

PRIMoRDiA Meets EasyHybrid3: Unlocking the Exploration of the Electronic Structure of Biomolecules Through Hybrid QC/MM Simulations On the Fly.

Igor Barden Grillo,[†] José Fernando Ruggiero Bachega,[‡] Martin J. Field,[¶] Larissa Pereira,[†] and Gerd Bruno Rocha[†]

[†]*Departament of Chemistry of Federal University of Paraíba, João Pessoa - PB, Brazil*

[‡]*Department of Pharmacosciences, Federal University of Health Sciences of Porto Alegre, Centro Histórico, Porto Alegre- RS,90050-170, Brazil.*

[¶]*Laboratoire de Chimie et Biologie des Métaux, UMR5249, Université Grenoble I, CEA, CNRS, 17 avenue des Martyrs, 38054 Grenoble Cedex 9, France*

[§]*Theory Group, Institut Laue-Langevin, 71 avenue des Martyrs CS 20156, 38042 Grenoble Cedex 9, France*

E-mail:

Introduction

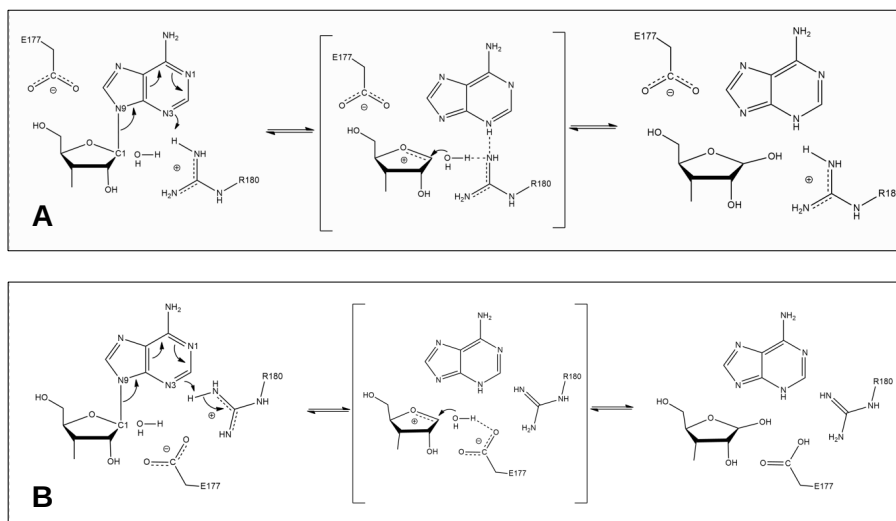


Figure 1

Methods

QC/MM Simulations Set Up

In the present simulations and analysis, we picked the PDB file containing the structure treated by the Fouzia Group. The system presents the substrate strongly bind on the active pocket to several hydrogen bonds and with a short distance from the catalytic histidine, being to a characteristic near attack conformation for the proton abstraction. Although, the distance from the cofactor is significant large. Other thing to be analyzed is the possibility of histidine activation by the aspartate.

Thus we perfomed preliminary relaxed surface two-dimensional scans to determine the protonation of the HIS-240 and to find a potential well structure with the nicotimide-ring closer to the substrate.

Respecting the initial protonation and charge states of the active pocket, showed in Figure 1, the system was prepared for QC/MM simulations with the following steps: i)

parametrization of the cofactor NADP; ii) parametrization of the substrate; ii) solvation of the complex, ions addition to neutralize the total charge and protein parameters generation with AMBER force field; iv) GROMACS energy minimization of the entire system. After these steps, the force field parameters and coordinates files were loaded in the pDynamo library, where the system were prepared, QC/MM atoms, and fixed atoms were defined and the further simulations were performed.

the library files for tleap were generated with parmchk2 from ambertools. The glucose-6-phosphate parameters were generated with am1-bcc charge scheme using antechamber program considering -2 of formal charge. We used the amberff99SB force filed for the protein atoms, tip3p model to build a solvation cubic box respecting the 12Å between the surface of protein complex until the box end. Also, seven sodium ions were added to neutralize the total system charge.

We build the complex parameters and coordinates using tleap program, the parameters files were converted to GROMACS format using the Parmed library for Python 3, for energy minimization. We performed this minimization using steepest descent algorithm, periodic boundary conditions in all dimensions, 50000 maximum number of steps and 1000 kJ/mol/nm force tolerance and 0.01 of energy step.

