JCTC Journal of Chemical Theory and Computation

Improving the QM/MM Description of Chemical Processes: A Dual Level Strategy To Explore the Potential Energy Surface in Very Large Systems

Sergio Martí and Vicente Moliner*

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

Iñaki Tuñón*

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

Received May 31, 2005

Abstract: Potential energy surfaces are fundamental tools for the analysis of reaction mechanisms. The accuracy of these surfaces for reactions in very large systems is often limited by the size of the system even if hybrid quantum mechanics/molecular mechanics (QM/MM) strategies are employed. The large number of degrees of freedom of the system requires hundreds or even thousands of optimization steps to reach convergence. Reactions in condensed media (such as enzymes or solutions) are thus usually restricted to be analyzed using low level quantum mechanical methods, thus introducing a source of error in the description of the QM region. In this paper, an alternative method is proposed, coupled to the use of a micro/macroiteration algorithm during the optimization. In these algorithms, the number of microsteps involved in the QM region optimization is usually much smaller than the number of macrosteps required to optimize the MM region. Thus, we define a new potential energy surface in which the gas-phase energy of the QM subsystem and the interaction energy with the MM subsystem are calculated at different computational levels. The high computational level is restricted to the gas-phase energy, which is only requested during the microsteps. The dual level strategy is tested for two reactions in solution (the Menshutkin and the oxy-Cope reactions) and an enzymatic one (the nucleophilic substitution of 1,2-dichloroethane in DhlA). The performance of the proposed computational scheme seems to be quite promising for future applications in other systems.

1. Introduction

The theoretical study of chemical processes taking place in condensed media (such as enzymes or solution) has become very promising in the last years because of the development of specific computational techniques. The strategy designed to afford such a study requires the consideration of at least three key steps. The first one is the selection and/or construction of the adequate potential energy function to describe the desired process. The quality of this potential energy function is crucial as far as it determines the possible molecular mechanisms of a given chemical reaction. The second step is the location and characterization of the relevant stationary structures (reactants, products, intermediates, and transition structures). This step is essential to discriminate among several mechanisms and to obtain all the information which the transition state theory can provide about the reaction (as, for example, transition vectors, rate constants,

* Corresponding authors phone: +34-6-354-48-80 ext 46100; fax: +34-6-354-45-64; e-mail: Ignacio.Tunon@uv.es (I.T.) and phone +34964728084; fax: +34964728066; e-mail: moliner@exp.uji.es (V.M.).

10.1021/ct0501396 CCC: $30.25
Published on Web 08/06/2005

kinetic isotopic effects, etc.). In any case, a third step involving molecular simulations is required in order to obtain well suited averaged macroscopic properties for the set of thermodynamic states involved in the studied chemical process. So far, one of the most important properties which can be obtained is the free energy profile, from which the activation and reaction free energies are computed. Nonetheless except in the simplest cases, the information obtained during the exploration of the potential energy surface is crucial to perform a proper sampling along adequate reaction coordinates.

The main specific problem related with the study of processes in enzymes or in solution is associated with the size of the systems being considered: the huge number of electrons and degrees of freedom. For the first problem a reasonable solution is provided by hybrid quantum mechanics/molecular mechanics (QM/MM) strategies.[1−3] Usually, electronic reordering due to breaking and forming bonds is located in a reduced portion of the system. One then needs to describe just the atoms of this region using quantum mechanics, while the rest of the system is described using MM potentials. In this way the number of electrons to be considered is drastically reduced. The successful application of this methodology is nowadays well reported in the scientific literature.[1−5] One could then wonder if it would be possible to perform such a reduction of the number of degrees of freedom to be considered, including only those atoms directly involved in the reaction. The answer is generally negative, consisting of a lot of examples of enzymatic and in solution processes where important changes in the environment appear as long as the reaction proceeds. For example, in nucleophilic substitution reactions in solution, solvent molecules reorganize following the reaction charge flow.[6−8] Another example is provided by the enzyme catechol O-methyl transferase, where the activation of the reactants require an important change in the coordination shell of the magnesium cation with respect to the X-ray structure.[9] So, meanwhile in most cases the number of electrons which need to be explicitly treated can be clearly reduced, whereas the same is not generally true for the degrees of freedom of the system. Thus, an adequate study of such kind of processes must consider the very high dimensionality of the potential energy surfaces involved. Fortunately there are efficient optimization algorithms capable of working with a very large number of degrees of freedom.[10]

This paper is organized as follows. In the following section we first summarize the characteristics of the micro/macroiterations optimization algorithm, which is the most successful, in its different variants, to study large systems described with QM/MM methods. Afterward, and taking advantage of the use of this algorithm, we propose a dual-level strategy that can be used to obtain more accurate descriptions of a chemical reaction in very large systems at a still reasonable cost. Finally, the method is tested for three different reactions, two in aqueous solution and one in an enzymatic active site.

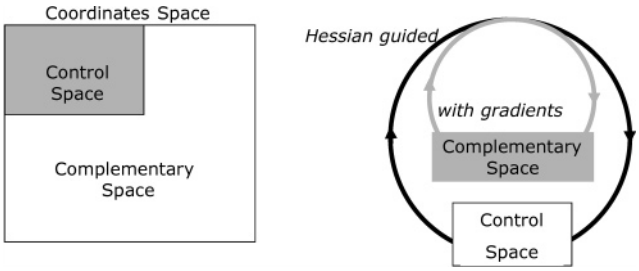


Figure 1. Schematic representation of the micro/macroiteration algorithm. The coordinate space is divided into two subsets: the control space and the complementary space. Optimization steps on the control space (microiterations) make use of a Hessian-based algorithm. At each step of this algorithm the complementary space is minimized using only the gradient vector (macroiterations).

