JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

FIREBALL/AMBER: An Efficient Local-Orbital DFT QM/MM Method for Biomolecular Systems

Jesús I. Mendieta-Moreno,[†,‡,¶] Ross C. Walker,[§] James P. Lewis,[∥] Paulino Gómez-Puertas,*[,‡] Jesús Mendieta,[‡,¶,⊥] and José Ortega*[,†,⊥]

[†]Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, ES-28049 Madrid, Spain

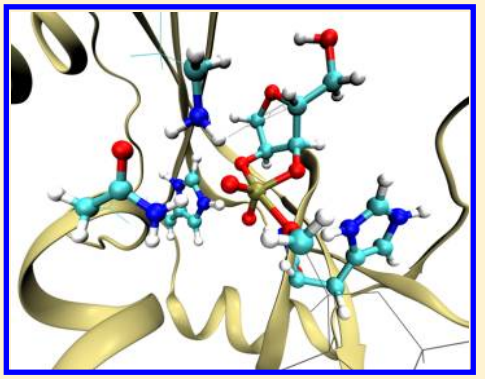
[‡]Molecular Modelling Group, Center of Molecular Biology "Severo Ochoa" (CSIC- UAM), ES-28049 Madrid, Spain

[¶]Biomol-Informatics SL, Campus UAM, ES-28049 Madrid, Spain

[§]San Diego Supercomputer Center and Department of Chemistry and Biochemistry, University of California—San Diego, La Jolla, California 92093, United States

[∥]Department of Physics, West Virginia University, Morgantown, West Virginia 26506-6315, United States

ABSTRACT: In recent years, quantum mechanics/molecular mechanics (QM/MM) methods have become an important computational tool for the study of chemical reactions and other processes in biomolecular systems. In the QM/MM technique, the active region is described by means of QM calculations, while the remainder of the system is described using a MM approach. Because of the complexity of biomolecules and the desire to achieve converged sampling, it is important that the QM method presents a good balance between accuracy and computational efficiency. Here, we report on the implementation of a QM/MM technique that combines a DFT approach specially designed for the study of complex systems using first-principles molecular dynamics simulations (FIREBALL) with the AMBER force fields and simulation programs. We also present examples of the application of this QM/MM approach to three representative biomolecular systems: the analysis of the effect of electrostatic embedding in the behavior of a salt bridge between an aspartic acid and a lysine residue, a study of the intermediate states for the triosephosphate isomerase catalyzed conversion of dihydroxyacetone phosphate into glyceraldehyde 3-phosphate, and the detailed description, using DFT QM/MM molecular dynamics, of the cleavage of a phosphodiester bond in RNA catalyzed by the enzyme RNase A.



1. INTRODUCTION

Quantum molecular-dynamics (MD) techniques have become a powerful tool for investigating the dynamical and electronic processes that take place at the atomic scale.[1,2] In these techniques, the atoms move along classical trajectories, but the forces on the atoms are obtained from the quantum electronic structure of the system. The continual improvement of these techniques and the sustained increases in available computational power are allowing the application of these "atomic microscopes" to increasingly more-complex problems and materials.

Biomolecular systems represent one of the most complex classes of materials. Their atomic-scale simulation usually involves 10^5−10^6 atoms with simulation times requiring nanoseconds to microseconds (10^9 integration time steps) to obtain meaningful data. Because of this complexity, molecular mechanics (MM) has traditionally been the method of choice to simulate biomolecules at the atomic scale. These methods use empirical force fields and are very computationally efficient, allowing simulation of the time evolution of a system containing 10^6 or more atoms. Thus, processes such as large-scale conformational changes[3] or direct ligand binding[4] can be studied with empirical force fields. However, the bonding forces are represented using simple (e.g., harmonic) potentials, which cannot be used to study electronic structure processes, such as enzyme reactions, in which the chemical bonds of the material are changing: the chemical bonds are intrinsically related to the quantum electronic structure of the material. Quantum-mechanics (QM) techniques are required to properly describe chemical reactions and other processes. Unfortunately, QM methods can only be used to simulate systems limited to a few hundred atoms. In order to simulate biomolecular systems, we need to combine QM and classical force fields MM methods, creating a hybrid QM/MM approach.[5,6] In these QM/MM methods, the QM approach is used to calculate the active region (the part of the system where the chemical activity is taking place, e.g., the reactant molecules and catalytic site residues) while the MM method is used to describe the environment (the remainder of the system).

Among other things, the accuracy of the QM/MM calculation is related to the level of theory used in the QM

Received: January 15, 2014
Published: April 22, 2014

   dx.doi.org/10.1021/ct500033w | J. Chem. Theory Comput. 2014, 10, 2185−2193

calculation; however, using high-level quantum-chemistry methods greatly increases the computational cost of the calculation. Thus, the use of *ab initio* methods prevents an exhaustive study of the conformational space of the reaction;[7] these high-level QM/MM techniques have been largely limited to searching for ground-state configurations (i.e., minimization of the energy). In most of the applications, however, it is critical to properly analyze the configurational space to understand the mechanism of the reaction. Utilizing low-level semiempirical QM methods will enable techniques that explore the mean force potential for the conformational space,[8] but at the price of reduced accuracy. Therefore, it is important for the QM techniques to present a proper balance between accuracy and computational efficiency. In this regard, density functional theory (DFT) is probably the most convenient methodology to perform quantum MD simulations for relatively large QM regions (10^2 atoms) within QM/MM simulations.

Here, we discuss the implementation of a QM/MM method based on the combination of FIREBALL[9−13] (QM) and AMBER[14,15] (MM). FIREBALL is a DFT MD technique that presents important advantages for QM/MM implementations, because of both its excellent balance between computational efficiency and accuracy and its perfect adaptation to be used in combination with MM methods. The main features of the FIREBALL technique and of its implementation in the AMBER suite of programs are presented in section 2. In section 3, we present three examples of the application of this QM/MM approach: the effect of electrostatic embedding in the behavior of a salt bridge and the analysis of two biomolecular reactions catalyzed by the enzymes triosephosphate isomerase (TIM) and ribonuclease A (RNase A). In the case of RNase A, the computational efficiency of the method allowed the generation of a free-energy surface map of the reaction (similar to those previously published for other proteins using the semiempirical PM3 Hamiltonian[8,16]), thus permitting a detailed analysis of alternative pathways.

2. QM/MM: FIREBALL/AMBER

2.1. QM: FIREBALL. For the QM calculation, we use a DFT technique (FIREBALL[9−13]) that is specifically designed for the study of complex systems using QM/MD simulations. The main characteristics of FIREBALL are (i) basis sets of short-range numerical atomic-like orbitals;[13,17] (ii) the technique is completely formulated in *real space* (no need for supercells) and does not require the use of grids for the charge density and potential calculations; (iii) self-consistency is achieved using the orbital occupation numbers;[12] (iv) a practical tabulation-interpolation scheme is used so that all the interactions/integrals required during the MD simulation can be calculated beforehand and stored in data tables,[13] speeding up the calculations.

