it was demonstrated that it does not produce reasonable structures with either the DAS or the SAP weight functions [34]. As a result, the use of the approach of Eq. (2.16) is recommended for all practical purposes, as was recently confirmed by Lin et al. [18]. While the resulting simulations do not conserve energy, one can make use of the fact that the forces are completely continuous, and integrate the forces *a posteriori* along the trajectory, to obtain a quantity that is conserved. This so-called bookkeeping quantity can be used with Eq. (2.16) and any of the available weight functions, to test simulation parameters such as the size of the transition region or the selected time step [34, 44].

2.2.3.1 Permuted Adaptive Partitioning (PAP)

This method includes all possible partitions of the transition region solvent molecules into two sets (QM and MM) contributing to the total energy in Eq. (2.14). If the number of solvent molecules in the transition region equals M at a certain time-step, then the maximum number of non-zero PAP partitions equals 2^M .

To obtain the weight functions σ of these partitions, the switching function λ is required each time-step for all M transition region molecules according to Eq. (2.3).

Each partition n in Eq. (2.14) consists of a set of QM solvent molecules $S_{QM}^{(n)}$ and a set of MM solvent molecules $S_{MM}^{(n)}$, which together sum to the total number of solvent molecules in the system. We can now define QM and MM fade-out functions for each partition as,

$$O_{QM}^{(n)}(\mathbf{r}) = \prod_{i \in S_{QM}^{(n)}} \lambda(\mathbf{r}_i : T) \quad , \quad O_{MM}^{(n)}(\mathbf{r}) = \prod_{i \in S_{MM}^{(n)}} 1 - \lambda(\mathbf{r}_i : T). \quad (2.17)$$

The fade-out functions earn their name because $O_{QM}^{(n)}(\mathbf{r})$ becomes zero when a QM solvent molecule diffuses into the environment, while the fade-out function of the set of MM molecules becomes zero when a MM molecule penetrates the QM region. The partition weight functions $\sigma^{(n)}(\mathbf{r})$ are defined as products of the fade-out functions,

$$\sigma^{(n)}(\mathbf{r}) = O_{QM}^{(n)}(\mathbf{r}) \cdot O_{MM}^{(n)}(\mathbf{r}). \quad (2.18)$$

The sum of the weights in Eq. (2.18) over all contributing partitions equals one.

The number of contributing partitions is often very large (the method scales as $O(2^M)$, see Fig. 2.9), but many of the weights are negligible, and in practice computation time is saved by computing only the partitions that have a non-negligible weight.

2.2.3.2 Sorted Adaptive Partitioning (SAP)

The SAP method drastically reduces the number of contributing partitions by constructing the weight functions $\sigma^{(n)}(\mathbf{r})$ in such a way that only $M + 1$ ‘ordered’ partitions contribute. The ‘ordered’ partitions are those in which all QM solvent molecules are closer to the QM core than the MM molecules (Fig. 2.10). This intrinsically means that each partition has a different number of QM solvent molecules. The main criterion for the weight functions $\sigma^{(n)}(\mathbf{r})$ in Eq. (2.14) is that when two solvent molecules are at similar distance from the QM core, then those partitions that describe one particle QM and the other particle MM have weights that smoothly approach zero. The expressions for the weight functions are fairly complex, and have been extensively discussed elsewhere [33, 34].

2.2.3.3 Difference-Based Adaptive Solvation (DAS)

Similar to SAP, in the DAS method only the ‘ordered’ partitions contribute. The weight functions used in DAS have a relatively simple form,

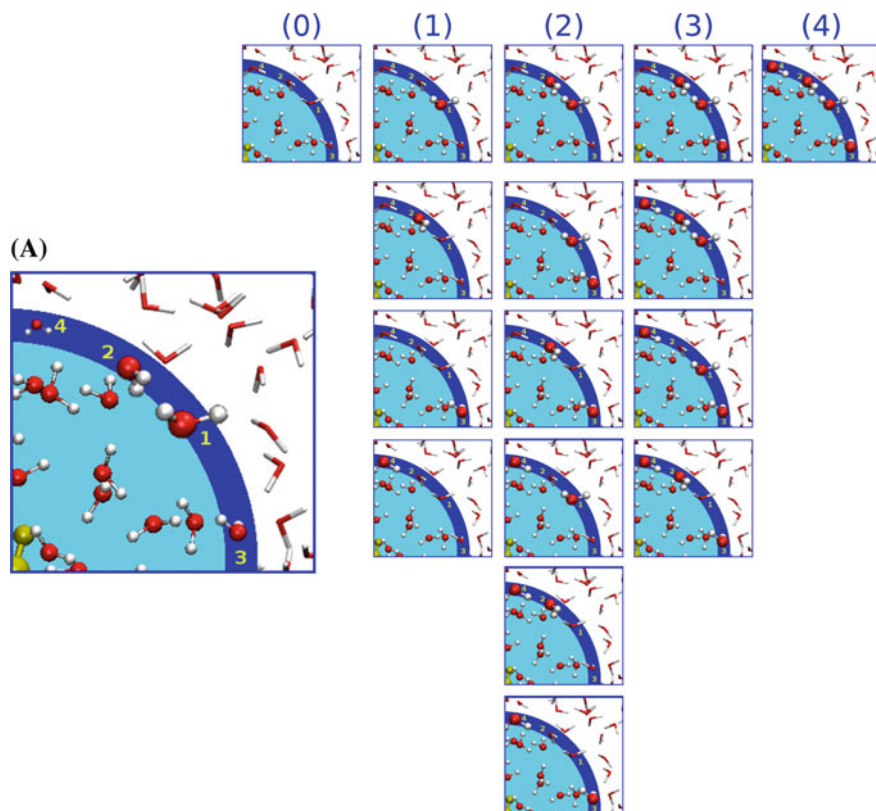


Fig. 2.9 Illustration of the PAP method for a methanol molecule (in the bottom left corner in yellow) solvated in water. QM molecules are shown in *Ball&Stick* while MM molecules are shown with *lines*. (A) QM characters of solvent molecules in a PAP simulation: In the transition zone, the size of a molecule indicates its proportion of QM character. On the *right*, all the computed partitions are shown, ordered by the number of QM solvent molecule, ranging from 0 to 4 in this example

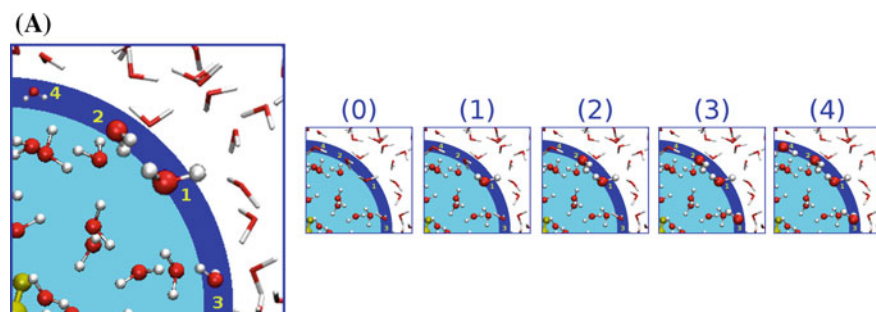


Fig. 2.10 SAP or DAS simulation (A) the percentage of QM nature is schematized by the size of the *Ball&Stick* drawing. In SAP, these weights are used to compute the energy, while in DAS they are used to compute the forces. On the *right*, the ‘ordered’ partitions are shown. As indicated in the text, they all have a different number of QM molecules, ranging from 0 to 4 in this example

$$\sigma^{(n)}(\mathbf{r}) = \max\left(\left\{\min\left(A_{MM}^{(n)}(\mathbf{r} : T)\right) - \max\left(A_{QM}^{(n)}(\mathbf{r} : T)\right), 0\right\}\right). \quad (2.19)$$

with $A_{MM}^{(n)}(\mathbf{r} : T)$ and $A_{QM}^{(n)}(\mathbf{r} : T)$ defined as sets of switching functions $\lambda(\mathbf{r}_i : T)$,

$$\begin{aligned} A_{QM}^{(n)}(\mathbf{r} : T) &= \left\{\lambda(\mathbf{r}_i : T) \mid i \in S_{QM}^{(n)}\right\}, \\ A_{MM}^{(n)}(\mathbf{r} : T) &= \left\{\lambda(\mathbf{r}_i : T) \mid i \in S_{MM}^{(n)}\right\}. \end{aligned} \quad (2.20)$$

The functions $\sigma^{(n)}(\mathbf{r})$ in Eq. (2.19) are effectively differences of $\lambda(\mathbf{r}_i : T)$ functions, contrary to PAP and SAP, which define $\sigma^{(n)}(\mathbf{r})$ as products of $\lambda(\mathbf{r}_i : T)$. As a result, the gradients of the weight functions in DAS are never greater than the gradients of the $\lambda(\mathbf{r}_i : T)$ functions, and as a result, a DAS simulation conserves the total energy better than PAP and SAP when a large time-step is applied. While the gradients of the DAS weight functions are not large, they do suffer from small discontinuities, due to the discrete maximum and minimum functions in Eq. (2.19).