2. Theory

2.1. The Micro/Macroiteration Optimization Algorithm. Most of the standard transition state localization and characterization techniques make use of the Hessian matrix.[11−17] While it is usually straightforward to obtain it in gas-phase calculations, this is not feasible when using a QM/MM based methodology: the large amount of degrees of freedom present in system models turns impracticable its evaluation. To overcome this problem, a partition of the full coordinates space of the system into a control space and complementary space subsets can be done: those atoms or molecules directly involved in the reaction process (plus may be some rounding molecules or residues) are included in the control space, while the rest defines the complementary space. Then, optimization of structures can be efficiently carried out in coupling iterations over these two subspaces: at each step of the control space Hessian guided optimization, the rest of the system is kept fully relaxed merely using gradient vectors. This strategy is also known as the micro/macroiteration method[18−22] (see Figure 1) and leads to stationary structures with the adequate number of negative eigenvalues for a reduced Hessian matrix (the one defined for the control space). In a typical application of the micro/macroiteration approach, the control space usually contains the coordinates of up to ∼100 atoms, while the complementary space can include a number of atoms 2 orders of magnitude larger. When searching stationary structures this is usually translated into about 10^1 Hessian guided optimization steps in the control space and up to 10^2-10^3 gradient based optimization steps in the complementary space at each control space movement. This last number of cycles is of course highly dependent on the size of the system and of the gradient of the considered structure. Anyway, a typical application in enzymes or condensed media can amount up to 10^3 or even more energy and gradients evaluations to fully relax a stationary structure, which means that only low cost computational methods can be employed to describe the QM region (except in those cases where the QM subsystem contains very few atoms). Whereas upcoming computational power would allow using higher electronic Hamiltonian approaches such as ab initio or based in density functional theory (DFT) to describe the quantum region, the usually large amount of gradient vector evaluations needed during

the macrosystem minimization procedure turns this way almost impracticable, still making semiempirical methods the most suitable one.

2.2. The Dual-Level Scheme. We here present a computational scheme to improve the description of QM/MM systems taking advantage of the characteristics of the micro/macroiteration algorithms. We propose a dual level method based on the combination of a high level (HL) and a low level (LL) description of the quantum region. A convenient decomposition of the total energy allows describing the QM gas-phase energy using a HL method, while the interaction with the environment is obtained at a LL one. In this way, the HL term must be evaluated only during the optimization of the control space, while the LL terms are devoted to avoid the more time consuming complementary space optimization. This method has shown to provide considerably better descriptions than standard QM/MM calculations at the semiempirical level, while the computational cost is still reasonable.[23]

The energy for a qm/mm system can be split into different terms

$$E = E_{\mathrm{qm}} + E_{\mathrm{qm/mm}} + E_{\mathrm{mm}} = \langle \Psi | \hat{H}_0 | \Psi \rangle + \langle \Psi | \hat{V} | \Psi \rangle + E_{\mathrm{mm}} \quad (1)$$

where $\hat{H}_0$ is the in vacuo Hamiltonian for the selected qm method, Ψ is the polarized wave function due to the presence of a nonrigid distribution of charges q_{mm} (representing the mm atoms), $\bar{V}$ is the coupling (electrostatic and van der Waals) operator between the qm and mm subsystems, and E_{mm} represents the force field energy. This equation must be solved at each optimization step.

However, having in mind the characteristics of the micro/macrooptimization scheme, there are several considerations that can be applied into the general qm/mm formulation. For each optimization step in the control space (microiteration), the energy function depending on the complementary space is minimized (gradient based macroiterations). Some strategies can be introduced to improve the treatment of the quantum subsystem taking advantage of the characteristics of this optimization procedure. Effectively, eq 1 can be also written as

$$E = \langle \Psi^0 | \hat{H}_0 | \Psi^0 \rangle + [\langle \Psi | \hat{H}_0 | \Psi \rangle - \langle \Psi^0 | \hat{H}_0 | \Psi^0 \rangle] + \langle \Psi | \hat{V} | \Psi \rangle + E_{\mathrm{mm}} = E^{\circ} + E_{\mathrm{pol}} + E_{\mathrm{int}} + E_{\mathrm{mm}} \quad (2)$$

where Ψ^0 stands for the unpolarized (gas phase) wave function of the qm subsystem, E_{pol} for the energy spent in polarizing this wave function, and E_{int} for the interaction between the (polarized) qm subsystem and the mm subsystem. During a micro/macroiteration algorithm only a part of the total energy function is needed at each type of iteration step as either the qm subsystem (in the macrosteps) or the mm subsystem (in the micro steps) is kept frozen. Then, the energy functions explored are

$$E_{\mathrm{micro}} = E_0 + E_{\mathrm{pol}} + E_{\mathrm{int}} \quad (3)$$

$$E_{\mathrm{macro}} = E_{\mathrm{pol}} + E_{\mathrm{int}} + E_{\mathrm{mm}} \quad (4)$$

Equations 3 and 4 allows for improving the quantum treatment of the qm subsystem just describing it in a different way during the micro- and the macroiterations. As far as the flexible mm region is large enough, the number of macroiterations largely exceeds the number of microiterations, a high level (HL) description can be selected only during the microiterations, precisely when the qm subsystem is being optimized, while a low level (LL) description can be used during the macroiterations. One of the first attempts to reduce the more time-consuming step, i.e. the environment relaxation, was accomplished by describing the qm subsystem as an ensemble of classical charges during the macroiterations:[15−17,24]

$$E_{\mathrm{micro}} = E_0^{\mathrm{HL}} + E_{\mathrm{pol}}^{\mathrm{HL}} + E_{\mathrm{int}}^{\mathrm{HL}} \quad (5)$$

$$E_{\mathrm{macro}} = \left[\sum \frac{q_{\mathrm{mm}} q_{\mathrm{qm}}^{\mathrm{fitted}}}{r_{\mathrm{mm,qm}}} + E_{\mathrm{LJ,qm/mm}} \right] + E_{\mathrm{mm}} \quad (6)$$

Albeit nuclear geometry of the core region and fitted charges ($q_{\mathrm{qm}}^{\mathrm{fitted}}$) are kept frozen during the environment optimization, skipping in this way the effect of the core electronic polarization as long as the classical mm electric field changes, this approach can lead to slower convergences (the global number of micro/macroiterations is increased) as far as the energy function contains an intrinsic discontinuity between the micro- and the macroiterations, due to the different treatment of the interaction and polarization terms.

