PRIMoRDiA 1.0v User Guide

Igor Barden Grillo
(barden. igor@gmail. com)
github. com/igorChem

June 9, 2020

Introduction

PRIMoRDiA (**PRI**MoRDiA **M**acromolecular **R**eactivity **D**escriptors **A**ccess) is a shared memory parallel software written in C++ for post electronic structure calculations, that efficiently reads output files from most used quantum mechanics packages, storing molecular information and processing it to generate several descriptors to evaluate the global and local reactivity of molecular systems. PRIMoRDiA supports the main reactivity descriptors of the Conceptual Density Functional Theory, the most famous and used reactivity theory, which works from response variables of the electronic structure of the molecules, as also other electrostatics properties.

This guide resumes:

1. Download and installation
2. Theoretical review
3. General usage
4. Specific options
5. Information needed from the quantum mechanical packages output (how to build their input files)
6. Input format and reactivity descriptors options.
7. List of generated files and descriptors.

Contents

Aknowlodgements

1 Download and Installation

Our Github repository, that contains the executable, this userguide and the test set, can be downloaded or cloned at https://github.com/igorChem/PRIMoRDiA1.0v.git. If you download you will have to extract the zip file. To clone the repository run the following command

```
computer@user:$ git clone https://github.com/igorChem/PRIMoRDiA1.0v.git
```

The executable, file named "PRIMoRDiA_1.0v_LINUX64", is compiled for linux machines with x86_64 architecture and ready to use in such configurations. To avoid run problems and ensure maximum performance, the system must have the amount of RAM at least equal to the files provided for analysis. This is relevant when dealing with macromolecules containing five thousand atoms or more.

The binary provided requires some shared libraries to be available in your library path. The majority are very common for Linux modern distros. If you would have some issues in running the executable, please send us an e-mail.

2 Reactivity Descriptors

Reactivity Descriptors (RDs) are convenient theoretical quantities that summarize the reactive behavior of a molecular system or its atoms/different regions. The most successful and employed RD are from the Conceptual Density Functional Theory, that provided mathematical definitions of already used chemical concepts[1]. This definitions are drawn from the fundamental differential equation of the DFT Equation 1, and its Taylor serie expansions, taking the energy and electron density derivatives with respect the number of electrons, and its Maxwell relations with other functional derivatives,

assigning to it an reactive behavior for the system[2].

$$dE = \mu dN + \int \rho(r)d\nu(r)dr \qquad (1)$$

$$\mathbf{\hat{H}\Psi = E\Psi} \qquad (2)$$

The descriptors can be firstly separated by global, a number that represent all molecular system, and local quantity that is function of position. All global quantities that are outputted by PRIMoRDiA are listed in Table 1, with its definition in CDFT and how the calculation occurs for each method available, Frozen Orbital Approximation (FOA) and Finite Differences (FD), that is the second possible way to categorize the RD. The third way to classify the access of these quantities is through the representation type, which is reserved only for the local RD, that can be represented by values at each point of a three dimensional grid or by a list of values, assigning the reactivity for each atomic center in the molecular system. The former representation type is called Condensed to atoms, often shortened in the literature as only condensed. This categorization types are summarized in Figure 1.
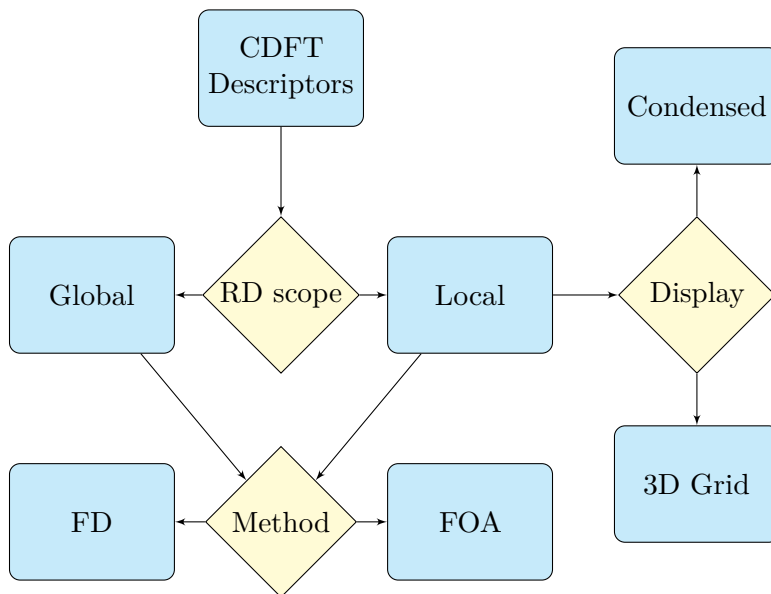
Figure 1: Classification of the reactivity descriptors.

The first CDFT RD that we can consider is the Electronic Chemical Potential (ECP), often represented by the Greek letter μ, that is the first derivative of the electrons energy in respect of the number of electrons with external potential ν constant, and represents the electron escape tendency, i.e., the system likeness to donate electrons. Pearson's Hard and Soft Acid and Base principles are

another chemical concept elevated to a more solid mathematical foundations by the CDFT development[3].

The hardness η is defined as either the second derivative of the energy with respect to the number of electrons or the derivative of ECP with respect to the number of electrons, at a constant external potential, and Softness S is defined as its reciprocal. Other global RDs can be derived to provide charge transfer quantities, which can be used to characterize chemical reactions. The maximum number of electrons ($nMax$) that the system could receive from an ideal donor and the stabilization energy when these electrons are received, which is called the total electrophilicity (ω), are examples of such global RDs Table 1.

There are two current calculation methods for the CDFT RD which are supported by PRIMoRDiA. The first is the Finite Differences, that for the global RDs takes the electronic energy of different charged states to estimate the derivative values. This is necessary because the fact of the discontinuity in the number of electrons, i.e., there is no fractional number of electrons, at least in the conventional quantum mechanics treatments. The electronic energy assessment is done for the original charged states of the molecular system of interest, one negative charged (minimum one more electron), and one positive charged (minimum one electron less). All these three state calculations are performed for the same nuclei coordinates, conserving the requirement of constant external potential of the partial derivative.

