

---

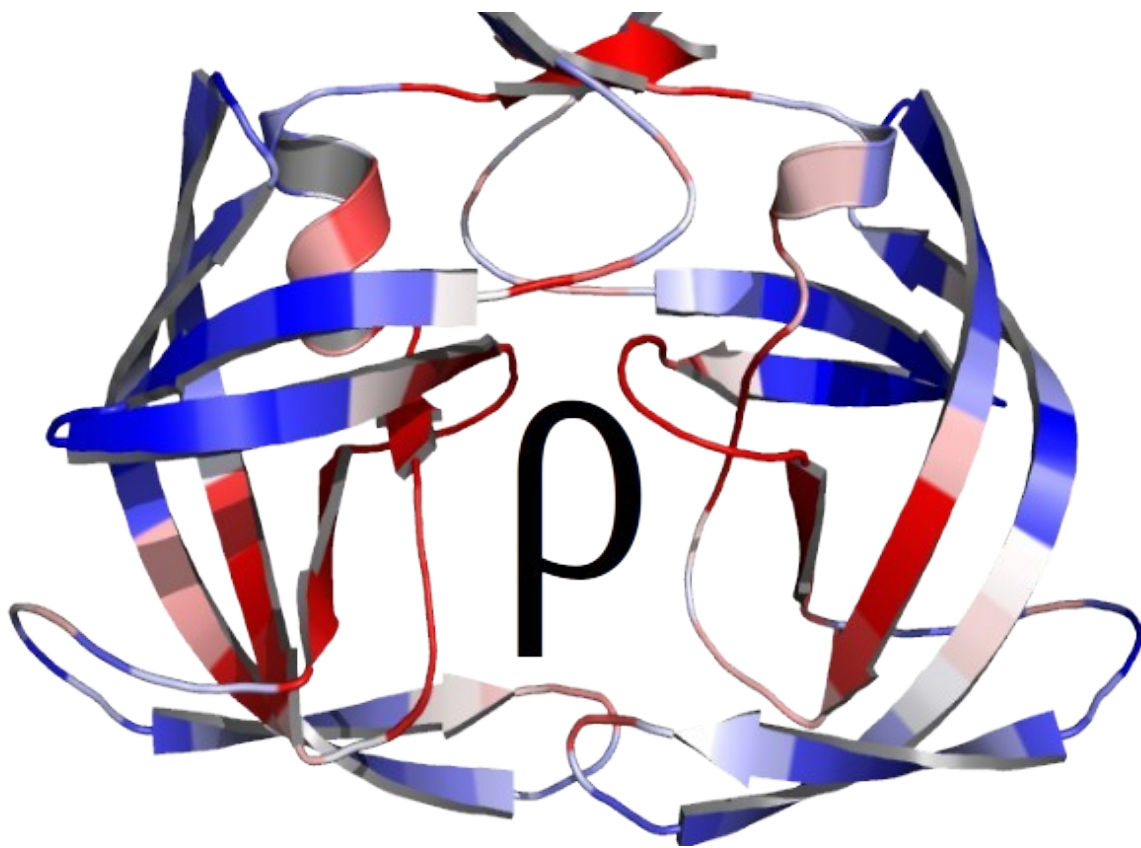
# PRIMoRDiA: Guia do Usuário

---

Igor Barden Grillo  
([barden.igor@gmail.com](mailto:barden.igor@gmail.com))  
[github.com/igorChem](https://github.com/igorChem)

13 de setembro de 2022

## Resumo



**PRIMoRDiA**  
**PRIMoRDiA Macromolecular Reactivity Descriptor Access**

## Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Download e Instalação</b>	<b>3</b>
2.1	Compilação com CMake (mais indicado) . . . . .	4
2.2	Binários Compilados . . . . .	5
<b>3</b>	<b>Versões do Programa</b>	<b>6</b>
<b>4</b>	<b>Descritores de Reatividade: Teoria</b>	<b>7</b>
4.1	Descritores Globais . . . . .	8
4.2	Métodos de Cálculo . . . . .	10
4.3	Descritores Locais . . . . .	11
4.3.1	Métodos de Representação . . . . .	11
4.3.2	Funções de Fukui . . . . .	12
4.3.3	Moleza e Eletrofilicidade Local . . . . .	16
4.4	Métodos de Dureza Local . . . . .	18
4.4.1	Aproximação do Modelo de Thomas-Fermi-Dirac . . . . .	19
4.4.2	Potencial de Fukui . . . . .	20
4.4.3	Aproximação Local do Potencial Químico Eletrônico . . . . .	21
4.4.4	Distribuição de Fukui . . . . .	22
4.5	Descritores de Banda e para Resíduos . . . . .	23
<b>5</b>	<b>Descritores Implementados</b>	<b>25</b>
<b>6</b>	<b>Programas Suportados e Informação Necessária</b>	<b>32</b>
<b>7</b>	<b>Execução do Programa e Opções Gerais</b>	<b>34</b>
<b>8</b>	<b>Opções do Input para o Cálculo dos Descritores</b>	<b>36</b>
8.1	Modos de Execução: 'Reaction' e 'Trajectory' . . . .	39
<b>9</b>	<b>Considerações Finais</b>	<b>41</b>

## 1 Introdução

PRIMoRDiA ( **PRIMoRDiA Macromolecular Reactivity Descriptors Access** ) é um software escrito em *C++*, desenvolvido para o pós-processamento eficiente de resultados de pacotes de química computacional, produzindo uma grande variedade de descritores quânticos, que são propriedades eletrônicas e de reatividade dos sistemas químicos.

O PRIMoRDiA traz implementado os principais descritores de reatividade da Teoria do DFT conceitual, incluindo as definições matemáticas para as teorias de elétrons de fronteira de Fukui e de Ácidos e Bases Duros e Moles de Pearson. O PRIMoRDiA foi desenvolvido com foco no cálculo para grandes moléculas que são relevantes para processos biológicos, e por isso há descritores modificados especificamente para atender as particularidades desses sistemas.

Esse guia traz informações gerais sobre a teoria por de trás dos descritores, e detalha o uso e a interpretação dos resultados gerados pelo programa. Para mais informações nós encorajamos o contato através dos e-mails fornecidos no repositório e nesse documento. PRIMoRDiA foi desenvolvido na Universidade Federal da Paraíba, no Laboratório de Química Quântica e Computacional.

Este Guia do Usuário esta em constante modificação. Novas versões dele serão disponibilizadas conforme identificado as necessidades dos usuários e novas funcionalidades sendo implementadas no programa. Esse documento traz informações atualizadas para até o lançamento de versão estável 1.25v.

## 2 Download e Instalação

PRIMoRDiA é um software de código livre sob a licença pública Mozilla 2.0. No repositório do Git Hub nós fornecemos versões estáveis do programa, com todos os arquivos fonte, dados para teste e binários já compilados. Nós indicamos fortemente a compilação do programa usando o *CMake*.

O programa foi desenvolvido para sistemas operacionais Linux, visando usuários acostumados a trabalhar na área de modelagem molecular. No entanto, para fins didáticos e de divulgação, é fornecido código e binário para o sistema Windows em outro repositório: [https://github.com/igorChem/PRIMoRDiA\\_WIN64](https://github.com/igorChem/PRIMoRDiA_WIN64). Lembrando que não há um suporte constante para essa versão.

Para baixar os dados do repositório é possível usar o seguinte comando do git:

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

Através desse comando, a pasta do repositório vai ser baixada no diretório de onde o seu terminal está aberto, podendo ser atualizado facilmente com o comando:

```
pc@user:$ git pull
```

Também, no próprio site é possível baixar um '.zip' contendo todos os dados do repositório. No entanto, é fortemente recomentado que o usuário baixe uma das versões estáveis, de preferência a última. Essas versões estáveis tem a garantia de estarem com todas as suas funcionalidades devidamente testadas.

*PRIMoRDiA é um programa que está em constante desenvolvimento e clonar o repositório em qualquer outro ponto que não seja as marcadas como estáveis pode gerar erros na utilização.*

Link para última versão: <https://github.com/igorChem/PRIMoRDiA1.0v/releases/tag/v1.25>

## 2.1 Compilação com CMake (mais indicado)

A compilação com o CMake é muito simples de se realizar, com poucos comandos é gerado um binário do PRIMoRDiA otimizado para a sua máquina. A desvantagem da compilação em relação a utilização dos binários prontos é a necessidade de instalação de algumas bibliotecas exigidas pelo PRIMoRDiA. Às vezes, o usuário também não tem os compiladores e o próprio CMake para o processo de compilação.

Em distribuições Linux mais populares baseadas em Debian, como por exemplo Ubuntu e Mint, isso pode ser facilmente resolvido instalando as bibliotecas a partir do repositório.

Por padrão, o PRIMoRDiA utiliza o g++-8 e a biblioteca Eigen 3. Para assegurar que sua máquina tenha o necessário para compilar o PRIMoRDiA, execute os comandos abaixo

```
pc@user:$ sudo apt install g++-8
pc@user:$ sudo apt install cmake
pc@user:$ sudo apt install libeigen3-dev
```

Depois de descomprimir a pasta baixada com a versão estável entre no diretório e faça a compilação usando os seguintes comandos:

```
pc@user:$ cd /caminho/da/pasta/PRIMoRDiA1.0v
pc@user:$ cmake .
pc@user:$ make
```

O processo de compilação deve ocorrer sem problemas e gerar um arquivo binário chamado "PRIMoRDIA\_1.25v ", caso a versão

1.25 seja a escolhida ( mais recomendado ). Com a linha de código mostrada abaixo é possível criar um atalho no arquivo ".bashrc" para executar o programa de qualquer diretório do seu sistema

```
pc@user:$ alias primordia='caminho/PRIMoRDiA1.0v/PRIMoRDiA\_1.25v'
```

## 2.2 Binários Compilados

No caso da impossibilidade de compilação, binários para a última versão estável é fornecido no repositório. O mesmo processo de colocar no caminho pode ser realizado. Problemas de falta de alguma biblioteca compartilhada podem acontecer, assim como a performance máxima não ser alcançada. Entretanto, qualquer problema com a execução dos binários é encorajado que os mesmos sejam reportados.

### 3 Versões do Programa

Até o momento da divulgação desse Guia de Usuário, o programa possui três versões estáveis. As mudanças que aparecem de forma mais sensível ao usuário estão no formato do input e na quantidade de dados extraído pelo software. O ideal é sempre dar preferência para a última versão estável.

1. **1.0v:**

Nove descritores globais e quatorze locais implementados. Descritores para macromoléculas, com descritores assinalados individualmente para resíduos de proteínas e DNA (formato do PDB).

2. **1.2v:**

Novo modelo de input, com mais controle; Scripts de R para análise estatística de trajetória de reação e dinâmica molecular, enfoque em análise nos átomos da coordenada de reação e resíduos requeridos no input. Ajustes gerais no código.

3. **1.25v:** Mais três métodos de dureza local implementados, e os mesmos descritores locais implementados para representação volumétrica e condensada. Ajustes na automatização de visualização dos resultados no Pymol. Correções e padronização de nomenclatura.

