Geometry Optimization and Transition State Search in Enzymes: Different Options in the Microiterative Method

XAVIER PRAT-RESINA,[1] JOSEP MARIA BOFILL,[2] ÀNGELS GONZÁLEZ-LAFONT,[1] JOSÉ M. LLUCH[1]

[1]*Departament de Química, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain*

[2]*Departament de Química Orgànica and Centre Especial de Recerca en Química Teòrica, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain*

Received 13 June 2003; revised 3 September 2003; accepted 15 December 2003
Published online 22 March 2004 in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/qua.20072

ABSTRACT: In the current article we present a systematic analysis of the different options in the so-called "microiterative method" used to locate minima and transition-state (TS) structures of big systems on quantum mechanics/molecular mechanics potential energy surfaces. The method splits the system in two parts: a core zone in which accurate second-order search is carried out, and an environment that is kept minimized with a cheap first-order algorithm. The different options studied here are: the alternating frequency between the environment minimization and the TS search in the core, the number of atoms included in each zone, and two alternative ways to reduce the computational cost in the calculation of core–environment interactions. The tests have been done at two different steps of the enzymatic mechanism of mandelate racemase: a proton transfer and a carbon configuration inversion step. The two selected TS structures differ in the number of atoms involved in their associated transition vectors; the proton transfer TS is an example of a local motion, whereas the carbon configuration inversion TS corresponds to a more global movement of several groups and residues, including an important number of atoms.
Int J Quantum Chem 98: 367–377, 2004

Key words: enzyme catalysis; Mandelate Racemase reaction mechanisms; microiterative method; QM/MM transition state structures; second derivatives direct optimization method

Correspondence to: J. M. Lluch; e-mail: lluch@klingon.uab.es
Contract grant sponsors: Ministerio de Ciencia y Tecnología and Fondo Europeo de Desarrollo Regional.
Contract grant numbers: BQU2002-00301 and BQU2002-00293.

Introduction

These last years, theoretical chemistry has focused its efforts on a deeper understanding of reactive systems constituted by thousands of atoms. Condensed phase reactions, enzymatic reactions, and solid state catalysis are some of the most successful areas. When so many degrees of freedom are taken into account, the eternal problem of describing the system with an adequate potential is especially challenging. In addition, an appropriate exploration of the configurational space is crucial to calculate thermodynamic and kinetic magnitudes and to understand the mechanism of a particular reactive process. Some work was done on the acceleration of that exploration of the configurational space by molecular dynamics techniques [1, 2].

Another important issue has to do with the need of building up a path connecting reactants to products before performing any dynamics calculations. In this way one tries to avoid trajectories getting trapped in regions far away from any of the reaction pathways. To that end some strategies were proposed to elucidate the reaction pathways, that is, the classical minimum energy path (MEP) [3], the recently developed temperature-depending MEP [4, 5], or an ensemble of transition paths that take into account dynamical effects [6].

In any case, the procedures mentioned above require the availability of very cheap potential energy expressions. However, the state of the art in enzymatic catalysis makes use of quantum mechanical/molecular mechanics (QM/MM) methods [7–9] or, as an alternative, of linear scaling methods [10, 11]. Although these potentials were originally developed to obtain a fast energy evaluation in big systems, their computational costs are still too high to afford the full exploration of the configurational space or even to completely locate a temperature-depending path. In that sense, the search of local saddle points (from here on, we refer to them as "transition states" [TS]) of a given reaction step in the potential energy surface (PES) is still a good strategy to verify the reliability of the reaction pathway. Moreover, when QM/MM potentials are so expensive as to render a posterior dynamics study prohibitive, the exploration of a local valley to locate stationary points might be the only way to obtain information about the reactivity of a chemical system.

The location of stationary points in chemical systems involving up to few tens of atoms is a very well stablished field [3]. The most effective algorithms are based on the Newton–Raphson equation that uses the second derivatives matrix (Hessian) [12]. On the other hand, when hundreds of atoms need to be moved, several problems emerge, due to the manipulation of these very big matrices, mainly, their computation, diagonalization, and storage. Some solutions were proposed [13, 14]; however, most avoid the usage of the Hessian and so increase the number of energy and gradient evaluations.

In high-dimensioned systems, there are many available parallel reaction paths at a given finite temperature. As a matter of fact, in systems such as enzymes, the PES becomes so shallow that special care must be taken to always perform the search in the same local valley. Therefore, we need efficient and tight algorithms. Furthermore, when expensive potentials such as the QM/MM ones are used, the process used should contain as few steps as possible. For all these reasons, we decided to take advantage of the efficiency of the microiterative method as a possible solution to locate not only TS but also stationary points in general in systems with many degrees of freedom.

The microiterative method is a strategy first used by Maseras and Morokuma [15] in the Integrated Molecular Orbital Molecular Mechanics (IMOMM) scheme and then applied by several groups to enzymatic systems [16–21]. The method splits the system in two parts: a core zone, in which an accurate second-order search is done; and an environment that is kept minimized with a cheap first-order method. The separation makes the sum of the expenses of the two processes considerably lower than that of a single global search.

