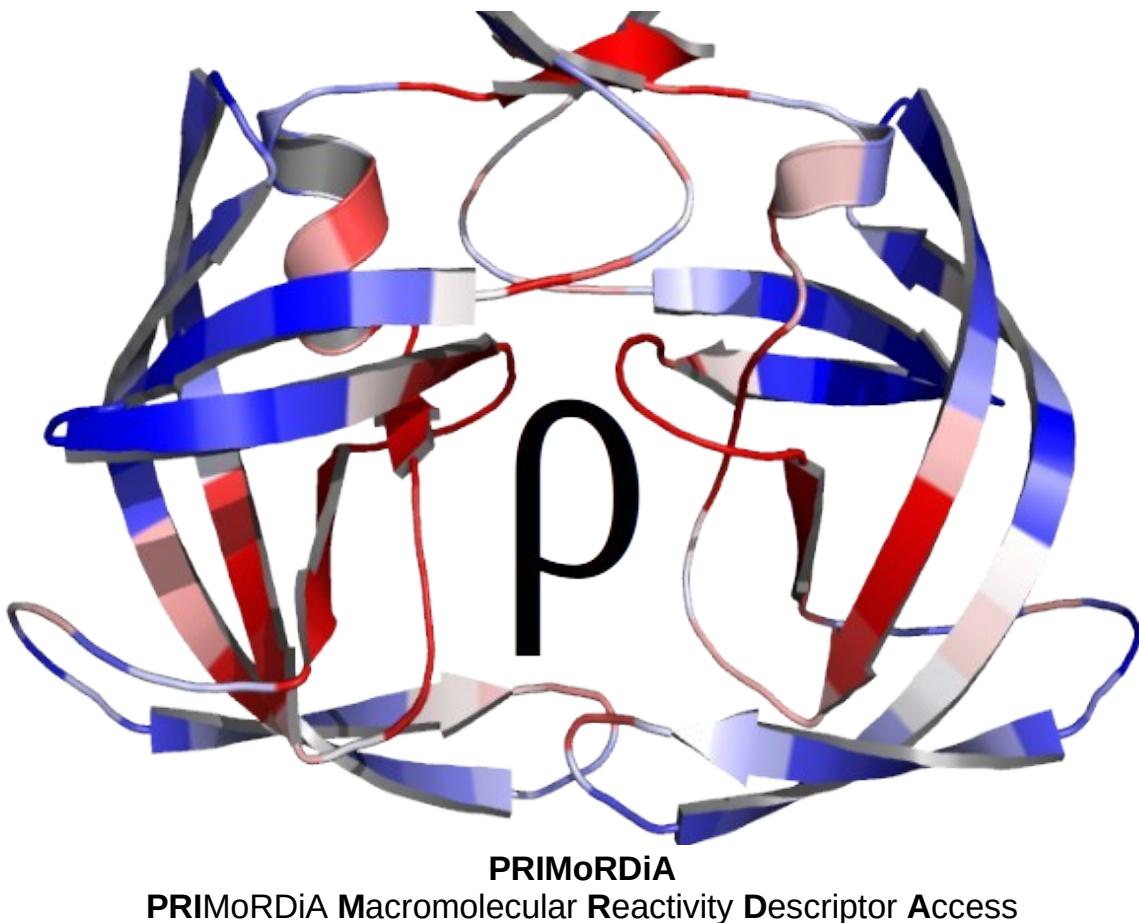

PRIMoRDiA Tutoriais

Igor Barden Grillo
(PRIMoRDiA.software@gmail.com)
github.com/igorChem

29 de janeiro de 2023



Sumário

1 Tutorial 1: Aproximação de Orbitais Congelados	6
1.1 Contextualização	6
1.2 Preparação dos Arquivos e Execução do Programa .	7
1.3 Análises dos Resultados	9
1.3.1 Descritores Locais Condensados	11
1.3.2 Descritores Locais Volumétricos	12
2 Tutorial 2: Aproximação de Diferenças Finitas	18
2.1 Contextualização	18
2.2 Preparação dos Arquivos e Execução do Programa .	18
2.3 Análises dos Resultados	19
3 Tutorial 3: Descritores de Banda	26
3.1 Contextualização	26
3.2 Preparação dos Arquivos e Execução do Programa .	26
3.3 Análises dos Resultados	29
3.3.1 Descritores Condensados	30
3.3.2 Representação Volumétrica	35
3.4 Representação Especial Para Resíduos e Densidade de Estados	41
4 Tutorial 4: Análise de Força Ácida	44
4.1 Contextualização	44
4.2 Arquivos Necessários e Execução	44
4.3 Análise dos Resultados	45
5 Tutorial 5: Análise de Reação Enzimática	52
5.1 Contextualização	52
5.2 Arquivos necessários e Execução	53
5.3 Análise dos Resultados	55
6 Tutorial Extra: Google Colabs	66

Introdução

Esse documento contém os tutoriais para o programa PRIMoRDiA. PRIMoRDiA (**P**RIMoRDiA **M**acromolecular **R**eactivity **D**escriptors **A**ccess) é um software escrito em *C++*, desenvolvido para o pós-processamento eficiente dos resultados de pacotes de química computacional, produzindo uma grande variedade de descriptores quânticos: propriedades eletrônicas e reatividade de sistemas químicos.