The other calculation method is defined according to Koopmans Theorem, which states that energies of the charge states can be approximated using Ionization Potential (IP) and Electron Affinity (EA). These quantities can be estimated using the energies of the Highest energy Occupied Molecular Orbital (HOMO) for IP and the Lowest energy Unoccupied Molecular Orbital (LUMO) for EA. In contrast with finite differences, the approximation from Koopmans theorem requires just a single point calculation to obtain the global RDs. In this approximation the relaxation energetic effect for the internal molecular orbital are not accounted because is considered negligible and thus called the Frozen Orbital Approximation.

Table 1: Global Reactivity Descriptors calculated by the by PRIMoRDiA software

Global RD	CDFT Definition	FOA	FD	Ref.
IP	-	$-E_{HOMO}$	$E_{N-1} - E_N$	
EA	-	$-E_{LUMO}$	$E_N - E_{N+1}$	
μ	$\left(\frac{\partial E}{\partial N}\right)_\nu$	$\frac{E_{HOMO}+E_{LUMO}}{2}$	$\frac{E_{N-1}-E_{N+1}}{2}$	[4]
η	$\left(\frac{\partial^2 E}{\partial N^2}\right)_\nu$	$E_{LUMO} - E_{HOMO}$	$E_{N-1} + E_{N+1} - 2E_N$	[5]
Softness(S)	$1/\eta$	-	-	[5]
ω	$\frac{\mu^2 S}{2}$	-	-	[6]
$nMax$	$-\frac{\omega}{\eta}$	-	-	[6]

Depending on the representation type, local RD are assigned to points in the three dimensional space, called voxels (three-dimensional analogous of a pixel). This representations type can be accessed in graphical packages requiring surfaces with that hold the same values (iso-surfaces), or rendering several of its iso-surfaces, which is called volume. Commonly, a set of voxels forming a three dimensional grid covering all space where the atoms from molecular system are enclosed is written to a cube file. PRIMoRDiA generates this cube files for each local RD supported if was required in grid arguments a integer greater than zero. This files are formatted as standard GAUSSIAN cube files, where the main graphical packages can open. This local RD represented as scalar fields are functions of three-dimensional coordinate r.

The Fukui indices are the first local RD in CDFT and its calculations are required for the great majority of others local RD. Which is also called the Fukui function, the derivative of electron density in respect to the number of electrons (Equation 3), that also is extracted from the Equation 1[7]. The regions in the molecular system where the Fukui function values are at maximum is indicated where are the great propensity to the electron density to increase as the number of electrons vary.

$$f(r) = \left(\frac{\partial \rho(r)}{\partial N}\right)_\nu \qquad (3)$$

The Fukui function characterizes intramolecular reactivity, which makes it useful for identifying the selectivity of a given electrophilic, nucleophilic or radical attack, each of which are represented by a given Fukui function. Fukui functions were initially developed to be calculated using FD in relation to the electron densities of charged states. With these descriptors, it is possible to say how the electron density will change when another system approaches and thus promote a change in the number of electrons. In other words, the system part where the descriptor is at its maximum is where the electron density will concentrate to form a new bond.

Due to the discontinuity of the electrons, the FD calculus method was the first used to estimate the derivative, set in Equation 3. The chosen method caused to the derivative to be defined in two ways, one at the right of the $\lim_{N \to 0}$ and to the left. The electrophilic attack susceptibility (EAS), which is the left Fukui function (f^-), reveals the regions where an electrophilic species is most likely to form a bond. The same explanation applies for nucleophilic attack susceptibility (NAS), right Fukui function (f^-), and for radical attack susceptibility (RAS), zero point Fukui function (f^0). The FD uses the electrons densities of different charged states to calculate the Fukui function, for EAS the difference between electron density $\rho(r)$ obtained from the molecular system with original charge (N electrons) and the electron density from the molecular system with decreased number of electrons (N-1, at least)Equation 4.

$$f^-(r) = \rho(r)_N - \rho(r)_{N-1} \tag{4}$$

The electron density is calculated from the sum of i occupied molecular orbitals $\phi(r)$ square module (Equation 5), and the molecular orbitals are obtained from the linear combination of the atomic orbitals χ Equation 6. The atomic orbitals can be a cartesian slater basis set (Equation 7), which is for instance the standard for the semiempirical methods in MOPAC.

$$\rho(r) = \sum_i^{occ} |\phi(r)_i^2| \tag{5}$$

$$\phi(r)_i = \sum_i c_i \chi(r) \tag{6}$$

$$\chi(r)_{kmn} = N x^k y^m z^n \exp(-\zeta r) \tag{7}$$

For NAS, the right Fukui function $f^+(r)$, the FD method defines the derivative as the difference of electron density of the molecular system with an increased number of electrons (N+1, at least) and the original charged state (N) electron density. Often the difference in charge is one to approximate at maximum to the limit, but this can change when the system needs a greater perturbation to show the more reactivity regions through these functions. For the FD option in PRIMoRDiA is required to indicate the charge difference used to calculated the other charged states to normalize the resulted Fukui functions to one.

$$f^+(r) = \rho(r)_{N+1} - \rho(r)_N \tag{8}$$

The RAS. or zero point Fukui function, is defined as the arithmetic average of the EAS and NAS (Equation 9), which brings the possible interpretation of this descriptor to be a averaged Fukui function.

$$f^0(r) = \frac{\rho(r)_{N+1} - \rho(r)_{N-1}}{2} \qquad (9)$$

A fourth Fukui function can be considered, the Dual Fukui or Dual Descriptor, that can indicate simultaneous the electrophilic and nucleophilc attack susceptibility regions. This descriptor is calculated as the difference between NAS and EAS (Equation 10), also mitigating the Fukui function values in regions that present both behaviors, which make it to be a net Fukui function as well. Also, is claimed that this RD carries less errors because the possible cancellation in the difference between left and right Fukui functions ([8]).