In the current work, we study some of the possible strategies that can be adopted in the microiterative method. To our knowledge, no similar systematic study has been published so far. The article is presented in two main parts. First we present the microiterative method as a possible solution to our problem, discuss the different options that this method can offer, and mention the big Hessian manipulation problems. Second we present the results, briefly describe the enzymatic system that we have used to perform our systematic tests, discuss the results obtained with the different options of the method, and, finally, present the conclusions.

Method

LOCATION OF STATIONARY POINTS IN BIG SYSTEMS

Most of the program packages designed to study biological systems have several standard minimizers. The algorithms were conceived for cases in which potential energy, usually molecular mechanics, is computationally cheap and the main computer demand is the storage of enormous vector arrays and the slow matrix diagonalizations. Examples of the methods are steepest descent, conjugate gradient, or the more accurate Adopted Basis Newton–Raphson (ABNR) [22], Truncated Newton [23], and Limited Memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [24] (see Ref. [25] for a recent comparison of some of the methods). Usually they are applied to minimize an experimental Protein Data Bank (PDB) structure before running molecular dynamics trajectories or a normal mode analysis. However, none of the methods were originally designed to search for TS structures. We need efficient methods to locate directly TS structures in enzymatic catalysis using the information provided by the second derivatives. In addition, the methods must be developed to use the rather expensive QM/MM potentials but with a demand of a reasonable computer cost.

In a recent article [26], we pointed out how important the accurate location of TS is. Sometimes, when the appropriate reaction coordinate is more complicated than a unique distance or an angle, the direct location of the stationary point making use of second derivatives is a good strategy. This prevents wrong conclusions when both MEP analysis and posterior free energy calculation along the reaction path are made.

A compromise between the need for using a second derivative method and the drawbacks of moving many degrees of freedom is the microiterative method. In this case, a Newton-like method [12] is used to locate a stationary point in a small core zone where the number of atoms to be moved is small enough to avoid computational problems. The rest of the atoms of the system belong to the environment zone, which is minimized with an inexpensive method. Both processes are alternated until consistency is reached (see Fig. 1). In our particular case, we make use of the rational function optimization method (RFO) [27–29] as a second-order method for the search in the core, whereas for the environment, the L-BFGS method is applied. Although the scheme can be applied to locate any kind of stationary points, from here on we refer only to TS structures.

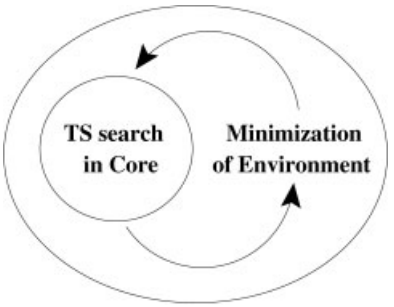

FIGURE 1. Scheme of the microiterative method.

The core and environment zones do not have to match the QM and MM regions, respectively. The quantum region is selected in a QM/MM system as the set of atoms whose interactions need to be described by a QM potential, usually when bond breaking or charge transfer is involved. This selection defines a certain PES. Conversely, the selection of the core zone is a strategy used to locate stationary points, and it can change when looking for different stationary points in a reaction mechanism. The criterium to select the core zone is not based on the interaction energy but on geometrical criteria, that is, the core zone must include the atoms whose movement may play an important role when looking for the stationary point at the current chemical step. Depending on the reaction type, a big core (even bigger than the QM zone) must be chosen, whereas in some other cases, a core with few atoms is good enough to easily reach the stationary point.

POSSIBLE OPTIONS IN THE MICROITERATIVE METHOD

Different possible options can be chosen in the general microiterative scheme, some of which, as described below, may be crucial in order to reach the accurate convergence as quickly as possible.

Although some authors [15, 16] perform a full minimization of the environment at every TS search step, it is not evident that this is the best way to proceed. We could run a full TS search until convergence before minimizing the environment again. As a matter of fact, there is as yet no agreement as to which is the most suitable alternating frequency

between the two processes: minimization of the environment, and the Newton step in the core.

Another important aspect is how big the core zone must be in order to reach convergence as quickly as possible. When the core is big, the coupling between the two zones is minimized. This would tend to reduce the number of steps required to converge. On the other hand, a big core region implies a bigger Hessian matrix and, as a consequence, a more expensive initial Hessian calculation (see below for a discussion of other problems related to big Hessian matrices).

CORE–ENVIRONMENT INTERACTION

Another option in the microiterative method that deserves more discussion is how to handle the interaction between the core and the environment. The environment zone is usually bigger than the core zone, and it needs to be optimized with a cheap minimizer, such as conjugate gradient, ABNR, or L-BFGS. Although the methods imply low memory requirements, their poor efficiency consequently requires many steps to reach convergence. Depending on the QM level, the minimization could demand a high computational effort if a full QM/MM calculation is performed at every step.