In this approach, the energy functional is written as

$$E_{\mathrm{TOT}}[\rho(\vec{r})] = E_{\mathrm{BS}} - E_{\mathrm{ee}}[\rho(\vec{r})] + E_{\mathrm{xc}}[\rho(\vec{r})] - \int \rho(\vec{r}) V_{\mathrm{xc}}[\rho(\vec{r})]\, \mathrm{d}\vec{r} + E_{\mathrm{ion-ion}} \tag{1}$$

where $\rho(\vec{r})$ is the *input* electron density, which is determined in a self-consistent way in terms of the orbital occupation numbers[12,18,19] (see below). E_{BS} is a sum over occupied eigenstates, of the effective one-electron Hamiltonian,

$$\left(-\frac{1}{2}\nabla^2 + V[\vec{r};\rho]\right)\psi_n = \varepsilon_n \psi_n \tag{2}$$

$$E_{\mathrm{BS}} = \sum_n \varepsilon_n \tag{3}$$

The potential V in eq 2 is the sum of the ionic potential, $V_{\mathrm{ion}}(\vec{r})$, (typically represented by a pseudo-potential), the electrostatic potential due to the electron density $\rho(\vec{r})$, and the exchange-correlation potential V_{xc}:

$$V[\vec{r};\rho] = V_{\mathrm{ion}}(\vec{r}) + \int \frac{\rho(\vec{r'})\, \mathrm{d}\vec{r}}{|\vec{r} - \vec{r}'|} + V_{\mathrm{xc}}[\vec{r};\rho] \tag{4}$$

In eq 1, E_{ee} is an average electron−electron energy,

$$E_{\mathrm{ee}}[\rho(\vec{r})] = \frac{1}{2}\int\int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}'|}\, \mathrm{d}\vec{r}\, \mathrm{d}\vec{r}' \tag{5}$$

$E_{\mathrm{ion-ion}}$ is the ion−ion interaction energy

$$E_{\mathrm{ion-ion}} = \frac{1}{2}\sum_{i,j} \frac{Z_i Z_j}{|\vec{R}_i - \vec{R}_j|} \tag{6}$$

(Z_i is the nuclear or pseudo-potential charge on atom i at position $\vec{R}_i$), and $E_{\mathrm{xc}}[\rho]$ is the exchange-correlation energy.

In the FIREBALL method, optimized atomic-like orbitals, $\phi_\mu(\vec{r})$, are used as the basis set to solve eq 2 and to determine the density $\rho(\vec{r})$ ($\mu \equiv (i,l,m)$, i is the atomic site, l represents the atomic subshell—e.g., 3*s*, 4*s*, 3*p*, 3*d*, etc.—and m is the magnetic quantum number); these orbitals are strictly localized in real space.[13,17] The computational efficiency of FIREBALL is associated with the possibility to choose $\rho(\vec{r})$ in the above equations as a sum of atomic-like densities, $\rho_i(\vec{r})$.[12] In particular, in the FIREBALL approach, the electron density $\rho(\vec{r})$ is written in terms of the atomic-like orbitals, $\phi_\mu(\vec{r})$, as

$$\rho(\vec{r}) = \sum_i \rho_i(\vec{r}) = \sum_\mu n_\mu |\phi_\mu(\vec{r})|^2 \tag{7}$$

where n_μ represents the charge on the orbital ϕ_μ. In this way, four-center integrals are not required for the calculation of E_{TOT} or the forces $\vec{F}_i$, and all the two- and three-center interactions are tabulated beforehand and placed in interpolation data tables, which are no larger than two-dimensional.[9,13] All the matrix elements (integrals) required during the MD simulation are evaluated by looking up the necessary information from the data tables (which are read at the beginning of the calculation and stored in memory throughout the MD simulation).

In practice, the atomic densities ρ_i are approximated to be spherically symmetric around each atomic site i (i.e., $n_{ilm} = n_{ilm'}$). Self-consistency is achieved by defining output orbital charges n_μ^{out} from the occupied eigenvectors ψ_n of eq 2); and imposing that, in the self-consistent solution, n_μ^{out} and the input orbital charges n_μ coincide.[12] Different methods can be used to project the output electron density from eq 2 into the form given by eq 7, using, for example, Löwdin orbitals[12,20,21] or natural atomic orbitals.[22,23]

QM MD simulations can be performed once the forces

$$\vec{F}_i = -\frac{\partial E_{\mathrm{TOT}}}{\partial \vec{R}_i} \tag{8}$$

on each atom i are evaluated.[13]

2.2. Implementation of FIREBALL in AMBER. The implementation of FIREBALL in AMBER is based on the work of Walker

et al.[24] for the AMBER MD engine *sander*. In QM/MM methods, the system is divided into a chemically active region (A) and the environment (E). In the frontier between these two subsystems, there is a region where the QM and MM calculations are modified in some way. In particular, this region may contain additional atoms (e.g., *link* atoms), or atoms with special properties; the mission of these atoms is to saturate the broken bonds that appear in the QM atoms at the frontier with the MM region (E).

In our QM/MM implementation, the total energy of the system is

$$E_{\mathrm{TOT}} = E_{\mathrm{QM}}(A) + E_{\mathrm{MM}}(E) + E_{\mathrm{QM-MM}}(A, E) \tag{9}$$

where $E_{\mathrm{QM}}(A)$ is the QM (FIREBALL) total energy for subsystem A (including the additional atoms, e.g., the link atoms), $E_{\mathrm{MM}}(E)$ is the MM energy for subsystem E and $E_{\mathrm{QM-MM}}(A,E)$ represents the interaction between the two subsystems. In MM methods, the energy of the system is described in terms of a classical potential energy (the force field); this potential energy usually contains bond terms, nonbonded terms, and electrostatic terms. There are several force field approaches such as AMBER,[25] CHARMM,[26] GROMACS,[27] GROMOS,[28] OPLS-AA,[29] with multiple parametrizations of each for biomolecular systems.

In similarity to the MM energy, the coupling term, $E_{\mathrm{QM-MM}}(A,E)$ also contains bonded, nonbonded, and electrostatic interactions. Regarding the bonded term, there are many cases in which the frontier between the QM (A) and MM (E) regions goes through covalent bonds, resulting in broken bonds that need to be properly saturated in the QM calculation. There are several approaches to deal with this problem (e.g., see Senn and Thiel[30]). In the present QM/MM implementation, we have adopted the link atom approach, as discussed in Walker et al.[24] Also, the nonbonded interactions between QM and MM atoms are described using the Lennard-Jones parameters from the MM force field.

In biomolecular systems, the effect of the electrostatic potential due to the environment on the active region is essential, since it may modify the physical−chemical properties of the active site. Thus, in our QM/MM implementation, we consider the electrostatic interaction between the active region and the environment at the QM level, including, in the QM Hamiltonian, the electrostatic potential due to the partial charges on the MM atoms (electrostatic embedding).