Contrary to PAP and SAP, DAS is most often applied in combination with the non-Energy conserving approach. In this approach, the expression for the forces (Eq. (2.16)) does not contain the gradient of the weight functions $\sigma^n(\mathbf{r})$, and the above-mentioned discontinuities pose no problems. However, when the weights in Eq. (2.19) are used in combination with the energy conserving approach of Eq. (2.15), this discontinuity has to be dealt with to achieve energy conservation in the simulation. In that situation, continuous maximum and minimum functions can be used,

$$\begin{aligned} \max(A) &= \frac{1}{k} \ln\left(\sum_{a \in A} e^{ka}\right), \\ \min(A) &= 1 - \left(\frac{1}{k} \ln\left(\sum_{a \in A} e^{k(1-a)}\right)\right). \end{aligned} \quad (2.21)$$

where A is a set of fractional numbers between 0 and 1, and k is a large constant, typically $k = 250$. The function $\max(A)$ is a function that returns the maximum element out of the set A , and $\min(A)$ is a function that returns the minimum. When these continuous functions are applied in Eq. (2.19), the sum of the weights of the contributing partitions slightly deviates from normalization but the error is less than 0.01.

2.2.3.4 Size-Consistent Multi-partitioning (SCMP)

The Size-Consistent Multi Partitioning (SCMP) method is a scheme where the number of contributing partitions N is predefined by the user, and remains the same at every time-step. The partitions themselves are defined such that each partition

treats the same number (N_{QM}) of molecules quantum mechanically (QM), as illustrated on Fig. 2.11. One can note here that SCMP includes only partitions that are in the same column in Fig. 2.9, while in SAP and DAS the contributing partitions are in the same row. Since each partition energy takes roughly the same time to calculate, this scheme is able to run efficiently in parallel.

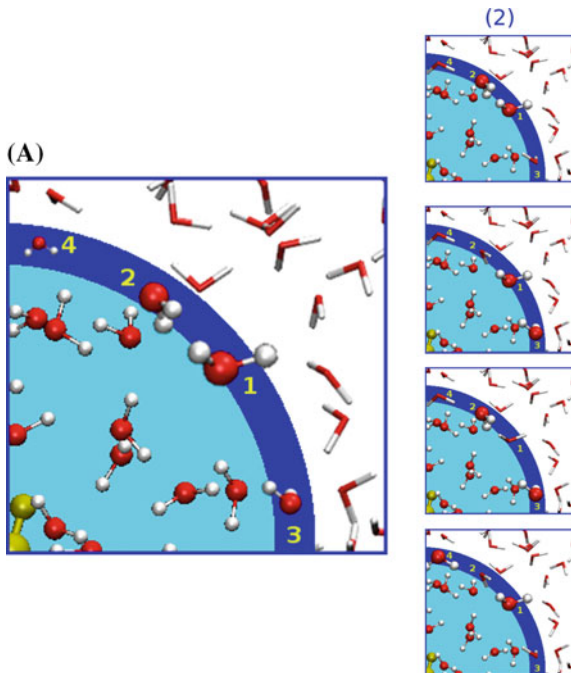
In order to obtain the weight functions of SCMP, again, at each time-step λ switching functions are required for all molecules. The SCMP method defines four different transition regions labeled T_{out}^{QM} , T_{in}^{QM} , T_{out}^{MM} , and T_{in}^{MM} . Each molecule has thus four λ values associated to it. The analogue of the PAP fade-out functions in SCMP is

$$O_{QM}^{(n)}(\mathbf{r}) = \prod_{i \in S_{QM}^{(n)}} \lambda(\mathbf{r}_i : T_{out}^{QM}), \quad O_{MM}^{(n)}(\mathbf{r}) = \prod_{i \in S_{MM}^{(n)}} 1 - \lambda(\mathbf{r}_i : T_{out}^{MM}). \quad (2.22)$$

In addition, the SCMP method introduces fade-in functions according to,

$$I_{QM}^{(n)}(\mathbf{r}) = 1 - \prod_{i \in S_{QM}^{(n)}} \lambda(\mathbf{r}_i : T_{in}^{QM}), \quad I_{MM}^{(n)}(\mathbf{r}) = 1 - \prod_{i \in S_{MM}^{(n)}} 1 - \lambda(\mathbf{r}_i : T_{in}^{MM}). \quad (2.23)$$

Fig. 2.11 Example of an SCMP simulation in which $N = 4$ partitions are considered with $N_{QM} = 2$ QM water molecules. (A) Forces computed for the molecules. The fraction of QM character is schematized by the size of the *Ball&Stick* drawing. On the right, the ‘ordered’ partitions are shown



Here, $I_{QM}^{(n)}(\mathbf{r})$ and $I_{MM}^{(n)}(\mathbf{r})$ are the fade-in functions of partition n for the QM and MM molecules, respectively. Finally, the weight function of partition n is defined as,

$$\sigma^{(n)}(\mathbf{r}) = \frac{O_{QM}^{(n)}(\mathbf{r}) \cdot O_{MM}^{(n)}(\mathbf{r}) \cdot I_{QM}^{(n)}(\mathbf{r}) \cdot I_{MM}^{(n)}(\mathbf{r})}{\sum_n^N O_{QM}^{(n)}(\mathbf{r}) \cdot O_{MM}^{(n)}(\mathbf{r}) \cdot I_{QM}^{(n)}(\mathbf{r}) \cdot I_{MM}^{(n)}(\mathbf{r})}, \quad (2.24)$$

where the sum of all weight functions is normalized to 1.

At each time-step an updating algorithm is used that ensures that significant partitions (partitions where QM solvent molecules are close to the center) are contributing to the forces in Eq. (2.16). Although less intuitive, the fade-in functions are used because they ensure that for the most-compact partition (the partition where all the QM molecules are closer to the center than the MM molecules) either $I_{QM}^{(n)}(\mathbf{r}) = 0$ or $I_{MM}^{(n)}(\mathbf{r}) = 0$, leading to a zero weight for this partition. As a result, when one replaces the least significant partition with the most-compact one, the energy and the forces are continuous during the exchange. For the precise algorithm of SCMP simulations, see Ref. [35].

Exploiting the fact that the calculations run efficiently in parallel, the SCMP method allows one to consider a fairly large number of partitions (typically on the order of 48 instead of around 10 for DAS or SAP), leading to a very smooth change in QM character of molecules inside of the transition region.

2.2.4 Choosing the Boundary Position in Adaptive QM/MM

On top of the standard QM/MM issues, to define the shape and size of the active region is also a challenge in both the constrained and adaptive approaches. Up to now, all methods define a spherical active region with a center and a radius. Most works center the active zone on a particular atom. This approach can straightforwardly be extended to a spherical region around the center of mass of a set of atoms. In a recent contribution, Lin et al. showed that in simulations of biomolecules, it can be desirable to use a fixed point in space as the center of the QM region. In order to retain conservation of momentum, they assigned a very heavy virtual atom as the QM center [45].

The size of the active region has been determined in three different ways: Distance-based (Distance-Adaptive Multi-Scale: DAMS), number-based (Number-Adaptive Multi-Scale: NAMS) [19, 46, 47], and density-based (Density-Based Adaptive QM/MM: DBA) [48]. The size-selection methods can in theory all be used in combination with any of the above adaptive QM/MM approaches.

Distance and Number-Based Adaptive Multi-Scale (DAMS and NAMS) DAMS is the original and most often used approach, where the radius of the active sphere remains fixed during the simulation. Csányi and co-workers define the ideal fixed radius as the distance where the description of the solvent molecules no longer affects the forces on the central atom. The number of solvent molecules in the active region is adjusted on the fly.

Conversely, in NAMS the number of solvent molecules in the active zone is constant. The QM/MM boundary is adjusted to the density of solvent molecules inside the active region, and the radius of the active region is updated on the fly.

Density Based Adaptive QM/MM (DBA)

The DBA approach is an elegant and systematic way to adaptively choose the size of the active region. It has been applied in combination with the Abrupt method. In this approach, the QM/MM boundary is not an empirically selected distance from a QM region, but is based on the overlap of electron densities between a subset of molecules at the core of the active region, with the rest of the solvent molecules. To have a fast method, the molecular electron density of a molecule is pragmatically defined as the sum of the atomic densities. The electron density and the reduced density gradient are used to determine if the interaction between two molecules can be defined as covalent or not. In practice, this amounts to a maximum distance of around 3.9 Å between molecules. This is comparable to the commonly chosen distance of 4 Å in some distance-based schemes. An advantage of DBA over the distance-based scheme is that orientation plays a role in determining the QM character of the water molecule. As a result, the number of QM molecules can be slightly smaller than in DAMS with a comparable cut-off distance, making DBA computationally more favorable. The DBA approach has thus far been applied only in geometry optimizations, but extension to molecular dynamics (even coupling to a continuous adaptive method) would be relatively straightforward.

