CHAPTER THREE

A Practical Quantum Mechanics Molecular Mechanics Method for the Dynamical Study of Reactions in Biomolecules

Jesús I. Mendieta-Moreno*,†, Iñigo Marcos-Alcalde†, Daniel G. Trabada*, Paulino Gómez-Puertas†, José Ortega*,[1], Jesús Mendieta†,‡

*Departamento de Física Teórica de la Materia Condensada and Condensed Matter Physics Center (IFIMAC), Universidad Autónoma de Madrid, Madrid, Spain
†Molecular Modelling Group, Center of Molecular Biology "Severo Ochoa" (CSIC-UAM), Madrid, Spain
‡Biomol-Informatics SL, Campus UAM, Madrid, Spain
[1]Corresponding author: e-mail address: jose.ortega@uam.es

Contents

Abstract

Quantum mechanics/molecular mechanics (QM/MM) methods are excellent tools for the modeling of biomolecular reactions. Recently, we have implemented a new QM/MM method (FIREBALL/AMBER), which combines an efficient density functional theory method (FIREBALL) and a well-recognized molecular dynamics package (AMBER), offering an excellent balance between accuracy and sampling capabilities. Here, we present a detailed explanation of the FIREBALL method and FIREBALL/AMBER implementation. We also discuss how this tool can be used to analyze reactions in biomolecules using steered molecular dynamics simulations. The potential of this approach is shown by the analysis of a reaction catalyzed by the enzyme triose-phosphate isomerase (TIM). The

Advances in Protein Chemistry and Structural Biology, Volume 100
ISSN 1876-1623
http://dx.doi.org/10.1016/bs.apcsb.2015.06.003

conformational space and energetic landscape for this reaction are analyzed without *a priori* assumptions about the protonation states of the different residues during the reaction. The results offer a detailed description of the reaction and reveal some new features of the catalytic mechanism. In particular, we find a new reaction mechanism that is characterized by the intramolecular proton transfer from O1 to O2 and the simultaneous proton transfer from Glu 165 to C2.

1. INTRODUCTION

One of the more interesting challenges at present in Biophysics and Molecular Biology is to understand how the proteins work at the atomic level. Processes such as conformational changes (Karplus & McCammon, 2002) or ligand binding (Gilson & Zhou, 2007) can be studied by molecular mechanics (MM) simulations using empirical force fields. Biomolecular reactions, however, require chemical bonds to be broken and formed, processes that are not properly described using MM force fields. Chemical bonds are intrinsically related to the quantum electronic structure of the material and therefore quantum mechanics (QM) techniques are required in these cases (Martin, 2004; Marx & Hutter, 2009). Unfortunately, the computational cost of QM calculations for biomolecular systems is prohibitive (Adcock & Mccammon, 2006). By describing the active region with a QM method and the surroundings with a MM method, QM/MM approaches offer the required accuracy to model biomolecular reactions with reasonable computational costs (Field, Bash, & Karplus, 1990; Senn & Thiel, 2009; Warshel & Levitt, 1976).

The accuracy and computational efficiency of QM/MM calculations critically depend on the QM method used. In most of the biomolecular applications, a detailed study of the conformational space for the reaction is essential to understand, e.g., the catalytic mechanisms. Computationally efficient QM methods, such as semi-empirical approaches, enable wider conformational sampling (Martín-García, Mendieta-Moreno, López-Viñas, Gómez-Puertas, & Mendieta, 2012), but usually do not offer enough accuracy. Conversely, high-accuracy quantum chemistry methods greatly increase the computational costs, significantly reducing the conformational sampling capabilities (Zhang, 2005). Recently, we have implemented a new QM/MM technique (Mendieta-Moreno et al., 2014) based on the combination of FIREBALL (QM) (Demkov, Ortega, Sankey, & Grumbach, 1995; Jelínek, Wang, Lewis, Sankey, & Ortega, 2005; Lewis et al., 2001, 2011;

Sankey & Niklewski, 1989) and AMBER (MM) (Case et al., 2012; Salomon-Ferrer, Case, & Walker, 2012). FIREBALL is a density functional theory (DFT) molecular dynamics (MD) technique that is very well suited for QM/MM biomolecular applications. This method allows molecular dynamics (MD) simulations involving relatively large QM regions ($\sim 10^2$ atoms) with good accuracy (Mendieta-Moreno et al., 2014). AMBER is a well-recognized MD package for the simulation of biomolecules which offers a wide range of MD tools. In particular, the *sander* program included in AMBER offers powerful techniques for exhaustive sampling of the conformational space using biased MD approaches (umbrella sampling, replica exchange, nudged elastic band, targeted MD, steered MD, …), which can be used directly in the QM/MM applications (Case et al., 2012).

This chapter describes the FIREBALL/AMBER implementation and highlights its most interesting features. In Section 2, detailed descriptions of FIREBALL and the FIREBALL/AMBER implementation are provided. This method can be used to analyze reactions in biomolecules sampling the conformational space by means of long QM/MM MD simulations. As explained in Section 3, we use steered MD to generate of the order of $\sim 10^6$ individual structures that allow us to calculate free energy maps for the reaction. In Section 4, we analyze the reversible interconversion between dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GAP) catalyzed by the triose-phosphate isomerase (TIM) using these techniques. Finally, in Section 5 we present our conclusions.

2. DESCRIPTION OF THE METHOD

2.1 QM Method: FIREBALL

In QM/MM MD methods, we need to obtain the quantum electronic structure of the system for each time step ($t_i = m_i \Delta t$; $\Delta t \sim 10^{-15}$ s) in the simulation. This is the reason why these methods are so computationally demanding. The most practical QM/MD methods are based on DFT that offers a considerable simplification of the quantum many-body problem (Hohenberg & Kohn, 1964; Kohn & Sham, 1965; Martin, 2004). In the following, we will use the acronym DFT/MM to specify a QM/MM technique in which a DFT method is used for the QM calculation. FIREBALL is a DFT method that is specifically designed for the study of complex systems using MD simulations (Demkov et al., 1995; Jelínek et al., 2005; Lewis et al., 2001, 2011; Sankey & Niklewski, 1989). In this method, the electronic structure is obtained solving the DFT equations (see below) using

a basis set of atomic-like orbitals localized in the atoms of the system. These orbitals are short-ranged, numerical, and are optimized so that reasonably good accuracy can be obtained with a low number of orbitals per atom (Basanta, Dappe, Jelínek, & Ortega, 2007; Sankey & Niklewski, 1989). Recently, we have developed basis sets of optimized atomic-like orbitals for biomolecular systems containing H, C, N, O, P, S, and Mg atoms. In particular, our minimal basis set for biomolecules includes one s orbital for H; sp^3 orbitals for C, N, and O; and sp^3d^5 orbitals for P, S, and Mg. In order to optimize the basis set, we have analyzed different sets of organic molecules relevant for biological systems and have considered intermolecular energies and distances as well as intramolecular distances. In particular, we have used the S66 reference dataset (Řezáč, Riley, & Hobza, 2011) for intermolecular energies and distances as well as the ionic hydrogen bond dataset (Řezáč & Hobza, 2012) for complexes featuring ionic hydrogen bonds (hydrogen bonds between ionic and neutral groups). Details will be published elsewhere.