In this point, the dual level method that we propose is based on two different quantum treatments for the qm region, one LL method (usually a semiempirical Hamiltonian) and one HL method (usually, DFT or ab initio correlated methods). The main difference is that we combine these two descriptions into a unique dual level energy surface:

$$E = E_0^{\mathrm{HL}} + E_{\mathrm{pol}}^{\mathrm{LL}} + E_{\mathrm{int}}^{\mathrm{LL}} + E_{\mathrm{mm}} \quad (7)$$

According to this definition of the dual level energy surface, the part of the global energy that is explored during the micro- and macrosteps are

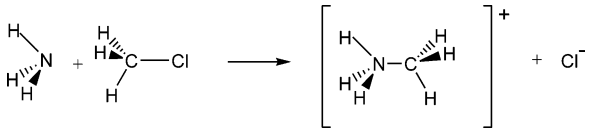
$$E_{\mathrm{micro}} = E_0^{\mathrm{HL}} + E_{\mathrm{pol}}^{\mathrm{LL}} + E_{\mathrm{int}}^{\mathrm{LL}} \quad (8)$$

$$E_{\mathrm{macro}} = E_{\mathrm{pol}}^{\mathrm{LL}} + E_{\mathrm{int}}^{\mathrm{LL}} + E_{\mathrm{mm}} \quad (9)$$

Then, this computational scheme allows a high level description of the qm subsystem during its optimization while avoiding the discontinuities between micro- and macrosteps. In other words, during the environment optimization, the core region is described by a computationally low cost Hamiltonian (such as semiempirical ones); meanwhile a high level correction in the pure gas-phase term is included during the microiterations, keeping the interaction and polarization terms at a low description level. Of course, this computational scheme is not only applicable to the energy but also to the gradient vector and the Hessian matrix:

$$H_{ij} = \frac{\partial^2 E}{\partial x_i \partial x_j} = \frac{\partial^2 E_0^{\mathrm{HL}}}{\partial x_i \partial x_j} + \frac{\partial^2 E_{\mathrm{pol}}^{\mathrm{LL}}}{\partial x_i \partial x_j} + \frac{\partial^2 E_{\mathrm{int}}^{\mathrm{HL}}}{\partial x_i \partial x_j} + \frac{\partial^2 E_{\mathrm{mm}}}{\partial x_i \partial x_j} \quad (10)$$

Scheme 1



It is important to stress here the differences with respect to the popular ONIOM methods.[25,26] Briefly, the strategy here proposed is an electronic embedding method, where the qm region is described using two different computational levels, being the low-level employed to obtain the interaction between the two subsystems. A more detailed discussion on the differences is presented in ref 23.

This dual level strategy, that can be generally denoted as HL:LL/MM calculations, has been coded into the Fortran90 DYNAMO library.[27,28] In the first choice, eqs 5 and 6, the microiteration is obtained by successive calls to Gaussian03[29] for both the energy and the fitted ESP charges[30,31] (considering the nuclear geometry of the core rounded by a distribution of point charges). For the second method, eqs 7−9, Gaussian03 is used just for the high level gas-phase calculation (given the geometry of the core region), meanwhile low level qm/mm and gas-phase calculations are carried out by DYNAMO. In the case of the macroiterations, the L-BFGS-B minimization routines[10] have been incorporated. On the other hand, microiterations are guided by the BAKER algorithm:[32,33] the positive eigenvectors of the Hessian matrix are minimized, while those with a negative eigenvalue are maximized. Control over sequential optimizations steps has been introduced through eigenvector overlapping, allowing the system to undo wrong steps. The BAKER algorithm has been slightly modified from the one already implemented in DYNAMO, to be able to work in a micro/macroscheme.

3. Results and Discussion

Preliminary results of our dual level scheme (eqs 7−9) were presented in ref 23, where the transition state of the chorismate to prephenate rearrangement catalyzed by BsCM was localized, and the kinetic isotopic effects were computed and compared to experiment.[34,35] In this work we present a more systematic test of the dual level QM/MM scheme, exploring the potential energy surface for different reactions in water solution and in an enzymatic environment.

Menshutkin Reaction in Aqueous Solution. The Menshutkin reaction[36] is a S_N2 process in which neutral reactants yield charged products, and then it is a prototypic example of a chemical process where the solvent effects play a decisive role on the process stabilizing the charge separation. We have selected the simplest example of this kind of reaction, as shown in Scheme 1.