2.3 Method Comparison

While no general overview paper exists that compares all methods discussed above on the same grounds, several contributions have appeared that compare subsets of methods. In 2007, Heyden et al. compared the performance of Abrupt, Hot Spot, ONIOM-XS, PAP, and SAP for a system of argon in argon with two different classical (MM) potentials [33]. The observables compared were energy conservation, temperature conservation, and radial distribution functions. In 2013, the authors of this chapter compared the performance of Abrupt, SAP, DAS, BF, FIRES, and BF for a system of water in water with four permutations of two QM and two MM potentials [17]. They compared the results obtained with the various QM/MM simulations with a reference calculation, full QM. In 2015, Pezeshki et al. reviewed ONIOM-XS, PAP, SAP, DAS, Hot-Spot and BF [18]. However, they did not perform any new method comparison with respect to Ref. [33]. Below, we will summarize the results of these studies, separated into a section about dynamical performance (energy/temperature conservation, time correlation functions), a section about structural results, and a section about timings. The comparison is summarized in Table 2.2.

Table 2.2 Performance of the various methods. They are classified following the previous section

	$E^{(a)}$	Momentum ^(b)	Structure ^(c)	$N^{(d)}$
FIRES [41]	✗	✓	✗	1
BEST [42]	✗	✓	✓	1
Abrupt [30]	✗	✓	✓	1
ONIOM-XS [30]	✗	✓	✓	2
LOTF [31]	✗	✓	✓	1
Hot-spot [16]	✗	✗	✓	1
BF [32]	✗	✗	✓	1
PAP [33]	✓	✓	✗	$\approx 2^M$
SAP [33]	✓	✓	✗	$M + 1$
DAS [34]	✗	✓	✓	$M + 1$
SCMP [35]	✗	✓	✓	User defined

The criteria of comparison are: (a) conservation of the total energy in a simulation in the microcanonical (NVE) ensemble (b) conservation of the total momentum (c) quality of radial distribution functions extracted from an MD simulation in the canonical ensemble (d) scaling in terms of number of QMMM partitions, M being the number of molecules in the transition region

2.3.1 Dynamical Performance

The comparative study by Heyden et al. [33] revealed that regarding the energy conservation, the quality of the performance of the methods can be ordered $\text{PAP} = \text{SAP} > \text{ONIOM-XS} > \text{Abrupt} > \text{Hot Spot}$ (with a simulation time-step of 0.1 fs) as shown in Fig. 2.12. This is in agreement with expectation, since the definition of the potential energy in PAP and SAP is designed to be continuous, so that energy conservation is possible—provided a small enough time-step is applied. While PAP and SAP are continuous and employ the energy conserving approach (Eq. 2.15), the ONIOM-XS potential energy (Eq. 2.11) is discontinuous, thereby sacrificing energy conservation. On the other hand, the energy in ONIOM-XS is considerably more continuous than the energy in the Abrupt approach, and indeed the latter performs worse in energy and temperature conservation. The Hot Spot method, besides its discontinuity, breaks Newton’s third law, which further reduces the performance in energy conservation.

The second comparative study [17] listed only temperature conservation, ranking the tested methods as $\text{FIRES} > \text{DAS} > \text{Abrupt} > \text{BF}$. The ranking is again in agreement with expectation. The FIRES method applies forces that are exact derivatives of the potential energy, which only has minor discontinuities, and is therefore the best at total energy (and thus temperature) conservation. The three adaptive methods, DAS, Abrupt, and BF, are all combined with the ‘non-energy conserving’ approach. Still, DAS conserves the energy better since it is a continuous approach while Abrupt and BF have a completely discontinuous (sudden) energy definition. BF (like Hot Spot) furthermore violates the Newton’s third law, which is why the temperature conservation is even worse.

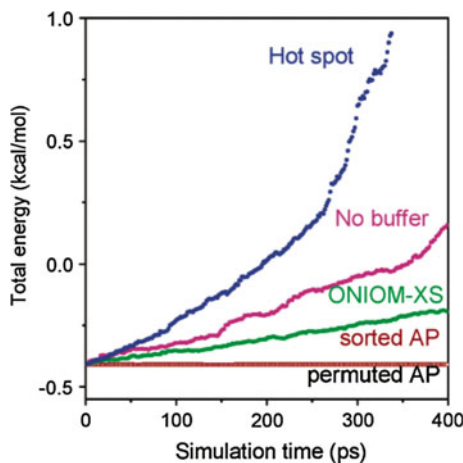


Fig. 2.12 Total energy during an MD simulation in the microcanonical ensemble. The system consists of 171 argon atoms in a cubic periodic box with a length of 20 Å. One atom is chosen to be the active center, and the radius of the active zone is 5 Å. Trajectories are computed using the hot spot (*blue*), ONIOM-XS (*green*), PAP (*red*, permuted AP), and SAP (*black*, sorted AP) methods with a transition region 0.5 Å thick. The ‘No buffer’ *pink* line corresponds to an Abrupt simulation within the nomenclature of this chapter. Total energy data from the simulation using the PAP method are directly underneath the data from the SAP simulation. Reproduced from Ref. [33] with permission

Reference [17] also computed the residence time of the water molecules in the first coordination shell of the central QM water. This dynamical quantity will clearly be affected by any errors in the equations of motion, and this can be seen from the ranking of the adaptive methods, which is the same as specified above for temperature conservation (DAS > Abrupt > BF). Being a constrained (non-diffusive) method, FIRES is not expected to properly reproduce a quantity that is determined by diffusion. Surprisingly, FIRES performs exceptionally well when the first solvation shell is buried deep enough in the active (QM) region, that is when the transition zone starts beyond the second solvation shell.

To sum up, only few adaptive methods conserve the energy. However, there is a very strong ranking in the quality of the dynamical performance: constrained > adaptive continuous > adaptive non-continuous.

2.3.2 Recovery of Structures

The results for structure recovery reveal a ranking that is almost opposite to the ranking found for dynamical performance. In Ref. [33], the radial distribution of the argon atoms with respect to the central argon atom were all very similar, and presumably of good quality, especially when a thermostat was applied in the

simulations. The oxygen-oxygen radial distribution function around the central QM water oxygen in Ref. [17] showed that the most reliable structures were obtained from the simulations using Abrupt, BF and DAS, with a small preference for BF (Fig. 2.13). The structures obtained with SAP strongly deviated from the reference full QM simulation. The FIRES radial distributions showed a large density peak at the QM/MM boundary, which was larger if the difference between the QM and MM potential was larger. We expect that this density peak is partially related to the problem with the partition function that was addressed by the BEST approach. Using electrostatic QM/MM coupling, the authors of the BF method demonstrated that the introduction of the buffer is crucial to recover reasonable structural properties. On two systems (water in water and an anion Cl^- in water), they compare the BF radial distribution functions measured at the center of the QM region with the full QM, the full MM and the adaptive QM/MM method referred to as Abrupt in this chapter (see Fig. 2.14). Without introduction of the buffer, the radial distribution is strongly affected by the adaptive QM/MM boundary with the apparition of a peak at the boundary. With the introduction of the buffer, the QM radial distribution is recovered in the active zone and the spurious peak disappears.

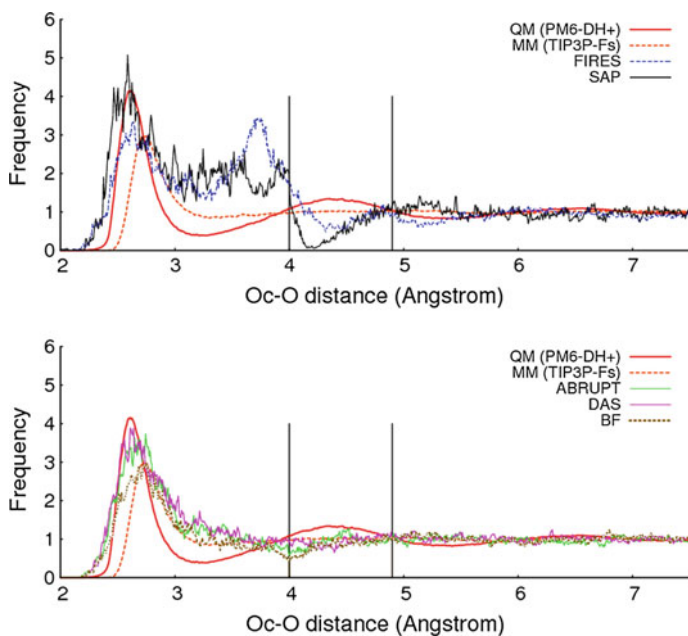


Fig. 2.13 Oc-O radial distribution function (RDF) (Oc is central oxygen) for water from QM/MM simulations (PM6-DH +/TIP3P-Fs) with an active region up to 4.0 Å (in FIRES simulation the active region is ≈ 4 Å). *Top* (FIRES, SAP) and *bottom* (Abrupt, BF and DAS). The vertical lines represent the inner and outer borders of the transition region. Adapted from Ref. [17] with permission

