



Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: [www.elsevier.com/locate/bbagen](http://www.elsevier.com/locate/bbagen)

## Review

Recent advances in QM/MM free energy calculations using reference potentials<sup>☆</sup>Fernanda Duarte<sup>\*</sup>, Beat A. Amrein, David Blaha-Nelson, Shina C.L. Kamerlin<sup>\*</sup>

Science for Life Laboratory, Department of Cell and Molecular Biology (ICM), Uppsala University, BMC Box 596, S-751 24 Uppsala, Sweden

## ARTICLE INFO

## Article history:

Received 24 May 2014

Received in revised form 6 July 2014

Accepted 7 July 2014

Available online xxxx

## Keywords:

Multiscale modeling

QM/MM free energy calculation

Averaging potential

Reference potential

Mean field approximation

## ABSTRACT

**Background:** Recent years have seen enormous progress in the development of methods for modeling (bio)molecular systems. This has allowed for the simulation of ever larger and more complex systems. However, as such complexity increases, the requirements needed for these models to be accurate and physically meaningful become more and more difficult to fulfill. The use of simplified models to describe complex biological systems has long been shown to be an effective way to overcome some of the limitations associated with this computational cost in a rational way.

**Scope of review:** Hybrid QM/MM approaches have rapidly become one of the most popular computational tools for studying chemical reactivity in biomolecular systems. However, the high cost involved in performing high-level QM calculations has limited the applicability of these approaches when calculating free energies of chemical processes. In this review, we present some of the advances in using reference potentials and mean field approximations to accelerate high-level QM/MM calculations. We present illustrative applications of these approaches and discuss challenges and future perspectives for the field.

**Major conclusions:** The use of physically-based simplifications has shown to effectively reduce the cost of high-level QM/MM calculations. In particular, lower-level reference potentials enable one to reduce the cost of expensive free energy calculations, thus expanding the scope of problems that can be addressed.

**General significance:** As was already demonstrated 40 years ago, the usage of simplified models still allows one to obtain cutting edge results with substantially reduced computational cost. This article is part of a Special Issue entitled Recent developments of molecular dynamics.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

## 1. Introduction

Recent years have seen enormous progress in the development of new methods for modeling molecular systems [1–6]. The introduction of massively parallelized computer architectures [7–9], together with the availability of more efficient algorithms, has allowed for a spectacular increase in terms of the size of the molecular systems studied [10,11] as well as the length of the simulations that can be performed [6,12]. For instance, the development of approximate methods for first-principles quantum chemistry [1,3,13] has made it possible to carry out electronic structure calculations for systems as big as millions of atoms [13–15]. Concurrently, the introduction of alternative hardware to perform

molecular dynamics (MD) simulations [7–9] now allows one to run single trajectories of  $\mu$ s [16,17] and, more recently, even MD simulations of ms in length [12,18]. Despite the many advances in the field, some of these technologies have failed to make the expected impact on the scientific community, in part due to their still excessive computational cost. Additionally, in the case of recent developments in first-principles quantum chemistry [1,3,13], their more difficult implementation and limited accessibility compared to conventional codes remain bottlenecks. Finally, despite the fact that running ms trajectories is computationally impressive, the computational cost involved discourages running sufficient replicas in order to test for reproducibility.

To perform large scale free energy calculations on complex systems, two key requirements need to be fulfilled: 1) the approach should be sufficiently accurate to describe the system under study in a physically meaningful way and 2) sufficient resources are needed to adequately sample the configurational space to draw unambiguous conclusions that are not dependent on starting structure. Therefore, one is stuck in balancing the cost of the simulations with the resources available. The availability of more powerful computer resources has partially resolved these issues allowing simulations of larger and larger systems. Examples of this include the ribosome system of approximately 2.64 million

**Abbreviations:** ABF, adaptive biasing force; CG, coarse-grained; EVB, empirical valence bond; FEP, free energy perturbation; LRA, linear response approximation; MD, molecular dynamics; PD, paradynamics; PMF, potential of mean force; QM/MM, quantum mechanics/molecular mechanics; QTCP, QM/MM thermodynamic cycle perturbation; US, umbrella sampling; REMD, replica exchange molecular dynamics; RS, reactant state; SCC-DFTB, self-consistent charge density functional tight binding; TS, transition state

<sup>☆</sup> This article is part of a Special Issue entitled Recent developments of molecular dynamics.

<sup>\*</sup> Corresponding authors.

<http://dx.doi.org/10.1016/j.bbagen.2014.07.008>

0304-4165/© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Please cite this article as: F. Duarte, et al., Recent advances in QM/MM free energy calculations using reference potentials, Biochim. Biophys. Acta (2014), <http://dx.doi.org/10.1016/j.bbagen.2014.07.008>

atoms [11] or the complete shell of southern bean mosaic virus, which in explicit solvent led to a system that comprised more than 4.5 million atoms and a total simulation time of about 100 ns [10]. Even with contemporary computer power, as the complexity of the systems of interest increases, these two requirements become more and more difficult to fulfill, further pushing the need for custom codes and specialized hardware. An alternative solution, in order to extend the range of problems that can be addressed, is the introduction of some level of simplification to the models being used. The use of simplified models is particularly important in the case of QM/MM free energy calculations, where even current computational resources do not allow for both a full high-level QM description of the reacting system and enough sampling of the conformational space [19] to obtain reasonable convergent free energies. Often the only solution to this problem has been to move to cheaper (but less quantitatively precise) semi-empirical models [5,19–21].

The use of simplified models in biomolecular simulations dates back to the 1970s, when the pioneering work of Levitt and Warshel [22] introduced the use of coarse-grained (CG) methods to the study of protein-folding processes. These models allowed one to significantly reduce the number of degrees of freedom of the system under study and therefore run longer simulations that would otherwise have been impossible using the resources available at that time. The first example of this was on bovine pancreatic trypsin inhibitor, BPTI [22], a small (<100 amino acids) single polypeptide chain protein of known conformation at that time (Fig. 1). In that study, the protein side chains were represented by spheres with an effective potential (implicitly representing the average potential of the solvated side chains) and the main chain was represented by virtual bonds between the  $C_\alpha$ 's. Using such a model, it was possible to correctly fold this protein in about 1000 minimization cycles, as outlined in Ref. [22] and illustrated in Fig. 1.

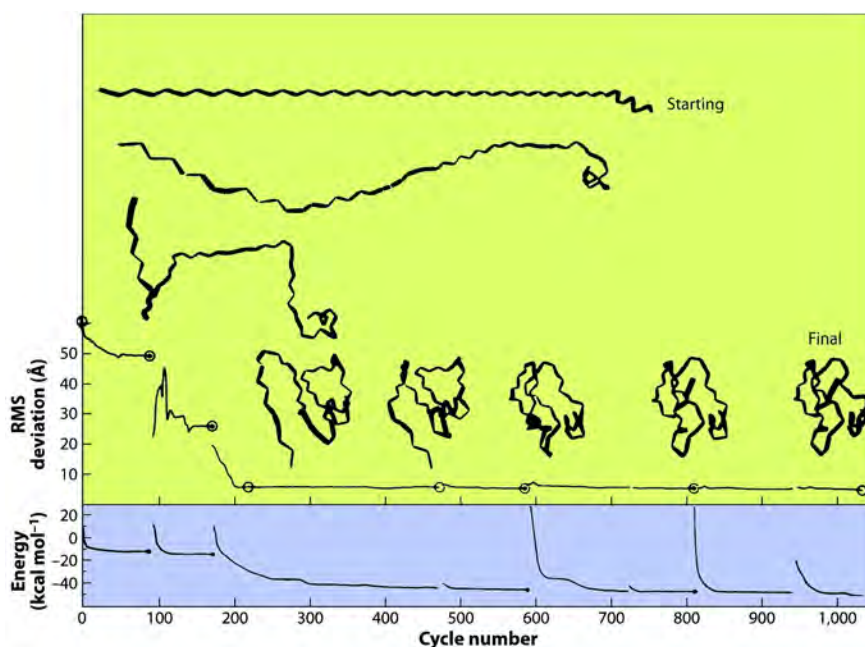
Similar approaches have subsequently been used in a variety of processes, including DNA and RNA folding [23,24], assemblies of membrane proteins [25] and vesicle formation [26]. More recently, the idea of using a simplified model as a reference potential has been expanded to a wide range of chemical problems [27–31], long time-scale conformational dynamics of proteins [32], and other related processes [33,34].

Having addressed the issue of cost vs. accuracy of the calculations, the second problem is the need for extensive conformational sampling. In principle, one would expect that the evaluation of a standard unbiased trajectory would be sufficient to visit the different regions of the conformational space multiple times. However, this requires the unbiased trajectory to be extremely (and inefficiently) long, as the system under study will spend a large fraction of the time in regions of phase space that have already been visited. A number of enhanced and rare event sampling techniques have been developed in order to reduce this problem: umbrella sampling [36], thermodynamics integration [37], replica exchange molecular dynamics (REMD) [38], the adaptive biasing force (ABF) method [39], transition path sampling [40], accelerated MD [41], metadynamics (MTD) [42] and paradynamics [28], just to name a few examples (for further information on some of these approaches, we refer readers to e.g. Ref. [43]). When combined with simplified models, these techniques have been shown to be capable of overcoming some of the limitations associated with computational cost in rational ways.

Earlier works have already discussed the methodological aspects of QM/MM approaches in detail (e.g. [20,44]) and their applications in the modeling of complex biological systems (e.g. [5,45,46]). In the present review, we hope to complement these excellent works, now focusing on a particular case study, namely the use of simplified reference potentials for performing higher-level hybrid QM/MM calculations. We will first provide a generalized introduction to this concept, followed by a discussion of the state-of-the-art of QM/MM free energy calculations using reference potentials as well as illustrating some recent applications. We will then conclude with an overview and discussion of open challenges and future perspectives for the field.

## 2. Approximating the high-level behavior of complex systems with low-level models

When using a simplified model to study complex systems, it is essential to use a model that captures the physics of the full explicit



**Fig. 1.** Simulation of bovine pancreatic trypsin inhibitor (BPTI) folding from an extended starting conformation with the terminal helix. Each residue has only one degree of freedom, i.e. the torsional angle  $\alpha$  between the four successive  $C_\alpha$  atoms. In all cases,  $\alpha = 180^\circ$ , except for residues 48 to 58 where  $\alpha = 45^\circ$ . No other knowledge whatsoever about the native protein was used for the simulation. In order to avoid nonproductive changes in the protein conformation, after a cycle of minimization, thermal fluctuations were introduced under the condition that each mode has an average kinetic energy of  $kT/2$  (where  $k$  is the Boltzmann constant and  $T$  is the absolute temperature).

Reprinted by permission from Macmillan Publishers Ltd: Nature [22], copyright 1975. Also adapted with permission from [35].

system and also to understand how the data obtained from the two models relate to each other [47]. This can be achieved using a simplified model as a *reference* in order to obtain the dynamical features of interest of the more complex system (here referred to as the *target* system) and then evaluating the cost of moving from the reference model to the target system and adding this as a correction to key states [27,48]. For example, if the dynamical feature of interest is the free energy of moving between the two states in the energy surface of the target system ( $\Delta g_{\text{TARGET}}$ ), one could first evaluate the corresponding free energy using the simplified reference model ( $\Delta g_{\text{REF}}$ ) and then evaluate the free energy difference between the simplified and the target systems ( $\Delta\Delta g_{\text{REF} \rightarrow \text{TARGET}}$ ) using a range of possible approaches that will be discussed in more detail below. This approach, which dates back to early protein folding studies [22,49], is described schematically in Fig. 2.

