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:

Abstract

Hybrid simulations are powerful computational chemistry tools to obtain structural and thermodynamics information about enzymatic catalysis reactions. In addition to such relevant applications, these calculations also provide the resolved electronic structure, which is require to estimate the electronic energy in the SCF process. Recently, our group is using such information to obtain several quantum chemical descriptors to profile the steering forces underlying biological processes. In our's group recent main publications we presented results about: how semiempirical Hamiltonian could be used to obtain the Fukui functions from polypeptides with modified descriptors, applying them to ligand-binding process; to profile enzymatic catalysis reaction coordinates; to study the main interactions on the active site of main protease SARS-COV-II, comparing with from the SARS-COV-I. The main goal of this work is to present our software PRIMoRDiA and its workflow with QC/MM hybrid software as the new pDynamo3 and its Graphical interface EasyHybrid3. PRIMoRDiA is an application to deal efficiently with large outputs of quantum chemical programs and to calculate more than thirty global and local quantum chemical descriptors.

DISCLAIMER: this work is associated with a poster presentation on CE-CAM event: *Present and Future of Hybrid Quantum Chemical and Molecular Mechanical Simulations.* Please, do not use for nothing than to have the full details of our work. This text follow a manuscript formatting, but the central goal is to complete the information of the poster, so writing errors and lack of references may occur.

Also, pDynamo3 was used under the permission of professor Martin Field, the software new version is under development. EasyHybrid3 has no official stable version as well, but will be soon launched.

Introduction

The developing of a computational protocols such the quantum chemical/quantum mechanical hybrid schemes, allowed the study of complex process that occurs in large biomolecular systems. This turned the Hybrid QC/MM methods the state of art and it comes to the energetic and structural studies of enzymatic catalysis. Although, in the great majority of those simulations the resolved electronic structure is trow away, i.e. the wave function information is not utilized even though the computational cost to obtain it already was paid.

PRIMoRDiA software is a new application released to the community prepared to deal with large output files from quantum chemical packages, transforming the electronic structure information in quantum chemical descriptors, global and local, that describes the reactivity propensity and/or provide some electronic properties that can be to used to train correlation models.

We already used PRIMoRDiA in some studies for profile and characterize reaction paths of enzymatic reactions, from trajectories generated with QC/MM simulations, such as the energy profiling of three enzymatic reactions: *Triosephosphate Isomerase*, *Haloalkane Dehalogenase* and *Adenosine Kinase.* As well the theoretical characterizations of the *Shikimate*

Dehydrogenase, depicting the roles of the active site amino-acids.

Thus, we already show that reactivity descriptors, these theoretical quantities calculated with the resolved electronic structure, can aid us to understand the chemistry of such complex reaction.

In the present work, we chosen to analysis the reactivity from a set of sampled structures from QC/MM simulations of the complex between the Ricin Toxic Chain A (RTA) enzyme complex with a fragment of its natural subtract, that is a motive of the ribossome RNA. This system is very interesting to study for some reasons: the inhibition of this enzyme turns eliminate the risks associated with working the castor plant oil, which have several industrial and green chemistry applications; and secondly, there are two proposed mechanism of the cleavage catalysis, each one involving the needing to scan at least two reaction coordinates simultaneously.