Different strategies were proposed to perform faster QM/MM energy evaluations [19, 20, 30–33]. The strategies are based on the idea that an exact self-consistent field (SCF) evaluation of the QM zone might not be needed at each minimization step in order to calculate the energy of the system. The methods were not only applied to the location of stationary points [19, 20], but some work had already been done in QM/MM Monte Carlo to explore the environment configurational space, avoiding a full SCF evaluation at every step of the simulation [30–33].

To summarize the different approximations adopted to accelerate the QM/MM calculation, we have to keep in mind the general QM/MM electronic embedding scheme [7]. The full Hamiltonian can be expressed as:

$$\mathbf{H} = \mathbf{H}_{\mathrm{QM}} + \mathbf{H}_{\mathrm{MM}} + \mathbf{H}_{\mathrm{QM/MM}} \tag{1}$$

where $\mathbf{H}_{\mathrm{QM}}$ is the Hamiltonian describing the set of atoms whose interactions are computed using QM and $\mathbf{H}_{\mathrm{MM}}$ is the MM Hamiltonian. The crucial aspect is the calculation of $\mathbf{H}_{\mathrm{QM/MM}}$ that describes the interaction between the two regions treated at the two different levels,

$$\mathbf{H}_{\mathrm{QM/MM}} = V_{\mathrm{QM/MM}}^{\text{van der Waals}} - \sum_{i}^{\text{electrons}} \sum_{C}^{\text{classical}} \frac{Q_C}{r_{iC}} + \sum_{K}^{\text{nuclei}} \sum_{C}^{\text{classical}} \frac{Z_K Q_C}{R_{KC}}, \tag{2}$$

where Q_C are the MM charges, Z_K are the effective nuclear charges of quantum atoms, and r_{iC} and R_{KC} are the distances from the MM charges to the electrons and quantum nuclei, respectively. The second term of the right side of Eq. (2) describes the interaction between the MM charges Q_C and the electrons. The Q_C charges will polarize the wavefunction. So when MM atoms are moved, all the terms in Eq. (1) must be recomputed in order to consider the effect of quantum atoms on the MM zone. Indeed, when the QM level is highly accurate, the energy computation becomes unaffordable.

A possible attempt to speed up the full QM/MM calculation is to consider a quantum atom as a classical atom during the MM atoms movement. This can be carried out associating point charges to the quantum atoms, for example, computing the electrostatic potential (ESP) fitted charges [34] from an electron density obtained by means of a full QM/MM calculation. So, in Eq. (2) we can join the second and third term, giving the expression shown in Eq. (3):

$$\mathbf{H}_{\mathrm{QM/MM}} = V_{\mathrm{ESP/MM}}^{\text{van der Waals}} + \sum_{K}^{\text{quantum}} \sum_{C}^{\text{classical}} \frac{Q_{\mathrm{ESP}K} Q_C}{R_{KC}}. \tag{3}$$

The ESP charges for the quantum atoms are constant during the environment minimization and they will be recomputed at the next core step or at the next full QM/MM evaluation. In this case it is evident that the core zone must always include all the quantum atoms.

In the current work, the ESP charges are taken from a parametric method 3 (PM3) wavefunction instead of using an ab initio one [19], but this does not include any systematic error in the obtained results.

It is important to note that the ESP/MM strategy reaches a geometry that is not a stationary point on the real QM/MM surface. It is not evident that the interaction between classical charges and ESP

charges is equivalent to the second and the third term in Eq. (2). A better approximation is a method that is based on the original QM/MM expression, hereafter called "1SCF method" in the current paper.

If we take a look at the QM/MM Eq. (2), we realize that the QM/MM interaction Hamiltonian only contributes to the total one-electron core quantum Hamiltonian but not to the two-electron integrals. This means that for a fixed geometry of the quantum atoms, if we move the MM atoms, the two-electron integrals do not have to be recomputed and the one-electron integrals can be easily updated at the new Q_C configuration. If we save the last converged wavefunction (Ψ_{frozen}) and perform only one SCF cycle, the Fock matrix will not be diagonal, but we will obtain a good and cheap approximation to the exact QM/MM energy [see Eq. (4)]:

$$\mathbf{E}_{\text{1SCF}} = \langle \Psi_{\text{frozen}} | \mathbf{H}_{\text{QM}} - \sum_{i}^{\text{electrons}} \sum_{C}^{\text{classical}} \frac{Q_C}{r_{iC}} | \Psi_{\text{frozen}} \rangle + \mathbf{H}_{\text{MM}} + V_{\text{QM/MM}}^{\text{van der Waals}} + \sum_{K}^{\text{nuclei}} \sum_{C}^{\text{classical}} \frac{Z_K Q_C}{R_{KC}}. \quad (4)$$

Our proposal is to use the method to save computing time during the minimization of the environment in the microiterative scheme. The efficiency of the strategy was evaluated by Evans and Truong [35] in a Monte Carlo sampling of the environment. Keeping the wavefunction frozen gives a good approximation as long as the perturbation of the classical charges Q_C is small. That is, if the distribution of charges Q_C during the minimization does not change too much or they are not too close to the QM part to polarize the real $|\Psi>$ in a very intense manner, the 1SCF method works well.