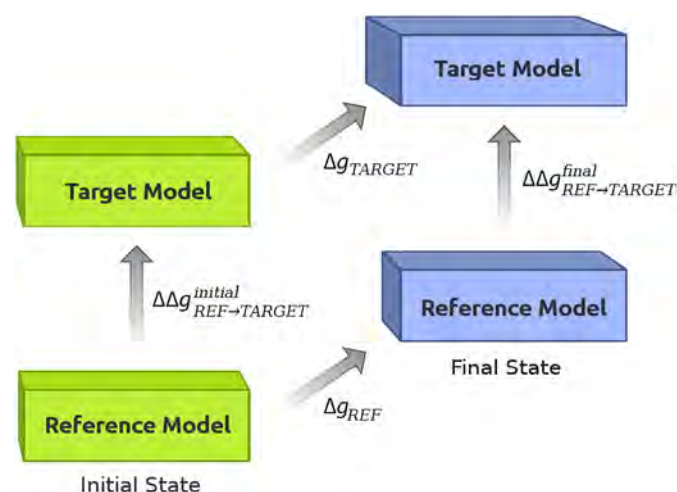
Following the thermodynamic cycle presented in Fig. 2, it is possible to formally describe the free energy correction necessary when using a simplified model as a reference potential for an explicit or high-level model [27]. This derivation (Eqs. (1)–(4)) was originally presented in Ref. [27] and extended in Refs [50,51]. Here we present its theoretical background to general readers.

We start by defining the configurational partition function of the system under study as follows:

$$Q_{\alpha}(x) = c_{\tau} \int_{-x}^{+x} \exp(-\Delta g_{\alpha}(x)\beta) dx. \quad (1)$$

Here,  $\alpha$  refers to either the simplified ( $Q_{\text{REF}}(x)$ ) or explicit ( $Q_{\text{TARGET}}(x)$ ) model,  $x$  is the reaction coordinate,  $\beta = 1/k_B T$ , where  $k_B$  is the Boltzmann constant and  $T$  is the absolute temperature,  $\tau$  denotes all of the coordinates perpendicular to the reaction coordinate,  $\Delta g_{\alpha}(x)$  is the resulting free-energy profile along the reaction coordinate, and  $c_{\tau}$  is a constant. From this, the partition function at the reactant state (RS) can be defined as:

$$Q_{\alpha}(\text{RS}) = c_{\tau} c_x \int_{-\infty}^{x^{\ddagger}} \exp(-\Delta g_{\alpha}(x)\beta) dx = c_{\tau} c_x \exp(\Delta G_{\alpha}(\text{RS})\beta). \quad (2)$$



**Fig. 2.** A thermodynamic cycle used to evaluate the cost of moving from a reference to a target model for a generic process. Having calculated the features of interest of the corresponding reference model (for the initial and final states) ( $\Delta g_{\text{REF}}$ ), the corresponding energy difference ( $\Delta\Delta g_{\text{REF} \rightarrow \text{TARGET}}$ ) is obtained from a perturbation between the two models.

Adapted with permission from Ref. [35].

A similar expression is obtained for the partition function of the product state (PS). The overall free energy of the reaction will then be simply defined as:

$$\Delta G_{\alpha}(\text{RS} \rightarrow \text{PS}) = \Delta G_{\alpha}(\text{PS}) - \Delta G_{\alpha}(\text{RS}). \quad (3)$$

The approach above is a generalization that can be applied to a range of (bio)chemical problems. In the specific case of evaluating the activation free energy of a chemical process, the following modified version of Eq. (3) can be used (for details see Ref. [27,51]):

$$\Delta g_{\alpha}^{\ddagger} = \Delta g_{\alpha}(x_{\alpha}^{\ddagger}) - \Delta g_{\alpha}(x_{\text{RS}}) + \beta^{-1} \ln \left[ \frac{1}{\Delta x^{\ddagger}} \int_{-\infty}^{x^{\ddagger}} \exp\left\{-\left(\Delta g_{\alpha}(x_{\alpha}^{\ddagger}) - \Delta g_{\alpha}(x_{\text{RS}})\right)\beta\right\} dx \right]. \quad (4)$$

The last term in Eq. (4) is a correction term that allows for the incorporation of ground state entropic effects into the transition state theory rate constant.

Once the activation ( $\Delta g_{\text{REF}}(x^{\ddagger})$ ) and reaction ( $\Delta G_{\text{REF}}(\text{RS} \rightarrow \text{PS})$ ) free energies have been calculated using the simplified model, the next step is to evaluate the free energy of moving between this simplified (or reference) potential to the corresponding target potential at key stationary points (i.e.  $\Delta\Delta g_{\text{REF} \rightarrow \text{TARGET}}(\text{RS} \rightarrow \text{PS})$  and  $\Delta\Delta g_{\text{REF} \rightarrow \text{TARGET}}(x^{\ddagger})$ , respectively). Once the relevant contributions have been calculated, one can obtain the respective free energies for the target system as follows:

$$\Delta G_{\text{TARGET}}(\text{RS}) = \Delta G_{\text{REF}}(\text{RS}) + \Delta\Delta G_{\text{REF} \rightarrow \text{TARGET}}(\text{RS}) \quad (5)$$

and

$$\Delta g_{\text{TARGET}}(x^{\ddagger}) = \Delta g_{\text{REF}}(x^{\ddagger}) + \Delta\Delta g_{\text{REF} \rightarrow \text{TARGET}}(x^{\ddagger}). \quad (6)$$

Eqs. (5) and (6) can, in principle, be evaluated by a single free energy perturbation/umbrella sampling (FEP/US) approach using the mapping potential,  $\varepsilon_m$ , which drives the system from the reference potential to the target potential [52]. In such a case, the mapping potential can be written as a linear product of the two potentials:

$$\varepsilon_m = (1 - \lambda_m)E_{\text{REF}} + \lambda_m E_{\text{TARGET}} \quad (0 \leq \lambda_m \leq 1) \quad (7)$$

where  $\lambda_m$  is changed in fixed increments ( $\lambda_m = 0/m, 1/m, \dots, m/m$ ) from 0 to 1. The free-energy associated with moving between the two potentials ( $E_{\text{REF}}$  and  $E_{\text{TARGET}}$ ) can then be evaluated using the standard (FEP) formula [53]:

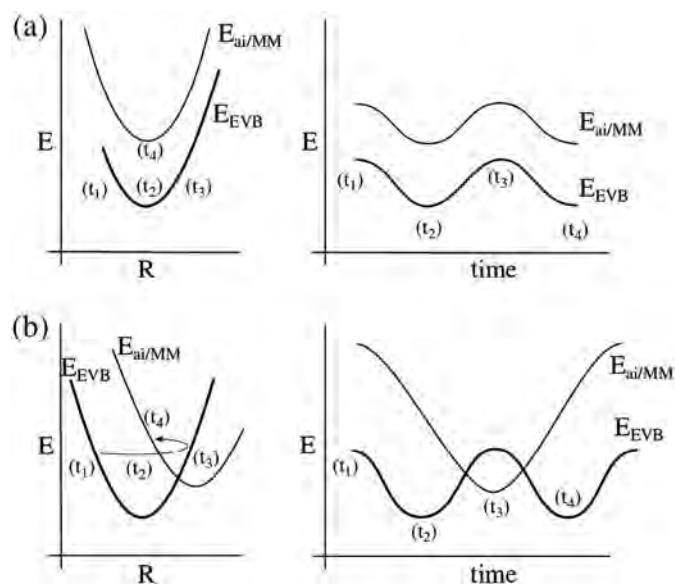
$$\delta G(\lambda_m \rightarrow \lambda_{m+1}) = -kT \ln \left[ \langle \exp\{-(\varepsilon_{m+1} - \varepsilon_m)/kT\} \rangle_{\varepsilon_m} \right] \quad (8)$$

where  $\langle \rangle$  designates an average over molecular dynamics (MD) trajectories propagated over  $\varepsilon_m$ . The final free energy for moving from  $E_{\text{REF}}$  and  $E_{\text{TARGET}}$  is taken as a sum of all free-energy increments:

$$\Delta G(\lambda_{n+1}) = \sum_{m=1}^n \delta G(\lambda_m \rightarrow \lambda_{m+1}). \quad (9)$$

In cases where the difference between the reference and target surfaces is significant (see Fig. 3) it might be impractical to use the expressions of Eqs. (7)–(9). Even in the case where similar surfaces are obtained for both reference and target surfaces, the use of a full FEP approach to gradually move from one surface to the other can be extremely computationally expensive [50]. An alternative, which also avoids the convergence problems inherent to FEP calculations [28], is to evaluate this correction term using the linear response approximation (LRA) [50]. The energy difference between





**Fig. 3.** Illustrating the factors that determine the energy difference between reference and target surfaces. The upper figure shows the situation when the two surfaces are similar. The lower figure shows the situation when the two surfaces are significantly different. In the last case, it is much harder to obtain converging results for  $\Delta G_{\text{REF} \rightarrow \text{TARGET}}$ . Reprinted with permission from [50]. Copyright 2002 American Chemical Society.

the reference and target potentials can be expressed within an LRA framework as follows [51]:

$$\begin{aligned} \Delta \Delta G_{\text{REF} \rightarrow \text{TARGET}}(\text{RS}) &= -kT \ln(Q_{\text{TARGET}}(\text{RS})/Q_{\text{REF}}(\text{RS})) \\ &= -kT \ln[\langle \exp\{-(E_{\text{TARGET}} - E_{\text{REF}})\beta\} \rangle_{\text{REF}}] \\ &\approx \frac{1}{2} (\langle E_{\text{TARGET}} - E_{\text{REF}} \rangle_{\text{REF}} + \langle E_{\text{TARGET}} - E_{\text{REF}} \rangle_{\text{TARGET}}). \end{aligned} \quad (10)$$

Similarly,  $\Delta \Delta G_{\text{REF} \rightarrow \text{TARGET}}(x^\ddagger)$  can be evaluated by the same LRA approach, but now considering the partial partition function at  $x^\ddagger$ :

$$\begin{aligned} \Delta \Delta G_{\text{REF} \rightarrow \text{TARGET}}(x^\ddagger) &= -kT \ln(q_{\text{TARGET}}(x^\ddagger)/q_{\text{REF}}(x^\ddagger)) \\ &= -kT \ln[\langle \exp\{-(E_{\text{TARGET}}(x^\ddagger) - E_{\text{REF}}(x^\ddagger))\beta\} \rangle_{\text{REF}}] \\ &\approx \frac{1}{2} (\langle E_{\text{TARGET}}(x^\ddagger) - E_{\text{REF}}(x^\ddagger) \rangle_{\text{REF}} + \langle E_{\text{TARGET}}(x^\ddagger) - E_{\text{REF}}(x^\ddagger) \rangle_{\text{TARGET}}). \end{aligned} \quad (11)$$

Note here that the lowercase  $q$  in Eq. (11) refers to the partition function at the TS, in contrast to the uppercase  $Q$  in Eq. (10), which referred to the partition function at the minima. While both approaches are viable, the LRA has been shown to be particularly powerful as it allows one to obtain reasonable results even in cases where the target and reference potentials are significantly different [28,51].

