Hybrid QM/MM Potentials of Mean Force with Interpolated Corrections

J. Javier Ruiz-Pernía, Estanislao Silla, and Iñaki Tuñón*

Departament de Química Física/IcMol, Universidad de Valencia, 46100 Burjasot, Valencia, Spain

Sergio Martí and Vicent Moliner

Departament de Ciències Experimentals, Universitat Jaume I, Box 224, 12080 Castellón, Spain

Received: January 27, 2004; In Final Form: April 2, 2004

Because of its computational cost, QM/MM simulations are usually carried out using low-quality Hamiltonians, such as semiempirical, which are not always able to provide an accurate potential energy surface. We here propose a simple but efficient way to obtain corrected quantum mechanics/molecular mechanics (QM/MM) potentials of mean force (PMF) for chemical processes in condensed media. By means of dual-level calculations on the QM subsystem, we evaluate a correction energy term by employing either the polarized or the unpolarized wave functions. This energy term is evaluated as a function of the distinguished reaction coordinate biased in the calculation of the PMF. Using a mapping coordinate and splines under tension, its derivatives can be readily included to perform molecular dynamics simulations. The structures selected to evaluate the energy correction are chosen from a reaction path obtained in the condensed media, ensuring then that they are representative of the ensemble of structures sampled during the simulation. We have tested the proposed scheme with two prototypical examples: the Menshutkin reaction in aqueous solution and the chorismate rearrangement to prephenate catalyzed by *Bacillus subtilis* chorismate mutase. In both cases the use of interpolated corrections clearly improves the quality of the results.

1. Introduction

When studying chemical reactions in condensed media, one must afford a double problem: on one hand it is necessary to use quantum mechanics to describe the electronic changes, and on the other hand one must include a large portion of the full system to take into account long-range effects. As an efficient compromise between accuracy and efficiency, hybrid quantum mechanics/molecular mechanics (QM/MM) methods have become very popular to simulate such chemical processes.[1−11] The inherent characteristic of all these methods is the division of the full system in two subsystems. One, the part in which the chemical changes are taking place, is described using quantum mechanics. For the rest of the system a molecular mechanics description is chosen. In this way, one can obtain a reasonable potential energy surface (PES), which includes the more important effects on the energetics of the process under study, at a moderate computational cost.

However, knowledge of this PES is not enough to describe correctly chemical reactions in condensed media. When studying chemical reactivity in solution or in enzymatic active sites, one is faced with a substantial difference with respect to gas phase reactivity. In this last case, both the reactant and the transition state usually correspond to single structures and the thermodynamic properties can be obtained by applying statistical thermodynamics to the energy levels of these structures. In solution or enzymatic environments there are a large number of conformations accessible to the environment and then one could find a myriad of stationary structures that could be assigned as reactants or transition structures for a particular process. Thus, the statistical treatment must include the exploration of a significant ensemble of minima and transition structures appearing on the PES to properly define the reactant and transition states. This ensemble can be generated using different simulation techniques such as Monte Carlo or molecular dynamics. Information obtained from these simulations can be then used to derive thermodynamic information and in particular the activation free energy, which can be related to the rate of a chemical reaction through the use of transition state theory. The problem is that the correct evaluation of all the contributions to the free energy using high-level quantum treatments is nowadays unaffordable, because current simulation techniques require, in principle, the evaluation of the quantum subsystem wave function thousands or even millions of times. Several approximations can be used to search a solution. Roughly speaking, one must renounce to include the quantum subsystem flexibility in the simulation or alternatively one is then compelled to use a low-level electronic description.

One possible strategy is to sample only the conformational changes taking place in the MM part, discarding then the internal movements of the quantum subsystem. Continuum solvent models[12−14] could be considered an extreme case where discrete molecules have been finally replaced by an averaged description. For enzymatic reactions, the quantum mechanics-free energy perturbation (QM-FEP) method[15] is an example of this kind of technique. In this approach the reaction path is usually obtained for a gas phase model of the active site. Then high-level quantum treatments are not computationally forbidden. Coordinates of the chemical system and point charges derived from quantum calculations can be then used in classical simulations where only the changes in the environment are sampled. Free energy perturbation (FEP) is then used to obtain the free energy

* To whom correspondence should be addressed. E-mail: ignacio.tunon@uv.es.

10.1021/jp049633g CCC: $27.50
Published on Web 05/25/2004

change as the sum of the gas phase reaction energy and environment contributions to the free energy. The main drawbacks of this strategy are that contributions arising from the movement of the chemical system are not considered and the fact that the whole environment is not incorporated into the calculation of the reaction path and then the two subsystems are not fully coupled.

Free energies profiles can be also obtained as a potential of mean force (PMF) appearing along a particular reaction coordinate.[16] The umbrella sampling method[17] is used to place the chemical system at different values of the reaction coordinate that cannot be sampled frequently enough by thermal fluctuations. Simulations are then carried out by sampling all the degrees of freedom of the system except for the biased reaction coordinate. Once the reaction coordinate has been fully explored from reactants to products, the total probability distribution function is obtained and thus the free energy profile is calculated. The main advantage of this technique is then the inclusion of all contributions to the free energy. This strategy has an obvious drawback: it requires the evaluation of the energy and gradients of the system at each simulation step, and then the wave function of the quantum subsystem must be obtained thousands of times. These calculations are thus expensive and nowadays restricted to semiempirical Hamiltonians[18−20] or empirical valence bond methods.[21,22]

