Hybrid Quantum Mechanics/Molecular Mechanics Simulations with Two-Dimensional Interpolated Corrections: Application to Enzymatic Processes

J. Javier Ruiz-Pernía,[†] Estanislao Silla,[†] Iñaki Tuñón,*[,†] and Sergio Martí[‡]

Departament de Química Física/IcMol, Universidad de Valencia, 46100 Burjasot, Valencia, Spain, and Departament de Ciències Experimentals, Universitat Jaume I, Box 224, 12080 Castellón, Spain

Received: June 7, 2006; In Final Form: July 10, 2006

Hybrid quantum mechanics/molecular mechanics (QM/MM) techniques are widely used to study chemical reactions in large systems. Because of the computational cost associated with the high dimensionality of these systems, the quantum description is usually restricted to low-level methods, such as semiempirical Hamiltonians. In some cases, the description obtained at this computational level is quite poor and corrections must be considered. We here propose a simple but efficient way to include higher-level corrections to be used in potential energy surface explorations and in the calculation of potentials of mean force. We evaluate a correction energy term as the difference between a high-level and a low-level calculation on the QM subsystem, employing either the polarized or the gas-phase wave function, obtained as a function of two geometrical coordinates relevant in the process considered. Through the use of two-dimensional bicubic splines this correction energy is included in the simulations, ensuring the continuity and derivability of the energy function. We have tested the proposed scheme with two prototypical examples: the chorismate to prephenate rearrangement catalyzed by *Bacillus subtilis* chorismate mutase and the catechol methylation catalyzed by catechol O-methyltransferase. In both cases the use of interpolated corrections clearly improves the energetic and geometric descriptions of the reaction.

1. Introduction

The study of chemical processes in condensed phases (such as enzymes, solid surfaces, or solutions) requires the development of specific methodological tools to deal with the special characteristics of the systems being considered: the huge number of electrons and degrees of freedom. A reasonable solution is provided by hybrid quantum mechanics/molecular mechanics (QM/MM) strategies,[1−17] which have become a very useful tool for studying large systems. In these methods, one part of the system is described by quantum mechanics while the rest of the system is described by molecular mechanics. The QM region is usually restricted to that part of the total system where bond breaking and forming are taking place, but inclusion of neighboring regions can be important to describe charge-transfer effects.

Although QM/MM methodologies are in extensive use for a diversity of problems, there are still some limitations that prevent their general applicability. To obtain relevant thermodynamic information, such as free energy changes, simulation techniques (molecular dynamics or Monte Carlo) require the calculation of the energy (and gradients) of the system thousands or even millions of times to sample adequately the relevant structures. Then, when a QM/MM method is used, the wave function must be evaluated a large number of times, which can be unaffordable if the QM subsystem contains more than a few atoms. Several approaches can be used to solve this problem. Roughly speaking, one may choose not to include the quantum subsystem flexibility in the simulation, or alternatively, one may use a low-level electronic description. The quantum mechanics free energy perturbation (QM-FEP) method[18,19] is an example of the first kind of computational strategies, where only the conformational changes taking place in the MM part are sampled, discarding the internal movements of the quantum subsystem. In this approach the reaction path is usually obtained for a gas-phase model of the active site, and then high-level quantum treatments are not computationally forbidden. Otherwise, if the QM region is going to be sampled during the simulations (as made when obtaining potentials of mean force associated with a particular distinguished reaction coordinate), then low-level QM treatments are required to keep the computational effort within reasonable limits. These include semiempirical Hamiltonians,[20−22] empirical valence bond,[23] or density functional tight binding methods.[24] The price to be paid is, in general, the loss of accuracy of the energy function employed. Normally, extensive testing is required to compare the performance of the low-level method versus high-level descriptions for gas-phase model reactions.

Some techniques have been developed to improve the quality of the low-level descriptions of the QM part in QM/MM calculations without increasing the computational cost too much. One obvious solution is to develop a new semiempirical Hamiltonian with parametrization optimized to describe the process under study. This strategy is usually known as specific reaction parameters (SRPs), and it has been adopted in a number of cases.[14,25,26] However, it should be taken into account that different parametrizations could be required to treat different chemical steps or even to improve the activation or the reaction energy. Another possibility is to include correction terms to the potential energy surface based on valence-bond theory.[27,28] We recently developed a correction scheme specifically designed for the calculation of improved potentials of mean force (PMFs).[29] This proposal is an extension of the interpolated

* Author to whom correspondence should be addressed. E-mail: tunon@uv.es.
[†] Universidad de Valencia.
[‡] Universitat Jaume I.

10.1021/jp063520a CCC: $33.50
Published on Web 08/12/2006

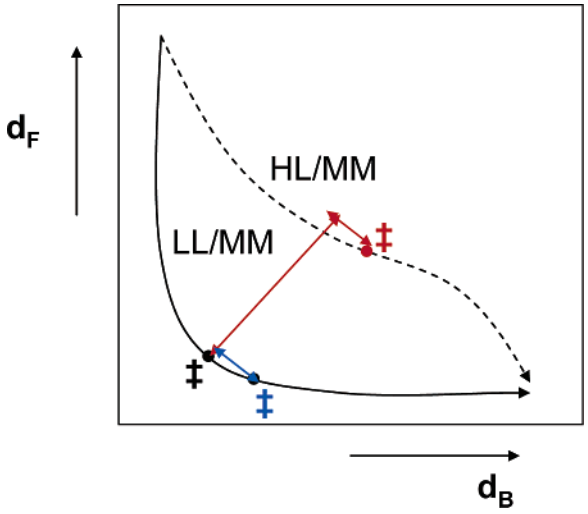


Figure 1. Qualitative illustration of the minimum energy paths and transition structures on the potential energy surface obtained as a function of the bond-breaking (d_B) and bond-forming (d_F) distances using a high-level (HL) or a low-level (LL) description of the QM subsystem. The HL transition structure (in red) is displaced with respect to the low-level one (in black) essentially in a direction orthogonal to the antisymmetric combination $d_B - d_F$. A possible transition structure (in blue) corresponding to the use of one-dimensional corrections along the distinguished reaction coordinate is also shown.