This reaction has been thoroughly studied by means of different theoretical strategies involving continuum and QM/MM models,[37−41] and it provides a very good example to test the reliability of the proposed approach. In our calculations the quantum subsystem, composed by methyl chloride plus ammonia, was placed in a water box of 31.4 Å of side containing a total of 1031 TIP3P water molecules.[42] To explore the potential energy surface and locate the transition structure we used a control space, for which the Hessian matrix is obtained, including only the coordinates of the QM subsystem. Structures were converged to a gradient norm lower than 0.1 kJ·mol^{-1}·Å^{-1}. We used different descriptions for our system: AM1/MM, where the quantum subsystem is described by means of the AM1 Hamiltonian; MP2/6-31+G**/MM, where the quantum subsystem is described at the MP2 level of theory; MP2/6-31+G**:CHARGE/MM, where the quantum subsystem is described at the MP2 level of theory and the coupling with the environment is made through the use of MK-ESP charges[30,31] (see eqs 5 and 6), and, finally, MP2/6-31+G**:AM1/MM, where the quantum subsystem is described at the MP2 level of theory and the coupling with the environment (polarization and interaction) is obtained through the AM1 Hamiltonian. The Lennard-Jones parameters for the qm/mm interaction are taken from the OPLS-AA potential,[43,44] except for the chlorine atom, for which we used those of ref 38. We employed a switched cutoff of 14.5 Å for all interactions. The most important characteristics of the transition states located with these methodologies are presented in Table 1.

Table 1. Transition Structures for the Menshutkin Reaction in Water Obtained at Different Computational Levels[a]

	AM1/MM	MP2:AM1/MM	MP2:CHARGE/MM	MP2/MM
d_{NC} (Å)	1.960	2.121	2.080	2.245
d_{CCl} (Å)	2.105	2.187	2.230	2.094
NCCl (deg)	177.7	177.8	177.0	176.4
ν (cm^{-1})	659i	536i	657i	467i
t_{rel}[b]	1.0	22.4	22.4	954

[a] See text. The most significant geometrical parameters (the distances associated with the broken and formed bonds and the angle between these two bonds) are provided together with the frequency of the transition vector and the relative CPU time of the different computational strategies. [b] Relative CPU time per optimization step. A standard optimization step is defined as 1 microiteration followed by 50 macroiterations.

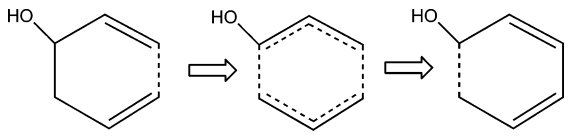
The AM1/MM calculation provides a transition structure with a too short NC distance when compared to the MP2/MM result. This result is clearly improved when the description is made using dual level schemes, either using MP2:AM1/MM or MP2:CHARGE/MM. In general, when compared to the MP2/MM description both dual levels provide a more advanced transition structure, where the NC distance is shorter and the CCl distance is larger, probably due to the underestimation of the interaction between the solute and the solvent molecules. This effect is more important in the MP2:CHARGE/MM calculations than in the MP2:AM1/MM one, which is reflected also in the value of the transition frequency. With respect to the computational efficiency, both dual schemes provide very similar CPU times, and the cost is much more modest than for a pure MP2 optimization.

Once we demonstrated the ability of the MP2:AM1/MM dual scheme to improve the description of the transition state of the model Menshutkin reaction, we traced down steepest descent reaction paths up to the reactant and product valleys. These calculations were followed by complete optimization to obtain stationary reactant and product structures. Table 2

Table 2. Relative Potential Energies (in kcal/mol) for the Reactants, Products, and Transition Structures of the Model Menshutkin Reaction in Aqueous Solution Obtained at the AM1/MM and MP2:AM1/MM Levels

	AM1/MM	MP2:AM1/MM
reactants	0.0	0.0
transition structure	23.1	16.0
products	−18.9	−43.9

Scheme 2



compares the relative potential energies of these stationary structures obtained using MP2:AM1/MM and AM1/MM levels.

The experimental activation and reaction free energies are 23.5 and −34 kcal/mol.[45,46] The potential energy differences presented in Table 2 cannot be directly compared to these experimental free energies as far as entropic effects are expected to make an important contribution. Effectively, during the reaction a charge separation process is taking place, and then the solvent disorder is considerably diminished. However, when the AM1/MM potential energy surface is used to trace the potential of the mean force corresponding to this reaction, the activation free energies are overestimated and the reaction free energies severely underestimated (obtaining 29.3 and −10.4 kcal/mol, respectively).[47] Taking into account these results and from the comparison of the energetic values presented in Table 2, the improvement obtained using the MP2:AM1/MM potential energy surface, at a still moderate computational cost, is quite evident. Assuming that the contribution of the entropy to the activation and reaction free energies are the same at both computational levels we can estimate these quantities in the MP2:AM1/MM approach as

$$\Delta G(\text{MP2:AM1/MM}) = \Delta G(\text{AM1/MM}) + [\Delta E(\text{MP2:AM1/MM}) - \Delta E(\text{AM1/MM})] \quad (11)$$

The results for the activation and reaction free energies are 22.1 and −35.4 kcal/mol, respectively, in very good agreement with the experimental estimations.

Oxy-Cope Reaction in Aqueous Solution. The rearrangement of 1,5-hexadiene-3-ol to 2,4-hexadiene-1-ol (see Scheme 2) is an oxy-Cope reaction, a pericyclic type reaction that has recently received much attention because its implication in antibody catalyzed processes.[48−50]

Solvent molecules, acting as proton donors with the alcohol oxygen atom, increase the energy barrier, while proton acceptor solvent molecules decrease the energy barrier. Bulk electrostatic effects (as obtained from continuum models) slightly increase the energy barrier. These solvent effects can be rationalized in terms of electron withdrawing or electron donor substituent effects.[51]

We selected this reaction as a model of an intramolecular process to test our proposed computational schemes. Location and characterization of the transition structure in aqueous solution has been carried out at the AM1/MM, MPW1K:AM1/MM, MPW1K:CHARGE/MM, and MPW1K/MM levels. Density functional calculations employed the MPW1K hybrid functional,[52] that has recently shown to provide reasonable energy barriers,[53] with the 6-31G* basis set. The quantum system, composed by the substrate (see Scheme 2), was placed in the center of a water box of 31.4 Å of side containing a total of 1026 TIP3P water molecules. The Lennard-Jones parameters of the qm subsystem were taken from the OPLS-AA force field. In all cases we employed a switched cutoff of 14 Å for all interactions. Also in this example we used a control space that includes only the coordinates of the QM subsystem. As before, the convergence criteria was a gradient norm lower than 0.1 $kJ \cdot mol^{-1} \cdot Å^{-1}$. The main results are collected in Table 3.