O PRIMoRDiA traz implementado os principais descriptores 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 de descriptores e outras propriedades eletrônicas mais comuns para grandes moléculas que são relevantes para processos biológicos, e por isso há descriptores modificados especificativamente para atender as particularidades desses sistemas.

Também oferecemos um guia de usuário com informações gerais sobre a teoria e detalhes sobre o uso e interpretação dos resultados gerados pelo programa. PRIMoRDiA foi desenvolvido na Universidade Federal da Paraíba, no Laboratório de Química Quântica e Computacional.

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

```
computer@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:

```
computer@user:$ git pull
```

Também, no próprio site é possível baixar um '.zip' contendo todos os dados do repositório. No entanto, é fortemente recomendado 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>

Informações mais específicas sobre compilação, instalação e uso dos binários já compilados, consulte o guia de usuário fornecido no repositório.

Sobre os tutoriais nesse documento, eles cobrem praticamente todas as funcionalidades do programa, mas sobretudo o que fazer com os resultados do PRIMoRDiA, já que a montagem do input e execução normalmente só necessitam de dois comandos em um terminal. O PRIMoRDiA trabalha com três modos de cálculo dos descritores, classificados pelo tipo de aproximação realizada. Os três primeiros tutoriais tratam exatamente desses três modos específicos de obter os descritores, tutoriais 4 e 5 trabalham aplicações em possíveis objetos de estudo e o tutorial 6 ensina como executar o programa na plataforma de computação em nuvem do Google para suprir a falta de uma versão do programa para Windows ou para usuários com poucos recursos computacionais.

A lista de tutoriais está logo abaixo:

1. Aproximação de Orbitais Congelados:
2. Aproximação de Diferenças Finitas:
3. Descritores de Banda
4. Análise de Força Ácida
5. Análise de Reação Enzimática
6. Uso do Google Colabs

O primeiro tutorial é muito importante de ser feito, pois ele introduz várias informações sobre a montagem do input e execuções básicas no software Pymol para análise dos resultados de forma gráfica. Também, nesse mesmo tutorial, é desenvolvida a interpretação dos descritores globais e o efeito dos métodos de estrutura eletrônica nessas quantidades, informações que são relevantes para todos os métodos e modos de cálculo.

O segundo tem pouca diferença no que diz respeito a montagem do input e execução do programa, e por isso aproveitamos para prover mais detalhes sobre o uso do Pymol. O terceiro é o principal tutorial para quem deseja trabalhar com macromoléculas, como proteínas/enzimas, introduzindo como utilizar os dois métodos implementados para o cálculo de descritores locais e outros vários parâmetros para a análise dos resultados.

O quarto tutorial pretende demonstrar o uso dos descritores dentro do contexto de uma pergunta de pesquisa, trazendo como exemplo a investigação da força ácida de diferentes amino-ácidos, mostrando como os diversos descritores podem ser utilizados e combinados para extrair informações de cálculos quânticos feitos para esses sistemas.

O quinto tutorial traz uma aplicação bastante avançada do software, se assemelhando a análises que fazemos nas nossas recentes publicações. Esse tutorial aplica os descritores de Banda, que são o terceiro tipo de modo de cálculo implementado exclusivamente no PRIMoRDiA para lidar com macromoléculas, analisando uma trajetória que corresponde a um caminho de reação de catálise enzimática.

No último tutorial, replicamos o primeiro na plataforma de computação em nuvem do Google. A ideia é fornecer um exemplo de uso da plataforma em conjunto ao binário já compilado do PRIMoRDiA fornecido no repositório. A partir desse primeiro exemplo todos os outros tutoriais podem ser executados com os recursos da plataforma que fornece gratuitamente 12GB de RAM e 100GB de armazenamento.

1 Tutorial 1: Aproximação de Orbitais Congelados

Nesse tutorial vai ser demonstrado como utilizar o programa para obter os descritores de reatividade usando as informações dos orbitais moleculares de fronteira. Por ser o primeiro tutorial, vai ser exemplificado a obtenção dos descritores para a molécula de acroleína com a estrutura eletrônica calculada com os quatro pacotes de química computacional que é suportado pelo PRIMoRDiA.

1.1 Contextualização

As teorias de reatividade sempre foram muito mais exploradas dentro da química orgânica através do estudo de orbitais moleculares de fronteira, os famigerados HOMO e LUMO. Esse estudo compreende a análise da distribuição especial desses orbitais nas moléculas e seus valores de energia.

E por que esses orbitais são importantes? Primeiro, os orbitais moleculares servem para descrever a densidade de probabilidade de encontrar um elétron no espaço tridimensional; Segundo, durante as reações químicas são os elétrons que são transferidos entre as moléculas, reorganizando as ligações químicas e portanto as geometrias. Esses orbitais de fronteira, como HOMO (Highest Occupied Molecular Orbital) e LUMO (Lowest Unoccupied Molecular Orbital), são os que descrevem a distribuição espacial do elétrons mais propensos a serem transferidos, ou dos níveis de energias virtuais mais propensos a receber elétrons.

Quando as teorias de reatividade foram demonstradas através do tratamento matemático baseado na Teoria de Densidade Funcional (DFT), os orbitais de fronteira se tornaram uma aproximação para funções locais e valores de energia para os descritores de reatividade. Essa é a aproximação de Orbitais Congelados, Frozen Orbital Approximation (FOA) em inglês. Toda a parte teórica dessa aproximação, parte conceitual e derivação dos descritores de reatividade podem ser encontrados no nosso Guia do usuário e nas referências citadas lá.

Essa aproximação tem algumas vantagens importantes, como necessitar de um protocolo de cálculo mais rápido que de diferenças finitas, e a consolidação dos conceitos baseados nos orbitais moleculares. A principal desvantagem é que por algumas vezes os orbitais HOMO e LUMO não são os que resultam na mais acurada representação da reatividade do sistema.

Nesse tutorial, vamos mostrar como calcular os descritores de reatividade/quânticos usando FOA, como gerar e analisar os resultados com o software gráfico Pymol, e a interpretação das tabelas com valores para cada átomo. Também, por ser o primeiro tutorial, vai ser dado mais atenção a detalhes técnicos sobre o input e scripts gerados pelo programa, interpretação de descritores e explicação de

conceitos que serão úteis para os próximos tutoriais.

1.2 Preparação dos Arquivos e Execução do Programa

Basicamente, o que o PRIMoRDiA faz é extrair as propriedades eletrônicas de arquivos de saída (outputs) de pacotes de métodos de química quântica, e com essas informações fazer o tratamento matemático necessário para obter os descritores de reatividade. Esses descritores podem ser globais, representando uma propriedade que vale para todo o sistema molecular considerado no cálculo quântico, ou descritores locais, que por sua vez descrevem tanto uma propriedade para um dado ponto no espaço tridimensional quanto para cada átomo.

Os descritores locais para os átomos são chamados de condensados, onde a parte do resultado da função local calculada para os descritores é atribuída para os átomos individualmente, tornando essa quantidade ligada as coordenadas do núcleo desse átomo.

Para exemplificar o funcionamento e uso do programa, a estrutura eletrônica da molécula de acroleína foi calculada nos quatro pacotes de química quântica que são suportados pelo PRIMoRDiA (até a presente versão), GAMESS, MOPAC, ORCA e Gaussian, cada um com método de cálculo diferente. A representação gráfica do arranjo molecular da acroleína é mostrado logo abaixo na Figura 1.

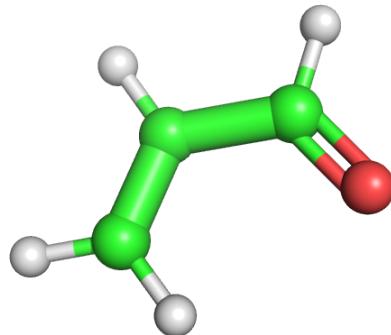


Figura 1: Representação de esferas-e-bastões para a estrutura da molécula de acroleína.

Os arquivos que o PRIMoRDiA precisam estar na pasta comprimida `Tutorials_Files`, que por sua vez tem as pastas para cada tutorial. Além desses arquivos, é necessário um arquivo de texto para indicar ao PRIMoRDiA o que fazer, famoso arquivo de input. Para gerar esse arquivo a partir dos arquivos de estrutura eletrônica que estão na pasta basta rodar o programa com a flag `-input` e os parâmetros específicos para o cálculo que se queira executar. Vamos executar o comando a baixo, dentro do diretório com os arquivos mencionados.

