Published in final edited form as:

J Phys Chem B. 2011 June 23; 115(24): 7950–7962. doi:10.1021/jp201217b.

ParaDynamics: An Effective and Reliable Model for Ab Initio QM/MM Free Energy Calculations and Related Tasks

Nikolay Plotnikov¹, Shina Caroline Lynn Kamerlin^{2,3,4}, and Arie Warshel^{1,*}

¹ Department of Chemistry (SGM418), University of Southern California, 3620 McClintock Ave., Los Angeles CA-90089, U. S. A

² Department of Organic Chemistry and Department of Biochemistry and Biophysics, Arrhenius Laboratory, Stockholm University, S-10691, Stockholm, Sweden

³ Stockholm Center for Biomembrane Research (CBR)

⁴ Swedish e-Science Research Center (SeRC)

Abstract

Recent years have seen tremendous effort in the development of approaches with which to obtain quantum mechanics/molecular mechanics (QM/MM) free energies for reactions in the condensed phase. Nevertheless, there remain significant challenges to address, particularly the high computational cost involved in performing proper configurational sampling and in particular in obtaining *ab initio* QM/MM (QM(ai)/MM) free energy surfaces. One increasingly popular approach that seems to offer an ideal way to progress in this direction is the elegant metadynamics (MTD) approach. However, in the current work we point out the subtle efficiency problems associated with this approach, and illustrate that we have at hand what is arguably a more powerful approach. More specifically, we demonstrate the effectiveness of an updated version of our original idea of using a classical reference potential for QM(ai)/MM calculations [*J. Phys. Chem. B*. 102 (1998), 2293], which we refer to as “paradynamics” (PD). This approach is based on the use of an empirical valence bond (EVB) reference potential, which is already similar to the real *ab initio* potential. The reference potential is fitted to the *ab initio* potential by an iterative and, to a great degree, automated refinement procedure. The corresponding free energy profile is then constructed using the refined EVB potential, and the linear response approximation (LRA) is used to evaluate the QM(ai)/MM activation free energy barrier. The automated refinement of the EVB surface (and thus the reduction of the difference between the reference and *ab initio* potentials) is a key factor in accelerating the convergence of the LRA approach. We apply our PD approach to a test reaction, namely the S_N2 reaction between chloride ion and methyl chloride, and demonstrate that, at present, this approach is far more powerful and cost effective than the metadynamics approach (at least in its current implementation). We also discuss the general features of the PD approach in terms of its ability to explore complex systems, and clarify that it is not a specialized approach limited to only accelerating QM(ai)/MM calculations with proper sampling, but rather, can be used in a wide variety of applications. In fact, we point out that the use of a reference (CG) potential coupled with its PD refinement, as well as our renormalization approach, both provide very general and powerful strategies that can be used very effectively to explore any property that has been studied by the MTD approach.

I. Introduction

Obtaining reliable *ab initio* free energies for reactions in the condensed phase has long been one of the holy grails of modern computational chemistry (see e.g. Refs. ^{1–10}). As a result, significant research effort has been vested into this problem, resulting in a number of promising strategies^{2, 4, 5, 9–14} (many of which^{4, 5, 10, 12, 14, 15} exploit our idea¹⁶ of utilizing a classical (or just simplified) potential as a reference for *ab initio* hybrid quantum mechanical / molecular mechanical (QM(ai)/MM) calculations). However, despite progress in this area, there are still some significant challenges, in particular the high computational cost associated with the repeated calls to the *ab initio* (which are necessary in order to perform the extensive sampling which is required to obtain physically meaningful results). Unfortunately, our strategies have been very slow to be accepted by the wider scientific community.

An increasingly popular idea that has emerged in recent years, and seems to solve a part of these problems is the “metadynamics” (MTD) approach, which can be combined with QM(ai)/MM approaches in order to study processes that are dominated by changes in chemical structure, such as chemical reactions. The MTD approach is partially an adaptation of van Gunsteren and coworkers’^{17–19} important idea of locally elevating the minima of the potential (note that, in this work, “potential” refers to the potential energy surface) by filling them with Gaussians, which introduces the concept of “memory” to the simulation, preventing the system from returning to areas of the configurational space that have already been visited, and therefore improving sampling. Since its original inception in 2002²⁰, the MTD approach has skyrocketed in popularity (possibly, amongst other reasons, because of its seemingly automatic implementation), and has been applied to a wide range of problems, including photoinduced ring transformations²¹, intramolecular reactions in complex systems²², rational catalyst design²³ and organometallic reactivity²⁴, to name just a few examples (see also the review in Ref. ²⁵). The present work will attempt to clarify the actual cost of and problems with the MTD approach, and to compare and contrast it with a recent refinement of our reference potential approach. Our refined approach will be referred to in this work as “paradynamics” (PD) in order to both highlight its similarities to the MTD approach, as well (perhaps) as to clarify that it is, at present, more effective than the MTD approach in its current implementation.

II. Background

In order to prepare the reader with the proper perspective to be able to compare the MTD approach to the PD approach using an empirical valence bond (EVB) reference potential, we begin by providing a description of the main points of the MTD approach.

The MTD approach involves the clever idea of devising a potential that is much shallower than the original system potential, where the simulation can thus avoid becoming trapped in the many fine minima of the real potential. A similar idea was already introduced quite early to the field, i.e. that of devising a coarse-grained (CG) model for protein folding by Levitt and Warshel in 1975²⁶. However, the MTD approach moves further in this direction, by trying to generate a potential that will completely flatten the surface of the system. This is achieved by iteratively adding a potential, $E'(r)$, to the real potential, in the form of a sum of Gaussians that gradually approaches $-E(r)$ (where $E(r)$ is the original potential). In other words, at the k^{th} iteration, the MTD is run on the potential given by:

$$E_k(r) = E_{QM}(r) + E'_k(r) \quad (1)$$

where

$$E'_k(r) = - \sum_{i_k > i_{k-1}} A_i e^{-\alpha_i \cdot r^2} \quad (2)$$

However, despite its elegant derivation and implementation, this approach ends up with having major similarities to the reference potential approach used by Warshel and coworkers (see the discussion below). At any rate, the MTD approach places repulsive markers in a coarse time-line in the space that is spanned by a small number of chemically relevant collective variables (which are used to define the reaction coordinate). These markers are then placed on top of the underlying free energy landscape, in order to push the system to rapidly accumulate in the initial basin, by discouraging it from revisiting points in the configurational space that have already been sampled (note that this involves some key aspects of the basic principle of the local elevation method of van Gunsteren and coworkers mentioned above). In this way, the system can “escape” over the lowest transition state as soon as the growing biasing potential and the underlying free energy well counterbalance each other, effectively allowing the simulation to “escape” free energy minima²⁰. To better illustrate this principle, one can consider, as an example, the case described in Fig. 1A. In the MTD one divides the range of interest spanned by the selected RC (from a minimum to maximum value) into smaller intervals, and identifies the case where the dynamics starts to lead to the closest minima at V_0 , which are effectively sampled. At this point, a potential V_1' is built at the V_0 minimum (using Gaussians as the functions of the chosen RC), and the dynamics is propagated on $V_0 + V_1'$. This allows for the sampling of higher energy points on the FES, i.e. those that lie on the height of the minima at $V_0 + V_1'$. At the next iteration, the potential V_2' is constructed, and the region near the minima of $V_0 + V_2'$ is effectively sampled. This procedure is iteratively repeated until the maxima of the free energy surface are all sampled, at which point the sum of $V_0 + V_n'$ is basically flat. The resulting flat potential allows for convenient sampling of the entire V_0 .