Table 3. Transition Structures for the Oxy-Cope Reaction in Water Obtained at Different Computational Levels[a]

	AM1/MM	MPW1K:AM1/MM	MPW1K:CHARGE/MM	MPW1K/MM
d_{CC} (Å)	1.646	1.857	1.969	1.948
$d_{CC(OH)}$ (Å)	1.716	1.967	1.871	1.876
ν (cm^{-1})	533i	456i	501i	471i
n_{micro}	42	31	54	40
t_{rel}	1.0	17.9	16.3	610

[a] See text. The distances associated with the broken CC(OH) bond and the new CC bond are provided together with the frequency of the transition vector, the number of microiterations needed to reach convergence, and the relative CPU time per optimization step of the different computational strategies.

The AM1 Hamiltonian gives too short distances for the broken and formed bonds, a feature already observed in other pericyclic reactions, for example, in the Claisen rearrangement.[54,55] Both dual schemes, MPW1K:AM1/MM and MPW1K:CHARGE/MM, give results in good agreement with the pure DFT/MM calculation. The geometrical agreement is somewhat better when using fitted charges to estimate the interaction with the environment. In the case of MPW1K:AM1/MM the transition structure presents a more advanced character, being the distance of the new C−C bond shorter than the distance of the broken bond (the C−C(OH) one). This is a feature also observed in the AM1/MM structure. However, the curvature of the PES, as given by the imaginary frequency associated with the transition vector, is better reproduced by the MPW1K:AM1/MM dual scheme. With respect to the computational efficiency, both dual schemes employ a similar CPU time per optimization step. As in the previous example the cost is about 20 times larger than at the AM1/MM level but much more modest than at the ab initio/MM or DFT/MM levels. It is important to note that, as explained in the preceding section, the convergence of the dual levels based on fitted charges is poorer, because the potential energy surface is not unique and then there is a discontinuity between the micro- and macroiterations. This discontinuity is avoided defining a HL:LL/MM surface with a semiempirical Hamiltonian as low level, which allows for the polarization of the wave function during the macrosteps. This is clearly reflected in the lower number of microiterations needed to reach convergence.

Table 4. Potential Energy Barriers (in kcal/mol) for the Oxy-Cope Reaction in Gas Phase and in Aqueous Solution at Different Computational Levels

	AM1 (gas)	AM1/MM	MPW1K (gas)	MPW1K:AM1/MM	MPW1K:CHARGE/MM
$\Delta E^{\ddagger}$	35.2	40.1	24.1	30.6	29.8

From the transition structures located at the AM1/MM and dual levels we traced down a steepest descent reaction path to the reactant valleys. The energy barriers obtained as the difference between the transition structure and reactants energy are given in Table 4, together with the gas-phase values. The AM1 Hamiltonian overestimates the energy barrier both in gas phase and in solution, a feature also found in other pericyclic reactions.[55] The effect of the solvent observed on the PES exploration, excluding enthalpic and entropic temperature effects, is to increase the energy barrier significantly: between 4.9 and 6.5 kcal/mol. The differences in the estimation of the solvent effect on the energy barrier among the different approaches are not larger than 1.6 kcal/mol. Both dual schemes, using the AM1 Hamiltonian or fitted charges to obtain the interaction between the qm and mm subsystems, provide a very similar energy barrier, although, as said before, the convergence of the MPW1K:AM1/MM scheme is better than using the MPW1K:CHARGE/MM one.

Haloalkane Dehalogenase. Haloalkane dehalogenase from *Xanthobacter Autotrophicus* GJ10 (DhlA) catalyzes the conversion of dichloroethane (DCE) to 2-chloroethanol and chloride.[56] The complete reaction takes place in two steps.[57,58] In the first one, dichloroethane undergoes a S_N2 displacement of a chloride anion by means of the carboxylate group of Asp124, resulting in an ester covalently bound to the enzyme. In a second process, a crystal water molecule hydrolyzes the ester (see Scheme 3).

The first step of the reaction has been the subject of different theoretical studies, in which the activation free energy has been estimated from QM/MM calculations based on semiempirical or empirical valence bond Hamiltonians.[59−61] Semiempirical descriptions of this S_N2 process leads to an important overestimation of the activation free energy, which has been usually corrected with ab initio estimations of the error introduced in the potential energy barrier. We have now explored the PES for this process at the PM3/MM and B3LYP:PM3/MM levels, using in the later the 6-31+G* basis set. With this purpose the 1,2-dichloroethane and part of the Asp124 residue were chosen to be the qm subsystem, while the rest of the enzyme and the water molecules were in the mm subsystem (see ref 60 for details). A link atom was added to the qm subsystem in order to complete its valence. This link atom was placed between CA and CB atoms of Asp124. The mm subsystem was described using the OPLS-AA potential for the enzyme and the flexible TIP3P potential for the water molecules. The Lennard-Jones parameters of the qm subsystem were also taken from the OPLS potential, except for the chlorine atoms, for which we used those of ref 38. A switched cutoff radius of 12 Å was used for all kind of interactions. To explore the potential energy surface and locate the transition structure, the control space, for which the Hessian matrix is obtained, includes only the coordinates of the qm subsystem. As before, the convergence criteria was a gradient norm lower than 0.1 $kJ \cdot mol^{-1} \cdot Å^{-1}$. We first located the transition structure (see Figure 2) using the PM3/MM and B3LYP:PM3/MM descriptions, and afterward we traced the corresponding reaction paths up to the reactant valleys. The main results concerning both structures are collected in Table 5.