We show in Figure 1 the two possible mechanisms, the first being the proton transference to the arginine to the adenine nitrogen, the cleavage of the ribose bond followed by an nucleophilic attack of the water oxygen in the ribose carbon and the proton abstraction of this water by the same arginine, restoring its protonation state.
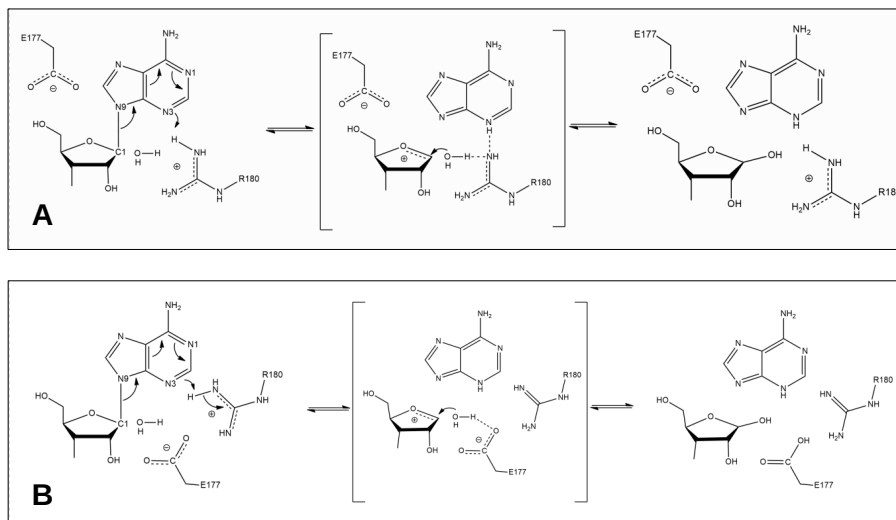
Figure 1: Proposed mechanisms for the cleavage of a ribossome RNA motive catalyzed by the RTA. A) Proton transference to the arginine to the adenine nitrogen, the cleavage of the ribose bond followed by an nucleophilic attack of the water oxygen in the ribose carbon and the proton abstraction of this water by the same arginine, restoring its protonation state. B) Same arginine from before, followed by the catalytic glutamic acid activation of the water before the nucleophilic attack of the ribose break bond.

The second mechanism considers the action of an glutamic acid residue, activating the water before the nucleophilic attack, keeping the same role for the arginine but without restoring its protonation state.

In a relaxed survey of the reactions coordinates we could consider a range of its combinations. Generally, we could use the initial distances of between the atoms to have a good guess of the most probable reaction coordinates to make the scan. Thus, this includes to have a reliable guess of the starting coordinates.

In fact, for the complex we are dealing with, the natural substrate it the ribossome itself, which is impractical to include in any time of simulations where we want to apply QC/MM hybrid potential to capture bond breaking and formation. Thus, our complex has as substrate an artificial analogous molecule, imitating the GAGA motive of the *Sarcin-Ricin Loop* region, located in the portion 285 of the ribossomic RNA. Then, this makes more difficult to define how the chosen starting geometries are reliable.

Thus, we will make a structural analysis based on the distance variation during the QC/MM molecular dynamics, and compare with structures sampled from the statistical analysis of the reactivity descriptors obtained for each saved frame.

Methods

In this study, we performed QC/MM molecular dynamics sampling of the Ricin Toxic Chain A (RTA) enzyme complex with a fragment of its natural subtract, that is a motive of the Ribossome RNA. This was followed by the energy refinement in MOPAC software to obtain the files with resolved electronic structure, using a larger QC region. We passed these files for PRIMoRDiA calculate the quantum chemical descriptors, and generate R scripts for statistical analysis and Pymol Scripts for visualization of the descriptors.

We performed an distance analysis between the two possible reaction coordinates, using the EasyHybrid3 core scripts, where we used to find the most probable distances in the QC/MM molecular dynamics runs.