Most DFT methods use periodic boundary conditions in some way (plane-waves basis sets, representation of the potential and electron density on a periodic grid, use of fast Fourier transforms, etc.), since these techniques were first developed in Solid State Physics to analyze the paradigmatic case of the crystalline solid with atoms arranged in a periodic structure. In biomolecular systems, however, the atoms are not arranged in any periodic way and there is no advantage to use any artificial periodicity in the calculations. FIREBALL is a DFT technique completely formulated in *real-space*, i.e., it does not require the use of supercells (periodic boundary conditions). Due to this property, FIREBALL is well suited for biomolecular simulations in combination with MM calculations (that are also real-space methods).

In the FIREBALL method, the electronic structure is obtained in each time step of the simulation via solution of the DFT Schrödinger-like equation (Martin, 2004):

$$\hat{H}\psi_i(\mathbf{r}) = \left(-\frac{\hbar^2}{2m_e}\nabla^2 + V_{\text{ion}}(\mathbf{r}) + V_{\text{ee}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}). \quad (1)$$

The first term represents the kinetic energy of the electrons, V_{ion}, is the ionic potential:

$$V_{\text{ion}}(\mathbf{r}) = \sum_{\alpha}\frac{-eZ_\alpha}{|\mathbf{r}-\mathbf{R}_\alpha|}, \quad (2)$$

(eZ_α is the nuclear or pseudopotential charge on atom α at position $\mathbf{R}_\alpha$; $-e$ is the charge of the electron),

$$V_{ee}(\mathbf{r}) = e\int \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d^3r' \tag{3}$$

is the average electron potential (i.e., the Hartree potential). Finally, the exchange-correlation potential V_{xc} is the functional derivative of the DFT exchange-correlation energy $E_{xc}[\rho(\mathbf{r})]$, $V_{xc} = \delta E_{xc}/\delta\rho(\mathbf{r})$ (Martin, 2004). In these equations, $\rho(\mathbf{r})$ is the electron density (Foulkes & Haydock, 1989; Harris, 1985), which is determined in a self-consistent way in terms of the orbital occupation numbers (Demkov et al., 1995).

In the FIREBALL method, a basis set of numerical atomic-like orbitals, ϕ_μ, is used to solve Eq. (1) (Basanta et al., 2007; Sankey & Niklewski, 1989):

$$\psi_i(\mathbf{r}) = \sum_\mu c_{i\mu}\phi_\mu(\mathbf{r}-\mathbf{R}_\alpha); \tag{4}$$

μ is a combined index, $\mu \equiv (\alpha,l,m)$ that refers to the particular atom, α (at $\mathbf{R}_\alpha$), atomic subshell l (e.g., 3s,4s,3p,3d,), and angular component m (e.g., p_x,p_y,p_z). The electron density $\rho(\mathbf{r})$ is also written in terms of the orbitals ϕ_μ as a sum of atomic-like densities, $\rho_\alpha(\mathbf{r})$, for the different atoms in the system (Demkov et al., 1995):

$$\rho(\mathbf{r}) = \sum_\alpha \rho_\alpha(\mathbf{r}) = \sum_\mu n_\mu |\phi_\mu(\mathbf{r}-\mathbf{R}_\alpha)|^2. \tag{5}$$

In practice, the atomic densities ρ_α are approximated to be spherically symmetric around each atomic site $\mathbf{R}_\alpha$. The electron density $\rho(\mathbf{r})$ in Eq. (5) is determined in a self-consistent way in terms of the orbital occupation numbers n_μ (Demkov et al., 1995; García-Vidal et al., 1994; Schönhammer, Gunnarsson, & Noack, 1995). For this purpose, output orbital occupation numbers n_μ^{out} are obtained from the output electron density $\rho^{out}(\mathbf{r})$ as obtained from the occupied eigenvectors ψ_i in Eq. (1):

$$\rho^{out}(\mathbf{r}) = \sum_{i\in occ} |\psi_i(\mathbf{r})|^2 \rightarrow \sum_\mu n_\mu^{out} |\phi_\mu(\mathbf{r}-\mathbf{R}_\alpha)|^2. \tag{6}$$

Different methods can be used to project [arrow in Eq. (6)] the output electron density from Eq. (1) into the form given by Eq. (5) using, for example, Löwdin Orbitals (Carlson & Keller, 1957; Demkov et al., 1995; Löwdin, 1950) or Natural Atomic Orbitals (Reed, Weinstock, & Weinhold,

1985). Equations (1–6) are iteratively solved until self-consistency: in the self-consistent solution n_μ^{out} and the input orbital occupation numbers n_μ must coincide (Demkov et al., 1995). In each of these iterations, the eigenvectors ψ_i are obtained from the diagonalization of the $N \times N$ Hamiltonian matrix in Eq. (1), $H_{\mu,\nu} = \langle \phi_\mu | \hat{H} | \phi_\mu \rangle$, where N is the total number of orbitals ϕ_μ in the basis set.

Once these equations have been solved, the total energy and forces are calculated as

$$E_{\text{tot}}[\rho(\mathbf{r})] = E_{\text{sum}} - E_{\text{ee}}[\rho(\mathbf{r})] + E_{\text{xc}}[\rho(\mathbf{r})] - \int \rho(\mathbf{r}) V_{\text{xc}}[\rho(\mathbf{r})] d^3r + E_{\text{ion–ion}}; \tag{7}$$

$$\mathbf{F}_\alpha = -\frac{\partial E_{\text{tot}}}{\partial \mathbf{R}_\alpha}. \tag{8}$$

In Eq. (7), E_{sum} is a sum over the occupied eigenvalues, $E_{\text{sum}} = \sum_{i \in \text{occ}} \epsilon_i$, of the self-consistent solution of Eq. (1), E_{ee} is the average electron–electron interaction energy:

$$E_{\text{ee}}[\rho(\mathbf{r})] = \frac{e^2}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3r d^3r', \tag{9}$$

$E_{\text{ion–ion}}$ is the ion–ion interaction energy:

$$E_{\text{ion–ion}} = \frac{e^2}{2} \sum_{i,j} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}, \tag{10}$$

and $E_{\text{xc}}[\rho(\mathbf{r})]$ is the exchange-correlation energy (Martin, 2004).

The computational efficiency of FIREBALL is related to both the use of optimized basis sets of atomic-like orbitals and the representation of $\rho(\mathbf{r})$ as a sum of atomic-like densities[1] (Eq. 5). In this way, four-center integrals are not required for the solution of Eq. (1) or the calculation of E_{tot} and the forces $\mathbf{F}_\alpha$, and all the two- and three-center interactions are tabulated beforehand and placed in interpolation data tables which are no larger than two-dimensional (Lewis et al., 2011; Sankey & Niklewski, 1989). A practical tabulation-interpolation scheme is used, and all the matrix elements required during the MD simulation are evaluated by looking up the necessary information from the data tables, which are read at the beginning

[1] This approximation is related to the use of a self-consistent version of the Harris–Foulkes functional (Demkov et al., 1995; Foulkes & Haydock, 1989; Harris, 1985).

of the calculation and stored in memory throughout the MD simulation. In order to also use this tabulation-interpolation scheme for the calculation of the exchange-correlation contributions, which are highly nonlinear, different approximations have been developed (Horsfield, 1997; Jelínek et al., 2005; Lewis et al., 2001; Sankey & Niklewski, 1989).