## 4 Descritores de Reatividade: Teoria

Os descritores de reatividade (DR) são quantidades teóricas que são convenientes para resumir propriedades de reatividade de sistemas moleculares, como um todo ou de diferentes regiões e/ou átomos. Os DRs de mais sucesso na literatura são os baseados na Conceptual Density Functional Theory (CDFT), que consegue englobar diferentes teorias de reatividade química através de definições matemáticas extraídas do DFT, relacionando essas quantidades com conceitos químicos já bem estabelecidos[1].

Esses descritores são obtidos da densidade eletrônica calculada a partir de sistemas moleculares, sendo propriedades que só podem ser obtidas a partir de cálculos quânticos. Outras propriedades eletrônicas, como a própria densidade eletrônica, energia total e cargas parciais, são exemplos de quantidades calculáveis a partir da densidade eletrônica e podem servir de descritores. O conjunto de todas essas propriedades calculadas pelo PRIMoRDiA são chamadas pelo nosso grupo de Descritores Quânticos Moleculares (DQM) ou só Descritores Quânticos.

Os DRs definidos dentro da CDFT provem da equação diferencial total do DFT (Equação 1) e suas derivadas[2], que relaciona a energia eletrônica ( $E$ ) com a densidade eletrônica ( $\rho(r)$ ) do estado de referência e o potencial externo ( $\nu(r)$ ). Os DR da CDFT podem ser classificados como sendo globais, um número que representa uma propriedade eletrônica para todo o sistema, ou locais, uma função da posição do sistema.

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

Como os principais descritores são derivadas, o cálculo se dá através de dois métodos de aproximação, que no caso são a Aproximação de Orbitais Congelados (AOC) e de Diferenças Finitas (DF). Os descritores locais podem ser separados ainda pelo método de representação/visualização, Condensado para átomos ou em uma representação volumétrica de uma grade tridimensional, com cada voxel representada por um valor. Esse esquema de classificação dos descritores pode ser visualizado na Figura 1.

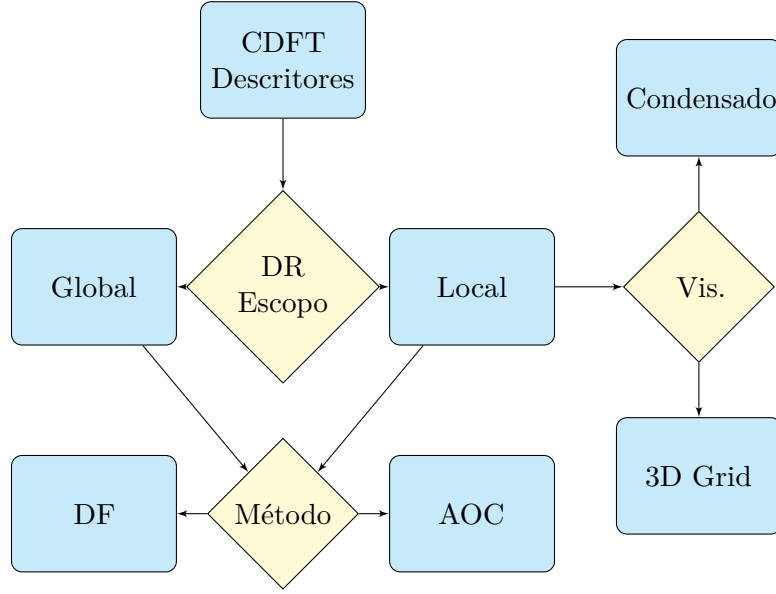


Figura 1: Classificação dos descritores de Reatividade do CDFT.

#### 4.1 Descritores Globais

O potencial químico eletrônico (PQE) é um exemplo de descritor de reatividade global, que no DFT é o multiplicador de Lagrange (representado pela letra  $\mu$ ) que minimiza a energia eletrônica em relação a densidade, dado o potencial externo constante. Na tabela Tabela 1 é apresentada a definição matemática para os descritores de reatividade global e as formas de cálculo nos dois métodos de aproximação.

Tabela 1: Descritores de Reatividade implementados no PRIMoRDiA

DR global	Definição em CDFT	AOC	DF	Ref.
IP	-	$-E_{HOMO}$	$E_{N-1} - E_N$	
EA	-	$-E_{LUMO}$	$E_N - E_{N+1}$	
$\mu$	$\left(\frac{\partial E}{\partial N}\right)_\nu$	$\frac{E_{HOMO} + E_{LUMO}}{2}$	$\frac{E_{N-1} - E_{N+1}}{2}$	[3]
$\eta$	$\left(\frac{\partial^2 E}{\partial N^2}\right)_\nu$	$E_{LUMO} - E_{HOMO}$	$E_{N-1} + E_{N+1} - 2E_N$	[4]
Softness( $S$ )	$1/\eta$	-	-	[4]
$\omega$	$\frac{\mu^2 S}{2}$	-	-	[5]
$nMax$	$-\frac{1}{\eta}$	-	-	[5]

O PQE estabelece a propensão dos sistema de doar elétrons, e quando os valores entre diferentes sistemas moleculares se igualam a transferência de elétrons cessa, caracterizando o estado de equilíbrio[4]. Por causa disso é que  $\mu$  é batizado em analogia ao



potencial químico molar da termodinâmica clássica.

O princípio de Ácidos e Bases Duros e Moles (*HSAB: Hard and Soft Acid and Base*) de Pearson[6] é uma das teorias de reatividade química que ganhou fundamentação teórica e matemática dentro do CDFT. Pearson criou conceitos batizados de dureza e moleza química para explicar a afinidade entre ácidos e bases pelo modelo de Lewis, baseado na observação de características atômicas e moleculares e as séries de reatividade[7]. Desse estudo foi definido que ácidos Duros são pequenos receptores de elétrons e pouco polarizáveis, que por sua vez tendem a interagir quimicamente com bases duras, que são doadores de elétrons com as mesmas características de volume e polarizabilidade. Essa interação entre espécies consideradas quimicamente "duras" por Pearson ocorrem por interações controladas por carga, formando ligações de caráter iônico.

Já os ácidos e bases moles são espécies polarizáveis e volumosas, onde a interação é governada pela deslocalização dos orbitais moleculares de fronteira, havendo uma doação de elétrons do orbital ocupado de mais alta energia (HOMO) da base para o orbital virtual de mais baixa energia do ácido (LUMO). Nesse processo de interação de espécies moles há uma transferência de carga líquida da base para o ácido e também tendo maior chance de formação de ligações com caráter covalente. Logo, é possível estabelecer regras complementares para entender as principais forças que levam uma reação química a acontecer usando esse princípio de Pearson.

No entanto, até a consolidação do CDFT, o HSAB era uma teoria sem respaldo matemático e uma forma sistemática de determinar escalar de dureza e moleza. No CDFT, a dureza é primeiramente encarada como uma propriedade do sistema que representa a sua resistência em doar elétrons, sendo também a característica de sistemas pouco polarizáveis e de pequeno volume, ou seja, onde os elétrons ficam mais aglutinados aos núcleos.

A definição matemática do descritor global associada a dureza química no CDFT é a derivada do PQE em respeito ao número de elétrons, o mesmo que a derivada segunda da energia em respeito ao número de elétrons, e comumente representado com a letra grega  $\eta$ . Rapidamente, a moleza química global é definida como o recíproco da dureza, normalmente representado pela letra  $S$ .

Outros descritores globais que podem ser obtidos no CDFT são basicamente sobre esses processos de transferência de carga, que geralmente são utilizados para racionalizar reações químicas. Como por exemplo: O máximo número de elétrons que podem ser recebidos por um sistema por um doador ideal  $nMax$ ; a estabilização de energia quando o sistema recebe esse número de elétrons, que pode ser chamado de eletrofilicidade e representado na literatura do CDFT com a letra grega  $\omega$ [8].

## 4.2 Métodos de Cálculo

Como já citado acima, a maioria dos descritores de reatividade são derivadas da energia e/ou da densidade eletrônica em respeito ao número de elétrons. O número de elétrons é sempre inteiro, e portanto não existe uma função contínua dessa quantidade, sendo necessário que essas derivadas sejam aproximados através de métodos matemáticos.

O método mais comum para a determinação numérica de derivadas, seja qual for, é o métodos de diferenças finitas (DF). Para as quantidades do CDFT, os cálculos tem que ser realizados utilizando dados da estrutura eletrônica de três estados de carga do sistema. Para os descritores globais, as informações necessárias são as energias eletrônicas desses três estados de carga, o estado de referência, o carregado positivamente (pelo menos um elétrons a menos) e o carregado negativamente (pelo menos um elétron a mais).

Para os descritores locais essa lógica se aplica da mesma forma para a densidade eletrônica, ou carga parcial no caso da representação condensada por átomos. Todos os cálculos de estrutura eletrônica devem ser realizados para as mesmas coordenadas atômicas do estado de referência, isto é, para a mesma posição dos núcleos, respeitando a condição de potencial externo constante requerido por essas derivadas.

O outro método de cálculo é chamado de Aproximação de Orbitais Congelados (*Frozen Orbital Approximation* em inglês)[9], que utiliza as informações dos orbitais moleculares de fronteira, HOMO e LUMO, e desconsidera a variação de energia/distribuição espacial dos orbitais internos, o que justifica o nome do método.

Para os descritores globais em particular, os cálculos são realizados com a energia dos orbitais moleculares HOMO e LUMO, e também é conhecido como Teorema de Koopmans[10], que estabelece que a energia do HOMO é a aproximação do Potencial de Ionização (PI) e a energia do LUMO como sendo a aproximação da Afinidade Eletrônica (AE). A grande diferença é dessa aproximação para a de diferenças finitas é que só requer uma avaliação de estrutura eletrônica (comumente chamado de *Single Point* em inglês). E em relação aos descritores locais, a obtenção das densidades de probabilidade desses orbitais requer um esforço computacional irrisório frente ao necessário para a densidade eletrônica total.

Logo, AOC tem um custo computacional significativamente menor que o DF, sendo o último se tornando proibitivo para grandes moléculas. Quanto a acurácia, a aproximação de DF por definição é teoricamente maior, já que considera a perturbação de toda a densidade eletrônica quando o sistema participa de processos de transferência de carga. No entanto, a AOC é estimada por orbitais moleculares que geralmente tem grande importância na explicação

da reatividade química das moléculas, pois da a distribuição de probabilidade de localização dos elétrons que são mais suscetíveis aos processos químicos.

### 4.3 Descritores Locais

A definição matemática de um descritor local é uma função do espaço tridimensional  $r$ , ou seja, dado um ponto no espaço o descritor assume certo valor. A densidade eletrônica  $\rho$  é uma dessas quantidades locais por definição, onde o valor da densidade é definida pelo produto dos orbitais moleculares ocupados  $\phi_{occ}(r)$  (Equação 2), que por sua vez são combinações lineares de orbitais atômicos  $\phi$ , onde os coeficientes  $c_i$  são determinados por métodos de química computacional.

$$\rho(r) = \sum_i^{occ} |\psi(r)_i|^2 \quad (2)$$

$$\psi(r)_i = \sum_i c_i \phi(r) \quad (3)$$

#### 4.3.1 Métodos de Representação

A forma prática de representar esses descritores locais é definir uma grade tridimensional, onde cada ponto  $r$  relaciona coordenadas tridimensionais do sistema químico com um valor. Esses pontos são conhecidos como *voxel*, o análogo tridimensional do pixel, e geralmente são reunidos em um arquivo formatado chamado de *Gaussian Cube*, ou simplesmente *cube*, com a extensão ".cube". Esse arquivo tem um padrão que é reconhecido pela grande maioria dos softwares de visualização molecular, podendo a partir deles criar representações gráficas de iso-superfícies, renderizar mais de uma dessas iso-superfícies atribuindo diferentes cores e opacidades, o que resulta em uma representação volumétrica.