$$\Delta f^{\pm} = f^+(r) - f^-(r) \qquad (10)$$

A FOA definition also exists for local RD that are derived from Fukui functions, based on the fact that a infinitesimal change in the number of electrons the inner molecular orbitals would have a irrelevant contribution to the electron density and thus for the Fukui function itself. In EAS case the left Fukui function is defined in ways of M one-electron molecular orbitals (Equation 11), that is the electron density derivative in respect to the number od electrons from the limit of δ decreasing of electrons tending to zero. The equations is divided in one part for the contribution of the HOMO and other, that are neglected in the FOA, for the inner molecular orbitals. The EAS for FOA turns to be the density of only the HOMO (Equation 12). The same occurs for the NAS, but the LUMO (M+1) is included, and the HOMO join the sum of the inner orbitals that are neglected (Equation 13), and NAS finally defines as the density of LUMO (Equation 14). For RAS, the definition often follows the ones that were cited earlier in the text, being the average of EAS and NAS, but also can be thought as the density of the Single Occupied Molecular Orbital (SOMO) (Equation 15).

$$f^-(r) = \lim_{\delta \to 0} \frac{\partial \rho_{M+\delta}(r)}{\partial N} = |\phi_M(r)|^2 + \sum_{i=1}^{M-1} \frac{\partial}{\partial N}|\phi_i(r)|^2 \qquad (11)$$

$$f^-(r) = |\phi_{HOMO}(r)|^2 \qquad (12)$$

$$f^+(r) = \lim_{\delta \to 0^+} \frac{\partial \rho_{M+\delta}(r)}{\partial N} = |\phi_{M+1}(r)|^2 + \sum_{i=1}^{M} \frac{\partial}{\partial N}|\phi_i(r)|^2 \qquad (13)$$

$$f^+(r) = |\phi_{LUMO}(r)|^2 \qquad (14)$$

$$f^0(r) = |\phi_{SOMO}(r)|^2 \qquad (15)$$

The other representation type used to these local RD is the condensed, as already cited, where the reactivity values are assigned to each atomic center in the molecule. The Equation 11 shows the generic condensation method for a Fukui function, where its values are integrated to each point r in space that is considered to the $k_t h$ atom enclose.

$$f_k^a = \int_{\Omega k} f(r)_k^a dr \qquad (16)$$

In the FD case, the condensation is done around the partial gross charge of each atom, obtained originally from the Mulliken's Population analysis. EAS is defined for the K_{th} atom as the difference of the partial charge in neutral and in the positive charged system (Equation 17), and NAS as the difference from the partial charge of negatively charged and the state with original charge (Equation 18).

$$f_k^- = q_{k[N]} - q_{k[N-1]} \qquad (17)$$

$$f_k^+ = q_{k[N+1]} - q_{k[N]} \qquad (18)$$

For FOA, the condensation is done using calculating the square of the molecular orbital i correspondent to the required by the descriptor, as in Equation 19 showing the module square of molecular orbital ψ i for the k_{th} atom, as the sum of the square of molecular orbital linear combination coefficients corresponding to the atomic orbitals (AO) of k_{th} atom, plus the sum of product between these coefficients among them and its corresponding element in the overlap matrix S, which represent the product of the one-electron atomic orbitals of the basis set.

$$|\psi_{\mathbf{k}}^{\mathbf{i}}|^{\mathbf{2}} = \sum_{\nu \in \mathbf{k}}^{\mathbf{AO}} |\mathbf{C}_{\nu \mathbf{i}}|^{\mathbf{2}} + \sum_{\mu \notin \nu}^{\mathbf{AO}} |\mathbf{C}_{\nu \mathbf{i}} \mathbf{C}_{\mu \mathbf{i}}| \mathbf{S}_{\mu \nu} \qquad (19)$$

$$\rho(\mathbf{r})_{\mathbf{k}} = \sum_{\mathbf{i}}^{\mathbf{occ}} |\psi_{\mathbf{k}}^{\mathbf{i}}|^{\mathbf{2}} \qquad (20)$$

Thus, EAS is calculated for each atom k as in Equation 21, using the square module of HOMO, and NAS the LUMO Equation 22.

$$\mathbf{f_k^-} = |\psi_k^{\mathbf{HOMO}}|^2 = \sum_{\nu \in \mathbf{k}}^{\mathbf{AO}} |\mathbf{C_{\nu HOMO}}|^2 + \sum_{\mu \notin \nu}^{\mathbf{AO}} |\mathbf{C_{\nu HOMO} C_{\mu HOMO}}| \mathbf{S_{\mu\nu}}$$
(21)

$$\mathbf{f_k^+} = |\psi_k^{\mathbf{LUMO}}|^2 = \sum_{\nu \in \mathbf{k}}^{\mathbf{AO}} |\mathbf{C_{\nu LUMO}}|^2 + \sum_{\mu \notin \nu}^{\mathbf{AO}} |\mathbf{C_{\nu LUMO} C_{\mu LUMO}}| \mathbf{S_{\mu\nu}}$$
(22)

We defined the Fukui functions for FD, FOA calculations methods, represented by a grid, volumetric, and condensed to atom. These calculations procedures can be extended for the RDs, with all of them are defined in combinations of FD, FOA, volumetric and condensed representation.

The global RD softness can be defined locally replacing the electron number in the original derivative for the electron density, which is in concept the local representation of electrons. Thus local softness, $s(r)$ is defined in the Equation 23, that shows the Maxwell relations that leads to the definition where the local softness is the product of its global counterpart and the Fukui function.

$$s(r) = \left(\frac{\partial \rho(r)}{\partial \mu}\right)_\nu = \left(\frac{\partial N}{\partial \mu}\right)_\nu \left(\frac{\partial \rho(r)}{\partial N}\right)_\nu = Sf(r)$$
(23)

As the Fukui function is divided and the other local RD, the local softness also have the same calculation methods, as in the Equation 24 for electrophilic attack susceptibility, Equation 25 for nucleophilic attack susceptibility, and Equation 26 both cases simultaneously. The dual local softness $S^\pm(r)$ is calculated in PRIMoRDiA.

$$s^-(r) = Sf^-(r)$$
(24)

$$s^-(r) = Sf^-(r)$$
(25)

$$s^\pm(r) = Sf^\pm(r)$$
(26)

Also, other local RD that have being often used nowadays is the Hyper softness, defined in Equation 27, and are also calculated by PRIMoRDiA.

$$s^2(r) = S2f^0(r)$$
(27)