We take the values of the descriptors summed over their belonging atoms, and used to make a Principal Component Analysis, to find the most relevant quantum chemical descriptors to analysis the reactivity of the system over the sampled structures. Then we take the main descriptors to sample the frame with the most probable value in they. We summarized Figure 2 all these computational protocols and analysis that was used in this work.
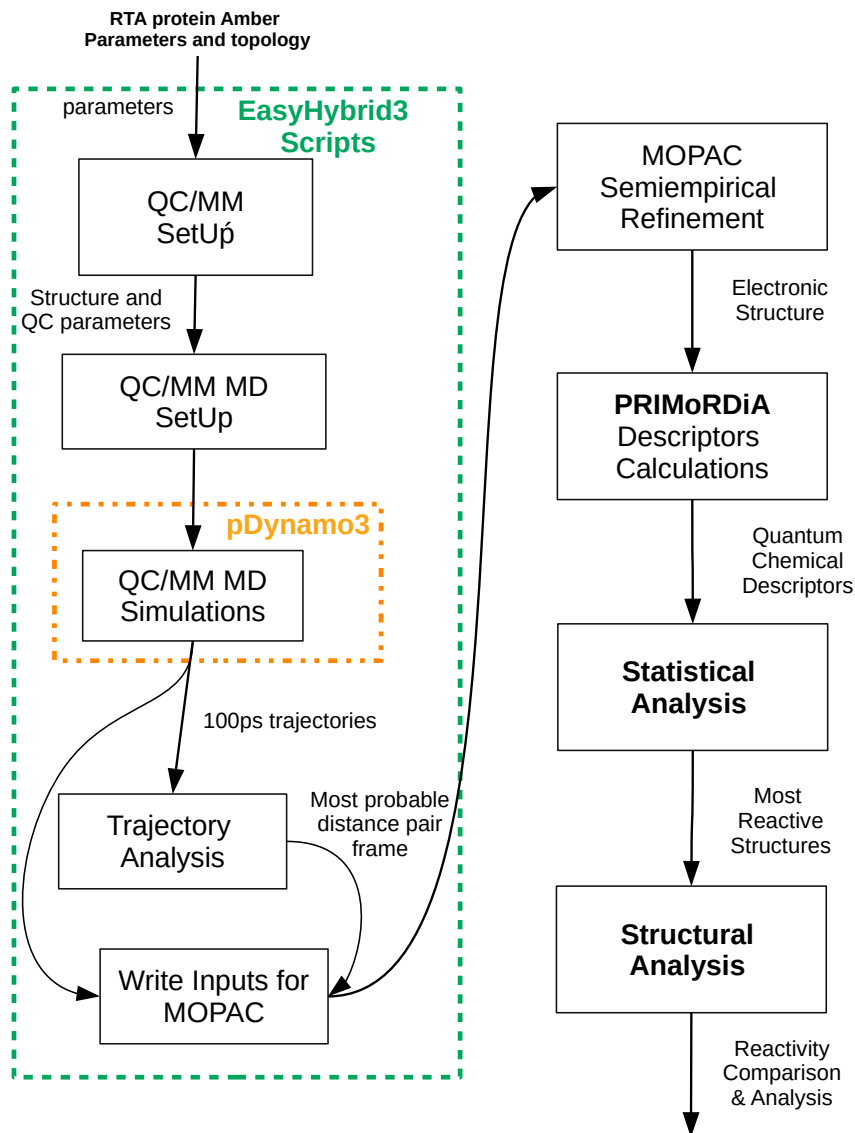
Figure 2: Flowchart depicting the pupeline and compuational protocols used in this work.

To do this, we picked the parameters/topology and prepared coordinates built in a master dissertation work within our group. The nucleotide parametrization was done using the CHARMM General Force Field (CGenFF), in the web-server: cgenff.umaryland.edu. This system was built using the VDM tool QwikMD, using the CHARMM protein force field for the RTA amino-acids. The solvation box was built with same tool cited above, with the TIP3P model and a size of 20Å from the most distances coordinates of the complex until to end of the box. We neutralized the net charge of the system with Na+ ions plus more Na+

and Cl- to mimic a NaCl concentration of 0.15mol/L.

QC/MM Simulations Set Up

The coordinates and the force field parameters were loaded and treated by the library classes and functions of pDynamo[1] and EasyHybrid[2]. Firstly, the box was set to a sphere, taking the nitrogen of the catalytic arginine (NH1) 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. The distances were chosen to consider all substrate atoms 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. We included only the most important moieties to the proposed catalysis mechanism for the molecular dynamics runs, which includes: the side chain of catalytic residues arginine and glutamic acid, the water molecule near the substrate, and the ribose and adenine atoms of the second nucleotide of the pseudo ligand.