corrections methodology developed by Truhlar et al.[30−32] for gas-phase dynamical calculations. In this method, the energy difference between structures computed at the low level and a chosen high level is written as a function of the distinguished reaction coordinate used to follow the chemical transformation and conveniently interpolated through the use of cubic splines.[33] While this procedure has been shown to introduce a systematic improvement in the results at a very low additional computational cost, it still relies on the assumption that the low-level energy surface is qualitatively reasonable. Effectively, one of the most important limitations for this method is the dependence on the AM1/MM minimum energy path, which can sometimes dramatically differ from the one obtained using more reliable and costly methods. In other words, this simple one-dimensional correction scheme allows for displacements of the transition state along the reaction coordinate but not in other directions. An approach to introduce corrections beyond the one-dimensional strategy in the context of enzymatic reactions was already made by Field and co-workers.[34] In this paper, we present an extension of this methodology following our previously proposed scheme to obtain the correction term.[29] The correction energy is expressed now as a function of two geometrical coordinates relevant in the description of the chemical reaction. For example, if the process under study can be described using an antisymmetric combination of bond-forming and bond-breaking distances ($d_B - d_F$), then the correction energy is mapped on a two-dimensional potential energy surface (PES) obtained as a function of these distances (d_F;d_B). Figure 1 illustrates this situation, showing the minimum energy paths and transition structures corresponding to a QM/MM calculation using a low-level quantum treatment (LL/MM) or a high-level one (HL/MM). As it can be seen, in general, the transition structure described on the high-level surface can be found in an advanced or delayed position along the reaction coordinate ($d_B - d_F$) or along an orthogonal direction. These displacements can be especially important when the high-level transition state is considerably more associative or dissociative than the low-level one; this is when the low-level energy surface has important qualitative drawbacks. In principle, this correction scheme is obviously more general than our previous procedure, and it is not designed specifically for the calculation of PMFs, but it can also be employed for PES-based applications (localization of stationary points and calculation of minimum energy paths for example).

In this paper, we present a two-dimensional interpolated correction scheme and validate it exploring two well-known enzymatic reactions: the *S*-adenosyl-L-methionine (SAM)-dependent methylation of catecholate catalyzed by catecholate O-methyltransferase (COMT) and the chorismate to prephenate rearrangement, a [3,3]-sigmatropic process catalyzed by *Bacillus subtilis* chorismate mutase (BsCM). Both reactions have been the subject of numerous experimental and theoretical studies, some of them in our group, and thus they are excellent examples to be used in the calibration of new methodologies. As shown below, these two reactions display a very different behavior when including corrections to the potential energy surface: While in one of them the transition structure is orthogonally displaced with respect to the distinguished reaction, the other does not suffer major structural changes. In addition, both systems share a common feature making all the computational treatment easier: the absence of covalent bonds between the substrate and the enzyme, avoiding the use of specific techniques to describe the boundary between the QM and the MM subsystems.

2. Methodology

The QM/MM potential energy is 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 can be written as the sum of three terms[2]

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

The first and second contributions must be obtained from the solution of the corresponding Hamiltonians

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

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

where the superscript zero indicates the gas-phase wave function or Hamiltonian. The gas-phase Hamiltonian includes the usual electronic and nuclear terms, while the Hamiltonian describing the QM subsystem embedded in the MM subsystem 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 interaction term, in a general system where there is no covalent bond between QM and MM regions, can be just expressed as the sum of electrostatic and van der Waals contributions, the last one not included in the self-consistent field procedure as it does not usually involve electronic coordinates. If we consider both terms, then the sum of the QM and QM/MM energies can be written as

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

Normally, molecular simulations or energy optimizations of very

large system (containing around 10^4 atoms) require of a very large number of energy and energy gradient (and in some cases Hessians) evaluations (around 10^{5-6}) Obviously, a low-level (LL) quantum method is required except in those cases where the quantum subsystem comprises a very small number of atoms. The problem associated with LL methods is that the description of the chemical event can be quantitatively or even qualitatively incorrect. This can be especially important if errors are associated with those geometrical parameters displaying important variations during the chemical reaction under study. Consider the picture presented in Figure 1; the LL treatment has important differences with respect to the HL one in the descriptions of the bond-breaking and bond-forming processes. In this example, the introduction of a one-dimensional correction along a distinguished reaction coordinate $d_B - d_F$ would only slightly improve the description of the process as far as the topology of the PES is significantly modified and the HL transition structure appears considerably displaced along the symmetric combination $d_B + d_F$. A more general correction scheme can be developed by introducing an additional correction energy term obtained as a function of two geometrical parameters $\lambda_1(r_q)$ and $\lambda_2(r_q)$ that are a function of the Cartesian coordinates of the QM subsystem. (In our example a natural selection for $\lambda_1(r_q)$ and $\lambda_2(r_q)$ would be the distances of the breaking and forming bonds, but in general $\lambda_1(r_q)$ and $\lambda_2(r_q)$ are chosen to provide as much information as possible about the chemical process under study.) Then, a new PES is defined using the following energy function

$$E'(r_q,r_m) = E^{LL}_{QM}(r_q) + E^{ele,LL}_{QM/MM}(r_q,r_m) + E^{vdw}_{QM/MM}(r_q,r_m) + E_{MM}(r_m) + \Delta E_{corr}(\lambda_1,\lambda_2) \quad (6)$$