SOLUTIONS TO THE MANIPULATION OF A BIG HESSIAN

As mentioned above, the reason standard Newton methods are not used in many systems with a great number of atoms is the computational difficulties encountered in a big Hessian manipulation. Microiterative methods overcome this problem, but we must still take into account that an expensive QM method or an increase in the core size provokes these computational problems to reappear.

Three main problems arise when working with a big Hessian, that is, the initial Hessian calculation, its diagonalization at every step of the process, and its storage. The initial Hessian computational effort is related to the QM level, whereas the two other issues are related only to the size. In any case, as described in the results section, a core size bigger than a thousand atoms is not useful, so that the storage problem is not a real problem in modern computers. Moreover, the full diagonalization can be avoided with a standard iterative diagonalization that provides the two lowest eigenpairs needed to calculate the displacement in the RFO scheme [29, 36].

Depending on the QM level, a compromise must be found between the quality of the initial Hessian and the steps required to converge. That is, a very cheap second-derivatives matrix could be used, but this would imply a larger number of steps. Because the number of steps required to converge with a bad initial Hessian can increase very rapidly, our recommendation would be to use the highest quality Hessian possible. Of course, in a TS search, the last statement is crucial. It was also shown [18, 37] that an adequate initial Hessian eliminates the known problems of coupling between Cartesian coordinates.

Any of the approximations proposed above to save computing time during the minimization of the environment in a microiterative scheme also can be used when calculating the energy derivatives with respect to the coordinates of the classical atoms that belong to the core zone. In this sense, we have seen that the 1SCF approximation is a good compromise between computational cost and accuracy.

Results and Discussion: Tests on Mandelate Racemase

MANDELATE RACEMASE ENZYME

We have performed several series of tests on a model of mandelate racemase enzyme, which we had already studied in previous works [26, 38, 39]. The enzyme catalyzes the reversible interconversion of the (S)- and (R)-enantiomers of mandelate and other, β,γ-unsaturated α-hydroxy carboxylates. Thus, propargylglycolate was found to be a moderately good substrate for racemization. In Figure 2, the active site of the enzyme shows that many residues can interact directly with the substrate.

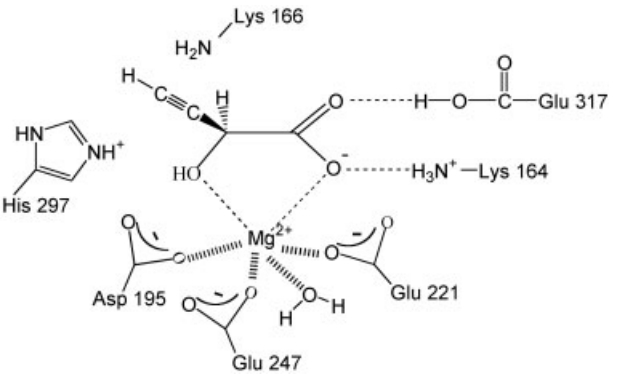

FIGURE 2. Mandelate racemase active site including propargylglycolate substrate.

Our previous results [38, 39] indicate that the racemization of propargylglycolate can take place via three parallel mechanisms. Two of them (mechanisms I and II) involve six steps through six TS structures and five intermediates each. In contrast, the third (mechanism III) is a one-step mechanism. From this number of steps we have selected two as representative examples of opposite cases. The first is step 1 of mechanism II, which consists of a proton transfer from Glu317 to a carboxylate oxygen of propargylglycolate. The second is step 4 of mechanism I, which essentially involves the configuration inversion of the stereogenic carbon of propargylglycolate. Analysis of the components of the transition vectors associated with the respective TS structures demonstrates the difference between the two steps, which, for the sake of clarity, we will call the "proton transfer step" and the "carbon configuration inversion step." Few atoms are expected to rearrange during the proton transfer step, whose transition vector mainly includes the motion of the transferring proton and a small contribution of the proton donor oxygen atom of Glu317. That is, the changes throughout this step are restricted to a small local zone involving a very reduced number of atoms. Conversely, many atoms are involved in the components of the transition vector corresponding to the carbon configuration inversion step: the stereogenic carbon atom and the atoms of propargylglycolate directly attached to it, along with the proton of His297 that is transferred to the substrate. In addition, some participation of Lys166 at the first stages of the step is also expected. This step, then, involves a global motion of several groups and residues, including an important number of atoms.

We have chosen the PM3 semiempirical Hamiltonian, but with Austin Model 1–specific-reaction parameter (AM1-SRP) parameters for magnesium atom, to define the QM zone. The zone contains the substrate, the magnesium atom, a water coordinated to it, and part of the lateral chains of all the residues shown in Figure 2 (see Ref. [38] for more details). This means 80 atoms, including hydrogen link atoms. The rest of the enzyme and solvation waters are treated with the AMBER united atom force field. Then a total of 3962 atoms including link atoms constitute the whole QM/MM model. Only 1298 atoms (within a sphere of 15 Å around the magnesium atom) are moved during the location of the stationary points.