Up to this point, it has been implicitly assumed that the position of the transition state (TS)  $x^\ddagger$  is identical for the two free energy surfaces, i.e.  $x^\ddagger = x^\ddagger_{\text{REF}} = x^\ddagger_{\text{TARGET}}$ . However, this is not necessarily true as the two TS can be different on the two surfaces using the different potentials, and therefore this approximation may no longer be valid [51]. To resolve this problem, it is therefore necessary to include the free energy difference involved when moving from  $x_{\text{REF}}^\ddagger$  to  $x_{\text{TARGET}}^\ddagger$  in Eq. (6):

$$\Delta G_{\text{TARGET}}(x_{\text{TARGET}}^\ddagger) = \Delta G_{\text{REF}}(x_{\text{REF}}^\ddagger) + \Delta \Delta G_{\text{REF} \rightarrow \text{TARGET}}(x_{\text{REF}}^\ddagger) + \Delta \Delta G_{\text{TARGET}}(x_{\text{REF}}^\ddagger \rightarrow x_{\text{TARGET}}^\ddagger). \quad (12)$$

The last term of this expression can be evaluated by calculating the potential of mean force (PMF) of moving from  $x_{\text{REF}}^\ddagger$  to  $x_{\text{TARGET}}^\ddagger$  on the

target surface, which can be done by defining two sets of constraints (one for  $x_{\text{TARGET}}^\ddagger$  and the other for  $x_{\text{REF}}^\ddagger$ ) and pulling the system from one constraint to the other. Using the LRA formulation, the last two terms from Eq. (12) can be evaluated in one step:

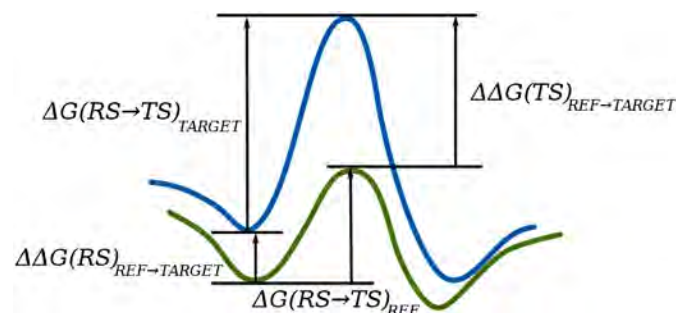
$$\begin{aligned} \Delta \Delta G_{\text{REF} \rightarrow \text{TARGET}}(x_{\text{REF}}^\ddagger \rightarrow x_{\text{TARGET}}^\ddagger) &= \Delta G_{\text{TARGET}}(x_{\text{TARGET}}^\ddagger) - \Delta G_{\text{REF}}(x_{\text{REF}}^\ddagger) \\ &= \frac{1}{2} (\langle E_{\text{TARGET}}(x_{\text{REF}}^\ddagger) - E_{\text{REF}}(x_{\text{REF}}^\ddagger) \rangle_{\text{REF}} + \langle E_{\text{TARGET}}(x_{\text{TARGET}}^\ddagger) - E_{\text{REF}}(x_{\text{TARGET}}^\ddagger) \rangle_{\text{TARGET}}). \end{aligned} \quad (13)$$

In this way, one can quantitatively evaluate the perturbation necessary when moving between reference and target potentials within a QM/MM (or any other related) framework.

### 3. Performing QM/MM free energy calculations using a reference potential

The use of a reference potential can be particularly useful in the study of complex reactions in solution and/or enzymes, where one is interested in describing the chemical reaction pathway while including the solvent and/or enzyme environment (Fig. 4). To achieve this, the challenge is again two-fold: first, one needs to use an accurate and computationally efficient description of the bond breaking/forming process. Second, one needs to take into account the effect of the environment and efficiently explore the complex energy landscape of the system in order to avoid fictitious minima.

In the first case, one option is to treat all components of the molecular system with an accurate level of theory, e.g. using a high-level quantum mechanical (QM) approach. However, this strategy is severely limited by the poor scaling of conventional QM approaches. For example, the computational cost of the Kohn–Sham density functional theory (DFT) grows cubically with the number of particles [54], thus making it quickly impractical as the complexity of the system increases. In this respect, there have been a number of promising advances in the field to substantially reduce this computational challenge, in most of the cases by using a reformulation of DFT and/or applying localization constraints [13]. Some of these approaches have been shown to scale almost linearly with system size [13,55,56], thus opening the door towards the application of DFT approaches to larger systems [13–15]. Among these developments, a few examples of note are constrained DFT [55], where an experimentally or physically motivated constraint is applied to the density during the minimization of the DFT energy functional [57]; orbital-free DFT [58,59], which approximates the kinetic energy of non-interacting electrons in terms of their density; and orbital-free embedded DFT, in its different variants, such as those by Goodpaster et al. [60] and Wesolowski et al. [61]. Another DFT alternative includes the semi-empirical self-consistent charge density functional tight binding (SCC-DFTB) methods [62,63], which involves the approximation



**Fig. 4.** Thermodynamic cycle used to evaluate the free energy difference between a high-level surface (blue) and a lower level surface (green). The corresponding free energy barrier for the high-level surface,  $\Delta G_{\text{TARGET}}(x^\ddagger)$ , is estimated as the free energy barrier of the reference potential  $\Delta G_{\text{REF}}(x^\ddagger)$  plus the free energy of moving from the reference to the target surface  $\Delta \Delta G(x^\ddagger)$  (shown by black arrows). Reprinted with permission from [51]. Copyright 2006 American Chemical Society.

and parametrization of interaction integrals. These methods have shown great success, but also some limitations, in comparison to higher-level models, in particular in terms of energy convergence between exact and linear-scaling methods (see Ref. [64,65] and references cited therein). Another active area of research has been the development of DFT-algorithms for graphics processing units (GPU) [66–69]. Compare to standard CPU implementations these algorithms have shown a significant speed-up in gradient and energy calculations. The remarkable speedups attained by these implementations have made it possible to even carry out *ab initio* MD simulation of large systems at a very low cost using standard desktop computers [70].

In the second case, an efficient sampling method is required in order to explore the broad phase space including the influence of the environment. The use of molecular mechanics (MM) approaches, which are based on classical potentials, can be extremely helpful as they allow inclusion of environmental effects (either solvent molecules or parts of the enzyme) and improve the sampling of the energy surface on which the reaction pathway is calculated. However, since MM methods do not describe the electronic rearrangements involved in the breaking and forming of chemical bonds, they cannot be used for the study of chemical processes. Consequently, one needs to find a balance between an accurate description of the electronic process and efficient modeling of the complex environment of the reaction. In this regard, one of the most widely used approaches is a multi-layer approach (Fig. 5), in which the interesting part of the system (usually where the chemical reaction takes place) is described at the electronic level by (high-level) QM approaches, with the remainder of the system being modeled by MM (or lower-level QM) methods.

These combined quantum mechanical/molecular mechanical (QM/MM) approaches [71–73], which are part of the reason for their 2013 Nobel Prize in Chemistry [74], have currently become one of the most popular approaches for studying chemical reactivity in solution and in enzyme active sites [5,20]. Depending on the level of theory used, QM/MM approaches can be classified into two different types. The first of these employs a semi-empirical description of the reactive part, using approaches such as AM1/d [75], orthogonalization corrections (OMx-D) [76,77], SCC-DFTB, or the empirical valence bond (EVB) approach [52,78]. The second type relies on the use of *ab initio* (wave-function based) or, more commonly, DFT methods to describe the QM region. Other *ab initio* QM/MM methods, using a valence bond approach, have also been developed. Among these approaches, one can mention the mixed molecular

orbital and valence bond (MOVb) method by Gao et al. [79] and the hybrid *ab initio* VB/MM method by Shurki et al. [80]. Additionally, these methods can be further sub-divided as “additive” or “subtractive” approaches, in terms of the scheme used to express the Hamiltonian of the system. Here we will only briefly describe the additive scheme, as it is currently the most widely used QM/MM approach. In this scheme the effective QM/MM Hamiltonian of the system can be defined by:

$$\hat{H}_{QM/MM} = \hat{H}_{QM(QM)} + \hat{H}_{MM(MM)} + \hat{H}_{QM/MM(QM-MM)} \quad (14)$$

where  $\hat{H}_{QM(QM)}$  is the energy of the QM region (calculated at the QM level of theory),  $\hat{H}_{MM(MM)}$  is the energy of the MM region (calculated at the MM level of theory or using a lower level QM approach), and  $\hat{H}_{QM/MM}$  is the coupling between the two regions. The last term in this expression is crucial, and includes the bonded, electrostatic and van der Waals interactions between the atoms in the two regions. Without correct treatment of this coupling, completely unphysical energies are obtained [20,36]. For a detailed description of both additive and subtractive QM/MM approaches, we refer the reader to Ref. [20].

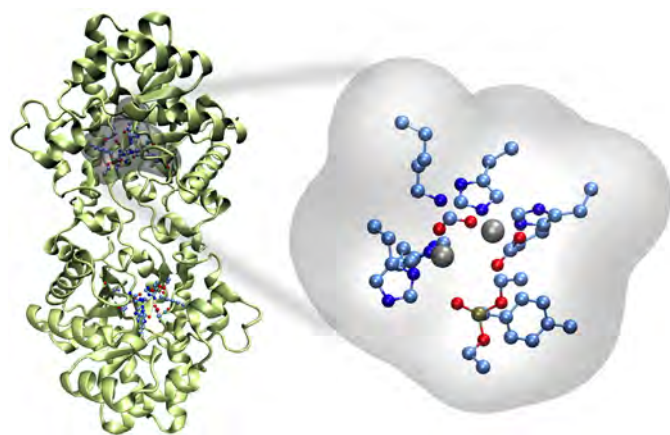
One of the major challenges inherent to *all* QM/MM approaches is the high computational cost needed for the repeated evaluation of the energies and forces in the QM region. Additionally, due to the large number of atoms included in these calculations, the number of local minima available for the system substantially increases [81], making QM/MM calculations based on a single minimized structure unreliable (as they may not be representative of the chemical process of interest). Due to these limitations, there have been efforts to develop QM/MM free energy methods [19,28,30,31,51,82–85] which aim to avoid direct sampling at high levels of theory while still giving an accurate estimate of the free energy changes during the reaction. This approach, initially developed by Warshel et al. [27], has been employed and modified by a number of other authors [19,30,31,82–84,86,87] by using the thermodynamic cycle of Fig. 2, auxiliary Monte Carlo (MC) simulations with an approximate Hamiltonian [82–84] or a combination of both [87] approaches.

To finish this section, we will briefly describe two other related approaches that use a reference potential. The first of them is the QM/MM thermodynamic cycle perturbation (QTCP) approach of Ryde et al. [30,88]. This method can be considered a combination of the QM/MM free energy method (QM/MM-FE) of Yang et al. [89] and the reference potential approach of Warshel et al. [27,50]. Here, as in the QM-FE method of Yang et al. [89], the reaction pathway is optimized (in this case using a semiempirical QM/MM reference potential) and a number of configurations are selected along the reaction pathway. Then, following the approach of Warshel et al. [27,50], the free energy change from the reference to the target potential is evaluated via FEP for each of the selected QM configuration along the reaction pathway. In this way, a high-level QM/MM PMF can be obtained. Ryde et al. have applied the QTCP approach, for example, to study the enzymes catechol O-methyltransferase (COMT) [30], cytochrome c peroxidase [90], and [FeFe]-hydrogenases [91].

The second approach is the molecular mechanics based importance function (MMBIF) developed by Iftimie and Schofield [92]. In this method, a reference MM Hamiltonian is used to guide the MC sampling, from which different configurations are generated. These configurations are then accepted or rejected into the QM/MM ensemble (from which the QM/MM free energies can then be calculated) according to the Metropolis–Hastings algorithm with the probability  $\min\{1, \exp(-\Delta\Delta E/k_B T)\}$ , where  $\Delta\Delta E$  is the difference between the target and reference energies for configuration  $x$ :

$$\Delta\Delta E = (E_{TARGET}(x_{new}) - E_{REF}(x_{new})) - (E_{TARGET}(x_{old}) - E_{REF}(x_{old})). \quad (15)$$

MMBIF has been applied to the study of the proton transfer reaction of malonaldehyde in an aprotic nonpolar solvent using DFT as a target



**Fig. 5.** Illustrating an example of the division of an enzyme into QM (reactive part) and MM (environment) regions. Shown here is the specific example of the zinc-containing bacterial phosphotriesterase from *Brevundimonas diminuta*, in complex with the substrate analog diethyl 4-methylbenzylphosphonate (PDB ID: 1DPM). Here, a potential QM region has been highlighted on the enzyme.

potential, showing an overall reduction of three orders of magnitude in the CPU simulation time compared to a direct *ab-initio* MC simulation [92].

Finally, the simplest way towards this correction is to calculate the reaction profile at a low level of theory and then subsequently perform single point calculations with a higher-level approach on a sample of individual structures to get the higher-level reaction profile [93–98]. Strictly, these structures should be reweighted, as it is not certain that the phase space sampled by the low-level methods is the same as with the higher-level methods [19]. Regardless of the preferred approach, the use of a reference potential is clearly a promising way to overcome the computational costs associated with directly performing all the configurational sampling with a high-level approach.

#### 4. Free energy calculations using updated mean charge distributions and reference potentials

As mentioned in the previous section, calculating free energies within a QM/MM framework is computationally demanding. In standard QM/MM calculations using the additive scheme (Eq. (14)), the bottleneck is the calculation of the QM internal energy and QM/MM electrostatic interactions ( $\hat{H}_{QM(QM)}$ ,  $\hat{H}_{QM/MM, Elec}$ , Eq. (14)), as they require a self-consistent field (SCF) calculation for each different QM or MM conformation. The cost associated with this makes it extremely challenging to perform extended sampling.