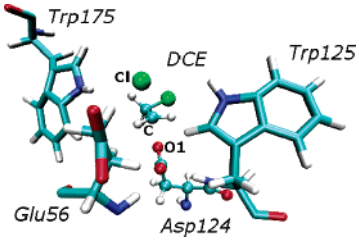


Figure 2. Transition structure for the nucleophilic reaction between DCE and Asp124 in the active site of DhlA. Only some relevant residues are shown for the sake of clarity.

In this case we provide not only intramolecular geometrical parameters but also some important hydrogen bond distances established between the qm and the mm subsystems. In particular the leaving chlorine anion can form hydrogen bonds with the H^{ϵ} of Trp125 and Trp175, while the nucleophilic oxygens of Asp124 (O1 is the nucleophilic oxygen and O2 is the other one) can establish H-bonds with H^{N} atoms of Glu56 and Trp125. The two computational levels give a very similar description of the reactant structure, both regarding intra- and intermolecular parameters. However, important differences appear in the transition structure. Using the dual level scheme, the transition structure presents significantly larger distances for the broken and formed bonds (the CCl and the CO1 distances). The difference is especially

Scheme 3

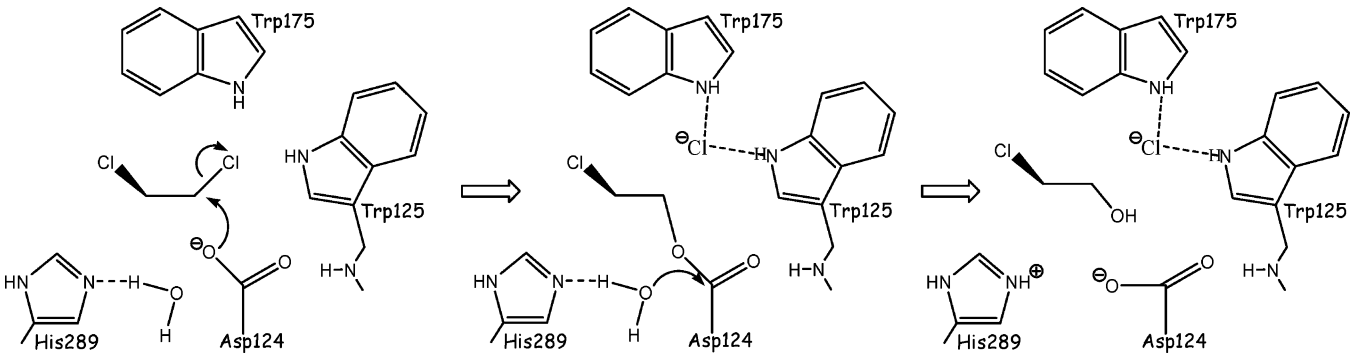


Table 5. Relevant Distances (in Å) and Relative Potential Energies (in kcal/mol^{-1}) for the Reactants and Transition Structure of the Nucleophilic Displacement of Chlorine by Aspartate in the Active Site of DhlA as Obtained from PM3/MM and B3LYP:PM3/MM Calculations

	PM3/MM		B3LYP:PM3/MM	
	R	TS	R	TS
d_{O1C}	2.883	1.905	2.905	2.017
d_{CCl}	1.784	2.256	1.835	2.491
$d_{Cl-Trp175}$	2.875	2.498	2.735	2.444
$d_{Cl-Trp125}$	3.701	2.623	3.777	2.569
$d_{O1-Trp125}$	1.810	2.861	1.888	2.866
$d_{O2-Glu56}$	1.842	1.853	1.829	1.853
ΔE	0	32.9	0	18.2

important in the CCl distance, which is about 0.23 Å larger in the B3LYP:PM3 transition structure than in the PM3 one. This intramolecular change has also consequences on the interaction pattern established with the residues of the active site. In particular, the H-bond distances established by the chlorine leaving group with Trp125 and Trp175 are shorter in the dual level description than in the PM3/MM exploration. This result can be important in the analysis of enzyme effects as it indicates that the use of the semiempirical description can lead to an underestimation of the enzymatic stabilization of the transition structure. We must keep in mind that much of the recent debate about the origin of an enzyme's ability to speed up chemical reactions is based on the comparison between the magnitude of transition state stabilization and reactant state destabilization.[9,62]

The energetic description of the reaction seems to be clearly improved with our dual treatment. Effectively, potential of mean force estimations of the activation free energy using PM3/MM energy functions lead to a severe overestimation. For example, in ref 60 we recently obtained a PM3/MM activation free energy of 28.4 kcal/mol to be compared to the experimentally derived value of 15.3 kcal/mol. If we look now at the potential energy barriers given in Table 5 we can see that the potential energy barrier provided by the B3LYP:PM3/MM method is about 14.7 kcal/mol lower than the PM3/MM value. Assuming that other contributions to the activation free energy remain essentially constant we could estimate (see eq 11) a dual level activation free energy of about 13.7 kcal/mol in much better agreement with the experimental value.