For the same reasons of the definitions of local softness, the local version of total electrophilicity calculated as a product of its global counterpart and the Fukui functions[9]. PRIMoRDiA can calculate the local electrophilicity $\omega^+(r)$, which is the stabilization energy when the local r receives electrons. The "left" electrophilicity (Equation 29) is less often used because the concept of total electrophilicity brings with the stabilization energy when the system receives electrons from a ideal donor, and thus makes less sense to use this local quantity to measure nucleophilicity. Although, the Dual Fukui function combined with the total electrophilicity is used as a multiphilic local descriptor Equation 30, and is obtained with PRIMoRDiA as well[10].

$$\omega^-(r) = \omega f^-(r) \tag{28}$$

$$\omega^+(r) = \omega^+ f(r) \tag{29}$$

$$\Delta\omega(r) = \omega f^\pm(r) \tag{30}$$

The global hardness is a important descriptor to study interaction of non-covalent nature, and thus a local version is expected to do the same to aiding the selectivity studies. Although, in contrast with the local softness, there is no local version to the quantity in the numerator of its derivative Equation 31. The mathematical development of CDFT leaded to several solutions of methods to approximate the local Hardness. Thus, the next section will bring the four implemented methods considered relevant and practical to be interpreted.

$$\eta(r) = \left(\frac{\partial\mu}{\partial\rho(r)}\right)_\nu \tag{31}$$

2.1 Local Hardness Methods and Electrostatic Properties

Through the reactivity theories, Klopman and Fukui summarized that all chemical reactions are either orbital-controlled or charged-controlled. Pearson and the other authors of CDFT defined the orbital-controlled reactions as result of soft-soft interactions, where the reaction is favored by a electron delocalization followed by charge transfer and intense frontier molecular orbital overlapping. The charged-controlled reactions would be result of hard-hard interactions, with the inverse behavior of the soft-soft, approximation and bonding governed by Coulombic Forces, more typical of long range, electrostatic, weak interactions or species with high charge density.

Thus Hardness, and its local counterpart are a very important quantity to the rationalization of all chemical processes with charge-controlled character, and also are intimately related with molecular electrostatic properties. PRIMoRDiA provides four methods to estimate local hardness that will be discussed in this section, Molecular Electrostatic Potential (MEP) grids based on partial charges, and production of formatted pdb files with electron density and partial charges for macromolecules to promote the visualization of these properties for biological relevant structures.

The local hardness is a CDFT quantity that presents ambiguity in its definition because it integration do not result in the global hardness, as for the local softness. The global counterpart of local hardness is only obtained by the integration of its product with the Fukui function (Equation 32), due to the identity showed by Equation 33.

$$\eta = \int \eta(r)f(r) \tag{32}$$

$$\partial\rho(r) = f(r)dN \tag{33}$$

Although, a search of local hardness working equations continued despite that, seeking a valid mathematical definition in CDFT that result in calculable reactivity descriptor to predict preferential sites for charge-controlled reactions. From the mathematical DFT definitions, the local hardness derivative is written as proportional to the integral in all space of second derivative of total electronic functional $F[\rho]$ with the respect of electron density $\rho(r)$ and $\rho(r')$ (Equation 34). The total electronic functional is part of the total energy functional $E[\rho]$, being dependent only of electrons and independent of external potential (Equation 35). Then, $F[\rho]$ is divided in three other functionals: kinetic electron energy $T[\rho]$, classical Coulomb interaction $J[\rho]$ and exchange-correlation $E_{xc}[\rho]$ (Equation 36).

$$\eta(r) = \left(\frac{\partial\mu}{\partial\rho(r)}\right)_\nu = \frac{1}{2N}\int \frac{\delta^2 F[\rho]}{\delta\rho(r)\delta\rho(r')}\rho(r')dr' \tag{34}$$

$$E[\rho] = \int \rho(r)d\nu(r)dr + F[\rho] \tag{35}$$

$$F[\rho] = T[\rho] + J[\rho] + E_{xc}[\rho] \tag{36}$$

Thus a total local hardness is defined as a sum of the quantity for each of these functionals that compose the total electronic energy functional Equation 37. In the Thomas-Fermi-Dirac (TFD) model

for these functionals, the local hardness of kinetic and exchange-correlation energy functionals have a abrupt fall-off of its values for the molecules outer regions, where lies the interest of predict the hard-hard interactions preferential sites in molecules. Thus the relevant form of the local hardness can be approximate only by the $\eta_J(r)$, defined in Equation 38, which replaces the $F[\rho]$ by $J[\rho]$, that is defined in Equation 40.

$$\eta_F(r) = \eta_T(r) + \eta_J(r) - \eta_K(r) \tag{37}$$

$$\eta^{TFD}(r) = \frac{1}{2N} \int \frac{\delta^2 J[\rho]}{\delta\rho(r)\delta\rho(r')} \rho(r')dr' = \frac{1}{2N} \int \frac{1}{|r - r'|} \rho(r')dr' \tag{38}$$

$$\eta^{\mathbf{TFD}}(\mathbf{r}) = \frac{\mathbf{1}}{\mathbf{2N}} \int \frac{\mathbf{1}}{|\mathbf{r} - \mathbf{r'}|} \rho(\mathbf{r'})\mathbf{dr'} \tag{39}$$

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr' \tag{40}$$

Thus its defined the first method of local hardness calculated by PRIMoRDiA, and its condensed to atom counterpart defining this quantity for each k_{th} atom in the system defined in Equation 41. The calculations are performed as a sum of the electron density for each other l_{th} atom pondered by its distance R_{kl} to atom k.

$$H_k = \sum_{l \neq k}^{atoms} \frac{\rho(l)}{R_{kl}} \tag{41}$$

The second method approximates the electron density by the left Fukui function, which is stated to represent the regions where the electrons density has more propensity to have electrostatic interactions Equation 42. This type of method is highly useful to speed up the calculations when using FOA, which obtains the Fukui function from only one QM single point and only calculates in function of the frontier molecular orbitals instead of calculate the entire electron density. Also, PRIMoRDiA provides a condensed to atom version of this second method, which also only replaces the electron density of atom l by its left Fukui function value Equation 43. The condensed versions for these two methods are much faster to calculate than for the grids.