One straightforward alternative for accelerating QM/MM simulations is the use of a mean-field approximation, by which the solvent environment is represented by an average solvent potential which is then added to the solute Hamiltonian [99]. In its earliest incarnations, such an approach was implicitly implemented in, for example, the QM/Langevin dipole (QM/LD) [49,100] and in various continuum models [101]. In the last decade, the use of an average solvent potential to accelerate QM/MM has been exploited in different ways. This includes the work of Warshel et al. [99], Yang et al. [31,45], and Aguilar et al. [102,103], among others [104,105]. This approach has been used to study activation energies in solution [106–108], enzymes [98,109–111], solvation free energies [87,112,113], and binding free energies [113].

The common point in all these approaches is the calculation of a partially [99] (Fig. 5C) or totally [102] averaged MM potential, which is normally evaluated by some variation of collecting independent positions of all solvent atoms over a number of frames of an MD trajectory, and then averaging over all solvent positions, which are sent to the QM program as point charges [108]. Even though the use of a point-charge approximation is straightforward and in most cases works quite well [108], it has shown some limitations, such as its tendency to overestimate short-range electrostatic (ES) interactions [114], as well as divergence problems during the SCF iterations [104]. This problem has been addressed by re-scaling the MM charges [99,115] or by smearing the MM charges using Gaussian distributions [116]. Another common aspect of these approaches is the use of a fixed structure of the QM system throughout the MD simulation. This approximation is extremely helpful to overcome convergence problems associated to FEP calculations (see discussion in Ref. [19,112,113]). However, its use implies that the entropy of the QM system is ignored [81] and that the increase in the size of the QM system does not ensure improvement of the results, *i.e.* just increasing the size of the QM system does not necessarily ensure that better results will be obtained [19,117]. These approaches can also be extended to cases where the solute is allowed to fluctuate as well [99]. Recent works proposed a simpler form of the Zwanzig equation in which differences in interaction energy, rather than total energy differences, are factored into the FEP formula [113].

To put this in context with an illustrative example, we will briefly describe the approach developed by Warshel and coworkers [99,118] to accelerate QM/MM free energy calculations using an updated mean charge distribution coupled with a classical reference potential. This is a multi-step protocol, where the free energy for a given process is first

calculated by performing classical MD simulations using a reference potential and then refined using a higher-level method. In the case of solvation free energy calculations, these can be addressed by using the given fixed solute with a reasonable charge set as a reference and then refining the calculated free energy while taking the real fluctuating MM environment over the course of the QM/MM simulation into account.

In order to do this refinement, Warshel et al. [99] developed a seemingly simple approach which allows one to calculate the QM/MM electrostatic interactions without the need for repetitive SCF calculations. This approach, which maps the effect of the fluctuating MM environment on a grid of point charges, can be summarized in the following three steps: (1) initial evaluation of QM charges,  $Q^{(n)}$  (where  $(n)$  designates the  $n$ th step), followed by (2) MD simulation over  $m$  steps to allow the MM environment to fluctuate in the potential ( $E_{QM/MM}^{el}(Q^{(1)}) + E_{vdW}$ ); and finally, (3) all  $m$  snapshots of solvent coordinates from  $m$  MD steps are stored. Then, in order to perform the averaging, the charge of these atoms is scaled by  $1/m$  and all the scaled solvent charges are sent to the QM program to reproduce an average solvent potential on the solute. The latter is used to obtain the corresponding solute polarization and a new set of solute charges  $Q^{(2)}$  (Fig. 6). The average solvent potential can be evaluated in a computationally less costly way [99] by partitioning the MM environment into two (or more [107]) regions: in the first region (Region I) the average potential is calculated on a grid of  $m \times N_{\text{extern}}$  point charges (where  $N_{\text{extern}}$  are the MM atoms situated within a predefined cutoff radius, scaled by  $1/m$ ) and included into the Hamiltonian within the QM program. In the second region (Region II), the average potential coming from  $N - N_{\text{extern}}$  MM atoms is represented by a dipole using the following relationship:

$$\xi_o = \frac{2q}{|\mathbf{r}_{OR}|^3} \mathbf{r}_{OR} \quad (16)$$

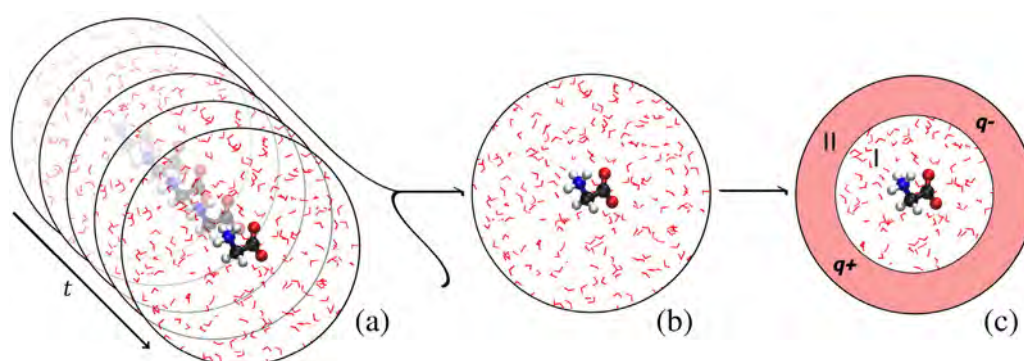
where  $\xi_o$  is electric field at point O (the geometrical center of QM system) and  $\mathbf{r}_{OR}$  is pointing along  $\xi_o$  to charge  $q$ . The validity of this averaging approach has been demonstrated in Ref. [99], with substantial savings in computational cost compared to standard QM/MM approaches using repetitive SCF calculations. The use of this multi-step protocol, which includes the updated charge mean distribution, the use of a classical reference potential, and the LRA approach (to get the energy difference between the two potentials, Eq. (12)) has demonstrated to lead to computational time savings of up to 1000× in QM(ai)/MM-FEP calculations of solvation free energies [99] as well as showing promise for the correct treatment of protein electrostatics in benchmark  $pK_a$  calculations [118].

#### 5. Automated refinement of the reference potential for QM/MM free energy calculations

One of the big challenges when allowing the solute to fluctuate is that the difference between the reference and target potentials can be quite large (and, in the case of chemical reactions, the transition states have different positions on the two potentials). In order to partially avoid this problem, one could freeze out the solute motions and do free energy calculations on static structures, as described in Section 4. The problems, however, start in the case of chemical reactions where the solute potential is allowed to fluctuate.

We recently developed an approach to overcome this problem, which we describe as “paradynamics” (PD) [28,119]. In this context, PD provides an alternative gradual and controlled way to make the physically-based reference potential as close as possible to the target potential in order to reduce the discrepancies between the two potentials during the actual simulations [119] (for a detailed description of this method we refer the reader to Refs [28,119]). The PD approach [28] also follows the cycle shown in Figs. 2 and 3. It starts with the evaluation of the energy difference between the reference and the target





**Fig. 6.** Schematic view of the averaging of the solvent potential over  $m$  steps of a standard classical MD simulation. (a) MD sampling of the MM subsystem is performed while keeping the QM region fixed; (b) all snapshots of the trajectory are combined to build a finite MM ensemble; (c) Simplified form of representing the solvent potential: in Region I it is calculated from the average of the explicit molecules, while in Region II the average potential is represented by two point charges.

Adapted with permission from [99]. Copyright 2008 American Chemical Society.

potential, which is done for each of the geometries generated by MD trajectories at the RS (PS) and at the TS and in both the reference and the target potentials. This evaluation can be done by using either the FEP approach or the LRA perturbation between the different potentials. While both approaches have been shown to give nearly identical results, as the LRA approach is just an endpoint approach, it has been shown to give significant time savings compared to the full FEP treatment [119]. This step is followed by iterative refinement of the reference potential in order to minimize the energy difference between the two potentials, as described below and shown in Fig. 7.

While in principle any reference potential could be used for such calculations, e.g., either a pure MM Hamiltonian [30,86] or an empirical valence bond potential (EVB) potential [50,51], we started by using the EVB potential for simplicity. The EVB approach (which has been discussed in detail elsewhere [52,120]) has proven to be a highly useful and efficient approach in a number of studies [28,51,99,118,119,121], as it is fast enough to allow both extensive conformational sampling and provides sufficient chemical information to describe chemical processes

under study. Additionally, the use of the energy gap between the two diabatic states in the EVB model ( $\Delta\epsilon = \epsilon_j - \epsilon_i$ ) as a reaction coordinate has been shown to be particularly powerful in simulations of chemical reactions, especially in complex systems, where the selection of a proper reaction coordinate can be a major bottleneck [120] in the process.

When an EVB potential is used as a reference potential, the refinement procedure involves the refinement of the EVB parameters, which is done by seeking the minimum of the least-squares function [28]:

$$\Theta(\mathbf{p}, \mathbf{r}) = k_1 \sum_{i=1}^N \left( E_i^{\text{EVB}}(\mathbf{p}, \mathbf{r}) - E_i^{\text{QM}}(\mathbf{r}) \right)^2 + k_2 \sum_{i=1}^N \sum_{j=1}^3 \left( \frac{\partial E_i^{\text{EVB}}(\mathbf{p}, \mathbf{r})}{\partial x_j} - \frac{\partial E_i^{\text{QM}}(\mathbf{r})}{\partial x_j} \right)^2 \quad (17)$$

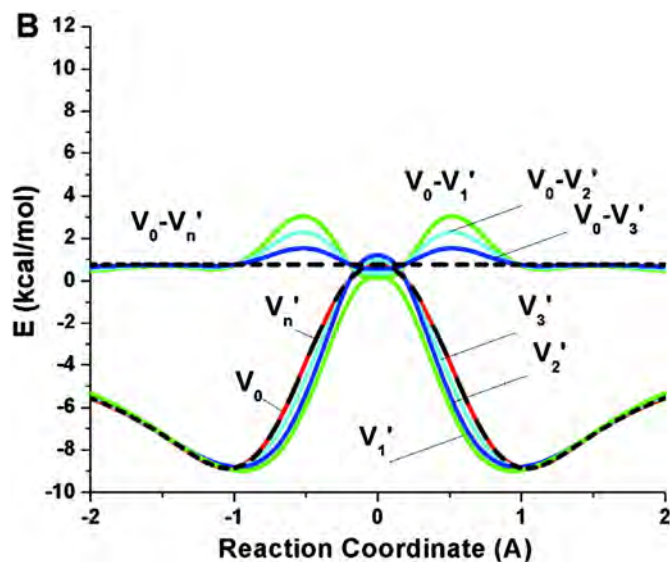
$E^{\text{EVB}}$  and  $E^{\text{QM}}$  in Eq. (17) denote the EVB and QM energies respectively,  $\mathbf{p}$  is the vector of the EVB parameters,  $i$  runs over the configurations generated during the MD simulation, and  $k_1$  and  $k_2$  are the weighting factors that are applied to the energies. For each EVB parameter,  $\mathbf{p}$ , an iterative search for the value that minimizes  $\Theta(\mathbf{p}, \mathbf{r})$  is carried out either by using a simplified Newton–Raphson approach or by refining the vector of EVB parameters,  $\mathbf{p}$ , in the optimal steepest descent approach [119].

While the refinement scheme given above refers specifically to an EVB potential, another refinement scheme with another kind of reference potential (such as a pure MM or semiempirical potential [30, 86]), can also be used by using the full FEP or LRA to evaluate the perturbation between the different potentials. For these cases, a recent approach has been developed for the refinement step, which involves the use of Gaussian functions. Such Gaussian functions have been previously used in different contexts [43,122,123], for example in the local elevation method [122] and in the MTD [43] approach. However, in contrast to these methods, the PD approach does not require expensive iterative sampling to build the negative potential [119]. It should be pointed out in any comparison of PD and MTD that MTD is “only” an enhanced sampling approach, whereas PD is a combined multi-scale enhanced sampling method. This makes a direct comparison of the performance of the two approaches more difficult.