Dentro da linguagem dos descritores quânticos adotada pelo PRIMoRDiA, essa representação é conhecida como volumétrica. O PRIMoRDiA calcula 21 propriedades eletrônica/descritores de reatividade que podem ser escritas no disco como arquivos *Cube* formatados. Para isso, é necessário que o usuário defina a resolução da grade tridimensional diferente de zero e forneça os dados de estrutura eletrônica necessários para o cálculo desses descritores quânticos. Mais detalhes sobre os parâmetros, input e os descritores em si são dados nas próximas sessões desse guia.

Como já dito antes, cada voxel representa um valor do descritor local para um dado ponto  $r$ . Para facilitar a interpretação química e manipulação numérica dessas propriedades é possível integrar elas em uma base comum  $\Omega$ . Como por exemplo, uma fração da densidade eletrônica pode ser atribuída a um átomo do sistema

molecular através de uma análise populacional, como a de Muliken ou Lowdin por exemplo. Esses métodos numéricos somam os coeficientes dos orbitais moleculares, ou elementos da matriz densidade, que são atribuídos aos orbitais atômicos, particionando esses valores para cada átomo. Esse processo é conhecido como condensação, gerando uma representação condensada para átomos, e geralmente abreviada para somente "condensada".

No caso genérico mostrado na Equação 4, o descritor  $f(r)$  pode ter parte sua atribuída ao átomo  $k$  através da integração dos valores de  $f(r)$  nas funções de base referente a esse átomo. O PRIMoRDIA calcula os descritores e escreve um arquivo com a lista de valores dos descritores locais ordenadas pelo índice dos átomos no arquivo fornecido para o software.

$$f_k^a = \int_{\Omega_k} f(r) dr \quad (4)$$

#### 4.3.2 Funções de Fukui

A função de Fukui pode ser considerada o primeiro descritor de reatividade local, definido no CDFT como a derivada da densidade eletrônica em respeito a variação do número de elétrons (Equação 5), sendo requerida no cálculo da maioria dos outros descritores locais. Há outros descritores locais anteriores, como a próprio superdeslocalizabilidade dos orbitais de fronteira criados por Kenichi Fukui que inspirou a função e os índices de Fukui do CDFT[11]. No entanto, essas outros propriedades tiveram seus conceitos consolidados no CDFT através de um ferramental matemático, tornando essas outras quantidade obsoletas.

Ao que se refere a reatividade, a função de Fukui indica os locais onde há a maior propensão da densidade eletrônica variar quando o sistema participa de um processo de transferência de carga. Pelas questões já citadas, os descritores que são derivadas em respeito ao número de elétrons devem ser aproximados por algum método.

$$f(r) = \left( \frac{\partial \rho(r)}{\partial N} \right)_\nu \quad (5)$$

Os índices de Fukui surgem da aproximação da função de Fukui pelo método de diferenças finitas, onde a derivada é aproximada pelos limites a esquerda e a direita em relação a variação do número de elétrons  $N$ . A aproximação a esquerda assume um valor menor que  $N$ , o mínimo possível sendo  $N - 1$  e o valor da densidade no estado de referência. A diferença entre o estado de referência e o estado carregado positivamente corresponde ao índice de Fukui que descreve susceptibilidade local do sistema de receber um ataque

eletrofílico, sinalizada como  $f^-(r)$  como mostrado na Equação 6.

$$f^-(r) = \rho(r)_N - \rho(r)_{N-1} \quad (6)$$

A interpretação do cálculo de  $f^-(r)$  é que a diferença de densidade eletrônica entre esses diferentes estados de carga vai revelar as regiões onde a densidade eletrônica sobra quando o sistema perde um elétron. Essa densidade eletrônica excedente é a mais susceptível a se "desprender" do sistema e servir para formar uma ligação com um eletrófilo ou de ser transferida.

A aproximação a direita considera um estado de carga com um número de elétrons maior que  $N$ , sendo no mínimo  $N + 1$ . A diferença entre o estado de carga negativamente carregado e o de referência é o índice de susceptibilidade local do sistema de receber um ataque nucleofílico, representada por  $f^+(r)$  mostrado na Equação 7.

$$f^+(r) = \rho(r)_{N+1} - \rho(r)_N \quad (7)$$

A interpretação do cálculo de  $f^+(r)$  é que a diferença de densidade eletrônica entre esses diferentes estados de carga vai revelar as regiões onde a densidade eletrônica se concentra quando o sistema ganha um elétron, indicando a região mais provável onde um nucleófilo irá formar uma ligação ou transferir densidade eletrônica.

A valor médio desses dois índices é a derivada no ponto 0, e é utilizada para indicar a susceptibilidade local do sistema de receber um ataque radical. O que na pratica é observada na equação desse índice é a reatividade média entre a susceptibilidade a um ataque nucleofílico e a um ataque eletrofílico, ou seja, a reatividade média do sistema independente do tipo de ataque.

$$f^0(r) = \frac{\rho(r)_{N+1} - \rho(r)_{N-1}}{2} \quad (8)$$

Esses descritores tem alguns nomes e termos em inglês que podem variar. Isso ocorre por uma questão de equilíbrio entre precisão dos termos e seu entendimento químico. Como por exemplo a *susceptibilidade a ataque eletrofílico* que nada mais é que a nucleofilicidade local do sistema. O primeiro é o termo mais técnico referente ao que foi calculado e o último tem um compromisso maior com o entendimento da quantidade teórica em questão.

Isso é importante dizer pois diferentes versões do programa e de trabalhos publicados do nosso grupo adotam diferentes nomenclaturas para esses descritores locais, que com o tempo foi se considerando cada vez mais a transparência e simplificação da nomenclatura do que a sua precisão em relação ao que é definido matematicamente.

Isso tudo é detalhada na próxima sessão do guia do usuário. Assim como "índices" e "funções locais" são similares e vão ser invocados dependendo da escolha dos autores.

Um quarto índice de Fukui que pode ser considerado é o índice Dual de Fukui ou somente descritor Dual, que possui esse nome por simultaneamente indicar os locais do sistema suscetíveis a ataques eletrofílicos e nucleofílicos. O descritor dual  $\Delta f^\pm$  é calculado pela diferença  $f^+(r)$  e  $f^-(r)$  (Equação 9), isto é, a reatividade local efetiva do sistema. Também, esse descritor é indicado como sendo em algumas vezes mais acurado que  $f^+(r)$  e  $f^-(r)$  individualmente devido ao cancelamento de erros que eventualmente ocorre quando há uma subtração entre essas quantidades[12]. No PRIMoRDiA, nós adotamos o nome "Netphilicity", no lugar de descritor dual.

$$\Delta f^\pm(r) = f^+(r) - f^-(r) \quad (9)$$

As funções de Fukui também podem ser obtidas com a aproximação de orbitais congelados, que se baseia que a variação infinitesimal no número de elétrons do sistema provoque uma mudança irrelevante nos orbitais moleculares internos. No caso de  $f^-(r)$  a sua definição de cálculo em função da densidade dos  $M$  níveis de energia (orbitais moleculares/orbitais de Khon-Sham) é dada na Equação 10 e a aproximação acaba sendo a densidade do orbital molecular de fronteira ocupada de maior energia, o vulgo HOMO (Equação 11).

$$f^-(r) = \lim_{\delta \rightarrow 0^-} \frac{\partial \rho_{M+\delta}(r)}{\partial N} = |\phi_M(r)|^2 + \sum_{i=1}^{M-1} \frac{\partial}{\partial N} |\phi_i(r)|^2 \quad (10)$$

$$f^-(r) = |\phi_{HOMO}(r)|^2 \quad (11)$$

Para  $f^+(r)$  é bem semelhante, onde a derivada é tomada do limite de uma variação positiva (a direita) do número de elétrons(Equação 12), resultando no nível de energia  $M+1$ , que é o nível de energia virtual de menor energia, o vulgo LUMO (Equação 13).

$$f^+(r) = \lim_{\delta \rightarrow 0^+} \frac{\partial \rho_{M+\delta}(r)}{\partial N} = |\phi_{M+1}(r)|^2 + \sum_{i=1}^M \frac{\partial}{\partial N} |\phi_i(r)|^2 \quad (12)$$

$$f^+(r) = |\phi_{LUMO}(r)|^2 \quad (13)$$

Para o cálculo de  $f^0(r)$  pelo AOC é feita pela média entre  $f^+(r)$  e  $f^-(r)$ , ou seguindo a lógica do uso dos orbitais moleculares de

fronteira,  $f^0(r)$  pode ser aproximado pelo orbital molecular unicamente ocupado, abreviado em inglês como SOMO (Single Occupied Molecular Orbital) (Equação 14).

$$f^0(r) = |\phi_{SOMO}(r)|^2 \quad (14)$$

As funções/índices de Fukui ainda podem ser representados de forma condensada para átomos para cada uma das aproximações possíveis. No caso de diferenças finitas, a densidade eletrônica no ponto  $r$  é substituída pela carga parcial de cada átomo, que vai depender da análise populacional utilizada. A função  $f^-$  é definida para cada átomo  $k$  como mostrado na Equação 15 e  $f^+$  na Equação 16. Consequentemente,  $f^0(k)$  tomado como a média de  $f_k^+$  e  $f^-(k)$ , e  $\Delta f^\pm$  como a diferença.

$$f^-(k) = q_{k[N]} - q_{k[N-1]} \quad (15)$$

$$f^+(k) = q_{k[N+1]} - q_{k[N]} \quad (16)$$

Para a AOC, os orbitais de fronteira é que devem ser condensados para os átomos, e isso é feito de forma a particionar esses orbitais moleculares por contribuição dos orbitais atômicos AO de índice  $\nu$  e sua sobreposição com os outros orbitais ( $\mu$ ) do mesmo átomo  $k$  contabilizando como o produto cruzada ponderado pela matriz de sobreposição, também chamada de integrais de um elétron  $S_{\nu\mu}$ . Esse processo pode ser realizado para a condensação da densidade de probabilidade para qualquer orbital molecular  $\psi^i$ , como mostrado na Equação 31.

Para AOC, a condensação por átomos é realizada usando a densidade de probabilidade dos orbitais de fronteira, cujo a contribuição é dada pelos coeficientes dos orbitais atômicos AO de índice  $\nu$  e sua sobreposição com os outros orbitais ( $\mu$ ) do mesmo átomo  $k$  contabilizando como o produto cruzada ponderado pela matriz de sobreposição, também chamada de integrais de um elétrons  $S_{\nu\mu}$ . Esse processo pode ser realizado para a condensação da densidade de probabilidade para qualquer orbital molecular  $\psi^i$ , como mostrado na Equação 17.

$$|\psi^i(k)|^2 = \sum_{\nu \in k}^{AO} \left\{ |C_{\nu i}|^2 + \sum_{\mu \notin \nu}^{AO} |C_{\nu i} C_{\mu i}| S_{\nu\mu} \right\} \quad (17)$$

Portanto, para  $f^-(k)$  é utilizado a densidade de probabilidade do

HOMO ( Equação 18 ) e para  $f^+(k)$  o LUMO ( Equação 18 ).

$$f^-(k) = |\psi^{HOMO}(k)|^2 = \sum_{\nu \in k}^{AO} \left\{ |C_{\nu HOMO}|^2 + \sum_{\mu \notin \nu}^{AO} |C_{\nu HOMO} C_{\mu HOMO}| S_{\mu\nu} \right\} \quad (18)$$

$$f^+(k) = |\psi^{LUMO}(k)|^2 = \sum_{\nu \in k}^{AO} \left\{ |C_{\nu LUMO}|^2 + \sum_{\mu \notin \nu}^{AO} |C_{\nu LUMO} C_{\mu LUMO}| S_{\mu\nu} \right\} \quad (19)$$

Até aqui definimos as formas de cálculo para as funções de Fukui, na sua forma condensada e volumétrica, e para os dois tipos de aproximação possível.