$$H(r) = \int \frac{f^-(r')}{|r - r'|} \tag{42}$$

$$H_k = \sum_{l \neq k}^{atoms} \frac{f^-(l)}{R_{kl}} \tag{43}$$

The main source of the ambiguity for local hardness CDFT definition is that there is no point-wise version of chemical potential. The third local method explained in this section, called in PRIMoRDiA by the *"LCP"* (Local Chemical Potential) keyword, defines the local hardness working equation based on a mathematical expression for local chemical potential. This was the first local hardness method implemented in the program because integrates to the global hardness and has local inversion relations with Fukui function, which holds the same type of information of local softness, thus extending the Pearson HSAB theory from global to local. Through mathematical properties of functionals that can be normalized, the electronic energy derivative with respect to the electron density can be written as in Equation 44

$$\frac{\delta E_\nu[\rho]}{\delta \rho(r)} = \frac{1}{N} \int \rho(r') \frac{\delta E_\nu[\rho]}{\delta \rho(r')} dr' \tag{44}$$

The electronic chemical potential μ can be recognized in the last equation, as show in Equation 45, and thus can be assumed a local version skipping the integration and becoming Equation 46. Using this definition for local chemical potential a working equation for local hardness results in Equation 47.

$$\mu = \frac{1}{N} \int \rho(r) \frac{\delta E_\nu[\rho]}{\delta \rho(r)} dr \tag{45}$$

$$\mu(r) = \frac{\rho(r)}{N} \frac{\delta E_\nu[\rho]}{\delta \rho(r)} dr' = \frac{\rho(r)}{N} \mu \tag{46}$$

$$\eta(r) = \left(\rho(HOMO) - \frac{\rho(r)}{N} \right) \frac{\mu}{2N} + \frac{\rho(r)}{N} \eta \tag{47}$$

The last method of local hardness that is calculated by PRIMoRDiA is based on the distribution of its global counterpart, written in terms of Ionization Potential and Electron Affinity, in function of the left and right Fukui functions[11]. The definition proposed in Equation 48, comes from a more chemical intuition nature, considering for instance that a hard electrophilic species would be one with high IP, i.e., with very low tendency to donate electrons. Possibly being positively charged and seeking other electron rich species regions to bond driven by electrostatic interaction instead

of electron density delocalization, which would be a softness type of interaction.

$$\eta(r) = IPf^-(r) - EAf^+(r) \tag{48}$$

A hard nucleophilic would have a large ionization potential and small electron affinity, and thus would interact with the hard electrophilic species type through theirs hardest regions. This methods is the most simple to calculate and thus the fastest, needing only the evaluation of Fukui function and the global RDs. A other great advantage of this definition is that its integration gives the global hardness. Although, as this method results to be proportional to the Fukui function the described reactivity happens to be redundant.

In Table 2 is summarized the local hardness methods supported by PRIMoRDiA, its keywords, definitions and references. For all the four definition there is a condensed to atom working equation based on the electrons density for the atom or condensed Fukui function.

Table 2: Local Hardness methods supported by PRIMoRDiA

Keyword	CDFT Definition	ref
mepEE	$\frac{1}{2N} \int \frac{\rho(r')}{\|r-r'\|}$	[12]
mepFukui	$\int \frac{f^-(r')}{\|r-r'\|}$	[13]
LCP	$\left(\rho(HOMO) - \frac{\rho(r)}{N} \right) \frac{\mu}{2N} + \frac{\rho(r)}{N}\eta$	[14]
fukui	$IPf^-(r) - EAf^+(r)$	[11]

2.2 Band and Residues Reactivity Descriptors

Macromolecules relevant for biological processes, such as Protein and DNA-like structures, present a larger number of degenerated states in energy levels near HOMO-LUMO, which hinders the complete reactivity description of these systems when the FOA is used due to the contribution neglecting of other relevant molecular orbitals[15]. Also, for FD method, the minimum charge perturbation in the Fukui function calculation often is not sufficient to provide a measurable response of electron density. In the last case, greater charge perturbations are applied in the QM calculations, as proposed by Khandogin and coworkers [16], to correct this problem. Although, this create more difficulties in these large structure calculations convergence, that are often already charged.

For small molecules with significant frontier molecular orbital degeneracy some solutions were already proposed in literature for Fukui function calculation using FOA. The Frontier Effective Reactivity Molecular Orbital (FERMO), propose a earlier visualization of the molecular orbitals to chose those with the best shape to match the reaction that are occurring [17]. This is clearly not a reasonable

approach for large systems. Some studies indicate to make a average mean of Fukui function by the degenerated states [14, 18], and Pino-Rios and coworkers [19] establish a energy weighted left Fukui function.

We extended this two latter concepts for all Fukui functions, and thus for all reactivity descriptors that are dependent, and implemented in PRIMoRDiA with energy and band size control by the user. In the run mode 3, more explained in section 4 of this userguide, there are two special implementations of the local RDs, the Band Density (BD) Fukui functions and the Energy Weighted (EW), where for the former, the left Fukui function is defined in Equation 49, and the right in Equation 51. In the case of EW version, the definition for the left Fukui function in Equation 50 as well for the right Fukui function, showed in Equation 52.

$$f_{BD}^- = \sum_{i=b}^{HOMO} |\psi_i|^2 \tag{49}$$

$$\mathbf{f_{EW}^-} = \sum_{\mathbf{i=b}}^{\mathbf{HOMO}} \mathbf{e}^{-|\mathbf{E_i} - \mathbf{E_{HOMO}}|} |\psi_\mathbf{i}|^\mathbf{2} \tag{50}$$

$$\mathbf{f_{BD}^+} = \sum_{\mathbf{i=LUMO}}^{\mathbf{b}} |\psi_\mathbf{i}|^\mathbf{2} \tag{51}$$

$$\mathbf{f_{EW}^+} = \sum_{\mathbf{i=LUMO}}^{\mathbf{b}} \mathbf{e}^{-|\mathbf{E_i} - \mathbf{E_{LUMO}}|} |\psi_\mathbf{i}|^\mathbf{2} \tag{52}$$

The b parameter specifies the band size, i.e. the number of molecular orbitals near HOMO-LUMO to be used in the calculations. Energy control barriers also are enabled by default allowing only the accounting of molecular orbitals within energy difference of 1eV from HOMO or LUMO. This also can easly be modified in input files before running PRIMoRDiA. The visualization of your macromolecule density of states is encouraged to have a better control parameters guess for the Band RDs, and can be acomplished with scripts written by PRIMoRDiA to plot graphs using R statistical package. This descriptors are also summarized by the residues, where the condensed values for each atom of the residues are summed and written in a separated file.