In the PD approach the least-squares function is minimized, using the optimal steepest descent with respect to the parameters  $\alpha_k$  and  $A_k$  of a number ( $i$ ) of Gaussians ( $A_i \exp(-\alpha(\xi - \xi_i)^2)$ ):

$$\Theta(\{a_k\}\{A_k\}) = \sum_{i=1}^N \left( T^{\text{TARG}}(\mathbf{r}_i) - E^{\text{TARG}}(\mathbf{r}_i) \right)^2 \quad (18)$$

The result of this fitting is a function,  $\Gamma$ , which approximates the target potential in the range of the reaction coordinate where the PES was



**Fig. 7.** In the PD approach, the reference potential,  $V_1'$ , is iteratively and systematically refined ( $V_2'$ ,  $V_3'$ , ...,  $V_n'$ ) to reproduce the real potential,  $V_0$ . The iterative refinement of the reference potential,  $V_n'$ , is aimed at reaching a situation where  $V_0$  and  $V_n'$  almost coincide. Once this is achieved, the computationally expensive sampling on the real potential can be substituted by sampling the reference potential instead. Reprinted with permission from [28]. Copyright 2011 American Chemical Society.

performed:

$$\Gamma_{\text{TARGET}} = \sum_k A_k \exp(-\alpha_k(x - \xi_k)^2). \quad (19)$$

Following this procedure for both the reference and the target potentials, it is possible to refine the original reference potential using the obtained r-functions:

$$E'_{\text{REF}} = E_{\text{REF}} - \Gamma_{\text{REF}} + \Gamma_{\text{TARGET}}. \quad (20)$$

In this way, the refined reference potential ( $E'_{\text{REF}}$ ) is close enough to the target potential in a given range of the reaction coordinate. Then, the mapping potentials used in order to go from RS to PS can be expressed as:

$$E_m = (1 - \lambda_m)E_{\text{RS}} + \lambda_mE_{\text{PS}} \quad (21)$$

where  $E_{\text{RS}} = E_{\text{QM}} - \Gamma_{\text{QM}} + E_{\text{CONS}}(\xi_{\text{RS}})$  and  $E_{\text{PS}} = E_{\text{QM}} - \Gamma_{\text{QM}} + E_{\text{CONS}}(\xi_{\text{PS}})$ . Here, one can use  $E_{\text{CONS}}(\xi_{\text{RS/PS}})$  as a harmonic potential centered at the reaction coordinate position  $\xi_{\text{RS/PS}}$ . Then, the PMF is evaluated by using a modification of Eq. (8) (using the FEP/US approach):

$$\Delta g(\xi) = \Delta G_m - kT \ln \left\langle \delta(x - \xi) \exp\left(-\frac{E_{\text{TARG}}(x) - E_m(x)}{kT}\right) \right\rangle_{E_m}. \quad (22)$$

Furthermore, the results from different simulation windows are combined by:

$$\Delta \bar{g}(\xi) = \sum_{\text{frames}} \frac{N_i(\xi)}{\sum_{\text{frames}} N_i(\xi)} \Delta g_i(\xi) \quad (23)$$

where  $N_i(\xi)$  is the number of times the MD visited a particular reaction coordinate value,  $\xi$ , while propagating on the  $i$ -th mapping potential.

The generalized features of PD approach allow the use of this approach not only to explore complex systems by acceleration QM(ai)/MM calculations with proper sampling, but also in a wide variety of applications where a reference potential is used (e.g. when the reference potential is a CG model), thus providing a very powerful strategy to explore any dynamical property of a complex system in an efficient way.

A key point that should be kept in mind when defining the level of theory of the target potential is its reliability to describe the chemical properties of the system of interest. Currently, most *ab initio* QM/MM approaches use DFT methods to describe the chemical process. Even though DFT has been shown to be highly efficient compared to other *ab initio* approaches (see discussion in Section 2), it has also been shown to suffer a number of limitations [124–126], which can be more or less critical depending on the particular characteristics of the chemical process under study. One example of this is the extensive use of the B3LYP density functional in the field of chemistry; this functional has proved to be accurate enough in a number of studies. However, it is now clear that its success has, in many cases, been simply the result of some cancelation of errors [124]. Attempts to solve these errors, by inclusion of physically motivated corrections have been done (e.g. the rCMB3LYP and LC-BLYP functionals), however they seem to come at the cost of a slightly worse description of the chemistry. On the other hand, some recent empirically-driven functionals, such as M06-2X, have shown excellent performance, but at the cost of a high level of parameterization [125]. Therefore, one should carefully examine the different higher-level approaches available as a target potential to choose an accurate, but also computationally efficient, approach. In summary, the quality of the target potential will be directly related to the quality of the QM approach used and, therefore, knowledge about

the methods used will be an essential aspect when using simplified models to describe a complex process.

## 6. Sample applications

The reference potential approach has been adopted and implemented in a number of methods aimed at QM(ai)/MM calculations of activation free energies in enzymes [30,45,51,98,109–111], aqueous solution [106], and the evaluation of solvation free energies [87,107,112,113], binding free energies [113,127],  $pK_a$  calculations [118], and kinetic isotope effects [83]. In this section, we present three illustrative examples of the applications of this approach.

### 6.1. The chloride/chloromethane $S_N2$ reaction: implementing the paradynamics refinement approach

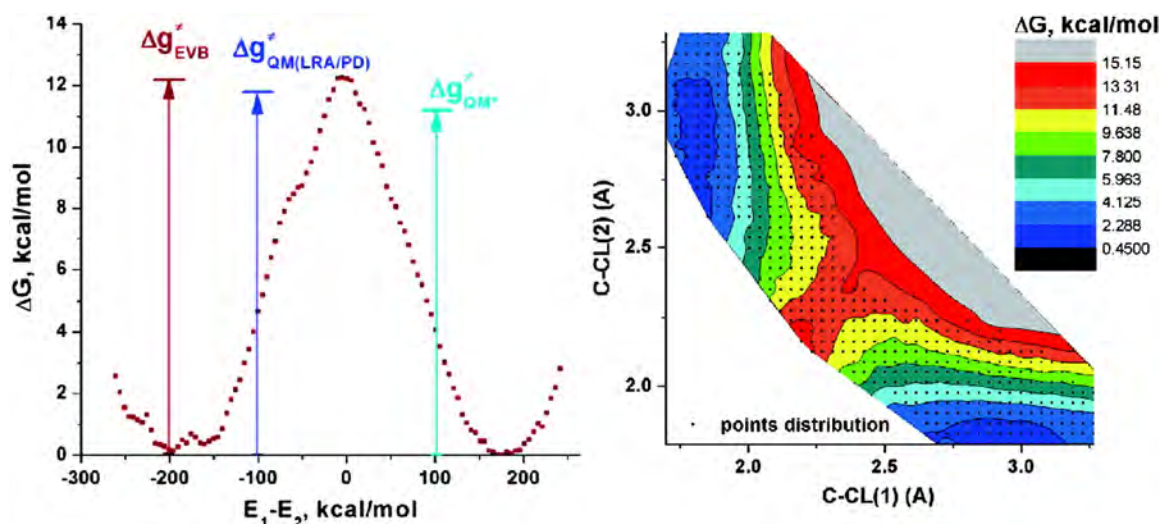
The  $S_N2$  reaction between chloride ion and chloromethane has been extensively used as a prototype reaction, not only to understand solvation effects, but also as a model system to validate a number of theoretical approaches [106,128,129] including the PD approach [28] presented in Section 4. For the PD study of this reaction, an EVB potential was used as a reference potential and the automated PD parameter refinement procedure (Eq. (17)) was used to find the set of EVB parameters that minimize the free energy difference between the reference (EVB) and the target potential (chosen to be the PM3 level of theory to reduce computational cost). Those parameters were then used to obtain the activation free energy in the reference potential  $\Delta g_{\text{EVB}}(x^\ddagger)$  (using Eqs. (22) and (23)) and the correction  $\Delta \Delta g_{\text{EVB}} \rightarrow \text{QM}(x^\ddagger)$  (Eq. (6)) was calculated using the LRA approach. Using this multistep approach, it was possible to obtain an activation barrier of 11.7 kcal/mol, which was in good agreement with the actual QM barrier of 11.2 kcal/mol (at the PM3 level of theory).

Another aspect analyzed in that work was the selection of the reaction coordinate, where the applicability of the PD approach was demonstrated not only when the EVB energy gap ( $\Delta \epsilon$ ) is used as the reaction coordinate, but also in cases where other reaction coordinates are used, for example with the conventional  $S_N2$  geometrical coordinate ( $r(\text{C}-\text{Cl}_1) - r(\text{C}-\text{Cl}_2)$ ) or the 2-D reaction coordinate ( $r(\text{C}-\text{Cl}_1)$  and  $r(\text{C}-\text{Cl}_2)$ ) (Fig. 8). In contrast to the MTD approach, which requires a separate run for each choice of reaction coordinate, the PD approach only requires a single post-processing job after the PD refinement in order to generate the refined free energy surface. In this aspect it is important to mention that variations of the PD approach have also been used to generate multidimensional surfaces, for example to examine the free energy landscape for both the conformational change and the chemical step in adenylate kinase [130]. Finally, this study also tested the efficiency of PD in comparison to the popular MTD approach [42] and it was shown that the PD approach was 200 times more computationally efficient than the MTD approach. The precise acceleration factor depends on the cost of the QM method used, as the main bottleneck is still the cost of the repeated QM calculations in the MTD approach.

### 6.2. Exploring the stability of different anionic tautomers of uracil

Anionic states of nucleic acid bases have been suggested to play a role in radiation damage processes. Several studies have focused on the stability of these species, mainly in the gas phase [131,132], but also in solution [107]. Haranczyk et al. [107] have used the reference potential approach in order to calculate the stability of five different anionic tautomers in the solvent phase (Fig. 9). Their methodology for the calculation of solvation free energy involves the two-step approach described in Section 3. In this particular case, the first step involved the calculation of the free energy of solvation using the free energy perturbation-adiabatic charging (FEP/AC) approach [52,133], where the charges of the solute molecule are obtained using a continuum solvent model. In the second step, the classical free energy of solvation is





**Fig. 8.** (left) The EVB free energy surface obtained for a model  $S_N2$  reaction in the gas phase (the symmetrical reaction between chloride ion and chloromethane) after the refinement of the EVB parameters. The refined EVB free energy profile is shown in red. Also shown for comparison are a red arrow representing the EVB free energy barrier, a blue arrow shows the barrier obtained through the LRA-correction after PD refinement, and a cyan arrow designating the real QM barrier. As can be seen from this figure, the difference between even the uncorrected EVB activation barrier using a refined potential and the “real” higher-level QM activation barrier is less than 1 kcal/mol. (right) 2-D free-energy surface constructed by EVB FEP/US after the PD refinement. The surface is defined in terms of the distances between the central carbon atom and the leaving and attacking nucleophiles (i.e.,  $r(C-Cl_1)$ ,  $r(C-Cl_2)$ ). Reprinted with permission from [28]. Copyright 2011 American Chemical Society.

refined by taking into account the effect of the explicit solvent (as illustrated in Fig. 6). From this study it was shown that three anionic tautomers (U2, U4, and U1; see Fig. 8) are more stable than the canonical one. Comparison of these results with the one obtained using only a continuum model showed significant differences, suggesting the need for the explicit treatment of solvent on these studies.

### 6.3. Dehalogenation of dichloroethane by haloalkane dehalogenase

The  $S_N2$  step in the mechanism of the reaction catalyzed by haloalkane dehalogenase (Dh1A) has been the subject of many theoretical studies [19,134–136]. Warshel et al. [51] have studied this process in order to exemplify the use of the EVB approach as a reference potential for QM/MM calculations in the condensed phase. More recently, Ryde et al. [19] have also used this system as a benchmark for their QM/MM thermodynamic cycle perturbation approach (QTCP) approach. As discussed before both approaches share several features.