#### 4.3.3 Moleza e Eletrofilicidade Local

Descritores globais podem ser definidos de forma local substituindo o número de elétrons pela densidade eletrônica, que de fato é a representação local do número de elétrons. Um exemplo clássico é a moleza local  $s(r)$ , definida na Equação 20 como a derivada da densidade eletrônica em respeito ao potencial químico com potencial externo constante. Pelas relações de Maxwell mostradas também na Equação 20, essa definição pode ser obtida pelo produto moleza global e da função de Fukui. Assim a moleza local é a moleza global distribuída para as regiões de maior reatividade do sistema, sendo uma combinação de um descritor global e um local[13].

$$s(r) = \left( \frac{\partial \rho(r)}{\partial \mu} \right)_{\nu} = \left( \frac{\partial N}{\partial \mu} \right)_{\nu} \left( \frac{\partial \rho(r)}{\partial N} \right)_{\nu} = S f(r) \quad (20)$$

Como a função de Fukui é dividida em diferentes índices nos métodos de cálculo em relação ao tipo de ataque que o sistema pode sofrer, o cálculo de moleza local acaba sofrendo a mesma partição, obtendo as formas descritas nas Equação 21 – Equação 24.

$$s^-(r) = S f^-(r) \quad (21)$$

$$s^-(r) = S f^-(r) \quad (22)$$

$$s^0(r) = S f^0(r) \quad (23)$$

$$s^{\pm}(r) = S f^{\pm}(r) \quad (24)$$



Também, há a hiper-moleza, que é a derivada da moleza local em relação ao número de elétrons, substituindo a função de Fukui pelo descritor dual  $f^\pm(r)$  e a moleza global pelo seu quadrado[14]. A principal vantagem desse descritor é que ele pode ser utilizado para avaliar simultaneamente a moleza local em regiões eletrofílicas e nucleofílicas.

$$s^2(r) = S^2 f^\pm(r) \quad (25)$$

A eletrofilicidade local surge da mesma forma que a moleza local, podendo ser calculada como o produto da eletrofilicidade total pelas funções de Fukui[15], como mostrado na Equação 26. Logo o mesmo efeito de divisão por tipo de ataque também se apresenta. A eletrofilicidade local particiona a propriedade global referente que é a estabilização de energia eletrônica dada no processo de recebimento de elétrons pelo sistema. Por isso, a definição padrão é dada usando  $f^+(r)$ , que descreve os locais mais propensos a ataques eletrofílicos. Quando utilizado o descritor dual  $f^\pm(r)$  no lugar de  $f^+(r)$  surge o descritor multifílico  $\Delta\omega(r)$ , que serve para descrever o mesmo tipo de processo independente do tipo de ataque[16].

$$\omega(r) = \omega f^+(r) \quad (26)$$

$$\Delta\omega(r) = \omega f^\pm(r) \quad (27)$$

Pela definição matemática esses descritores contém informação local redundante em relação às funções de Fukui, já que eles indicarão como regiões mais reativas aquelas com os valores máximos do índice de Fukui aplicado no cálculo. No entanto, a combinação da informação dos descritores globais permite uma comparação de reatividade local para diferentes moléculas, já que as funções de Fukui são normalizadas e não são capazes de dizer se certo átomo  $\alpha$  na molécula A é mais reativo que o átomo  $\beta$  pertencente à molécula B[17]. As funções de Fukui representam a reatividade intramolecular, e descritores locais como moleza local e eletrofilicidade local podem ser utilizados para uma comparação de reatividade local intermolecular.

Os descritores globais servem para uma comparação de reatividade entre sistemas moleculares diferentes, rendendo uma análise intermolecular, os locais para uma comparação de reatividade por região da molécula, e quando os dois são combinados é possível comparar a reatividade de regiões de moléculas diferentes, ou seja uma comparação intermolecular local.

## 4.4 Métodos de Dureza Local

Através do desenvolvimento das teorias de reatividade, Klopman e Fukui resumiram as forças que governam os processos químicos em dois tipos: Controlados-por-Orbital e Controlados-por-Carga[18, 11]. Pearson definiu que os processos Controlados-por-Carga são protagonizados por ácidos e bases duras, sendo portanto interações do tipo duro-duro, e os processos Controlados-por-Orbital são portanto do tipo mole-mole[8].

O primeiro tem a aproximação das espécies governada por forças Coulombianas, resultado em ligações de maior caráter iônico que covalente. Já as interações mole-mole se caracterizam por um significativa transferência de carga, e sobreposição de orbitais moleculares de fronteira, resultando em ligações de caráter mais covalente que iônico.

Vimos que existe uma definição de moleza local e naturalmente é esperado que exista a dureza local. De fato, se pegarmos o recíproco da Equação 20 obtemos a definição mostrada na Equação 28, uma derivada válida no CDFt. No entanto, a integração de  $\eta(r)$  não resulta na quantidade global sem que haja a multiplicação por uma função kernel, como mostrado na Equação 29. Isso introduz uma ambiguidade na condição de descritor local, que quando integrado em todo o espaço  $r$ , deveria representar em sua totalidade a quantidade global.

$$\eta(r) = \left( \frac{\partial \mu}{\partial \rho(r)} \right)_\nu \quad (28)$$

$$\eta = \int \eta(r) f(r) \quad (29)$$

O CDFt é uma área bastante rigorosa quanto a sua fundamentação matemática e desde a definição a dureza local é alvo de debates e desenvolvimentos, com correntes de pesquisadores buscando uma definição matemática que respeite a condição de integração e outra corrente buscando a sua efetividade na descrição da regioselectividade de reações controladas-por-carga. A partir desse "problema da dureza local", surgiu diversas "equações de trabalho", que são definições matemáticas para o cálculo dessa propriedade para a sua aplicação.

Com o intuito de tornar disponível todas as ferramentas possíveis de cálculo de propriedades oriundas da estrutura eletrônica de macromoléculas, o PRIMoRDiA tem implementado as equações de trabalho mais difundidas na literatura para que o usuário decida qual é a mais apropriada para o seu objeto de estudo. Na versão 1.0v o software foi lançado com quatro versões de dureza local. Na última versão estável correspondente a esse guia de usuário (1.25v), foi implementado algumas variações da dureza local e

disponibilizados mais algumas propriedades eletrônicas locais que podem ser utilizadas para estudar processos de natureza duro-duro. A lista completa dessas quantidades é apresentada na sessão posterior desse guia.

Como essa sessão do guia do usuário é sobre a teoria básica dos descritores quânticos, vai ser demonstrado algumas deduções importantes dessas equações de trabalho, a origem de suas aproximações e o que elas podem representar na interpretação da reatividade de sistemas químicos.

#### 4.4.1 Aproximação do Modelo de Thomas-Fermi-Dirac

A primeira equação de trabalho surge logo que a derivada parcial mostrada na Equação 28 foi definida por Berkowitz em 1985[19], como uma função do Funcional de energia eletrônica  $F[\rho]$  através da relação mostrada na Equação 29. A dureza local fica como uma quantidade proporcional a integral do produto entre a densidade eletrônica e da derivada segunda do funcional de energia eletrônica total em respeito a densidade eletrônica. Quando uma propriedade já é apresenta uma derivada em relação a  $\rho(r)$ , a aplicação de mais uma derivada em respeito a densidade eletrônica vai ser em relação a um  $r'$  representando um ponto no espaço tridimensional diferente de  $r$ .

$$\eta(r) = \left( \frac{\partial \mu}{\partial \rho(r)} \right)_\nu = \frac{1}{2N} \int \frac{\delta^2 F[\rho]}{\delta \rho(r) \delta \rho(r')} \rho(r') dr' \quad (30)$$

Essa definição surge pela relação que existe entre o funcional de energia total  $E[\rho]$  e  $F[\rho]$ , dada na Equação 31, onde é separado a contribuição devido ao potencial dos núcleos dos átomos  $v(r)$  da parte eletrônica. Por sua vez,  $F[\rho]$  é dividida em três contribuições eletrônicas independentes: funcional de energia cinética  $T[\rho]$ ; funcional de interação coulombiana clássica  $J[\rho]$ ; e o funcional de troca-correlação eletrônica  $E_{xc}[\rho]$  (Equação 32). Logo a dureza local pode ser dividida também em três, cada uma correspondente a um funcional de energia eletrônica que compõe  $F[\rho]$  (Equação 33).

$$E[\rho] = \int \rho(r) d\nu(r) dr + F[\rho] \quad (31)$$

$$F[\rho] = T[\rho] + J[\rho] + E_{xc}[\rho] \quad (32)$$

$$\eta_F(r) = \eta_T(r) + \eta_J(r) - \eta_K(r) \quad (33)$$

No modelo de Thomas-Fermi-Dirac (TFD) para essas funcionais, a dureza local referente ao funcional de energia cinética e o de

troca-correlação podem ser negligenciados, já que esses mesmos funcionais assumem valores ínfimos nas regiões dos sistemas químicos onde ocorrem as reações, fazendo com que a Equação 30 possa ser aproximada somente usando o funcional de interação elétron-elétron clássico, como mostrado na Equação 34. Essa aproximação facilita bastante a definição de uma primeira equação de trabalho para dureza local, já que a derivada de  $J[\rho]$  (Equação 35) em respeito a densidade eletrônica é trivial e o cálculo de  $\eta(r)$  fica como mostrado na Equação 36,

$$\eta^{TFD}(r) = \frac{1}{2N} \int \frac{\delta^2 J[\rho]}{\delta \rho(r) \delta \rho(r')} \rho(r') dr' \quad (34)$$

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(r) \rho(r')}{|r - r'|} dr dr' \quad (35)$$

$$\eta^{TFD}(r) = \frac{1}{2N} \int \frac{\rho(r')}{|r - r'|} dr' \quad (36)$$

Essa é a primeira equação de trabalho para dureza local, utilizada até hoje, podendo ser encontrada com o "TFD", indicando a aproximação do modelo explicado. Portanto, a dureza local pode ser estimada como a integral da densidade eletrônica nas posições  $r'$  do sistema, que são todas que não forem  $r$  que é a posição para a qual a propriedade esta sendo calculada. A quantidade ainda é normalizada pelo número de elétrons  $N$ , mas ainda sua integral não é correspondente a quantidade global, pela própria definição da derivada. No PRIMoRDiA, tanto essa aproximação quanto a estimativa completa de  $\eta(r)$  com as outras duas funcionais estão disponíveis.