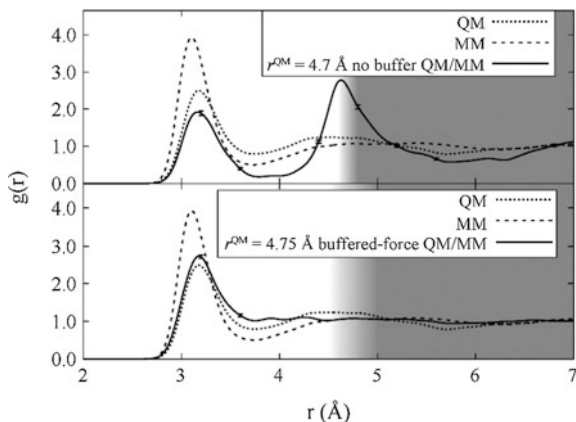


Fig. 2.14 Cl-O radial distribution function $g(r)$ for a chlorine anion solvated in water in a QM/MM (BLYP/fTIP3P) simulation. The Cl^- anion is the active center. Each panel shows reference full QM (*dotted*) and MM (*dashed*) results, as compared with an adaptive QM/MM simulation (*solid*). The top panel shows results of a Abrupt simulation, while the *bottom panel* shows the results of a BF simulation. The *unshaded* parts of the plots represent the QM regions, with the gradient in shading corresponding to the hysteresis range, a refinement used by the authors to reduce the fluctuations in the set of molecules that constitute the active and transition region. Reproduced from Ref. [32] with permission

2.3.3 Timings

In general, one can roughly rank the timings of the approaches discussed in Sect. 2.2 by the number of partitions that are computed each time-step. This would result in a speed ranking of constrained > discontinuous > continuous. The continuous methods compute many partitions each time-step, where PAP computes by far the most, and will therefore be the slowest among all. On the other hand, the viewpoint has been taken [19] that these partition-based calculations are trivially parallelizable, and as long as a large number of computing cores are available, more partitions do not lead to more wall time. At that point, the deciding issue becomes the size of the active and transition regions, the required time-step for the simulation, and for a minor part the overhead of the adaptive algorithm.

Most QM/MM methods are implemented in different program packages, which makes it difficult to compare their timings on the same ground. In Ref. [17] the timings for a subset of QM/MM methods were presented, all using FlexMD, [44] and are reproduced in Table 2.3. FlexMD is a QM/MM python library developed with the specific purpose of multi-scale simulations on aqueous systems in mind, and it is available within the ADF software [49]. The FlexMD simulations, using a relatively fast semi-empirical QM description (PM6-DH+), yield a speed ranking of Abrupt > FIRES > BF > DAS > SAP. The discontinuous adaptive methods compete with the constrained methods, while, regardless of parallelization of the partition calculations, the continuous adaptive methods are the slowest. In the

Table 2.3 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

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

All simulations are run in serial, except DAS and SAP, which are un 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. Reproduced from Ref. [17] with permission

BF and DAS simulations, the size of the largest set $S_{QM}^{(n)}$ in the corresponding partition (n) is the same, and the slower performance of DAS can only be attributed to overhead. The very slow timings of SAP are caused by the fact that the steep partition weight functions $\sigma^{(n)}(\mathbf{r})$ require the simulation to be performed with a time-step of 0.1 fs, smaller than the 0.5 fs used for other methods in this example.

2.4 Proton Transfer at the Boundary of Adaptive QM/MM

The previous sections presented the state of art in adaptive QM/MM methods. Since their debut in 1996 with the Hot Spot method, they have largely reduced the demand on computational resources, while improving structural properties and energy conservation. These methods have already been successfully applied to simulate the solvation structure of ions in liquid water [41, 43, 50, 51] and even aqueous chemical reactions [19, 23]. The validation procedure is usually based on comparison with experiments [50] or with the corresponding fully QM simulation [19, 23, 41].

However, as stated in the introduction, the bulk of the aqueous reactions of interest involve proton transfer in some form. Even with advanced adaptive QM/MM methods it is still not trivial to perform simulations on proton transfer processes, due to their delocalized nature. Therefore, considering an example system in which proton transfer and diffusion occur can serve as a tough “crash test” for adaptive QM/MM, and should demonstrate the limitations of the methods. We specifically focus on diffusion and Grothuss shuttling, which are particularly non-local, and will affect thermodynamic reaction quantities. The impact of the introduction of an adaptive QM/MM boundary will be quantified through comparison with fully QM simulations on the same system. To underline the limits of the methodology, the behavior of the proton transfer events and their deviation from expectation will be thoroughly discussed.

The generally accepted notion is that the hydronium ion, once formed, exists in two preferred conformations: the dimer or Zundel ion (O_2H_5^+), or the Eigen cation (O_4H_9^+) [52]. In the dimer, a single step proton transfer between the contributing oxygen atoms is nearly barrierless, and occurs on a very short time-scale [53, 54]. This behavior is generally separated from the so-called forward hopping motion, where the proton

moves to a new donor oxygen atom without hopping back on a short time-scale. This forward hopping motion is attributed to a rearrangement of the solvent shell around the hydronium ion (re-solvation) [55]. The hydroxide ion also requires resolution to precede a forward hopping step. Molecular dynamics simulations indicate that the hydroxide ion in its stable state is involved in four hydrogen bonds to proton donors (hypercoordination) [8, 56] in agreement with experiments [57]. Only when one of the donated hydrogen bonds breaks while at the same time an acceptor hydrogen bond is in place [5], a proton transfer event can result in a stable new hydroxide ion.

The proton transfer reaction selected for study is the diffusion of a hydroxide ion in a solution of methanol in water ($pH \gg 7$). The model contains a methanol molecule dissolved in a cubic periodic water box of 15 Å side containing 112 water molecules and one (Na^+ , OH^-) dissociated ion pair. This gives a total density of 1.03 kg/L and an approximate pH around 13.3.

Based on its satisfactory performance, as discussed in Sect. 2.3, the DAS method was selected for the adaptive QM/MM simulations. The center of the active region is defined by the position of the methanol oxygen atom O_{Me} , and the active region has a radius of 6.3 Å around it. The transition zone extends from 6.3 to 7.2 Å and ensures a smooth QM/MM transition for the water molecules. This setup is depicted in Fig. 2.15.

The QM and MM descriptions were selected based on two requirements: (1) The MM force field should allow proton shuttling in the environment region, and

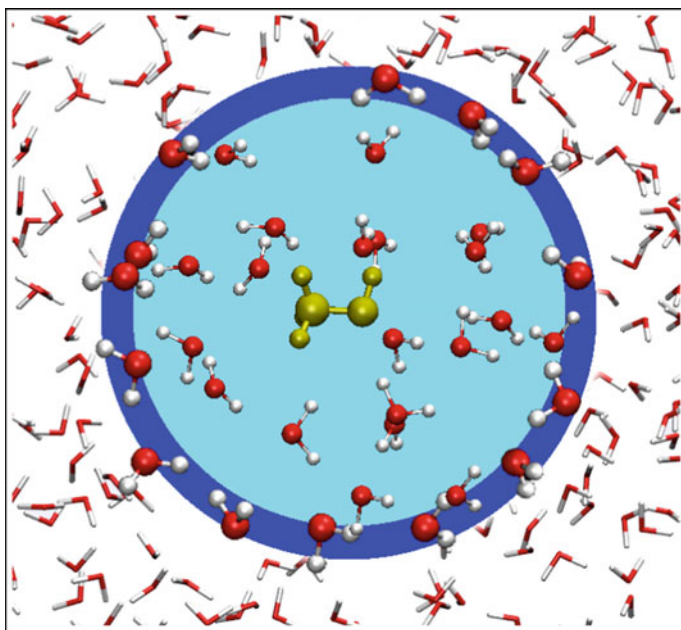


Fig. 2.15 Methanol solvated by an alkaline solution. Methanol is represented in yellow. The light blue disk (a sphere in 3D) corresponds to the active (QM) zone, the blue ring to the transition zone

(2) The difference between the QM and the MM description of water should not be too large, to minimize the artifacts at the QM/MM boundary. The reactive force field ReaxFF [58] is used as the MM level (as implemented in the ADF package), [49] with parameters specifically developed for proton transfer in water. In this way the OH^- ion is in principle allowed to diffuse in and out of the active region, while still retaining its reactive qualities. All QM calculations are done at the semi-empirical PM6-DH+ level of theory, [59, 60] as implemented in the MOPAC2012 program [61]. Earlier calculations [17] showed that the PM6-DH+ and ReaxFF descriptions of water are very similar, thereby minimizing the chance of artifacts due to differences between the potentials. PM6 offers a real improvement to AM1 or PM3 for the description of liquid water [59, 62]. This comes from the fact that the core-core interaction has been modified, analogously to the strategy proposed by the authors of PM3-MAIS or PM3-PIF [63, 64]. While able to describe proton transfer, the PM6-DH+ potential is not optimized for proton transfer processes [65], and is known to overestimate the methanol-water interaction [62]. As a result, our setup is perfectly capable of demonstrating the pitfalls of adaptive QM/MM, but it is not expected to provide meaningful physical insights into the selected process. The interaction between the QM and MM molecules is described with mechanical embedding, which means that the electron density is not polarized by the MM charges. However, the use of the charge equilibration scheme Qeq [66] in ReaxFF, ensures polarization of the QM region by the MM environment, even if this effect is purely classical.