In the current work we have looked for the TS structures of the two above-mentioned steps, starting from the highest energy point of the profile built up along an adequate reaction coordinate. We preferred to test our algorithm on the location of TS rather than on minima. The reason is that although the algorithm is applicable in both cases, the TS search is always more problematic, and this will help us to discriminate and discuss between the several options considered here.

All the results presented in the current work have been done with our own source code implemented for the energy and gradient evaluations in the QM/MM module (ROAR 1.2 [40]) of AMBER package [41].

Here we compare the results obtained with the microiterative method using the different options mentioned above. We must stress that some uncertainty accompanies the quantitative value of most of the results obtained in the current work. The number of iterations required to reach the TS structures is an example. The convergence of each minimization of the environment or each core search is fulfilled when the suitable convergence criterion of the gradient norm is reached. However, sometimes the low-gradient region can be attained quickly, but the algorithm can use some time in the quasiconverged region, conferring some aleatory character to the crossing of the gradient norm threshold [root mean square (RMS) = 0.005 kcal/(mol · Å)]. Indeed, this fact also affects the total central processing unit (CPU) time spent in each TS structure location. Nevertheless, we think that several important qualitative trends emerge from the analysis of our numerical results. The tests were run in a 2.0-Ghz Pentium IV computer.

CORE/ENVIRONMENT SIZE

We carried out the series of tests corresponding to the proton transfer step when a full TS search in

TABLE I

Results testing the core/environment size for the proton transfer step (full TS search).

Core/ environment	Core/environment iterations	Energy (kcal/mol)	Total CPU time (s)	Hess CPU time (s)	% Hess time
3/1295	(186/1859)	−7453.26	17772	150	1
12/1286	(219/2056)	−7453.26	20317	604	3
23/1275	(172/725)	−7453.27	8994	1158	13
80/1218	(174/479)	−7453.28	9791	4017	41
138/1160	(206/765)	−7453.27	15551	6989	45
194/1104	(187/596)	−7453.28	17084	10002	59
399/899	(205/530)	−7453.28	27926	20263	73

the core until convergence was performed, before again minimizing the environment. Each test corresponds to a microiterative search of the TS structure using a particular core/environment distribution of the 1298 moving atoms. The results are presented in Table I. From left to right, the different columns indicate, respectively, the number of atoms included in the core/environment regions, the number of iterations made in each zone, the final energy of the located TS structure, the total CPU time of the location, the CPU time devoted to calculate the initial Hessian, and its percentage over the total CPU time.

The first test includes three atoms in the core: the shifting proton and both the proton donor and acceptor oxygen atoms. The successives tests progressively increase the number of atoms around the first three that are incorporated in the core region. The third column shows that the final energy always has the same value, which indicates that, in this case, whatever core/environment partition that includes the three atoms directly involved in the proton transfer leads to the right TS structure. However, each test behaves in a different way, based on the number of iterations and the CPU time. When the core is small a lot of environment iterations are required, because of the huge coupling between the two regions and the great size of the environment region. Conversely, the coupling is low when the core is big (which in addition, implies a small environment region), significantly reducing the number of environment iterations. On the other hand, it seems that there exists a range of intermediate core sizes involving a reduced number of core iterations. This probably results from a compromise between a lower core/environment coupling and the efficiency of the RFO method in handling a progressively increasing number of core atoms as the size of the core grows. The last results show clearly that when the core/environment partition is selected in an adequate way, the part of the Hessian matrix of the entire system corresponding to the core–core and environment–environment diagonal blocks are numerically much more important than the core–environment nondiagonal ones. A Hessian of this type is the most suitable to be used for any type of optimization [42].

As for the CPU time, there is a good correlation between the total number of iterations and the difference between the total CPU time and the time devoted to the calculation of the initial Hessian. So, for instance, the core/environment partition 80/1218 converges with the smallest number of total iterations (653) taking the least CPU time (5774 s), excluding the Hessian calculation. On the other hand, as expected, the CPU time corresponding to the calculation of the initial Hessian monotonically increases very quickly with the core size, requiring the 1% of the total CPU time when just three atoms are included in the core but as much as the 73% for a core with 399 atoms. Considering all the factors described above, it can be easily understood why there exists an interval of medium core sizes that minimizes the total CPU time, as seen in the fourth column, the optimal partition being 23/1275 (within the discrete, limited series studied here).

Exhibited in Table II are the results of the series of tests corresponding to the carbon configuration inversion step when one performs a full TS search in the core until convergence before again minimizing the environment. The first test has seven atoms in the core, including the stereogenic carbon atom and several atoms of His297 and Lys166. As for the number of iterations and the CPU time, the results follow the same trends as the proton transfer step (Table I). The difference lies on the final energies. It

TABLE II

Results testing the core/environment size for the carbon configuration inversion step (full TS search).