To overcome the quantitative limitations imposed by the use of semiempirical Hamiltonians in the description of the PES, several methodologies have been proposed. One obvious solution is to develop a new parametrization exclusively for the process under study (specific reaction parameters, SRP).[23−25] This has been adopted in a number of cases, but it is not always easy to improve simultaneously the reaction and activation energies. Moreover, continuity problems can arise when dealing with multistep processes, as far as different parametrizations may be required. Another possibility is to include correction terms to the potential energy surface based on valence bond theory.[26] This strategy can be successful if the semiempirical Hamiltonian gives a good qualitative PES. Another strategy has been proposed to obtain ab initio QM/MM free energy profiles using a simple reference potential (for example, empirical valence bond).[27] Here we present a new approach to correct low-level energy functions for PMF calculations. The proposal is based on the interpolated corrections scheme proposed by Truhlar et al.[28−30] for dynamical calculations of gas phase chemical reactions. This method provides an efficient way to combine information computed at a lower level of electronic structure calculations with selected results computed with a higher level for a reduced number of geometries. The theory has been tested for a number of cases, showing that reasonable accuracy can be reached at a reduced computational cost.[28] The implementation shown here is based on the partition of the total QM/MM energy into three components: the QM, the MM, and the interaction terms. Dual-level calculations are carried out just for the QM energy or for both the QM and the interaction components. The main advantage of this scheme is that it can be easily modified for PMF calculations in condensed media, if forces are also conveniently corrected. Adaptation of the original method of Truhlar et al. is described in the following section, and afterward we show the results of its application to a chemical reaction in solution and in an enzyme. In particular, we have selected two paradigmatic reactions, which have been thoroughly studied with different methodologies: the Menshutkin reaction in aqueous solution and the chorismate rearrangement to prephenate catalyzed by *Bacillus subtilis* chorismate mutase.

2. Theory

As explained before, PMF calculations are obtained from simulations based on a QM/MM PES. The potential energy is obtained as a function of the coordinates of the molecular mechanics atoms (r_m) and of the quantum atoms (r_q). For a given configuration of the system (r_q, r_m) the energy is then

$$E(r_q,r_m) = E_{QM}(r_q) + E_{QM/MM}(r_q,r_m) + E_{MM}(r_m) \quad (1)$$

where the first term of the right-hand side is the energy of the quantum subsystem, the second term includes the interaction and polarization between the subsystems and the last one is the energy of the classical subsystem. The dependence on the electronic coordinates has been omitted for the shake of simplicity. In terms of the quantum subsystem wave function, the first and second contributions can be expressed as

$$E_{QM}(r_q) = \langle \Psi^0(r_q)|\hat{H}^0(r_q)|\Psi^0(r_q)\rangle \quad (2)$$

$$E_{QM}(r_q) + E_{QM/MM}(r_q,r_m) = \langle \Psi(r_q,r_m)|\hat{H}(r_q,r_m)|\Psi(r_q,r_m)\rangle \quad (3)$$

where the subscript zero indicates the gas phase wave function or Hamiltonian. When the quantum subsystem interacts with the rest of the system, then the Hamiltonian includes the corresponding interaction term:

$$\hat{H}(r_q,r_m) = \hat{H}^0(r_q) + \hat{V}_{int}(r_q,r_m) \quad (4)$$

This last term is usually expressed as the sum of electrostatic and van der Waals contributions. For the first one, the point charge model is normally preferred because it can be easily implemented in standard quantum codes. The van der Waals term is usually obtained as a sum of center to center contributions, and it is not included in the SCF solution of the QM subsystem.

$$E_{QM}(r_q) + E_{QM/MM}(r_q,r_m) = \langle \Psi(r_q,r_m)|\hat{H}^0(r_q)|\Psi(r_q,r_m)\rangle + \langle \Psi(r_q,r_m)|\hat{V}_{ele}(r_q,r_m)|\Psi(r_q,r_m)\rangle + V_{vdw}(r_q,r_m) = E_{QM}(r_q) + E^{ele}_{QM/MM}(r_q,r_m) + E^{vdw}_{QM/MM}(r_q,r_m) \quad (5)$$

A typical QM/MM PMF calculation requires the evaluation of the quantum energy and its derivatives about 10^5-10^6 times. Obviously, a quantum low-level (LL) method must be selected except in those cases where the number of atoms in the quantum subsystem is very small. This means that the potential energy surface used for the simulation can contain large errors. The problem can be dramatic if these errors appear associated to the geometrical coordinates showing large changes during the chemical process in which we are interested. This is the case when our low-level method is not able to correctly describe the energetic of a bond-breaking or a bond-forming process. The PMF obtained as a function of the bond-breaking or bond-forming distances (or a combination) would be clearly inaccurate. A way to improve the quality of the results is to obtain a corrected energy surface (E_{corr}) including an energy correction term (ΔE) as a function of the distinguished reaction coordinate ($\zeta = \zeta(r_q)$) used in the PMF:

$$E_{corr}(r_q,r_m) = E^{LL}(r_q,r_m) + \Delta E(\zeta) \quad (6)$$

This correction term is obtained as the difference between the energy provided by the low-level method (LL) and a high-level one (HL) for a particular configuration of the system (r_q, r_m) obtained along the chosen reaction coordinate. This correction

term can be calculated, in principle, from the unperturbed (eq 2) or the perturbed (eq 3) wave function of the quantum subsystem:

$$\Delta E^{U}(\zeta) = E^{HL}_{QM}(\zeta;r_q) - E^{LL}_{QM}(\zeta;r_q) \quad (7)$$

$$\Delta E^{P}(\zeta) = (E^{HL}_{QM}(\zeta;r_q) + E^{HL}_{QM/MM}(\zeta;r_q,r_m)) - (E^{LL}_{QM}(\zeta;r_q) + E^{LL}_{QM/MM}(\zeta;r_q,r_m)) = (E^{HL}_{QM}(\zeta;r_q) + E^{ele,HL}_{QM/MM}(\zeta;r_q,r_m)) - (E^{LL}_{QM}(\zeta;r_q) + E^{ele,LL}_{QM/MM}(\zeta;r_q,r_m)) \quad (8)$$