In all simulations, a time-step of 0.5 fs is used. We base this choice on earlier contributions [17, 44] that used the same QM and MM descriptions as applied here. We found that a time-step of 0.5 fs was sufficient to describe the O-H vibrations in water, and resulted in energy conserving simulations. We furthermore note that the proton transfer process—as the proton transfers from an elongated bond with the donor oxygen to an elongated bond with the acceptor oxygen—occurs in 5–20 time-steps, which is sufficient to properly describe the process. A Nose-Hoover thermostat was employed to keep the temperature at 290 K. To serve as benchmarks, we first performed two fully QM molecular dynamics simulations of the entire system, with the initial OH^- position at (1) 8.85 Å and (2) 3.02 Å from the oxygen atom in methanol (O_{ME}), and one fully MM simulation. From the QM simulations we extracted five frames that were used to run five QM/MM simulations. To obtain a comparable amount of data for both types of simulations (QM and QM/MM) and to ensure they both spanned overlapping regions of phase space, four additional fully QM simulations (3), (4), (5) and (6) were started from positions taken from the last frame of the five QM/MM simulations. All simulations were run for 20 ps, of which the first 10 ps are considered equilibration, and excluded from analysis. Analysis of the five QM/MM simulations was performed over the entire 50 ps of equilibrated data available. Analysis of the QM data was performed over the trajectories from simulations (2), (3), (4), (5) and (6), also yielding a total of 50 ps of equilibrated data. The average time consumptions per time step for QM and QM/MM simulations are 24.82 s and 14.14 s, respectively. Note that the QM/MM simulations with DAS presented here are much slower than the ones in Sect. 2.3.3. This is simply because a bigger QM zone was used. Three different types of quantities were

computed: (i) the probability to find the OH^- at a certain distance from the QM core, (ii) the radial distribution of water oxygen atoms around the hydroxide ion, and (iii) the rate of proton transfer, from the time-correlation of the OH^- oxygen identity. We found that the structural data was easily converged within the first 10 ps, but the relatively low mobility of the OH^- ion encourages some caution in interpreting the data on its distribution over the simulation space. To estimate the convergence of the distribution of the distance from the OH^- to the methanol, we compared the probabilities to find the OH^- ion somewhere in the solution over two different time intervals: Between 10 and 20 ps and between 10 and 19 ps. We did this for all the QM and QM/MM simulations. The sum of the absolute values of the difference was found to be approximately 1 % of the maximum probability, which was deemed so small as to have a negligible influence on the results.

2.4.1 Influence of QM/MM Boundary on OH^- Diffusion

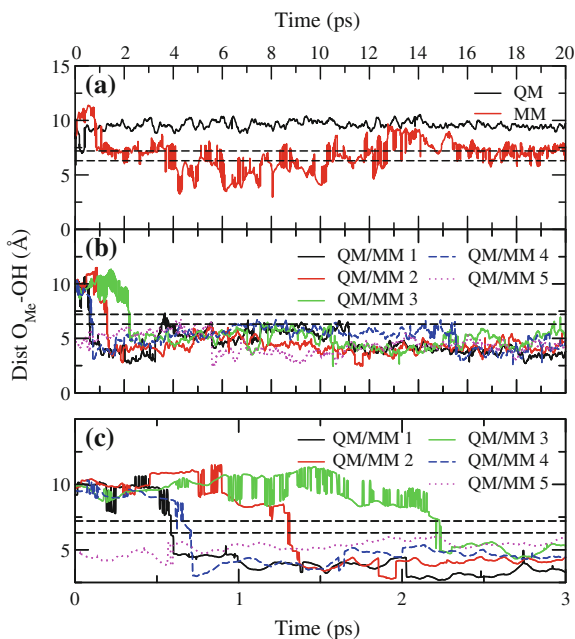
We started the QM/MM simulations with the hydroxide far away from O_{Me} , at approximately 10 Å, which is well inside the environment region. The hydroxide ion is therefore initially described MM. At the start of all five QM/MM simulations, we observed fast consecutive proton transfer occurrences at the start of the simulations, involving the hydroxide ion and neighboring water molecules. An initial fast and frequent shuttling of the proton back and forth between two oxygen atoms is then suddenly followed by a large displacement of the OH^- ion, always in the direction of the central methanol molecule. Fig. 2.16c displays the distance between the OH^- ion and the methanol O_{Me} atom, which suddenly decreases within the first 2.5 ps of the simulation. The OH^- ion then finds itself inside the active region boundary, where it stays for the remainder of the simulation.

It is worth noting that the sodium cation Na^+ follows a Brownian diffusion and thus moves much more slowly. In all simulations, it stayed close to its initial location at ~ 9 Å of the methanol molecule.

Of the six fully QM reference simulations, only one has a starting geometry with the OH^- ion located far away from the central methanol molecule. The evolution of the distance of the OH^- ion from the methanol molecule is depicted in Fig. 2.16a, and it reflects a very different behavior. The simulation starts with the OH^- ion located at 8.9 Å distance from the O_{Me} atom. Unlike in the QM/MM simulations, the OH^- ion did not approach the methanol molecule, but remains within a range of approximately 8 to 10 Å. This region corresponds to approximately 30 % of the total volume of the simulation box.

The deviating behavior of the OH^- ion in the QM/MM simulations has two separate aspects: (i) the fast migration of the OH^- ion towards the boundary of the QM/MM transition zone, and then (ii) the sudden jump into the QM zone of the OH^- ion. Since the mechanical embedding scheme is applied in the QM/MM simulations, it could be that a long range MM interaction causes the quick diffusion of the OH^- ion towards the central methanol molecule. However, the results of a

Fig. 2.16 Distance from the oxygen in methanol to the OH^- ion in (a) QM and MM simulations, and (b) QM/MM simulations. The first 3 ps of (b) is shown in more detail in (c)



fully MM simulation show that the MM potential does not steer the OH^- ion towards the MeOH molecule (Fig. 2.16a). The OH^- ion does move more freely through the solution than it does in the fully QM simulation, and like in the QM/MM simulation we observe many proton shuttling events (Figs. 2.16a). This is the direct result of a much higher proton transfer frequency in MM simulation. Including the proton shuttling events, we find proton transfer rates of 0.48^{-1} and 43 ps^{-1} in the QM simulation and MM simulations respectively. In our QM/MM simulations, the OH^- is initially positioned at the corner of the box, which is around 10 Å distance from MeOH. Due to the limited size of the box, the highly mobile MM OH^- can quickly migrate to the boundary of the transition region at 7.2 Å. We observed a few occasions in which the ion re-entered the transition region, but these were quickly followed by a proton shuttling event launching it back, and overall the OH^- ion stays trapped inside the active region. The apparent preference of the OH^- ion for the active region creates an unphysical pH gradient in the QM/MM solution, and increases the probability of a reaction with the solute present at the core of the active region.

We note that an artifact such as the one described above strongly depends on the choice of QM and MM potentials. The simplest solution to the above problems would be to develop force fields that are very similar to the QM description of proton transfer in water. However, we suspect that there will be a limit to the accuracy with which a simple force field can describe such a complex process. As an alternative solution, we suggest the introduction of a second active region, centered on the diffusive OH^- ion itself. In this manner, the OH^- ion is described at the QM level

everywhere in the simulation box, and an un-physical preference for one region over another is avoided. Developing such a method is not be without its difficulties since the hydroxide anion does not always correspond to the same oxygen and proton.

2.4.2 Reliability of QM/MM Equilibrated Structures

We now focus on average structural differences between a QM and a QM/MM equilibrated system. We select three relevant structural indicators, namely the distance of the OH^- ion from the QM core, the radial distribution of water around water in the active region, and the radial distribution of water around the OH^- ions when it is located inside the active region.

The average position of the OH^- ion in QM/MM simulations is closer to the QM core than it is in the fully QM reference simulations, as shown in Fig. 2.17. In the QM reference simulations, the OH^- ion spends some time in the area that corresponds to the transition region in the QM/MM simulations. The difference between the two results is related to our earlier observation that in the QM/MM simulation the OH^- ion is quickly shuttled back into the active region whenever it enters the transition region.

The radial distribution of the oxygen atoms of the water molecules in the active region around one another in the active region is depicted in Fig. 2.18a. Overall, the water structure obtained in the QM/MM simulations agrees well with that of the QM reference. The first solvation peak in the QM/MM simulations is almost on top of that in the QM simulations, and the second solvation peak is a fraction further away. More significant is the increased density between the first and second solvation shell in the QM/MM simulations. This corresponds to a more compressed aqueous environment in the active region. Integration of the first solvation peak in the radial distribution function $g(r)$ yields an average water coordination number of 4.23 for the QM/MM water molecules in the active region, versus a coordination number of 3.43 in the same region in the QM simulation. This amounts to an

Fig. 2.17 Average histogram of distance from O_{Me} to OH^- after 10 ps of equilibration for both QM (*black solid*) and QM/MM (*red empty*) simulations, the histogram of the only MM simulation is also shown in *dashed orange*