It should be noted that although MTD has been formulated in an elegant way, the impression that this is a new powerful way in order to solve the sampling challenge is not entirely correct. That is, while it is clearly frequently easier to evaluate the FES using a flat reference potential, this may obscure the fact that the main computational cost arises from flattening the potential, as this involves evaluating the energy by expensive QM calculations a countless number of times. Since the key issue is the overall efficiency we may look at related formulations. Thus, we start by clarifying that the philosophy of this approach is actually almost identical (at least formally) to the overall philosophy of our earlier reference potential approach (see above). In fact, the construction of a potential that makes the landscape flat is similar to the use of a simplified folding model as a reference potential²⁷, and the general use of a reference potential for accelerated sampling has been a key part of our QM/MM free energy perturbation (QM/MM-FEP) studies for a long time⁸, long before the appearance of MTD in the field. The same approach has also been used by us in quantizing nuclear motions^{28, 29}, as well as in moving from CG to full potentials in studies of protein folding³⁰, as well as related problems^{3, 31–34}. The formal similarity between the two approaches becomes clearer once, while developing the reference potential, we also take into account the idea of improving this potential so as to minimize the difference (by some measure) between the reference and real potentials (see e.g. Refs. ³). This point can be best illustrated in Fig. 1, which compares and contrasts the MTD and PD approaches. From this figure, it can be seen that, in the case of MTD, as outlined above, this approach samples the energy landscape while incrementally adding Gaussian potentials to the real potential of the system. This is repeated many times until the *sum* of both potentials reaches zero (which

results in a flat potential, as shown in Fig. 1A). In contrast to this, in our PD approach, our initial reference potential (in this case an EVB potential) is already quite similar to the real potential of the system, and instead we iteratively and systematically refine the parameters of the reference potential, until the *difference* between the two potentials approaches zero (see Fig. 1B). A comparison of the computational cost of the two approaches will be found in Section IV.4.

In any case, the key issue here is that while it is currently very fashionable to use approaches where the best RC is not assumed *a priori*^{35–37}, using chemical knowledge can in many cases be superior to a blind search (even though many workers prefer black box approaches). Thus, whilst of course the iterative approach of the MTD, which places Gaussians in the minima to prevent re-sampling of the conformational space is a highly effective approach, this problem can also be trivially solved by roughly mapping the surface, placing Gaussians on the approximate minima, and refining this by use of any steepest descent minimization approach, at a fraction of the computational cost. It should be noted that this work focuses on QM/MM simulations as an illustrative example of the power of the PD approach. However, we will also discuss some other general directions where the PD approach (and, in particular, its renormalized version) offers a powerful way to address problems where it could be assumed that MTD would offer the best strategy.

III. The Paradynamics Approach

III.1. General Outline

The aim of PD is to provide an efficient and reliable way to perform QM(*ai*)/MM free energy calculations. The starting point for this is the use of the EVB approach (which has been discussed in detail elsewhere^{38, 39}) as a reference potential, which has proven to be highly successful in the past (e.g. Refs. ^{3, 16, 32, 40}). This method has several features, the first of which involves having a reference potential which is reasonably close to the real potential, allowing one to use a perturbation treatment, similar to the approach of Ref. ³. Secondly, the EVB energy gap offers what is probably the best RC for studies of chemical processes both in solution and in proteins, as was also recently observed in a study that used the energy gap in MTD calculations⁴¹ (see also Section IV.5). Now the use of a reference potential is formally identical to using a potential that is the inverse of the real potential, as is done by the MTD (see Fig. 1). This requires, however, that the two potentials (i.e. the EVB and the *ab initio* potentials) are as close as possible. As will be outlined below, this can be achieved in several steps by means of our PD approach. Here, we will start with a brief description of EVB free energy calculations, followed by a consideration of the procedure for moving from the EVB to the *ab initio* FES, and finally, we will outline the procedure for refining the EVB parameters.

III.2. The EVB Free Energy Surface

The EVB potential surface is obtained by constructing diabatic surfaces (ϵ_i) that approximate the reactant (RS), product (PS) and different intermediate states (the transition state (TS) in particular), and mixing these surfaces with off-diagonal terms that lead to the actual ground state surface, E_{EVB} (for details see e.g. Refs. ^{38, 39, 42}). The relevant activation free energy, $\Delta g_{\text{EVB}}^\ddagger$, is evaluated using a mapping potential, ϵ_m^{map} , which drives the system from one diabatic state to another as we accumulate sampling along the RC between the RS and PS. In the simplest case, which involves only two diabatic states, the mapping potential can be written as a linear product of the reactant (ϵ_1) and product (ϵ_2) potentials:

$$\epsilon_m^{\text{map}} = (1 - \lambda_m)\epsilon_1 + \lambda_m\epsilon_2 \quad (3)$$

where λ_m is incrementally changed from 0 to 1 ($(n+1)$ values) in n fixed increments $\Delta\lambda$. The free energy increment associated with moving between the two adjacent mapping potentials ε_m^{map} and ε_{m+1}^{map} is evaluated using the (FEP) formula⁴³:

$$\delta G_{map}(\lambda_m \rightarrow \lambda_{m+1}) = -kT \ln \left[\langle \exp(-(\varepsilon_{m+1}^{map} - \varepsilon_m^{map})/kT) \rangle_{\varepsilon_m^{map}} \right] \quad (4)$$

where $\langle \rangle$ designates an average over molecular dynamics (MD) trajectories propagated over ε_m^{map} . The final free energy for moving from ε_1 to ε_2 is taken as a sum of all free energy

$$\Delta G_{map}(\lambda_{n+1}) = \sum_{m=1}^n \delta G_{map}(\lambda_m \rightarrow \lambda_{m+1}) \quad (5)$$

However, Eq. 5 would only provide the free energy barrier, $\Delta g_{map}^\ddagger$, for moving from ε_1 to the mapping potential ε_{TS}^{map} , forcing the system to remain near the EVB TS. Therefore, in order to construct the full EVB FES, we use as our RC, x , the energy gap between the two diabatic states (i.e. $x = \varepsilon_1 - \varepsilon_2$), and calculate the free energy of the system at a particular point on the RC using:

$$\Delta G_{EVB}(x) \cong -kT \ln \left[\langle \delta(\Delta\varepsilon - x) \exp(-(E_{EVB} - \varepsilon_m^{map}(x))/kT) \rangle_{\varepsilon_m^{map}} \right] + \Delta G_{map}(\lambda_m) \quad (6)$$

(see e.g. Refs. ⁴⁴ and ⁴⁵ for further discussion).

III.3. Evaluating the Correction for Moving from the EVB to the *AB INITIO* Potential

Once the EVB surface has been constructed as outlined in Section III.2, we obtain, the activation free energy barrier, $\Delta g_{EVB}^\ddagger$, on the EVB FES, using:

$$\Delta g_{EVB}^\ddagger = \Delta G_{EVB}(x_{EVB}^{TS}) - \Delta G_{EVB}(x_{EVB}^{RS}) \quad (7)$$

Similarly, the activation free energy barrier on the QM FES, $\Delta g_{QM}^\ddagger$, is given by:

$$\Delta g_{QM}^\ddagger = \Delta G_{QM}(x_{QM}^{TS}) - \Delta G_{QM}(x_{QM}^{RS}) \quad (8)$$

In principle, one can proceed in a similar manner to that of Section III.2, and directly obtain the QM FES with a modified form of Eq. 6 where E_{EVB} is substituted by E_{QM} ^{8, 16}. However in this work we adopt the approach of Ref. ² and use the thermodynamic cycle depicted in Fig. 2, where the corresponding points on the EVB FES, $\Delta G_{EVB}(x_{EVB})$, are used as references for $\Delta G_{QM}(x_{QM})$. The terms on the right hand side of Eq. 8 can be calculated by means of the following relationship:

$$\Delta G_{QM}(x_{QM}^i) = \Delta G_{EVB}(x_{EVB}^i) + \Delta \Delta G_{EVB \rightarrow QM}(x_{EVB}^i \rightarrow x_{QM}^i) \quad (9)$$