The electrostatic interaction between the partial charges q_k on the MM atoms and the QM subsystem presents two contributions: the interaction with the QM nuclear (or pseudo-potential) charges, Z_i (see below), and the interaction with the QM electron cloud. The electrostatic potential due to the environment E changes the electronic structure of the QM region (i.e., the QM Hamiltonian); in particular, the MM charges q_k yield the following contributions to the QM Hamiltonian matrix element between orbitals α and β:

$$H^{E}_{\alpha,\beta} = \sum_k h^k_{\alpha\beta} \tag{10}$$

$$h^k_{\alpha\beta} = q_k \int \frac{\phi_\alpha(\vec{r} - \vec{R}_\alpha)\phi_\beta(\vec{r} - \vec{R}_\beta)}{|\vec{r} - \vec{R}_k|}\, \mathrm{d}\vec{r} \tag{11}$$

where $\phi_\alpha(\vec{r}-\vec{R}_\alpha)$ and $\phi_\beta(\vec{r} - \vec{R}_\beta)$ are the basis set orbitals α and β centered on the atoms placed at $\vec{R}_\alpha$ and $\vec{R}_\beta$. The contribution $h^k_{\alpha\beta}$ represents the electrostatic interaction between the overlap charge $\phi_\alpha(\vec{r} - \vec{R}_\alpha)\phi_\beta(\vec{r} - \vec{R}_\beta)$ and a point charge q_k at $\vec{R}_k$. In order to include these interactions in a practical way in the QM calculation, we use the following approximation:[12]

$$h^k_{\alpha\beta} = q_k \left(\frac{g_\alpha}{|\vec{R}_k - \vec{R}_\alpha|} + \frac{g_\beta}{|\vec{R}_k - \vec{R}_\beta|} \right) \tag{12}$$

with *effective overlap charges* g_α, g_β

$$g_\alpha = \frac{1}{2} S_{\alpha\beta} - \frac{p_{\alpha\beta}}{|\vec{R}_\alpha - \vec{R}_\beta|} \tag{13}$$

$$g_\beta = \frac{1}{2} S_{\alpha\beta} + \frac{p_{\alpha\beta}}{|\vec{R}_\alpha - \vec{R}_\beta|} \tag{14}$$

which includes monopole and dipole far-field effects; this is the same approximation as used in the FIREBALL code to include long-range contributions.[12] In eqs 13 and 14, $S_{\alpha\beta}$ is the overlap:

$$S_{\alpha\beta} = \int \phi_\alpha(\vec{r} - \vec{R}_\alpha)\phi_\beta(\vec{r} - \vec{R}_\beta)\, \mathrm{d}\vec{r} \tag{15}$$

and $p_{\alpha\beta}$ is the component along the $(\vec{R}_\beta - \vec{R}_\alpha)$ direction of the dipole moment $\vec{P}_{\alpha\beta}$, with respect to the midpoint, $\vec{R}_m$, between atoms α and β:

$$\vec{P}_{\alpha\beta} = \int [\vec{r} - \vec{R}_m] \phi_\alpha(\vec{r} - \vec{R}_\alpha)\phi_\beta(\vec{r} - \vec{R}_\beta)\, \mathrm{d}\vec{r} \tag{16}$$

$(\vec{R}_m = (\vec{R}_\beta + \vec{R}_\alpha)/2)$. Thus, the electrostatic potential due to the environment is taken into account through the addition of these contributions (eq 10) to the QM Hamiltonian, changing the electronic structure of the system.

The electrostatic potential due to the environment also interacts with the nuclear (or pseudo-potential) charges Z_i of the QM atoms, yielding the following contribution to the $E_{\mathrm{QM-MM}}$ energy:

$$E^{\mathrm{nuc}}_{\mathrm{QM/MM}} = \sum_k \sum_i \frac{q_k Z_i}{|\vec{R}_k - \vec{R}_i|} \tag{17}$$

where q_k are the partial charges on the MM atoms.

In order to handle these long-range electrostatic interactions (eqs 10 and 17), we use periodic boundary conditions, as described in Walker et al.,[24] either using the Ewald or Particle Mesh Ewald (PME) methods. In particular, in the calculations described below, we have used the PME method with a cutoff of 10 Å for both the MM-MM and QM-MM interactions.

An important feature of FIREBALL/AMBER when compared to other QM/MM implementations is the optimized calculation time. The calculation time is of the same order of magnitude to that obtained with QM/MM methods based on semiempirical Hamiltonians (e.g., PM3[31] or DFTB[32]), thus allowing the analysis of a large conformational space using long QM/MM MD simulations. Finally, we mention that the FIREBALL code is implemented in AMBER as a library for the *sander* MD program. This presents the advantage that all the capabilities of the *sander* program (umbrella sampling, replica exchange, nudged elastic band, targeted MD, steered MD, ...) can be used directly in the QM/MM applications. This implementation will be included in the next major AMBER release.

3. RESULTS

In this section, we present the application of this QM/MM method to three representative examples. In the first one, we check the effect of the water environment on a salt bridge