```
computer@user:$  
/path/to/PRIMoRDiA/PRIMoRDiA_1.25v -input -op 1 -p mopac -grid 40
```

Nesse caso, o PRIMoRDiA vai procurar na pasta os arquivos com a extensão ".aux", pois é requerido cálculos de descritores locais para arquivos de saída do mopac. Depois de executar esse comando vai aparecer no diretório dois arquivos novos: primordia.input e primordia.log. Se isso não ocorrer como descrito ou houver alguma mensagem de erro, verifique se o caminho para o executável está correto. Lembrando que é necessário trocar o "/path/to/primordia/" para o endereço da pasta onde o executável está localizado. Abra o arquivo "primordia.input", o resultado deve ser como mostrado na seguinte caixa de texto (Listing 1).

Listing 1: Input gerado pelo comando do PRIMoRDiA.

```
#RT normal  
#PR eband 1 extrard pymols  
1 acrolein.aux true 10 mopac
```

Esse comando é útil para quando há vários arquivos para serem calculados na pasta. Vamos entender as opções utilizadas juntamente com a flag

- **-op:**
Indica a opção de cálculo que vai ser passada no próximo argumento.
- **1:**
Opção de cálculo passada, ela indica que os descritores vão ser obtidos com FOA.
- **-p:**
Indica o programa de origem dos arquivos de saída.
- **mopac:**
É a keyword que indica que os arquivos de saída na pasta vieram do pacote de química quântica MOPAC.
- **-grid:**
flag que serve para indicar a resolução da grade para os resultados dos descritores em representação volumétrica. Quanto maior o valor maior a resolução, mas também maior o custo computacional.
- **40:**
Resolução da grade.

No entanto, há outros arquivos de saída na pasta oriundos de outros pacotes de química computacional, e para incluí-los é necessário edição manual do arquivo para que fique igual ao da seguinte caixa de texto (Listing 2).

Listing 2: Input editado para execução do tutorial.

```
#RT normal
#PR extrard pymols
1 acrolein.aux true 40 mopac
1 acrolein_orca.out true 40 orca
1 acrolein_gam.log true 40 gamess
1 acrolein_gauss.fchk true 40 gaussian
```

A primeira linha do input sempre deve começar com "#RT", que por padrão é seguido pela keyword "normal", que serve para indicar qual o Run Type (tipo de cálculo). Na segunda linha "#PR" indica que os próximos parâmetros vão ser para controlar aspectos gerais dos cálculos.

O parâmetro "pymols" indica que queremos que o programa escreva scripts para automatizar a visualização dos resultados no Pymol. Todas as linhas no input que começarem com "#" não vão ser consideradas, ou seja, é possível escrever comentários começando a linha dessa forma. As linhas que começarem com os inteiros, 1, 2 ou 3, serão consideradas para cálculo utilizando a opção correspondente.

Para executar o PRIMoRDiA é muito simples, a partir do nosso input criado e editado é só rodar o seguinte comando