Core/ environment	Core/environment iterations	Energy (kcal/mol)	Total CPU time (s)	Hess CPU time (s)	% Hess time
7/1291	(145/2061)	−7443.25	19041	346	2
17/1281	(150/1965)	−7442.31	18677	840	4
23/1275	(140/1862)	−7442.31	18172	1136	6
34/1264	(102/1524)	−7442.32	15453	1686	11
80/1218	(113/1827)	−7442.34	20401	3978	19
138/1160	(62/731)	−7440.50	17809	9025	51
194/1104	(102/891)	−7440.49	23960	12826	54
643/655	(291/508)	−7440.53	58303	42504	73

can be seen that three distinct sets of energies are obtained, with a difference of roughly 3 kcal/mol between the test with seven core atoms and the tests of 138 or more core atoms. This indicates that three different progressive approximations to the TS structure are located, a number of 138 or more core atoms (in this discrete, limited series) being required to converge to the right TS structure as the core size increases. (We consider that the right TS structure of the corresponding potential energy valley is the one that would be obtained including all the 1298 moving atoms in the core).

This fact highly contrasts with the proton transfer step where a core of three atoms is sufficient. This is a consequence of the global character of the changes involved in the carbon configuration inversion step (in front of the local character of the proton transfer step), whose TS search requires a core including all the atoms that participate significantly in the reaction coordinate to work well. Then, at first glance, it would seem that the best core/environment partition could be 34/1264, which requires 15453 s of total CPU time. However, this partition does not lead to a good enough approximation to the right TS structure. The optimal partition for this step is, instead, 138/1160, which takes more total CPU time, (17809 s) because of the bigger CPU time required to calculate the initial Hessian, although it clearly involves fewer iterations and takes less time in the location of the TS structure once the initial Hessian has been calculated. Indeed, this partition leads to the right TS. That is, as the different atoms that define the reaction coordinate are incorporated in the core, the description of the TS of the corresponding potential energy valley is progressively improved, until a good enough approximation to the right TS is reached.

Finally, it must be mentioned that the high total CPU time of the test corresponding to the partition 643/655 is due not only to the CPU time required to calculate the initial Hessian but also to the time employed to partially diagonalize this big matrix Hessian at each iteration.

CORE TS SEARCH/ENVIRONMENT MINIMIZATION SWITCH

The results of the series of tests corresponding to the proton transfer and the carbon configuration inversion steps, when a full minimization of the environment is carried out at every TS search step in the core, are given in Tables III and IV, respectively. Compared with Tables I and II, it can be observed that the minimization of the environment at every TS search step in the core clearly reduces the number of core iterations, although it in turn tends to augment the number of environment iterations. As a result, the difference between both strategies does not seem to be very significant, al-

TABLE III

Results testing the core TS search/environment minimization switch for the proton transfer step.

Core/ environment	Core/environment iterations	Energy (kcal/mol)
3/1295	(96/1896)	−7453.26
12/1286	(105/2075)	−7453.25
23/1275	(42/1211)	−7453.27
80/1218	(46/976)	−7453.29
138/1160	(105/1088)	−7453.27
194/1104	(62/752)	−7453.27
399/899	(80/793)	−7453.28

TABLE IV

Results testing the core TS search/environment minimization switch for the carbon configuration inversion step.

Core/ environment	Core/environment iterations	Energy (kcal/mol)
7/1291	(88/2092)	−7443.25
17/1281	(143/1923)	−7442.31
23/1275	(115/1908)	−7442.31
34/1264	(66/1631)	−7442.32
80/1218	(69/1861)	−7442.33
138/1160	(52/758)	−7440.50
194/1104	(73/920)	−7440.50
643/655	(117/710)	−7440.54

though the minimization of the environment at every TS search step tends to increase the total number of iterations in the two reaction steps studied here. As expected, the energy column shows that the final TS structure for each core/environment partition is independent of the strategy used.

In addition, our experience shows that since determining the location of TS points is complicated, the strategy of a full TS search before the next minimization is recommended to ensure at a primary stage of the search that the final convergence will be reached. On the contrary, when the minimization is performed at every TS search step, we have no information about the desired success of the search until the process is completed.

CORE/ENVIRONMENT INTERACTION

ESP/MM Results

We have tested the ESP/MM approach during the environment minimization in the proton transfer step using different core/environment partitions. We have performed a full TS search in the core until convergence before again minimizing the environment. Unfortunately, most of the calculations are unable to reach convergence. The few that do converge require a huge amount of iterations and total CPU time. For instance, for the partition 138/1160, the ESP/MM approach requires 947/4942 core/environment iterations and 52690 s of total CPU time, as opposed to 206/765 iterations and 15551 s when the QM/MM calculations are carried out. The problems in the convergence of the microiterative method are probably due to the fact that the TS search in the core and the environment minimization follow different PES.