The set of configurations (r_q, r_m) used to obtain the correction term should be as representative as possible of the configurations to be explored during the simulations. A natural choice is to select the configurations from the reaction path of the reaction under study obtained in the presence of the environment. The first step is then to locate and characterize a possible transition structure. For this purpose we use the GRACE program[31,32] in which a Hessian-guided search is carried out for a particular coordinates subspace (the control space) whereas the rest (the complementary space) is minimized through the gradients. From this transition structure the intrinsic reaction coordinate (IRC) paths are traced down to reactant and product valleys, by means of a routine based on the method of Gordon et al.[33] Several structures are picked up from this IRC to calculate the energy difference between the high-level and the low-level methods.

$$\Delta E^{U}_{IRC}(\zeta) = E^{HL}_{QM}(\zeta;r_q) - E^{LL}_{QM}(\zeta;r_q) \qquad r_q \in \mathrm{IRC} \quad (9)$$

$$\Delta E^{P}_{IRC}(\zeta) = (E^{HL}_{QM}(\zeta;r_q) + E^{HL}_{QM/MM}(\zeta;r_q,r_m)) - (E^{LL}_{QM}(\zeta;r_q) + E^{LL}_{QM/MM}(\zeta;r_q,r_m)) \qquad r_q, r_m \in \mathrm{IRC} \quad (10)$$

In this way we obtain the value of the correction term (using either the unperturbed or the perturbed scheme) for a set of values of ζ. Then, following the work of Truhlar et al.,[28−30] a spline under tension[34] is used to interpolate this correction term at any value of ζ. In this way we obtain a continuous function in ζ, with continuous first and second derivatives, which are necessary to perform molecular dynamics simulations. However, before the spline fit is carried out we need to change our variable. Effectively, for an arbitrary reaction, ζ could be defined in the interval $[-\infty, +\infty]$ (for example, in the case of a bimolecular reaction where in the reactant and product state the reacting fragments are at infinite distance). A change of variable is made, defining a new one, z, as

$$z = \frac{2}{\pi}\arctan\left(\frac{\zeta - \zeta_0}{L}\right) \quad (11)$$

The new variable allows us to map the energy correction term (ΔE^{X}_{IRC}) onto the finite interval $[-1, +1]$, transforming in this way an extrapolation problem into an interpolation one. The constants ζ_0 and L are chosen for centering and scaling the mapping function in the range of interest. A detailed discussion about the proper choice of these two constants can be found in ref 28 but, in many cases, ζ_0 can be given the value of the distinguished reaction coordinate in the transition structure and L one-fourth of the range of ζvalues to be mapped. This choice usually provides the largest variation of the mapping function in the interval of ζ relevant for the process under study.

The final potential energy surface employed in the simulation would then be

$$E_{corr}(r_q,r_m) = E^{LL}(r_q,r_m) + \mathrm{spline}[\Delta E^{X}_{IRC}(\zeta),z] \quad (12)$$

This procedure can be easily implemented in any simulation package. At this stage the proposed correction scheme can be used if only energies are required, as for example, in Monte Carlo simulation methods. To run molecular dynamics, we need to add the force term arising from this correction to the energy. This force can be easily calculated using the chain rule, and then the forces on the corrected energy surface (F_{corr}) are obtained as

$$F_{corr,x_i} = -\frac{\partial E_{corr}(r_q,r_q)}{\partial x_i} = -\frac{\partial E^{LL}(r_q,r_q)}{\partial x_i} - \frac{\partial \mathrm{spline}[\Delta E^{X}_{IRC}(\zeta),z]}{\partial z}\frac{\partial z}{\partial \zeta}\frac{\partial \zeta}{\partial x_i} \quad (13)$$

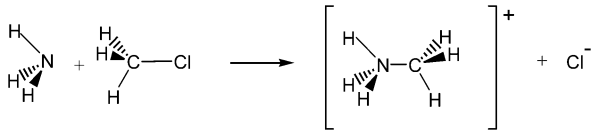
PMFs generated on the corrected surface can be denoted as PMF with interpolated corrections (PMF-IC). Depending on the choice of the energy term used to calculate the correction, we can distinguish between unperturbed (eq 7) and perturbed (eq 8) interpolated corrections (PMF-UIC and PMF-PIC).

3. Results and Discussion

In this section we present the results of the application of interpolated corrections to a reaction in aqueous solution, the Menshutkin reaction between methyl chloride and ammonia, and an enzymatic reaction, the chorismate rearrangement into prephenate catalyzed by BsCM. We selected these two processes because they have been thoroughly studied by different methodologies and experimental data are avalaible.

The Menshutkin Reaction in Water. The Menshutkin reaction is a S_N2 process in which neutral reactants yield charged products. Solvent effects play a decisive role on the process stabilizing the charge separation and then affecting both the activation and the reaction free energies. Historically, this was the first reaction for which solvent effects on rate constants were systematically gathered and compared.[35] We have selected the simplest example of this kind of reaction, as shown in Scheme 1.

SCHEME 1



This reaction has been thoroughly studied by means of continuum models[36−39] and also employing QM/MM techniques,[40−42] and then it provides a very good example to test the reliability of the proposed approach.