Para a condensação para os átomos, a densidade eletrônica é calculada para cada átomo usando o somatório do produto dos orbitais moleculares, e as coordenadas dos núcleos substituindo os valores de  $r$  e  $r'$ .

#### 4.4.2 Potencial de Fukui

Uma equação de trabalho que veio logo a seguir foi a que aproxima a densidade eletrônica na Equação 36 pela função de Fukui, ao mesmo tempo que retira o fator de normalização em relação ao número de elétrons pelo fato da função de Fukui já ser normalizada, resultando na Equação 37. Esse descritor também é conhecido como potencial de Fukui, para  $r = 0$  é considerado a dureza na posição do núcleo do átomo, já que a sua definição também é equivalente a derivada da dureza em respeito a variação do potencial externo. Além disso, essa aproximação é justificada pelo fato de que a função de Fukui é considerada por representar a porção mais reativa da densidade

eletrônica, o que em teoria aumentaria o poder de predição da reatividade do sistema.

$$\eta^f(r) = \int \frac{f^-(r')}{|r - r'|} \quad (37)$$

Como o potencial de Fukui depende do cálculo das funções de Fukui, esse descritor se dividi em três e o métodos de cálculo dependem da aproximação. Portanto, o primeiro potencial de Fukui que definimos usando a  $f^-(r)$  é o potencial de Fukui a esquerda  $\eta^{f^-}(r)$  que está implementado no PRIMoRDiA junto com o potencial de Fukui a direita  $\eta^{f^+}(r)$  e potencial de Fukui médio  $\eta^{f^0}(r)$ . Na aproximação de Orbitais congelados esses descritores dependem de regiões onde há maior interação com a densidade dos orbitais de fronteira, e os resultados acabam se diferenciando bastante quando é feito com diferenças finitas. A grande vantagem computacional nesse descritor é que na AOC o cálculo é significativamente mais rápido que a primeira equação de trabalho apresentada, que requer o cálculo da densidade eletrônica total.

#### 4.4.3 Aproximação Local do Potencial Químico Eletrônico

A terceira equação de trabalho surge da tentativa de solucionar a ambiguidade da definição da dureza local através do desenvolvimento de uma versão local do potencial químico eletrônico. Isso por que a derivada do PQE local em respeito ao número de elétrons é uma definição válida para dureza local (Equação 38), e sua integral daria a derivada do PQE global  $\mu$  em respeito ao número de elétrons, que é a própria dureza global, como mostrada na Equação 39.

$$\eta(r) = \left( \frac{\partial \mu(r)}{\partial N} \right)_\nu \quad (38)$$

$$\eta = \int \left( \frac{\partial \mu(r)}{\partial N} \right)_\nu dr = \left( \frac{\partial \mu}{\partial N} \right)_\nu \quad (39)$$

O grande questão teórica nesse caso é que o PQE é uma descritor que indica o equilíbrio químico do sistema em relação a sua variação no número de elétrons, e o sistema em equilíbrio não apresentaria tendência de uma transferência interna de elétrons de uma região para a outra. Isso é o que de fato define como um descritor global, mostrando que o valor é idêntico em todas as regiões do sistema e por isso ele está em equilíbrio.

Mesmo assim, uma definição matemática para a derivada do PQE em relação ao número de elétrons para um dado  $r$  foi desenvolvida como um artifício matemático para chegar em uma definição válida da dureza local que seja integrável para a propriedade global. Para

isso Gal e colaboradores utilizaram de relações matemáticas intrínsecas aos funcionais que sejam normalizáveis[20], fazendo com que a derivada do funcional de energia eletrônica total  $E[\rho]$  em respeito a densidade eletrônica seja escrita como na Equação 40.

$$\frac{\delta E[\rho]}{\delta \rho(r)} = \frac{1}{N} \int \rho(r') \frac{\delta E[\rho]}{\delta \rho(r')} dr' \quad (40)$$

O potencial químico eletrônico acaba sendo escrito como uma integral sobre o espaço de uma outra quantidade local, nesse caso a densidade eletrônica, como mostrado na Equação 41. Logo, retirando a integral na Equação 41, o PQE se torna uma função de  $r$ , isto é, uma função local, que pode ser escrita como na Equação 42.

*Basicamente o PQE local é a quantidade local distribuída pela densidade eletrônica normalizada, e portanto integrando perfeitamente para a quantidade global.*

$$\mu = \frac{1}{N} \int \rho(r) \frac{\delta E[\rho]}{\delta \rho(r)} dr \quad (41)$$

$$\mu(r) = \frac{\rho(r)}{N} \frac{\delta E[\rho]}{\delta \rho(r)} = \frac{\rho(r)}{N} \mu \quad (42)$$

A partir dessa definição de  $\mu(r)$ , é possível escrever a derivada do PQE local em relação ao número de elétrons resultando na dureza local  $\eta(r)$ , como mais uma equação de trabalho mostrada na Equação 43. Na primeira parcela da equação a parte local é dada pela diferença da densidade do HOMO pela densidade eletrônica normalizada, cujo a integral é zero. Na segunda parcela a integral é igual a dureza global, resolvendo a ambiguidade da definição da dureza local.

$$\eta(r) = \left( \rho(HOMO) - \frac{\rho(r)}{N} \right) \frac{\mu}{2N} + \frac{\rho(r)}{N} \eta \quad (43)$$

#### 4.4.4 Distribuição de Fukui

O último desenvolvimento de equação de trabalho tratada aqui é a que distribui a dureza global usando as funções de Fukui a esquerda  $f^-(r)$  e a direita  $f^+(r)$ . Nesse caso, a dureza global é dividida nas suas parcelas de Potencial de Ionização e Afinidade Eletrônica (Equação 44), e a dureza local é definida como o a diferença do produto entre PI e  $f^-(r)$  e do produto entra EA e  $f^+(r)$  (Equação 45).

$$\eta = PI - AE \quad (44)$$

$$\eta(r) = PIf^-(r) - AEf^+(r) \quad (45)$$

Essa equação de trabalho foi proposta por Meneses e colaboradores[21], e se baseia na noção que espécies eletrofílicas tem PI altos e um nucleófilo teria uma EA baixo, maximizando  $\eta(r)$  onde as funções de Fukui a esquerda indicasse a densidade eletrônica mais reativa. Apesar de se basear em conceitos de reatividade química bem estabelecidos e cumprir com a exigência de integração para quantidade global, essa equação de trabalho depende bastante de propriedades que correlacionam com interações mole-mole, como pontos máximos da função de Fukui e energia de orbitais de fronteira ressonantes.

#### 4.5 Descritores de Banda e para Resíduos

Macromoléculas relevantes para os processos biológicos, como proteínas e fitas de DNA e RNA, apresentam uma estrutura eletrônica peculiar. Isso implica que essas estruturas tenham os níveis de energia de fronteira com valores bem próximos, dificultando a explicação da sua reatividade somente pela análises dos orbitais HOMO e LUMO[22].

Além disso, por serem estruturas moleculares com muitos átomos, a perturbação mínima no número de elétrons é bem menos sensível[23]. Logo, é possível chegar a conclusão que o cálculo dos descritores apresenta dificuldades significativas para esses sistemas.

Inspirado em soluções para sistemas de moléculas com alto nível de degenerescência, foi desenvolvido e implementado no PRIMoRDiA modificações para as funções de Fukui a esquerda e a direita para a aproximação de Orbitais Congelados, e logo essas modificações refletem na definição de todas as outros descritores locais.

A primeira modificação é a simples combinação de um certo número de orbitais moleculares ocupados para  $f^-(r)$  a partir do HOMO (Equação 46), e de orbitais moleculares virtuais a partir do LUMO para  $f^+(r)$  (Equação 47). Esse método de cálculo foi batizado de "Band Density"(BD), pois se assemelha ao cálculo de densidade eletrônica, mas só para a banda de orbitais escolhida.

$$f_{BD}^-(r) = \frac{1}{b} \sum_{i=b}^{HOMO} |\psi_i(r)|^2 \quad (46)$$

$$f_{BD}^+(r) = \frac{1}{b} \sum_{i=LUMO}^b |\psi_i(r)|^2 \quad (47)$$

A segunda modificação é uma combinação desses orbitais dando pesos diferentes aos orbitais escolhidos baseado na diferença de energia do orbital em questão com o HOMO no caso de  $f^-(r)$

(Equação 48), e para o LUMO no caso de  $f^+(r)$  (Equação 49). Essa modificação foi batizada de "Energy Weighted"(EW), e é mais flexível para exploração de reatividade, sendo que os coeficientes de orbitais com energia muito diferente não interferem de forma significativa.

$$f_{EW}^-(r) = \sum_{i=b}^{HOMO} e^{-|E_i - E_{HOMO}|} |\psi_i(r)|^2 \quad (48)$$

$$f_{EW}^+(r) = \sum_{i=LUMO}^b e^{-|E_i - E_{LUMO}|} |\psi_i(r)|^2 \quad (49)$$

Esses mesmos descritores são implementados para a versão condensada, assinalando os valores para cada átomo do sistema. Isso é relevante especialmente para os biopolímeros, como no caso de proteínas, pois sítios são definidos por conjunto de resíduos de amino-ácidos. Esses sítios podem ser de catálise, um canal de íons, um dobradiça de movimento e etc. Portanto, é de grande interesse no estudo desses sistemas a representação de propriedades calculáveis para esses resíduos.

O PRIMoRDiA tem implementado todos os descritores em versão condensada para resíduos de biopolímeros, desde que a opção de cálculo dos métodos modificados BD ou EW e seja fornecido um arquivo do tipo PDB. A forma genérica desse processo de condensação é dada de forma bem simples, como mostrado na Equação 50, somando os valores dos descritor  $\Gamma$  dos átomos  $k$  pertencentes ao resíduo  $R$ .

$$\Gamma(R) = \sum_{k \in R}^{atoms} \Gamma(k) \quad (50)$$

Detalhes do uso e aplicação desses descritores modificados são mostrados e discutidos na próxima sessão.



## 5 Descritores Implementados

Devido a atualizações importantes na versão estável 1.25v, alguns descritores locais foram modificados, com processos de normalização, outros adicionados e mudanças de nomenclatura. Essa parte do guia do usuário serve para tabelar os descritores que podem ser obtidos com PRIMoRDiA, o básico de sua interpretação, nome no arquivo e etc.

Também há detalhes sobre representação dos descritores locais, onde há algumas propriedades eletrônicas e descritores que só são relevantes para uma representação volumétrica e portanto escritos em arquivos do tipo Cube (extensão '.cube'). Como por exemplo o potencial eletrostático molecular ou densidade eletrônica dos estados de carga catiônico e aniônico utilizados para os cálculos de diferenças finitas. Por outro lado, existem propriedades que só fazem sentido quando associadas aos centros atômicos, como carga parcial.

Na primeira versão, os métodos de dureza local são calculados individualmente para a representação volumétrica dependendo de qual keyword utilizada, enquanto que para a representação condensada todos os métodos eram calculados e escritos no arquivo de saída. Já para a versão 1.25, todos os métodos possíveis de dureza local vão ser produzidos se as condições do cálculo forem supridas, como por exemplo métodos de dureza local que necessitam da avaliação da densidade eletrônica total.