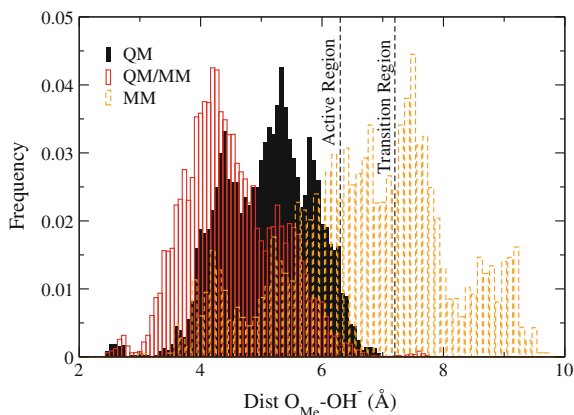
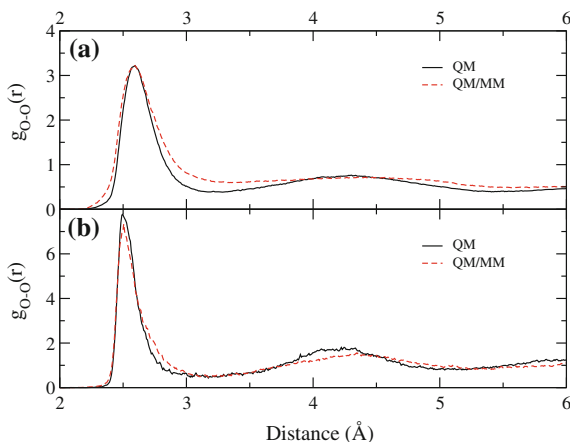


Fig. 2.18 Average oxygen-oxygen radial distribution around (a) all oxygen atoms in the range of 6.3 Å from O_{Me} in the QM and in the QM/MM simulations, and (b) all oxygen atoms around the OH^- ion in QM and QM/MM simulations



average increase in coordination with 0.80 water molecules. The integration of the radial distribution function over the entire QM region yields 43.21 water molecules for the QM/MM simulation, a 10.2 % increase over 39.22 water molecules in the same region in QM simulations. We attribute the higher QM/MM water density to the differences between the QM and the MM chemical potentials of water. The optimized geometries of a water dimer have optimized oxygen-oxygen bond lengths r_{O-O} of 2.79 Å with the QM potential, and 2.89 Å with the MM potential. This suggests that indeed the MM water prefers a lower density than the QM water, and therefore exerts a pressure on the QM region. The radial distribution of water oxygen atoms around the OH^- oxygen atom in the active region (Fig. 2.18b) does not appear to be significantly affected by the increased density in the active region. Integration of the $g(r)$ for the distribution of oxygen atoms around the hydroxide ion reveals only a small increase in coordination, from 5.13 in the QM simulation to 5.46 in the QM/MM simulations. The overall altered environment of the OH^- ion may still affect its reactive behavior.

2.4.3 Proton Transfer in the Active Region

Both our QM and QM/MM simulations show that hydroxide migration occurs in step-wise hops followed by long resting periods. The average hopping rates (counting all proton transfer events, not only the forward-hopping motions) are 1.67 and 0.48 ps^{-1} for QM/MM and QM simulations respectively.

Another way to quantify the proton hopping rate is to compute the decorrelation with time of the location of the hydroxide ion at a certain oxygen atom. Let us define a characteristic function $h_i(t)$ which is 1 if a certain oxygen atom O_i can be

identified as the hydroxide ion at time t , and zero when it is not. We can then evaluate the auto-correlation function of $h_i(t)$ as,

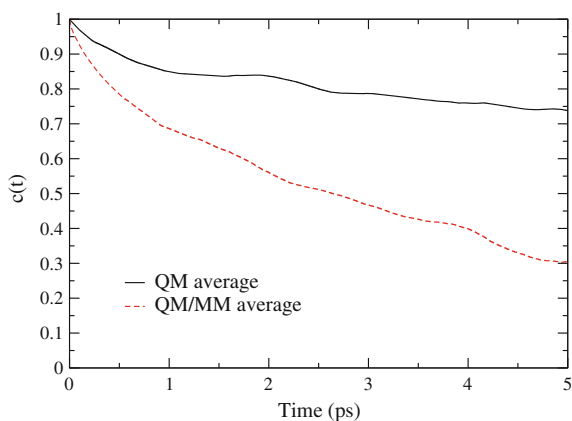
$$c(t) = \frac{1}{t_{\max}} \int_{t'=0}^{t_{\max}} \sum_{i \in O} h_i(t') h_i(t' + t) dt', \quad (2.25)$$

where the time-intervals $\{t', t' + t\}$ are extracted from all the QM (or QM/MM) trajectories after equilibration.

Decorrelation is not complete on the time-scales of our simulations ($c(t)$ in Fig. 2.19 does not decay to zero). Again, we observe a considerably faster decay (higher hopping-rate) in the QM/MM simulation than in the fully QM simulation. In both cases, the OH^- ion remains mainly in the active region.

The increased water density in the active region of the QM/MM simulation (discussed in Sect. 2.4.2) may explain the increased proton transfer rate observed in the active region of the QM/MM simulation. The generally accepted picture of proton transfer to an OH^- ion (shown in Fig. 2.20a) states that the OH^- ion is generally involved in hydrogen bonds to five neighboring water molecules, four of which are prospective proton donors. A water molecule, by contrast, has on average four hydrogen bonds. A forward hopping proton transfer event is preceded by a fluctuation of the solvent structure around the hydroxide ion, known as re-solvation. This involves cleavage of one of the proton donating hydrogen bonds, after which the hydroxide ion finds itself under more favorable conditions to become a water molecule. If a low-barrier proton transfer then occurs, the product state is accommodated by the solvation shell (presumably the solvation structure of the formed hydroxide ion simultaneously adjusts towards a coordination number of five), and the new conformation is stable. In the QM/MM simulations, the coordination number of the water molecules is significantly higher than in the QM reference simulations, shown in Fig. 2.20b. As a result, the water molecules adjacent to the hydroxide ions on average have more coordinated water molecules, and will therefore more often have a solvation structure that is akin to that of a

Fig. 2.19 Average auto-correlation function of OH^- for QM and QM/MM simulations



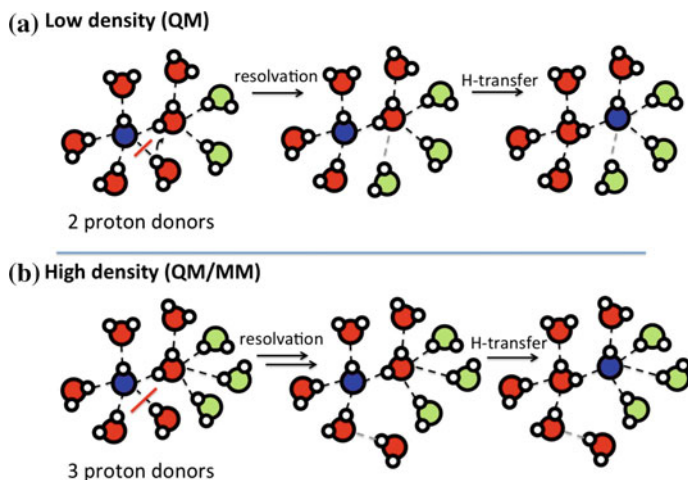


Fig. 2.20 Schematic representation of re-solvation accompanying a proton transfer event to a hydroxide ion (*blue*) at regular and high density. At high density many of the proton donors have sufficient coordinated water molecules (*green*) to stabilize the product conformation

hydroxide ion (5 hydrogen bonds). This can be viewed as favorable re-solvation to accommodate the product of a proton transfer event, thereby increasing the mobility of the hydroxide ion.

2.4.4 What Did We Learn?

To illustrate the capabilities and limitations of adaptive QM/MM methods, we focused on a highly challenging problem: the diffusion of a hydroxide ion from the bulk to the reactive center. We evaluated the performance of Difference-based Adaptive Solvation (DAS) as a representative adaptive QM/MM method in the description of proton diffusion, using an aqueous solution of NaOH and methanol as our model system. Even for our very small model system, the QM/MM simulation is faster than fully QM by approximately a factor of two. We find that performing a QM/MM simulation of a non-local process like proton transfer results in an amplification of the known artifacts induced by the QM/MM boundary.

The differences between the QM and the MM potential can result in strong artifacts in the diffusivity of a hydroxide ion, which in our simulations remained largely trapped in the active (QM) region. The nature of the artifacts in the diffusivity will strongly depend on the nature of the QM and MM potentials and their comparative description of the diffusion of a proton. As a possible solution we suggest the introduction of an extra active region that adapts its position to the location of the reactive OH^- ion.

Inside the active region of a QM/MM system, we found a significantly higher proton transfer rate with respect to the fully QM reference. We attribute this second discrepancy to a general problem—also found in conventional QM/MM—that the interaction between QM and MM molecules introduces artifacts. In our simulations, the water density inside the QM region was somewhat higher than the water density in the fully QM simulations (both were performed with fixed total volume in the NVT ensemble). We expect that this density difference stems from the chemical potential difference between QM and MM molecules, as a result of which the two regions exert a pressure on one another [67]. The resulting smaller volume of the QM region does not significantly affect the solvation shell of the OH^- ion, but the water molecules have a denser solvation structure. Because an OH^- ion generally prefers a higher coordination number than a water molecule, the water molecules that can serve as proton donors to a neighboring OH^- ion are better equipped for their future role as OH^- ion, resulting in the observed higher proton transfer rate.

Overall, we believe that the presented methods for a multi-scale description of reactions in water represent the future of simulations of aqueous reactions. However, our results show that for the successful description of proton transfer events, several challenges still lie in our path.