```
computer@user:$ /path/to/PRIMoRDiA/PRIMoRDiA_1.25v -f primordia.input
```

A flag -f indica que o próximo argumento é um arquivo de input formatado para o programa. O arquivo de input pode ter qualquer nome e não é necessário que tenha a extensão ".input". Depois de rodar esse comando o programa deve indicar o tempo total de execução, e no mesmo diretório onde os dados e o input se encontram, vários novos arquivos de diferentes tipos devem aparecer. Vamos explicar eles a analisar na próxima parte do tutorial.

1.3 Análises dos Resultados

Os descritores de reatividade são calculados como quantidades globais, propriedades de todo o sistema, e locais, que é um mapeamento da reatividade sobre a topologia do sistema molecular. O PRIMoRDiA escreve arquivos com extensão ".GRD" para os descritores globais de cada entrada do sistema, e ".lrd" para os locais condensados e ".cube" para os locais volumétricos. Também escreve um arquivo com o final ".global", onde todos os descritores globais calculados são reunidos.

Se abrirmos esse arquivo, primordia.global, encontraremos dados como mostrado na Figura 2. Esse arquivo já é formatado para ter o espaçamento ideal para ser copiado para um planilha, como mostrado na Figura 3.

GRD	HOMO_E	LUMO_E	T_Energy	Energy_CAT	Energy_AN	IP	EA	ECP	Hardness	Softness	Electrophilicity	GAP	N_MAX	HOF
acrolein	-10.25700000	-0.37600000	-26.34006077	0.00000000	0.00000000	10.25700000	0.37600000	-5.31650000	4.94050000	0.20240866	2.86055786	9.881		
acrolein_orca	-10.72560000	1.49950000	-5192.03767000	0.00000000	0.00000000	10.72560000	-1.49950000	-4.61305000	6.11255000	0.16359784	1.7406999			
acrolein_gam	-10.74306072	1.49390586	-8032.83699098	0.00000000	0.00000000	10.74306072	-1.49390586	-4.62457743	6.11848329	0.16343920	1.74771389			
acrolein_gauss	-7.00216526	-1.85905344	-191.90932743	0.00000000	0.00000000	7.00216526	1.85905344	-4.43060935	2.57155591	0.38886963	3.81681362			

Figura 2: Resultados dos descritores globais para o input trabalhado nesse tutorial.

GRD	HOMO_E	LUMO_E	T_Energy	Energy_CAT	Energy_AN	IP	EA	ECP	Hardness	Softness	Electrophilicity	GAP	N_MAX	HOF
acrolein	-10.257	-0.376	-26.34006077	0	0	10.257	0.376	-5.3165	4.9405	0.20240866	2.86055786	9.881		
acrolein_orca	-10.7256	1.4995	-5192.03767	0	0	10.7256	-1.4995	-4.61305	6.11255	0.16359784	1.7406999			
acrolein_gam	-10.74306072	1.49390586	-8032.836991	0	0	10.74306072	-1.49390586	-4.62457743	6.11848329	0.16343920	1.74771389			
acrolein_gauss	-7.00216526	-1.85905344	-191.9093274	0	0	7.00216526	1.85905344	-4.43060935	2.57155591	0.38886963	3.81681362			

Figura 3: Resultados dos descritores globais transpostos para uma planilha.

Para o Potencial de Ionização (IP), os valores ficaram em torno de 10 eV, com exceção dos resultados calculados a partir da saída GAUSSIAN, que utilizou o método DFT e apresentou 7 eV. Este valor é a energia necessária para extrair um elétron do sistema e é a base para outros descritores de reatividade junto com a afinidade eletrônica (EA), que é a mudança de energia quando o sistema recebe um elétron.

O Potencial Químico Eletrônico (ECP) é o descritor do DFT conceitual que mede a propensão de um sistema molecular para doar elétrons, o que é muito semelhante entre os métodos de *ab initio* e DFT usados aqui para a acroleína, mostrando a molécula mais propensa a doar elétrons quando os descritores provêm do cálculo semi-empírico.

O ECP frequentemente é um valor negativo que vem da derivada da energia eletrônica em relação ao número de elétrons, ou seja, os sistemas requerem energia para doar elétrons. Assim, interpretamos esses resultados como os menos negativos, pois requerem menos energia para doar elétrons e, assim, reagir por meio do processo de transferência de carga.

A dureza é a segunda derivada da energia eletrônica em relação ao número de elétrons e portanto a derivada do ECP em relação ao número de elétrons. Esses descritores indicam a resistência do sistema em doar elétrons, e um sistema com maior dureza é mais estável quimicamente.

Os maiores valores de dureza calculados para a acroleína foram com os métodos Hartree-Fock e o menor com DFT. Assim, podemos concluir que o método DFT indicou a acroleína com maior propensão em doar seus elétrons e participar de uma reação química através desse processo. A dureza sendo a derivada do ECP indica como ela mudará conforme a densidade de elétrons se move do sistema. Assim, um sistema pode ter um alto ECP, mas a alta dureza tornará a transferência da densidade do elétron mais difícil.

Energia de estabilização de um sistema para receber todos os elé-