The correction term, $\Delta E_{corr}(\lambda_1,\lambda_2)$, is obtained as the difference between the energy provided by the LL method and a HL one for a particular configuration of the system (r_q,r_m) obtained during the LL/MM exploration of the PES. Thus, several structures are calculated at the LL/MM level, fixing our system at given values of the $\lambda_i(r_q)$ coordinates and minimizing the rest of the coordinates of the system (both the remaining coordinates of the QM subsystem and all the coordinates of the MM subsystem), providing in this way a global representation as rich as achievable of the chemical process. From our previous paper,[29] this correction term can be calculated using an unperturbed (eq 7) or a perturbed (eq 8) wave function scheme

$$\Delta E^{u}_{corr}(\lambda_1,\lambda_2) = E^{HL}_{QM}(\lambda_1,\lambda_2;r_q) - E^{LL}_{QM}(\lambda_1,\lambda_2;r_q) \quad (7)$$

$$\Delta E^{p}_{corr}(\lambda_1,\lambda_2) = (E^{HL}_{QM}(\lambda_1,\lambda_2;r_q) + E^{HL}_{QM/MM}(\lambda_1,\lambda_2;r_q,r_m)) - (E^{LL}_{QM}(\lambda_1,\lambda_2;r_q) + E^{LL}_{QM/MM}(\lambda_1,\lambda_2;r_q,r_m)) \quad (8)$$

In other words, in the unperturbed scheme only the gas-phase energy of the QM subsystem is corrected, while in the perturbed scheme the polarization and interaction energies are also corrected. It should be stressed that in both cases the configurations (r_q,r_m) used to compute the correction term are picked up from the full potential energy surface under study obtained in the presence of the environment.

At this point, the correction energy is a discontinuous set of correction energy values obtained for particular values of the coordinates $\lambda_1(r_q)$ and $\lambda_2(r_q)$. To ensure the continuity of the energy function, gradients, and force constants, a two-dimensional cubic spline[35] is used to interpolate this correction energy term at any value of $\lambda_1(r_q)$ and $\lambda_2(r_q)$. Although there are several ways to perform the two-dimensional interpolation,[35] we have chosen to make use of two consecutive cubic-spline-based interpolations, thus precomputing and storing only all the derivative information in just one of the directions: The cubic spline is evaluated in one direction (λ_1, for example), and then with these new values the cubic spline in the other direction (λ_2) is obtained.[35] In this way we obtain a continuous function in $\lambda_1(r_q)$ and $\lambda_2(r_q)$, with continuous first and second derivatives both in the interval and in the boundaries, which are necessary to perform molecular dynamics simulations and energy optimizations. In any case, it is convenient that the potential energy samplings were performed covering a wider extension of the potential energy surface than the final region that is going to be explored in subsequent calculations.

The final energy function employed is then

$$E'(r_q,r_m) = E^{LL}(r_q,r_m) + 2D\_spline\{\Delta E_{corr}(\lambda_1,\lambda_2)\} \quad (9)$$

It is important to stress here that sometimes the coordinates selected may be defined in the interval $[-\infty, +\infty]$. In such case it may be useful to avoid the extrapolation problem using a mapping variable defined in a finite interval. This was discussed in our previous work [29] and explained in detail in ref 32.

Finally, to run molecular dynamics we need to add the force term arising from the correction energy. In our implementation, we make use of the continuous and well-defined gradient expression provided by the two-dimensional spline interpolation algorithm, which must be projected into the coordinate's space depending on the nature of the $\lambda_i(r_q)$ terms. This can be easily made through the use of the corresponding Wilson's B matrix[36]

$$B_{ij} = \frac{\partial \lambda_i}{\partial r_{q,j}} \quad (10)$$

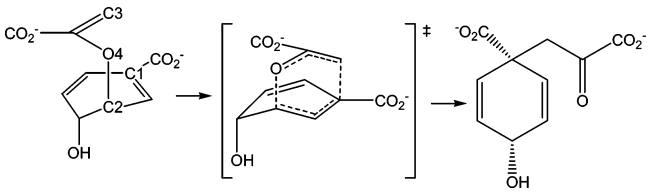
In the rest of the paper, results obtained using either the perturbed or the unperturbed interpolated correction schemes will be denoted as PIC or UIC, respectively. The prefix 1D or 2D is employed to distinguish between the present two-dimensional procedure and our previous one-dimensional one. Although it has not been explicitly considered here, the correction scheme can also be applied for those cases where the division between the QM and the MM subsystems implies a covalent bond, just considering the appropriate definition of the QM region and of the interaction energy to perform higher-level calculations.

3. Results

In this section we present the results of the application of the two-dimensional interpolated corrections for two prototypical systems: the chorismate rearrangement into prephenate catalyzed by BsCM and the S_N2 methylation activated by COMT.

Chorismate to Prephenate Rearrangement Catalysis. The conversion of chorismate into prephenate is a Claisen intramolecular process, which takes place through a chairlike aromatic transition state. During the rearrangement, the bond between atoms C2 and O4 is broken, and a new bond is formed between atoms C1 and C3 (Scheme 1). The reaction is one of the few samples of biologically catalyzed sigmatropic processes known and has been identified as part of the metabolic pathway of some organisms for obtaining aromatic amino acids. In particular, we study the reaction catalyzed by *Bacillus subtilis* chorismate mutase (BsCM). In addition to experimental data,[37] there are a large number of theoretical studies for this chemical process, either in gas phase or including different environments.[38-44]

SCHEME 1



The presence of such references makes the system ideal for benchmarking purposes of new computational methods.

Our computational model for describing the enzymatic system consists of the trimeric protein in which only one of the three possible active sites has been occupied by a chorismate molecule. The initial coordinates were taken from the Protein Data Bank (PDB code 1COM).[45] This enzyme−substrate complex was immersed in a cubic box of 55.8 Å on a side of water molecules, and all the overlapping waters were then removed (those with and oxygen lying within 2.8 Å of a non-hydrogen atom). Once solvated, those residues 17 Å far from the chorismate in the active site (∼13 500 atoms from a total of ca. 17 200) were selected to be kept frozen during the rest of the simulations. The quantum method selected for the chorismate substrate was the semiempirical AM1 Hamiltonian. For the rest of the system, the OPLS-AA molecular mechanics force field[46] was chosen, combined with the TIP3P potential for water molecules.[47] Periodical boundary conditions were applied for evaluating the nonbonding interactions with a switched cutoff decay from 8 to 12 Å. All of the low-level calculations (both LL/MM and gas-phase ones) have been performed using the DYNAMO[48] library, while the high-level correction term was computed by means of the Gaussian03[49] package of programs.