Regarding the calculation of the forces $\mathbf{F}_\alpha$, Eq. (8), the most difficult term is the contribution from E_{sum}. Using a variant of the Hellmann–Feynman theorem (Sankey & Niklewski, 1989), this contribution is calculated in terms of the derivatives of the Hamiltonian matrix elements, $H_{\mu,\nu}$. These matrix elements are precalculated and stored in data tables and thus the required derivatives $\partial H_{\mu,\nu}/\partial \mathbf{R}_\alpha$ are readily available (see Sankey & Niklewski, 1989 for details).

2.2 FIREBALL/AMBER

In QM/MM methods, the system is divided into a chemically active region (e.g., the reactant molecules and catalytic site residues) and the *environment* (i.e., rest of the protein, water molecules, ions). The active region (or QM region) typically contains about 10^2 atoms and is described using a QM approach. For practical reasons, the environment ($\sim 10^5$–10^6 atoms), or MM region, is described using a MM method. In the study reactions in biomolecules, it is particularly important to properly take into account the effect of the electrostatic potential of the environment on the electronic structure of the active region, so we pay special attention to this contribution in the following discussion. In our FIREBALL/AMBER method (Mendieta-Moreno et al., 2014), we consider the electrostatic interaction between the active region and the environment at the QM level, including in the QM Hamiltonian $H_{\mu,\nu}$ the electrostatic potential due to the force-field partial charges, eq_k, on the MM atoms. This yields:

$$H_{\mu,\nu} = H^{\text{A}}_{\mu,\nu} + \delta H^{\text{E}}_{\mu,\nu} \tag{11}$$

$H^{\text{A}}_{\mu,\nu}$ is the contribution of the atoms in the active region to the Hamiltonian matrix element, and $\delta H^{\text{E}}_{\mu,\nu}$ the electrostatic potential contribution due to the MM partial charges eq_k in the environment:

$$\delta H^{\text{E}}_{\mu,\nu} = \sum_k h^k_{\mu,\nu} = \sum_k -eq_k \int \frac{\phi_\mu(\mathbf{r}-\mathbf{R}_\alpha)\phi_\nu(\mathbf{r}-\mathbf{R}_\beta)}{|\mathbf{r}-\mathbf{R}_k|} d^3\mathbf{r} \tag{12}$$

$\phi_\mu(\mathbf{r}-\mathbf{R}_\alpha)$ and $\phi_\nu(\mathbf{r}-\mathbf{R}_\beta)$ are atomic-like basis set orbitals centered on the atoms placed at $\mathbf{R}_\alpha$ and $\mathbf{R}_\beta$, and $h^k_{\mu,\nu}$ is the contribution to the Hamiltonian

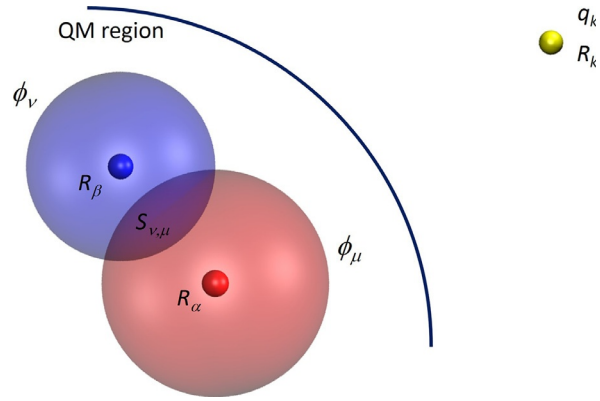

Figure 1 Scheme representing the electrostatic interaction between the MM charge, eq_k, at $\mathbf{R}_k$ and the overlap charge, $-eS_{\mu,\nu}$, yielding the contribution $h^k_{\mu,\nu}$ to the QM Hamiltonian matrix element $H_{\mu,\nu}$ between orbitals $\phi_\mu(\mathbf{r}-\mathbf{R}_\alpha)$ and $\phi_\nu(\mathbf{r}-\mathbf{R}_\beta)$ (see Eqs. 12 and 13).

matrix element $H_{\mu,\nu}$ due to the electrostatic potential of a point charge eq_k at $\mathbf{R}_k$ (see Fig. 1). For practical reasons (Demkov et al., 1995), $h^k_{\mu,\nu}$ is approximated as:

$$h^k_{\mu,\nu} = -eq_k\left(\frac{g_\mu}{|\mathbf{R}_k-\mathbf{R}_\alpha|}+\frac{g_\nu}{|\mathbf{R}_k-\mathbf{R}_\beta|}\right), \tag{13}$$

with *effective overlap charges* g_μ, g_ν

$$g_\mu = \frac{1}{2}S_{\mu,\nu}-\frac{p_{\mu,\nu}}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|}, \tag{14}$$

$$g_\nu = \frac{1}{2}S_{\mu,\nu}+\frac{p_{\mu,\nu}}{|\mathbf{R}_\alpha-\mathbf{R}_\beta|}; \tag{15}$$

$S_{\mu,\nu}$ is the overlap:

$$S_{\mu,\nu} = \int \phi_\mu(\mathbf{r}-\mathbf{R}_\alpha)\phi_\nu(\mathbf{r}-\mathbf{R}_\beta)d^3\mathbf{r}, \tag{16}$$

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

$$\mathbf{P}_{\mu,\nu} = \int [\mathbf{r}-\mathbf{R}_\mathrm{m}]\phi_\mu(\mathbf{r}-\mathbf{R}_\alpha)\phi_\nu(\mathbf{r}-\mathbf{R}_\beta)d^3\mathbf{r} \tag{17}$$

($\mathbf{R}_\mathrm{m}=(\mathbf{R}_\alpha+\mathbf{R}_\beta)/2$). Physically, the approximation of Eq. (13) includes monopole and dipole far-field effects (Demkov et al., 1995).

The electrostatic potential due to the environment also interacts with the nuclear (or pseudopotential) charges of the QM atoms, eZ_α, yielding the following contribution to the total energy (Mendieta-Moreno et al., 2014):

$$E_{\mathrm{QM-MM}}^{\mathrm{nuc}} = \sum_{\alpha,k} \frac{e^2 q_k Z_\alpha}{|\mathbf{R}_k - \mathbf{R}_\alpha|}. \tag{18}$$

In addition to the electrostatic term, the interaction between the QM and MM regions also presents bonded and nonbonded contributions, in similarity with the MM force field (Senn & Thiel, 2009). In most of the applications, the frontier between the QM and MM regions goes through covalent bonds, resulting in bonds that need to be properly saturated in the QM calculation. In the calculations presented in Section 4, we have used the link atom method (Field, Albe, Bret, Proust-De Martin, & Thomas, 2000; Walker, Crowley, & Case, 2008), adding H atoms to saturate the broken bonds at the frontier between the QM and MM regions. The nonbonded van der Waals interactions between QM and MM atoms are described using the Lennard-Jones parameters from the MM force field, and the long-range electrostatic interactions are calculated using the Particle Mesh Ewald method (Walker et al., 2008). We have used Löwdin Orbitals to determine the electron charges in the QM region, n_μ. Finally, in the simulations presented in Section 4, we have used the BLYP exchange-correlation functional (Becke exchange (Becke, 1988) with Lee–Yang–Parr correlation (Lee, Yang, & Parr, 1988)); these exchange-correlation contributions are calculated with the help of the multicenter weighted exchange-correlation density approximation (McWEDA) (Jelínek et al., 2005).

