Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation

Iakov Polyak, Tobias Benighaus, Eliot Boulanger, and Walter Thiel

Citation: The Journal of Chemical Physics **139**, 064105 (2013);
View online: https://doi.org/10.1063/1.4817402
View Table of Contents: http://aip.scitation.org/toc/jcp/139/6
Published by the American Institute of Physics

Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation

Iakov Polyak, Tobias Benighaus,[a] Eliot Boulanger, and Walter Thiel[b]
Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D–45470 Mülheim an der Ruhr, Germany

(Received 6 May 2013; accepted 18 July 2013; published online 9 August 2013)

The dual Hamiltonian free energy perturbation (DH-FEP) method is designed for accurate and efficient evaluation of the free energy profile of chemical reactions in quantum mechanical/molecular mechanical (QM/MM) calculations. In contrast to existing QM/MM FEP variants, the QM region is not kept frozen during sampling, but all degrees of freedom except for the reaction coordinate are sampled. In the DH-FEP scheme, the sampling is done by semiempirical QM/MM molecular dynamics (MD), while the perturbation energy differences are evaluated from high-level QM/MM single-point calculations at regular intervals, skipping a pre-defined number of MD sampling steps. After validating our method using an analytic model potential with an exactly known solution, we report a QM/MM DH-FEP study of the enzymatic reaction catalyzed by chorismate mutase. We suggest guidelines for QM/MM DH-FEP calculations and default values for the required computational parameters. In the case of chorismate mutase, we apply the DH-FEP approach in combination with a single one-dimensional reaction coordinate and with a two-dimensional collective coordinate (two individual distances), with superior results for the latter choice. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4817402]

I. INTRODUCTION

Free energy is a key thermodynamic quantity to characterize chemical processes. It governs the relative stability of different species and the rate of chemical reactions. Knowledge of the potential energy of the system along the reaction coordinate (RC) is not sufficient to determine the reaction rate because of the entropic contributions to the free energy. In systems that obey classical statistical mechanics, one needs information about all accessible configurations of the system through the partition function to calculate the free energy exactly. The Helmholtz free energy is given by

$$A = -\frac{1}{\beta}\ln(Z), \tag{1}$$

where Z is the canonical ensemble partition function of the system and $\beta = \frac{1}{k_B T}$ is available from the Boltzmann constant k_B and the temperature T. Free energy differences can be expressed in terms of ensemble averages that can be approximately evaluated with the use of sampling techniques, such as molecular dynamics (MD) or Monte Carlo (MC) simulations.[1]

There are several well-established procedures to calculate the free energy, e.g., umbrella sampling,[2] thermodynamic integration,[3] and free energy perturbation (FEP).[4] For example, FEP can be used to determine the free energy difference between a perturbed and an unperturbed state of the system, which are described by two different Hamiltonians, through the sampling of the potential energy difference between them.

Regardless of the chosen procedure, the configurational phase space needs to be sampled extensively to obtain accurate free energies. This will become computationally demanding when going to ever larger systems and to ever more accurate and time-consuming methods for computing the potential energy during the sampling. Nowadays, classical force fields are widely used to describe thermodynamic properties of large biomolecular systems. If electronic effects are important, e.g., as in chemical reactions, one can apply hybrid quantum mechanical/molecular mechanical (QM/MM) methods,[5] in which the electronically relevant part of the system is treated quantum-mechanically, while the remainder is described by a classical force field. QM calculations require significantly more computational time than MM calculations, and therefore extensive sampling of large systems is demanding at the QM/MM level, especially when using first-principles QM methods. Due to this limitation, there have been many efforts[6–26] to develop QM/MM free energy methods, which aim at avoiding direct sampling at high levels of theory while still giving an accurate estimate of the free energy changes during the reaction.

A powerful approach, initially proposed and developed by Warshel and co-workers,[6–12] and also employed in a modified form by Ryde and co-workers,[13,14,27] makes use of thermodynamical cycles; an initial estimate of the free energy is determined by sampling with some approximate reference Hamiltonian and then corrected by evaluating via FEP the free energy change when going from the approximate reference Hamiltonian to the target QM/MM Hamiltonian. In some of these studies,[12,27] the reference potential has been generated using semiempirical QM methods. Another approach[22–24] is to accelerate the sampling of configurational phase space by using auxiliary MC simulations performed with an

[a]Permanent address: Lanxess Deutschland GmbH, 51369 Leverkusen, Germany.
[b]Electronic mail: thiel@mpi-muelheim.mpg.de

approximate Hamiltonian; the resulting final MC structures are subjected to MC update tests, which are based on the phase space overlap of the two Hamiltonians, thus significantly increasing the rate of the overall convergence. The two approaches have also been combined.[25]

There are also QM/MM free energy calculations that conduct a direct sampling of the whole phase space of the full QM/MM system on a single potential surface using umbrella sampling,[28] thermodynamic integration,[29] or umbrella integration.[30] Such calculations usually employ efficient semiempirical methods as QM component and trajectories of less than 100 ps (sufficient to obtain converged results in the investigated enzymatic systems according to standard statistical tests[29]). In a recent study,[27] the use of semiempirical QM/MM sampling for evaluating the entropic contributions was however considered questionable, because the phase space showed only weak overlap with the one derived from higher-level methods. In the dual-level approach of Tuñon and co-workers,[31,32] higher-level single-point calculations are employed to determine correction terms for the semiempirical QM/MM energy and gradient as a continuous function of a distinguished reaction coordinate, and free energy calculations are then done on the resulting surface using umbrella sampling.

The QM/MM-FE technique developed by Yang *et al.*[26] is based on the FEP method and targets an especially efficient QM/MM sampling. In this approach, the reaction path is divided into windows, and in each of them the geometry of the QM region is obtained by a restrained QM/MM optimization. This geometry is then kept fixed during the sampling which is performed only for the MM region, with the QM atoms being represented by partial charges (derived by an ESP fit of the electrostatic potential). The perturbations are defined by the exchange of the two subsequent geometries of the QM region. This procedure offers an inexpensive way to directly obtain the free energy profile of a reaction at the QM/MM level since the sampling of the MM region is effectively done at the MM level. The conceptual drawback of this approach is the lack of sampling in the QM region, and hence the entropic QM contribution can only be evaluated at the stationary points within the rigid-rotor harmonic-oscillator approximation of statistical thermodynamics.

A more general formulation of the QM/MM-FE approach proposed by Rod and Ryde[13,14] and named QTCP (quantum-mechanical thermodynamic-cycle perturbation) uses the FEP method both for evaluating the MM $\rightarrow$ MM perturbation along the reaction coordinate and for estimating the vertical MM $\rightarrow$ QM/MM free energy differences in a thermodynamical cycle.

In this paper, we present a modified version of the QM/MM-FE method, in which the phase space of the QM region is freely sampled, except for the RC which is the subject of the perturbation. The sampling of the QM region combines MD simulations at the efficient semiempirical QM/MM level with first-principles QM/MM energy evaluations (using *ab initio* or density functional QM methods). We therefore call this approach Dual Hamiltonian Free Energy Perturbation (DH-FEP). In Secs. II–III, we first describe the method and its implementation. Thereafter we validate it for two test systems: a two-dimensional analytic model potential and the enzymatic reaction catalyzed by chorismate mutase.