Where the first term of Eq. 9 is obtained from EVB calculations and the second term is the free energy change associated with moving from the EVB to the QM FES at the relevant

point on x^i (i stands for RS, TS or PS). The resulting QM free energy barrier is therefore related to the EVB barrier by:

$$\Delta g_{QM}^\ddagger = \Delta g_{EVB}^\ddagger + \Delta \Delta g_{EVB \rightarrow QM}^\ddagger \quad (10)$$

where the terms on the right hand side of Eq. 10 are the difference between the corresponding right hand side terms of Eq. 9 at the TS and at the RS. The free energy of the transfer between EVB and QM potentials (i.e. the second term of Eq. 9) can then be directly obtained by means of FEP calculations using the mapping potential:

$$\mathcal{E}_m^{map} = (1 - \lambda_m)E_{EVB} + \lambda_mE_{QM} \quad (0 \leq \lambda_m \leq 1) \quad (11)$$

(see Ref. ²), and then using Eqs. 4 and 5. Alternately, convergence problems in the FEP can be overcome as was done in Ref. ³, by using the LRA formulation:

$$\Delta \Delta G_{EVB \rightarrow QM}(x_{EVB}^i \rightarrow x_{QM}^i) \cong \frac{1}{2} \left(\langle E_{QM} - E_{EVB} \rangle_{E_{QM}} + \langle E_{QM} - E_{EVB} \rangle_{E_{EVB}} \right) \quad (12)$$

This is particularly important in the event that the EVB and QM potential surfaces differ significantly from each other. However, since the point of the refinement procedure is to bring the two potentials as close as possible to each other, once the EVB parameter refinement is complete, the end-points approach of the LRA should be a good approximation to performing full FEP.

III.4. Performing Iterative PD Fitting to an *AB INITIO* Potential

After the above “introduction” we can move to the key point of this work which is the reduction of the difference between the EVB and the QM(*ai*)/MM surfaces. It should be made clear to the reader that EVB parameters have been fitted to experimental results and/or to reliable *ab initio* calculations for more than 2 decades (see e.g. Ref. ⁴⁴). This fitting can, in principle, be done manually to any desired QM potential (whether it is semi-empirical, *ab initio* or hybrid QM/MM, etc). In order to facilitate this process, we adopt the minimum least-square fitting algorithm, which, to a great extent, provides an automated way to refine the default EVB parameters so as to fit them to the QM potential of interest. One should keep in mind, however, that each QM method has its own intrinsic pros and cons, which sets limitations to its applicability, both in terms of reliability and/or feasibility. Consequently, each QM potential has its own characteristic potential and FES, which may differ significantly from each other, from the experimental results, and from the initial EVB potential. The refinement procedure ensures the proximity of the EVB reference potential to the QM potential of interest. This in turn ensures that the points on the EVB FES that are obtained by sampling the EVB potential during MD or Monte Carlo simulations to the first approximation already represent the QM FES. Thus, a major amount of computer time is saved by running MD trajectories on the EVB potential, while constructing the FES, since EVB calculations of the energy and its derivatives are faster than any electronic structure calculation.

The first step in our procedure involves automatically fitting the EVB/MM potential to the corresponding QM(*ai*)/MM potential. It is worth mentioning here that since the EVB method includes solvation effects in the energies of the diabatic states through the uncoupled electrostatic, VdW and induced dipoles terms, it can be sufficient to refine the parameters in the gas phase, and to use this parameter set in subsequent calculations (see e.g.

Refs. ⁴⁴ and ⁴⁶). This is achieved by iteratively refining the EVB parameters for the reaction in solution (as well as also the gas-phase reaction). Our treatment considers the difference, $\mathfrak{Z}$, between the EVB/MM and QM/MM energies, as well as the first derivatives at various configurations generated during MD runs on either the QM or EVB surfaces. $\mathfrak{Z}$ can be represented by following the idea of Ref. ¹:

$$\mathfrak{Z}(\vec{p}, r) = k_1 \sum_{i=1}^N \left(E_i^{\text{EVB}}(\vec{p}, r) - E_i^{\text{QM}}(r) \right)^2 + k_2 \sum_{i=1}^N \sum_{j=1}^3 \left(\frac{\partial}{\partial x_j} E_i^{\text{EVB}}(\vec{p}, r) - \frac{\partial}{\partial x_j} E_i^{\text{QM}}(r) \right)^2 \quad (13)$$

where E^{EVB} and E^{QM} denote EVB and QM energies respectively, $p \rightarrow$ is the vector of the EVB parameters, i runs over the configurations generated during the MD run and k_1 and k_2 are weighting factors which are applied to the energies and derivatives respectively. For each EVB parameter, p , we iteratively search for the value that minimizes $\mathfrak{Z}$ using an augmented steepest descent approach, with some Newton-Raphson character, using the diagonal second derivatives:

$$p_{k+1} = p_k - g \left[\partial \mathfrak{Z}(p_k) / \partial p \right] / \left[\partial^2 \mathfrak{Z}(p_k) / \partial p^2 \right] \quad (14)$$

where g is the steepest descent scaling factor, which is increased or decreased at every iteration, depending on whether the value of $\mathfrak{Z}$ increases or decreases. The corresponding derivatives are calculated numerically by means of Eq. 15 and 16:

$$\frac{\partial}{\partial p} \mathfrak{Z}(p) = \frac{\mathfrak{Z}(p+dp) - \mathfrak{Z}(p-dp)}{2dp} \quad (15)$$

$$\frac{\partial^2}{\partial p^2} \mathfrak{Z}(p) = \frac{[\mathfrak{Z}(p+dp) - \mathfrak{Z}(p)] - [\mathfrak{Z}(p) - \mathfrak{Z}(p-dp)]}{dp^2} \quad (16)$$

Note, however, that the relevant first and second derivatives could also in principle be evaluated analytically; however, the EVB calculations are sufficiently cost-efficient to permit the calculating of the derivatives using the numerical approach outlined in Eqs. 15 and 16. The iterative approach described above comprises the refinement procedures used to bring the EVB potential close to the QM potential (see also Fig. 1B), and effectively flattens the QM potential, in the sense that the high energy points of the QM potential are sampled on the underlying EVB potential.

IV. Results and Discussion

In order to best illustrate our PD approach, we will confine ourselves to a simple test case, namely the symmetric S_N2 reaction between chloride ion and methyl chloride ($\text{CH}_3\text{Cl} + \text{Cl}^-$, see Fig. 3). All calculations, unless otherwise specified, were performed using the MOLARIS software package⁴⁷. The interface between MOLARIS and the relevant QM program (in this case MOPAC2009⁴⁸) was externally provided through automated python scripts.

IV.1. Performing the QM Calculations

The test system was initially studied in the gas phase, using MOPAC2009⁴⁸. Geometries of the RS (the dipole complex) and the TS were optimized at the PM3 level of theory^{49, 50}. The enthalpic contribution to the activation free energy was found to be 9.7 kcal/mol. In addition, despite its approximated nature (see discussion in e.g.^{51–53}) the harmonic entropic contribution to the activation free energy was evaluated by means of the vibrational analysis in the harmonic approximation, with the rigid rotator approximation, and was estimated to be ~1.6 kcal/mol, giving a total $\Delta g_{QM}^\ddagger$ of 11.2 kcal/mol at 298K. Normal modes calculated using PM3 were used to constrain the MD runs near the TS and to prevent the system from sliding downhill along the vibrational mode corresponding to the single imaginary frequency. Particular care was taken in order to leave other vibrational modes, and, in particular, those of high intensities, intact even with the imposed constraints. Specifically, large distance constraints of 250 kcal mol⁻¹ Å⁻¹ were imposed on the C-Cl atom pairs to keep the relevant atoms at a distance of approximately 2.3Å.