We have already applied this two types of descriptor modification, each one in one published work. The BD approach was employes in the interactions calculations of the HIV-1 protease enzyme with sixteen inhibitors with experimental data [20]. The EW was successfully applied in the theoretical characterization of three enzymatic reaction mechanism obtained with hybrid quantum mechanical / molecular mechanics calculations [21].

3 Programs Supported and Information needed

PRIMoRDiA can parse and work with information from output files of the most common quantum mechanical calculation packages. In the present version, such software are GAMESS, GAUSSIAN, MOPAC and ORCA. The main options in the input files to run QM calculations to compute the maximum of properties provided by PRIMoRDiA for each of these packages will be discussed here. s

In the following text box (Listing 1) is presented a GAMESS input example for a acrolein molecule to generate in log file the information needed to run all options. The most important modification in standard input option for GAMESS run is the nprint option set to 3 (NPRINT=3) in the $contlr group. This option will guarantee that one-electron atomic integrals to be printed to calculate the condensed to atoms local RD and also the other needed information that also are printed if the option 7 would to be selected, which already is the default.

Listing 1: Gamess Input Example

```
 $SYSTEM MWORDS=20 $END
 $CONTRL RUNTYP=Energy $END
 $CONTRL SCFTYP=RHF $END
 $CONTRL ICHARG=0  MULT=1 NPRINT=3 $END
 $BASIS GBASIS=N311 NGAUSS=6 $END
 $BASIS NDFUNC=1 NFFUNC=1 $END
 $BASIS DIFFSP=.TRUE. $END
 $SCF DIRSCF=.TRUE. $END
 $DATA
Molecule specification
C1
C 6.000000 0.811276 0.464703 0.000085
C 6.000000 -0.666979 0.602652 -0.000083
C 6.000000 -1.468478 -0.468578 0.000045
H 1.000000 -1.067523 1.614725 -0.000234
H 1.000000 -2.550922 -0.381565 -0.000017
H 1.000000 -1.040888 -1.467756 0.000243
H 1.000000 1.374133 1.423218 0.000372
O 8.000000 1.403786 -0.597661 -0.000081
 $END
```

For ORCA software (at least after the 3.x version) enclosed in the %output group the options: print[p_mos] 1 and print[p_overlap] 5 (Listing 2). This will allow the printing of the atomic basis set information and atomic one-electrons integrals.

Listing 2: ORCA Input Example

```
!  HF
! PrintBasis 6-311+G(2d,2p)
%output
print[p_mos] 1
print[p_overlap] 5
end #output
* xyz 0    1
C   0.811276 0.464703 0.000085
C  -0.666979 0.602652 -0.000083
C  -1.468478 -0.468578 0.000045
H  -1.067523 1.614725 -0.000234
H  -2.550922 -0.381565 -0.000017
H  -1.040888 -1.467756 0.000243
H   1.374133 1.423218 0.000372
O   1.403786 -0.597661 -0.000081
*
```

In MOPAC is necessary to use the aux keyword to generate the .aux where are all relevant information to the PRIMoRDiA calculations. Also the "allvecs" and "large" keywords, to print the full molecular orbitals coefficients. In addition, for the small molecules (up to 300 atoms), the "graphf" option is required for finite different calculations and total electron density cube generation for mopac output.

4   Program Execution and General Options

After copying the binary files to your system, you can use the program as follows:

```
computer@user:$ /path/to/PRIMoRDiA/primordia -f input -np 4 -log -verbose
```

This example above is the command to run the reactivity descriptors calculations, where the files and options are given in the input file. To run this mode you need to put after binary path the *"-f"* flag followed by the name of input file. The other options can be set in any order after the mandatory and positional ones.

By default, PRIMoRDiA uses the maximum number of threads in your CPU node, but it can be set as you please using the *"-np"* flag followed an integer indicating the maximum number of threads to be used, as explicit in the example above. If you want to generate a log file with PRIMoRDiA run details, possible warning or errors causes, run with *"-log"* after the mandatory flags and options. To print these log information to the console screen you have to use the *"-verbose"* flag.

In the Table 3 present other run options in its required arguments,

the *"-help"* can be used to print these resume information to a rapid access in the console screen, as presented in the next example.

```
computer@user:$ /path/to/PRIMoRDiA/primordia --help
```

```
Starting PRIMoRDiA software!
Calculations starting at: Thu Jun  4 10:43:56 2020

PRIMoRDiA help page
PRIMoRDiA Macromolecular Reactivity Descriptor Access
The program must be run as follows:
/path/to/executable [option run] [file\_name] [other options]
options run:
--help/-h : Display this help message
-f    : Reactivity descriptors run option
-ed   : Electron density cube file generation run option
-mo   : Molecular Orbital cube file generation run option
-input: Produce input from the name list in the current folder
-cubed: cube file differences and similarity index calculation
-cdiff: Calculates the similarity index from a list of cube files
Generic options is the options must be placed after all the other arguments
Generic options:
-np [n] : program runs using n threads
-log    : program produces a log file of its operations
-verbose: program prints to the console messages about its operations

Total execution time of PRIMoRDiA program: 0.0136223 seconds
```

Table 3: PRIMoRDiA run options ans its arguments

Mode	Flag	Argument 1	Argument 2	Argument 3	Argument 4
Reactivity descriptors	-f	input file	-	-	-
Electron density	-ed	QM output	grid size	QM program	-
Molecular orbital	-mo	QM output	MO number	grid size	QM program
Pair Cube differences	-cubed	cube file 1	cube file 2	-	-
Cube differences	-cdiffs	input	-	-	-
Cube integral	-int	cube file			

The *"-ed"* flag tells the program to process the following argument as the QM output and generate from its information the total electrons density cube file for that molecular system, with a given grid size and the last argument signalizing the QM program name. The same is true for generate the molecular orbital cube files, with an extra argument for the molecular orbital number. Two identical cube files will be generated for molecular orbitals to load each phase in Pymol.

The integral of the difference between to cube files can be required through the *"-cubed"* flag, providing the cube file names after that.