II. METHOD

A. QM/MM-FE

According to Zwanzig,[4] the free energy difference between a perturbed (2) and an unperturbed (1) state can be expressed as

$$\Delta A = A_2 - A_1 = -\frac{1}{\beta}\ln\int P_1(\mathbf{r})\exp\{-\beta[E_2(\mathbf{r}) - E_1(\mathbf{r})]\}\mathrm{d}\mathbf{r}, \tag{2}$$

where $E(\mathbf{r})$ is the potential energy and $P_1(\mathbf{r})$ is the probability of finding the unperturbed system in the configuration $\mathbf{r}$. For a QM/MM Hamiltonian, the energy is decomposed into three parts and therefore we have

$$\Delta A = -\frac{1}{\beta}\ln\int P_1(\mathbf{r}) \times \exp\{-\beta[\Delta E_{QM} + \Delta E_{QM-MM} + \Delta E_{MM}]\}\mathrm{d}\mathbf{r}. \tag{3}$$

In the QM/MM-FE method introduced by Yang and co-workers,[26] the perturbation is defined as the exchange of two neighboring QM structures that result from restrained optimizations of points along the reaction path. The underlying assumption is that the QM and MM degrees of freedom (DOFs) can be treated separately and that the sampling needs to be done only over the MM DOFs, whereas the contributions to the free energy arising from the fluctuations of the QM region around its "optimum reaction path" are assumed to be constant along the RC. The expression for the free energy difference between "windows" A and B along the RC is[26]

$$\Delta A(R_c) = \Delta E_{QM}(\mathbf{r}_{QM}^{min}) - \frac{1}{\beta}\ln\int P\left(R_c^A\right)\exp\{-\beta[E_{QM/MM}\left(\mathbf{r}_{QM}^{min}\left(R_c^B\right)\right) - E_{QM/MM}\left(\mathbf{r}_{QM}^{min}\left(R_c^A\right)\right)]\}\mathrm{d}\mathbf{r}_{MM}. \tag{4}$$

Corrections for zero-point vibrational energies and entropic contributions are only included at the stationary points using the rigid-rotor harmonic-oscillator approximation,

$$\Delta A_{QM} - \Delta E_{QM} = \Delta E_{QM}^{ZPE} + \Delta U_{QM}^{th} - T\Delta S_{QM}. \tag{5}$$

The QM/MM-FE method outlined above involves two major assumptions. The first one is conceptual, namely not to sample the QM region, which causes a truncation of the accessible configurational space and may thus lead to an underestimation of the entropic contributions. The second and less critical one arises from the implementation: the representation of the QM atoms by ESP charges to allow for an efficient sampling (technically at the MM level).

B. Dual Hamiltonian free energy perturbation

In our approach, we do not separate the QM and MM DOFs but define the perturbation in terms of a pre-determined RC ξ on the potential energy surface. The RC is split into

discrete windows, each having a specific ξ_i value assigned, so that ξ_i and ξ_{i+1} are two constraints defining two neighboring windows along the RC,

$$\Delta E_{pert}^{\xi_i \to \xi_{i+1}} = E(\mathbf{r}', \xi_{i+1}) - E(\mathbf{r}', \xi_i), \tag{6}$$

where $\mathbf{r}'$ represents any configuration that fulfills the constraint ξ_i. We thus have a constrained Hamiltonian and can write the free energy along the RC as

$$A(\xi_i) = -\frac{1}{\beta} \ln \int \exp\{-\beta E(\mathbf{r}', \xi_i)\} \mathrm{d}\mathbf{r}'. \tag{7}$$

In standard notation,[4] the free energy perturbation between two subsequent points is given by

SHAKE is used to compute E(r', \xi_{i+1}), see Page 5.

$$\Delta A^{\xi_i \to \xi_{i+1}} = -\frac{1}{\beta} \ln \int P_i(\mathbf{r}', \xi_i) \times \exp\{-\beta[E(\mathbf{r}', \xi_{i+1}) - E(\mathbf{r}', \xi_i)]\} \mathrm{d}\mathbf{r}'. \tag{8}$$

In practice, the integration is replaced by a discrete sum over MD steps. In the limit of complete sampling over all $\mathbf{r}'$ we obtain

$$\Delta A^{\xi_i \to \xi_{i+1}} = -\frac{1}{\beta} \ln \left[\frac{1}{N} \sum_{i=1}^{N} \exp\{ -\beta \Delta E_{pert}^{\xi_i \to \xi_{i+1}} \} \right]. \tag{9}$$

Applying this approach directly in combination with high-level QM methods would be expensive. Therefore we look for an approximation that will make our computations efficient. The integration step size in the MD simulation is usually chosen rather small to ensure a stable and accurate propagation of the system. Two consecutive points are thus rather close in geometry and ΔE_{pert} does not vary much, i.e., the step size is ideal for the MD run, but not for sampling ΔE_{pert} efficiently. Therefore we adopt a procedure, in which ΔE_{pert} is computed regularly only after skipping a pre-determined number of steps; this also decreases the correlation between subsequent configurations. The intermediate MD steps are disregarded during the computation of the free energy, which is thus determined from a limited number of configurations. This allows us to introduce the next approximation: ΔE_{pert} is evaluated with a computationally demanding high-level QM method at the selected steps (which is affordable because of the relatively small number of such calculations), while the sampling is performed at the semiempirical QM/MM level. Denoting the low-level and high-level Hamiltonian by *Ham1* and *Ham2*, Eq. (8) can then be reformulated accordingly,

Using wrong weight!!!

$$\Delta A^{\xi_i \to \xi_{i+1}} = -\frac{1}{\beta} \ln \int P_i^{Ham1}(\mathbf{r}', \xi_i) \times \exp\{-\beta[E^{Ham2}(\mathbf{r}', \xi_{i+1}) - E^{Ham2}(\mathbf{r}', \xi_i)]\} \mathrm{d}\mathbf{r}'. \tag{10}$$

Using a cumulant expansion,[33] the free energy difference can be expressed as a function of the central moments of the energy difference distribution,

$$\Delta A = \langle \Delta E \rangle - \frac{\beta}{2} \sigma^2 + O(\beta)^2. \tag{11}$$

We use this expansion to overcome the problem of possible random occurrences of low ΔE_{pert} values in the trajectory, which may adversely affect the direct exponential average. In practice, we neglect all higher-order terms (as in Ref. 33), and the free energy difference is calculated as a sum of the average value and the variance of the energy difference distribution.

In actual applications, the reaction path is obtained from a sequence of restrained optimizations for suitably defined "windows," each one with a given RC value ξ_i. A semiempirical QM/MM MD simulation is then performed for each window with the constrained RC value ξ_i. Every x number of steps, the RC is perturbed to ξ_{i+1} and ΔE_{pert} is evaluated using a high-level QM Hamiltonian; note that the system is always propagated at RC $= \xi_i$. The ΔE_{pert} value obtained is then tested for equilibration as described in Ref. 29 by ensuring that there is no trend in the coarse-grained average and variance, and by checking the distribution for normality and lack of correlation. If the test for trend reveals non-stationarity of ΔE_{pert} or its variance, some MD steps from the beginning of simulation (and rarely from the end) are dropped until the resulting data becomes stationary. If the above analysis results in less than 400 equilibrated data points, further sampling is performed for the given window. The free energy difference between RC values ξ_i and ξ_{i+1} and the related confidence interval are then calculated based on the cumulant expansion.[33] Finally, the free energy profile of the reaction is obtained by summing up all the free energy differences between adjacent windows.

So far our development has been in terms of a one-dimensional RC (e.g., an internal coordinate or a linear combination of internal coordinates) that gives rise to a single constraint ξ_i. However, our formalism, in particular Eq. (10), remains valid when using a more general collective coordinate, for example a collection of several (N) independent internal coordinates $\{d_j(i)\}$ that are individually and simultaneously constrained during the sampling. A typical case is a one-dimensional RC defined as a linear combination of two distances, where the corresponding collective coordinate is composed of these two distances ($N = 2$). The use of a collective coordinate may lead to improved results, when the individual constraints are chosen appropriately and reflect the most relevant changes during the reaction.