3. DYNAMICAL ANALYSIS OF REACTIONS IN BIOMOLECULES

Biomolecular systems have a large number of atoms and work at physiological temperature. Therefore, they present a high number of degrees of freedom, and for a given macroscopic condition, there is a huge number of different accessible structures. In principle, MD simulations allow us to take into account the high number of degrees of freedom in these systems. However, long DFT/MM MD simulations for biomolecular systems are still rare due to their computational cost; DFT/MM studies of reactions in biomolecules are usually performed using minimization techniques. Note that

this approach assumes that a single energy-minimized structure represents each state, not taking into account the plethora of different accessible structures.

In order to analyze a given biomolecular reaction, we must first determine the geometrical coordinates that are related to the key structural changes taking place in the reaction. The values of these coordinates define different states of the system, each of which corresponding to an ensemble of different *microscopic* structures. Therefore, instead of a single atomic structure and energy value, an ensemble of different structures and energy values is associated with each state. Thus, we have to analyze the variation of the free energy as a function of the chosen reaction coordinates. Alternatively, the corresponding potential of mean force can also be used. For this purpose, several methodologies, such as WHAM, thermodynamic integration, metadynamics, free energy perturbation, are available (Jorge, Garrido, Queimada, Economou, & MacEdo, 2010; Kumar, Rosenberg, Bouzida, Swendsen, & Kollman, 1992; Laio & Parrinello, 2002; Park, Khalili-Araghi, Tajkhorshid, & Schulten, 2003; Zwanzig, 1954).

In DFT/MM studies of enzymatic reactions, these statistical considerations can play an important role but are infrequently taken into account (principally when a large portion of the conformational space is to be studied) due to the high computational cost of DFT methods. The use of semiempirical QM methods such as AM1 and PM3 (Stewart, 2004) can reduce this cost and make a wider conformational sampling possible, but they have a low accuracy, especially for the calculation of barriers involved in reactions. The FIREBALL/AMBER implementation allows us to sample large portions of the conformational space with a higher accuracy and a similar computational cost, using DFT/MM MD simulations.

In our approach, we typically define two director geometrical parameters which adequately represent the different states of the process under study. For example, if an enzymatic reaction is essentially characterized by the formation of one bond and the disruption of another, the bond distances for these two bonds could be used as reaction coordinates (see, e.g., Mendieta-Moreno et al., 2014). Once the appropriate reaction coordinates have been defined, we explore the conformational space using an adaptively biased MD approach that presents some characteristics of steered MD as well as umbrella sampling. In this approach, we generate several MD trajectories where one of the reaction coordinates is constrained with an appropriate harmonic potential while the other reaction coordinate is adiabatically

changed by means of another harmonic potential whose position is slowly moving with constant speed during the simulation time, as in steered MD. Combining a set of equidistant parallel MD trajectories, we can perform a highly dense sampling of the conformational space of interest. In a typical calculation, we obtain around 2×10^6 individual structures, each with their own values for the energy and set of reaction coordinates.

In the next step, we create a uniform grid of points in the plane defined by the two reaction coordinates and assign each of the individual structures in our sampling to the closest grid point. In this way, each grid point is associated to a group of atomic structures; typically, each of these groups contains $\sim 10^3$ different atomic structures, on average. With all this information, the free energy landscape for the reaction is calculated as follows. First, we calculate the value of the function Q for each point (x,y) on the grid

$$Q(x,y) = \frac{1}{N_i(x,y)} \sum_i \exp\left(\frac{-E_i(x,y)}{k_B T}\right), \tag{19}$$

x and y are the values of the reaction coordinates, $E_i(x,y)$ are the potential energy values for all the atomic structures corresponding to the grid point (x,y), and $N_i(x,y)$ is the number of atomic structures in the sum. The free energy landscape is then obtained as

$$F(x,y) = -k_B T \quad \ln[Q(x,y)]. \tag{20}$$

Alternatively, the energetic landscape can also be obtained from the following potential of mean force:

$$P(x,y) = \frac{\sum_i E_i(x,y) \exp\left(\frac{-E_i(x,y)}{k_B T}\right)}{\sum_i \exp\left(\frac{-E_i(x,y)}{k_B T}\right)}. \tag{21}$$

The final map (see, e.g., Fig. 5) is obtained applying a 3D local regression (LOESS) method to the data to remove the irregularities inherent in the statistical sampling data (Cleveland & Devlin, 1986), since the variations in this energetic landscape should be smooth as a function of the reaction coordinates.

4. CATALYTIC MECHANISM OF TIM

4.1 Introduction

As an example of our approach to analyze enzymatic reactions, we study the reversible interconversion between DHAP and GAP catalyzed by the TIM. This enzyme has been extensively studied by many different techniques including theoretical calculations. However, some details of its catalytic mechanisms are still not fully elucidated. It is generally accepted that the transfer of a proton from C1 atom of DHAP to the glutamic 165 yields an enediolate intermediate which can be detected also in the nonenzymatic reaction (Hall & Knowles, 1975). However, there is no such general agreement about the catalytic mechanism by which the enediolate intermediate yields glyceraldehyde, the final product (see Fig. 2).

Several hypotheses have been proposed for the proton transfer paths during the second step of the reaction. In one of the proposed paths (Bash et al., 1991), His 95 could act as a catalytic acid donating a proton to the carboxyl oxygen in C2. After this step, the proton of the alcohol group in O1 is transferred to His 95 and C2 accepts the proton from Glu 165 recovering the enzyme the original configuration. RMN data support, however, that the imidazole ring of His 95 is not charged over the entire pH range of the enzymatic activity (Lodi & Knowles, 1991); thus, the proposed path would imply the formation of an imidazolate anion, in spite of the very high pK_a value (≈14) associated with this process (Walba & Isensee, 1961). Other paths have been proposed in order to avoid the formation of the energetically unfavorable imidazolate anion, including the transfer of the proton from Glu 165 to the carboxyl oxygen in C2 and the subsequent protonation of C2 from the C1 alcohol group through Glu 165 (Cui & Karplus, 2002). Other residues such as Lys 12 (Go, Amyes, & Richard, 2010) and Glu 97

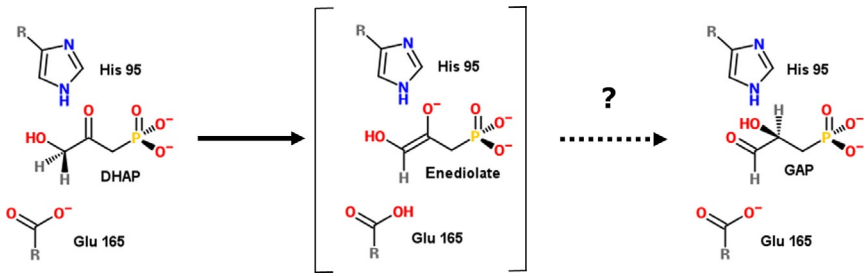

Figure 2 Scheme for the conversion of DHAP to GAP catalyzed by TIM.