2.5 Conclusion

We have presented in some detail the state of the art in available methods for the QM/MM simulations of chemical reactions in solution. We distinguish three classes of methods, each with a different approach to the diffusion of solvent molecules across the QM/MM boundary. The first class, constrained QM/MM, prevents this diffusion, and as a consequence dynamical properties are correctly described only well within the QM region. In addition, recovery of the correct structural properties at the QM/MM boundary still poses a challenge. Nonetheless, the constrained methods have the non-negligible advantage of simplicity and low computation cost, equivalent to that of a conventional QM/MM simulation.

The second class, discontinuous adaptive QM/MM, adapts the description of a solvent molecule (QM versus MM) as it diffuses across the QM/MM boundary, and does this in a discontinuous manner. The straight-forward Abrupt method ‘abruptly’ modifies the description of a solvent molecule as it diffuses from the QM active region into the MM environment. This results in strong discontinuities in potential energy and forces, and in some cases involves structural artifacts at the QM/MM boundary. To reduce these effects, several strategies have been developed over the past 20 years. The basic feature is the transition region between the active and environment regions, and the partitioning of the system into QM and MM molecules in two different manners. If a weighted average of the resulting energies or forces is then applied, the molecules in the transition region can be assigned fractional QM (and MM) character. These methods have led to the first successful applications of the adaptive QM/MM methods for solution chemistry. While the

dual partition approach can reduce discontinuities, it cannot remove them, and as a consequence the total energy is not conserved, and structural properties can be recovered only using a strong thermostat to compensate the anomalous heating of the system.

The third and more recent class contains the continuous adaptive methods. Like the second class they define a transition region, but they are not limited to the computation of only two different QM/MM partitions for each given configuration. By computing many partitions, these methods achieve a continuous description of the forces exerted on the solvent molecules as they cross the transition region. The accompanying increase in computational load is easily compensated by efficient parallelization of the numerous QM/MM computations that are required at each molecular dynamics time-step. The methods constitute a formal improvement, but they are not exempt from defects. While an energy conserving application is possible, it is generally sacrificed to obtain accurate molecular structures. In addition, the complex definitions of the energies and/or forces can result in fast changes, requiring a small time-step for proper integration.

A comparative performance of the three classes of QM/MM methods leads to the following conclusions. In terms of computation performance, a ranking can be approximately based on the number of required QM/MM calculations, which increases with the complexity of the method. For the constrained QM/MM methods only one calculation suffices, the discontinuous methods require at most two QM/MM calculations, and the continuous methods are the most demanding, with a computational cost at least proportional to the number of solvent molecules in the transition region. When ranking energy conservation, the continuous methods are at the top, closely followed by the constrained methods. Energy conservation suggests a good description of dynamical properties, but the constrained methods fail by construction when dynamical properties involving diffusion near the QM/MM boundary are desired. Finally, in reproducing the QM molecular structures, the discontinuous methods perform at least as well as the continuous ones. Overall, the best compromise between diffusion properties (solvation life-times) and structural properties (radial distribution) is provided by non-energy-conserving continuous adaptive methods such as DAS.

We exemplified the strengths and limitations of the DAS method on a challenging system by describing the diffusion of a hydroxide ion in water, in the presence of a methanol molecule. The active region is centered on the methanol molecule, and the hydroxide ion can freely diffuse in and out this active region. We used the semi-empirical Hamiltonian PM6-DH+ to describe the active region, while in the environment region all molecules are treated with the ReaxFF force field, which can describe a Grothuss shuttling mechanism for hydroxide diffusion. For the sake of simplicity, the mechanical embedding scheme was used. When comparing our DAS-QM/MM with fully QM and fully MM simulations, the main artifacts we observed resulted directly from the QM/MM interactions, not from the adaptive nature of the simulations. The difference between the QM and the MM description of proton transfer slows the diffusion of the hydroxide ion down once it enters the active region, thereby effectively trapping it inside. The hydroxide diffusion in the

active region is still faster than in the fully QM simulation, which is a result of anomalous interactions at the QM/MM boundary.

In summary, the large choice of adaptive schemes available today allows QM/MM simulations of solution chemistry with the main artifacts similar to those of conventional QM/MM schemes. We expect that a careful choice of the QM/MM scheme combined with a QM region around each active center (here methanol *and* hydroxide anion) should cure the main sources of error.

References

1. Shelton DP (2000) Chem Phys Lett 325:513. doi:[10.1016/S0009-2614\(00\)00734-X](https://doi.org/10.1016/S0009-2614(00)00734-X)
2. Marcus Y (2009) Chem Rev 109(3):1346. doi:[10.1021/cr8003828](https://doi.org/10.1021/cr8003828)
3. von Grotthuss CJT(1806) Ann Chim (Paris) 58:54. Available at: <http://gallica.bnf.fr/ark:/12148/bpt6k6573748h/f54.image.r=Annales%20de%20chimie%20Paris.langEN>
4. Tielrooij KJ, Timmer RLA, Bakker HJ, Bonn M (2009) Phys Rev Lett 102:198303. doi:[10.1103/PhysRevLett.102.198303](https://doi.org/10.1103/PhysRevLett.102.198303)
5. Hassanali A, Giberti F, Cuny J, Kühne TD, Parrinello M (2013) PNAS 110(34):13723. doi:[10.1073/pnas.1306642110](https://doi.org/10.1073/pnas.1306642110)
6. Roberts ST, Mandal A, Tokmakoff A (2014) J Phys Chem B 118(28):8062. doi:[10.1021/jp501145p](https://doi.org/10.1021/jp501145p)
7. Giberti F, Hassanali AA, Ceriotti M, Parrinello M (2014) J Phys Chem B 118(46):13226. doi:[10.1021/jp507752e](https://doi.org/10.1021/jp507752e)
8. Tuckerman ME, Marx D, Parrinello M (2002) Nature 417(6892):925. doi:[10.1038/nature00797](https://doi.org/10.1038/nature00797)
9. Warshel A, Levitt M (1976) J Mol Biol 103(2):227. doi:[10.1016/0022-2836\(76\)90311-9](https://doi.org/10.1016/0022-2836(76)90311-9)
10. Thole BT, van Duijnen PT (1980) Theor Chim Acta 55(4):307. doi:[10.1007/BF00549429](https://doi.org/10.1007/BF00549429)
11. Field MJ, Bash PA, Karplus M (1990) J Comput Chem 11(6):700. doi:[10.1002/jcc.540110605](https://doi.org/10.1002/jcc.540110605)
12. Gao J (1995) Reviews in computational chemistry. In: Lipkowitz K, Boyd DB (eds) Wiley, pp 119–185. doi:[10.1002/9780470125847.ch3](https://doi.org/10.1002/9780470125847.ch3)
13. Sherwood P (2000) Modern methods and algorithms of quantum chemistry: proceedings. In: Grotendorst J (ed) NIC series, vol 3, 2nd edn. NIC-Directors, Jülich, pp 257–277
14. Carloni P, Rothlisberger U, Parrinello M (2002) Acc Chem Res 35(6):455. doi:[10.1021/ar010018u](https://doi.org/10.1021/ar010018u)
15. Magistrato A, DeGrado WF, Laio A, Rothlisberger U, VandeVondele J, Klein ML (2003) J Phys Chem B 107(17):4182. doi:[10.1021/jp027032o](https://doi.org/10.1021/jp027032o)
16. Kerdcharoen T, Liedl KR, Rode BM (1996) Chem Phys 211(1–3):313. doi:[10.1016/0301-0104\(96\)00152-8](https://doi.org/10.1016/0301-0104(96)00152-8)
17. Buló RE, Michel C, Fleurat-Lessard P, Sautet P (2013) J Chem Theory Comput 9(12):5567. doi:[10.1021/ct4005596](https://doi.org/10.1021/ct4005596)
18. Pezeshki S, Lin H (2015) Mol Simul 41(1–3):168. doi:[10.1080/08927022.2014.911870](https://doi.org/10.1080/08927022.2014.911870)
19. Park K, Gotz AW, Walker RC, Paesani F (2012) J Chem Theory Comput 8:2868. doi:[10.1021/ct300331f](https://doi.org/10.1021/ct300331f)
20. Mones L, Jones A, Goetz AW, Laino T, Walker RC, Leimkuhler B, Cányi G, Bernstein N (2015) J Comput Chem 36: 633. doi:[10.1002/jcc.23839](https://doi.org/10.1002/jcc.23839)
21. Pezeshki S, Davis C, Heyden A, Lin H (2014) J Chem Theory Comput 10:4765. doi:[10.1021/ct500553x](https://doi.org/10.1021/ct500553x)
22. Tunon I, Martins-Costa MTC, Millot C, Ruiz-Lopez MF (1997) J Chem Phys 106(9):3633. doi:[10.1063/1.473457](https://doi.org/10.1063/1.473457)