trons possíveis é medida pela eletrofilicidade total. O número máximo correspondente de elétrons que o sistema pode receber é dado pelo descriptor nMax e, finalmente, a diferença entre os orbitais moleculares da fronteira, HOMO e LUMO, 'gap' é o último descriptor da lista. Para o método FOA, o gap e a dureza são a mesma quantidade, embora, para o método de diferenças finitas isso não é verdade e as duas informações possam ser relevantes.

1.3.1 Descritores Locais Condensados

Agora, vamos analisar e interpretar os descritores locais, que são quantidades atribuídas a um ponto no espaço tridimensional. Existem duas maneiras comuns de representar os descritores locais para moléculas, que são implementadas no PRIMORDiA. O primeiro é o condensado para átomos, onde os valores são atribuídos a cada átomo e são escritos como uma lista. Todas as entradas do arquivo de input produzem um arquivo com extensão '.lrd', especificamente para métodos FOA (opção 1 no arquivo de input), o sufixo do arquivo 'FOA.lrd'. Portanto, devemos ter quatro desses arquivos em nosso diretório, como podemos verificar usando o comando 'ls' no terminal Linux. Na imagem abaixo (Figura 4), as informações desses arquivos foram transferidas para uma tabela do libreoffice calc.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	acrolein																
2	n	atom	charge	ucleophilicity	Electrophilicity	RAS	Netphilicity	Softness	Hardness_A	Hardness_B	Hardness_C	Hardness_D	Multiphilic	Electrophilic	Fukushima	electron_dens	Softness_dual
3	1	C	0.33294001	0.05813121	0.24245778	0.1502945	0.18452667	0.03042091	5.494981225	0.53949859	0.830690877	0.52727682	0.69356451	0.29764723	4.03881718	0.03730929	
4	2	C	-0.30585	0.14870409	0.18610597	0.16740503	0.03740188	0.03388423	5.3493368	0.47364046	0.42232891	1.45528197	0.10699025	0.5323669	0.31778527	4.41101912	0.00750740
5	3	C	-0.17526001	0.01019188	0.40576901	0.20834444	0.39484913	0.04217072	4.7529305	0.4325497	0.44301328	0.04056396	1.12948879	0.16072573	0.41659135	3.8540127	0.07992089
6	4	H	0.1586	0.0449569	0	0.0219784	-0.00439569	0.00044486	1.03643919	0.48783043	0.44440701	0.04508659	-0.01257413	0	0.00439569	0.8415022	-0.00088973
7	5	H	0.14487	0.00872356	0	0.00436178	-0.00872356	0.00088286	1.05040768	0.43184318	0.31656707	0.08947755	-0.02495425	0	0.00872356	0.85514018	-0.00176572
8	6	H	0.164	0.00251001	0	0.02515001	-0.00251001	0.00025402	1.03093172	0.48845316	0.43045971	0.02574517	-0.00718003	0	0.00251001	0.83602856	-0.00050805
9	7	H	0.1086	0.10010896	0	0.05005448	-0.10010896	0.01013146	1.03447281	0.49526144	0.5067473	0.1268176	-0.28636747	0	0.10010896	0.89140894	-0.02026292
10	8	O	-0.42789	0.7663518	0.1656403	0.46599605	-0.60071151	0.09432164	7.89040052	0.35164999	0.16773862	7.7981897	-1.71837003	0.47382365	0.85218647	6.80067944	-0.12158921
11	acrolein_orca																
12	n	atom	charge	ucleophilicity	Electrophilicity	RAS	Netphilicity	Softness	Hardness_A	Hardness_B	Hardness_C	Hardness_D	Multiphilic	Electrophilic	Fukushima	electron_dens	Softness_dual
13	1	C	-0.070785	0.00856342	0.24890483	0.54868806	0.39879645	0.29978323	0.06524224	8.3736050	0.44605424	3.4921144	0.52183363	0.65110281	0.17759495	5.39937145	0.05979225
14	2	C	-0.008105	0.24890483	0.54868806	0.39879645	0.29978323	0.06524224	8.3736050	0.44605424	3.4921144	0.52183363	0.95510125	0.64918335	5.5735377	0.04904389	
15	3	C	-0.56422698	0.20665308	0.72754597	0.64094952	0.68689288	0.10486756	8.35048071	0.37308623	0.22478756	8.82808963	1.51195954	0.17168054	0.96361895	5.54246155	0.14209992
16	4	H	0.247532	0	1E-08	0	0	0.77831762	0.4617565	0.36954316	2E-08	0	0	0.50932434	0		
17	5	H	0.26056701	0	0	0	0	0	0.78434717	0.41462748	0.33547358	0	0	0	0.50285669	0	
18	6	H	0.28144801	0	1E-08	0	0	0	0.71389428	0.45748622	0.3507261	1E-08	0	0	0.46716262	0	
19	7	H	0.212835	0	1E-08	0	0	0	0.69959303	0.48190146	0.23150457	2E-08	0	0	0.65174309	0	
20	8	O	-0.35926399	0.10140029	0.22133924	0.16136977	0.11993895	0.02639975	15.36692881	0.31593567	0.1829073	1.41947716	0.20877772	0.38528519	0.20468475	0.0192175	
21	acrolein_gam																
22	n	atom	charge	ucleophilicity	Electrophilicity	RAS	Netphilicity	Softness	Hardness_A	Hardness_B	Hardness_C	Hardness_D	Multiphilic	Electrophilic	Fukushima	electron_dens	Softness_dual
23	1	C	-0.073806	0.00609686	0.2812824	0.14361255	0.27503137	0.03247192	8.6915237	0.47368662	0.3389816	0.4854781	0.48067615	0.49133172	0.17446733	5.68444676	0.04495091
24	2	C	-0.107531	0.26555519	0.61913133	0.44234326	0.35357615	0.07229623	8.77926891	0.44390831	0.19666325	3.77779943	0.61719494	0.08206443	0.64130931	5.83886492	0.0577882