(Samanta, Murthy, Balaram, & Balaram, 2011) have also been proposed to be involved in the catalytic mechanism to avoid the formation of the imidazole anion. The internal proton transfer from O1 to O2 without involving other groups has also been considered, but it seems to be energetically unfavorable in relation with other paths, at least while Glu 165 stays protonated (Cui & Karplus, 2002).

The uncertainties about the proton transfer mechanisms could be due to the fact that the energetic profiles obtained for different proton transfer paths were calculated by means of energy-minimization techniques in which the protonation state was defined *a priori*. The use of MD techniques to calculate the potential of mean force for a conformational space defined by reaction coordinates that do not presuppose any particular protonation state for the different residues can help to clarify the proton transfer mechanisms. The use of MD simulations (e.g., at $T = 300$ K) can also help to take into account important dynamical effects, such as the formation of low-barrier hydrogen bonds where the hydrogen can freely move in the space between the two heavy atoms (Cleland & Kreevoy, 1994; Ishikita & Saito, 2014).

During the reaction catalyzed by TIM, the orbital hybridization of the carbon atom C1 changes from sp^3 in DHAP to sp^2 in GAP, while the orbital hybridization of the carbon atom C2 changes from sp^3 to sp^2 (see Fig. 3). The molecular orbital rearrangement modifies the geometry of C1 and C2 carbons from planar (sp^2) to tetrahedral (sp^3), and vice versa (Fig. 3). Therefore, the dihedral angles for the C1 and C2 atoms are a good choice of reaction coordinates to monitor the reaction. Using these parameters to define the conformational space, no assumption is made about the protonation state of the different residues involved in the reaction.

4.2 Results

After 10 ns of classical MD stabilization ($T = 300$ K) of the system based on the X-ray structure of TIM (pdb: 1NEY) (Jogl, Rozovsky, McDermott, & Tong, 2003), we define the atoms to be included in the QM region (see Fig. 4). This region includes the substrate (DHAP); the side chain of catalytic residues Glu 165, His 95; the main chain amino group of residue 97 that stabilizes His 95; the main chain amino group of residues 171, 211, and 233 which stabilize the phosphate group of DHAP; and the side chain of Lys 12. This last side chain has been included in the QM region because the mutation of this residue has a dramatic effect on the catalysis

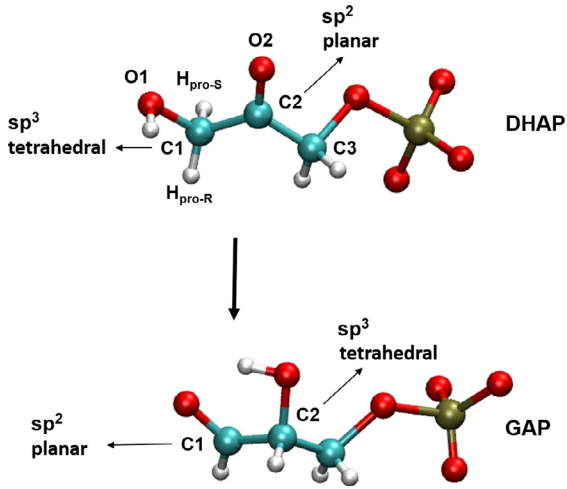

Figure 3 Orbital hybridization and geometry of dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GAP). The atoms involved in the two dihedral angles used as reaction coordinates are O1–C2–$H_{pro\text{-}S}$–C1 for the C1 dihedral angle and C1–C3–O2–C2 for the C2 dihedral angle.

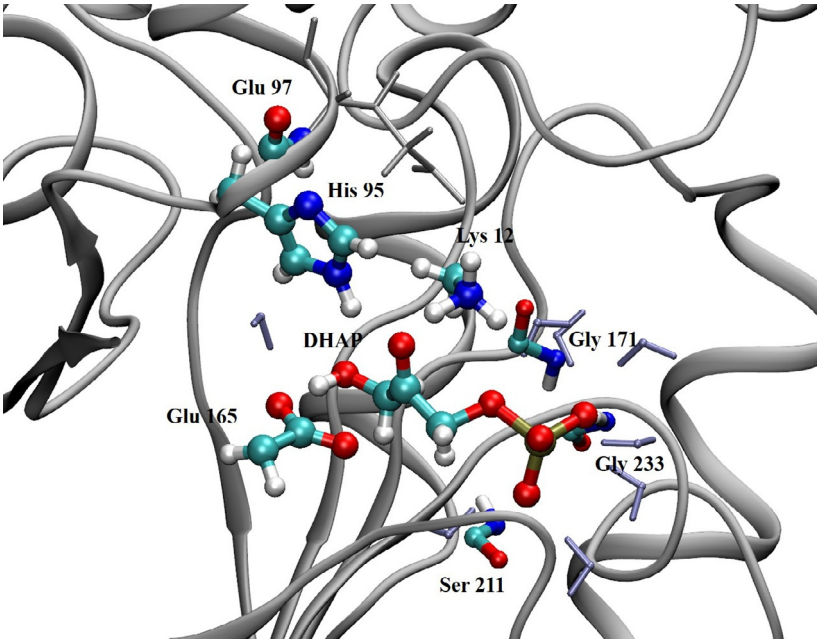

Figure 4 Active center for the reaction catalyzed by triose-phosphate isomerase. Atoms in QM region are represented in ball and sticks, the rest of the protein is presented in gray, and some solvent molecules present in the active site in blue (gray in the print version).

modifying the ratio K_{cat}/K_m (Go, Koudelka, Amyes, & Richard, 2010). Figure 4 shows a global vision of the QM region inside the protein structure.

In order to adapt the system from classical MD to DFT/MM MD, a 100-ps nonrestraint DFT/MM MD simulation was performed. Once the system is stable, we can initiate the sampling of the conformational space defined by the C1 and C2 dihedral angles (α_1 and α_2, respectively) previously established. Figure 5 shows the energetic landscape for the reaction catalyzed by TIM obtained using the techniques described in Sections 2 and 3. This landscape presents three energy minima, which can be associated with the substrate (DHAP), the intermediate state, and the product of the reaction (GAP), respectively.

The lowest energy basin corresponds to DHAP, the initial state of the reaction. The energy value of this minimum was taken as energy reference. This minimum is centered around $\alpha_1 = 30°$, which corresponds to sp^3 hybridization for the C1 atom. However, as a result of the interaction