In our calculations the quantum subsystem (described using as low-level method the AM1 Hamiltonian[19]) is composed of methyl chloride plus ammonia. Lennard-Jones parameters of the QM subsystem were taken from ref 40. For water molecules we selected the TIP3P potential.[43] We then used a combination of GRACE[31,32] and CHARMM[44] programs to obtain a transition structure in aqueous solution imposing the collinearity of the nitrogen, carbon, and chlorine atoms. The control space, for which the Hessian matrix is obtained, includes only the coordinates of the QM subsystem. From this transition structure IRC was traced down to product and reactant valleys. We then

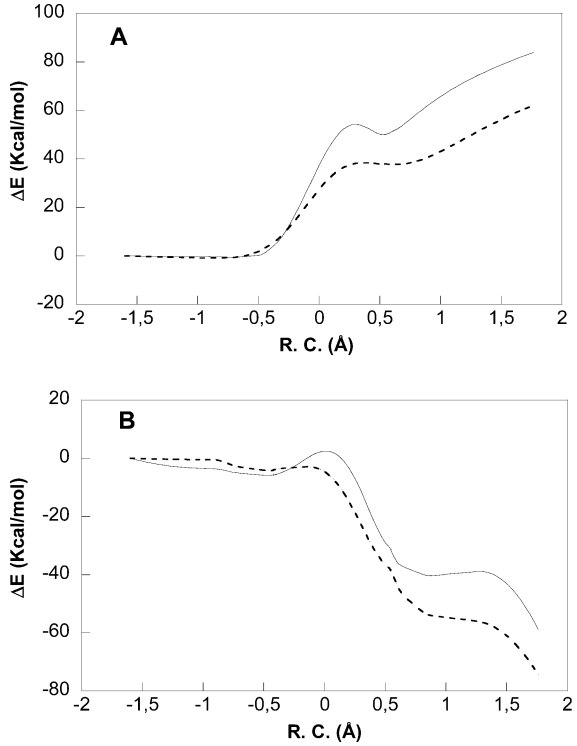


Figure 1. Relative values of E_{QM} (A) and $E_{QM} + E^{ele}_{QM/MM}$(B) for the Menshutkin reaction in water calculated at the AM1 (normal line) and MP2(fc)/6-31+G** (dashed line) levels, as a function of the reaction coordinate. The difference between both curves is the correction energy term included in the PMF-IC simulations.

selected 47 structures homogeneously distributed along the IRC, including the transition structure and the infinitely separated reactant and products molecules. For these structures we calculated E_{QM} and $E^{ele}_{QM/MM}$ at the MP2(fc)/6-31+G** level using the Gaussian98 package.[45] We obtained the correction energy term using the unperturbed and perturbed scheme (eqs 7 and 8) as a function of a distinguished reaction coordinate, in this case the antisymmetric combination of the bond-forming (N−C) and bond-breaking (C−Cl) distances:

$$\zeta = r_{CCl} - r_{CN}$$

Figure 1 shows the relative values of E_{QM} (A) and $E_{QM} + E^{ele}_{QM/MM}$ (B), calculated at the AM1 and MP2 levels, as a function of the reaction coordinate. The van der Waals interaction energy, which is the same at both computational levels, is omitted. It is interesting to note that the AM1 and MP2 curves diverge as the reaction coordinate advances toward the product state, both in Figure 1A,B. This is due in part to the large error in the AM1 value of the electron affinity of the chlorine atom (the experimental electron affinity of chlorine is 83.4 kcal mol^{-1}, whereas the calculated value at the AM1 level is 66.7 kcal mol^{-1}). Thus, the energy correction term is very important when the reaction is advanced and the correction will have a noticeable effect on both the activation and reaction free energies.

Corrected and uncorrected QM/MM PMFs (see Figure 2) were obtained using the DYNAMO library.[46] For this purpose the reacting system was placed in a simulation box of 32.0 Å with 1030 TIP3P water molecules. A switched cutoff radius of 15.5 Å was employed for all kind of interactions. Molecular dynamic simulations with periodic boundary conditions were performed at 300 K using the NVT ensemble with a time step of 1 fs. The umbrella sampling technique was used to compute the PMF following the distinguished reaction coordinate ζ. The

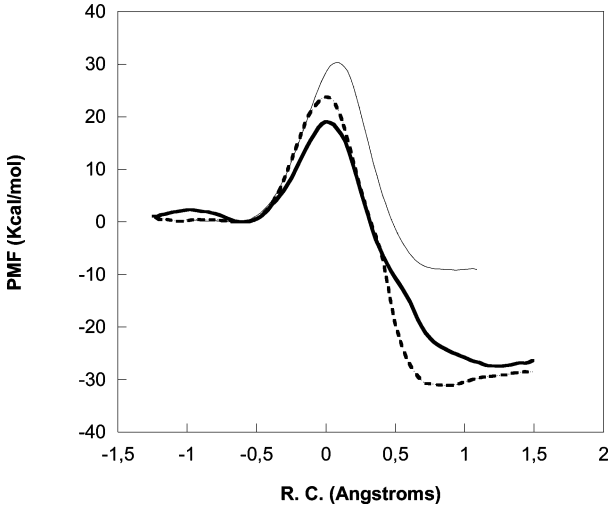


Figure 2. PMFs for the Menshutkin reaction in water obtained at the AM1/TIP3P level (normal line) and with corrections obtained with the unperturbed scheme (PMF-UIC, dashed line) and with the perturbed scheme (PMF-PIC, bold line).