In the Figure 3, the quantum region for the molecular dynamics is show in green (carbons only) sticks, the line representation is the MM free atoms, and in gray are represented the fixed atoms. All MD simulations were done using Leap Frog integrator implemented in pDynamo3, with a 0.001ps integration step, at 300.15K. For a equilibration step, we employ 20000 steps (20ps) and to sample the properties 100000 steps (100ps), saving frame coordinates for each 1ps. We replicate these runs using the the Hamiltonians present in the software with parameters for the atoms in the considered system: AM1, AM1dPhot, RM1, PM3, pddgPM3 and PM6.
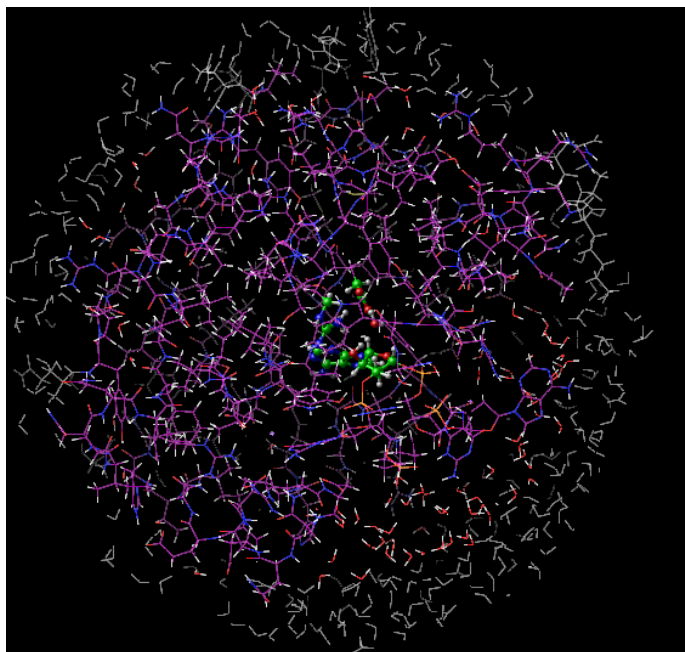
Figure 3: Divisions of the hybrid system for the molecular dynamics simulations. Quantum region: carbon green and sticks; molecular mechanis mobile atoms: carbon in purple and lines; molecular mechanics fixed atoms all in gray.

After the molecular dynamics sampling, we used the Easyhybrid core scripts to generate the Mopac inputs for energy refinement considering a larger QC/MM regions, which includes some side chains of the residues around the attack site, which is represented in Figure 4. We done the refinement for the coordinates obtained with the AM1/MM hybrid potential MD, that showed to be the most consistent in the analysis. For these 100 sample structures, we performed the energy refinement for all base Hamiltonians present in mopac: AM1, RM1, PM3, PM6 and PM7.
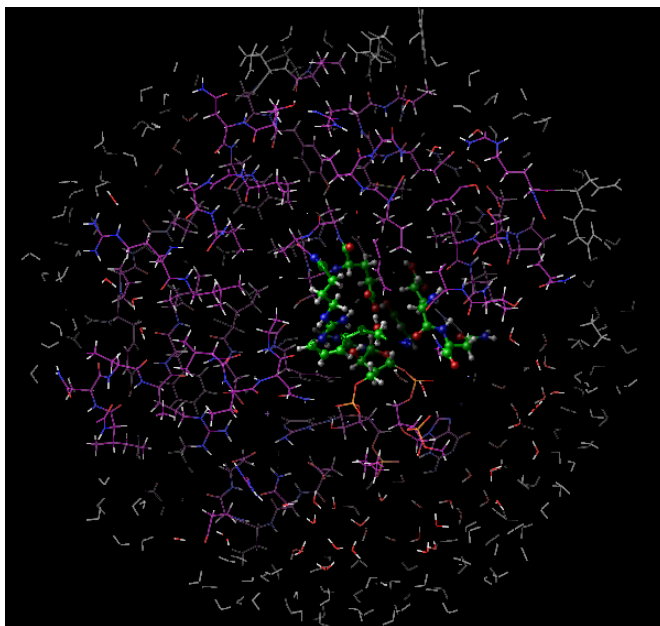
Figure 4: Divisions of the hybrid system for the molecular dynamics simulations. Quantum region: carbon green and sticks; molecular mechanis mobile atoms: carbon in purple and lines; molecular mechanics fixed atoms all in gray.