DH-FEP is related to the several existing methods[11,12,27,31,32] in the sense that it uses a reference potential in order to perform efficient sampling, while obtaining the free energy difference at a higher theory level. It differs from previously proposed dual-level free energy methods in that we do not evaluate and apply high-level perturbation corrections after the low-level sampling is finished,[11,12,27] nor do we perform a semiempirical QM/MM sampling with a pre-calculated first-principles correction function along the reaction path.[31,32] Instead, our goal is to approximate an accurate high-level QM/MM sampling by using efficient semiempirical QM/MM MD simulations and directly evaluating first-principles QM/MM perturbation energies at a relatively small number of selected MD steps. We thus avoid a perturbation treatment in the method space, based on the assumption that there is a sufficient overlap in the phase space of the low-level and high-level methods used. Both our approach and the methods based on a thermodynamic cycle may suffer from a possibly weak

overlap of the two underlying phase spaces. In Ref. 12 this problem is approached by refining the reference potential, while we try to tackle it by finding a suitable semiempirical method that will represent the high-level QM method phase space well and/or by using an appropriate collective reaction coordinate (see Sec. IV). DH-FEP thus shares some basic strategic ideas with the MM based importance function method of Iftimie *et al.*,[22,23] which uses a classical MM potential to guide a first-principles MC simulation, but the computational framework is of course entirely different in these two approaches.

The convergence of QM/MM free energy perturbations based on semiempirical QM/MM simulations has recently been studied by Heimdal and Ryde.[27] The main distinction from our approach is the use of a thermodynamic cycle to account for the differences between low-level and high-level QM/MM methods via FEP. Within this framework, the so-called QTCP-free calculations are conceptually similar to our approach in the sense that only the reaction coordinate is kept fixed (rather than the whole QM system); it is not specified, however, whether the atoms involved in the reaction coordinate are fixed to their initial Cartesian coordinates or whether a constraint is applied, which impedes direct comparisons. In the QTCP-free calculations,[27] the error bars for the perturbation along the reaction coordinate are rather small (as in our approach, see below), whereas those for the perturbation in the method space (which have no counterpart in our approach) are quite large.

III. COMPUTATIONAL DETAILS

The DH-FEP method was implemented in a developmental version of the Chemshell package.[34] Constraints were imposed using the SHAKE procedure,[35] which was extended to include the difference of two distances between four different atoms. When evaluating ΔE_{pert} during the MD simulations, the SHAKE procedure is applied twice for the four atoms involved in the reaction coordinate, first to satisfy the constraint on the unperturbed system (RC value ξ_i), and then to satisfy the constraint on the perturbed system (RC value ξ_{i+1}); thereafter the potential energy is computed for the two resulting structures. For the calculation of free energy differences and the statistical validation of sampled data, we used a supplementary program written by Kästner for our original QM/MM-FE implementation.[33]

In the QM/MM calculations, we employed the following codes: MNDO2005[36] for the semiempirical QM methods OM3 (orthogonalization model 3)[37,38] and SCC-DFTB (self-consistent-charge density functional tight binding),[39] TURBOMOLE 6.3[40] for the *ab initio* QM method RI-MP2 (resolution-of-identity Møller-Plesset second-order perturbation theory),[41,42] and DL-POLY[43] for the CHARMM22 force field.[44]

IV. ASSESSMENT

We assess our method using two examples. The first one involves an analytic potential function that allows an exact evaluation of the free energy and can thus be used to validate our ansatz for calculating free energy differences along a pre-defined RC. We use two potential functions that differ slightly from each other, one of which is used for sampling and the other one for evaluating the perturbation energy differences, in order to test the importance of configurational phase space overlap between the two potentials. The second example addresses the evaluation of the activation free energy in the enzymatic reaction catalyzed by chorismate mutase: here we examine the performance of our method for a chemically meaningful QM/MM system and compare the results to experimental data.

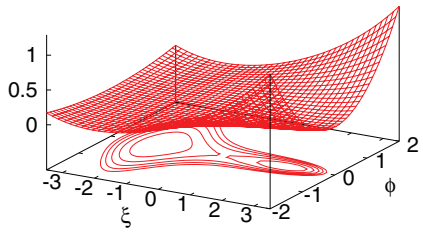

FIG. 1. 3D plot and contour plot of the analytic potential $E_1(\xi, \phi)$ with a contour spacing of 0.005. All values in atomic units.

A. Analytical model potential

For numerical validation of our method, we use a two-dimensional model potential taken from Ref. 30, for which the free energy can be computed analytically: $E_1(\xi, \phi) = f(\xi) + k(\xi)\phi^2$ with $f(\xi) = b - c\xi^2 + (c^2/4b)\xi^4$ and $k(\xi) = k_{min} + 2db/c + \sqrt{(8d^2b)/c}\xi + d\xi^2$. The 3D plot and a contour plot of the potential are shown in Fig. 1. The RC is represented by ξ while ϕ is an additional degree of freedom, along which the surface will be sampled to compute the free energy; on the RC, we always have $\phi = 0$. This model potential has two minima with $E_1 = 0$, which have different surroundings and thus differ in free energy (lower at the minimum with a broader potential because of higher entropic contributions). The free energy along the RC can be evaluated analytically as $A_1(\xi) = f(\xi) + \ln(k(\xi))/2\beta + \text{const}$.

We used the same parameters as in our previous work.[30] In atomic units, the barrier is chosen to be $b = 0.01$, the minima are placed at $\xi_{min} = \pm 2$ by assigning $c = 0.005$, while the width of E_1 in the direction of ϕ is defined by setting $d = 0.01$ and $k_{min} = 0.01$.

Constrained Metropolis MC simulations[45] were carried out on this model potential in the NVT ensemble at a temperature of 298.15 K. The path from $\xi = -3$ to $\xi = 3$ was split into windows separated by a width of 0.05. Fifty thousand MC trial steps with a maximum step size of 0.05 were performed for each window along the RC, with each new run starting at $\phi = 0$ and the RC being constrained to ξ_i. At each step, both ϕ and ξ were shifted in a random direction. If the step was accepted, the ξ value was replaced first with ξ_i and then with ξ_{i+1}, and the energies at both points were evaluated. Thereafter the next step was performed starting from ξ_i. The free energy difference was calculated from the direct exponential average of all sampled ΔE_{pert} values.

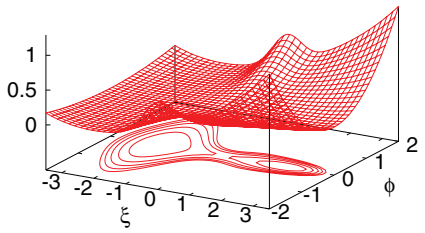

FIG. 2. 3D plot and contour plot of the analytic potential $E_{2a}(\xi, \phi)$ with a contour spacing of 0.005. All values in atomic units.

The resulting activation and reaction free energies are in excellent agreement with the analytic results. Compared with the analytic values of 28.250 and 3.512 kJ/mol for the activation and reaction free energies, the errors were as small as 0.043 and 0.049 kJ/mol, respectively, which clearly validates the FEP ansatz for calculating free energy differences along the reaction coordinate. With this justification in hand, we now test the approximation of using two different potentials for sampling and for evaluating ΔE_{pert} at the sampled geometries.

For this purpose, we constructed two new model potentials that differ in the transition state region but are the same at both minima. This choice is motivated by the intended QM/MM applications, where we expect low-level QM methods to mimic high-level QM methods more closely near the minima than near the transition states.

We first introduced into E_1 a term that depends in a Gaussian fashion on ξ and quadratically on ϕ, being zero at $\phi = 0$. In the resulting function $E_{2a}(\xi, \phi) = f(\xi) + k(\xi)\phi^2 + a \exp\{-\xi^2/(2s)\}\phi^2$, we chose the parameters as $a = 0.1$ and $s = 0.2$ (see Fig. 2).