IV.2. Performing an Initial EVB Study Using the Default Parameters

In parallel, we performed EVB free energy perturbation/umbrella sampling (FEP/US) calculations using the non-polarizable version of the ENZYME force field⁴⁷ and the default EVB parameters (see Tables S1 and S2), using the resonance structures shown in Fig. 4. That is, the initial EVB runs were performed using the standard EVB parameters, which are given in the Supporting Information (SI) (Tables S1 and S2, note that as the reaction under consideration is a symmetrical reaction, the relevant parameters for the RS and PS are identical). To improve sampling near the RS and PS, medium harmonic constraints (K=5 kcal mol⁻¹ Å⁻¹) were introduced in both states, to keep the respective atoms (2 and 6 in the RS, 1 and 2 in the PS) at a distance of ~3Å. The constraint energy was included in the corresponding diabatic energies. Additionally, an angle constraint between the Cl⁻-C-Cl atoms was used at both the RS and TS (see the SI for full parameter sets before and after refinement). The barrier height on the EVB free energy surface, $\Delta g_{EVB}^\ddagger$, was adjusted to 12 kcal/mol (in accordance with experiment⁵⁴), using an H₁₂ of 50 kcal/mol (for further details see the SI). The resulting EVB free energy surface is shown in Fig. 5.

In the next step, we evaluated the relevant terms to calculate $\Delta\Delta G_{EVB \rightarrow QM}$ using the LRA approach outlined in Eq. 12, in which we run MD trajectories on both the EVB ($\langle \rangle_{EVB}$) and PM3 ($\langle \rangle_{E_{QM}}$) potentials at the RS and TS (see discussion in Section III and Ref.⁵). Fig. 6 shows the fluctuations in $E - E$ during the MD run, with the results of the calculations being summarized in Table 1. As seen from this table, we obtained a correction of

$\Delta\Delta g_{EVB \rightarrow QM}^\ddagger = -7.0$ and the resulting $\Delta g_{QM}^\ddagger$ is about 5 kcal/mol. The relatively poor agreement with the estimated barrier of 11.2 kcal/mol reflects the large difference between the EVB and PM3 surfaces, and the corresponding difficulties in obtaining a reliable estimate of the free energy change for transferring between the two potentials using the LRA approach when the two potentials are very different.

IV.3. Performing the PD Refinement

As discussed in Section III, once we reduce the difference (or, in other words, $\langle (E_{QM} - E_{EVB})/RT \rangle$) between the EVB and QM potentials, it is possible to obtain a reliable QM(ai)/MM surface. Thus, we utilized the automated PD parameter refinement procedure outlined in Section III.4 in order to find a set of EVB parameters that minimize the energy difference between the EVB and PM3 surfaces. It is important to note that although the refinement protocol is formally automated, we strongly believe that it is important to apply human intuition and logic to the trends and problems that could arise during the refinement procedure. That is, for instance, we noticed in particular that the parameter fitting for the

diabatic states is simpler if the mixing term (H_{12}) is relatively small in the RS and PS (as, in this case, the energetics near the RS and PS have very little effect from H_{12} and are determined by the parameters for the corresponding diabatic states only, as shown in Fig. 7A). Running MD trajectories on the PM3 potential at the RS and PS generates the structures used during the refinement. Once a significant improvement in the EVB parameters has been achieved, the structures generated by running MD on the PM3 potential at the TS can be introduced into the refinement procedure (note that for the system under consideration here, the Cl atoms were frozen and a $500 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ positional constraint was applied to the carbon atom in the Z-direction along the Cl-Cl axis, thus constraining C-Cl bonds at $2.3 \text{ \AA}$). Here, the best form for the mixing term was found to be:

$$H_{12} = A \exp(-\mu(r_{12} - r_{26})^2) \quad (17)$$

where r_{12} is the distance between atoms 1 and 2 and r_{26} the distance between atoms 2 and 6 (see Fig. 4 and the SI for the value of the parameters). We refined the H_{12} parameters and added the Cl-C-Cl angle bending contribution (see Table S2 in the supporting information) to the energies of diabatic states. The resulting difference between the EVB and PM3 potentials using the refined EVB parameters is outlined in Fig. 7B. Once a set of refined EVB parameters had been obtained, we ran MD trajectories on the mapping potential given by Eq. 3, and used Eq. 6 to obtain the free energy profile shown in Fig. 8, which gave a $\Delta g_{\text{EVB}}^\ddagger = 12.12 \text{ kcal/mol}$. Calculating the corresponding second terms on the right side of Eq. 9 for RS and TS, using Eq. 12, we obtained the results presented in Table 1. The corresponding MD runs used for LRA correction are given in Fig. 9. The correction for $\Delta \Delta g_{\text{EVB} \rightarrow \text{QM}}^\ddagger = -0.38 \text{ kcal/mol}$ resulted in $\Delta g_{\text{QM}}^\ddagger = 11.74 \text{ kcal/mol}$, which gives much better agreement with the estimated value presented in Section IV.1.

IV.4. Efficiency Benchmarks

As one of the main points in our discussion is the efficiency of PD in comparison to the currently popular MTD approach, it is therefore important to provide a detailed timing comparison between the two. As was pointed out in the discussion above, the most time consuming part of the QM(ai)/MM calculations are the repeated calls to the QM. Thus, the overall simulation time is roughly proportional to the *total* number of QM calls necessary to obtain $\Delta g_{\text{QM}}^\ddagger$. The length of a single QM call depends on the method being used and on the specific type of calculation requested. That is, when evaluating the total simulation time for the QM free energy calculations, we need to take into account whether the calls to QM are a simple single point energy (SPE) calculation, the time required for which is denoted here as t_{SPE} , or a longer job which also includes an additional calculation of the energy derivatives (denoted “Force” for simplicity), as well as the calculation of the charges by fitting of the electrostatic potential. We also denote the time required for a “Force” job as t_{Force} . Here, we should keep in mind that MTD is run on the QM/MM potential, and therefore, in the MTD approach, every QM call is a “Force” job. In comparison, in PD, only half of the total QM calls (i.e. those for trajectories running on the QM surface) require more expensive “Force” calculations. Additionally, it is worth also noting that, in a gas-phase MTD study of the same system⁴¹, the total number of the “Force” calls required for a full QM(ai)/MM study was $2 \cdot 10^6$, whereas in our PD study, we required 6000 “Force” calls, and 6000 SPE calls (i.e. a total of 12000 calls to the QM). In order to estimate the total time required for calculating an *ab initio* QM/MM free energy barrier, we ran jobs requesting both “Force” and SPE calculations at different levels of theory using the Gaussian03⁵⁵ software package for our system in the gas-phase. Fig. 10A presents t_{SPE} and t_{Force} required for the corresponding job types on a single core of a Quad-Core AMD Opteron 2356 CPU, with 4

Gb available memory. Of course, the calculations can be parallelized with different levels of efficiency, but since this varies between different approaches, our benchmark timings are reported here for a sequential job on one CPU. In order to make the real cost of the calculation clear to the reader, we estimated the total number of CPU-hours for a full QM(ai)/MM study by our PD approach, and compared this with an estimate for the computational cost of studying the same system using the MTD⁴¹ approach, using Eq. 18:

$$\frac{t_{\text{MTD}}}{t_{\text{PD}}} = \frac{N_{\text{MTD}} t_{\text{Force}}}{0.5 N_{\text{PD}} (t_{\text{Force}} + t_{\text{SPE}})} \quad (18)$$

here, the total time needed for the MTD and PD calculations are designated as t_{MTD} and t_{PD} , respectively, and N_{MTD} and N_{PD} denote the total number of QM calls for each approach respectively. As mentioned above, in the case of MTD, this involves $2 \cdot 10^6$ “Force” calls, as per Ref. ⁴¹, and for our PD approach, this involves 12000 total QM calls, of which 6000 are SPE calculations and 6000 are “Force” calculations. From this relationship, it can be seen that our PD approach is ~200 times computationally more efficient than the MTD approach (though, clearly, the precise factor depends on the QM method used).