Reactivity Descriptors

he resultant electronic structure files were provided to PRIMoRDIA software[?] . PRIMoRDiA is built over a specialized code for calculations of electronic properties for macromolecules, implementing several global and local descriptors. We focused on the local reactivity descriptors due to the nature of the study. The first to consider are the Fukui functions, that rises from the derivative of the electron density in respect of the number of electrons (Equation 1)[3].

$$f(r) = \left(\frac{\partial \rho(r)}{\partial N} \right)_{\nu} \tag{1}$$

Due to the discontinuity of in the number of electrons this derivative is taken from different limits giving the origin to local functions that describe the susceptibility of the system

to be attacked by a electrophile or a nucleophile. This functions can be calculated from the molecular orbitals HOMO (Equation 2) and LUMO (Equation 3), by the approximation of frozen orbitals, respectively named as electrophilicity and nucleophilicity in PRIMoRDiA. The descriptors are calculated for each atom k and thus summed over by their belonging amino-acid residue.

$$f^-(k) = |\psi^{HOMO}(k)|^2 = \sum_{\nu \in k}^{AO} \left\{ |C_{\nu HOMO}|^2 + \sum_{\mu \notin \nu}^{AO} |C_{\nu HOMO} C_{\mu HOMO}| S_{\mu\nu} \right\} \qquad (2)$$

$$f^+(k) = |\psi^{LUMO}(k)|^2 = \sum_{\nu \in k}^{AO} \left\{ |C_{\nu LUMO}|^2 + \sum_{\mu \notin \nu}^{AO} |C_{\nu LUMO} C_{\mu LUMO}| S_{\mu\nu} \right\} \qquad (3)$$

This functions can be averaged (Equation 4) or subtracted (Equation 5), giving origin to the radical and dual descriptor respectively, being the latter called netphilicity in the PRIMoRDiA implementation. The Fukui functions are the basis of majority of the local functions used in the chemical reactivity theories. Although, for macromolecules, PRIMoR-DiA implement modifications on the calculation electrophilicity and nucleophilicty through these functions, accounting the contribution of other molecular orbitals with similar energy, as the protein systems are highly degenerated and may present their reactivity distributes along several atoms.

$$f^0(k) = \frac{f^+(k) + f^-(k)}{2} \qquad (4)$$

$$\Delta f^\pm(k) = f^+(k) - f^-(k) \qquad (5)$$

For this study, we employed the Energy Weighted method found on PRIMoRDIA, using a 3eV of energy band to combine the molecular orbitals. These descriptors based on the frontier molecular orbitals are related to the soft-soft type of interactions that governs the chemical processes, characterized by high polarization of electron density, net charge transfer

and covalent bonds formation[?] . To complement the description of the soft-soft interactions on the target systems, we calculated the local softness[4] (Equation 6) and local multiphilicity Equation 7[?] , that combines global descriptors with the Fukui functions, distributing polarizability and electron transfer information for the atoms of the system.

$$s^{\pm}(k) = S f^{\pm}(k) \tag{6}$$

$$\Delta\omega(k) = \omega f^{\pm}(k) \tag{7}$$

The another type of interaction are the hard-hard ones, that are characterized by electrostatic forces, low-polarizability and non-covalent interactions[?] . In this case, we could use the local hardness from the CDFT as a appropriate descriptor. Although, this quantity present an ambiguous definition because the proposed derivative does not integrate to the global hardness, and there is no direct definition of local chemical potential. Besides that, this concept are used through some work equations, that are mathematical definitions that are based on the CDFT but seeking a practical way to depict hard-hard interactions without rigorous theoretical concerns.