Next we shifted the zero of the new term along the ϕ axis, thus slightly changing the minimum energy path in the region of the transition state. The new function was $E_{2b}(\xi, \phi) = f(\xi) + k(\xi)\phi^2 + a \exp\{-\xi^2/(2s)\}(\phi + \Delta)^2$ (see Fig. 3). We confirmed that MC calculations of the reaction and activation free energies for these modified potentials were as accurate as before (for the $E_1(\xi, \phi)$ potential, see above) when the energy differences were evaluated with the same potential that was used for sampling.

We then ran MC simulations with the same parameters as before, with the sampling done on the $E_{2a}(\xi, \phi)$ potential and the evaluation of ΔE_{pert} done on the $E_{2b}(\xi, \phi)$ potential. As expected, the results deteriorate with increasing values of the shift parameter Δ that governs the deviation from the $E_{2a}(\xi, \phi)$ sampling potential. In the sequence $\Delta = 0.05, 0.1,$ and 0.2, the error in the activation free energy rises from 0.45 kJ/mol via 2.33 kJ/mol to 9.74 kJ/mol. Due to the deliberate choice of the shape of the potentials (see above), the error in the reaction free energy grows much more slowly, from 0.00 kJ/mol via 0.17 kJ/mol to 0.92 kJ/mol, respectively.

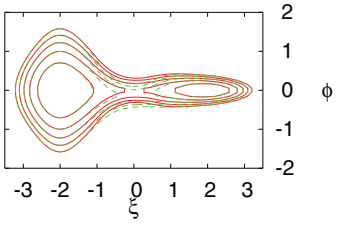

FIG. 3. Contour plot of the analytic potentials $E_{2a}(\xi, \phi)$ (solid lines) and $E_{2b}(\xi, \phi)$ (dashed lines) with a contour spacing of 0.005 and $\Delta = 0.2$. All values in atomic units.

The drastic rise of the error in the activation free energy confirms the importance of having sufficient overlap between the configurational phase space accessible on the two surfaces. At the transition state, the ϕ values that can be sampled on the $E_{2a}(\xi, \phi)$ potential range from -0.2 to 0.2 due to the steep rise of energy along the ϕ axis. Therefore, as soon as the transition state on the $E_{2b}(\xi, \phi)$ potential is moved close to the border of the ϕ values accessible at the $E_{2a}(\xi, \phi)$ level, we no longer sample the correct configurational space, and hence the computed activation free energy can no longer be trusted.

The DH-FEP method is thus clearly sensitive to the degree of the overlap of configurational phase space between the two potentials that are used for sampling and for evaluating ΔE_{pert} at the sampled geometries. Therefore the geometrical correspondence of the two potentials along the RC must be carefully checked prior to free energy calculations.

B. Chorismate Mutase

As second example we chose a "real-life" QM/MM system and calculated the activation free energy of the Claisen rearrangement of chorismate to prephenate, catalyzed by the *Bacillus subtilis* Chorismate Mutase (BsCM) enzyme. This reaction is a key step on the shikimate pathway of the aromatic amino acid synthesis in plants, fungi, and bacteria. It has been intensively investigated theoretically.[46] One peculiar trait of this system is the lack of covalent bonds between the substrate and the protein environment during the whole reaction, making it a rather convenient model for testing QM/MM methods. Experimentally, the entropic contribution to the activation free energy has been determined[47] to be $T\Delta S = -11.4 \pm 1.5$ kJ/mol at $T = 300$ K, which may serve as a reference value for assessing the results from QM/MM free energy calculations. In our present work on BsCM, we first focus on technical issues relevant to the proposed DH-FEP approach: we test the number of steps that may be skipped between two subsequent ΔE_{pert} evaluations, as well as the overall number of MD steps needed to obtain converged results, and we address the problem of configurational phase space overlap between the two potentials and how this affects the results.

In the QM/MM calculations, we treated the substrate (24 atoms) at the QM level (OM3, SCC-DFTB, RI-MP2/SVP) and the rest of the system comprising the protein and the solvent shell (13421 atoms in total) with the CHARMM22 force field.[44] The initial preparation of the system has been described elsewhere.[48] The first MD snapshot from the previous study[48] was subjected to further MD sampling using CHARMM33b1,[49,50] and six independent new snapshots were randomly chosen from this MD run.

Following standard conventions, we first defined the RC as the difference between the lengths of the breaking C–O and the forming C–C bond (see Fig. 4). Potential energy profiles

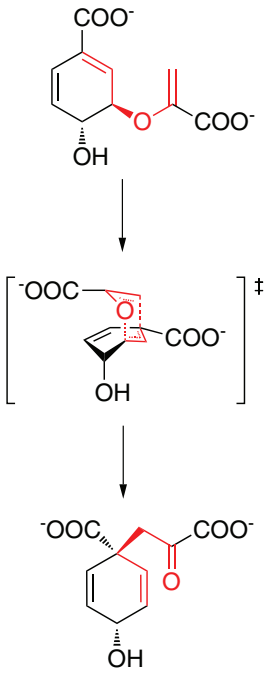

FIG. 4. Claisen rearrangement of chorismate to prephenate in chorismate mutase. The two parallel red dashed lines in the transition state indicate the forming and the breaking bonds. The difference between the corresponding distances is the reaction coordinate.

were calculated at all applied QM/MM levels (see above) for all the snapshots, via a series of restrained optimizations with the RC being sequentially changed from −2.4 Å to 2.4 Å in steps of 0.05 Å. For some of the snapshots, the reaction pathways were calculated several times in forward and backward direction until any unevenness was removed from the potential energy profile. Subsequent transition state optimizations and intrinsic reaction coordinate computations confirmed that the chosen RC is perfectly adequate and a valid reference to perform the FEP calculations.

In DH-FEP applications, two parameters need to be set, namely the number of steps skipped between two subsequent perturbations (x) and the total number of ΔE_{pert} evaluations to be performed. We have tested these options in a single MD run for an arbitrarily chosen window (at RC = −1.15 Å). The system was heated up to 300 K in steps of 10 K during 3 ps and then equilibrated for 20 ps, before the sampling was performed for 25 ps with ΔE_{pert} evaluated at every step. In this and all further MD calculations, the step size was 1 fs. All MD simulations were run in the canonial ensemble using the Nosé-Hoover chain thermostat[51,52] with a chain length of 4 and a characteristic time for the first thermostat of 0.02 ps. We used OM3/CHARMM both for sampling and for evaluating ΔE_{pert}. Results for different values of x with the number of ΔE_{pert} evaluations fixed to 1000 are shown in Fig. 5. We depict both the direct exponential average of all ΔE_{pert} evaluations taken (dashed) and the values obtained via cumulant expansion from a reduced number of ΔE_{pert} values (solid) selected after applying the statistical test on the lack of trend. The error bars shown in Fig. 5 refer to the latter; they were evaluated according to Ref. 33. The two sets of data do not deviate significantly, reflecting the lack of trend in most

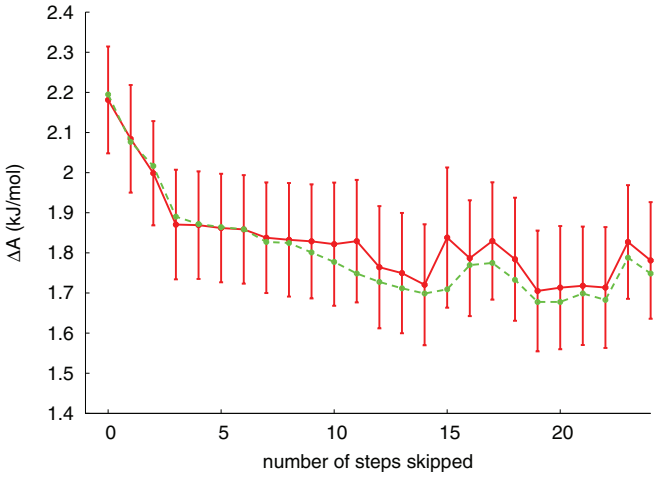

FIG. 5. Free energy difference between two windows calculated with a different number of steps skipped between the ΔE_{pert} evaluations, with the overall number of these evaluations fixed to 1000. The values in red (solid line) were obtained after subjecting the data to statistical tests for lack of trend and decorrelation. The values in green (dashed line) were obtained from direct exponential averaging of all data points. Data were taken starting from the end of a 25 ps OM3/CHARMM MD sampling run of one of the windows along the CM reaction profile (see text for further details).