IV.5. Choosing an Appropriate Reaction Coordinate

At this stage, it is important to comment on our choice of RC, x , in this work. In general, the choice of RC is of crucial importance, as it is not only important to select a RC that properly describes the relevant process of interest, but also, a poorly selected RC will greatly affect the convergence of free energy calculations, as the chosen RC is often used to construct the mapping potential. Now it may be tempting to describe reaction progress by means of the popular approach of using a geometric coordinate, and, in fact, the “collective variables” that are normally chosen to define the FES in MTD tend to be geometric variables such as bond distances, angles, torsions, etc. However, while such a description may be effective in gas-phase studies, it becomes rapidly problematic when attempting to model the multi-dimensional RC in solution and in proteins (particularly as it is important in such cases to be able to efficiently capture the effect of solvent polarization).

Fortunately, the PD sidesteps all problems associated with the use of a geometric RC by following our earlier treatment³⁹ of using as our RC the electronic energy gap (E_{gap}) between the two diabatic states (see discussion in Section III.1). This approach, which we believe is at present the best RC when studying chemical processes in solution and in enzymes, is particularly powerful when one tries to represent the entire many dimensional solvent space by a single RC (see e.g. discussion in Ref. ⁴⁴). Specifically, the energy gap coordinate, $x(\mathbf{R}, s)$ naturally includes the dependence on the solute geometrical coordinates, $\mathbf{R}$, as well as on the solvent coordinate s , reflecting the solute-solvent interactions, which are dominated by the electrostatic term⁴⁴. As the reaction involves changes in both $\mathbf{R}$ (due to changes in the chemical structure of the solute) and s (due to the response of the solvent to the change in the solute charge distribution), it is crucial to correctly represent their coupling at each point along the reaction path. The use of just the solute coordinate, $\mathbf{R}$, requires extreme caution and longer relaxation times in order for the solvent to adjust to the new charge distribution on the solute atoms, and it also misses the so-called non-equilibrium solvation barrier (see Refs. ^{44, 56}). In contrast, the energy gap represents the entire configurational space of the system (both solute and solvent), and therefore an EVB mapping potential driving the system from the reactant to products along this coordinate allows for more efficient sampling. Additionally, the energy gap coordinate appears to accurately capture the physics of the solvent reaction field, a fact that is now recognized by many workers. In comparison, it is unlikely that a blind MTD search will find the optimal

solvent coordinate (see section V). This was specifically illustrated by a recent study⁴¹, which examined the efficiency of the energy gap as an universal RC for the simulation of chemical reactivity, using the same test case as that presented in the current work. In this work, the authors demonstrated that not only does E_{gap} (which can be used in systems described with any quantum chemistry Hamiltonian) provide significantly more efficient sampling than a geometrical RC (see above), as well as allowing better TS localization, but, more importantly for the cases in which E_{QM} can be reproduced within 2-3RT (which we demonstrate is the case with our PD approach), the use of E_{gap} allows one to calculate the free energy barrier with two orders of magnitude less computational effort than performing MTD calculations on the same system.

V. Additional Clarifications About the Relationship between the PD and MTD Approaches

The present work focuses on comparing the MTD and PD approaches in the specific case of QM/MM calculations. However, we would like to emphasize that the PD approach is *not* a specialized method that performs best in the case of one-dimensional problems. That is, like the MTD approach, it is an extremely generalized approach that can be used in countless complex situations (and is possibly even more generalized than MTD). While we are not questioning the power of the MTD approach, we would nevertheless like to clarify that assuming that the PD is a specialized esoteric approach would be unjustified, but rather, that it is in fact very generally applicable.

To elaborate on the dimensionality issue, we start from a rather trivial clarification of the fact that the PD is not actually a highly specialized approach to construct a 1-D free energy profile, using E_{gap} as the reaction coordinate. That is, the advantages of E_{gap} compared to other definitions of the reaction coordinate were already presented in Section IV.5. Here, we would like to point out that while E_{gap} is in our opinion the most reliable reaction coordinate, nevertheless, the PD approach is not *limited* to using this reaction coordinate. In order to illustrate this fact, we have provided the PD-refined EVB free energy profile along the conventional geometrical coordinate used to define S_N2 reactions (i.e. the difference in distance between the electrophilic center and the leaving and attacking nucleophiles respectively) in Fig. 11. This profile closely represents the QM FES constructed using the more expensive approach of Ref. ⁴¹ in the whole range of the RC between the RS and the PS. In Fig. 12, we also provide the 2-D PD-refined EVB FES (defined in terms of the distances between the central carbon atom and the leaving and attacking nucleophiles), where it can be seen that the coordinates of the relevant stationary points (RS, TS and PS) are close to the PM3 optimized values (i.e. 1.806/2.843, 2.189/2.189 and 2.843/1.806 Å respectively). Both surfaces were, as mentioned, obtained using EVB FEP/US *after* performing the full PD refinement, for the test system considered in this work (see Fig. 3). Once the PD refinement has been performed, it is trivial to obtain the 2-D FES by post-processing a single computational job, in contrast to the MTD approach, which would require a separate run for each choice of RC. Additionally, it should be pointed out that the PD approach has already been used to generate a multi-dimensional surface⁵⁷, in a case where a renormalized reference potential was used to examine the energy landscape for both the conformational change and the chemical step in adenylate kinase. Finally, it is also tempting to assume that in order to successfully use our PD approach, it is essential to know the end points of the simulation. However, even in cases where the RC is not known, it would be fairly trivial to run MC simulations with a soft potential in order to find low-energy points, and to then use them for further PD refinement.

Following on from this, one can assume that the impressive accomplishments of the MTD approach are exclusive to this method. As an example, it would be tempting to assume that

the elegant studies of phase transitions in crystals done by the MTD approach (e.g. Ref. ⁵⁸) can not be achieved using the PD. However, we have been studying crystals for a long time, including thermal expansion⁵⁹ and photochemistry in crystals⁶⁰, and it would be rather straightforward to use our MCA program⁶¹ in order to first empirically simulate crystal properties, and to then use the PD approach to move to a QM periodic potential.

Finally, we would like to point out that in recent years, with increasing computational power, more and more attempts have been made to study complex systems using brute-force all atom approaches (e.g. Ref. ⁶² to name a single example amongst countless more in the literature). While such studies are, of course elegant, they suffer from the fact that to truly obtain insight into complex systems, it is necessary to run more than just a single trajectory, and the computational cost of studying long timescale processes in such systems using brute-force all atom approaches with proper sampling is at present simply prohibitive on even the most powerful supercomputers available. In contrast, as early as 1975²⁶, Warshel and Levitt demonstrated the power of using coarse-grained representations in biology by using a simplified potential to successfully simulate protein folding. Since then, such simplified representations have been used to address even more challenging problems, and we have recently reviewed the applications of coarse-grained simulations to chemistry and biology in great detail (for further reading, see Refs.^{63, 64} and references cited therein). Clearly, the PD approach is a broadly general approach with which to refine the simplified representation of a system relative to the results obtained using a higher-level potential, and such an approach has broad applicability to a wide range of problems from protein folding to, as mentioned above, even solid-state chemistry. One notable example of this is our recent work on adenylate kinase⁵⁷, in which we used a renormalization approach to examine the dynamics of the system on the ms timescale. We would like to point out that once the simplified potential has been renormalized, it is possible to run a ms trajectory in the space of a week on a standard desktop computer, while maintaining a reasonable physical description of the system. This is in sharp contrast to the prohibitory amount of computational time and power that would be required to run a similar trajectory using an all-atom brute force approach. Finally, in the case of a truly complex system with many minima, it is possible to examine the nature of the surface by initial exploration with a high-level approach, but of course having a physically meaningful CG model will make the search much more effective.