First at all, the chemical conversion process of the chorismate into prephenate was studied within the enzymatic environment to determine a valid transition structure and the corresponding minima using the GRACEFUL algorithm.[50] For this search, first- and second-order optimization algorithms were applied within the framework of the control/complementary spaces:[16] The Baker eigenvector following scheme was used to guide the initial search for the transition structure of the chemical rearrangement, using a Hessian matrix defined for the set of coordinates of the QM subsystem or control space. Each optimization step in the control space is carried out after a full relaxation of the environment (classical molecular mechanics protein and water molecules, the complementary space). Once a transition structure has been located and characterized, a steepest-descent energy path[51] is traced down toward the corresponding reactants and products. This procedure allows us to inspect which internal coordinates are closely coupled to the reaction process, to be used as distinguished coordinates for calculating the energy correction terms, and to set up the strategy for obtaining the potential of mean force (sampling reaction coordinate, range of values, number of simulation windows, etc.). For the chorismate rearrangement we selected the bond-breaking and bond-forming distances: d(C2,O4) as λ_1 and d(C1,C3) as λ_2. A potential energy surface is then obtained by imposition of restraints at different values of the λ_i coordinates and minimizing the rest of the system. For this particular problem, we varied λ_1 (the bond-breaking distance) in the range from 1.4 to 2.5 Å with an increment of coordinate of 0.1 Å and λ_2 (the forming bond distance) between 1.4 and 3.9 Å with the same step size. This means a total of 11 × 25 points on the AM1/MM potential energy surface. The corresponding geometries were selected for the computation of the correction energy term with both the unperturbed and the perturbed schemes using the B3LYP/6-31G(d) method as high-level, a theoretical method well-suited for describing this kind of sigmatropic processes.[38] Obviously, higher QM levels could be used, but the main purpose of this work is to show the potentiality and implementation of the procedure. The correction energy terms were then interpolated through the use two-dimensional bicubic splines and added to the AM1/MM energy surface to obtain corrected and continuous energy surfaces to be used in PMF calculations.

The resulting energy surfaces are shown in Figure 2. A quite important displacement of the transition structure toward larger values of the new CC bond distance and, to a lesser extent, toward larger values of the broken OC bond distance on both corrected surfaces can be observed (perturbed and unperturbed correction schemes). With this, we have a clear displacement of the transition structure orthogonal to the natural distinguished reaction coordinate ($\zeta = d_{CO} - d_{CC}$), and then any correction obtained exclusively as a function of this reaction coordinate ($\Delta E_{corr}(\zeta)$) cannot capture all the differences between the LL and HL energy surfaces. Otherwise, the perturbed and unperturbed schemes provide a very similar description.

To evaluate the PMFs, we performed molecular dynamic simulations using the DYNAMO library. In these simulations we employed the AM1/MM energy function adding the two-dimensional interpolated corrections obtained from the previous PES's. The umbrella-sampling technique[52] was applied to the ζ coordinate by means of a harmonic restraining potential. The value of the umbrella constant was 2500 kJ mol^{-1} Å^{-1}. The classical equations of movement have been integrated using a step size of 1 fs within the canonical ensemble (NVT), making use of a Langevin thermostat to keep the temperature at 300 K and the Verlet algorithm for updating the velocities. Approximated values for the activation free energies are shown in Table 1 together with the potential energy barriers derived from Figure 2. Note that these last values have not been averaged and then are provided only for qualitative purposes. Averaged values of the bond-breaking and bond-forming distances corresponding to the transition and reactant states are also provided. The activation free energy values have been obtained as PMF differences between the maxima and the reactant's minimum, without incorporating either the reaction coordinate motion in the reactant state or the fact that the selected internal coordinate to perform the PMF is curvilinear instead of rectilinear. These terms are usually negligible for this reaction, but further discussion about their magnitude can be found for example in ref 53. If we compare the geometric description obtained with our 2D correction schemes to the AM1/MM results, then we can observe that when the corrections are added the reaction is described as a significantly more dissociative process, with the CC bond formation lagging behind the CO bond breaking. Through the use of the perturbed or the unperturbed correction schemes, the transition state appears at slightly larger values of the CO distance (about 0.2 Å larger than the AM1/MM value) and at significantly larger values of the CC distance (about 0.6−0.7 Å larger that the averaged AM1/MM value). These results are in agreement with our qualitative discussion on the PES differences and also with the conclusions obtained from the comparison of gas-phase transition structures located at the AM1 and higher QM levels.[38] The 1D-PIC scheme is not able to provide a good geometrical description of the transition state, resulting only in a slight advance along the reaction coordinate. This drawback of the one-dimensional correction method was already noted in our previous works.[29,54] Differences in the reactant states are much less drastic, and in this case the results