datasets. ΔA converges with increasing x, showing that the decreasing dependency between subsequent ΔE_{pert} calculations improves the quality of the sampling. For x between 0 and 4, the free energy is clearly not converged, while values above 10 seem to be a reasonable choice. In this study, we adopted $x = 14$ (i.e., we evaluate ΔE_{pert} at every 15th step) since ΔA fluctuates around some average value for higher x. In an additional test, we have confirmed that this remains true up to $x = 149$, i.e., when extending the time between ΔE_{pert} evaluations up to 150 fs (see Fig. S1a of the supplementary material[53]).

Concerning the second option, Fig. 6 shows the variation of ΔA against the overall number of steps taken, with a fixed

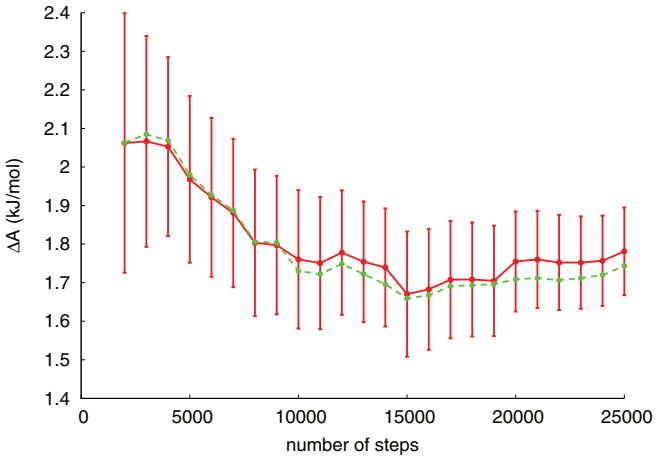

FIG. 6. Free energy difference between two windows calculated with 14 steps skipped between two ΔE_{pert} evaluations, with the overall number of these evaluations being varied. The values in red (solid line) were obtained after subjecting the data to statistical tests for lack of trend and decorrelation. The values in green (dashed line) were obtained from direct exponential averaging of all data points. Data were taken during a 25 ps OM3/CHARMM MD sampling run of one of the windows along the CM reaction profile (see text for further details).

value of $x = 14$. ΔA seems to converge after MD sampling times of around 10 ps. As expected, the error bar for ΔA decreases with increasing sampling time, i.e., with the number of ΔE_{pert} evaluations performed. This kind of convergence is confirmed by further test calculations with sampling times up to 105 ps (see Fig. S1b of the supplementary material[53]). In the following, we normally limit ourselves to 10 ps of sampling and use $x = 14$ throughout. We note in this context that more extensive sampling will often be hardly affordable in practice when the ΔE_{pert} evaluations are carried out with a high-level QM method.

Next we assess the accuracy of our method by comparing it to the well-established thermodynamic integration (TI) method. Both the sampling and ΔE_{pert} evaluations were performed at the OM3/CHARMM level. Since we focus on the activation free energy, we only considered the first 50 windows from the energy profile, covering the reactant minimum and transition state areas. The MD calculations were done for four snapshots in the following way: in every window, the system was first heated up to 300 K in steps of 10 K during 3 ps, then equilibrated for 25 ps, and finally sampled for 15 ps, with ΔE_{pert} being computed at every 15th step.

The results from the OM3/CHARMM FEP runs were in good agreement with those from TI calculations performed for the same snapshots with the same MD parameters: for all four snapshots tested, the activation free energies agreed to within 0.8 kJ/mol, which is of the same order as the error estimate[29] of 1.0 kJ/mol for the TI values with the currently adopted setup. The computed activation free energies $\Delta A^{\ddagger}$ for the four snapshots range between 66.5 and 71.5 kJ/mol, hence the snapshot-dependent fluctuations are significantly larger than the uncertainties in the TI and FEP calculations (both run on the same single potential surface). We have also tested the convergence of the OM3/CHARMM FEP results for one particular snapshot with regard to the MD sampling time in the FEP procedure: when prolonging the sampling time per window from 15 to 105 ps, the resulting free energy profiles remain virtually identical (see Fig. S2 of the supplementary material[53]), the activation free energies agree to within 0.2 kJ/mol, and the associated uncertainties decrease from 0.7 to 0.3 kJ/mol.

We now test the central DH-FEP approximation, namely the use of two different QM Hamiltonians in the QM/CHARMM calculations: OM3 or SCC-DFTB for sampling and MP2/SVP for evaluating ΔE_{pert}. As shown in Subsection IV A for the analytic model potential, reasonably accurate DH-FEP results can be expected only if the two QM methods yield reasonably similar geometries along the RC. To check this crucial DH-FEP issue, we define two criteria of geometrical correspondence: first, the interatomic distances entering the expression for the RC, and second, the root-mean-square deviation (RMSD) between the geometries of the whole QM region along the RC.

For a given value of RC defined as the difference of the distances in the forming C–O and the breaking C–C bond, restrained QM/MM optimizations (as well as constrained QM/MM dynamics) with two different QM methods will give different individual C–O and C–C distances, and therefore comparison of these distances can be a straightforward way to examine the geometrical correspondence of the two QM methods. Fig. 7 shows that the optimized distances from OM3/CHARMM nearly coincide with those from QM(MP2/SVP)/CHARMM up to RC = −1.4 Å, but start to deviate thereafter, with the difference growing up to 0.4 Å at RC = 0. The optimized C–O and C–C distances from SCC-DFTB/CHARMM show the opposite behavior: they differ from the QM(MP2/SVP)/CHARMM distances somewhat up to RC = −1.1 Å, but then follow them closely up to RC = 0 except for the region of RC = {−0.6 Å, −0.2 Å}. Concerning the RMSD values for the optimized QM regions relative to the QM(MP2/SVP)/CHARMM geometries along the RC: they vary from 0.04 to 0.06 Å for SCC-DFTB (being lowest in the region of RC = {−1.2 Å, 0.0 Å}) while they range from 0.06 to 0.09 Å for OM3 (being lowest for RC = {−1.15 Å, −0.7 Å}).

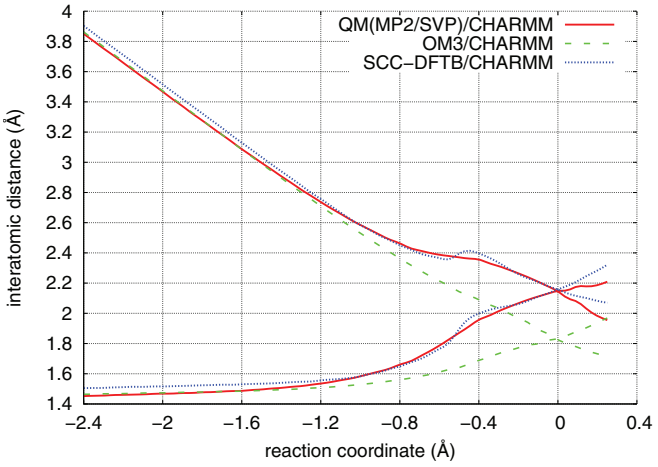

FIG. 7. Optimized C–C and C–O distances along the RC for the three different QM methods.