The cubes must have the same dimensions and grid size. The value of Carbo index is also outputted in these calculations, indicating the similarity index between the grids from zero to one. Several cubes differences can be calculated at once using the *"-cdiffs"* flags and providing the name of an input file, that must contain in the first line the number of comparisons and each other line the pair of cube names. This run mode will output a file with the ".reportdiffs" extension with the similarity index values. Finally, the integral of a cube file can be calculated using *"-int"* flag.

5   Input options

In this section will be showed the option and requirements to build input and run Reactivity Descriptor calculations in PRIMoRDiA. The Listing 3 shows an input example to run using the *"-f"* flag, where in the first line are set general options applied to all the run, and in each one of the other lines is for set the structures to be calculated. The first number in each of these lines must be among 1 and 3, signalizing the general type of the run: 1 is to run a generic structure using Frozen Orbital Approximation; 2 is to run using Finite Differences approximation; 3 is specific to run macromolecules using the special reactivity descriptors using the frontier molecular orbitals. As presented in Listing 3, the three modes can be required in the same input file, in any order you wish. In the case of the first character in the line do not be 1, 2 or 3, PRIMoRDiA will ignore that line.

Listing 3: Example of input for reactivity descriptor calculations.

```
dos extrard eband 3 pymols
1 mol.aux 40 mopac mep
2 mol.log cation.log anion.log 40 1 gamess
3 trp.aux n 40 10 trp.pdb 28.95  21.86  27.90 10
```

The run control parameters set in the first line turn on the outputting of extra files and RD and in the case of *"eband"* option modify the energy criteria for the band reactivity descriptors, that must be given after the keyword. If none of these options are required by the user, some word must be written in the line. The *"dos"* option output file and script written in R language to generate a plot of the molecular density of states. The *"extrard"* option requires that the cube files of all possible local RD to be outputted. As default, only the Fukui functions and the local hardness cube files are generated, if some local hardness method was required. Also, for all run modes, if the grid size required was not greater than zero the cube files are not written in the output. The *pymols* generate script input to load grids and pdbs withe the condensed RD in PYMOL, setting colors and iso-surfaces to simplify the reactivity maps generation with this graphical package.

Each run mode have theirs requirements, positional and optional arguments. In the next subsections these details will be covered. The possible methods of local hardness and names od programs are the same for all run mode options.

The possible keywords for the local hardness methods are:

- LCP

- mepEE

- mepFukui

- fukui

- any word (to avoid the calculation)

The programs keyword accept are:

- mopac

- gamess

- orca

- gaussian

5.1 Option 1: Generic FOA

The run mode option 1 calculates the reactivity descriptors using Frozen Orbital Approximation requiring the QM output file for a single point calculation for the given molecular structure. In Listing 4 is show an input file example for a run only using option 1. In the first line the *"extrard"* signalizes the outputting of extra local RD cube files, as explained earlier. In second line, the first character indicates the run mode, followed by the name of QM output file, the third positional argument is the local hardness method, 40 in the case, fourth the grid size and the fifth last mandatory argument is the program name, in the case *"gaussian"*. The *"mep"* option must be placed after the mandatory arguments if wanted PRIMoRDiA to generate cube files with the molecular electrostatic potential approximated by the atomic partial charges.

Listing 4: Example of input for reactivity descriptor calculations using Frozen Orbital Approximations.

```
extrard
1 acrolein.fchk potential 40  gaussian mep
1 acrolein.aux potential_fukui 60  mopac mep
1 acrolein.log default  40  gamess  mep
1 acrolein.out default 0  orca mep
```

In all cases a file with ".lrd" extension will be generated with all reactivity descriptors possible to be calculated for each atom in the

molecule. Grid size number can be set to zero, as in the fifth line, to generate only condensed to atoms results. The possible local hardness and program keywords were cited in the earlier section. For this options is required for GAUSSIAN outputs path of ".fchk" file, and in the folder where is this file the ".log" with the overlap matrix also is needed. For MOPAC, the ".aux" is required, GAMESS the ".log" and ORCA ".out".

5.2 Option 2: Generic FD

The second option in input using *"-f"* flag, is for the reactivity calculations using Finite difference method for generic structures. The differences in input for mode 2 in comparison with 1 are the addition of file names of positive charged state, as third argument, negatively charged state as the fourth argument, local hardness method name as fifth, grid size as sixth and the net charge difference between the sates as seventh position in the line. For this option the mopac output required is that with ".mgf" extension, ".aux" and ".out" also were required to be in the same folder.

Listing 5: Example of input for reactivity descriptor calculations using Finite Difference Approximations.

```
extrard
2 mol.mgf cation.mgf anion.mgf n        60 1 mopac mep
2 mol.log cation.log anion.log default 40 1 gamess mep
```

5.3 Option 3: Macromolecules

The option 3 is for calculate the special frontier molecular orbital Band RD, built for macromolecules, requiring protein data bank files and other specific parameters. This options has the valid keywords to be placed in the first lines: *"dos"*, *"pymols"* and *"eband"*, which already were explained in beginning of the section four. The Listing 6 show a input file example for option 3 possible uses and required arguments.

Listing 6: Example of input for reactivity descriptor calculations using Frozen Orbital Approximations for Macromolecules.

```
extrard dos eband 3.4 pymols
3 trp.aux n 40 10 trp.pdb mopac 28.95  21.86  27.90 10 EW
3 trp.log potential_fukui mopac 0 30 trp.pdb 0 0 0 0 elecdens
```

After the run mode number, the name of QM output file name is required as the second argument, followed by the local hardness methods, grid size, maximum number of orbitals to be used (if set to 0 all can used), pdb file name, and program name. The eight mandatory argument are the x-axis coordinate in Angstrom for the box center where the grid calculations will be placed, followed by

the y-axis, z-axis and the size of the side of this box as the eleventh argument. If the box side size is set to zero, in the case where the grid size is greater than zero, the box will be selected by the program enclosing all the macromolecules atoms. This can have to effects in the calculations, if the grid size is set to a small number the rendering in graphical packages may be very poor in definition of the shapes. If the chosen value is big, the calculations can be very costly, specially for the local hardness methods.

The Band Density RD method is the default method for each structure that follows option 3 in the input and can be changed by putting the "EW" word after the positional arguments. Other possible argument is "elecdens" which enables the condensed electron density calculation.

6 Output files

The Table 4 lists all file types that can be generated by PRIMoRDiA.