25	3	C	-0.54746097	0.20837279	0.64109028	0.38543499	0.10477928	8.33698293	0.3811117	0.23160617	3.84272923	1.51253275	0.187670876	0.95337421	5.52910177	0.141446	
26	4	H	0.27123001	0	1E-08	0	0	0	0.7179924	0.47259819	0.37152194	2E-08	0	0	0.46939686	0	
27	5	H	0.26720201	0	0	0	0	0	0.74361409	0.41969533	0.34082147	0	0	0	0.48614274	0	
28	6	H	0.28829199	0	1E-08	0	0	0	0.68948957	0.48215219	0.35417602	2E-08	0	0	0.45078949	0	
29	7	H	0.22368699	0	1E-08	0	0	0	0.59392891	0.49271314	0.23081399	2E-08	0	0	0.62716783	0	
30	8	O	-0.32143301	0.08810155	0.1764545	0.13227802	0.08835294	0.02161941	15.31295794	0.32614372	0.18843421	1.21010673	0.15441567	0.30839197	0.20208548	0.014424539	0.01444033
31	crolein_gauss																
32	n	atom	charge	ucleophilicity	Electrophilicity	RAS	Netphilicity	Softness	Hardness_A	Hardness_B	Hardness_C	Hardness_D	Multiphilic	Electrophilic	Fukushima	electron_dens	Softness_dual
33	1	C	0.22238933	0.0180785	0.28532501	0.15206643	0.26651716	0.05913402	3.1388650	0.49140908	0.78963154	-0.39873875	1.01724633	0.1990324	0.25463977	4.8983123	0.10364043
34	2	C	-0.12598789	0.22457265	0.14716403	0.28185863	0.35272285	0.41825103	3.50628036	0.41825103	0.42232892	0.50854528	0.56169766	0.30211787	5.56387603	-0.03010186	
35	3	C	-0.29290217	0.01925081	0.40288878	0.2198198	0.4013797	0.08521633	3.06082484	0.43886784	0.64670879	-0.64670879	1.53106887	0.16045463	0.347595	5.15912048	0.15599038
36	4	H	0.1406536	0.01139882	1E-08	0.00569491	-0.01139881	0.0221633	0.417659	0.44310103	0.46600881	0.07916138	-0.04350713	0	0.0089375	0.69697705	-0.00444265
37	5	H	0.16184655	0.0031596	0	0.00065798	-0.00025587	0.01466002	0.36753884	0.439551872	0.32166033	0.00921457	-0.00502778	0	0.00107743	0.64663505	-0.00051174
38	6	H	0.18443182	0.00635514	0	0.00317757	-0.00635514	0.00123566	0.38858684	0.4499626	0.40554109	0.04449973	-0.0245638	0	0.00697536	0.60988491	-0.00247132
39	7	H	0.1056417	0.14217657	0	0.07108229	-0.14217657	0.02744048	0.46551041	0.45037211	0.42904196	0.99554385	-0.54266147	1E-08	0.10597094	0.84657181	-0.05528815
40	8	O	-0.39634469	0.60178937	0.2421949	0.42199124	0.35956494	0.16409993	5.9527261397	0.19254574	0.37635734	-1.37250505	0.9244128	0.72506543	0.3095383	-0.13893537	
41																	

Figura 4: Descritores locais condensados.

Esses valores são mapeados para as coordenadas de cada núcleo atômico, conforme mostrado na figura acima(Figura 4), apresentando o oxigênio da acroleína como o mais reativo para receber um ataque eletrofílico.

Os átomos de carbono da dupla ligação também apresentam valores

significativos de nucleofilicidade. O valor mais alto de eletrofilicidade é do carbono 3. Este tipo de representação fornece um meio de comparação quantitativa, intermolecular e intramolecular.

Na Figura 5, mostramos os valores mapeados para os centros atômicos, para os descritores de eletrofilicidade e nucleofilicidade, calculados com os diferentes pacotes de química computacional e métodos de estrutura eletrônica.

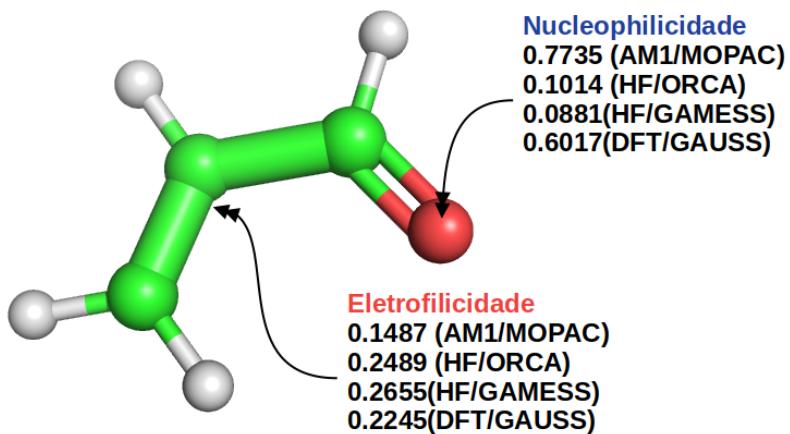


Figura 5: Átomos mais reativos da molécula acroleína demonstrados pelos descritores.

1.3.2 Descritores Locais Volumétricos

Para a representação volumétrica os arquivos Cubes gerados são utilizados para a renderização dos descritores no espaço tridimensional. Nas próximas imagens, mostraremos como gerar esta representação gráfica para os descritores de reatividade mais utilizados através do software Pymol.

Você pode baixar Pymol do repositório (para Linux). Recomendamos fortemente que você faça os tutoriais básicos do Pymol, especialmente se desejar usá-lo para gerar suas figuras científicas, como fizemos várias vezes em nossas publicações.

Os próximos passos consistem em gerar a renderização de volume em Pymol, um objeto que representa graficamente o campo escalar no espaço tridimensional, atribuindo a cada voxel de um mesmo valor a mesma cor e opacidade. A opacidade é um recurso importante, pois é necessário definir pelo menos dois níveis de valores do campo escalar com opacidades distintas para que haja a rendereização de uma volume. Se esses dois ou mais níveis tiverem a mesma opacidade, apenas o mais próximo das bordas da grade aparecerá.

O PRIMoRDia escreve scripts para automatizar a visualização desses arquivos no Pymol, vamos executar um deles para a acroleína

calculada com semi-empírico. Abra o Pymol e digite o comando que está em destaque na figura abaixo (Figura 6).

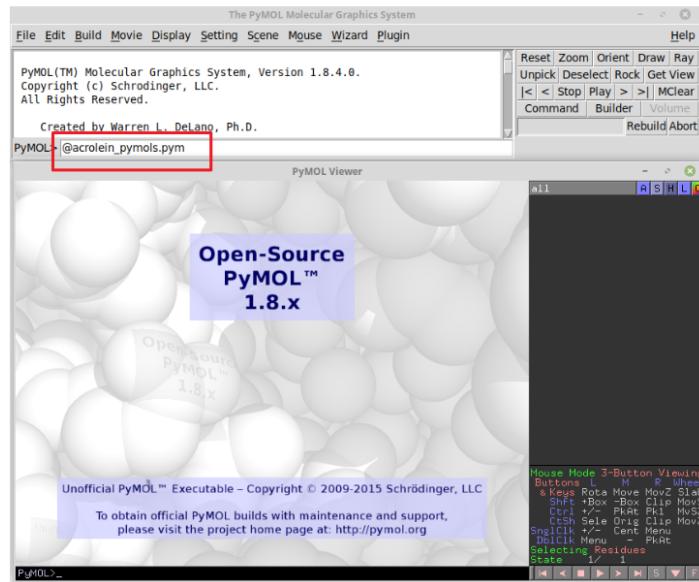


Figura 6: Janela do Pymol e comando para executar dentro do programa para visualizar os descritores na representação volumétrica.

O comando em destaque na caixa vermelha na Figura 6 está repetido abaixo