23. Várnai C, Bernstein N, Mones L, Csányi G (2013) *J Phys Chem B* 117(40):12202. doi:[10.1021/jp405974b](https://doi.org/10.1021/jp405974b)
24. Nielsen SO, Buló RE, Moore PB, Ensing B (2010) *Phys Chem Chem Phys* 12:12401. doi:[10.1039/c004111d](https://doi.org/10.1039/c004111d)
25. Potestio R, Fritsch S, Espanol P, Delgado-Buscalioni R, Kremer K, Everaers R, Donadio D (2013) *Phys Rev Lett* 110:108301/1. doi:[10.1103/PhysRevLett.110.108301](https://doi.org/10.1103/PhysRevLett.110.108301)
26. Praprotnik M, Delle Site L (2013) *Methods Mol Biol* (NY, U.S.) 924:567. doi:[10.1007/978-1-62703-017-5_21](https://doi.org/10.1007/978-1-62703-017-5_21)
27. Salazar MR (2005) *J Phys Chem A* 109(50):11515. doi:[10.1021/jp053551q](https://doi.org/10.1021/jp053551q)
28. Guthrie MG, Daigle AD, Salazar MR (2010) *J Chem Theory Comput* 6:18. doi:[10.1021/ct900449q](https://doi.org/10.1021/ct900449q)
29. Pezeshki S, Lin H (2011) *J Chem Theory Comput* 7:3625. doi:[10.1021/ct2005209](https://doi.org/10.1021/ct2005209)
30. Kerckhove T, Morokuma K (2002) *Chem Phys Lett* 355(3–4):257. doi:[10.1016/S0009-2614\(02\)00210-5](https://doi.org/10.1016/S0009-2614(02)00210-5)
31. Csányi G, Albaret T, Payne MC, De Vita A (2004) *Phys Rev Lett* 93:175503. doi:[10.1103/PhysRevLett.93.175503](https://doi.org/10.1103/PhysRevLett.93.175503)
32. Bernstein N, Várnai C, Solt I, Winfield SA, Payne MC, Simon I, Fuxreiter M, Csányi G (2012) *Phys Chem Chem Phys* 14:646. doi:[10.1039/c1cp22600b](https://doi.org/10.1039/c1cp22600b)
33. Heyden A, Lin H, Truhlar DG (2007) *J Phys Chem B* 111(9):2231. doi:[10.1021/jp0673617](https://doi.org/10.1021/jp0673617)
34. Buló RE, Ensing B, Sikkema J, Visscher L (2009) *J Chem Theory Comput* 5(9):2212. doi:[10.1021/ct900148e](https://doi.org/10.1021/ct900148e)
35. Watanabe HC, Kubar T, Elstner M (2014) *J Chem Theory Comput*. doi:[10.1021/ct5005593](https://doi.org/10.1021/ct5005593)
36. Bakowies D, Thiel W (1996) *J Phys Chem* 100(25):10580. doi:[10.1021/jp9536514](https://doi.org/10.1021/jp9536514)
37. Senn H, Thiel W (2007) *Atomistic approaches in modern biology*. In: Reiher M (ed) *Topics in current chemistry*, vol 268. Springer, Berlin, pp 173–290. doi:[10.1007/128_2006_084](https://doi.org/10.1007/128_2006_084)
38. Senn HM, Thiel W (2009) *Angew Chem Int Ed* 48(7):1198. doi:[10.1002/anie.200802199](https://doi.org/10.1002/anie.200802199)
39. Zhang Y, Lin H (2008) *J Chem Theory Comput* 4(3):414. doi:[10.1021/ct700296x](https://doi.org/10.1021/ct700296x)
40. Zhang Y, Lin H (2010) *Theor Chem Acc* 126(5–6):315. doi:[10.1007/s00214-009-0704-z](https://doi.org/10.1007/s00214-009-0704-z)
41. Rowley CN, Roux B (2012) *J Chem Theory Comput* 8(10):3526. doi:[10.1021/ct300091w](https://doi.org/10.1021/ct300091w)
42. Shiga M, Masia M (2013) *J Chem Phys* 139(4):044120. doi:[10.1063/1.4816629](https://doi.org/10.1063/1.4816629)
43. Hofer TS, Pribil AB, Randolf BR, Rode BM (2005) *J Am Chem Soc* 127(41):14231. doi:[10.1021/ja052700f](https://doi.org/10.1021/ja052700f)
44. Fleurat-Lessard P, Michel C, Buló RE (2012) *J Chem Phys* 137:074111/1. doi:[10.1063/1.4739743](https://doi.org/10.1063/1.4739743)
45. Pezeshki S, Davis C, Heyden A, Lin H (2014) *J Chem Theory Comput* 10:4765. doi:[10.1021/ct500553x](https://doi.org/10.1021/ct500553x)
46. Takenaka N, Kitamura Y, Koyano Y, Nagaoka M (2012) *Chem Phys Lett* 524:56. doi:[10.1016/j.cplett.2011.12.053](https://doi.org/10.1016/j.cplett.2011.12.053)
47. Takenaka N, Kitamura Y, Koyano Y, Nagaoka M (2012) *J Chem Phys* 137:024501/1. doi:[10.1063/1.4732307](https://doi.org/10.1063/1.4732307)
48. Waller MP, Kumbhar S, Yang J (2014) *ChemPhysChem* 15:3218. doi:[10.1002/cphc.201402105](https://doi.org/10.1002/cphc.201402105)
49. Velde G, Bickelhaupt F, Baerends E, Fonseca-Guerra C, Van Gisbergen S, Snijders J, Ziegler T (2001) *J Comput Chem* 22:931. doi:[10.1002/jcc.1056](https://doi.org/10.1002/jcc.1056)
50. Rode BM, Schwenk CF, Tongraar A (2004) *J Mol Liq* 110:105. doi:[10.1016/j.molliq.2003.09.016](https://doi.org/10.1016/j.molliq.2003.09.016)
51. Hofer TS, Tran HT, Schwenk CF, Rode BM (2004) *J Comput Chem* 25(2):211. doi:[10.1002/jcc.10374](https://doi.org/10.1002/jcc.10374)
52. Agmon N (1995) *Chem Phys Lett* 244:456. doi:[10.1016/0009-2614\(95\)00905-J](https://doi.org/10.1016/0009-2614(95)00905-J)
53. Komatsuzaki T, Ohmine I (1994) *Chem Phys* 180(23):239. doi:[10.1016/0301-0104\(93\)E0424-T](https://doi.org/10.1016/0301-0104(93)E0424-T)
54. Newton MD, Ehrenson S (1971) *J Am Chem Soc* 93(20):4971. doi:[10.1021/ja00749a001](https://doi.org/10.1021/ja00749a001)
55. Markovitch O, Chen H, Izvekov S, Paesani F, Voth GA, Agmon N (2008) *J Phys Chem B* 112(31):9456. doi:[10.1021/jp804018y](https://doi.org/10.1021/jp804018y)

56. Marx D, Chandra A, Tuckerman ME (2010) *Chem Rev* 110(4):2174. doi:[10.1021/cr900233f](https://doi.org/10.1021/cr900233f)
57. Botti A, Bruni F, Imberti S, Ricci M, Soper A (2005) *J Mol Liq* 117(1–3):81. doi:[10.1016/j.molliq.2004.08.013](https://doi.org/10.1016/j.molliq.2004.08.013)
58. Rahaman O, van Duin ACT, Goddard WA, Doren DJ (2011) *J Phys Chem B* 115(2):249. doi:[10.1021/jp108642r](https://doi.org/10.1021/jp108642r)
59. Stewart J (2007) *J Mol Model* 13(12):1173. doi:[10.1007/s00894-007-0233-4](https://doi.org/10.1007/s00894-007-0233-4)
60. Korth M (2010) *J Chem Theory Comput* 6(12):3808. doi:[10.1021/ct100408b](https://doi.org/10.1021/ct100408b)
61. Stewart JJP (2012) Mopac2012. Stewart computational chemistry, Version 7.263 W. <http://OpenMOPAC.net>
62. Marion A, Monard G, Ruiz-Lopez MF, Ingrosso F (2014) *J Chem Phys* 141(3):034106. doi:[10.1063/1.4886655](https://doi.org/10.1063/1.4886655)
63. Bernal-Uruchurtu MI, Ruiz-López MF (2000) *Chem Phys Lett* 330(1–2):118. doi:[10.1016/S0009-2614\(00\)01062-9](https://doi.org/10.1016/S0009-2614(00)01062-9)
64. Monard G, Bernal-Uruchurtu MI, van der Vaart A, Merz KM, Ruiz-López MF (2005) *J Phys Chem A* 109(15):3425. doi:[10.1021/jp0459099](https://doi.org/10.1021/jp0459099)
65. Wu X, Thiel W, Pezeshki S, Lin H (2013) *J Chem Theory Comput* 9(6):2672. doi:[10.1021/ct400224n](https://doi.org/10.1021/ct400224n)
66. Rappe AK, Goddard WA (1991) *J Phys Chem* 95(8):3358. doi:[10.1021/j100161a070](https://doi.org/10.1021/j100161a070)
67. Fritsch S, Poblete S, Junghans C, Ciccotti G, Delle Site L, Kremer K (2012) *Phys Rev Lett* 108:170602/1. doi:[10.1103/PhysRevLett.108.170602](https://doi.org/10.1103/PhysRevLett.108.170602)



<http://www.springer.com/978-3-319-21625-6>

Quantum Modeling of Complex Molecular Systems

Rivail, J.-L.; Ruiz-Lopez, M.; Assfeld, X. (Eds.)

2015, IX, 523 p., Hardcover

ISBN: 978-3-319-21625-6