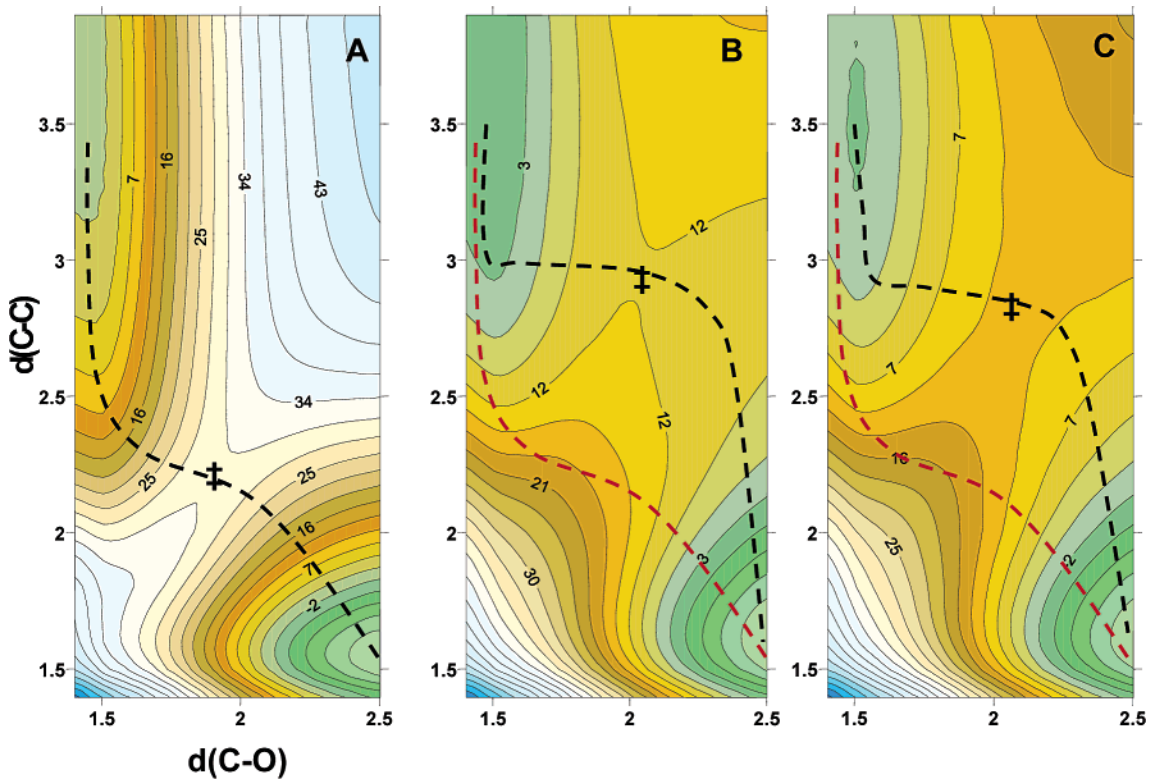


Figure 2. Potential energy surfaces for the chorismate to prephenate rearrangement in BsCM as a function of the bond-breaking (CO) and bond-forming (CC) distances (in Å): (A) obtained at the AM1/MM level, (B) adding 2D-UIC, and (C) adding 2D-PIC, both using B3LYP/6-31G(d) as the higher level. The position of the transition structure and the minimum energy path are qualitatively shown. The red line corresponds to the AM1/MM paths projected on the corrected surfaces. Isopotential energy lines are depicted each 3 kcal/mol.

TABLE 1: Activation Free Energies and Potential Energy Barriers (in kcal/mol) Obtained for the Chorismate to Prephenate Rearrangement in BsCM at the AM1/MM Level and Including 1D or 2D Interpolated Corrections Calculated at the B3LYP/6-31G(d) Level[a]

	2D-PIC	2D-UIC	1D-PIC	AM1/MM	expt.
$\Delta G^{\ddagger}$	10.9	8.8	19.8	29.3	15.4[37]
$\Delta E^{\ddagger}$	12.0	10.0		30.0	
$\langle d_{CO}\rangle_{TS}$	2.11	2.15	1.80	1.94	
$\langle d_{CC}\rangle_{TS}$	2.76	2.90	2.19	2.18	
$\langle d_{CO}\rangle_{R}$	1.50	1.48	1.45	1.46	
$\langle d_{CC}\rangle_{R}$	3.22	3.18	3.20	3.05	
ζ(TS)	−0.65	−0.85	−0.39	−0.24	
ζ(R)	−1.72	−1.70	−1.85	−1.59	

[a] Averaged geometrical values are provided for the bond-breaking and bond-forming distances and the antisymmetric combination (in Å) in the transition and reactant states.

of both correction procedures (1D and 2D) are more coincident. In any case, both the one-dimensional and the two-dimensional methods drastically reduce the error of the activation free energy. Effectively, the AM1 Hamiltonian clearly overestimates the energy of the transition state, providing an activation free energy that nearly doubles the experimental estimation. The use of 1D-PIC results in an activation free energy of 19.8 kcal/mol, 8.9 kcal/mol above the value obtained using 2D-PIC. This difference can be easily understood realizing that the 1D-PIC scheme is equivalent to following the AM1/MM reaction path on the HL/MM surface. This is represented in Figure 2, showing that the AM1/MM path projected on the HL/MM surface passes through regions with considerably higher energies. Finally, it is interesting to note that some minor differences appear between the perturbed and the unperturbed schemes. These differences can be rationalized considering that when using the UIC scheme the polarization and interaction energy are described at the low level. This can lead to a slight overestimation of the enzyme−substrate interaction energy because of the fact that the AM1 wave function is more localized than at higher levels. Taking into account that the interaction is maximized in the TS region,[40] this overestimation leads to a lower activation free energy and to a more dissociative structure.