The subsequent coordinates and the force field parameters were loaded and treated by the library classes and functions of pDynamo[?] and EasyHybrid[?]. Firstly, the box was set to a sphere, taking the hydride (H1) as the center and including all atoms of the residues with up to 30 Å of distance. The atoms of the residues within 22 Å from the center was considered as mobile atoms and beyond as fixed, i.e., they only interact with the other atoms but without to get their coordinates updated. This avoids unnecessary gradient calculations without losing the effect of this outer layer. The distances were chosen to consider the NADP molecule present in the other biding pocket, as well to provide a minimum water molecules layer. After that, a geometry optimization was performed using pDynamo conjugated gradient minimizer with a 0.1 kJ.mol⁻¹.Å⁻¹ for root mean square tolerance.

The most critical definition is the atoms that will be treated with the quantum chemistry potential. An unnecessary inclusions of atoms, in respect to the description of the reactions, could lead to unpractical simulations times for energy calculations employing DFT functionals or further free energy evaluations. Thus, we included only the most important moieties to the proposed catalysis mechanism, which includes: the nicotinamide ring atoms from NADP; all glucose-6-phosphate atoms; the imidazole ring atoms of the catalytic histidine-337 and the near histidine-175; amine group of the lysine-145 and the carbon bind with the nitrogen, which is a residue very near the oxygen O1 of the substrate; and finally the carboxylic oxygens and carbon of the aspartate-174, that could activate the HIS237. The pruned system and division set up is shown in the ??.

QC/MM Initial Relaxed Surface Scans

Before the simulations for the determination of the reaction path, we performed some preliminary simulations to determine the protonation state of the histidine-337 and subsequently the initial position of the NADP cofactor.

Firstly, the histidine protonation was addressed by two-dimensional relaxed surface scans. In this type of simulation we performed a survey in the reaction coordinates, forcing a displacing in the combined distance between the atoms with aid of harmonic potential restriction. For each step, the geometry optimization is carried and thus is why these survey is called as "relaxed".

Also, these simulations were performed using the algorithms implemented on pDynamo and automatized by the EasyHybrid core scripts. The first reaction coordinated considered was the displacement of the proton H01(G6P) from the O1(G6P) to the direction of the NE2(HIS337). The second one was the displacement of the proton HD1(HIS337) from the ND1(HIS337) to the oxygen OD1(ASP174). We treated these two reaction coordinates equally, applying an harmonic potential restriction energy with a $4000 \text{ kJ.mol}^{-1} \text{ \AA}^{-2}$ of force constant, scanning twenty steps in each direction with 0.05 \AA of distance step size increment, resulting in a two-dimensional 20X20 grid.

This step size and force constants were chosen to perform a fine grid survey on these two reaction coordinates. Reactions path of proton transference often presents big displacements on the reaction coordinate distances, being more easy to lose the transition state structure information.

We performed these calculations for the six semiempirical Hamiltonians implemented on pDynamo software that have parameters available for the atoms of the deined quantum region: AM1, AM1dPhot, PM3, pddgPM3, RM1 and PM6. The energy of each of the grid point optimized structure were calculated to generate the Potential Energy Surface (PES) in respect to these two reaction coordinates, as showed in the ?? for the PM3 Hamiltonian.

The PES shows two potential wells with the lowest energies, the initial positions, or the

reactants (R), and the other signalized by the red star on the figure. This second structure corresponds to the HE2(HIS337) transferred to the aspartate without the H01(G6P) transference to the histidine. The energy barrier for this saddle point (SP) is about 64 kJ/mol(15 Kcal/mol), where the system gains 21 kJ/mol of potential energy. Thus we pose that this energy gain and low barrier are consistent with the activation of the histidine by the aspartate.

The direct proton abstraction of the B6G by the histidine (path R-SP-P* in ??) presented a barrier of about 103 kJ/mol(25 kcal/mol), and after the activation (path *-SP-P**) the barrier is about 75 kJ/mol (18 kcal/mol), forming a most stable product (P**). Thus is plausible to expect that the aspartate activate the histidine before the substrate transfer his proton and hydride. Thus, for the subsequent relaxed surface scans, we used as initial coordinates the one corresponding to the * in ??.

Thus, starting with the histidine-337 deprotonated, we performed another two-dimensional relaxed surface scans using as the first reaction coordinate the displacement of the proton H01(G6P) from the O1(G6P) to the direction of the NE2(HIS337) and the second the displacement of the hydride H1(G6P) from the substrate carbon C1(G6P) to the cofactor carbon C4N(NADP). Despite the initial position of the substrate in the active pocket, the initial coordinates of the NADP cofactor are unfavorable for the reaction studies, with a 5.4 Å distance from C4N(NADP) to the H1(G6P). Thus, we employed a larger number of steps and larger step size for the second reaction coordinate.