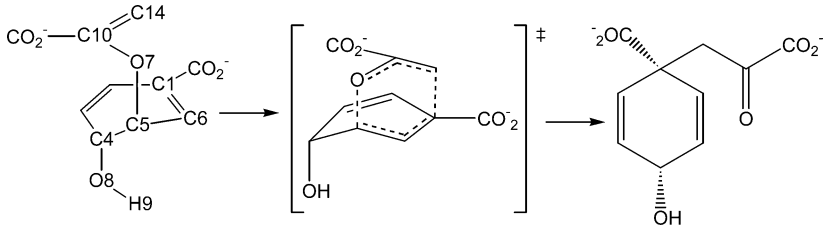
TABLE 1: Activation and Reaction Free Energies (kcal mol^{-1}) Obtained from the Uncorrected and Corrected PMFs and Experimental Values of the Menshutkin Reaction in Water and Bond Distances (Å) and Mulliken Charges (au) of the Transition State Averaged from the Simulation Window Located at the PMF Maximum

	PMF-AM1	PMF-UIC	PMF-PIC	exp
$\Delta G^{\ddagger}$	29.3	23.8	19.1	23.5[a]
ΔG_r	−10.4	−31.1	−27.4	−34[b]
$\langle d_{CN}\rangle_{TS}$	2.10	2.24	2.22	
$\langle d_{CCl}\rangle_{TS}$	2.10	1.99	2.01	
$\langle Q_{Cl}\rangle_{TS}$	−0.62	−0.43	−0.48	

[a] Reference 48. [b] Reference 41.

value of the force constant used for the umbrella sampling [(2.5−5.0) × 10^3 kJ mol^{-1} Å^{-2}] was determined to allow a full overlapping of the different windows traced in the PMF evaluation. The probability distributions, obtained from each individual window, were put together by means of the weighted histogram analysis method (WHAM).[17] The length of each window was 10 ps after equilibration, and the total number of windows was 81. To interpolate the energy correction term during the simulations, the mapping coordinate z was defined according to eq 11 using $\zeta_0 = -0.100$ and $L = 0.5$, determined from a previous uncorrected PMF. The activation and reaction free energies appear in Table 1, together with the experimental values. We calculated the activation free energy directly from the PMF difference between the transition and the reactant state. For a more detailed discussion about the relationship between activation free energies and PMFs, see ref 47. Some averaged properties of the transition state are also given. The corrected PMFs provide a much better agreement with the experimental activation[48] and reaction[41] free energies. The product state is largely stabilized when the correction energy term is included and the energy barrier is reduced. As a consequence the transition state is now much more advanced from both the geometrical and the electronic points of view. At the AM1 level the averaged Mulliken charge on the chlorine atom in the transition state is −0.62, which means that the transition state presents a large charge separation. Using the interpolated corrections, the averaged Mulliken charge on the chlorine atom is severely reduced, in absolute value, indicating a drastic change in the electronic nature of the transition state. When the perturbed and the unperturbed correction schemes are compared,

SCHEME 2



it seems that the latter gives a better agreement with the experimental free energies. However, it should be considered that the experimental activation free energy is taken from the Menshutkin reaction with iodine instead of chlorine[48] and then the agreement can be partly fortuitous. With respect to the product state, when the unperturbed correction scheme is used, a stable ion pair, with a dissociation free energy of about 2.6 kcal mol^{-1}, is obtained. For the perturbed correction scheme, a shallow free energy minimum is obtained for the reactants ion−dipole complex. These differences can be due to the fact that the corrections to the interaction energy are obtained for minimized structures. In such a case the interactions are expected to be enhanced with respect to the structures appearing during a simulation at 300 K.

It must be also considered that the correction scheme could be sensitive to the particular IRC followed. However, this effect can be expected to be small for corrections as large as those considered in this example (more than 20 kcal mol^{-1}). To estimate the effect of other configurations on the correction energy, we have calculated this quantity using molecular dynamics simulations carried out at different values of the reaction coordinate. At each position of the reaction path we selected several structures to obtain the correction energy. The standard deviations of the calculated correction term, due to fluctuations in coordinates of the system different from the reaction coordinate, were about 1.5 kcal mol^{-1}. A possible way to overcome the dependence of the correction energy on the particular configuration of the system selected at a particular value of the reaction coordinate is the averaging over different IRCs, as recently made by Alhambra et al. in the calculation of transmission coefficients.[49] Another possible strategy is the parametrization of the correction energy also as a function of other solute or solvent coordinates. For the case of the solvent a collective coordinate, such as the electrostatic potential on the solute, could be used. Method with such a spirit has been used by Cho and co-workers to obtain instantaneous frequency and frequency autocorrelation functions for small solutes in the solution.[50]

Claisen Rearrangement of Chorismate Catalyzed by BsCM. The conversion of (−)-chorismate to prephenate is a key step in the shikimate pathway for biosynthesis of amino acids in bacteria, fungi, and higher plants.[51] In addition, this is a rare example of a Claisen rearrangement catalyzed by enzymes, such as *Bacillus subtilis* chorismate mutase (BsCM). This reaction has been the subject of extensive experimental[52−55] and theoretical studies.[56−61] From the computational point of view, one of the most interesting features of this process is the fact that the reaction only involves one substrate without direct participation of the enzyme. Thus, this is an ideal system to be described using hybrid QM/MM methodologies, as the definition of the subsystems does not require cutting any covalent bond. The reaction takes place in a single chemical step in which a carbon−oxygen bond is broken and a new carbon−carbon bond is formed (see Scheme 2). Table 2 gives some gas phase values of the reactant and transition structures at the AM1[19] level (the low level) and at B3LYP/6-31G* (high level).[62] As can be seen, the process is concerted but asynchronous, the carbon−oxygen bond breaking being more advanced than the carbon−carbon bond forming. When the AM1 and B3LYP results are compared, it can be observed that the AM1 Hamiltonian provides too short distances for both bonds and an energy barrier clearly overestimated. At this point, it is worth mentioning that the use of single point energy corrections, as we propose here, can lead to an overestimation of the high-level energy barrier if the low-level and high-level paths differ noticeably.[63]