Going beyond geometry considerations, we performed a series of QM(MP2/SVP)/MM single-point energy calculations at the optimized OM3/MM and SCC-DFTB/MM geometries along the RC (see Fig. 8). None of the resulting two curves was exactly matching the QM(MP2/SVP)/MM

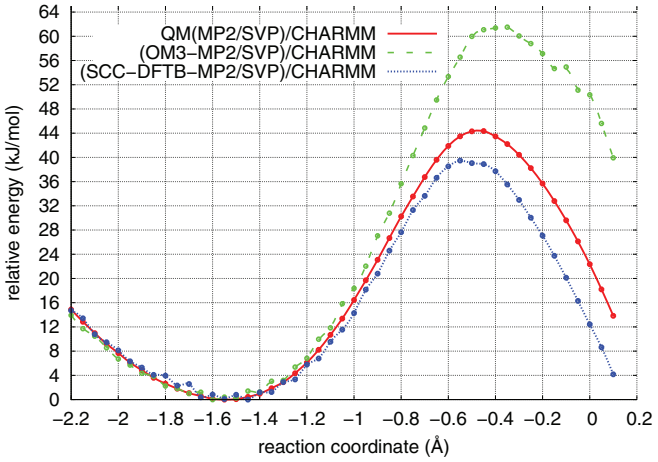

FIG. 8. Potential energy profile computed at the QM(MP2/SVP)/CHARMM level and QM(MP2/SVP)/CHARMM single-point energies at the optimized OM3 and SCC-DFTB structures along the reaction path.

The results look strange.
While QM/MM-FEP follows MEP very closely,
DH-FEP deviates very substantially.
Not sure if either method works properly.

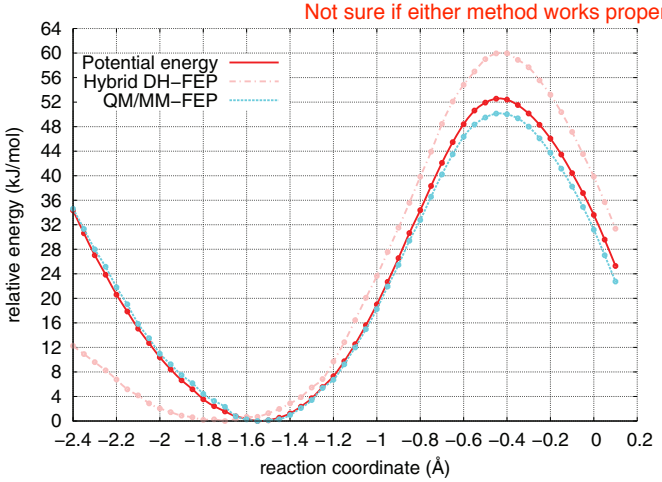

FIG. 9. Potential energy, DH-FEP, and QM/MM-FE profiles obtained for snapshot 6. The potential energy was computed at the QM(MP2/SVP)/CHARMM theory level. The DH-FEP profile was determined with a hybrid approach, in which the first part of the reaction path was sampled with OM3/CHARMM, and the second part with SCC-DFTB/CHARMM, while ΔE_{pert} was evaluated with QM(MP2/SVP)/CHARMM. The conventional QM/MM-FE profile was computed at the QM(MP2/SVP)/CHARMM level.

energy profile, but the relative energies computed at the SCC-DFTB/MM geometries were clearly closer to the QM(MP2/SVP)/MM reference values.

In an overall assessment of the QM/MM geometries for BsCM, SCC-DFTB thus seems superior to OM3 in reproducing the MP2-based results, and hence it should be a good choice for performing the sampling in QM/MM DH-FEP calculations. However, the corresponding QM(MP2/SVP//SCC-DFTB)/MM DH-FEP results (see Fig. S3 of the supplementary material[53]) were unsatisfactory: the DH-FEP free energy profile started rising much too fast at an early stage of the reaction close to the reactant state, and the activation free energy was too high compared with the QM(MP2/SVP)/MM reference value. Moreover, these calculations failed to reproduce the entropic contribution to the activation free energy that is known experimentally (see above). By contrast, the QM(MP2/SVP//OM3)/MM DH-FEP free energy profile was found to "behave" very well close to the reactant equilibrium, but to become quite different in shape from the QM(MP2/SVP)/MM reference curve closer to the TS, as expected from the geometry correspondence tests (see above).

Given the fact that neither OM3 nor SCC-DFTB provides sufficiently accurate QM/MM geometries along the whole RC, we decided to test a hybrid approach, running the MD sampling for the first part of the reaction (RC = {−2.4 Å, −1.25 Å}) with OM3/MM and using SCC-DFTB/MM for the second part (RC = {−1.25 Å, 0.1 Å}). To limit the computational effort for the MP2-based evaluation of ΔE_{pert}, the MD procedure was slightly changed: the heating was done in steps of 5 K during 6 ps, thereafter the system was equilibrated for 15 ps and sampled for 10 ps. We thus performed 1333 MP2/CHARMM calculations per window. The hybrid approach (dashed-dotted curve in Fig. 9) gave satisfactory results: the difference between $\Delta E^{\ddagger}$ and $\Delta A^{\ddagger}$ ranged from −2.0 to −18.0 kJ/mol for the individual snapshots, with an average value of −10.3 kJ/mol and a confidence interval for the barrier of about 1 kJ/mol. Taking into account the difference ΔE^{ZPE}_{QM} between the zero-point vibrational energies of TS and reactant (−4.2 kJ/mol for each snapshot in harmonic approximation) and assuming the thermal corrections ΔU^{th} to be negligible, we arrive at an average $T\Delta S^{\ddagger}$ value of −14.5 kJ/mol, which is close to the experimental result of −11.4 ± 1.5 kJ/mol.[47] It is obvious from Table I that the $\Delta A^{\ddagger}$ value fluctuates much less from snapshot to snapshot than the $\Delta E^{\ddagger}$ value, implying that the sampling was adequate. The fluctuations in the entropic contributions ($\Delta E^{\ddagger} - \Delta A^{\ddagger}$) thus mainly arise from the differences in the energy barriers for the individual snapshots.

The error estimates given in Table I account only for statistical fluctuations and incomplete sampling during the MD runs. They do not include errors caused by an insufficient overlap of the two underlying configurational spaces, as we do not apply an explicit reweighting of the semiempirical surface via FEP, as done, e.g., in Refs. 12 and 27. In the latter work, the errors associated with the perturbations along the reaction coordinate were fairly small (as in our case), while those associated with the perturbations in the method space (avoided in our approach) were rather large, thus raising general concerns about using semiempirical methods to provide the reference potential. We note that there was no attempt in Ref. 27 to evaluate the configurational space overlap between the chosen semiempirical and higher-level QM method prior to performing MD simulations, or to go beyond standard MNDO-type semiempirical methods. Doing so may enhance the quality of the reference potential in such dual-level free energy calculations.

For comparison, we also performed conventional QM/MM-FE calculations[26,29] for snapshot 6 (see Fig. 9). As expected from the lack of sampling in the QM region, the entropic contribution is underestimated: the free energy profile basically follows the potential energy profile, and the TS is even slightly lower, suggesting an entropic contribution with the wrong sign. Following the conventional procedure,[26,29] the entropic contribution for the QM region can be evaluated at the stationary points using the rigid-rotor

TABLE I. Free energy and potential energy barriers and entropic contributions to the barrier of the BsCM-catalyzed reaction for the six snapshots considered. All values in kJ/mol.