VI. Concluding Remarks

The present work clarifies the formulation and implementation of our PD approach, and compares it to the currently popular MTD approach. Significant effort is placed into highlighting the conceptual similarities between the earlier PD and subsequent MTD approaches, as well as to demonstrating the actual computational requirements of each approach. We demonstrate that the PD approach provides major savings in computational time due to the fact that it requires significantly fewer QM calls to accurately evaluate the activation free energy than the MTD approach. That is, since the main contributions to the barrier height are calculated by the EVB FEP/US approach, where MD is run on the EVB mapping potential, the calls to the QM program are only needed in order to evaluate the free energy difference between the EVB and QM surfaces at the key inflection points on the RC, namely the RS and TS.

We would like to reiterate that the core of the power of MTD lies in providing an elegant way to flatten the minima. However, we believe that intuitive knowledge of chemistry or biology would be more effective, as, for instance, an understanding of the solution energetics (e.g. the solvent coordinate) can provide a great amount of insight. That is, one of the goals of the MTD approach is to be able to perform free energy calculations which

sample a significant part of the surface with as little *a priori* information as possible, making it a broadly applicable approach. In contrast to this, as mentioned above, it might be assumed by some that the PD approach is a problematic approach in that the end points of the simulation need to be known. Fortunately, in most cases this is not an issue, as they are indeed known. However, even in those cases where the end points are not known, it is fairly straightforward to perform a very fast mapping between the two states using a QM method in order to obtain information about the initial surface, fit this to an EVB potential, and then perform the PD as outlined in this work. Additionally, there are cases where the pathway is not intuitive, and where it may seem like using a “black box” approach such as the MTD approach is more informative. However, even in such situations, applying the renormalization approach with a weak potential allows one to rapidly examine alternative pathways. Additionally, using for instance the EVB as a reference potential includes an enormous amount of chemical information, which is clearly an advantage and not a disadvantage. Finally, not only that our QCP approach²⁸ of running on a free particle potential is formally similar to the MTD after having filled the Gaussians, but also, our proposal⁸ for minimizing the variance between the two potentials is the equivalent of refining the Gaussians (no need to do it with “time dependence”). Therefore, overall, while it may be tempting to go for a black box approach, we demonstrate here that this can have a highly prohibitive computational cost, and that learning about a complex system without knowing the end points of the process is not necessarily the best approach for resolving this issue. At any rate, coming back to the illustrative example used in this work, we are not aware of any MTD calculations that have provided reliable QM(*ai*)/MM free energy surfaces with less effort than that required for our reference potential approach, and we are fairly sure that this observation can be extended to other potential applications.

It should also be pointed out that, to a large extent, the advantage of the PD approach comes from the wealth of chemical information provided by the EVB treatment (as mentioned above). However, the general idea of using a simple reference potential can also be exploited by an adaptation of our approach in which, rather than using the EVB as was done in this work, the reference potential is evaluated by means of a fast MO semi-empirical approach. In such a case, however, it may be more complicated to refine the semi-empirical integrals by means of the least squares refinement, unless one simply adds an EVB-type (or other simple analytical) correction potential. Interestingly, the PD option of refining the reference potential can also include using Gaussian type functions that will help minimize the difference between V_0 and V' . In any case, regardless of the precise reference potential used, the key fact demonstrated here is that despite the popularity of “black box” approaches, using a properly refined reference potential (combined with chemical intuition) allows for the performance of accurate *ab initio* QM/MM free energy calculations with proper sampling, at a fraction of the computational cost associated with related existing approaches, such as MTD.

The present work has mainly addressed the advantages of the PD approach over the MTD approach in the specific case of QM(*ai*)/MM calculations. However, it may here be also useful to address the fact that, as outlined in the introduction, the MTD approach has been used for many other applications, sometimes with quite impressive results. In this work, our aim has not been to perform a comparison of the relative advantages of the two approaches in terms of different applications (although it would be instructive to explore this issue), nor are we aiming to in any way discredit the MTD approach. Nevertheless, we would like to clarify several key points. First, we would like to note that we have invested major effort into exploring the actual performance of different sampling approaches in complex systems, which includes examining the performance of PMF, WHAM, FEP and related approaches both in ion channels⁶⁵ as well as in general simulations⁶⁶. We believe that providing such information is crucial for an unbiased assessment of what the optimal approach is. Second,

and perhaps more importantly, while it may seem as if the PD is specific to QM(ai)/MM calculations, it is in fact an umbrella concept which includes both our renormalization approach^{57, 63, 66–68}, as well as our general approach for improving CG models (e.g. ³¹). This general strategy has also been implemented in more mathematical versions⁶⁹ as well as in closely related strategies (e.g. ⁷⁰), and covers many problems that one could otherwise consider using the (computationally more expensive) MTD approach for. For instance, our renormalization approach allows us to explore extremely complex landscapes with high efficiency⁵⁷. Moreover, the general strategy of refining CG models to best describe an explicit complex system is, in our opinion, not only broadly applicable, but also perhaps more effective than the MTD approach, as it actually stores the information from one system upon moving to the next related system (an example of this would be obtaining a general CG model to model complex biological systems). In other words, the key issue here is not one of presenting a highly specialized approach to accelerate QM(ai)/MM calculations, but rather one of examining the relationship between the concepts of using a reference potential (PD) relative to constructing the equivalent of the negative of the reference potential by gradually flattening the original surface (as is done in MTD). Further examples of the broad applicability of our approach are seen in our use of a CG simplified folding model as a reference potential for exploring problems in protein folding, conformational changes, and related problems^{31, 63, 64}. In such cases, the CG model can provide key information about the relevant landscape, and one can then move to the full explicit surface by means of our thermodynamic cycle. One can, of course, then use the iterative refinement approach applied in this work, and minimize the LRA contribution by optimizing the CG model. Finally, perhaps the greatest testimony to the power of the reference potential approach is its increasing appreciation by other workers, who are now using it in different adaptations^{70–72}. In this respect, we would like to strongly emphasize that we do not in any way intend this work to be a general contest between the PD and MTD approaches. But, in our point of view, the fact that the PD generates a transferable reference potential that can be used in any other related system (in other words, the refined EVB can be used in any related reaction) is a major advantage of the PD approach.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

AW would like to thank the NIH (grants GM24492 and 1U19CA105010), and SCLK would like to thank the Swedish research council (VR, grant 2010-5026) for funding. All computational work was performed on the University of Southern California High Performance Computing and Communication Center (HPCC), whom we would like to thank for having provided us with access to computer time on the HPCC cluster.

References

1. Bentzien J, Muller RP, Florian J, Warshel A. *J Phys Chem B*. 1998; 102:2293–2301.
2. Štrajbl M, Hong G, Warshel A. *J Phys Chem B*. 2002; 106:13333–13343.
3. Rosta E, Klähn M, Warshel A. *J Phys Chem B*. 2006; 110:2934–2941. [PubMed: 16471904]
4. Rosta E, Haranczyk M, Chu ZT, Warshel A. *J Phys Chem B*. 2008; 112:5680–5692. [PubMed: 18412414]
5. Kamerlin SCL, Haranczyk M, Warshel A. *J Phys Chem B*. 2009; 113:1253–1272. [PubMed: 19055405]
6. Shurki A, Warshel A. *Adv Protein Chem*. 2003; 66:249–313. [PubMed: 14631821]
7. Hu H, Yang W. *Ann Rev Phys Chem*. 2008; 59:573–601. [PubMed: 18393679]
8. Muller RP, Warshel A. *J Phys Chem*. 1995; 99:17516–17524.