TABLE 2: Bond Distances (Å) of Transition (TS) and Reactant (RS) Structures and Activation and Reaction Energies (kcal mol^{-1}) for Chorismate Rearrangement in Gas Phase at the AM1 and B3LYP/6-31G* Levels

	B3LYP/6-31G*	AM1
$(d_{CO})_{RS}$	1.383	1.429
$(d_{CC})_{RS}$	4.404	4.124
$(d_{CO})_{TS}$	2.275	1.854
$(d_{CC})_{TS}$	2.719	2.164
$\Delta E^{\ddagger}$	44.1	52.7
$\Delta G^{\ddagger}$	41.1	52.1
ΔG_r	2.7	−7.9

As in the preceding example, we used a combination of GRACE[31,32] and CHARMM[44] programs to obtain a transition structure in the enzyme. In these calculations only the chorismate molecule is included in the QM subsystem (24 atoms) and described using the AM1 Hamiltonian. Again, only the coordinates of the QM atoms are used to obtain the Hessian matrix, whereas for the rest of the system only gradients are required. From the obtained transition structure, IRC was traced down to products and reactants valleys. We then selected 52 structures homogeneously distributed along the IRC and for these structures we calculated E_{QM} and $E^{ele}_{QM/MM}$ at the B3LYP/6-31G* and MP2/6-31G* levels using the Gaussian98 package.[45] These energies were then tabulated as a function of a distinguished reaction coordinate, here defined as the antisymmetric combination of the bond-forming (C1−C14) and bond-breaking (C5−O7) distances:

$$\zeta = r_{CO} - r_{CC}$$

Figure 3 shows the relative values of E_{QM} (Figure 3A) and $E_{QM} + E^{ele}_{QM/MM}$ (Figure 3B), calculated at the AM1 and B3LYP levels, as a function of this reaction coordinate. It can be clearly shown from these figures that in both the unperturbed and perturbed correction schemes, the maximum value of the correction energy term would be attained at the transition structure. Thus we can expect an important change in the activation free energies (improving the agreement with experiment) when the correction energy term is included in the simulations. Important corrections are also obtained for the products region, and AM1 and B3LYP energies agree quite well in the reactants region.

To interpolate the energy correction term during the simulations, we used a mapping coordinate z defined according to eq

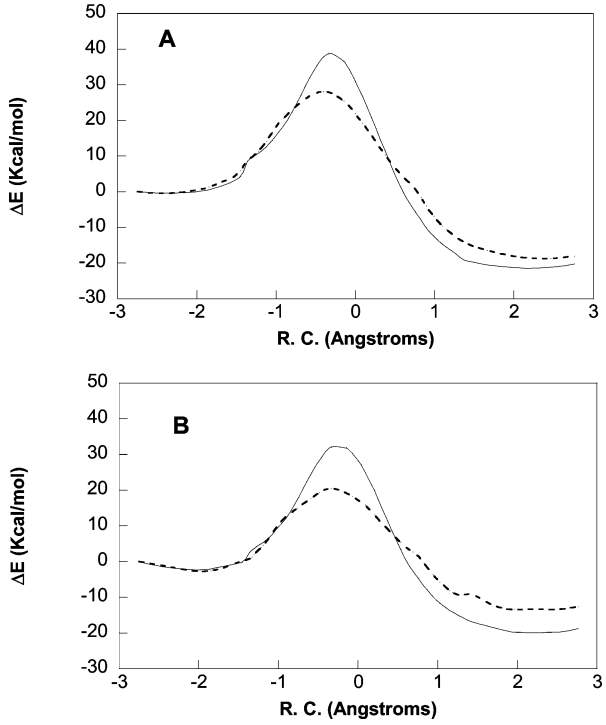


Figure 3. Relative values of E_{QM} (A) and $E_{QM} + E^{ele}_{QM/MM}$(B) for the chorismate rearrangement in BsCM calculated at the AM1 (normal line) and B3LYP/6-31G* (dashed line) levels, as a function of the reaction coordinate. The difference between both curves is the correction energy term included in the PMF-IC simulations.

TABLE 3: Activation Free Energies (kcal mol^{-1}) Obtained from the Uncorrected and Corrected PMFs of Chorismate Rearrangement in BsCM and Experimental Values and Bond Distances (Å) of the Reactant Transition States Averaged from the Simulations Ran at the PMF Maximum

	AM1	B3LYP/6-31G*		MP2/6-31G*		
	PMF	PMF-UIC	PMF-PIC	PMF-UIC	PMF-PIC	exp
$\Delta G^{\ddagger}$	29.3	19.6	19.8	12.7	14.3	15.4[a]
$\langle d_{CC}\rangle_{RS}$	3.05	3.25	3.20	3.10	3.14	
$\langle d_{CO}\rangle_{RS}$	1.46	1.45	1.45	1.45	1.45	
$\langle d_{CC}\rangle_{TS}$	2.18	2.18	2.19	2.16	2.16	
$\langle d_{CO}\rangle_{TS}$	1.94	1.79	1.80	1.70	1.69	

[a] Reference 54.