TABLE V

Results testing the QM(1SCF)/MM approach for the proton transfer step (full TS search).

Core/ environment	Core/ environment iterations	Energy (kcal/mol)	Total CPU time (s)
80/1218	(184/557)	−7453.28	8677
140/1158	(164/658)	−7453.28	12005
194/1104	(183/616)	−7453.27	14776
399/899	(190/525)	−7453.28	26250
643/655	(229/509)	−7453.29	44321

QM(1SCF)/MM Results

We present now the results of the series of tests in which only one SCF cycle was performed to approximate the QM/MM energy during the minimization of the environment. We performed a full TS search in the core until convergence before again minimizing the environment. The cases corresponding to the proton transfer and carbon configuration inversion steps are given in Tables V and VI, respectively. Although the 1SCF option has no definite effect on the total number of iterations, it is clear that the total CPU time is noticeably smaller. Indeed, this is due to the fact that each 1SCF energy evaluation during the environment minimization is faster than a complete QM/MM energy calculation. On the other hand, the energy column shows that the 1SCF option leads to the same TS structures as the complete QM/MM calculations. The results confirm that the 1SCF approach is reliable and pro-

TABLE VI

Results testing the QM(1SCF)/MM approach for the carbon configuration inversion step (full TS search).

Core/ environment	Core/ environment iterations	Energy (kcal/mol)	Total CPU time (s)
80/1218	(426/2426)	−7442.34	19842
140/1158	(84/839)	−7440.50	12407
194/1104	(105/841)	−7440.49	15525
643/655	(327/545)	−7440.54	47773

vides a cheaper alternative to be used in the microiterative method.

Conclusions

In the current paper we have studied how the efficiency of the microiterative method for locating TS structures in QM/MM PES of very high dimensionality can be optimized. We have run several series of calculations testing different options on the PES corresponding to two of the reaction steps of the mechanisms by which mandelate racemase enzyme catalyzes the reversible interconversion of the (S)- and (R)-enantiomers of the substrate propargylglycolate. We have chosen the PM3 semiempirical Hamiltonian (AM1-SRP for magnesium atom) to define the quantum part that involves 80 atoms, including hydrogen link atoms. In all, 3962 atoms constitute the whole QM/MM model, 1298 of which are moved during the location of the TS structures. The microiterative method divides the whole system in two parts: a core zone in which an accurate second-order search (a rational function optimization method in this paper) of the TS structure is done, and an environment that is kept minimized with a cheap first-order method (an L-BFGS method in this case). Our results show that the core must include at least all the atoms that participate significantly in the reaction coordinate of the corresponding reaction step. Otherwise, the right TS structures are not reached. Indeed, the threshold size of the core depends a lot on each particular reaction step. Above the minimum core size, there is an interval of medium core sizes that minimizes the total CPU time. This arises from a compromise among a lower core/environment coupling, the efficiency of the rational function optimization method in handling a progressively increasing number of core atoms, and the monotone augmentation of the CPU time required to calculate the initial Hessian matrix, as the core size grows. As a consequence, the use of a core as great as possible is not advised. In other words, above that threshold size of the core, which depends on the relevant motions taking place during the corresponding chemical step, it is not true that the larger is the core size, the more efficient is the TS search.

On the other hand, a considerable amount of CPU time can be saved if only one SCF cycle is performed to evaluate the potential energy during the environment minimization. This option is clearly faster than a complete QM/MM energy calculation and leads to the right TS structures. Conversely, the use of the ESP charges to simplify the calculation of the interaction energy between the QM and the MM regions seems to be less efficient. Finally, we have to remark that the location of the TS structures in enzyme catalysis is needed to define reliable reaction pathways along which a set of generalized free energies can be calculated. The corresponding generalized free energy barrier determines the canonical reaction rate constant according to the Variational Transition State Theory. In this sense, extreme accuracy in the geometrical or energetic parameters of the TS structures is not required. We require only that the right pathway be built up, starting from the TS structure, thus avoiding the use of reaction paths that run on wrong potential energy valleys.

ACKNOWLEDGMENTS

We are grateful for financial support from the Spanish "Ministerio de Ciencia y Tecnología" and the "Fondo Europeo de Desarrollo Regional," through project numbers BQU2002-00301 and BQU2002-00293.