These equations are recently being used in our's group works, and showed to be useful for biopolymers systems. In the PRIMoRDiA version used in this study there are four local hardness options implemented, as we used three to depict the charge-controlled interactions: i) The local chemical potential approximations[5], defined in Equation 8. ii) Approximation based on Electron-electron part of the molecular electrostatic potential[?] Equation 9. iii) Approximation based on the Fukui potential[6]Equation 10.

$$\eta(r) = \left(\rho(HOMO) - \frac{\rho(r)}{N} \right) \frac{\mu}{2N} + \frac{\rho(r)}{N} \eta \tag{8}$$

$$\eta^{TFD}(r) = \frac{1}{2N} \int \frac{\rho(r')}{|r - r'|} dr' \tag{9}$$

$$\eta^f(r) = \int \frac{f^-(r')}{|r - r'|} \qquad (10)$$

There is a fourth usual working equation for local hardness that is implemented in the software, where the global quantity is spread in the molecular system by the values of the Fukui functions[7]. In addition to these local reactivity descriptors from CDFT formalism, we also explored the electronic density and the frontier molecular orbitals localization. This data were first calculated by each atom of the whole protein structures and then summed for each of their belonging amino-acid residues. Such type of representations of these data are called as residue condensed in the PRIMoRDiA software, that provide it automatically for its user.

Results

QC/MM Molecular Dynamics Sampling

Reactivity Descriptors

Statistical Analysis

Structural and Reactivity Analysis

Final Considerations and Perpectives

The upcoming new version of the program, is planned to work directly with the electronic structure obtained with pDynamo software through EasyHybrid 3.0 interface, allowing such theoretical analysis on the fly for hybrid QC/MM simulations.

References

(1) Field, M. J. The pDynamo program for molecular simulations using hybrid quantum chemical and molecular mechanical potentials. *Journal of Chemical Theory and Computation* **2008**, *4*, 1151–1161.

(2) Bachega, J. F. R.; Timmers, L. F. S.; Assirati, L.; Bachega, L. R.; Field, M. J.; Wymore, T. GTKDynamo: A PyMOL plug-in for QC/MM hybrid potential simulations. 2013.

(3) Yang, W.; Parr, R. G.; Pucci, R. Electron density, Kohn–Sham frontier orbitals, and Fukui functions. *The Journal of Chemical Physics* **1984**, *81*, 2862–2863.

(4) Roy, R. K.; Krishnamurti, S.; Geerlings, P.; Pal, S. Local Softness and Hardness Based Reactivity Descriptors for Predicting Intra- and Intermolecular Reactivity Sequences: Carbonyl Compounds. *J. Phys. Chem. A* **1998**, *5639*, 3746–3755.

(5) Gál, T.; Geerlings, P.; De Proft, F.; Torrent-Sucarrat, M. A new approach to local hardness. *Phys. Chem. Chem. Phys.* **2011**, *13*, 15003–15.

(6) Cárdenas, C.; Tiznado, W.; Ayers, P. W.; Fuentealba, P. The Fukui potential and the capacity of charge and the global hardness of atoms. *The Journal of Physical Chemistry A* **2011**, *115*, 2325–2331.

(7) Meneses, L.; Tiznado, W.; Contreras, R.; Fuentealba, P. A proposal for a new local hardness as selectivity index. *Chemical physics letters* **2004**, *383*, 181–187.