11 using $\zeta_0 = 0.008$ and $L = 1.38$. Corrected and uncorrected QM/MM PMFs were then obtained using the DYNAMO library.[46] For this purpose the reacting system was placed in a simulation box of ~56 Å with the enzyme and 3835 TIP3P water molecules. Molecular dynamic simulations with periodic boundary conditions were performed at 300 K using the NVT ensemble with a time step of 1 fs and a cutoff radius of 13.5 Å. The umbrella sampling technique was used to compute the PMF following the distinguished reaction coordinate ζ defined above. The value of the force constant used for the umbrella sampling was 2.5×10^3 kJ mol^{-1} Å^{-2}. The length of each window was of 20 ps after equilibration with a total number of windows of 61. The probability distributions obtained from each individual window, are put together by means of the weighted histogram analysis method (WHAM).[17]

The activation free energies for the uncorrected and corrected profiles, both using B3LYP and MP2 as high levels, are given in Table 3, together with the experimental value. Selected averaged geometrical properties of the transition and reactant states are also provided. The activation free energy is considerably reduced when the correction terms are included in the simulation, improving the agreement with the experimental value. The averaged geometries of the reactant and transition states are also different. The reactant state in the enzyme is found at larger carbon−carbon distances when the PMF is corrected, just reflecting the tendencies observed for the gas phase geometries (the B3LYP gas phase reactant structure presents a larger carbon−carbon distance than the AM1 one). For the transition state, the main difference between the gas phase structures is a lengthening of both the C−C and the C−O distances in the high level with respect to the AM1 one (see Table 2). This means a geometrical correction along the symmetric combination of distances that is not included in our correction scheme. In our method we can get a reasonable energy but the geometry does not change in the expected direction. In fact, the carbon−oxygen distance is shortened when the PMF is corrected either at the B3LYP or at the MP2 levels. A unidimensional correction scheme is not flexible enough to get reasonable energies and structures when the differences between the high and low levels appear also along other coordinates. Obviously, the method only incorporates corrections for the distinguished coordinate, here the antisymmetric combination of the bond-forming and bond-breaking distances. Thus, this is a good example to show some of the limitations of the interpolated correction scheme here proposed. The use of bidimensional correction schemes (using in this case the C−C and the C−O distances) may be a good solution for such cases.

4. Conclusions

We have presented a correction scheme to obtain improved potentials of mean force in QM/MM studies of chemical processes in condensed media based on dual-level calculations of the QM subsystem. The proposal is an extension of the interpolated corrections methodology developed by Truhlar et al.[28−30] for dynamical calculations in gas phase in the framework of the variational transition state theory.

In the proposed method a low-level potential energy surface is corrected by adding an energy term. This term is obtained as the single-point difference with a high-level calculation and expressed as a function of the distinguished reaction coordinate to be biased in the calculation of the PMF. To be representative, the structures selected to carry out the difference between the two levels are taken from the IRC path in the condensed media. For this purpose the GRACE algorithm is employed to define a reduced Hessian matrix to guide the reaction path calculation. Two different correction schemes have been proposed using either the unpolarized (gas phase) wave function or the polarized one. By means of a mapping coordinate and the use of splines under tension, the energy correction and its derivatives can be readily included in any simulation package.

We have presented the application of the method to two paradigmatic reactions in condensed media: the Menshutkin reaction in water and the chorismate rearrangement catalyzed by BsCM. These two reactions have been thoroughly studied from the experimental and theoretical points of view and then are good examples to test the viability of the methodology. In both cases, the use of interpolated corrections leads to a drastic improvement of the data obtained in the simulations when compared to the PMFs on the basis of the use of a semiempirical Hamiltonian. Activation and reaction free energies compare much better to the experimental values. It must be finally pointed out that the computational cost of including the interpolated corrections is negligible when compared to the total computational cost of the PMF.

Our results are not conclusive with respect to the preference of the perturbed or unperturbed correction schemes. This

question, together with the use of corrections averaged over different reaction paths, bidimensional interpolation schemes, and also the consideration of geometry optimizations at the higher level will be the subject of future works.

Acknowledgment. We are indebted to DGI for project DGI BQU2003-4168, BANCAIXA for project P1A99-03 and Generalitat Valenciana for project GV01-324, which supported this research, and the Servei d'Informatica of the Universitat de València and Universitat Jaume I for providing us with computer capabilities. S.M. thanks UJI-BANCAIXA Foundation for a Postdoctoral fellowship. J.J.R.-P. thanks the Ministerio de Educación (Spain) for a FPU fellowship. We are grateful to D. G. Truhlar, J. M. Lluch, A. González-Safont, and M. García-Viloca for helpful comments and suggestions.