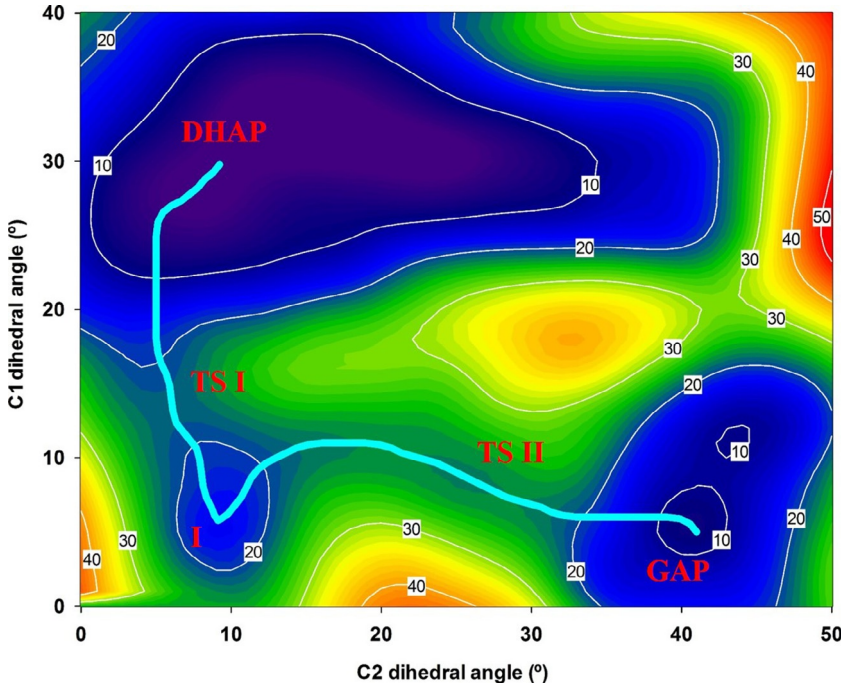

Figure 5 Free energy landscape in kcal/mol of TIM-catalyzed reaction between DHAP and GAP. The *x* and *y* axes are the two reaction coordinates, i.e., the C2 and C1 dihedral angles, α_2 and α_1, respectively. The most probable pathway is represented as a cyan (black in the print version) line.

between the negatively charged O2 atom of DHAP and some residues in the active site (Lys 12 and His 95), the range of values covered by the C2 dihedral angle α_2 in this basin is quite extensive. The second minimum presents C1 and C2 dihedral angle values corresponding to an sp^2 hybridization, $\alpha_2,\alpha_1 \sim 5°–10°$. This corresponds to the generally accepted enediolate intermediate of the reaction (see Fig. 2). The third minimum is centered around $\alpha_1 = 5°$ and $\alpha_2 = 40°$. These values are compatible with sp^2 hybridization for the C1 atom and sp^3 hybridization for the C2 atom, as corresponds to GAP, the final state of the reaction. Figure 5 also shows the lower energy-barrier path that corresponds with the most probable reaction pathway. Figure 6 shows the energy profile corresponding to this path.

The reaction path can be divided into two different steps. The first step goes from the initial state to the enediolate intermediate state through a 21-kcal/mol energy barrier. A similar barrier for this first step has been obtained in a free energy computational study (Hu, Lu, & Yang, 2007). The second step goes from the intermediate state to GAP, the reaction product. The energy barrier for this second step presents a value of 23 kcal/mol. We have also performed a DFT/MM MD simulation along our reaction path (using steered MD) in order to have a further insight into the reorganization of atoms during the catalytic mechanism (see Fig. 7).

The first step is mainly related to the rearrangement of the orbitals of the C1 carbon, which can be described through the variation of α_1 from 30°

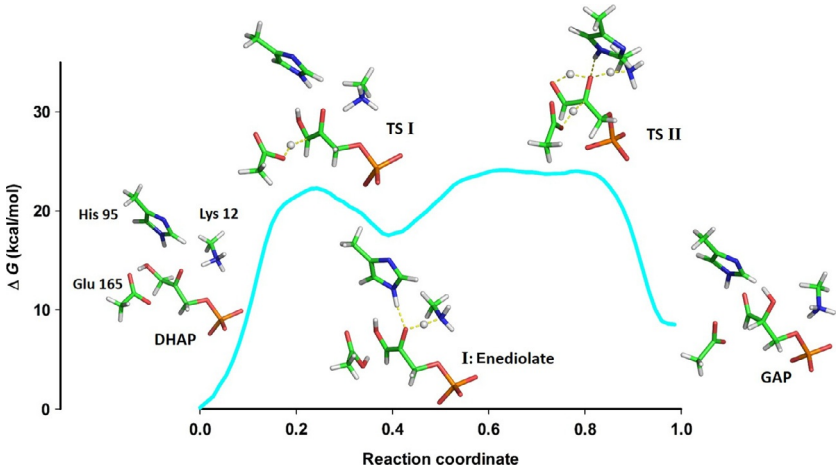

Figure 6 Energy profile of the proposed pathway for the reaction (see cyan (black in the print version) line in Fig. 5) and atomic structure for the most relevant states along the path.

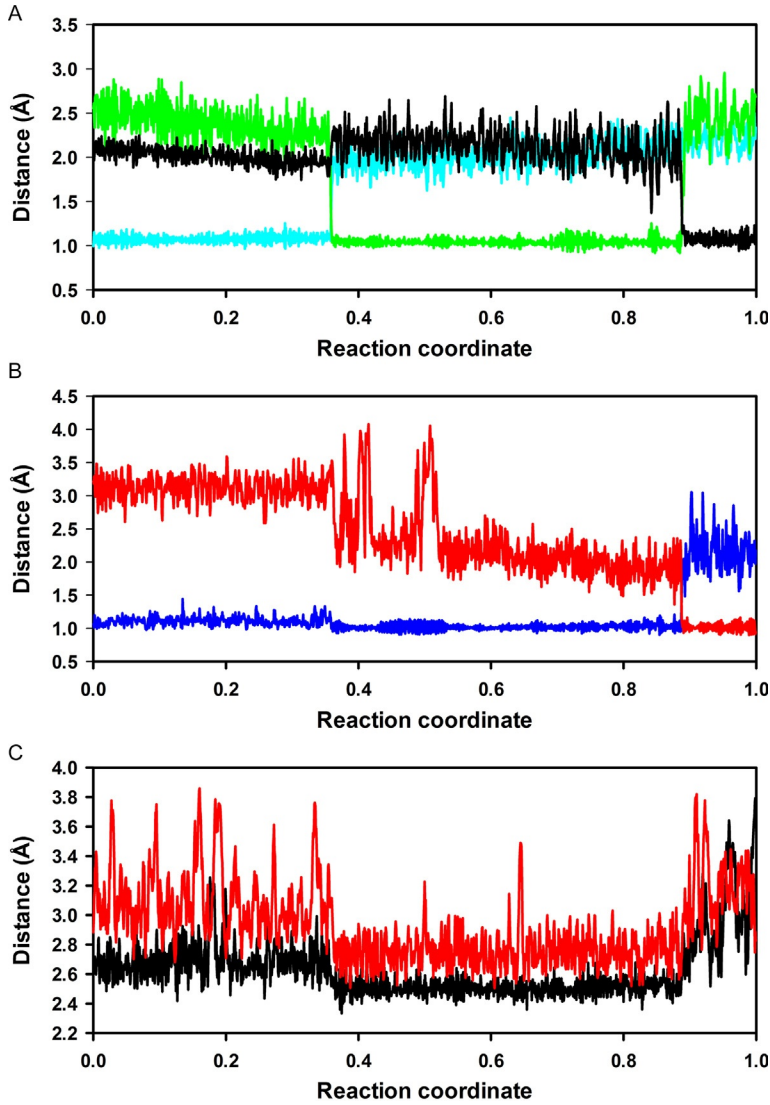

Figure 7 Distances measured during a DFT/MM MD simulation along the proposed pathway (Fig. 6). During the reaction, $H_{pro\text{-}R}$ is jumping from C1 to C2, while another H jumps from O1 to O2. (A) Distances between $H_{pro\text{-}R}$ and C1 (cyan (light gray in the print version)), the O in Glu 165 (green (dark gray in the print version)) or C2 (black). (B) Distances between H and O1 (blue (black in the print version)) or O2 (red (gray in the print version)); this H is initially bonded to O1. (C) Distances for the hydrogen bonds between O2 and nearby residues that stabilize the intermediate and transition states: distance between O2 and N of His 95 (red (gray in the print version)) and distance between O2 and N of Lys 12 (black).