```
@acrolein_pymols.pym
```

Depois da execução desse comando, todos os arquivos cube vão ser carregados, como mostrado na figura abaixo (Figura 7).

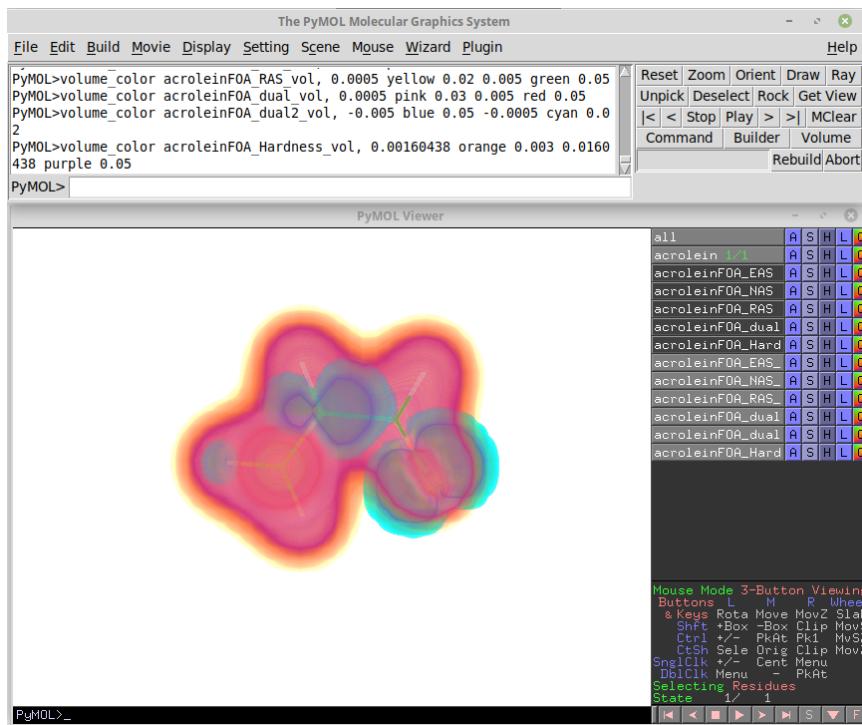


Figura 7: Visualização dos descritores volumétricos gerada pelo script produzido pelo PRIMoR-DiA.

Selecionando o objeto acroleinFOA_EAS_vol na janela lateral do Pymol, mostraremos somente o descritor de nucleofilicidade. Usando o comando no Pymol