For the first one direction we chosen 12 steps of 0.1 Å size, using a harmonic potential with a $3000 \text{ kJ.mol}^{-1} \text{ Å}^{-2}$ of force constant. For the second direction, the hydride transfer, we chosen 30 steps of 0.2 Å with a $1500 \text{ kJ.mol}^{-1} \text{ Å}^{-2}$ of force constant for the harmonic potential. For larger increment steps we employed smaller energy constants for the harmonic restrictions, to allow the reaction coordinate to relax more and not produce abnormal energy peaks. In the case of the reaction coordinate involving the NADP atom, we employed a larger step size, because the initial distance, and a lower force constant in the reaction coordinate

to allow all atoms of the nicotinamide ring to approximate the substrate.

With this analysis we performed a survey on these two reaction coordinates, focusing in finding a potential well where the co-factor would be in a more energetic favorable position for the reaction occur. The results of the 12X30 scan is shown in the Figure 2, where we clearly find a first potential well after a displacement of more than 2 Å in the hydride transfer coordinate, signalized by the red star on energy map. From this point, the finishing of the hydride transfer without changing in the proton transfer coordinate, present a barrier of about 150 KJ/mol. Otherwise, the finishing of the proton transfer would require 100 KJ/mol to reach the transition state.

Figure 2: Two-dimensional potential energy surface for the reaction coordinates of the proton H01(G6P) transference from the substrate to histidine-337 and the hydride transfer to the NADP.

Though, the displacements were large on the hydride transfers reaction coordinate, more specifically near the likely transition state structures. Thus, for a reaction path determination we need to calculate a more fine grid, with smaller steps. However, the results of Figure 2 provided to us an initial geometry with lower potential energy and with the NADP closer to the substrate, indicated by the red star. In the Figure 3 we show in detail this geometry of the active site, from the perspective of the reaction coordinates (Figure 3:A), with the atoms showed connected by the magenta dashed lines. In the Figure 3:B, we show the active site from the perspective of phosphate group of BG6, showing the several hydrogen bonds formed between the enzyme and the substrate.

Figure 3: Initial coordinates for the further simulations for reaction path determination. A) Initial position for the further simulations and the distances between the atoms of the reaction coordinates (magenta dashed lines). B) Initial position of the active site visualized from the perspective of the substrate phosphate group showing the hydrogen bonds between the enzyme and the G6P.

QC/MM Reaction Path Determination

From the previous simulations, we determined our guess of starting geometry, with the deprotonated histidine and the distance between the carbon C4N(NADP) and the hydride H1(G6P) of 2.7 Å. Again, we performed another two-dimensional relaxed surface scan using the hydride and proton transference coordinates from G6P to the NADP and histidine respectively. For the proton transference, we kept the same parameters, 12 steps of 0.1 Å employing distance dependent harmonic biasing potential with a $4000 \text{ kJ.mol}^{-1} \text{ Å}^{-2}$ of force constant. For the hydride transference reaction coordinate, we chosen 20 steps with 0.1 Å of size, using the same $1500 \text{ kJ.mol}^{-1} \text{ Å}^{-2}$ force constant.

This makes a 12X20 grid for the relaxed surface scan, resulting in the PES showed in the ??A calculated with the PM3/AMBER hybrid potential. In this PES description of the studied reaction, we can see clearly three potential wells, reactatns **R**, a transient state **R*** and the products **P**, and two saddle points **TS** and **TS**. Connecting these points, we can retrieve a plausible guess of the reaction path, which occurs with the transference of the G6P proton to the HIS337 first, with a barrier of 48.74 kJ/mol, followed by the hydride transfer of the G6P to NADP, presenting a 84.78 kJ/mol of energy barrier. The energy profile for the reaction path is showed in ??B for PM3 and the others semiempirical Hamiltonians.

In the ??, we show key structures of the reaction path and the energy profile refined with DFT/MM hybrid potential. We can observe that lysine-145 is near the O1 from G6P throughout the path, possibly playing a important role of charge stabilization. All other near residues are interacting with the substrate through hydrogen bonds, holding it in the active pocket. The B3LYP/6-311G*/AMBER refinement corroborates the two transition states barriers, lowering significantly the first and slightly increasing the second.

In summary, from these simulations, we undertand that the reaction mechanism catalyzed by the G6PDH occurs with the histidine-337 activated by the aspartate-174, the transference of the proton H01 from G6P to the histidine-337 followed by the hydride H1 transference from G6P to NADP. Besides the base role played by the histidine, we identified a possible

charge stabilizer role in lysine-145.