(sp^3) to 5° (sp^2). During this process, the Glu 165 acts as catalytic base accepting the $H_{pro\text{-}R}$ from C1 (see Figs. 2 and 6). The enediolate forms a H-bond between DHAP O2 and the ϵN of His 95. Also, we find that the H-bond formed between the keto group of DHAP and Lys 12 adopts the character of a low-barrier H-bond where a hydrogen is shared between the ζN of Lys and the DHAP O2. These two H-bonds are the main interactions stabilizing the intermediate state.

The second step of the path undergoes through the rearrangement of the orbitals of the C2 carbon, which can be described through the variation of α_2 from 10° (sp^2) to 40° (sp^3). Along this process, the low-barrier H-bond of Lys 12 and the H-bond of His 95 stabilize the increasingly negative charge of the O2, which reaches a maximum when the reaction passes over the saddle point. In this transition state, the H is donated from O1 to O2 through a direct intramolecular transfer. *Simultaneously*, the Glu 165 acts as a catalytic acid, transferring the former $H_{pro\text{-}R}$ to the enediolate C2 atom, yielding GAP. Once the proton rearrangement has finished, the interaction between His 95 and O2 is lost. In addition, the low-barrier H-bond formed by Lys 12 becomes a typical H-bond (Fig. 7).

4.3 Discussion

The reaction mechanism described here presents some important differences with other previously proposed mechanisms. The proton from the alcohol group in C1 is directly transferred to the carboxyl oxygen in C2. This proton transfer takes place simultaneously with the proton transfer from Glu 165 to C2 (see Fig. 7). Notice also that during this process, the formation of the energetically unfavorable imidazolate anion is not required.

The intramolecular proton transfer from O1 to O2 has been considered previously and was found energetically unfavorable in relation to the transfer through the imidazole anion (Cui & Karplus, 2002). This analysis was performed using an energy-minimization approach; interestingly, Glu 165 stays protonated during these calculations. The different results obtained in our MD simulations are most likely related to the different computational approaches (energy minimization vs. dynamical simulation). In our dynamical approach, we do not make any assumptions about the protonation states during the process, allowing the protons to move freely. Thus, the simultaneous transfer of the two protons in the second step of the reaction is an unexpected outcome of our dynamical exploration. Since this is a dynamical process, it is hard to observe with minimization studies.

Another important dynamical effect observed in our simulations is the formation of the low-barrier H-bond between O2 and Lys 12 (see Fig. 7) that stabilizes the second transition state, facilitating the simultaneous proton transfer.

5. CONCLUSIONS

We have presented a detailed description of a practical DFT/MM method for the analysis of reactions in biomolecules using MD simulations. This technique combines the local-orbital DFT method FIREBALL and the AMBER suite of programs for the simulation of biomolecular systems (Mendieta-Moreno et al., 2014). Due to the good balance between computational efficiency and accuracy, this method can be used to analyze reactions in biomolecules, sampling an appropriate conformational space by means of long DFT/MM MD simulations with relatively large QM regions ($\sim 10^2$ atoms). In our dynamical analysis of these reactions, the conformational space is explored using an adaptively biased MD approach that presents characteristics of steered MD and umbrella sampling. In a typical calculation, we obtain with this procedure around 2×10^6 individual structures. This information allows us to generate a free energy landscape for the reaction and get new insights into the possible reaction mechanisms.

As an example of this approach, we have studied the conversion between DHAP and GAP catalyzed by the TIM. In our analysis, we have chosen as reaction coordinates the dihedral angles for the DHAP C1 and C2 atoms. In this way, no assumption is made about the protonation state of the different residues present in the QM region along the MD simulations. The free energy landscape presents three energy minima, which can be associated with the substrate (DHAP), the intermediate state (enediolate), and the product of the reaction (GAP). The first step of the reaction (DHAP → enediolate) presents an energy barrier of 21 kcal/mol, in agreement with previous calculations (Hu et al., 2007). Regarding the second step (enediolate → GAP), we find a new reaction mechanism that is characterized by the intramolecular proton transfer from O1 to O2 and the simultaneous proton transfer from Glu 165 to C2. This is a dynamical mechanism not found in previous studies based on energy-minimization techniques. We also find the formation of a low-barrier H-bond between O2 and Lys 12 that stabilizes the transition state from the intermediate state to the final state.

This work highlights the importance of the appropriate exploration of the conformational space for the analysis of reactions in biomolecules, using DFT/MM MD simulations to take into account the dynamically accessible structures.

ACKNOWLEDGMENTS

This work is supported by the Spanish Ministerio de Economía y Competitividad (MINECO), projects MAT2014-59966-R (J.O.) and IPT2011- 0964-900000 (P.G-P.). 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.

REFERENCES

Adcock, S. A., & Mccammon, J. A. (2006). Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews*, *106*(5), 1589–1615.

Basanta, M. A., Dappe, Y. J., Jelínek, P., & Ortega, J. (2007). Optimized atomic-like orbitals for first-principles tight-binding molecular dynamics. *Computational Materials Science*, *39*, 759–766.

Bash, P. A., Field, M. J., Davenport, R. C., Petsko, G. A., Ringe, D., & Karplus, M. (1991). Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry*, *30*(24), 5826–5832.

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, *38*, 3098.

Carlson, B. C., & Keller, J. M. (1957). Orthogonalization procedures and the localization of wannier functions. *Physical Review*, *105*, 102–103.

Case, D. A., Darden, T. A., Cheatham, T. E., III, Simmerling, C. L., Wang, J., Duke, R. E., et al. (2012). *AMBER 12*. San Francisco: University of California.

Cleland, W. W., & Kreevoy, M. M. (1994). Low-barrier hydrogen bonds. *Science*, *264*, 1887–1890.

Cleveland, W. S., & Devlin, S. J. (1986). Locally weighted regression: An approach to regression analysis by local fittings. *Journal of the American Statistical Association*, *83*(403), 596–610.

Cui, Q., & Karplus, M. (2002). Quantum mechanical/molecular mechanical studies of the triosephosphate isomerase-catalyzed reaction: Verification of methodology and analysis of reaction mechanisms. *Journal of Physical Chemistry B*, *106*(7), 1768–1798.

Demkov, A. A., Ortega, J., Sankey, O. F., & Grumbach, M. P. (1995). Electronic structure approach for complex silicas. *Physical Review B*, *52*, 1618–1630.

Field, M. J., Albe, M., Bret, C., Proust-De Martin, F., & Thomas, A. (2000). The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *Journal of Computational Chemistry*, *21*(12), 1088–1100.

Field, M. J., Bash, P. A., & Karplus, M. (1990). A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *Journal of Computational Chemistry*, *11*, 700–733.

Foulkes, W. M. C., & Haydock, R. (1989). Tight-binding models and density-functional theory. *Physical Review B*, *39*, 12520–12536.

García-Vidal, F. J., Merino, J., Pérez, R., Rincón, R., Ortega, J., & Flores, F. (1994). Density-functional approach to lcao methods. *Physical Review B*, *50*, 10537–10547.