References

1. Elber, R.; Ghosh, A.; Cárdenas, A. Acc Chem Res 2002, 35, 396–403.
2. Schlick, T.; Skeel, R.; Brunger, A. T.; Kalé, L. V.; Board, J. A.; Hermans, J.; Schulten, K. J Comput Phys 1999, 151, 9–48.
3. Schlegel, H. B. J Comput Chem 2003, 24, 1514–1527.
4. Elber, R.; Shalloway, D. J Chem Phys 2000, 112, 5539–5545.
5. Crehuet, R.; Field, M. J. J Chem Phys 2003, 118, 9563–9571.
6. Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Annu Rev Phys Chem 2002, 53, 291–318.
7. Warshel, A.; Levitt, M. J Mol Biol 1976, 103, 227–249.
8. Singh, U. C.; Kollman, P. A. J Comput Chem 1986, 7, 718–730.
9. Field, M. J.; Bash, P. A.; Karplus, M. J Comput Chem 1990, 11, 700–733.
10. Goedecker, S. Rev Mod Phys 1999, 71, 1085–1123.
11. Titmuss, S. G.; Cummins, P. L.; Rendell, A. P.; Bliznyuk, A. A.; Gready, J. E. J Comput Chem 2002, 23, 1314–1322.
12. Fletcher, R. Practical Methods of Optimization, 2nd Edition; John Wiley & Sons: Tiptree, 1987.
13. Monard, G.; Prat-Resina, X.; González-Lafont, A.; Lluch, J. M. Int J Quantum Chem 2003, 93, 229–244.
14. Henkelman, G.; Jóhannesson, G.; Jónsson, H. In Progress on Theoretical Chemistry and Physics; Schwartz, S. D., Ed.; Kluwer Academic Publishers: New York 2000; p 269.

15. Maseras, F.; Morokuma, K. J Comput Chem 1995, 16, 1170–1179.
16. Turner, A. J.; Moliner, V.; Williams, I. H. Phys Chem Chem Phys 1999, 1, 1323–1331.
17. Hall, R. J.; Hindle, S. A.; Burton, N. A.; Hillier, I. H. J Comput Chem 2000, 21, 1433–1441.
18. Billeter, S. R.; Turner, A. J.; Thiel, W. Phys Chem Chem Phys 2000, 2, 2177–2186.
19. Zhang, Y.; Liu, H.; Yang, W. J Chem Phys 2000, 112, 3483–3492.
20. Murphy, R.; Philipp, D.; Friesner, R. Chem Phys Lett 2000, 321, 113–120.
21. Vreven, T.; Morokuma, K.; Farkas, Ö.; Schlegel, H. B.; Frisch, M. J. J Comput Chem 2003, 24, 760–769.
22. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187–217.
23. Derremaux, P.; Zhang, G.; Schlick, T.; Brooks, B. J Comput Chem 1994, 15, 532–552.
24. Liu, D. C.; Nocedal, J. Math Program 1989, 45, 503–528.
25. Das, B.; Meirovitch, H.; Navon, I. M. J Comput Chem 2003, 24, 1222–1231.
26. Prat-Resina, X.; González-Lafont, A.; Lluch, J. M. 2003, 632, 297–307.
27. Banerjee, A.; Adams, N.; Simons, J.; Shepard, R. J Phys Chem 1985, 89, 52–57.
28. Besalú, E.; Bofill, J. M. Theor Chem Acc 1998, 100, 265–274.
29. Prat-Resina, X.; Garcia-Viloca, M.; Monard, G.; González-Lafont, A.; Lluch, J. M.; Anglada, J. M.; Bofill, J. M. Theor Chem Acc 2002, 107, 147–153.
30. Gao, J. J Phys Chem 1992, 96, 537–540.
31. Tuñón, I.; Martins-Costa, M.; Millot, C.; Ruiz-López, M.; Rivail, J. J Comput Chem 1996, 17, 19–29.
32. Truong, T. N.; Stefanovich, E. V. Chem Phys Lett 1996, 256, 348–352.
33. Cubero, E.; Luque, F. J.; Orozco, M.; Gao, J. J Phys Chem B 2003, 107, 1664–1671.
34. Besler, B. H.; Merz, K. M. Jr.; Kollman, P. A. J Comput Chem 1990, 11, 431–439.
35. Evans, T. J.; Truong, T. N. J Comput Chem 1998, 19, 1632–1638.
36. Anglada, J. M.; Besalú, E.; Bofill, J. M.; Rubio, J. J Math Chem 1999, 25, 85–92.
37. Eckert, F.; Pulay, P.; Werner, H.-J. J Comput Chem 1997, 18, 1473–1483.
38. Garcia-Viloca, M.; González-Lafont, A.; Lluch, J. M. J Am Chem Soc 2001, 123, 709–721.
39. Prat-Resina, X.; Garcia-Viloca, M.; González-Lafont, A.; Lluch, J. M. Phys Chem Chem Phys 2002, 4, 5365–5371.
40. Cheng, A.; Stanton, R. S.; Vincent, J. J.; van der Vaart, A.; Damodaran, K. V.; Dixon, S. L.; Hartsough, D. S.; Mori, M.; Best, S. A.; Monard, G.; Garcia-Viloca, M.; Zant, L. C. V.; Merz, K. M. Jr. ROAR 2.0. The Pennsylvania State University, 1999.
41. Pearlman, D.; Case, D.; Caldwell, J.; Ross, W.; Cheatham, T.; Debolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. Comp Phys Comm 1995, 91, 1–41.
42. Bell, S.; Crighton, J. S.; Fletcher, R. Chem Phys Lett 1981, 82, 122–126.