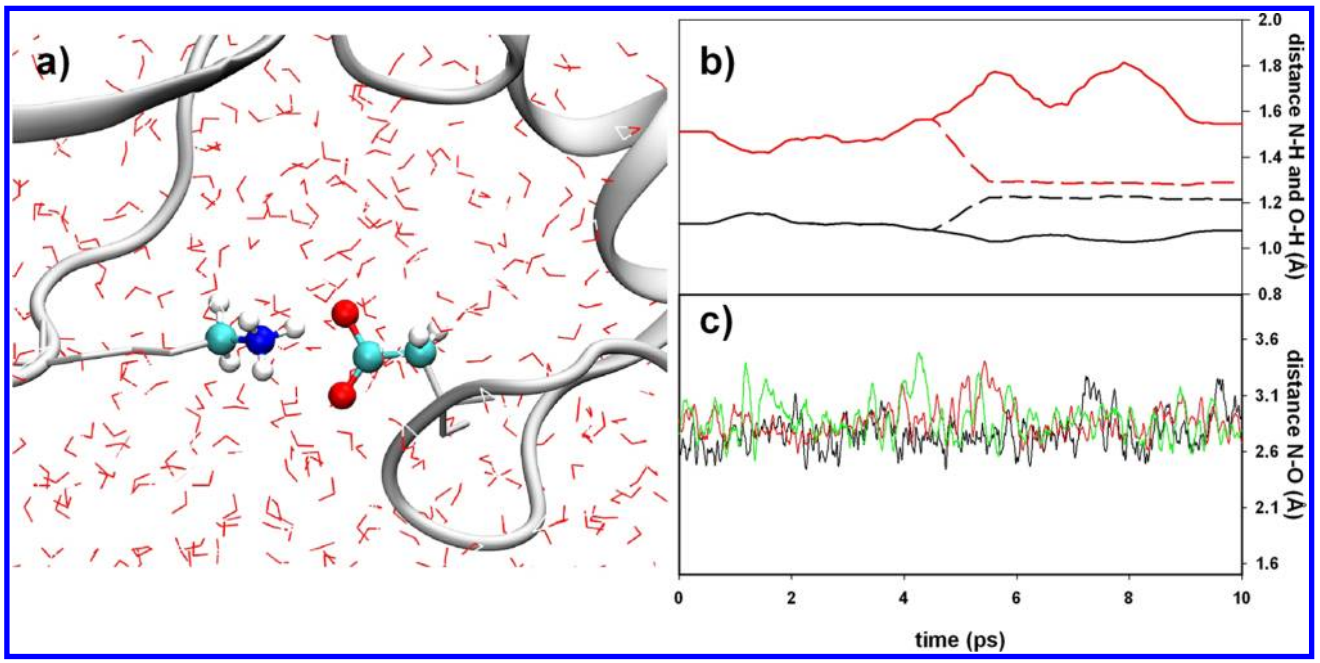


Figure 1. Molecular dynamics simulations for an aspartic–lysine salt bridge: (a) Simulated system: atoms in the QM region are represented as ball and sticks, the protein backbone is represented as gray ribbons and water molecules are represented as red lines. (b) Smoothed distance along the QM/MM molecular dynamics between the N and H atoms (black lines) and the O and H atoms (red lines) forming the salt bridge: electrostatic embedding (solid lines) and mechanical embedding (dashed lines). (c) Distance between the O and N atoms forming the salt bridge for three different calculations: (black line) both the aspartic and lysine residues are included in the QM calculation; (red line) the aspartic is included in the QM region and the lysine is included in the MM region; (green line) the lysine is included in the QM region and the aspartic is included in the MM region.

formed between an aspartic and a lysine residue. This simple example is used to analyze the performance of our electrostatic embedding. In the second example, we analyze the relative energies of different intermediate structures for the TIM-catalyzed conversion of dihydroxyacetone phosphate (DHAP) into glyceraldehyde 3-phosphate (GAP); this system has been investigated by Lennartz et al.[33] using different DFT and *ab initio* QM/MM methods, allowing us a comparison of our results to high-level QM/MM calculations. Finally, in the third example, we illustrate the use of DFT QM/MM molecular dynamics, using FIREBALL/AMBER, to study the conformational space for a well-known reaction: the cleavage of a RNA molecule by the RNase A enzyme.[34]

In these calculations, we use the BLYP exchange-correlation functional (Becke exchange[35] with Lee–Yang–Parr correlation[36]) and an optimized basis set of numerical atomic-like orbitals.[17] In particular, we use a basis set of sp^3 orbitals for the C, N, and O atoms, sp^3d^5 orbitals for P and a single s orbital for H (the basis set optimization will be described elsewhere).[37] Using this basis set (and functional), we performed FIREBALL calculations for the S22 Database,[38] which is specifically designed to test noncovalent interactions in biological molecules. We obtain a mean absolute deviation of 3.9 kcal/mol, to be compared with 3.49 or 2.31 kcal/mol as obtained in DFT calculations[39] using the B3LYP or PBE exchange-correlation functionals (see Table 4 in ref 39).

We should mention that the accuracy of our calculations depends on the choice of basis set (and exchange-correlation functional). The numerical atomic-like orbitals used in the present calculations were optimized to yield a good description of typical covalent bond distances in biomolecules, i.e., noncovalent weak interactions were not taken into account in the optimization of the basis set. It is to be expected that an appropriate basis set optimization, taking also into account noncovalent interactions, should increase the accuracy of these calculations. This will be the subject of future studies.

3.1. Salt Bridge. The interaction between two charged residues usually plays an important role in the structure and function of proteins. The energy of this interaction can be quite large under vacuum,[40] but it is reduced in the presence of other polar molecules (i.e., water or polar residues) in the environment. Thus, we have used a salt bridge formed by Asp and Lys residues as an initial simple test case for our QM/MM implementation, and, in particular, for our electrostatic embedding. For this purpose, we have used the salt bridge formed between Asp 69 and Lys 219 of the HIV-1 reverse transcriptase. In previous studies, it has been shown that this interaction plays an important role in AZT hypersusceptibility.[41]

Starting from a stable region taken from a previous classical MD simulation for this system,[41] we performed a QM/MM MD of 20 000 steps (Δt = 0.5 fs), defining as the QM region the atoms of the residues involved in the salt bridge (Figure 1a). In the QM/MM simulations with electrostatic embedding, we obtain a distance between the H atom and the O atom in the salt bridge of ∼1.6–1.7 Å, in the expected range for this type of interaction[42] (see solid red line in Figure 1b). If the electrostatic interaction due to the MM region is not included in the QM Hamiltonian (mechanical embedding, Figure 1b, dashed line) the distance between the H atom and the two electronegative atoms decreases dramatically, presenting an oscillatory behavior (in Figure 1b, these oscillations have been smoothed out) as a consequence of a very strong interaction, as expected in a vacuum environment.[40] This result indicates that the electrostatic embedding implemented in FIREBALL/AMBER correctly reproduces the effect of the polar environment for a

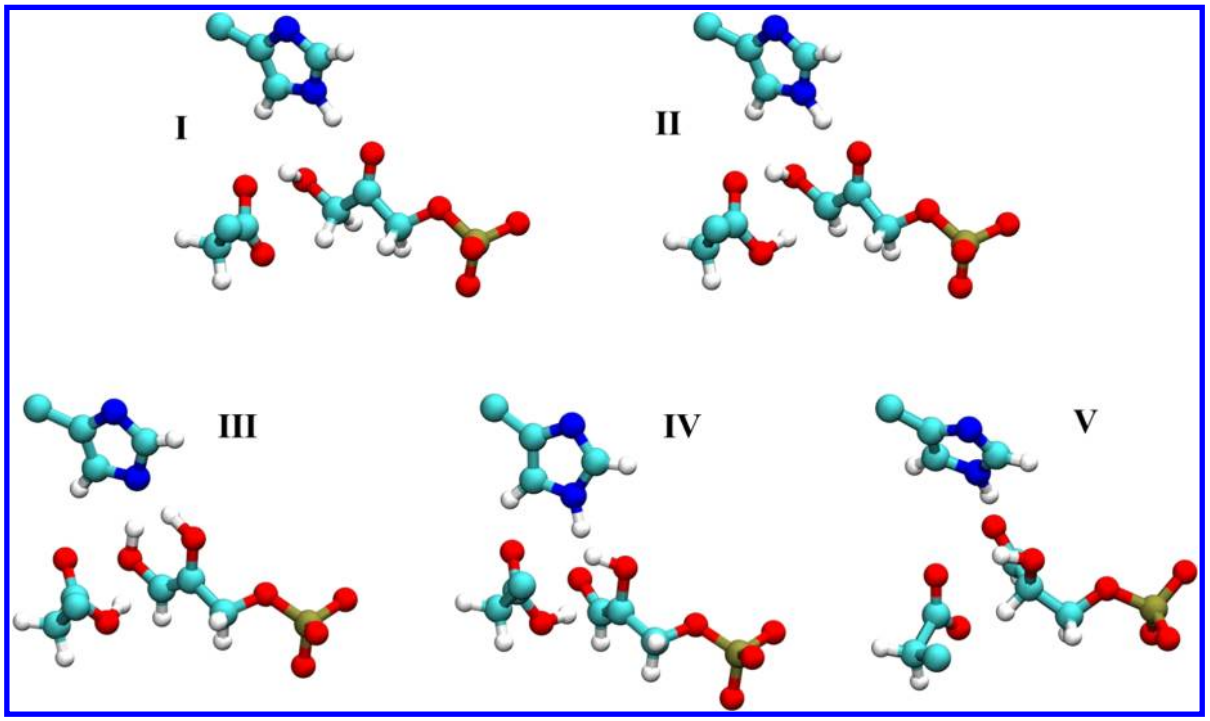


Figure 2. Intermediate structures obtained in our calculations for the triosephosphate isomerase (TIM) system.

salt bridge between amino acids. We have also analyzed the first solvation shell around the salt bridge. We find that, on average, there are four water molecules around the ammonium group. For the carboxylic group, we find two water molecules (on average) around the O atom involved in the hydrogen bond and three water molecules for the other O atom (see Figure 1a). We also find that the average number of water molecules in the first solvation shell around the carboxylic and ammonium groups is the same in QM/MM and classical molecular dynamics simulations.