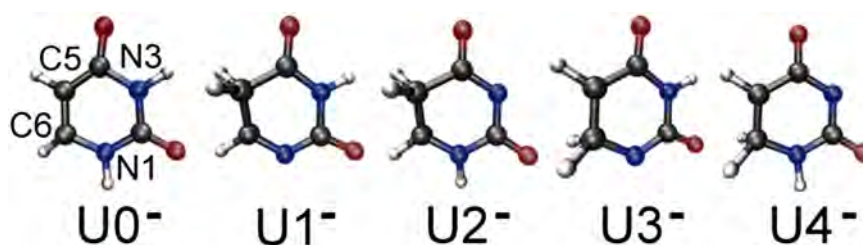
In the study of Warshel et al. [51] (which builds on earlier detailed mechanistic studies of the same system [135]), a combined approach was applied using the EVB approach as a reference potential, and the LRA approach, to evaluate the energy cost of moving from the EVB to the QM(B3LYP/DFT)/MM surfaces. These results were then used to evaluate the activation barrier using Eqs. (6) and (11), where it was

assumed that the TSs for both QM/MM and EVB potentials were similar, and also using Eq. (13), where the energy difference between the two TSs is also included. Both approaches produced very similar activation barriers (Fig. 10).

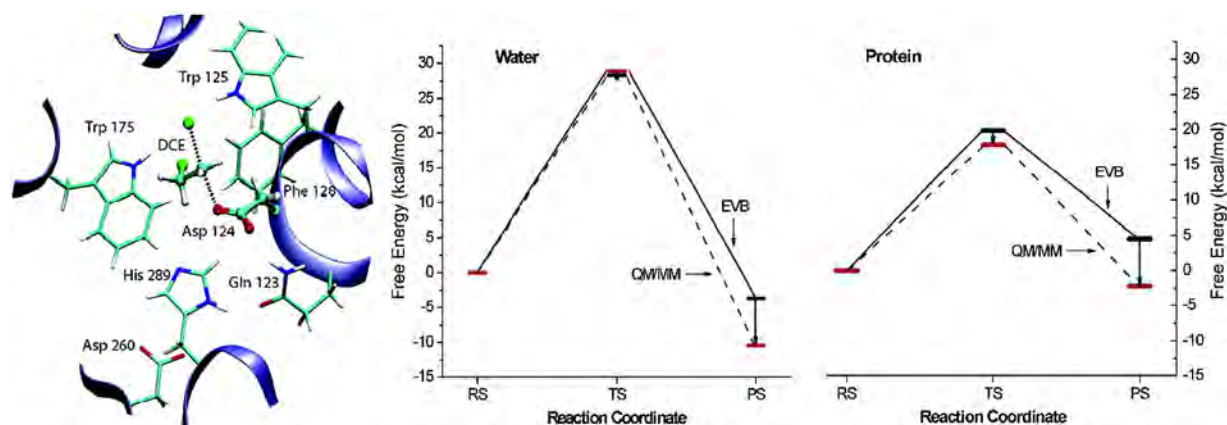
In the work by Ryde et al. [19] this system was used as a benchmark to compare the convergence when calculating the energy difference from the MM (or semiempirical) description to a QM/MM description, using the QTCP approach. For the QM calculations, the DFT level of theory (using three different functionals: PBE, B3LYP and TPSSH) was tested. It was concluded that simulations based the semiempirical QM/MM methods have slightly better convergence properties, however they require a longer simulation time (~10 ns). Additionally, it was shown that convergence is only obtained when electrostatic interactions between the QM region and its surroundings are ignored. Finally, it was demonstrated that the use of single point energy calculations, without any reweighting of the trajectories, gives poor results, at least when a large number of frames is used to obtain the average correction.

## 7. Conclusions and future perspectives

Hybrid QM/MM approaches have rapidly become one of the most popular computational tools for studying chemical reactivity in biomolecular systems. Despite tremendous recent advances in computer power and improvements in the scaling of conventional QM codes, the



**Fig. 9.** Anionic tautomers of uracil studied by Haranczyk et al. [107]. This figure was originally presented in Ref. [107] with color added for clarity.



**Fig. 10.** (A) TS of the  $S_N2$  reaction step in the active site of DhlA. (B) QM/MM activation free energy for the  $S_N2$  reaction step in water and in the active site of DhlA. This profile was obtained by moving from the EVB to the higher-level (DFT or *ab initio*) QM/MM surfaces, as outlined in Ref. [51]. This figure was originally presented in Ref. [51].

cost of high-level descriptions of the QM region still remains out of the reach of QM/MM free energy calculations [19] as they require extensive configurational sampling. Therefore, alternative approaches are needed to work around this problem. The use of simplified models to describe complex biological systems has a long and distinguished history, dating back to the early days of protein folding simulations [22] and continuing today with simulations on such large systems as the translocon [137], F1 ATPase [138,139] and the ribosome [11]. We present here some recent advances in using reference potentials and mean field approximations to accelerate high-level QM/MM calculations. In particular, we are optimistic about the promise of the paradynamics approach as an effective tool to reduce the cost of high-level QM/MM calculations by using a mostly automated parameter refinement procedure to limit the amount of sampling necessary on the high-level surface.

Despite the many recent advances in the field, a number of challenges and potential improvements still remain. In principle, approaches can be developed for on-the-fly fitting to the higher level potential during the simulation (rather than *a priori* parameterization). Another gain would be improving the robustness of approaches to calculate the perturbation between the two potentials. One example of a recent advance in this direction was the use of Gaussian-based correction potentials in order to improve the quality of the reference potential and the convergence of the calculations [119]. Another advance would be the incorporation of elements of the MTD sampling approach into the PD protocol [140]. Additionally, there have been many other recent breakthroughs in the field. For example, adaptive QM/MM approaches, in which the molecules are allowed to cross the boundaries between the QM and the MM regions to dynamically change their “character” (from QM to MM and vice versa) [141–143] have been increasingly developed in recent years. The simplest application of this approach assumes an abrupt change in the description of a molecule when it crosses a given cut-off, leading to sudden changes in the energy and forces involved. In order to avoid such artifacts, recent approaches have introduced a buffer zone between the QM and MM regions. Some include the permuted adaptive partitioning (PAP) [141,144] approach, the sorted adaptive partitioning (SAP) [141,144] approach, or the difference-based adaptive solvation (DAS) [145] approach. Clearly, this brief perspective cannot provide an overview of all QM/MM free energy approaches currently available. Overall, we believe that such approaches have broad applications in biomolecular simulations and the strategies outlined in this perspective will help overcome the quantitative limitations to existing semi-empirical QM/MM calculations without massive increases in computational cost.

## Acknowledgements

The European Research Council has provided financial support under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no. 306474. SCLK acknowledges support from the Swedish Research Council (Vetenskapsrådet, grant 2010–5026).

## References

- [1] S. Goedecker, Linear scaling electronic structure methods, *Rev. Mod. Phys.* 71 (1999) 1085–1123.
- [2] R. Schulz, B. Lindner, L. Petridis, J.C. Smith, Scaling of multimillion-atom biological molecular dynamics simulation on a petascale supercomputer, *J. Chem. Theory Comput.* 5 (2009) 2798–2808.
- [3] A.M.P. Sena, T. Miyazaki, D.R. Bowler, Linear scaling constrained density functional theory in CONQUEST, *J. Chem. Theory Comput.* 7 (2011) 884–889.
- [4] R.O. Dror, R.M. Dirks, J.P. Grossman, H.F. Xu, D.E. Shaw, Biomolecular simulation: a computational microscope for molecular biology, *Annu. Rev. Biophys.* 41 (2012) 429–452.
- [5] M.W. van der Kamp, A.J. Mulholland, Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology, *Biochemistry* 52 (2013) 2708–2728.
- [6] T.J. Lane, D. Shukla, K.A. Beauchamp, V.S. Pande, To milliseconds and beyond: challenges in the simulation of protein folding, *Curr. Opin. Struct. Biol.* 23 (2013) 58–65.
- [7] R. Susukita, T. Ebisuzaki, B.G. Elmgreen, H. Furusawa, K. Kato, A. Kawai, Y. Kobayashi, T. Koishi, G.D. McNiven, T. Narumi, K. Yasuoka, Hardware accelerator for molecular dynamics: MDGRAPE-2, *Comput. Phys. Commun.* 155 (2003) 115–131.
- [8] D.E. Shaw, M.M. Deneroff, R.O. Dror, J.S. Kuskin, R.H. Larson, J.K. Salmon, C. Young, B. Batson, K.J. Bowers, J.C. Chao, M.P. Eastwood, J. Gagliardo, J.P. Grossman, C.R. Ho, D.J. Ierardi, I. Kolossvary, J.L. Klepeis, T. Layman, C. McLeavey, M.A. Moraes, R. Mueller, E.C. Priest, Y.B. Shan, J. Spengler, M. Theobald, B. Towles, S.C. Wang, Anton, a special-purpose machine for molecular dynamics simulation, *Commun. ACM* 51 (2008) 91–97.
- [9] A.W. Gotz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born, *J. Chem. Theory Comput.* 8 (2012) 1542–1555.
- [10] M. Zink, H. Grubmüller, Mechanical properties of the icosahedral shell of southern bean mosaic virus: a molecular dynamics study, *Biophys. J.* 96 (2009) 1350–1363.
- [11] K.Y. Sanbonmatsu, C.S. Tung, High performance computing in biology: multimillion atom simulations of nanoscale systems, *J. Struct. Biol.* 157 (2007) 470–480.
- [12] V.A. Voelz, G.R. Bowman, K. Beauchamp, V.S. Pande, Molecular simulation of *ab initio* protein folding for a millisecond folder NTL9(1–39), *J. Am. Chem. Soc.* 132 (2010) 1526–1528.
- [13] J. VandeVondele, U. Borstnik, J. Hutter, Linear scaling self-consistent field calculations with millions of atoms in the condensed phase, *J. Chem. Theory Comput.* 8 (2012) 3565–3573.
- [14] D.R. Bowler, T. Miyazaki, Calculations for millions of atoms with density functional theory: linear scaling shows its potential, *J. Phys. Condens. Matter* 22 (2010) 074207.
- [15] L. Hung, E.A. Carter, Accurate simulations of metals at the mesoscale: explicit treatment of 1 million atoms with quantum mechanics, *Chem. Phys. Lett.* 475 (2009) 163–170.
- [16] P.L. Freddolino, F. Liu, M. Gruebele, K. Schulten, Ten-microsecond molecular dynamics simulation of a fast-folding WW domain, *Biophys. J.* 94 (2008) L75–L77.
- [17] D.L. Ensign, P.M. Kasson, V.S. Pande, Heterogeneity even at the speed limit of folding: large-scale molecular dynamics study of a fast-folding variant of the villin head-piece, *J. Mol. Biol.* 374 (2007) 806–816.