Gilson, M. K., & Zhou, H. X. (2007). Calculation of protein-ligand binding affinities. *Annual Review of Biophysics and Biomolecular Structure*, *36*, 21–42.

Go, M. K., Amyes, T. L., & Richard, J. P. (2010). Rescue of K12G triosephosphate isomerase by ammonium cations: The reaction of an enzyme in pieces. *Journal of the American Chemical Society*, *132*(38), 13525–13532.

Go, M., Koudelka, A., Amyes, T., & Richard, J. (2010). Role of Lys-12 in catalysis by triosephosphate isomerase: A two-part substrate approach. *Biochemistry*, *49*(25), 5377–5389.

Hall, A., & Knowles, J. R. (1975). The uncatalyzed rates of enolization of dihydroxyacetone phosphate and of glyceraldehyde 3-phosphate in neutral aqueous solution. The quantitative assessment of the effectiveness of an enzyme catalyst. *Biochemistry*, *14*(19), 4348–4353.

Harris, J. (1985). Simplified method for calculating the energy of weakly interacting fragments. *Physical Review B*, *31*, 1770.

Hohenberg, P., & Kohn, W. (1964). Inhomogeneous electron gas. *Physical Review*, *136*, B864–B871.

Horsfield, A. P. (1997). Efficient ab initio tight binding. *Physical Review B*, *56*, 6594–6602.

Hu, H., Lu, Z., & Yang, W. (2007). QM/MM minimum free energy path: Methodology and application to triosephosphate isomerase. *Journal of Chemical Theory and Computation*, *3*(2), 390–406.

Ishikita, H., & Saito, K. (2014). Proton transfer reactions and hydrogen-bond networks in proton transfer reactions and hydrogen-bond networks in protein environments. *Journal of the Royal Society Interface*, *11*(91), 20130518.

Jelínek, P., Wang, H., Lewis, J., Sankey, O., & Ortega, J. (2005). Multicenter approach to the exchange-correlation interactions in ab initio tight-binding methods. *Physical Review B*, *71*, 235101.

Jogl, G., Rozovsky, S., McDermott, A. E., & Tong, L. (2003). Optimal alignment for enzymatic proton transfer: Structure of the Michaelis complex of triosephosphate isomerase at 1.2-A resolution. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(1), 50–55.

Jorge, M., Garrido, N. M., Queimada, A. J., Economou, I. G., & MacEdo, E. A. (2010). Effect of the integration method on the accuracy and computational efficiency of free energy calculations using thermodynamic integration. *Journal of Chemical Theory and Computation*, *6*(4), 1018–1027.

Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, *9*, 646–652.

Kohn, W., & Sham, L. J. (1965). Self-consistent equations including exchange and correlation effects. *Physical Review*, *140*, A1133.

Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, *13*(8), 1011–1021 (John Wiley & Sons, Inc.).

Laio, A., & Parrinello, M. (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12562–12566.

Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, *37*, 785.

Lewis, J. P., Glaesemann, K. R., Voth, G. A., Fritsch, J., Demkov, A. A., Ortega, J., et al. (2001). Further developments in the local-orbital density-functional-theory tight-binding method. *Physical Review B*, *64*, 195103.

Lewis, J. P., Jelinek, P., Ortega, J., Demkov, A. A., Trabada, D. G., Haycock, B., et al. (2011). Advances and applications in the FIREBALL ab initio tight-binding molecular-dynamics formalism. *Physica Status Solidi B*, *248*, 1989–2007.

Lodi, P. J., & Knowles, J. R. (1991). Neutral imidazole is the electrophile in the reaction catalyzed by triosephosphate isomerase: Structural origins and catalytic implications. *Biochemistry*, *30*(28), 6948–6956.

Löwdin, P. (1950). On the nonorthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *The Journal of Chemical Physics*, *18*, 365.

Martin, R. M. (2004). *Electronic structure*. Cambridge (UK): Cambridge University Press.

Martín-García, F., Mendieta-Moreno, J. I., López-Viñas, E., Gómez-Puertas, P., & Mendieta, J. (2012). The role of Gln61 in HRas GTP hydrolysis: A quantum mechanics/molecular mechanics study. *Biophysical Journal*, *102*, 152–157.

Marx, D., & Hutter, J. (2009). *Ab initio molecular dynamics*. Cambridge (UK): Cambridge University Press.

Mendieta-Moreno, J. I., Walker, R., Lewis, J. P., Gomez-Puertas, P., Mendieta, J., & Ortega, J. (2014). FIREBALL/AMBER: An efficient local-orbital DFT QM/MM method for biomolecular systems. *Journal of Chemical Theory and Computation*, *10*, 2185–2193.

Park, S., Khalili-Araghi, F., Tajkhorshid, E., & Schulten, K. (2003). Free energy calculation from steered molecular dynamics simulations using Jarzynskis equality. *The Journal of Chemical Physics*, *119*, 3559.

Reed, A. E., Weinstock, R. B., & Weinhold, F. (1985). Natural population analysis. *The Journal of Chemical Physics*, *83*, 735.

Řezáč, J., & Hobza, P. (2012). Advanced corrections of hydrogen bonding and dispersion for semiempirical quantum mechanical methods. *Journal of Chemical Theory and Computation*, *8*, 141–151.

Řezáč, J., Riley, K. E., & Hobza, P. (2011). S66: A well-balanced database of benchmark interaction energies relevant for biomolecular structures. *Journal of Chemical Theory and Computation*, *7*, 2427–2438.

Salomon-Ferrer, R., Case, D., & Walker, R. (2012). An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science*, *3*, 198–210.

Samanta, M., Murthy, M. R. N., Balaram, H., & Balaram, P. (2011). Revisiting the mechanism of the triosephosphate isomerase reaction: The role of the fully conserved glutamic acid 97 residue. *ChemBioChem*, *12*(12), 1886–1896.

Sankey, O. F., & Niklewski, D. J. (1989). Ab initio multicenter tight-binding model for molecular-dynamics simulations and other applications in covalent systems. *Physical Review B*, *40*, 3979–3995.

Schönhammer, K., Gunnarsson, O., & Noack, R. M. (1995). Density-functional theory on a lattice: Comparison with exact numerical results for a model with strongly correlated electrons. *Physical Review B*, *52*, 2504.

Senn, H. M., & Thiel, W. (2009). QM/MM methods for biomolecular systems. *Angewandte Chemie*, *48*, 1198–1229.

Stewart, J. J. P. (2004). Optimization of parameters for semiempirical methods IV: Extension of MNDO, AM1, and PM3 to more main group elements. *Journal of Molecular Modeling*, *10*(2), 155–164.

Walba, H., & Isensee, R. W. (1961). Acidity constants of some arylimidazoles and their cations. *The Journal of Organic Chemistry*, *26*(8), 2789–2791.

Walker, R. C., Crowley, M. F., & Case, D. A. (2008). The implementation of a fast and accurate QM/MM potential method in Amber. *Journal of Computational Chemistry*, *29*, 1019–1031.

Warshel, A., & Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, *103*, 227–249.

Zhang, Y. (2005). Pseudobond ab initio QM/MM approach and its applications to enzyme reactions. *Theoretical Chemistry Accounts*, *116*, 43–50.

Zwanzig, R. W. (1954). High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics*, *22*(8), 1420–1426.