Catecholate Methylation Catalyzed by COMT. Catechol O-methyltransferase (COMT, EC 2.1.1.6) catalyzes the methyl transfer from *S*-adenosyl-L-methionine (SAM) to a negatively charged oxygen of a substituted catechol. COMT is important in the central nervous system where it metabolizes dopamine, adrenaline, noradrenaline, and various xenobiotic catechols.[55] This reaction involves an attack on a methyl group, originally bonded to the sulfur atom of the coenzyme SAM, by a catecholate nucleophilic oxygen atom in a direct bimolecular S_N2 process (Scheme 2). During the reaction a sulfur−carbon bond (SC) is broken and a new CO is formed. This reaction has been previously studied using different computational approaches,[56,57] including the calculation of the PMF associated with the antisymmetric combination $\zeta = d_{SC} - d_{CO}$.[57] The AM1/MM description of the reaction leads to a noticeable underestimation of the activation free energy barrier, an error that can be attributed to the semiempirical Hamiltonian.[57]

The initial coordinates for the study of the enzymatic process were taken from the X-ray crystal structure of a COMT−inhibitor complex,[58] where the inhibitor was manually changed to catecholate. The full enzyme−substrate system was placed inside a cubic box of 55.8 Å on a side of water molecules, removing those water molecules placed at less than 2.8 Å from any other non-hydrogen atom. The QM subsystem, treated at the AM1 level, consisted of the cofactor SAM and the substrate catecholate (63 atoms), while the MM subsystem, described by means of the OPLS-AA and TIP3P force fields, contained the reminder of the enzyme and water molecules. Periodic bound-

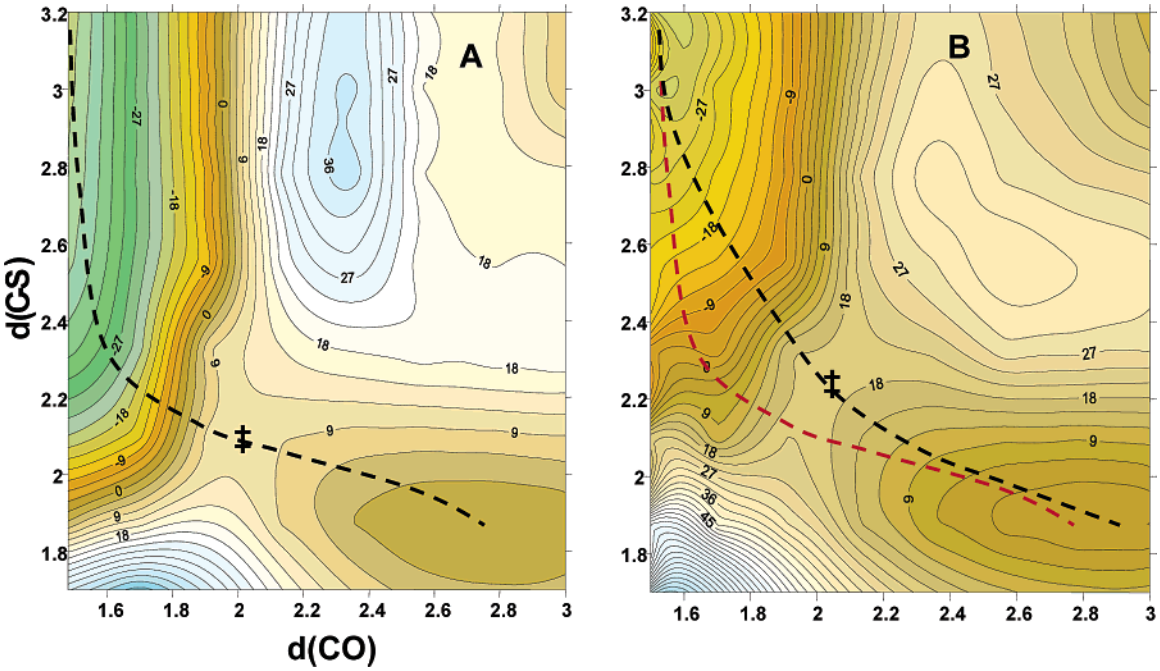
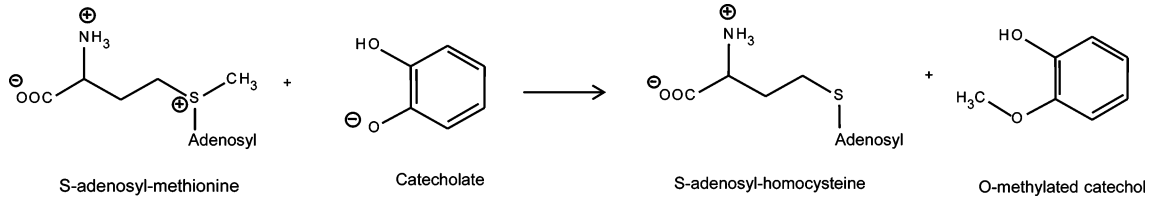


Figure 3. Potential energy surfaces for catecholate SAM-dependent methylation catalyzed by COMT as a function of the bond-breaking (CS) and bond-forming (CO) distances (in Å): (A) obtained at the AM1/MM level and (B) adding 2D-UIC using MP2/6-31+G(d) calculations. The position of the transition structure and the minimum energy paths are qualitatively shown. The red line corresponds to the AM1/MM path projected on the corrected surface. Isopotential energy lines are depicted each 3 kcal/mol.

SCHEME 2



ary conditions and a switched cutoff decay from 8 to 12 Å were employed for describing nonbonding interactions as in the preceding example.

The present study was started from a transition structure localization employing the GRACEFUL algorithm. For this purpose the control space was defined as the set of QM coordinates. Once a transition structure was located and characterized (by inspection of frequencies) the minimum energy path was traced down towards the corresponding reactant and product valleys. We used the bond-breaking and bond-forming distances (d(SC) as λ_1 and d(CO) as λ_2) to obtain a PES for the catecholate methylation. Structures on the surface were obtained by imposing restraints at different values of the λ_i coordinates and minimizing the rest of the system. For this particular system, we varied λ_1 (the bond-breaking distance) in the range from 1.7 to 3.2 Å with a coordinate increment of 0.1 Å and λ_2 (the bond-forming distance) between 1.5 and 3.0 Å with the same step size. The obtained geometries were then used for the computation of the correction energy term within the unperturbed scheme, selecting the MP2/6-31+G(d) method as an adequate higher-level according with our previous experience on this reaction.[57] The correction energy terms were then interpolated through the use two-dimensional bicubic splines (as a function of λ_1 and λ_2) and added to the AM1/MM energy surface to obtain corrected and continuous energy surfaces to be used in PES explorations and PMF calculations.