	Snapshots						Average	Exp.[47]
Snapshot number	1	2	3	4	5	6		
$\Delta A^{\ddagger}$	57.1±0.7	59.2±0.7	62.4±0.9	56.7±0.7	62.1±0.8	60.0±0.7	59.6±0.75	64.4
$\Delta E^{\ddagger}$	47.5	41.9	44.4	49.2	60.1	52.6	49.3	
$\Delta E^{\ddagger} - \Delta A^{\ddagger}$	− 9.6±0.7	− 17.3±0.7	− 18.0±0.9	− 7.5±0.7	− 2.0±0.8	− 7.4±0.7	− 10.3±0.75	− 11.4±1.5

harmonic-oscillator approximation; this gives a $T\Delta S^{\ddagger}_{QM}$ contribution of 2.5 kJ/mol, which is clearly too small to get close to the experimental value of the entropic contribution (see above). This example confirms that the degrees of freedom in the QM region should also be sampled to obtain a realistic entropic contribution to activation free energies in chemical reactions.

Our results with the hybrid approach indicate that the DH-FEP approach can provide free energies that closely mimic those from high-level QM/MM approaches, if the low-level QM/MM approach used for sampling yields realistic geometries along the RC (close to the high-level QM/MM geometries). However, such close matching of low-level and high-level geometries, e.g., from semiempirical and *ab initio* QM/MM calculations, may not always be achievable, as presently demonstrated for OM3 or SCC-DFTB versus MP2/SVP. In such cases, we can generalize the DH-FEP strategy by using more than one constraint, based on the observation that it is crucial to match the decisive geometrical variables entering the RC. In the case of BsCM, instead of only constraining the RC (i.e., the difference between the distances of the forming C–O and the breaking C–C bond), we now constrain the individual C–O and C–C distances to their reference values from QM(MP2/SVP)/MM restrained optimizations. This choice removes two DOFs of the QM region from sampling (rather than one DOF as before) and may thus entail the risk to underestimate the entropic contributions. This disadvantage is expected to be outweighed by the advantage of sampling a more appropriate configurational phase space, with better coverage of the region that is important in the high-level treatment.

We checked the performance of this collective coordinate approach by running DH-FEP calculations for snapshot 6, constraining both relevant C–O and C–C distances separately and using either OM3/CHARMM or SCC-DFTB/CHARMM for sampling throughout the whole reaction (see Fig. 10).

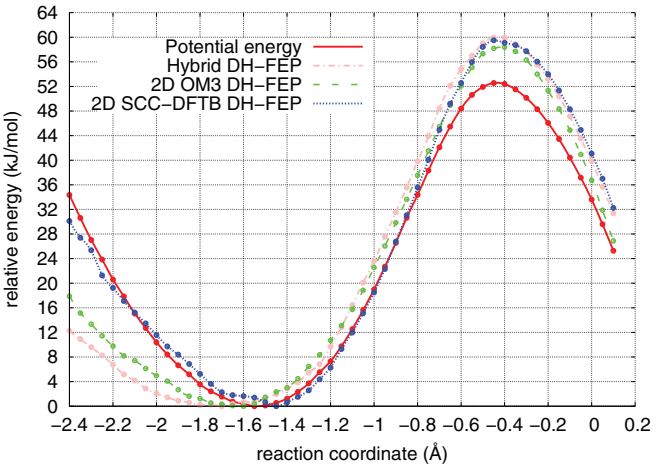

FIG. 10. Potential energy profile from QM(MP2/SVP)/CHARMM calculations and three free energy profiles computed for snapshot 6. Hybrid DH-FEP profile, sampling with OM3/CHARMM and SCC-DFTB/CHARMM for the first and second part of the reaction path, respectively (see text); 2D DH-FEP profiles, evaluated with the use of a two-dimensional collective coordinate, sampling with OM3/CHARMM and with SCC-DFTB/CHARMM. ΔE_{pert} obtained from QM(MP2/SVP)/CHARMM single-point calculations (see text).

Both DH-FEP calculations gave similar free energy profiles and reproduced the $\Delta A^{\ddagger}$ values that had previously been obtained with the hybrid DH-FEP approach. The use of a collective coordinate (here composed of the two relevant interatomic distances) in the DH-FEP calculations thus helps to overcome the limitations associated with the use of a single one-dimensional RC.[32]

The DH-FEP treatment may thus be improved by the judicious choice of an appropriate collective coordinate, thereby replacing the single constraint on the RC with two (or more) constraints on suitably chosen DOFs. This allows for successful applications even when there are appreciable differences between the low-level and high-level geometries along the reaction path. Obviously, a careful analysis of these differences is essential for identifying the DOFs that should enter the collective coordinate and be constrained in the DH-FEP calculations. Compared with the conventional QM/MM-FE procedure,[26,29] the DH-FEP approach, regardless of whether used with a single or a collective reaction coordinate, is expected to give a better estimate of the entropic contributions to the free energy profile, because of the explicit sampling of most of the QM region.

V. CONCLUSION

We have presented the DH-FEP method for evaluating free energies differences in large QM/MM systems. Compared with the conventional QM/MM-FE approach,[26,29] our method samples not only the MM region but also the QM region, i.e., the full configurational space except for the reaction coordinate. For the sake of computational efficiency, we introduced the approximation to use a less expensive low-level QM/MM method for sampling, while the perturbation energy differences ΔE_{pert} are evaluated through higher-level single-point QM/MM calculations performed at regular intervals, after skipping a pre-determined number of MD sampling steps. We examined the performance of our method using two test systems, namely a two-dimensional analytic model potential and a prototypical enzymatic reaction, the chorismate-to-prephenate conversion catalyzed by the BsCM enzyme.

Our implementation of the FEP approach was validated using the same potential for sampling and for evaluating ΔE_{pert} (i.e., a single Hamiltonian approach). The FEP results were shown to accurately reproduce the exact solutions for an analytic model potential and the activation free energy of the BsCM reaction obtained from standard thermodynamic integration.

In the numerical tests of the dual Hamiltonian approximation for the analytic model potential, the computed free energies were found to be quite sensitive to the overlap of the two surfaces in the region accessible to the sampling, thus calling for a careful analysis of the geometrical correspondence between the low-level and high-level methods chosen for DH-FEP calculations.

In the QM/MM tests for the enzymatic BsCM reaction, we first determined the necessary simulation parameters: we found that it was sufficient to evaluate ΔE_{pert} every 15 MD steps and to sample for at least 10 ps to obtain

results that are converged well enough. The subsequent DH-FEP QM/MM calculations employed the semiempirical OM3 and SCC-DFTB QM methods for sampling and the *ab initio* MP2/SVP approach for evaluating ΔE_{pert}. In the basic DH-FEP treatment, we constrained only the RC (defined as the difference between the distances of the forming C–O and the breaking C–C bond). The quality of the DH-FEP results was found to depend on the similarity between the low-level and high-level QM/MM structures along the RC: neither OM3 nor SCC-DFTB provided a good match to the MP2-based geometries over the entire RC, while being reasonably accurate in complementary regions of the reaction path. More realistic DH-FEP results could be obtained by a hybrid approach, in which the reaction path was divided into two regions, each described with the most suitable semiempirical method: the computed entropic contribution to the activation free energy was close to the experimental value.

Closer analysis of these DH-FEP QM/MM results for BsCM revealed that the crucial indicator of success is not the RMSD between the low-level and high-level QM/MM structures along the RC, but rather the match of the C–O and C–C distances used to define the RC (see above). Therefore, we applied the more general collective coordinate approach, with separate constraints on these two distances, to ensure an improved sampling of the relevant configurational space. The corresponding results were very close to the those from the hybrid approach, regardless of whether OM3 or SCC-DFTB was used for sampling. We thus recommend to use such a collective RC whenever the analysis of the low-level and high-level QM/MM structures along the RC reveals substantial discrepancies. A suitable collective RC can be defined by proceeding as follows. First, high-level QM/MM calculations are performed to locate the relevant transition state and the reaction path that connects it with the reactants and products. A natural choice for determining the reaction path is to follow the intrinsic reaction coordinate (IRC) starting from the optimized transition state, which can efficiently be done at the QM/MM level by an approximate microiterative scheme.[54] The IRC can then be used to identify the (small) set of internal coordinates, e.g., of individual interatomic distances, that undergo the most drastic changes along the reaction path and that should thus enter the collective RC for the subsequent DH-FEP calculations.