- [18] D.E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y.B. Shan, W. Wriggers, Atomic-level characterization of the structural dynamics of proteins, *Science* 330 (2010) 341–346.
- [19] J. Heimdal, U. Ryde, Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations, *Phys. Chem. Chem. Phys.* 14 (2012) 12592–12604.
- [20] H.M. Senn, W. Thiel, QM/MM methods for biomolecular systems, *Angew. Chem. Int. Ed.* 48 (2009) 1198–1229.
- [21] J.L. Gao, S.H. Ma, D.T. Major, K. Nam, J.Z. Pu, D.G. Truhlar, Mechanisms and free energies of enzymatic reactions, *Chem. Rev.* 106 (2006) 3188–3209.
- [22] M. Levitt, A. Warshel, Computer-simulation of protein folding, *Nature* 253 (1975) 694–698.
- [23] C. Hyeon, D. Thirumalai, Capturing the essence of folding and functions of biomolecules using coarse-grained models, *Nat. Commun.* 2 (2011) 487.
- [24] D. Thirumalai, Z.X. Liu, E.P. O'Brien, G. Reddy, Protein folding: from theory to practice, *Curr. Opin. Struct. Biol.* 23 (2013) 22–29.
- [25] A.Y. Shih, A. Arkhipov, P.L. Freddolino, K. Schulten, Coarse grained protein-lipid model with application to lipoprotein particles, *J. Phys. Chem. B* 110 (2006) 3674–3684.
- [26] J. Shillcock, Vesicles and vesicle fusion: coarse-grained simulations, in: L. Monticelli, E. Salonen (Eds.), *Biomolecular Simulations*, vol. 924, Humana Press, 2013, pp. 659–697.
- [27] R.P. Muller, A. Warshel, *Ab initio* calculations of free energy barriers for chemical reactions in solution, *J. Phys. Chem. 99* (1995) 17516–17524.
- [28] N.V. Plotnikov, S.C.L. Kamerlin, A. Warshel, Parodynamics: an effective and reliable model for *ab initio* QM/MM free-energy calculations and related tasks, *J. Phys. Chem. B* 115 (2011) 7950–7962.
- [29] A. Crespo, M.A. Martí, D.A. Estrin, A.E. Roitberg, Multiple-steering QM–MM calculation of the free energy profile in chorismate mutase, *J. Am. Chem. Soc.* 127 (2005) 6940–6941.
- [30] T.H. Rod, U. Ryde, Accurate QM/MM free energy calculations of enzyme reactions: methylation by catechol O-methyltransferase, *J. Chem. Theory Comput.* 1 (2005) 1240–1251.
- [31] H. Hu, Z. Lu, J.M. Parks, S.K. Burger, W. Yang, Quantum mechanics/molecular mechanics minimum free-energy path for accurate reaction energetics in solution and enzymes: sequential sampling and optimization on the potential of mean force surface, *J. Chem. Phys.* 128 (2008) 34105–34123.
- [32] M. Roca, B. Messer, D. Hilvert, A. Warshel, On the relationship between folding and chemical landscapes in enzyme catalysis, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 13877–13882.
- [33] I. Kim, A. Warshel, Coarse-grained simulations of the gating current in the voltage-activated Kv1.2 channel, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 2128–2133.
- [34] M. Roca, H. Liu, B. Messer, A. Warshel, On the relationship between thermal stability and catalytic power of enzymes, *Biochemistry* 46 (2007) 15076–15088.
- [35] S.C.L. Kamerlin, S. Vicatos, A. Dryga, A. Warshel, Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems, *Annu. Rev. Phys. Chem.* 62 (2011) 41–64.
- [36] G.M. Torrie, J.P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling, *J. Comput. Phys.* 23 (1977) 187–199.
- [37] J.G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* 3 (1935) 300–313.
- [38] U.H.E. Hansmann, Parallel tempering algorithm for conformational studies of biological molecules, *Chem. Phys. Lett.* 281 (1997) 140–150.
- [39] E. Darve, A. Pohorille, Calculating free energies using average force, *J. Chem. Phys.* 115 (2001) 9169–9183.
- [40] E. Weinan, E. Vanden-Eijnden, Transition-path theory and path-finding algorithms for the study of rare events, *Annu. Rev. Phys. Chem.* 61 (2010) 391–420.
- [41] D. Hamelberg, J. Mongan, J.A. McCammon, Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules, *J. Chem. Phys.* 120 (2004) 11919–11929.
- [42] A. Laio, M. Parrinello, Escaping free-energy minima, *Proc. Natl. Acad. Sci. U. S. A.* 99 (2002) 12562–12566.
- [43] D.M. Zuckerman, Equilibrium sampling in biomolecular simulations, *Annu. Rev. Biophys.* 40 (2011) 41–62.
- [44] O. Acevedo, W.L. Jorgensen, Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions, *Acc. Chem. Res.* 43 (2009) 142–151.
- [45] H. Hu, W. Yang, Free energies of chemical reactions in solution and in enzymes with *ab initio* quantum mechanics/molecular mechanics methods, *Annu. Rev. Phys. Chem.* 59 (2008) 573–601.
- [46] R. Lonsdale, K.E. Ranaghan, A.J. Mulholland, Computational enzymology, *Chem. Commun.* 46 (2010) 2354–2372.
- [47] S.C.L. Kamerlin, A. Warshel, Multiscale modeling of biological functions, *Phys. Chem. Chem. Phys.* 13 (2011) 10401–10411.
- [48] V. Luzhkov, A. Warshel, Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies, *J. Comput. Chem.* 13 (1992) 199–213.
- [49] M. Levitt, A simplified representation of protein conformations for rapid simulation of protein folding, *J. Mol. Biol.* 104 (1976) 59–107.
- [50] M. Strajbl, G. Hong, A. Warshel, *Ab initio* QM/MM simulation with proper sampling: “first principle” calculations of the free energy of the autodissociation of water in aqueous solution, *J. Phys. Chem. B* 106 (2002) 13333–13343.
- [51] E. Rosta, M. Klähn, A. Warshel, Towards accurate *ab initio* QM/MM calculations of free-energy profiles of enzymatic reactions, *J. Phys. Chem. B* 110 (2006) 2934–2941.
- [52] A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, Wiley, New York, 1991.
- [53] R.W. Zwanzig, High-temperature equation of state by a perturbation method. 2. Polar gases, *J. Chem. Phys.* 23 (1955) 1915–1922.
- [54] G.E. Scuseria, Linear scaling density functional calculations with Gaussian orbitals, *J. Phys. Chem. A* 103 (1999) 4782–4790.
- [55] B. Kaduk, T. Kowalczyk, T. Van Voorhis, Constrained density functional theory, *Chem. Rev.* 112 (2011) 321–370.
- [56] R.Z. Khaliullin, J. VandeVondele, J. Hutter, Efficient linear-scaling density functional theory for molecular systems, *J. Chem. Theory Comput.* 9 (2013) 4421–4427.
- [57] J. Rezac, B. Levy, I. Demachy, A. de la Lande, Robust and efficient constrained DFT molecular dynamics approach for biochemical modeling, *J. Chem. Theory Comput.* 8 (2012) 418–427.
- [58] L.W. Wang, M.P. Teter, Kinetic-energy functional of the electron-density, *Phys. Rev. B* 45 (1992) 13196–13220.
- [59] Y.A. Wang, N. Govind, E.A. Carter, Orbital-free kinetic-energy density functionals with a density-dependent kernel, *Phys. Rev. B* 60 (1999) 16350–16358.
- [60] J.D. Goodpaster, T.A. Barnes, T.F. Miller, Embedded density functional theory for covalently bonded and strongly interacting subsystems, *J. Chem. Phys.* 134 (2011) 164108–164117.
- [61] T.A. Wesolowski, A. Warshel, Frozen density-functional approach for *ab-initio* calculations of solvated molecules, *J. Phys. Chem.* 97 (1993) 8050–8053.
- [62] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, G. Seifert, Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties, *Phys. Rev. B* 58 (1998) 7260–7268.
- [63] M. Elstner, The SCC-DFTB method and its application to biological systems, *Theor. Chem. Acc.* 116 (2006) 316–325.
- [64] M. Gaus, Q. Cui, M. Elstner, DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB), *J. Chem. Theory Comput.* 7 (2011) 931–948.
- [65] D.R. Bowler, T. Miyazaki, O(N) methods in electronic structure calculations, *Rep. Prog. Phys.* 75 (2012) 036503.
- [66] I.S. Ufimtsev, T.J. Martinez, Graphical processing units for quantum chemistry, *Comput. Sci. Eng.* 10 (2008) 26–34.
- [67] M.A. Nitsche, M. Ferreria, E.E. Mocskos, M.C.G. Lebrero, GPU accelerated implementation of density functional theory for hybrid QM/MM simulations, *J. Chem. Theory Comput.* 10 (2014) 959–967.
- [68] K. Yasuda, Accelerating density functional calculations with graphics processing unit, *J. Chem. Theory Comput.* 4 (2008) 1230–1236.
- [69] X. Andrade, A. Aspuru-Guzik, Real-space density functional theory on graphical processing units: computational approach and comparison to gaussian basis set methods, *J. Chem. Theory Comput.* 9 (2013) 4360–4373.
- [70] I.S. Ufimtsev, T.J. Martinez, Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics, *J. Chem. Theory Comput.* 5 (2009) 2619–2628.
- [71] A. Warshel, M. Levitt, Levitt Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *J. Mol. Biol.* 103 (1976) 227–249.
- [72] A. Warshel, A. Bromberg, Oxidation of 4a,4b-dihydrophenanthrenes. 3. A theoretical study of large kinetic isotope effect of deuterium in the initiation step of the thermal reaction with oxygen, *J. Chem. Phys.* 52 (1970) 1262–1269.
- [73] A. Warshel, M. Karplus, Calculation of ground and excited-state potential surfaces of conjugated molecules. 1. Formulation and parametrization, *J. Am. Chem. Soc.* 94 (1972) 5612–5625.
- [74] The Nobel Prize in Chemistry 2013 – press release, Nobelprize.org, Nobel Media AB, 2013.
- [75] K. Nam, Q. Cui, J. Gao, D.M. York, Specific reaction parametrization of the AM1/d Hamiltonian for phosphoryl transfer reactions: H, O, and P atoms, *J. Chem. Theory Comput.* 3 (2007) 486–504.
- [76] W. Weber, W. Thiel, Orthogonalization corrections for semiempirical methods, *Theor. Chem. Acc.* 103 (2000) 495–506.
- [77] T. Tuttle, W. Thiel, OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application, *Phys. Chem. Chem. Phys.* 10 (2008) 2159–2166.
- [78] W. Wu, P. Su, S. Shaik, P.C. Hiberty, Classical valence bond approach by modern methods, *Chem. Rev.* 111 (2011) 7557–7593.
- [79] Y.R. Mo, J.L. Gao, *Ab initio* QM/MM simulations with a molecular orbital-valence bond (MOVb) method: application to an S(N)2 reaction in water, *J. Comput. Chem.* 21 (2000) 1458–1469.
- [80] A. Shurki, H.A. Crown, Hybrid *ab initio* VB/MM method – a valence bond ride through classical landscapes, *J. Phys. Chem. B* 109 (2005) 23638–23644.
- [81] M. Klähn, S. Braun-Sand, E. Rosta, A. Warshel, On possible pitfalls in *ab initio* quantum mechanics/molecular mechanics minimization approaches for studies of enzymatic reactions, *J. Phys. Chem. B* 109 (2005) 15645–15650.
- [82] R. Iftimie, D. Salahub, D. Wei, J. Schofield, Using a classical potential as an efficient importance function for sampling from an *ab initio* potential, *J. Chem. Phys.* 113 (2000) 4852–4862.
- [83] R. Iftimie, J. Schofield, Separation of quantum and classical behavior in proton transfer reactions: implications from studies of secondary kinetic isotope effects, *Int. J. Quantum Chem.* 91 (2003) 404–413.
- [84] P. Bandopadhyay, Accelerating quantum mechanical/molecular mechanical sampling using pure molecular mechanical potential as an importance function: the case of effective fragment potential, *J. Chem. Phys.* 122 (2005) 91102–91106.
- [85] I. Polyak, T. Benighaus, E. Boulanger, W. Thiel, Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation, *J. Chem. Phys.* 139 (2013) 64105–64116.