9. Zhang Y, Liu H, Yang W. *J Chem Phys.* 2000; 112:3483–3492.
10. Rod TH, Ryde U. *Phys Rev Lett.* 2005; 94:138302. [PubMed: 15904045]
11. Olsson MHM, Hong G, Warshel A. *J Am Chem Soc.* 2003; 125:5025–5039. [PubMed: 12708852]
12. Liu WB, Wood RH, Doren DJ. *J Phys Chem B.* 2003; 107:9505–9513.
13. Iftimie R, Salahub D, Schofield J. *J Chem Phys.* 2003; 119:11285–11297.
14. Crespo A, Marti MA, Estrin DA, Roitberg AE. *J Am Chem Soc.* 2005; 127:6940–6941. [PubMed: 15884923]
15. Pradipta B. *J Chem Phys.* 2005; 122:091102. [PubMed: 15836102]
16. Luzhkov V, Warshel A. *J Comp Chem.* 1992; 13:199–213.
17. Huber T, Torda AE, van Gunsteren WF. *J Comput Aided Mol Des.* 1994; 8:695–708. [PubMed: 7738605]
18. Christen M, Keller B, van Gunsteren WF. *J Biomol NMR.* 2007; 39:265–273. [PubMed: 17929172]
19. Bakowies D, van Gunsteren WF. *J Mol Biol.* 2002; 315:713–736. [PubMed: 11812142]
20. Laio A, Parrinello M. *Proc Natl Acad Sci U S A.* 2002; 99:12562–12566. [PubMed: 12271136]
21. Donadio D, Bernasconi M. *Phys Rev B.* 2005; 71:073307.
22. Ascianto E, Sagui C. *J Phys Chem A.* 2005; 109:7682–7687. [PubMed: 16834142]
23. Urakawa A, Iannuzzi M, Hutter J, Baiker A. *Chemistry.* 2007; 13:6828–6840. [PubMed: 17566132]
24. Michel C, Laio A, Mohamed F, Krack M, Parrinello M, Milet A. *Organometallics.* 2007; 26:1241–1249.
25. Ensing B, De Vivo M, Liu Z, Moore P, Klein ML. *Acc Chem Res.* 2006; 39:73–81. [PubMed: 16489726]
26. Levitt M, Warshel A. *Nature.* 1975; 253:694–698. [PubMed: 1167625]
27. Fan ZZ, Hwang JK, Warshel A. *Theor Chem Acc.* 1999; 103:77–80.
28. Hwang J-K, Warshel A. *J Phys Chem.* 1993; 97:10053–10058.
29. Hwang J-K, Warshel A. *J Am Chem Soc.* 1996; 118:11745–11751.
30. Fan ZZ, Hwang J-K, Warshel A. *Theor Chem Acc.* 1999; 103:77–80.
31. Messer BM, Roca M, Chu ZT, Vicatos S, Kilshtain A, Warshel A. *Proteins: Struct Funct Bioinform.* 2010; 78:1212–1227.
32. Wesolowski T, Muller RP, Warshel A. *J Phys Chem.* 1996; 100:15444–15449.
33. Roca M, Liu H, Messer B, Warshel A. *Biochemistry.* 2007; 46:15076–15088. [PubMed: 18052079]
34. Roca M, Messer B, Hilvert D, Warshel A. *Proc Natl Acad Sci USA.* 2008; 105:13877–13882. [PubMed: 18779576]
35. Chandler, D. *Finding transition pathways: Throwing ropes over rough mountain passes, in the dark.* World Scientific; Singapore: 1998.
36. Dellago C, Boluis PG, Geissler PL. *Adv in Chem Phys.* 2002; 123:1–78.
37. Bolhuis PG, Chandler D, Dellago C, Geissler P. *Ann Rev Phys Chem.* 2002; 59:291–318. [PubMed: 11972010]
38. Åqvist J, Warshel A. *Chem Rev.* 1993; 93:2523–2544.
39. Warshel, A. *Computer modeling of chemical reactions in enzymes and solutions.* John Wiley & Sons; New York: 1991.
40. Wesolowski T, Warshel A. *J Phys Chem.* 1994; 98:5183–5187.
41. Mones L, Kulhánek P, Simon I, Laio A, Fuxreiter M. *J Phys Chem B.* 2009; 113:7867–7873. [PubMed: 19432459]
42. Kamerlin SCL, Warshel A. *Faraday Discuss.* 2010; 145:71–106.
43. Zwanzig RW. *J Chem Phys.* 1954; 22:1420.
44. Hwang J-K, King G, Creighton S, Warshel A. *J Am Chem Soc.* 1988; 110:5297–5311.
45. King G, Warshel. *J Chem Phys.* 1990; 93:8682–8692.
46. Hong G, Rosta E, Warshel A. *J Phys Chem B.* 2006; 110:19570–19574. [PubMed: 17004821]

47. Lee FS, Chu ZT, Warshel A. *J Comp Chem*. 1993; 14:161–185.
48. Stewart, JJP. MOPAC2009. Colorado Springs; CO, USA: 2008.
49. Stewart JJP. *J Comput Chem*. 1989; 10:209–220.
50. Stewart JJP. *J Comput Chem*. 1989; 10:221–264.
51. Kamerlin SCL, Florián J, Warshel A. *ChemPhysChem*. 2008; 9:1767–1773. [PubMed: 18666265]
52. Rosta E, Kamerlin SCL, Warshel A. *Biochemistry*. 2008; 47:3725–3735. [PubMed: 18307312]
53. Kamerlin SCL, Warshel A. *J Phys Chem B*. 2009; 113:15692–15698. [PubMed: 19888735]
54. Barlow SE, van Doren JM, Bierbaum VM. *J Am Chem Soc*. 1988; 110:7240–7242.
55. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven JT, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone BMV, Cossi M, Scalmani G, Rega N, Petersson HNGA, Hada M, Ehara M, Toyota K, Fukuda JHR, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai MKH, Li X, Knox JE, Hratchian HP, Cross JB, Adamo JJC, Gomperts R, Stratmann RE, Yazyev O, Austin RCAJ, Pomelli C, Ochterski JW, Ayala PY, Morokuma GAVK, Salvador P, Dannenberg JJ, Zakrzewski SDVG, Daniels AD, Strain MC, Farkas DKMO, Rabuck AD, Raghavachari K, Foresman JVOJB, Cui Q, Baboul AG, Clifford S, Cioslowski BBSJ, Liu G, Liashenko A, Piskorz P, Komaromi RLMI, Fox DJ, Keith T, Al-Laham MA, Peng ANCY, Challacombe M, Gill PMW, Johnson WCB, Wong MW, Gonzalez C, Pople JA. Gaussian 03, Revision C03. 2004
56. Villà J, Warshel A. *J Phys Chem B*. 2001; 105:7887–7907.
57. Pislakov AV, Cao J, Kamerlin SCL, Warshel A. *Proc Natl Acad Sci U S A*. 2009; 106:17359–17364. [PubMed: 19805169]
58. Martonak R, Laio A, Bernasconi M, Ceriani C, Ralteri P, Zipoli F, Parinello M. *Comput Crystallography*. 2005; 220:489–498.
59. Warshel A, Lifson S. *Journal of Chemical Physics*. 1970; 53:582–594.
60. Warshel A, Shakked Z. *Journal of the American Chemical Society*. 1975; 97:5679–5684.
61. Huler, E.; Warshel, A. QCPE 325: Quantum Chemistry Program Exchange. Indiana University; 1976.
62. Voelz VA, Bowman GR, Beauchamp K, Pande VS. *J Am Chem Soc*. 2010; 132:1526–1528. [PubMed: 20070076]
63. Kamerlin SCL, Vicatos S, Warshel A. *Ann Rev Phys Chem*. 2011; 62:41–64. [PubMed: 21034218]
64. Kamerlin SCL, Warshel A. *PhysChemChemPhys*. In Press.
65. Kato M, Warshel A. *J Phys Chem B*. 2005; 109:19516–19522. [PubMed: 16853521]
66. Dryga A, Warshel A. *J Phys Chem B*. 2010; 114:12720–12728. [PubMed: 20836533]
67. Burykin A, Kato M, Warshel A. *PROTEINS: Struct Func Genet*. 2003; 52:412–426.
68. Liu H, Shi Y, Chen XS, Warshel A. *Proc Natl Acad Sci U S A*. 2009; 106:7449–7454. [PubMed: 19383795]
69. Brandt, A. *Principles of systematic upscaling*. Oxford University Press; Oxford: 2008.
70. Savelyev A, Papoian GA. *Proc Natl Acad Sci U S A*. 2010; 107:20340–20345. [PubMed: 21059937]
71. Hwang J-K, Pan J-J. *Chem Phys Lett*. 1995; 243:171–175.
72. Woods CJ, Manby FR, Mulholland AJ. *J Chem Phys*. 2008; 128:014109. [PubMed: 18190187]