Going beyond this type of RC-based DH-FEP approach, one may attempt to devise procedures that directly control the space being sampled, for example by using MC techniques with update criteria based on the overlap between the two configurational spaces as suggested previously in a different context.[22] Alternatively, one may implement a DH-FEP scheme, in which the geometries and energy differences are stored during MD sampling, with the energy differences being weighted according to phase space overlap criteria at the end. Generally speaking, it is advisable to examine whether there is sufficient similarity of the geometries and sufficient overlap of the configurational phase spaces obtained with the low-level and high-level QM/MM methods used in the DH-FEP approach. If this is the case, DH-FEP offers an efficient opportunity to calculate accurate free energy differences in large QM/MM systems.

This approach can become even more valuable with the increase of computer power that will allow for future large-scale sampling at more expensive first-principles QM/MM levels, which may then enable even more accurate free energy evaluations, e.g., with larger basis sets or coupled cluster QM methods.

[1] D. Frenkel and B. Smit, *Understanding Molecular Simulation. From Algorithms to Applications*, 2nd ed. (Academic Press, 2002).
[2] G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974).
[3] J. G. Kirkwood, J. Chem. Phys. **3**, 300 (1935).
[4] R. W. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).
[5] H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[6] R. P. Muller and A. Warshel, J. Phys. Chem. **99**, 17516 (1995).
[7] J. Bentzien, R. P. Muller, J. Florin, and A. Warshel, J. Phys. Chem. B **102**, 2293 (1998).
[8] M. Štrajbl, G. Hong, and A. Warshel, J. Phys. Chem. B **106**, 13333 (2002).
[9] M. H. M. Olsson, G. Hong, and A. Warshel, J. Am. Chem. Soc. **125**, 5025 (2003).
[10] E. Rosta, M. Klähn, and A. Warshel, J. Phys. Chem. B **110**, 2934 (2006).
[11] N. V. Plotnikov, S. C. L. Kamerlin, and A. Warshel, J. Phys. Chem. B **115**, 7950 (2011).
[12] N. V. Plotnikov and A. Warshel, J. Phys. Chem. B **116**, 10342 (2012).
[13] T. H. Rod and U. Ryde, Phys. Rev. Lett. **94**, 138302 (2005).
[14] T. H. Rod and U. Ryde, J. Chem. Theory Comput. **1**, 1240 (2005).
[15] J. Chandrasekhar, S. F. Smith, and W. L. Jorgensen, J. Am. Chem. Soc. **106**, 3049 (1984).
[16] J. Chandrasekhar, S. F. Smith, and W. L. Jorgensen, J. Am. Chem. Soc. **107**, 154 (1985).
[17] W. L. Jorgensen, Acc. Chem. Res. **22**, 184 (1989).
[18] R. V. Stanton, M. Perkyl, D. Bakowies, and P. A. Kollman, J. Am. Chem. Soc. **120**, 3448 (1998).
[19] P. A. Kollman, B. Kuhn, O. Donini, M. Perakyla, R. Stanton, and D. Bakowies, Acc. Chem. Res. **34**, 72 (2001).
[20] B. Kuhn and P. A. Kollman, J. Am. Chem. Soc. **122**, 2586 (2000).
[21] O. Donini, T. Darden, and P. A. Kollman, J. Am. Chem. Soc. **122**, 12270 (2000).
[22] R. Iftimie, D. Salahub, D. Wei, and J. Schofield, J. Chem. Phys. **113**, 4852 (2000).
[23] R. Iftimie, D. Salahub, and J. Schofield, J. Chem. Phys. **119**, 11285 (2003).
[24] P. Bandyopadhyay, J. Chem. Phys. **122**, 091102 (2005).
[25] C. J. Woods, F. R. Manby, and A. J. Mulholland, J. Chem. Phys. **128**, 014109 (2008).
[26] Y. Zhang, H. Liu, and W. Yang, J. Chem. Phys. **112**, 3483 (2000).
[27] J. Heimdal and U. Ryde, Phys. Chem. Chem. Phys. **14**, 12592 (2012).
[28] E. Rosta, M. Nowotny, W. Yang, and G. Hummer, J. Am. Chem. Soc. **133**, 8934 (2011).
[29] H. M. Senn, S. Thiel, and W. Thiel, J. Chem. Theory Comput. **1**, 494 (2005).
[30] J. Kästner and W. Thiel, J. Chem. Phys. **123**, 144104 (2005).
[31] J. Ruiz-Pernia, E. Silla, I. Tunon, S. Marti, and V. Moliner, J. Phys. Chem. B **108**, 8427 (2004).
[32] J. J. Ruiz-Pernia, E. Silla, I. Tunon, and S. Marti, J. Phys. Chem. B **110**, 17663 (2006).
[33] J. Kästner, H. M. Senn, S. Thiel, N. Otte, and W. Thiel, J. Chem. Theory Comput. **2**, 452 (2006).
[34] P. Sherwood, A. de Vries, M. Guest, G. Schreckenbach, C. Catlow, S. French, A. Sokol, S. Bromley, W. Thiel, A. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjovoll, A. Fahmi, A. Schäfer, and C. Lennartz, J. Mol. Struct.: THEOCHEM **632**, 1 (2003).
[35] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[36] W. Thiel, *MNDO2005*, version 7.0; Max-Planck-Institut für Kohlenforschung: Mülheim, 2005.
[37] M. Scholten, Ph.D. thesis, Universität Düsseldorf, 2003.
[38] N. Otte, M. Scholten, and W. Thiel, J. Phys. Chem. A **111**, 5751 (2007).
[39] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).
[40] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, Chem. Phys. Lett. **162**, 165 (1989).

[41]F. Weigend and M. Häser, Theor. Chem. Acc. **97**, 331 (1997).
[42]F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).
[43]W. Smith and T. Forester, J. Mol. Graphics **14**, 136 (1996).
[44]A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, J. Phys. Chem. B **102**, 3586 (1998).
[45]N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[46]F. Claeyssens, K. E. Ranaghan, N. Lawan, S. J. Macrae, F. R. Manby, J. N. Harvey, and A. J. Mulholland, Org. Biomol. Chem. **9**, 1578 (2011).
[47]P. Kast, M. Asif-Ullah, and D. Hilvert, Tetrahedron Lett. **37**, 2691 (1996).
[48]H. M. Senn, J. Kästner, J. Breidung, and W. Thiel, Can. J. Chem. **87**, 1322 (2009).
[49]B. Brooks, R. Bruccoleri, D. Olafson, D. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[50]B. R. Brooks, C. L. Brooks III, A. D. MacKerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, J. Comput. Chem. **30**, 1545 (2009).
[51]S. Nose, J. Chem. Phys. **81**, 511 (1984).
[52]W. Hoover, Phys. Rev. A **31**, 1695 (1985).
[53]See supplementary material at http://dx.doi.org/10.1063/1.4817402 for additional evaluation of the effect of sampling time on free energy convergence and for OM3/CHARMM and SCC-DFTB/CHARMM DH-FEP profiles.
[54]I. Polyak, E. Boulanger, K. Sen, and W. Thiel, "A microiterative intrinsic reaction coordinate method for large QM/MM systems," Phys. Chem. Chem. Phys. (published online).