- [86] R.H. Wood, E.M. Yezdimer, S. Sakane, J.A. Barriocanal, D.J. Doren, Free energies of solvation with quantum mechanical interaction energies from classical mechanical simulations, *J. Chem. Phys.* 110 (1999) 1329–1337.
- [87] C.J. Woods, F.R. Manby, A.J. Mulholland, An efficient method for the calculation of quantum mechanics/molecular mechanics free energies, *J. Chem. Phys.* 128 (2008) 14109–14117.
- [88] T.H. Rod, U. Ryde, Quantum mechanical free energy barrier for an enzymatic reaction, *Phys. Rev. Lett.* 94 (2005) 138302.
- [89] Y.K. Zhang, H.Y. Liu, W.T. Yang, Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined *ab initio* QM/MM potential energy surface, *J. Chem. Phys.* 112 (2000) 3483–3492.
- [90] J. Heimdal, P. Rydberg, U. Ryde, Protonation of the proximal histidine ligand in heme peroxidases, *J. Phys. Chem. B* 112 (2008) 2501–2510.
- [91] C. Greco, M. Bruschi, J. Heimdal, P. Fantucci, L. De Gioia, U. Ryde, Structural insights into the active-ready form of [FeFe]-hydrogenase and mechanistic details of its inhibition by carbon monoxide, *Inorg. Chem.* 46 (2007) 7256–7258.
- [92] R. Iftimie, J. Schofield, Efficient *ab initio* sampling methods in rate constant calculations for proton-transfer reactions, *J. Chem. Phys.* 114 (2001) 6763–6773.
- [93] Y. Zhang, J. Kua, J.A. McCammon, Role of the catalytic triad and oxyanion hole in acetylcholinesterase catalysis: an *ab initio* QM/MM study, *J. Am. Chem. Soc.* 124 (2002) 10572–10577.
- [94] L. Ridder, J.N. Harvey, I.M.C.M. Rietjens, J. Vervoort, A.J. Mulholland, *Ab initio* QM/MM modeling of the hydroxylation step in p-hydroxybenzoate hydroxylase, *J. Phys. Chem. B* 107 (2003) 2118–2126.
- [95] J.C. Hermann, L. Ridder, H.D. Holtje, A.J. Mulholland, Molecular mechanisms of antibiotic resistance: QM/MM modelling of deacylation in a class A beta-lactamase, *Org. Biomol. Chem.* 4 (2006) 206–210.
- [96] M. Valiev, J. Yang, J.A. Adams, S.S. Taylor, J.H. Weare, Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations, *J. Phys. Chem. B* 111 (2007) 13455–13464.
- [97] H. Yin, D. Wang, M. Valiev, Hybrid quantum mechanical/molecular mechanics study of the  $S_N2$  reaction of  $\text{CH}_3\text{Cl} + \text{OH}^-$  in water, *J. Phys. Chem. A* 115 (2011) 12047–12052.
- [98] R. Lonsdale, S. Hoyle, D.T. Grey, L. Ridder, A.J. Mulholland, Determinants of reactivity and selectivity in soluble epoxide hydrolase from quantum mechanics/molecular mechanics modeling, *Biochemistry* 51 (2012) 1774–1786.
- [99] E. Rosta, M. Haranczyk, Z.T. Chu, A. Warshel, Accelerating QM/MM free energy calculations: representing the surroundings by an updated mean charge distribution, *J. Phys. Chem. B* 112 (2008) 5680–5692.
- [100] J. Florián, A. Warshel, Langevin dipoles model for *ab initio* calculations of chemical processes in solution: parametrization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous solution, *J. Phys. Chem. B* 101 (1997) 5583–5595.
- [101] O. Tapia, O. Goscinski, Self-consistent reaction field theory of solvent effects, *Mol. Phys.* 29 (1975) 1653–1661.
- [102] I.F. Galván, M.L. Sánchez, M.E. Martín, F.J. Olivares del Valle, M.A. Aguilar, Geometry optimization of molecules in solution: joint use of the mean field approximation and the free-energy gradient method, *J. Chem. Phys.* 118 (2003) 255–263.
- [103] M.L. Sánchez, M.E. Martín, I. Fdez. Galván, F.J. Olivares del Valle, M.A. Aguilar, Theoretical calculation of the stark component of the solute–solvent interaction energy. Validity of the mean field approximation in the study of liquids and solutions, *J. Phys. Chem. B* 106 (2002) 4813–4817.
- [104] H. Nakano, T. Yamamoto, Variational calculation of quantum mechanical/molecular mechanical free energy with electronic polarization of solvent, *J. Chem. Phys.* 136 (2012) 134107–134124.
- [105] F.C. Cui, H. Li, Mean field QM/MM method: average position approximation, *J. Chem. Phys.* 138 (2013) 174114–174121.
- [106] H. Hu, W.T. Yang, Elucidating solvent contributions to solution reactions with *ab initio* QM/MM methods, *J. Phys. Chem. B* 114 (2010) 2755–2759.
- [107] M. Haranczyk, M. Gutowski, A. Warshel, Solvation free energies of molecules. The most stable anionic tautomers of uracil, *Phys. Chem. Chem. Phys.* 10 (2008) 4442–4448.
- [108] H. Nakano, T. Yamamoto, Accurate and efficient treatment of continuous solute charge density in the mean-field QM/MM free energy calculation, *J. Chem. Theory Comput.* 9 (2012) 188–203.
- [109] H. Hu, Z. Lu, W. Yang, QM/MM minimum free energy path: methodology and application to triosephosphate isomerase, *J. Chem. Theory Comput.* 3 (2007) 390–406.
- [110] T. Kosugi, S. Hayashi, Crucial role of protein flexibility in formation of a stable reaction transition state in an  $\alpha$ -amylase catalysis, *J. Am. Chem. Soc.* 134 (2012) 7045–7055.
- [111] J. Cao, R. Björnsson, M. Buhl, W. Thiel, T. van Mourik, Modelling zwitterions in solution: 3-fluoro- $\gamma$ -aminobutyric acid (3F-GABA), *Chemistry* 18 (2012) 184–195.
- [112] G. König, P.S. Hudson, S. Boresch, H.L. Woodcock, Multiscale free energy simulations: an efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-Boltzmann Bennett reweighting schemes, *J. Chem. Theory Comput.* 10 (2014) 1406–1419.
- [113] S.J. Fox, C. Pittcock, C.S. Tautermann, T. Fox, C. Christ, N.O.J. Malcolm, J.W. Essex, C.K. Skylaris, Free energies of binding from large-scale first-principles quantum mechanical calculations: application to ligand hydration energies, *J. Phys. Chem. B* 117 (2013) 9478–9485.
- [114] H. Hu, W. Yang, Development and application of *ab initio* QM/MM methods for mechanistic simulation of reactions in solution and in enzymes, *THEOCHEM* 898 (2009) 17–30.
- [115] Y. Zhang, T.-S. Lee, W. Yang, A pseudobond approach to combining quantum mechanical and molecular mechanical methods, *J. Chem. Phys.* 110 (1999) 46–54.
- [116] D. Das, K.P. Eurenium, E.M. Billings, P. Sherwood, D.C. Chatfield, M. Hodošček, B.R. Brooks, Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method, *J. Chem. Phys.* 117 (2002) 10534–10547.
- [117] M. Kaukonen, P. Söderhjelm, J. Heimdal, U. Ryde, Proton transfer at metal sites in proteins studied by quantum mechanical free-energy perturbations, *J. Chem. Theory Comput.* 4 (2008) 985–1001.
- [118] S.C.L. Kamerlin, M. Haranczyk, A. Warshel, Progress in *ab initio* QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies, *J. Phys. Chem. B* 113 (2009) 1253–1272.
- [119] N.V. Plotnikov, A. Warshel, Exploring, refining, and validating the paradynamics QM/MM sampling, *J. Phys. Chem. B* 116 (2012) 10342–10356.
- [120] S.C.L. Kamerlin, A. Warshel, The empirical valence bond model: theory and applications, *WIREs Comput. Mol. Sci.* 1 (2011) 30–45.
- [121] L. Mones, W.J. Tang, J. Florián, Empirical valence bond simulations of the chemical mechanism of ATP to cAMP conversion by anthrax edema factor, *Biochemistry* 52 (2013) 2672–2682.
- [122] T. Huber, A. Torda, W. Gunsteren, Local elevation: a method for improving the searching properties of molecular dynamics simulation, *J. Comput. Aided Mol. Des.* 8 (1994) 695–708.
- [123] H.S. Hansen, H.P. Hünenberger, Using the local elevation method to construct optimized umbrella sampling potentials: Calculation of the relative free energies and interconversion barriers of glucopyranose ring conformers in water, *J. Comput. Chem.* 31 (2010) 1–23.
- [124] A.J. Cohen, P. Mori-Sanchez, W.T. Yang, Insights into current limitations of density functional theory, *Science* 321 (2008) 792–794.
- [125] A.J. Cohen, P. Mori-Sanchez, W.T. Yang, Challenges for density functional theory, *Chem. Rev.* 112 (2012) 289–320.
- [126] J. Florián, B.G. Johnson, Comparison and scaling of Hartree-Fock and density-functional harmonic force-fields. 1. Formamide monomer, *J. Phys. Chem.* 98 (1994) 3681–3687.
- [127] F.R. Beierlein, J. Michel, J.W. Essex, A simple QM/MM approach for capturing polarization effects in protein–ligand binding free energy calculations, *J. Phys. Chem. B* 115 (2011) 4911–4926.
- [128] J. Chandrasekhar, S.F. Smith, W.L. Jorgensen, Theoretical-examination of the  $S_N2$  reaction involving chloride-ION and methyl-chloride in the GAS-phase and aqueous-solution, *J. Am. Chem. Soc.* 107 (1985) 154–163.
- [129] S. Re, K. Morokuma, ONIOM study of chemical reactions in microsolvation clusters:  $(\text{H}_2\text{O})_n\text{CH}_3\text{Cl} + \text{OH}^- (\text{H}_2\text{O}) (n + m = 1 \text{ and } 2)$ , *J. Phys. Chem. A* 105 (2001) 7185–7197.
- [130] A.V. Pislakov, J. Cao, S.C.L. Kamerlin, A. Warshel, Enzyme millisecond conformational dynamics do not catalyze the chemical step, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 17359–17364.
- [131] R.A. Bachorz, W. Klopper, M. Gutowski, X. Li, K.H. Bowen, Photoelectron spectrum of valence anions of uracil and first-principles calculations of excess electron binding energies, *J. Chem. Phys.* 129 (2008) 54309–54319.
- [132] P. Dedikova, P. Neogady, M. Urban, Electron affinities of small uracil–water complexes: a comparison of benchmark CCSD(T) calculations with DFT, *J. Phys. Chem. A* 115 (2011) 2350–2358.
- [133] A. Warshel, F. Sussman, G. King, Free-energy of charges in solvated proteins – microscopic calculations using a reversible charging process, *Biochemistry* 25 (1986) 8368–8372.
- [134] A. Soriano, E. Silla, I. Tunon, M.F. Ruiz-Lopez, Dynamic and electrostatic effects in enzymatic processes. An analysis of the nucleophilic substitution reaction in haloalkane dehalogenase, *J. Am. Chem. Soc.* 127 (2005) 1946–1957.
- [135] M.H.M. Olsson, A. Warshel, Solute solvent dynamics and energetics in enzyme catalysis: the  $S_N2$  reaction of dehalogenase as a general benchmark, *J. Am. Chem. Soc.* 126 (2004) 15167–15179.
- [136] L.S. Devi-Kesavan, J.L. Gao, Combined QM/MM study of the mechanism and kinetic isotope effect of the nucleophilic substitution reaction in haloalkane dehalogenase, *J. Am. Chem. Soc.* 125 (2003) 1532–1540.
- [137] A. Rychkova, A. Warshel, Exploring the nature of the translocon-assisted protein insertion, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 495–500.
- [138] J. Czub, H. Grubmüller, Torsional elasticity and energetics of F1-ATPase, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 7408–7413.
- [139] S. Mukherjee, A. Warshel, Electrostatic origin of the mechanochemical rotary mechanism and the catalytic dwell of F1-ATPase, *Proc. Natl. Acad. Sci. U. S. A.* 108 (2011) 20550–20555.
- [140] N.V. Plotnikov, Computing the free energy barriers for less by sampling with a coarse reference potential while retaining accuracy of the target fine model, *J. Chem. Theory Comput.* (2014) Article ASAP. <http://dx.doi.org/10.1021/ct500109m>.
- [141] A. Heyden, H. Lin, D.G. Truhlar, Adaptive partitioning in combined quantum mechanical and molecular mechanical calculations of potential energy functions for multiscale simulations, *J. Phys. Chem. B* 111 (2007) 2231–2241.
- [142] K. Park, A.W. Gotz, R.C. Walker, F. Paesani, Application of adaptive QM/MM methods to molecular dynamics simulations of aqueous systems, *J. Chem. Theory Comput.* 8 (2012) 2868–2877.
- [143] C. Varnai, N. Bernstein, L. Mones, G. Csanyi, Tests of an adaptive QM/MM calculation on free energy profiles of chemical reactions in solution, *J. Phys. Chem. B* 117 (2013) 12202–12211.
- [144] S. Pezeschi, H. Lin, Adaptive-partitioning redistributed charge and dipole schemes for QM/MM dynamics simulations: on-the-fly relocation of boundaries that pass through covalent bonds, *J. Chem. Theory Comput.* 7 (2011) 3625–3634.
- [145] R.E. Buló, B. Ensing, J. Sikkema, L. Visscher, Toward a practical method for adaptive QM/MM simulations, *J. Chem. Theory Comput.* 5 (2009) 2212–2221.