Figure 1. A conceptual comparison of meta- and paradynamics

(A) The MTD approach iteratively adds Gaussian potentials, $V_1', V_2', V_3' \dots V_N'$, to the real potential of the system, V_0 . At the final iteration, the addition of the Gaussian potential V_N' to the real potential creates a flat potential, $V_0 + V_N'$, which allows for sampling of the TS region. (B) In the PD approach, the reference potential (in this case an EVB potential), V_1' , is iteratively and systematically refined ($V_2', V_3' \dots V_N'$) to reproduce the real potential, V_0 . The iterative refinement of the EVB 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 (e.g. by EVB FEP/US), plus a minor LRA correction.

Figure 2. The thermodynamic cycle used in calculating the QM(AI) activation free energy
 The QM activation free energy (vertical dashed arrow) is calculated as the sum of the EVB barrier (vertical solid arrow) evaluated by EVBFEP/US method, and the perturbations between the QM and EVB surfaces at the RS and TS (horizontal solid arrows), which is evaluated using the LRA approach.

Figure 3. The S_N2 reaction between chloride ion and methyl chloride

This reaction was taken as the test system in the present work. The subscripts A and B are used to distinguish between the two chlorine atoms.

Figure 4. Describing the EVB resonance structures

The figure illustrates the nature of the EVB bonding along the reaction coordinate (e.g. the RS and TS), where, for each geometry, we use the two relevant diabatic states, with the indicated bonding pattern.

Figure 5. The initial EVB free energy surface

The figure depicts the free energy surface obtained for our S_N2 reaction, prior to the refinement of any EVB parameters. The corresponding EVB parameters are given in Table S1 of the Supplementary Information.

Figure 6. Fluctuations of the energy gap ($E_{\text{QM}} - E_{\text{EVB}}$) for trajectories run on the initial EVB and PM3 potentials

The energy fluctuations (blue dots) were obtained during MD runs on: (A) the original EVB potential at the RS; (B) the PM3 potential at the RS; (C) the EVB potential with constraints (see the main text) at the TS, and (D) the PM3 potential with constraints at the TS. The averages over these fluctuations (dashed orange lines) are used to obtain the LRA estimates of the free energy of transfer from the EVB to the PM3 surface ($\Delta\Delta G_{\text{EVB} \rightarrow \text{QM}}$) at the RS (A, B) and TS (C, D), according to Eq. (12).

Figure 7. Illustrating the effect of refining the EVB parameters by the PD approach

The blue line shows the energy difference between the EVB and PM3 potentials before the parameter refinement, while the red line is the result after a series of refinement iterations done for: (A) just the structures generated at the RS, where E_{EVb} has no contribution from H_{12} ($E_{\text{EVb}}(\text{RS}) = E_1$) and the only parameters refined are those of EVB diabatic states; and (B) both geometries lying near the RS, as well as those lying near the TS. Here the refinement is done for H_{12} , and the Cl-C-Cl angle deformation energy parameters, while other parameters are taken from the earlier iteration round, the results of which is depicted in Fig. 7A.

Figure 8. The EVB free energy surface obtained with the refined parameters

This figure depicts the free energy surface obtained for our S_N2 reaction, after the refinement of the EVB parameters. The refined parameters are provided in Table S2 of the Supplementary Information.

Figure 9. Fluctuations of the energy gap, $E_{QM} - E_{EVB}$, for trajectories run on the refined EVB and PM3 potentials

The relevant energy fluctuations (dotted lines) were obtained during MD runs on the refined EVB (red) and PM3 (blue) potentials at the RS (A) and TS (B). The average values are shown as dashed lines of the respective colors.

Figure 10. A comparison of the simulation times required for MTD and PD, at different levels of QM theory

(A) Time per QM call (in seconds) for different methods, for calculations of only the single point energy (SPE, black) as well as calculations that calculate both the single point energy and the energy derivatives (Force, red). The calculations of the energy and its derivatives were performed for the studied system using G03 on a single core of a Quad-Core AMD Opteron2356 CPU with 4GB available memory. (B) The estimated total number of CPU-hours required for a PD study of the test system presented in this work (black), comprising a total of 12000 QM calls, and for a corresponding MTD study of the same system (red), which would comprise $2 \cdot 10^6$ QM calls, based on Ref. ⁴¹. Note that the simulation time for method 11 (60667 CPU-hours) is out of scale. In both cases, the X-axis represents the relevant QM approach. The QM approaches used for the benchmark are: (1) PM3, (2) RHF/6-31G(d), (3) RMP2-FC/6-31G(d), (4) RB3LYP/6-31+G(d), (5) RmPW1PW91/6-31+G(d), (6) RMP2-FC/6-31+G(d), (7) RCISD-FC/6-31G(d), (8) RCISD-FC/6-31+G(d), (9) RMP2-FC/6-311+G(d,p), (10) RCCSD-FC/6-31G(d) and (11) RCCSD-FC/6-31+G(d).

Figure 11. A comparison of the free energy surfaces obtained using different approaches for the S_N2 reaction studied in this work

Shown here is the free energy surface in the reaction coordinate defined by $r(\text{C-Cl}_A)-r(\text{C-Cl}_B)$ (see Fig. 3) obtained by: EVB FEP/US after the PD refinement (blue), the MTD approach of Ref. ⁴¹(red) as well as the free energy values of the PM3 optimized reactant, transition and product states (cyan), obtained using the harmonic approximation.

Figure 12. 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 respectively (i.e. $r(\text{C}-\text{Cl}_A)$, $r(\text{C}-\text{Cl}_B)$ of Fig. 3). This surface was obtained by post-processing the same job as for Figs. 8 and 11.

Table 1

Evaluating the free energies for transferring from the EVB to the PM3 free energy surface, using the LRA approach outlined in Eq. 12. All energies are in kcal/mol.

	Default EVB	PD refinement
$\langle E_{\text{PM3}} - E_{\text{EVB}} \rangle_{\text{EVB(RS)}}$	18.79	4.08
$\langle E_{\text{PM3}} - E_{\text{EVB}} \rangle_{\text{PM3(RS)}}$	-0.12	0.68
$\Delta\Delta G(\text{RS})_{\text{EVB} \rightarrow \text{PM3}}$	9.34	2.38
$\langle E_{\text{PM3}} - E_{\text{EVB}} \rangle_{\text{EVB(TS)}}$	5.65	4.37
$\langle E_{\text{PM3}} - E_{\text{EVB}} \rangle_{\text{PM3(TS)}}$	-0.96	-0.36
$\Delta\Delta G(\text{TS})_{\text{EVB} \rightarrow \text{PM3}}$	2.34	2.00
$\Delta\Delta g_{\text{EVB} \rightarrow \text{PM3}}^\ddagger$	-7.00	-0.38