Os descritores globais seguem os mesmos, mas com algumas atualizações nos seus nomes no arquivo de saída. Na Tabela 2, estão citados todas as propriedades eletrônicas e descritores de reatividade nos arquivos de saída do PRIMoRDiA. Na coluna de nomeada como "nome do arquivo", estão indicados como o descritor fica identificado no arquivo de saída do PRIMoRDiA. Esses nomes já estão atualizados para a versão 1.25, visto que não tem mais sentido utilizar qualquer versão anterior a essa. Na coluna de observação é dado detalhes sobre a obtenção desses descritores e sua interpretação. Como é possível observar, alguns descritores são "crus", extraídos diretamente do processo de leitura do arquivo de saída de pacotes de química computacional, como energias eletrônicas e calor de formação, essa última sendo facilmente obtida nos cálculos com método semi-empírico no pacote MOPAC.

Tabela 2: Descritores globais calculados/extraídos pelo PRIMoRDiA

Descritor	Nome no arquivo	Definição Matemática	Observação
Energia eletrônica	T_ENERGY	-	Extraído do output do cálculo de estrutura eletrônica do estado de referência
Energia eletrônica do cátion	ENERGY_CAT	-	Extraído do output do cálculo de estrutura eletrônica do estado de carregado positivamente
Energia eletrônica do anion	ENERGY_AN	-	Extraído do output do cálculo de estrutura eletrônica do estado de carregado negativamente
Calor de Formação	HOF	-	Variação de entalpia de formação do sistema calculado por método do MOPAC. Extraído do output.
Potencial de Ionização	IP	$E_{N-1} - E_N$	Energia necessária para extrair um elétron. Negativo do HOMO na aproximação de orbitais congelados.
Afinidade Eletrônica	EA	$E_N - E_{N+1}$	energia liberada na absorção de um elétron. Negativo do LUMO na aproximação de orbitais congelados.
Potencial Químico Eletrônico ( $\mu$ )	ECP	$\left(\frac{\partial E}{\partial N}\right)_\nu$	Tendência do sistema de doar elétrons negativo da eletronegatividade de Muliken
Dureza ( $\eta$ )	hardness	$\left(\frac{\partial^2 E}{\partial N^2}\right)_\nu$	Resistência do sistema em variar o seu número de elétrons. Relacionado com a estabilidade química.
Moleza ( $S$ )	softness	$1/\eta$	Facilidade do sistema em variar o seu número de elétrons. Recíproco da dureza. Relacionado com a polarizabilidade, reatividade química e excitabilidade.
Eletrofilicidade total( $\omega$ )	electrofilicity	$\frac{\mu^2 S}{2}$	Energia de estabilização do sistema na absorção do máximo de elétrons.
Diferença HOMO-LUMO	GAP	$E_{LU} - E_{HO}$	Diferença de energia dos orbitais moleculares fronteira.
Número máximo de elétrons recebíveis	nMax	$-\frac{\omega}{\eta}$	Máximo de elétrons que o sistema pode receber de um doador ideal.

Na Tabela 3 estão listados todos os descritores locais obtidos pelo programa desde a sua primeira versão estável. Na Tabela 4 estão listados os descritores locais na versão condensada disponíveis desde a primeira versão. Na Tabela 5 estão tabelados os novos descritores locais volumétricos e na Tabela 6 os novos descritores locais em representação condensada disponíveis a partir da versão 1.25.

Algo a se notar é que há uma seleção de descritores em relação aqueles que foram mostrados na sessão teórica desse guia de usuário. Como por exemplo, o PRIMoRDiA não calcula todas as versões de moleza local, entendendo que são informações redundantes às funções de Fukui já calculadas, sendo que Moleza local média e moleza local dual já trazem a distribuição da moleza no sistema, tanto mostrando as regiões por tipo de ataque, eletrofílico ou nucleofílico, quanto a reatividade média geral.

Tabela 3: Descritores locais calculados no PRIMoRDiA desde a versão 1.0v e escritos em representação volumétrica, em função da coordenada tridimensional  $r$ .

Descritor	Nome do arquivo cube	Definição Matemática	Observação
Densidade eletrônica ( $\rho(r)$ )	Elec_Dens	$\sum_i^{occ}  \phi(r)_i ^2$	Se requeridos diferenças finitas ou dureza local.
Densidade eletrônica do cátion ( $\rho^{N-1}(r)$ )	Elec_Dens_Cation	$\sum_i^{occ}  \phi(r)_i ^2$	Escrito se foi utilizado diferenças finitas.
Densidade eletrônica do anion ( $\rho^{N+1}(r)$ )	Elec_Dens_Anion	$\sum_i^{occ}  \phi(r)_i ^2$	Escrito se foi utilizado diferenças finitas.
Susceptibilidade a ataque eletrofílico ( $f^-(r)$ )	left_Fukui	$\lim_{\partial N \rightarrow 0^-} \left( \frac{\partial \rho(r)}{\partial N} \right)_\nu$	Identifica as regiões nucleofílicas.
Susceptibilidade a ataque nucleofílico ( $f^+(r)$ )	right_Fukui	$\lim_{\partial N \rightarrow 0^+} \left( \frac{\partial \rho(r)}{\partial N} \right)_\nu$	Identifica as regiões eletrofílicas.
Susceptibilidade a ataque radical ( $f^0(r)$ )	zero_Fukui	$\lim_{\partial N \rightarrow 0} \left( \frac{\partial \rho(r)}{\partial N} \right)_\nu$	Reatividade média. Usado para ataques radicais.
Descritor Dual ( $f^\pm(r)$ )	net_Fukui	$\left( \frac{\partial^2 \rho(r)}{\partial N^2} \right)_\nu$	Identifica a tendência líquida.
Dureza Local ( $h(r)$ )	l_hardness_Vee	$\frac{1}{2N} \int \frac{\rho(r')}{ r - r' } dr'$	Equação de trabalho mais tradicional.
Dureza Local (Potencial Químico) ( $h(r)$ )	l_hardness_lcp	$\left( f^- - \frac{\rho(r)}{N} \right) \frac{\mu}{2N} + \frac{\rho(r)}{N} \eta$	Equação de trabalho que integra para a quantidade global.
Dureza Local (Potencial de Fukui) ( $\nu^-(r)$ )	Fukui_pot_left	$\int \frac{f^-(r')}{ r - r' } dr'$	Equação de trabalho que aproxima a densidade de carga para $f^-$ .
Moleza Dual Local $s^\pm(r)$	softness_dual	$S f^\pm(r)$	-
Hiper Moleza Local $s^{(2)}(r)$	hyper_softness	$S^{(2)} f^\pm(r)$	-
Multifilicidade Local $\omega^\pm(r)$	multiflicity	$\Delta \omega f^\pm(r)$	-

Tabela 4: Descritores locais calculados pelo PRIMoRDiA 1.0v e escritos em representação condensada

Descritor	Nome no arquivo	Definição Matemática	Observação
Densidade eletrônica ( $\rho(k)$ )	Electron_density		Integrado na base do átomo
Carga parcial ( $q(k)$ )	charge	-	Extraído do arquivo de saída do pacote de química computacional
Susceptibilidade a ataque eletrofílico ( $f^-(k)$ )	nucleophilicity	$\lim_{\partial N \rightarrow 0^-} \left( \frac{\partial \rho(k)}{\partial N} \right)_\nu$	-
Susceptibilidade a ataque nucleofílico ( $f^+(k)$ )	electrophilicity	$\lim_{\partial N \rightarrow 0^+} \left( \frac{\partial \rho(k)}{\partial N} \right)_\nu$	-
Susceptibilidade a ataque radical ( $f^0(k)$ )	radicality	$\lim_{\partial N \rightarrow 0} \left( \frac{\partial \rho(k)}{\partial N} \right)_\nu$	-
Descritor dual ( $f^\pm(k)$ )	netphilicity	$\lim_{\partial N \rightarrow 0} \left( \frac{\partial^2 \rho(k)}{\partial N^2} \right)_\nu$	-
Dureza local (potencial elétron-elétron) ( $h(k)^{\nu_{ee}}$ )	hardness_Vee	$\frac{1}{2N} \sum_{\alpha \neq k}^{atoms} \frac{\rho(\alpha)}{ r_k - r_\alpha }$	-
Dureza local (potencial químico) ( $h(k)^{LCP}$ )	hardness_lcp	$\left( f^-(k) - \frac{\rho(k)}{N} \right) \frac{\mu}{2N} + \frac{\rho(k)}{N} \eta$	-
Dureza local (potencial Fukui) ( $\nu_f^-(k)$ )	Fukui_pot_left	$\sum_{\alpha \neq k}^{atoms} \frac{f^-(\alpha)}{ r_k - r_\alpha }$	-
Moleza local dual ( $s^\pm(k)$ )	softness_dual	$S f^\pm(k)$	-
Hiper moleza local ( $s^{(2)}(k)$ )	hiper_softness	$S^{(2)} f^\pm(k)$	-
Multifilicidade local ( $\omega^\pm(k)$ )	multiphilicity	$\Delta \omega f^\pm(k)$	-
Localização de banda de fronteira	fukushima	-	Se os métodos de banda forem requeridos

Tabela 5: Descritores locais adicionados no PRIMoRDiA a partir da versão 1.25 escritos em representação volumétrica

Descritor	Nome do arquivo cube	Definição Matemática	Observação
Localização de banda de fronteira	mo_band	-	-
HOMO ( $\phi(r)^N$ )	HOMO	-	Fases positiva e negativa do orbital
LUMO ( $\phi(r)^{N+1}$ )	LUMO	-	Fases positiva e negativa do orbital
Potencial Fukui nucleofílico ( $\nu_f^+(r)$ )	Fukui_pot_right	$\int \frac{f^+(r')}{ r - r' } dr'$	-
Potencial Fukui radical ( $\nu_f^0(r)$ )	Fukui_pot_zero	$\int \frac{f^0(r')}{ r - r' } dr'$	-
Moleza Local média ( $s^0(r)$ )	softness_avg	$S f^0(r)$	-
Potencial eletrostático molecular ( $\Phi(r)$ )	mep	$\sum_{\alpha} \frac{Z_{\alpha}}{ r_{\alpha} - r' } - \int \frac{\rho(r')}{ r - r' } dr'$	-
Dureza local (TFD) ( $h(r)^{TFD}$ )	l_hardness_tfd	$h(r)_T + h(r)_{EX} + h(r)_{EE}$	Dureza local calculada com funcional completa

Tabela 6: Descritores locais adicionados no PRIMoRDiA a partir da versão 1.25 escritos em representação condensada

Descritor	Nome no arquivo	Definição matemática	Observação
Potencial Fukui nucleofílico ( $\nu_f^+(k)$ )	Fukui_pot_right	$\sum_{\alpha \neq k}^{atoms} \frac{f^+(\alpha)}{ r_k - r_\alpha }$	-
Potencial Fukui radical ( $\nu_f^0(k)$ )	Fukui_pot_zero	$\sum_{\alpha \neq k}^{atoms} \frac{f^0(\alpha)}{ r_k - r_\alpha }$	-
Potencial molecular eletrostático ( $\Phi(r)$ )	mep	$\sum_{\alpha \neq k} \frac{Z_\alpha}{ r_k - R_\alpha } - \int \frac{\rho(r)}{ r - R_\alpha } dr$	-
Moleza local média ( $s^\pm(k)$ )	softness_avg	$Sf^0(k)$	-
Dureza local (TFD) ( $h(k)^{TFD}$ )	hardness_TFD	$h(k)_T + h(k)_{EX} + h(k)_{EE}$	-

## 6 Programas Suportados e Informação Necessária

PRIMoRDiA pode ler e trabalhar com informações de arquivos de saída dos pacotes de química quântica computacional mais populares, que são: GAMESS, GAUSSIAN, MOPAC e ORCA. No entanto, dependendo quais descritores, representações e métodos de cálculo, esses pacotes precisam escrever informações específicas.