References and Notes

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(2) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *6*, 700−733.
(3) Théry, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269−282.
(4) Gao, J. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(5) Tuñón, I.; Millot, C.; Martins-Costa, M. T. C.; Ruiz-López, M. F. *J. Chem. Phys.* **1997**, *106*, 3633−3642.
(6) Field, M. J.; Albe, M.; Bret, C.; Proust-de Martin, F.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088−1100.
(7) Kollman, P. A.; Kuhn, B.; Donini, O.; Peräkylä, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72−79.
(8) Villà, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887−7907.
(9) Mullholland, A. J. *Theor. Comput. Chem.* **2001**, *9*, 597−653.
(10) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.
(11) Field, M. J. *J. Comput. Chem.* **2002**, *23*, 48−58.
(12) Rivail, J. L.; Rinaldi, D.; Ruiz-López, M. F. In *Theoretical and Computational Models for Organic Chemistry*; Formosinho, S. J., Arnaut, L., Csizmadia, I., Eds.; Kluwer: Dodrecht, The Netherlands, 1991.
(13) Tomasi, J.; Persico, M. *Chem. Rev.* **1994**, *94*, 2027.
(14) Cramer, C. J.; Thrular, D. G. In *Reviews in Computacional Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1995; Vol. 6, pp 1−72.
(15) Kollman, P. A.; Kuhn, B.; Donini, O.; Peräkylä, M.; Stanton, R.; Bakowies, D.; *Acc. Chem. Res.* **2001**, *34*, 72.
(16) Roux, B. *Comput. Phys. Commun.* **1995**, *91*, 275−282.
(17) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(18) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, 129−135.
(19) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(20) Stewart, J. J. P. *J. Comput. Aided Mol. Des.* **1990**, *4*, 1−105.
(21) Warshel, A.; Weiss, R. M. *J. Am. Chem. Soc.* **1980**, *102*, 6218−6226.
(22) Warshel, A. In *Computer modelling of chemical reactions in enzymes and solutions*; John Wiley & Sons: New York, 1991.
(23) González-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618−4627.
(24) Lau, E. Y.; Kahn, K.; Bash, P. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9937−9942.
(25) Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2002**, *106*, 1768−1798.
(26) (a) Alhambra, C.; Corchado, J. C.; Sanchez, M. L.; Gao, J.; Truhlar, D. G. *J. Am. Chem. Soc.* **2000**, *122*, 8197−8203 (b) Devi-Kesavan, L. A.; García-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2003**, *109*, 133−139.
(27) Strajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333−13343.
(28) Corchado, J. C.; Coitiño, E. L.; Chuang, Y.; Fast, P. L.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 2424−2438.
(29) Chuang, Y.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 1140−1149.
(30) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. *J. Chem. Phys.* **1995**, *103*, 5522.
(31) Moliner, V.; Turner, A. J.; Williams, I. H. *J. Chem. Soc. Chem. Commun.* **1997**, 1271−1272.
(32) Turner, A. J.; Moliner, V.; Williams, I. H. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323−1331.
(33) Schmidt, M. W., Gordon, M. S.; Dupuis, M. *J. Am. Chem. Soc.* **1985**, *107*, 2585−2589.
(34) (a) Renka, R. J. SIAM *J. Stat. Comput.* **1987**, *8*, 393. (b) Renka, R. J. *ACM Trans. Math. Software* **1993**, *19*, 81.
(35) Menshutkin, N. *Z. Phys. Chem.* **1890**, *5*, 589.
(36) Solà, M.; Lledos, A.; Duran, M.; Bertran, J.; Abboud, J. M. *J. Am. Chem. Soc.* **1991**, *113*, 2873.
(37) Chuang, Y.-Y.; Cramer, C. J.; Truhlar, D. G. *Int. J. Quantum Chem.* **1998**, *70*, 887−896.
(38) Castejon, H.; Wiberg, K. B. *J. Am. Chem. Soc.* **1999**, *121*, 2139−2146.
(39) Dillet, V.; Rinaldi, D.; Bertrán, J.; Rivail, J. L. *J. Chem. Phys.* **1996**, *104*, 9437−9444.
(40) Gao, J.; Xia, X. *J. Am. Chem. Soc.* **1993**, *115*, 9667−9675.
(41) Gao, J. *J. Am. Chem. Soc.* **1991**, *113*, 7796
(42) Naka, K.; Sato, H.; Morita, A.; Hirata, F.; Kato, S. *Theor. Chem. Acc.* **1999**, *102*, 165−169.
(43) Jorgensen, W. L.; Chandrasekhar, J.; Madura J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(44) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A.6; Gaussian, Inc.: Pittsburgh, PA, 1998.
(46) Field, M. J. *A practical Introduction to the Simulation of Molecular Systems*; Cambridge University Press: Cambridge, U.K., 1999.
(47) Schenter, G. K.; Garrett, B. C.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 5828−5833.
(48) Okamoto, K.; Fukui, S.; Shingu, H. *Bull. Chem. Soc. Jpn.* **1967**, *40*, 1920. (b) Okamoto, K.; Fukui, S.; Nitta, I.; Shingu, H. *Bull. Chem. Soc. Jpn.* **1967**, *40*, 2354.
(49) Alhambra, C.; Corchado, J.; Sánchez, M. A.; García-Viloca, Truhlar, D. G.; M.; Gao, J. *J. Comput. Chem.* **2003**, *24*, 177−190.
(50) Stratt, R. M.; Cho, M. *J. Chem. Phys.* **1994**, *100*, 6700−6706.
(51) Haslam, E. *Shikimic Acid: Metabolism and Metabolites*; John Wiley & Sons: New York, 1993.
(52) Chook, Y. M.; Ke, H.; Lipscomb, W. N. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8600−8603.
(53) Cload, S. T.; Liu, D. R.; Pastor, R. M.; Schultz, P. G. *J. Am. Chem. Soc.* **1996**, *118*, 1787−1788.
(54) Kast, P.; Asif-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691−2694.
(55) Mandal, A.; Hilvert, D. *J. Am. Chem. Soc.* **2003**, *125*, 5598−5599.
(56) Lyne, P. D.; Mulholland, A. J.; Richards, W. G. *J. Am. Chem. Soc.* **1995**, *117*, 11345−11350.
(57) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J.; Field, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 1709−1712.
(58) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Chem. Eur. J.* **2003**, *9*, 984−991.
(59) Hur, S.; Bruice, T. C. *J. Am. Chem. Soc.* **2003**, *125*, 1472−1473.
(60) Strajbl, M.; Shurki, A.; Kato, M.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 10228−10237.
(61) Ranaghan, K. E.; Ridder, L.; Szefczyk, B.; Sokalski, W. A.; Hermann, J. C.; Mulholland, A. J. *Mol. Phys.* **2003**, *101*, 2696−2714.
(62) (a) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B.* **1988**, *37*, 785−789. (b) Becke, A. D. *Phys. Rev. A.* **1988**, *38*, 3098−3100.
(63) Hu, W.-P.; Liu, Y.-P.; Truhlar, D. G. *J. Chem. Faraday Trans.* **1994**, *90*, 1715−1725.