As a final test about the goodness of our dual level PES, heavy atom kinetic isotopic effects on this latter enzyme catalyzed reaction have been computed. We will compare our results with the available experimental data of intrinsic chlorine KIEs, as these effects provide information about the single chemical step of the full dehalogenation catalyzed process. The required structures of reactants (Michaelis complex) and TS were taken from the previously located and characterized by means of both methods: PM3/MM and B3LYP:AM1/MM. Once the stationary structures on the PES were refined, the rigid-rotor/harmonic-oscillator approximations were used with the CAMVIB/CAMISO programs[63,64] to calculate semiclassical KIEs without the scaling of vibrational frequencies, as explained and applied in previous papers.[23,65] A core of 15 atoms was equivalent to the QM region, which corresponds to the substrate. This subset of atoms was used to define the Hessian for these KIE calculations, consistent with the "cut-off rule" and the local nature of isotope effects.[66] The calculated values were 1.0047 and 1.0071 at the PM3/MM and the B3LYP:AM1/MM levels, respectively. Both methods predict small, normal heavy-atom effects for ^{37}Cl substitutions, close to the experimentally determined[67] k_{35}/k_{37} equal to 1.0066. Nevertheless, as we expected, the use of a high level theory to describe the QM region renders a value much closer to the experimental results, as was previously observed for the chorismate mutase KIE calculations.[66] A better agreement with experiments has been obtained not only because the structures obtained at a higher level are more reliable but also because of the good quality of vibrational frequencies. This limitation of the semiempirical methods to predict good chlorine kinetic isotope effects was already observed by Paneth and co-workers[68] who obtained the best estimation by means of the ONIOM QM/MM method on a reduced active site model. Our final results give much confidence in our developed dual level method.

4. Conclusions

We have here presented a new dual level strategy to be used in qm/mm methods combined with a micro/macroiteration algorithm for searching stationary structures. A new potential energy surface is defined in which the gas-phase and the interaction energies of the qm subsystem are obtained using different computational levels: a high level and a low level, respectively.

We have analyzed the performance of dual level methods in two chemical reactions in solution (mensutkin and oxy-Cope processes in aqueous solution) and one enzymatic reaction (a S_N2 reaction catalyzed by DhlA haloalkane dehalogenase). When compared to typical semiempirical/MM computations, the dual level approach (HL:LL/MM) allows for a considerable improvement of the description of bond breaking and bond forming processes. We have tested two different implementations, one in which the low level method is a semiempirical description and another one in which the low level method is a set of potential-derived charges. While the computational efficiency of both implementations is very similar, the use of a qm method as low level ensures the continuity of the potential energy surface between micro- and macroiterations. This characteristic leads to a reduction in the number of steps needed to reach full convergence, and then it seems to be the most promising strategy of the two proposed here. Obviously, the computational scheme is not restricted to a particular choice of the high level and low level methods, and then extension to other qm methods will be available in future applications.

Acknowledgment. We thank DGI (Spain) for project BQU2003-04168-C03, BANCAIXA for project P1A99-03 and Generalitat Valenciana for projects GRUPOS04/28, GV04B-21, and GV04B-131.

References

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227.

(2) Gao, J. In *Methods and applications of combined quantum mechanical and molecular mechanical potentials*; Lipkowitz, Boyd, Eds.; VCH Inc.: New York, 1995.

(3) Gao, J.; Truhlar, D. G. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467.

(4) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Chem. Eur. J.* **2003**, *9*, 984.

(5) Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar D. G. *Science* **2004**, *303*, 186.

(6) Gertner, B. J.; Wilson, K. R.; Hynes, *J. Chem. Phys.* **1989**, *90*, 3537.

(7) Gertner, B. J.; Whitnell, R. M.; Wilson, K. R.; Hynes, J. T. *J. Am. Chem. Soc.* **1991**, *113*, 74.

(8) Kim, H. J.; Hynes, J. T. *J. Am. Chem. Soc.* **1992**, *114*, 10528.

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

(10) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. *J. Sci. Comput.* **1995**, *16*, 1190.

(11) Farkas, Ö.; Schlegel, H. B. *J. Chem. Phys.* **1999**, *111*, 10806.

(12) Schlegel, H. B. *J. Comput. Chem.* **2003**, *24*, 1514.

(13) Pu, J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2005**, *1*, 54.

(14) Nudged Elastic Band (NEB) methods make use of a discrete representation of the reaction path from the reactant structure to the product one, along a string of replicas which are usually built by linear interpolation. Then optimization algorithms (mostly gradient based) are used to relax each of the structures of the chain towards the minimum energy path (MEP), providing this way an idea of the energy profile and of the transition structure. The key of the method relies on a proper definition of the constraints between consecutive points of the chain. For further information see: (a) Henkelman, G.; Jóhannesson, G.; Jónsson, H. *Methods for Finding Saddle Points and Minimum Energy Paths*; Schwartz, S. D., Ed.; Kluwer Academic Publishers: 2000. (b) Olsen, R. A.; Kroes, G. J.; Henkelman, G.; Arnaldsson, A.; Jónsson, H. *J. Chem. Phys.* **2004**, *121*, 9776.

(15) Galván, I. F.; Sánchez, M. L.; Martín, M. E.; Olivares del Valle, F. J.; Aguilar, M. A. *J. Chem. Phys.* **2003**, *118*, 255.

(16) Galván, I. F.; Sánchez, M. L.; Martín, M. E.; Olivares del Valle, F. J.; Aguilar, M. A. *Comput. Phys. Commun.* **2003**, *115*, 244.

(17) Galván, I. F.; Martín, M. E.; Aguilar, M. A. *J. Comput. Chem.* **2004**, *25*, 1227.

(18) (a) Moliner, V.; Turner, A. J.; Williams, I. H. *Chem. Com.* **1997**, 1271. (b) Turner, A. J.; Moliner, V.; Williams, I. H. *Phys. Chem. Chem. Phys.* **1999**, *1*, 1323.

(19) Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H.; Frisch, M. *J. Comput. Chem.* **2003**, *24*, 760.

(20) Prat-Resina, X.; González-Lafont, A.; Lluch, J. M. *J. Mol. Struct. Theochem* **2003**, *632*, 297.