Para o cálculo dos descritores quânticos condensados é necessário que a matriz de *overlap*, ou a integral de 1-elétron. Já para o cálculo dos descritores locais é necessário das informações das funções de base. Para que essas informações sejam escritas nas saída dos arquivos do GAMESS, é necessário incluir no arquivo de input a sinalização da opção 'nprint=3' no grupo '\$contrl'. Na seguinte caixa de texto (Listing 1) é exemplificado o input para GAMESS para a molécula de acroleína.

Listing 1: Exemplo de input do Gamess

```
$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
```

Para o software ORCA, essas informação são obtidas incluído no grupo '%output' as opções: 'print[p\_mos] 1' e 'print[p\_overlap] 5'. Isso ativa a escrita das bases atômicas e das integrais de um elétron. O exemplo de input para a molécula acroleína com essas opções foi listada na seguinte caixa de texto (Listing 2)



Listing 2: Exemplo de input para o ORCA

```
! 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
*
```

Para o MOPAC é necessário a utilização da keyword 'aux' para a escrita do arquivo com a extensão do 'aux' onde todas as informações relevantes para o PRIMoRDiA estão. Outras keywords são importantes para a escrita completa dos orbitais moleculares: 'large' e 'allvecs'. Para o cálculo de diferenças finitas com outputs do MOPAC, é importante a keyword 'graphf', para o cálculo da densidade eletrônica total.

Para o GAUSSIAN são necessárias as keywords para ativar a escrita do arquivo fchk e para a escrita da matriz de *overlap* no arquivo de log. Essas keywords são respectivamente '%chk = NomeDoInput.chk' e 'iop(3/33=4) pop=full'. Abaixo, mostramos um exemplo de input para gaussian, no Listing 3.

Listing 3: Exemplo de input para o GAUSSIAN

```
%chk=Fe_DF0_s2_nuevo.chk
%mem=6GB
%nprocs=8
B3LYP/gen empiricaldispersion=gd3
scrf=(solvent=water,smd)
scf=save ginput iop(6/7=3)
iop(3/33=4) pop=full
....
```

## 7 Execução do Programa e Opções Gerais

Depois da instalação do programa no seu sistema, é possível utilizar o programa como mostrado a seguir

```
pc@user:$ /path/to/PRIMoRDiA/primordia -f input -np 4
```

O exemplo acima mostra o comando para rodar o cálculo dos descritores de reatividade a partir do arquivo de input. A flag *-f* indica que o próximo argumento é o arquivo de input. A flag *-np* indica que o próximo argumento é o número de threads máximo utilizado nos cálculos. Esse argumento é opcional e o número padrão de threads é o número máximo disponível na sua máquina. Usando a flag *-help*, um resumo geral das opções de uso do programa e seus argumentos, como exemplificado nas caixas de texto a seguir

```
pc@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
-np [n]  : program runs using n threads

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

As outras opções possíveis aceitas pelo programa estão listadas na Tabela 7, junto com os argumentos necessários (vermelho) e opcionais (azul). As saídas do Pacotes de Química Computacional (PQC) são os arquivos que contêm as informações de estrutura eletrônica requeridas pelo PRIMoRDiA, e para as opções *-ed*, que calcula a densidade eletrônica total e escreve em um arquivo 'cube', e *-mo*, que calcula a amplitude de probabilidade do orbital molecular indicado e escreve no arquivo 'cube'.

A resolução da grade (RS) é o número de pontos de uma dimensão da grade tridimensional, sendo que o total de voxels é RS elevado ao cubo. Quanto maior a resolução maior os detalhes que podem ser visualizados através de representações gráficas dessas propriedades

eletrônicas.

Para algumas opções também é requerido que o PQC seja indicado para a leitura correta do arquivo de saída. As keywords aceitas para os PQCs são: 'mopac', 'orca', 'gaussian', e 'gamess'. O programa é sensível ao tipo de caixa de texto, então é necessário utilizar as keywords exatamente como indicadas em letras minúsculas.

Tabela 7: PRIMoRDiA opções de input e seus argumentos. Em vermelho estão sinalizados os argumentos posicionais e em azul os que podem ser colocados em qualquer ordem depois dos posicionais.

Modo	Flag	Arg. 1	Arg. 2	Arg. 3	Arg. 4	Arg. 5
Descritores de reatividade	-f	input	-	-	-	
Densidade eletrônica	-ed	saída do PQC	resolução da grade	PQC	-	
Orbital molecular	-mo	saída do PQC	número do OM	resolução da grade	PQC	-
Gerador de input	-input	-op x	-grid RS	-p QCP	-bandmethod bm	-band b
Diferenças entre par de cubes	-cubed	arquivo cube 1	arquivo cube 2	-	-	
Diferença entre pares de cubes	-cdiffs	input	-	-	-	
Integral do cube	-int	arquivo de cube	-	-	-	-

<sup>a</sup>

<sup>a</sup> PQC: Pacote de Química Computacional; OM: Orbital Molecular; RS: Resolução da Grade

O programa também tem uma opção de gerar automaticamente um arquivo de input a partir de arquivos de saída de um QCP contidos na pasta de onde o PRIMoRDiA rodar. Edições posteriores pelo usuário geralmente são necessárias pós geração automática do Input, sendo que opção serve para agilizar bastante a montagem de input para um grande volume de entradas. Várias opções podem ser escolhidas no momento de montar esse input, no entanto, parâmetros globais devem ser ajustados no cabeçalho do input. Na próxima sessão do guia do usuário a estrutura do input é explicada e esses detalhes são descritos.

## 8 Opções do Input para o Cálculo dos Descritores

A principal opção de uso do programa é através da flag `'-f'`, que sinaliza o cálculo dos descritores quânticos implementados no PRIMoRDiA para os arquivos sinalizados no input, cujo o nome tem que ser obrigatoriamente fornecido após a flag citada.

Antes da versão 1.2 o arquivo de input era muito simples, como mostrado na caixa de texto Listing 4. A primeira linha era o espaço para todas as opções globais de cálculo serem colocadas e nas linhas seguintes começadas com inteiros 1, 2 ou 3, para sinalizar o cálculo dos descritores para o arquivo de estrutura eletrônica correspondente. É enfatizado que esse modelo de input só funciona na versão 1.0, e portanto já é bastante obsoleta.

Listing 4: Exemplo de input utilizado antes da versão 1.2.

```
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
```

Na próxima caixa de texto Listing 5, é apresentado um exemplo de input para o PRIMoRDiA a partir da versão 1.2. Na primeira linha é mandatório sinalizar o tipo de execução requerida, usando o marcador `"#RT"` seguido de uma das opções válidas: `'normal'`, `'reaction'` e `'trajectory'`. No modo `'normal'` de execução os descritores serão calculados para cada entrada no input de forma independente. Nos outros modos o PRIMoRDiA assume que há uma correlação entre as estruturas que são passadas para o cálculo e análises e agrupamentos de dados especiais são realizados.

Listing 5: Exemplo de input utilizado a partir da versão 1.2.