As a further test for the QM/MM coupling, we have also run simulations for this system, including one residue in the QM region and the other in the MM region, or vice versa. Figure 1c shows the distance between the O and N atoms in the salt bridge in these simulations (red and green lines), in comparison to the case with both residues in the QM region (black line). In the three cases, the O−N distance oscillates between 2.6−3.1 Å, as expected for the hydrogen-bond distance for this salt bridge.

3.2. TIM. In this example, the relative energies of different intermediate structures for the triosephosphate isomerase (TIM)-catalyzed conversion of dihydroxyacetone phosphate (DHAP) to glyceraldehyde 3-phosphate (GAP) are compared to the results by Lennartz et al.[33] using different DFT and *ab initio* QM/MM approaches. In our calculations for these proton-transfer reactions, the QM region contains 37 atoms (His95, Glu165, and DHAP). The atomic structures for the intermediate states were obtained following the same protocol as in ref 33. Figure 2 shows these intermediate structures, in agreement to the DFT QM/MM intermediate structures found previously.[33] In Table 1, we compare the DFT QM/MM results of Lennartz et al.[33] for the intermediate states with optimized geometries with our results for our optimized geometries; B3LYP/SVP+(37) and BP86/SVP+(37) correspond to calculations with the same QM region as in our work, whereas in BP86/SVP+(275), there are 275 atoms in the QM region. For the sake of completeness, in Table 1, we also show the range of values for the different DFT and *ab initio* (MP2) QM/MM calculations in ref 33.

Table 1. Relative Energies for the TIM Intermediate States (see Figure 2) in Our Work, Compared to Different DFT QM/MM Calculations[33]

	Relative Energy (kcal/mol)				
method[a]	I	II	III	IV	V
this work (37)	0.0	8.7	18	9.5	2.2
B3LYP/SVP+ (37)	0.0	6.4	15.8	8.5	3.9
BP86/SVP+ (37)	0.0	3.4	12.3	4.7	1.9
BP86/SVP+ (275)	0.0	9.2	15.4	8.7	2.5
range[b]	0.0	3.4−9.7	9.3−17.1	4.5−11.5	1.9−10.7

[a]Number given in parentheses indicates the number of atoms in the QM region. [b]Range of values obtained in the different DFT and *ab initio* QM/MM results of ref 33.

The results obtained in our calculations present an overall good agreement with the other DFT QM/MM results shown in Table 1. In particular, our results are quite similar to the B3LYP results. Also notice the good agreement of our results with the BP86 calculations with 275 atoms in the QM region.

3.3. RNase A. In biomolecular systems, perhaps the most important applications of QM/MM are studies in which molecular orbitals are rearranged, as it occurs in the active center of enzyme catalyzed reactions. As an example of our QM/MM implementation, we have analyzed the cleavage of a RNA molecule through a hydrolysis process, a reaction catalyzed by RNase A, the first protein whose catalytic molecular mechanism was described.[43] The process is divided in two steps: the first one consists of the cleavage of a phosphodiester bond, ending in a 2′,3′ cyclic nucleotide. In the second step, the 2′,3′ cyclic nucleotide is hydrolyzed to a 3′ nucleotide. The molecular mechanism is based in two histidine residues, His 12 and His 119, which are deprotonated and

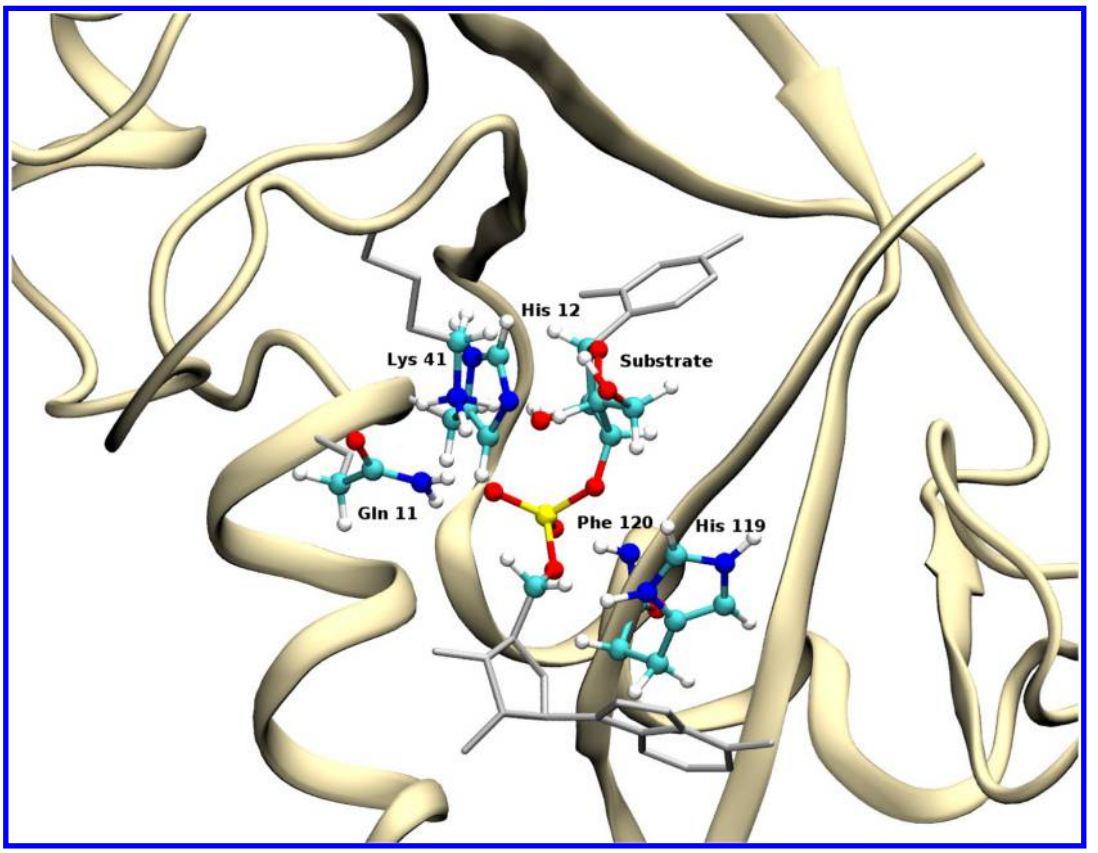


Figure 3. RNase A active center. Atoms in the QM region are represented in ball and sticks, the rest of the protein is presented in yellow, and the nucleic acid is shown as gray sticks. Water molecules are not represented for the sake of clarity.