The AM1/MM PES and that obtained by adding 2D-UIC are shown in Figure 3. In this case the change is not as evident as in the preceding example. The position of the transition structure is quite similar on both PES's with a small displacement along the direction orthogonal to $\zeta = d_{SC} - d_{CO}$. From the energetic point of view the reaction shows a significantly higher energy barrier on the corrected energy surface.

The PMFs for the methyl transfer were obtained along the antisymmetric combination of the distances describing the breaking and forming bonds ($\zeta = d_{SC} - d_{CO}$). Molecular dynamic simulations were carried out in the NVT ensemble at a reference temperature of 300 K, making use of a Langevin thermostat and the Verlet algorithm for updating the velocities. The umbrella-sampling approach was used to restrain the system by means of a parabolic energy penalty (with a force constant of 2500 kJ mol^{-1} Å^{-2}) centered at the desired value of the reaction coordinate. Table 2 presents the results corresponding to three different PMFs obtained using the AM1/MM energy function[57] and the 1D-UIC[59] and 2D-UIC correction schemes. The activation free energies presented in this table are directly obtained as the PMF difference between the maximum and the reactant's minimum. Other contributions to the activation free energy are expected to be very small (about 0.2 kcal/mol).[59] The averaged values of the bond-breaking and bond-forming distances in the reactants and transition states are also given together with the potential energy barriers directly taken from potential energy surfaces presented in Figure 3 (and thus without averaging). In this example, the geometrical differences between the AM1/MM and the corrected transition states are not dramatically large. The positions of both the reactants and the

TABLE 2: Activation Free Energies and Potential Energy Barriers (in kcal/mol) for Catecholate Methylation Catalyzed by COMT Obtained at the AM1/MM Level and Including 1D and 2D Unperturbed Interpolated Corrections Evaluated at the MP2/6-31+G(d) Level[a]

	2D-UIC	1D-UIC	AM1/MM	expt.
$\Delta G^{\ddagger}$	18.6	14.8	10.4	18.0[60]
$\Delta E^{\ddagger}$	17.0		10.0	
$\langle d_{CO}\rangle_{TS}$	2.16	2.03	2.06	
$\langle d_{CS}\rangle_{TS}$	2.14	2.14	2.13	
$\langle d_{CO}\rangle_{R}$	3.11	2.82	2.92	
$\langle d_{CS}\rangle_{R}$	1.93	1.81	1.83	
ζ(TS)	−0.02	0.11	0.07	
ζ(R)	−1.18	−1.01	−1.09	

[a] Averaged values are provided for the bond-breaking and bond-forming distances and the antisymmetric combination (in Å) in the transition and reactant states.

transition states are slightly less advanced along the reaction coordinate on the 2D-corrected surface, and this trend is reflected in the 1D-corrected results. If we focus on the free energies, then we can observe that the semiempirical estimation clearly underestimates the free energy barrier, while a very good quantitative agreement is obtained when including 2D-UIC. The use of 1D-UIC also means a considerable reduction of the error with respect to the experimental data, but the final result still underestimates the activation free energy. The reason can be understood considering the PES's presented in Figure 3. The maximum of the AM1/MM minimum energy path projected on the 2D-corrected surface presents an energy value very close to the true transition structure on this surface. However, the position of the reactant minimum on this AM1/MM path is somewhat advanced with respect to the true reactant structure and its energy is higher. As a result the energy barrier is underestimated.

4. Conclusions

We have presented a procedure to correct potential energy surfaces and potentials of mean force obtained using hybrid QM/MM methods to study chemical processes in condensed media. 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 between a high-level and a low-level calculation of the QM subsystem and expressed as a function of two geometrical parameters relevant in the chemical process under consideration. The correction energy is then fitted by means of two-dimensional bicubic splines, obtaining continuous energy, energy derivatives, and force constants, which are essential both for optimizations and characterizations of stationary structures and for molecular dynamics simulations. These calculations can be carried out with a negligible computational cost once the correction energies have been obtained. As far as many studies of chemical reactions require the previous analysis of the potential energy surface, the calculation of the correction energy can be routinely incorporated to the research protocols.

We have analyzed the performance of the correction scheme for two paradigmatic enzymatic reactions: the chorismate rearrangement catalyzed by BsCM and catecholate methylation catalyzed by COMT. In both cases, we obtained the correction energy term as a function of the bond-breaking and bond-forming distances. The results are compared to the uncorrected simulations and also to the use of one-dimensional interpolated corrections. The use of the new correction procedure clearly improves the quality of the semiempirical/MM results, both from a geometric and an energetic point of view.

Acknowledgment. We are indebted to DGI for project BQU2003-4168, BANCAIXA for project P1A99-03, and Generalitat Valenciana (GVA) for projects GV06-021 and GVA-COMP2006-079, which supported this research. J. J. Ruiz thanks the Ministerio de Educación, Spain, for a FPU fellowship and S. Martí thanks GVA for a Postdoctoral fellowship (CTESPP/2005/088). The authors also acknowledge fruitful discussions with M. Roca, V. Moliner, J. M. Lluch, M. García-Viloca, and A. González-Lafont.