(21) Monard, G.; Prat-Resina, X.; González-Lafont, A.; Lluch, J. M. *Int. J. Quantum. Chem.* **2003**, *93*, 229.

(22) Prat-Resina, X.; Bofill, J. M.; González-Lafont, A.; Lluch, J. M. *Int. J. Quantum. Chem.* **2004**, *98*, 367.

(23) Martí, S.; Moliner, V.; Tuñón, I.; Williams, I. H. *J. Phys. Chem. B* **2005**, *109*, 3707.

(24) Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483.

(25) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. *J. Mol. Struct. Theochem* **1999**, *21*, 461.

(26) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.

(27) Field, M. J.; Albe, M.; Bret, C.; Proust-de Martin, F.; Thomas, A. *J. Comput. Chem.* **2000**, *21*, 1088.

(28) Field, M. J. *A practical introduction to the simulation of molecular systems*; Cambridge University Press: 1999.

(29) Gaussian 03, by 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.; 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 Inc.: Pittsburgh, PA, 2003.

(30) Singh U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.

(31) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.

(32) Baker, J.; Kessi, A.; Delley, B. *J. Chem. Phys.* **1996**, *105*, 192.

(33) Baker, J. *J. Comput. Chem.* **1997**, *18*, 1079.

(34) Lia Addadi, E. K. J. *Biochemistry* **1983**, *22*, 4494.

(35) Gustin, D. J.; Mattei, P.; Kast, P.; Wiest, O.; Lee, L.; Cleland, W. W.; Hilvert, D. *J. Am. Chem. Soc.* **1999**, *121*, 1756.

(36) Menshutkin, N. Z. *Phys. Chem.* **1890**, *5*, 589.

(37) Solà, M.; Lledos, A.; Duran, M.; Bertran, J.; Abboud, J. M. *J. Am. Chem. Soc.* **1991**, *113*, 2873.

(38) Gao, J.; Xia, X. *J. Am. Chem. Soc.* **1993**, *115*, 9667.

(39) Dillet, V.; Rinaldi, D.; Bertran, J.; Rivail, J. L. *J. Chem. Phys.* **1996**, *104*, 9437.

(40) Chuang, Y.-Y.; Cramer, C. J.; Truhlar, D. G. *Int. J. Quantum Chem.* **1998**, *70*, 887.

(41) Castejon, H.; Wiberg, K. B. *J. Am. Chem. Soc.* **1999**, *121*, 2139.

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

(43) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(44) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(45) Okamoto, K.; Fukui, S.; Shingu, H. *Bull. Chem. Soc. Jpn.* **1967**, *40*, 1920.

(46) Okamoto, K.; Fukui, S.; Nitta, I.; Shingu, H. *Bull. Chem. Soc. Jpn.* **1967**, *40*, 2354.

(47) Ruiz-Pernia, J. J.; Silla, E.; Tunon, I.; Marti, S.; Moliner, V. *J. Phys. Chem. B* **2004**, *108*, 8427.

(48) Braisted, A. C.; Schultz, P. G. *J. Am. Chem. Soc.* **1994**, *11*, 2211.

(49) Black, K. A.; Leach, A. G.; Kalani, M. Y. S.; Houk, K. N. *J. Am. Chem. Soc.* **2004**, *126*, 9695.

(50) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Angew. Chem., Int. Ed.* **2005**, *44*, 904.

(51) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. unpublished results.

(52) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715.

(53) Zhao, Y.; Gonzalez-Garcia, N.; Truhlar, D. G. *J. Phys. Chem. A* **2005**, *109*, 2012.

(54) Ganem, B. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 936.

(55) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Theor. Chem. Acc.* **2001**, *105*, 207.

(56) Pries, F.; Kingma, J.; Krooshof, G. H.; Jeronimus-Stratingh, C. M.; Bruins, A. P.; Janssen, D. B. *J. Biol. Chem.* **1995**, *270*, 10405.

(57) Kennes, C.; Pries, F.; Krooshof, G. H.; Bokma, E.; Kingma, J.; Janssen, D. J. *Eur. J. Biochem.* **1995**, *228*, 403.

(58) Schanstra, J. P.; Kingma, J.; Janssen, D. B. *J. Biol. Chem.* **1996**, *271*, 14747.

(59) Nam, K.; Prat-Resina, X.; Garcia-Viloca, M.; Devi-Kesavan, L. S.; Gao, J. *J. Am. Chem. Soc.* **2004**, *126*, 1369.

(60) Soriano, A.; Silla, E.; Tunon, I.; Ruiz-Lopez, M. F. *J. Am. Chem. Soc.* **2005**, *127*, 1946.

(61) Olsson, M. H. M.; Warshel, A. *J. Am. Chem. Soc.* **2004**, 126, 15167.

(62) Martí, S.; Roca, M.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. *Chem. Soc. Rev.* **2004**, *33*, 98.

(63) Williams, I. *Chem. Phys. Lett.* **1982**, *88*, 462.

(64) Williams, I. *J. Mol. Struct. Theochem* **1983**, *11*, 275.

(65) Martí, S.; Moliner, V.; Tuñón, I.; Williams, I. H. *Org. Biomol. Chem.* **2003**, *1*, 483.

(66) Ruggiero, G. D.; Guy, S. J.; Martí, S.; Moliner, V.; Williams, I. H. *J. Phys. Org. Chem.* **2004**, *17*, 592.

(67) Grimsrud, E. P.; Taylor, J. W. *J. Am. Chem. Soc.* **1970**, *92*, 739.

(68) Lewandowicz, A.; Rudzinski, J.; Tronstad, L.; Widersten, M.; Ryberg, P.; Mattson, O.; Paneth, P. *J. Am. Chem. Soc.* **2001**, *123*, 4550.

CT0501396