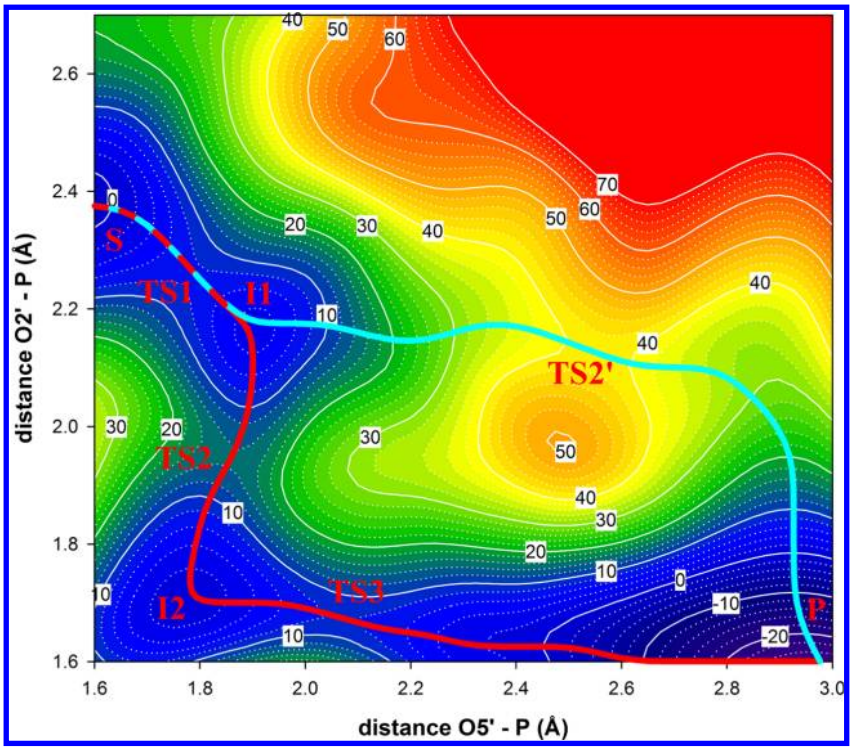


Figure 4. Free-energy surface for the RNase A reaction obtained using the FIREBALL/AMBER system (in kcal/mol). Axes represent the reaction coordinates (distances O2′−P and O5′−P, respectively). The two possible reaction trajectories are indicated: dissociative S_N1 pathway in cyan and associative S_N2 pathway in red. The positions of substrate (S), transition states (TS), intermediate states (I), and reaction product (P) in both pathways are indicated.

dx.doi.org/10.1021/ct500033w | J. Chem. Theory Comput. 2014, 10, 2185−2193

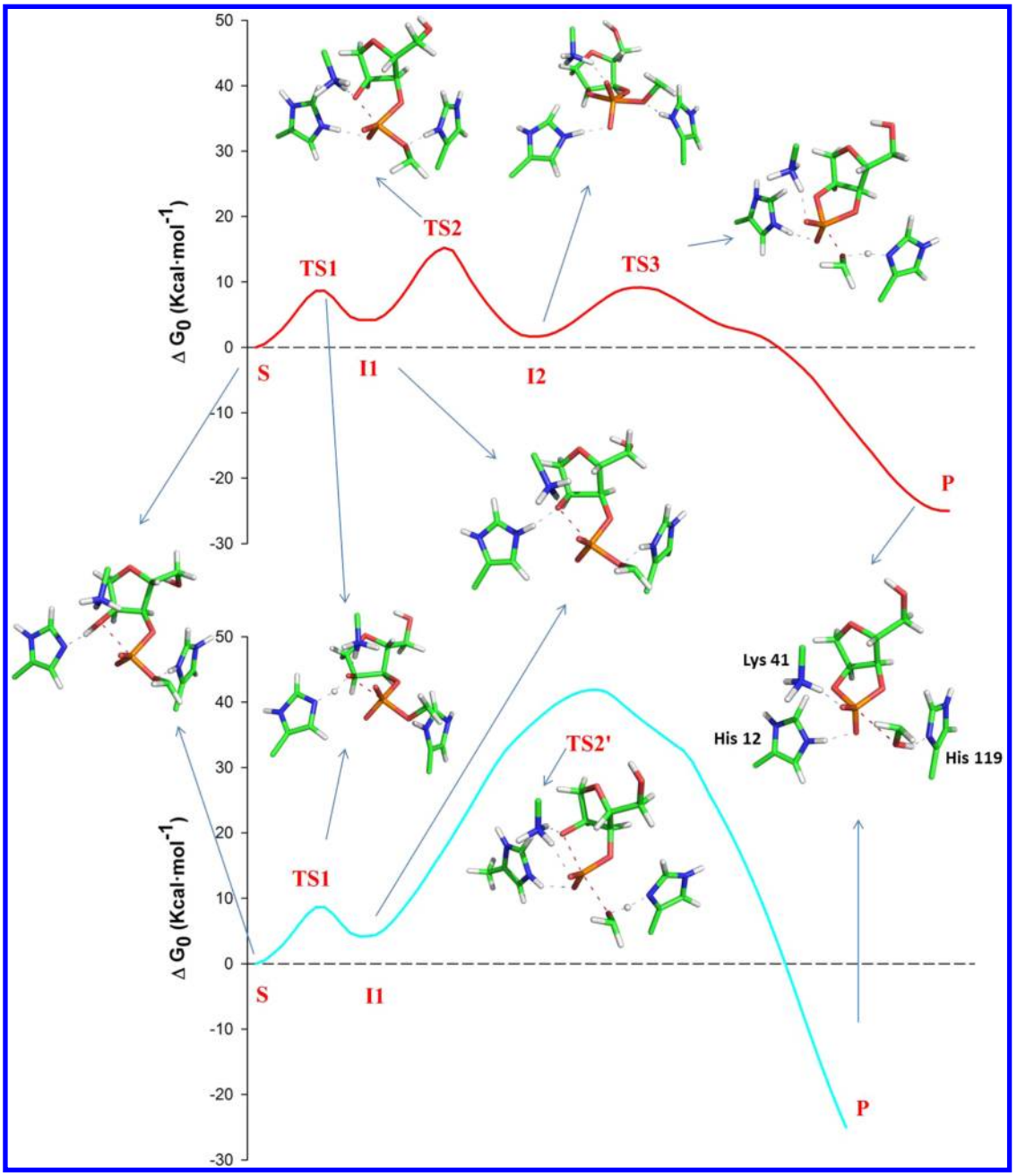


Figure 5. Analysis of the two possible trajectories of the RNase A reaction obtained from the energy map in Figure 4. Dissociative S_N1 pathway is colored in cyan and associative S_N2 pathway in red. The positions of substrate (S), transition states (TS), intermediate states (I), and product (P) are indicated. The atomic structure of the QM region for each state is depicted.

protonated, respectively, at the beginning of the reaction (see Figure 3). His 12 acts as the acceptor for the proton of the sugar O2′, being His 119 involved in the proton transfer to the O5′.[34] In the FIREBALL/AMBER simulation, the X-ray structure of RNase A with code 1RPG in the Protein Data Bank was used; the initial substrate coordinates were obtained from the position of the inhibitor present in the crystal.[44] The system was solvated using a box of TIP3 waters and neutralized with Cl^- ions. Before the QM/MM simulation, an unrestricted NPT classical MD of 10 ns was run for the equilibration of the structure, followed by 100 ps of equilibration in QM/MM MD. The atoms included in the QM region were as follows: the sugar moiety containing the free OH2′, the phosphodiester bond, the side chains of His12, His119, Lys41, Gln11, and the main chain of Phe120 (Figure 3). In total, 75 atoms were included in the QM calculation. For the MM region, the ff99SB force field was used[45] and a direct space cutoff of 10 Å was used.