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**, *11*, 700−733.
(3) Thery, V.; Rinaldi, D.; Rivail, J. L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269−282.
(4) Gao, J. L. *Acc. Chem. Res.* **1996**, *29*, 298−305.
(5) Tuñón, I.; Martíns-Costa, M. T. C.; Millot, C.; Ruiz-López, M. F. *J. Chem. Phys.* **1997**, *106*, 3633−3642.
(6) Wright, K.; Hillier, I. H.; Vaughan, D. J.; Vincent, M. A. *Chem. Phys. Lett.* **1999**, *299*, 527−531.
(7) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293−307.
(8) Villà, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 7887−7907.
(9) Mulholland, A. J.; Richards, W. G. *Proteins* **1997**, *27*, 9−25.
(10) Gao, J. L.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.
(11) Field, M. J. *J. Comput. Chem.* **2002**, *23*, 48−58.
(12) Poulsen, T. D.; García-Viloca, M.; Gao, J. L.; Truhlar, D. G. *J. Phys. Chem. B* **2003**, *107*, 9567−9578.
(13) Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425−443.
(14) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618−4627.
(15) García-Viloca, M.; Alhambra, C.; Truhlar, D. G.; Gao, J. L. *J. Comput. Chem.* **2003**, *24*, 177−190.
(16) Martí, S.; Moliner, V.; Tuñón, I. *J. Chem. Theory Comput.* **2005**, *1*, 1008−1016.
(17) Chakraborty, A.; Zhao, Y.; Lin, H.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *124*, 044315.
(18) Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974−2975.
(19) Kollman, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72−79.
(20) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129−S135.
(21) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(22) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−45.
(23) Warshel, A.; Aqvist, J. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 267−298.
(24) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260−7268.
(25) Alhambra, C.; Gao, J.; Corchado, J. C.; Villa, J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1999**, *121*, 2253−2258.
(26) Lau, E. Y.; Kahn, K.; Bash, P. A.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, 97, 9937−9942.
(27) Alhambra, C.; Corchado, J. C.; Sanchez, M. L.; Gao, J. L.; Truhlar, D. G. *J. Am. Chem. Soc.* **2000**, *122*, 8197−8203.
(28) Devi-Kesavan, L. S.; García-Viloca, M.; Gao, J. *Theor. Chem. Acc.* **2003**, *109*, 133−139.
(29) Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I.; Martí, S.; Moliner, V. *J. Phys. Chem. B* **2004**, *108*, 8427−8433.
(30) Nguyen, K. A.; Rossi, I.; Truhlar, D. G. *J. Chem. Phys.* **1995**, *103*, 5522−5530.
(31) Chuang, Y. Y.; Corchado, J. C.; Truhlar, D. G. *J. Phys. Chem. A* **1999**, *103*, 1140−1149.
(32) Corchado, J. C.; Coitiño, E. L.; Chuang, Y. Y.; Fast, P. L.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 2424−2438.
(33) Renka, R. J. *ACM Trans. Math. Software* **1993**, *19*, 81−94.
(34) Proust-De Martín, F.; Dumas, R.; Field, M. J. *J. Am. Chem. Soc.* **2000**, *122*, 7688−7697.
(35) Press: W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. In *Numerical Recipes in C*; Cambridge University Press: Cambridge, U. K., 1988; pp 120−122.
(36) Wilson, E. B.; Decius, J. C. In *Molecular Vibrations*; Dover Publications: New York 1980.
(37) Kast, P.; Asif-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691−2694.

(38) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *J. Am. Chem. Soc.* **2004**, *126*, 311−319.

(39) Lee, Y. S.; Worthington, S. E.; Krauss, M.; Brooks, B. R. *J. Phys. Chem. B* **2002**, *106*, 12059−12065.

(40) Szefczyk, B.; Mulholland, A. J.; Ranaghan, K. E.; Sokalski, W. A. *J. Am. Chem. Soc.* **2004**, *126*, 16148−16159.

(41) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J.; Field, M. J. *J. Am. Chem. Soc.* **2001**, *123*, 1709−1712.

(42) Zhang, X.; Bruice, T. C. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16194−16198.

(43) Woodcock, H. L.; Hodoscek, M.; Sherwood, P.; Lee, Y. S.; Schaefer, H. F.; Brooks, B. R. *Theor. Chem. Acc.* **2003**, *109*, 140−148.

(44) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejon, P.; Roitberg, A. E.; Estrin, D. A. *J. Phys. Chem. B* **2003**, *107*, 13728−13736.

(45) Chook, Y. M.; Gray, J. V.; Ke, H. M.; Lipscomb, W. N. *J. Mol. Biol.* **1994**, *240*, 476−500.

(46) Kaminski; G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105* 6474−6487.

(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(48) Field, M. J.; Albe, M.; Bret, C.; Proust-De Martin, F.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088−1100.

(49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.

(50) Albeit the code has been mainly rewritten using different scripting capabilities and search algorithms, the ideas behind GRACEFUL corresponds with those of Moliner, V.; Turner, A. J.; Williams, I. H. *Chem. Comun.* **1997**, 1271−1272.

(51) Melissas, V. S.; Truhlar, D. G. *J. Phys. Chem.* **1994**, *98*, 875−886.

(52) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.

(53) Schenter, G. K.; Garret, B. C.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 5828−5833.

(54) Ferrer, S.; Ruiz-Pernía, J. J.; Tuñón, I.; Moliner, V.; García-Viloca, M.; Gonzalez-Lafont, A.; Lluch, J. M. *J. Chem. Theory Comput.* **2005**, *1*, 750−761.

(55) Gulberg, H. C.; Marsden, C. A. *Pharmacol. Rev.* **1975**, *27*, 135−206.

(56) Strajbl, M.; Florian, J.; Warshel, A. *J. Phys. Chem. B* **2001**, *105*, 4471−4484.

(57) Roca, M.; Martí, S.; Andrés, J.; Moliner, V.; Tuñón, M.; Bertrán, J.; Williams, I. H. *J. Am. Chem. Soc.* **2003**, *125*, 7726−7737.

(58) Vidgren, J.; Svensson, L. A.; Liljas, A. *Nature* **1994**, *368*, 354−358.

(59) Roca, M.; Moliner, V.; Ruiz-Pernía, J. J.; Silla, E.; Tuñón, I. *J. Phys. Chem. A* **2006**, *110*, 503−509.

(60) Schultz, E.; Nissinen, E.; Kaakkola, S. *Biochem. Pharmacol.* **1989**, *38*, 3953−3956.