Table 4: File types generated by PRIMoRDiA

Extension	Description	Observations
.cube	Gaussian formatted files for scalar field data	Generated when grid calculations are required
.log	File with PRIMoRDiA internal information, warning messages and possible error details	Generated when *"-log"* flag is used
.global	File with global RDs for each structure in input file	-
.lrd	File with condensed local RDs	-
residues.lrd	File with total local RDs for each residue in the macromolecule	Only for run option 3
.pdb	pdb file of the macromolecule rewritten with the condensed local RD in the b-factor column	Only for run option 3
.R	Script for density of states graph generation by R software	Generated when dos opton is set in the input first line
.DOS	File with energies by molecular orbital	Generated when dos opton is set in the input first line
.pymol	Script for running inside of Pymol to automate reactivity maps generation	Generated when "pymols" opton is set in the input first line
.reportdiffs	File with Carbo index for sclaar field pair comparison	Generated when the flag "-cdiffs" is used

References

[1] P. Geerlings, F. De Proft, and W. Langenaeker, "Conceptual density functional theory," *Chemical reviews*, vol. 103, no. 5, pp. 1793–1874, 2003.

[2] R. G. Parr, R. A. Donnelly, M. Levy, and W. E. Palke, "Electronegativity: the density functional viewpoint," *The Journal of Chemical Physics*, vol. 68, no. 8, pp. 3801–3807, 1978.

[3] R. G. Pearson, "Recent advances in the concept of hard and soft acids and bases," *J. of Chem. Edu.*, vol. 64, no. 7, p. 561, 1987.

[4] A. J. M. Ribeiro, G. L. Holliday, N. Furnham, J. D. Tyzack, K. Ferris, and J. M. Thornton, "Mechanism and catalytic site atlas (m-csa): a database of enzyme reaction mechanisms and active sites," *Nucleic acids research*, vol. 46, no. D1, pp. D618–D623, 2017.

[5] R. G. Parr and R. G. Pearson, "Absolute hardness: companion parameter to absolute electronegativity," *Journal of the American Chemical Society*, vol. 105, no. 26, pp. 7512–7516, 1983.

[6] A. Cedillo and R. Contreras, "A local extension of the electrophilicity index concept," *Journal of the Mexican Chemical Society*, vol. 56, no. 3, pp. 257–260, 2012.

[7] R. G. Parr, R. G. Parr, W. Yang, and W. Yang, "Density functional approach to the frontier-electron theory of chemical reactivity," *J. Am. Chem. Soc.*, vol. 106, no. 14, pp. 4049–4050, 1984.

[8] J. I. Martínez-Araya, "Why is the dual descriptor a more accurate local reactivity descriptor than fukui functions?," *Journal of Mathematical Chemistry*, vol. 53, no. 2, pp. 451–465, 2015.

[9] S. Noorizadeh and H. Parsa, "Evaluation of absolute hardness: a new approach," *The Journal of Physical Chemistry A*, vol. 117, no. 5, pp. 939–946, 2013.

[10] J. Padmanabhan, R. Parthasarathi, M. Elango, V. Subramanian, B. Krishnamoorthy, S. Gutierrez-Oliva, A. Toro-Labbé, D. Roy, and P. Chattaraj, "Multiphilic descriptor for chemical reactivity and selectivity," *The Journal of Physical Chemistry A*, vol. 111, no. 37, pp. 9130–9138, 2007.

[11] L. Meneses, W. Tiznado, R. Contreras, and P. Fuentealba, "A proposal for a new local hardness as selectivity index," *Chemical physics letters*, vol. 383, no. 1-2, pp. 181–187, 2004.

[12] M. Berkowitz, S. K. Ghosh, and R. G. Parr, "On the concept of local hardness in chemistry," *Journal of the American Chemical Society*, vol. 107, no. 24, pp. 6811–6814, 1985.

[13] C. Cárdenas, W. Tiznado, P. W. Ayers, and P. Fuentealba, "The fukui potential and the capacity of charge and the global hardness of atoms," *J. Phys. Chem. A*, vol. 115, no. 11, pp. 2325–2331, 2011.

[14] T. Gal, P. Geerlings, F. De Proft, and M. Torrent-Sucarrat, "A new approach to local hardness," *Physical Chemistry Chemical Physics*, vol. 13, no. 33, pp. 15003–15015, 2011.

[15] K. Fukushima, M. Wada, and M. Sakurai, "An insight into the general relationship between the three dimensional structures of enzymes and their electronic wave functions: Implication for the prediction of functional sites of enzymes," *Proteins: Struct. Func.*, vol. 71, no. 4, pp. 1940–1954, 2008.

[16] J. Khandogin, K. Musier-Forsyth, and D. M. York, "Insights into the regioselectivity and rna-binding affinity of hiv-1 nucleo-capsid protein from linear-scaling quantum methods," *Journal of molecular biology*, vol. 330, no. 5, pp. 993–1004, 2003.

[17] F. La Porta, J. Giacoppo, P. Ramos, M. Guerreiro, and T. Ramalho, "Computational insights into the role of the frontiers orbital in the chemistry of tridentate ligands," *American Journal of Chemistry*, vol. 2, no. 5, pp. 255–262, 2012.

[18] M. Torrent-Sucarrat, P. Salvador, M. Solà, and P. Geerlings, "The hardness kernel as the basis for global and local reactivity indices," *Journal of computational chemistry*, vol. 29, no. 7, p. 1064, 2008.

[19] R. Pino-Rios, O. Yañez, D. Inostroza, L. Ruiz, C. Cardenas, P. Fuentealba, and W. Tiznado, "Proposal of a simple and effective local reactivity descriptor through a topological analysis of an orbital-weighted fukui function," *Journal of computational chemistry*, vol. 38, no. 8, pp. 481–488, 2017.

[20] I. B. Grillo, G. A. Urquiza-Carvalho, E. J. F. Chaves, and G. B. Rocha, "Semiempirical methods do fukui functions: Unlocking a modeling framework for biosystems," *Journal of Computational Chemistry*, 2020.

[21] I. B. Grillo, G. Urquiza-Carvalho, J. F. R. Bachega, and G. B. Rocha, "Elucidating enzymatic catalysis using fast quantum chemical descriptors," *Journal of Chemical Information and Modeling*, 2020.