In order to obtain a useful and complete view of the first step of the reaction (Figure 3) in RNase A, we decided to construct a free-energy surface map with our QM/MM implementation, using as reaction coordinates the distances O2′−P and O5′−P. The free-energy surface was calculated using a combination of steered MD along the reaction coordinates and umbrella sampling. This technique was demonstrated to be useful in a previous analysis for the proteins HRas[8] and F1-ATPase,[16] using the semiempirical PM3 Hamiltonian. In this way, a large conformational space can be explored to find the lower energy

trajectory and the energy barrier of the reaction. The energetic landscape obtained for RNase A is shown in Figure 4. In order to generate this map, we had to run QM/MM MD simulations corresponding to a total simulation time of 2 ns (time step Δt = 0.5 fs, i.e., 4 × 10^6 steps); this was possible due to the computational efficiency of FIREBALL/AMBER, which is comparable to semiempirical approaches.

From this energy map for the RNase A reaction (Figure 4), a large amount of information can be extracted. The most noteworthy one is the existence of two possible reaction pathways, whose energy profiles are also represented in Figure 5. The pathway showing the lowest energy barrier (red line in Figures 4 and 5) has the characteristics of a bimolecular nucleophilic substitution (S_N2 reaction or associative mechanism), being a complex pathway presenting two intermediate states (I1 and I2) and three transition states (TS1 to TS3) with their corresponding energy barriers. The alternative pathway is somehow simpler (one intermediate state, I1, and two transition states, TS1 and TS2′) and it is geometrically compatible with an unimolecular nucleophilic substitution (S_N1 reaction or dissociative mechanism). This S_N1 pathway takes place through a higher energy barrier and, therefore, it is unfavorable with respect to the S_N2 reaction.

Both pathways, S_N1 and S_N2, start with the activation of the OH2′ group of the ribose moiety, which donates the proton to His 12. The energy barrier corresponding to this proton transfer is ∼8 kcal/mol; the negative charge of O^-2' in the first transition state (TS1) is stabilized by the interaction with the Lys 41 residue. The positive charge of the protonated His 12 provides additional stability to the intermidiate structure I1. Then, in the S_N2 associative pathway (red line in Figures 4 and 5), the nucleophilic O^-2' group attacks the phosphorus of the phosphodiester bond, giving rise to the pentavalent transition state characteristic of a S_N2 reaction, with an energy barrier of ∼15 kcal/mol. The positive imidazole ring of the His 12 residue moves and interacts with an equatorial oxygen of the transition state, contributing to its stabilization. The resulting structure, I2, stabilized by the interaction with His 12 and Lys 41, corresponds to the dianionic intermediate described in the theoretical study of Glennon and Warshel.[46] The next energy barrier (TS3) corresponds to the proton transfer from His 119 to the leaving O5′ and the formation of the 2′,3′ cyclic nucleotide phosphate, this being the final product of the first step of the entire reaction.

The S_N1 dissociative pathway (cyan line in Figures 4 and 5) shares the first steps (TS1 and I1) with the S_N2 reaction. In the S_N1 pathway, once the proton from the OH2′ has been transferred to His 12 (I1 intermediate), the protonated His 119 donates its proton to the bridging O5′ of the phosphate, cleaving the bond. This cleavage generates the trigonal transition state, characteristic of a S_N1 reaction. Although His 12 and Lys 41 partially stabilize transition state TS2′, the resulting energy barrier (∼40 kcal/mol) is notably higher than the one associated with transition state TS2 in the S_N2 pathway. We conclude from the analysis of the energy map that the most probable trajectory for the reaction is the S_N2 associative pathway.

For the sake of comparison, we have also performed some calculations for this system using several semiempirical QM/MM approaches. The advantage of these techniques is computational efficiency, allowing a detailed analysis of the conformational space for the biomolecular reaction (e.g., calculation of energy maps, such as the one shown in Figure 4), at the price of a reduced accuracy. In the present case, we find that these semiempirical approaches do not properly describe the key proton transfer processes for this reaction, because of their poor description of the P pentavalent transition state. In particular, the transfer of H from OH2′ to His 12 (I1 state, see Figure 5) does not occur (AM1[47] or PM3[31] methods, which do not include d-orbitals for P), or takes place only at very short O2′−P distances (AM1d[48] or PM6,[49] with P d-orbitals). This is due to the lack of formation of a transition state with the appropriate geometry. In the case of DFTB (using the only available basis set, which includes P atoms[50]), the initial state is already unstable.

In conclusion, the calculations presented here for RNase A highlight the main advantage of FIREBALL/AMBER: it offers the possibility to run DFT QM/MM molecular dynamics simulations for biomolecules with a computational efficiency similar to semiempirical QM/MM approaches, but with a much greater accuracy.

4. CONCLUSIONS

In summary, we have presented an efficient quantum mechanics/molecular mechanics (QM/MM) technique that combines the local-orbital molecular dynamics (MD) density functional theory (DFT) method FIREBALL and the AMBER suite of programs for the simulation of biomolecular systems. This technique presents an excellent balance between computational efficiency and accuracy, allowing the analysis of a large reaction conformational space using long DFT QM/MM molecular dynamics simulations. Also, this technique can be used to study biomolecular systems including several hundred of atoms in the QM region. We have tested this approach for three representative biomolecular systems: a salt bridge between aspartic and lysine residues and two enzymatic reactions (the conversion of dihydroxyacetone phosphate (DHAP) to glyceraldehyde 3-phosphate (GAP) catalyzed by triosephosphate isomerase (TIM) and the cleavage of RNA by RNase-A). In particular, in the case of RNase we have been able to analyze the first step of this enzymatic reaction exploring a large conformational space using long DFT QM/MM molecular dynamics simulations (total simulation time = 2 ns). We finally mention that the FIREBALL/AMBER method presents a much higher accuracy than QM/MM methods based on semiempirical Hamiltonians, with a comparable computational speed, which represents a major step forward for the computer simulation of bioprocesses.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: pagomez@cbm.csic.es (P. Gómez-Puertas).
*E-mail: jose.ortega@uam.es (J. Ortega).