```
#RT normal
#PR eband 1 extrard pymols dos Rscript
1 acrolein.aux false 0 mopac
1 acrolein_orca.out false 40 orca
2 benzene.out benzene_cat.out benzene_an.out true 40 1 mopac
3 2jof.aux true 40 0 2jof.pdb mopac 32.335 1.325 0.876 10 EW
3 2jof.aux false 0 10 2jof.pdb mopac 0 0 0 0 BD
```

Em linhas começadas com `'#PR'` devem ser colocadas os parâmetros globais do cálculo, isto é, opções de output e critérios de corte energético que podem ser controlados pelo usuário e que valerão para todas as entradas do arquivo de input. Não é necessário colocar todos em uma única linha, a única exigência é que as linhas comecem com `'#PR'`. Também, todas as linhas que comecem com `"#"`, e não sejam as opções padrão do software, são ignoradas e podem ser utilizados pelo usuário para adicionar comentários.

Os parâmetros globais e as keywords válidas nas linhas começadas com "#PR", estão resumidos na lista a baixo:

- ***eband:***

Após essa keyword é esperado um valor de energia em eV para definir o corte de seleção dos orbitais moleculares para o cálculo das funções de Fukui. O valor padrão do software é 1.0, e é indicado utilizar entre 1.0 e 5.0. Essa opção é relevante para inputs que tenham entradas para cálculos com os descritores de banda, usando tanto o método *EW* quanto o *BD*.

- ***extrard***

Essa keyword faz com que descritores locais extras sejam escritos em arquivos ".cube". Por padrão, alguns descritores locais, que são de certa forma redundantes em relação a outros, não são escritores na representação volumétrica de forma a diminuir o número de arquivos gerados pelo PRIMoRDiA. Esses descritores não padrão são: potencial de Fukui a direita; potencial de Fukui zero; moleza local média; hiper moleza local; e multifilicidade. Todos os descritores locais são escritos na representação condensada, sem exceção.

- ***dos:***

Essa keyword sinaliza para o PRIMoRDiA para escrever dados de energia eletrônica dos orbitais moleculares em um arquivo separado e gerar um script para a visualização da densidade de estados, salvando em um arquivo de imagem ".png".

- ***pymols:***

Essa keyword ativa a escrita de scripts para a automatização da visualização dos resultados no software gráfico Pymol.

- ***Rscript:***

Essa Keyword ativa a escrita de scripts para a automatização da produção de gráficos estatísticos para análise de trajetórias e coordenadas de reação.

- ***composite:***

Essa keyword ativa o cálculo e a escrita de descritores de dureza local especiais, onde os valores são ponderados pelo volume molecular (no caso da dureza global), e volume atômico (no caso da dureza local). Esses descritores ainda estão em desenvolvimento, e não foram testados e validados em estudos publicados. No momento estão disponíveis na versão 1.2, para facilitar o seu estudo.

Todas as entradas no input que comecem a linha com o inteiro "1", indicam o cálculo dos descritores quânticos usando a aproximação de orbitais congelados. Nesse modo de cálculo, a entrada deve conter obrigatoriamente 5 argumentos na linha. Vamos pegar a primeira entrada do exemplo mostrado na Listing 5:

- **#1: 1**  
Inteiro sinalizando o modo de cálculo
- **#2 `acrolein.aux`**  
Nome do arquivo na pasta contendo as informações de estrutura eletrônica
- **#3: `false`**  
keyword indicando se os métodos de dureza local que requerem o cálculo da densidade eletrônica devem ser computados e escritos na representação volumétrica. Nesse caso, a keyword indica para não calcular. A keyword que ativa o cálculo é "true".
- **#4: 0**  
Inteiro que define a resolução/granularidade da grade dos arquivos ".cube" para representação volumétrica. O valor 0 desativa o cálculo dos descritores na representação volumétrica. Quanto maior o valor mais demorado o cálculo. Geralmente para Moléculas pequenas 40 é o suficiente para uma boa relação resolução/custo computacional.
- **#5: `mopac`**  
Keyword indicando o pacote de química computacional que o arquivo de saída se originou. Keywords aceitas são: "mopac", "gamess", "orca" e "gaussian".

A terceira entrada mostrada na Listing 5 é um exemplo para cálculo pelo método de diferenças finitas, sinalizado pelo inteiro "2" no começo da linha. A diferença aqui para a primeira e segunda entrada no input é que o nome de mais dois arquivos têm de ser fornecidos, que são os estados carregados positivamente e negativamente. Além disso, tem um argumento a mais, o sétimo na linha, que é o que sinaliza a diferença de carga entre os estados carregados e o de referência. O mínimo de carga é 1, mas para sistemas maiores com grande número de elétrons, essa perturbação pode não ser suficiente. Por isso o usuário pode calcular a estrutura eletrônica desses sistemas com perturbações de carga maior, e se isso ocorrer a grandeza da perturbação deve ser indicada ao programa de forma que o mesmo saiba normalizar os descritor de forma apropriada.

Já para o uso dos descritores adaptados para macromoléculas, os descritores de banda através dos métodos *Band Density* e *Energy Weighted*, é necessário inciar a entrada com o inteiro "3", como na quarta e quinta entrada no exemplo mostrada na Listing 5. Já nesse modo, a vários outros argumentos necessários explicados abaixo:

- "3": Indica a opção de cálculo dos descritores para o PRIMoRDiA
- "2jof.aux": Nome do arquivo de output que contém a informação de estrutura eletrônica
- "false": Keyword para opção de dureza local para ser calculada na versão volumétrica

- "0": Granularidade dos arquivos de cube a serem gerados para os descritores volumétricos, zero indica que esses cálculos de grid não serão gerados.
- "0": Número máximo de orbitais moleculares a serem utilizados pelo método BD
- "2jof.pdb": Nome do PDB com as informações de referência
- "0": Coordenada do eixo x para o centro da caixa a ser gerado para os descritores volumétricos
- "0": Coordenada do eixo y para o centro da caixa a ser gerado para os descritores volumétricos
- "0": Coordenada do eixo z para o centro da caixa a ser gerado para os descritores volumétricos
- "0": Tamanho do lado da caixa a ser gerada a partir do centro dado pelas coordenadas acima
- "BD": Keyword para sinalizar o método de cálculo de descritores para macromoléculas ponderado por energia

PRIMoRDiA fornece a opção de o cálculo dos arquivos cube serem feitos para um espaço delimitado do sistema. Isso por que grandes moléculas, como proteínas, necessitariam de uma granularidade alta para que a resolução fosse apreciável, no entanto esses descritores/propriedades podem ser mais interessantes serem analisada para sítios específicos do sistema. Assim a granularidade pode ser relativamente baixa, diminuindo o custo do cálculo, e ainda sim a região de interesse apresentar uma boa resolução na representação volumétrica gerada nos softwares gráficos.

Para cálculos específicos aos resíduos desses biopolímeros é necessário fornecer um arquivo do tipo PDB, que contém diversas informações dessas estruturas além das suas coordenadas tridimensionais. Também esse arquivo é base para o PRIMoRDiA poder escrever arquivos do mesmo tipo com informações de descritores calculadas e escritas nele. Isso permite uma visualização dessas propriedades em softwares gráficos por paleta de cores na colorização dos átomos.

## 8.1 Modos de Execução: 'Reaction' e 'Trajectory'

O acompanhamento das propriedades calculadas pelo PRIMoRDiA pode ser exaustivo para simulações que geram um grande número de estruturas. Muitas vezes o maior interesse na análise desses resultados é para algumas partes específicas do sistema e de forma numérica, como a variação nos átomos que participam de uma reação, ou em resíduos de interesse em estruturas geradas por uma simulação de dinâmica molecular.

Para isso o programa fornece scripts em R para análise estatística desses resultados, desde que no arquivo de input sejam fornecidos dados das coordenadas de reação e/ou resíduos para serem monitorados pelo programa. Outra coisa é na montagem do input, sendo necessário fornecer somente um prefixo do nome dos arquivos de estrutura eletrônica, caso eles estejam organizados com um sufixo mais um número indicando a ordem do ponto na trajetória.

Na caixa de texto Listing 6 é apresentado um exemplo de um input para a análise de uma reação química, oriunda de simulações híbridas QM/MM de uma catálise enzimática. O tipo de execução é sinalizado pela keyword na primeira linha `"#RT"` na variação `"reaction"`. Os parâmetros globais são indicados a seguir. Como o PRIMoRDiA entende que é uma trajetória de reação, os scripts para o pacote R são especiais para análise de coordenada de reação. Nas linhas começadas com a keyword `"Reaction"`, é onde as informações da coordenada de reação e da trajetória são inseridas: `"RC1"` indica que as informações a seguir são para a primeira coordenada de reação; 98, 107, 51 são os índices dos átomos da coordenada de reação; `"dimX"` indica que o próximo inteiro na linha a ser passado da o tamanho da trajetória, o que é importante para o programa saber o nome dos arquivos para ler na pasta. Se os resultados forem de uma simulação de varredura bidimensional, então também será necessário um parâmetro indicando o tamanho dessa segunda coordenada: `"dimY"`.

Listing 6: Exemplo de input para análise de coordenadas de reação.

```
#RT reaction
#PR eband 2 Rscript pymols
#Reaction RC1 98 107 51
#Reaction dimX 15
#TRJ residues 2 4 7
3 sys.aux none 30 10 sys.pdb mopac 0 0 0 0 EW
```

A keyword `"#TRJ"` pode ser tanto utilizada no tipo de execução `'reaction'` quanto no `'trajectory'`, sendo nesse último mandatória. Depois dessa keyword vem `'residues'` indicando que os próximos números inteiros indicam os resíduos no PDB por ordem numérica a serem monitorados pelo programa. Na última linha do input de exemplo, é dada a entrada genérica com os parâmetros que vão ser aplicados para os arquivos na pasta que comecem com `"sys"` e terminem com `".aux"`, e entre eles tenham números inteiros de 0 a 14. Exemplos nos tutoriais.

Para análise somente da trajetória, que é útil para o tratamento de estruturas vindas de simulações de dinâmica molecular, algumas keywords importantes mudam. No Listing 7, mostramos um arquivo de input com parâmetros de exemplo somente para analisar trajetória. Para indicar o tamanho da trajetória é utilizada a



keyword "*frames*", e também há a possibilidade de se indicar o número do frame para o início da lista dos arquivos na pasta, usando a keyword "*start*". Essa última keyword também pode ser útil quando se quer descartar uma parte inicial das estruturas amostradas por simulações de dinâmica molecular.

Listing 7: Exemplo de input para análise de trajetória.

```
#RT trajectory
#PR eband 2 Rscript pymols
#TRJ frames 15 start 1
#TRJ residues 2 4 7
3 sys.aux none 30 10 sys.pdb mopac 0 0 0 0 EW
```

## 9 Considerações Finais

PRIMoRDiA é um programa em constante desenvolvimento por ser novo e estar se consolidando na comunidade científica. Esse guia de usuário é uma tentativa de fornecer todas as informações necessárias para entender da teoria até a execução de todas as funcionalidade do programa e também de se divulgar para todos os cientistas do nosso país.

Quanto a erros de execução, o programa fornece um arquivo de log chamado "primordia.log", que descreve as ações do programa e nomeia os arquivos que estão sendo trabalhados. Isso pode ajudar a identificar em que etapa os erros podem estar se originando. As versões estáveis do programa são exaustivamente testadas e problemas podem se originar no uso de versões intermediárias que fazem parte do desenvolvimento do código.

Outra fonte comum de erros é informação passada de forma equivocada para o programa, tanto na linha de comando como no input. Em situações de erros, o nome do arquivo que se queira abrir pode estar errado ou o mesmo não estar na pasta. O arquivo também pode estar corrompido ou com informações incompletas. Além disso, passar números em campos de input onde se espera texto e vice-versa geralmente provoca erros.

Exemplos do uso e aplicação do programa estão demonstrados nos tutoriais, material que complementa o presente manual. Dúvidas naturalmente surgem, pois o assunto dos descritores e aplicação do programa tem a intenção de transitar entre os universos da química teórica a biologia molecular. Por isso, encorajamos fortemente o contato com o nosso grupo, caso uma dessas dúvidas surja, e nos ajude a melhorar nosso material didático.

## Agradecimentos

We gratefully thanks the Brazilian science/education institutes by the physical structure and financial support: INCT-INAMI, CNPq, CAPES, FAPESQ-PB, PRONEX-FACEPE and FINEP. Universidade Federal da Paraíba (UFPB) and the computer resources of Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP). Research project Bioinformática Estrutural de Proteínas: Modelos, Algoritmos e Aplicações Biotecnológicas (Edital Biologia Computacional 51/2013, processo AUXPE1375/2014 da CAPES). G.B.R. acknowledges support from the Brazilian National Council for Scientific and Technological Development (CNPq grant no. 309761/2017-4).

## Referências

- [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] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] R. G. Pearson, “Hard and Soft Acids and Bases,” *J. Am. Chem. Soc.*, vol. 83, no. 22, pp. 3533–3339, 1963.
- [8] R. G. Pearson, “Hard and soft acids and bases-the evolution of a chemical concept,” *Coord. Chem. Rev.*, vol. 100, no. C, pp. 403–425, 1990.
- [9] W. Yang, R. G. Parr, and R. Pucci, “Electron density, kohn–sham frontier orbitals, and fukui functions,” *The Journal of Chemical Physics*, vol. 81, no. 6, pp. 2862–2863, 1984.
- [10] R. Manne and T. Åberg, “Koopmans’ theorem for inner-shell ionization,” *Chemical Physics Letters*, vol. 7, no. 2, pp. 282–284, 1970.
- [11] K. Fukui, “Theory of orientation and stereoselection,” in *Orientation and Stereoselection*, pp. 1–85, Springer, 1970.
- [12] 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.
- [13] C. Lee, W. Yang, and R. G. Parr, “Local softness and chemical reactivity in the molecules CO, SCN- and H<sub>2</sub>CO,” *J. Mol. Struc. THEOCHEM*, vol. 163, no. C, pp. 305–313, 1988.

- [14] C. Sandoval-Yañez, C. Mascayano, and J. I. Martínez-Araya, "A theoretical assessment of antioxidant capacity of flavonoids by means of local hyper-softness," *Arabian journal of chemistry*, vol. 11, no. 4, pp. 554–563, 2018.
- [15] 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.
- [16] 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.
- [17] R. K. Roy, S. Krishnamurti, P. Geerlings, and S. Pal, "Local Softness and Hardness Based Reactivity Descriptors for Predicting Intra- and Intermolecular Reactivity Sequences: Carbonyl Compounds," *J. Phys. Chem. A*, vol. 5639, no. 97, pp. 3746–3755, 1998.
- [18] G. Klopman, "Chemical reactivity and the concept of charge- and frontier-controlled reactions," *J. Am. Chem. Soc.*, vol. 90, no. 2, pp. 223–234, 1968.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] J. Khandogin, K. Musier-Forsyth, and D. M. York, "Insights into the regioselectivity and rna-binding affinity of hiv-1 nucleocapsid protein from linear-scaling quantum methods," *Journal of molecular biology*, vol. 330, no. 5, pp. 993–1004, 2003.