Author Contributions

⊥These authors contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Enrique Abad for fruitful discussions. This work is supported by the Spanish MICINN (Contract Nos. IPT2011-0964-900000, SAF2011-13156-E, FIS2010-16046), the European Commission (Contract Nos. DIVINOCELL FP7 HEALTH-F3-2009-223431 and COST-CMTS Action CM1002) and the Comunidad de Madrid (S2009/MAT-

1467). The computational support of the "Centro de Computacion Cientifica-CCC-UAM" is acknowledged. Work at Biomol-Informatics was partially financed by the European Social Fund. R.C.W. acknowledges funding from the National Science Foundation (NSF), through the Scientific Software Innovations Institutes (Program No. NSF SI2-SSE (NSF114876)) and a fellowship from NVIDIA, Inc. J.P.L. acknowledges support from the Department of Energy (Grant No. DE-FG02-10ER16164); this funding was instrumental in adding new algorithms to the FIREBALL code for enhancing calculations of catalytic reaction pathways.

REFERENCES

(1) Martin, R. *Electronic Structure*; Cambridge University Press: Oxford, U.K., 2004.
(2) Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics*; Cambridge University Press: Oxford, U.K., 2009.
(3) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(4) Gilson, M. K.; Zhou, H.-X. *Ann. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21−42.
(5) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(6) Field, M.; Bash, P.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700−733.
(7) Zhang, Y. *Theor. Chem. Acc.* **2005**, *116*, 43−50.
(8) Martín-García, F.; Mendieta-Moreno, J. I.; López-Viñas, E.; Gómez-Puertas, P.; Mendieta, J. *Biophys. J.* **2012**, *102*, 152−157.
(9) Lewis, J. P.; Jelnek, P.; Ortega, J.; Demkov, A. A.; Trabada, D. G.; Haycock, B.; Wang, H.; Adams, G.; Tomfohr, J. K.; Abad, E.; Wang, H.; Drabold, D. A. *Phys. Status Solidi B* **2011**, *248*, 1989−2007.
(10) Jelínek, P.; Wang, H.; Lewis, J.; Sankey, O.; Ortega, J. *Phys. Rev. B* **2005**, *71*, 235101.
(11) Lewis, J.; Glaesemann, K.; Voth, G.; Fritsch, J.; Demkov, A.; Ortega, J.; Sankey, O. *Phys. Rev. B* **2001**, *64*, 195103.
(12) Demkov, A. A.; Ortega, J.; Sankey, O. F.; Grumbach, M. P. *Phys. Rev. B* **1995**, *52*, 1618−1630.
(13) Sankey, O. F.; Niklewski, D. J. *Phys. Rev.* **1989**, *40*, 3979.
(14) Case, D.; Darden, T.; Cheatham, T. E., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A.; Kolossváry, I.; Wong, K.; Paesani, F.; Vanicek, J.; Wolf, R.; Liu, J.; Wu, X.; Brozell, S.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Cui, G.; Roe, D.; Mathews, D.; Seetin, M.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. et al. *AMBER 12*, 2012.
(15) Salomon-Ferrer, R.; Case, D.; Walker, R. *WIREs Comput. Mol. Sci.* **2012**, *3*, 198−210.
(16) Martín-García, F.; Mendieta-Moreno, J. I.; Marcos-Alcalde, I. N.; Gómez-Puertas, P.; Mendieta, J. *Biochemistry* **2013**, *52*, 959−966.
(17) Basanta, M.; Dappe, Y.; Jelínek, P.; Ortega, J. *Comput. Mater. Sci.* **2007**, *39*, 759−766.
(18) García-Vidal, F.; Merino, J.; Perez, R.; Rincon, R.; Ortega, J.; Flores, F. *Phys. Rev. B* **1994**, *50*, 10537.
(19) Schönhammer, K.; Gunnarsson, O.; Noack, R. *Phys. Rev. B* **1995**, *52*, 2504.
(20) Löwdin, P. *J. Chem. Phys.* **1950**, *18*, 365.
(21) Carlson, B.; Keller, J. M. *Phys. Rev.* **1957**, *105*, 102−103.
(22) Reed, A. E.; Weinstock, R. B.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 735.
(23) Trabada, D. G. Efectos dinámicos en la superficie-SiC(100). *Ph.D. Thesis*, Universidad Autónoma de Madrid: Madrid, Spain, 2009.
(24) Walker, R.; Crowley, M.; Case, D. *J. Comput. Chem.* **2008**, *29*, 1019−1031.
(25) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(26) Brooks, B. R.; Brooks, C. L., III.; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; et al. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(27) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845−854.
(28) Kunz, A.-P. E.; Allison, J. R.; Geerke, D. P.; Horta, B. A. C.; Hünenberger, P. H.; Riniker, S.; Schmid, N.; van Gunsteren, W. F. *J. Comput. Chem.* **2012**, *33*, 340−353.
(29) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *7863*, 11225−11236.
(30) Senn, H. M.; Thiel, W. *Angew. Chem.* **2009**, *48*, 1198−1229.
(31) Stewart, J. J. P. *J. Mol. Model.* **2004**, *10*, 155−164.
(32) de M. Seabra, G.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E. *J. Phys. Chem. A* **2007**, *111*, 5655−5664.
(33) Lennartz, C.; Schafer, A.; Terstegen, F.; Thiel, W. *J. Phys. Chem. B* **2002**, *106*, 1758−1767.
(34) Cuchillo, C. M.; Nogués, M. V.; Raines, R. T. *Biochemistry* **2011**, *50*, 7835−7841.
(35) Becke, A. D. *Phys. Rev. B* **1988**, *38*, 3098.
(36) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
(37) Mendieta-Moreno, J. I. et al. Unpublished work, 2014.
(38) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.
(39) Korth, M.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 2929−2936.
(40) Honig, B. H.; Hubbellt, W. L. *Proc. Natl. Acad. Sci. U. S. A.* **1984**, *81*, 5412−5416.
(41) Kisic, M.; Mendieta, J.; Puertas, M. C.; Parera, M.; Martínez, M. a.; Martinez-Picado, J.; Menéndez-Arias, L. *J. Mol. Biol.* **2008**, *382*, 327−341.
(42) Bordo, D.; Argos, P. *J. Mol. Biol.* **1994**, *243*, 504−519.
(43) Findly, D.; Herries, D. G.; Mathias, A. P.; Rabin, B. R.; Ross, C. A. *Nature* **1961**, *190*, 781−784.
(44) Zegers, I.; Maes, D.; Dao-Thi, M. H.; Poortmans, F.; Palmer, R.; Wyns, L. *Protein Sci.* **1994**, *3*, 2322−2339.
(45) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712−725.
(46) Glennon, T. M.; Warshel, A. *J. Am. Chem. Soc.* **1998**, *7863*, 10234−10247.
(47) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *4*, 3902−3909.
(48) Imhof, P.; Noé, F.; Fischer, S.; Smith, J. C. *J. Chem. Theory Comput.* **2006**, *2*, 1050−1056.
(49) Stewart, J. J. P. *J. Mol. Model.* **2006**, *13*, 1173−1213.
(50) Yang, Y.; Yu, H.; York, D.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2008**, *4*, 2067−2084.