```
draw 1000,800,antialias=2
```

Se renderiza uma imagem com resolução 1000X800 pixels, e com parâmetro para aumentar a qualidade da imagem. seguido do comando de

```
png acrolein_eas.png
```

salva a imagem em um arquivo em formato *png* na pasta, como mostrada abaixo na (Figura 8)

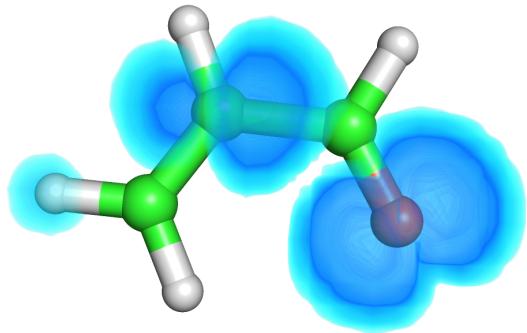


Figura 8: Exemplo de imagem salva para o descriptor de nucleofilicidade para a molécula de acroleína.

O mesmo processo pode ser repetido para os outros objetos de volume. Também é possível usar a interface do Pymol para alterar os valores de superfície, cores e opacidades. Abaixo, a dureza local é objeto desse exemplo da Figura 9.

E na Figura 10, mostramos os descritores para a acroleína considerando todos os métodos de cálculo testados, mostrando os lugares suscetíveis da molécula para receber ataques eletrofílicos (azul) e nucleofílicos (rosa).

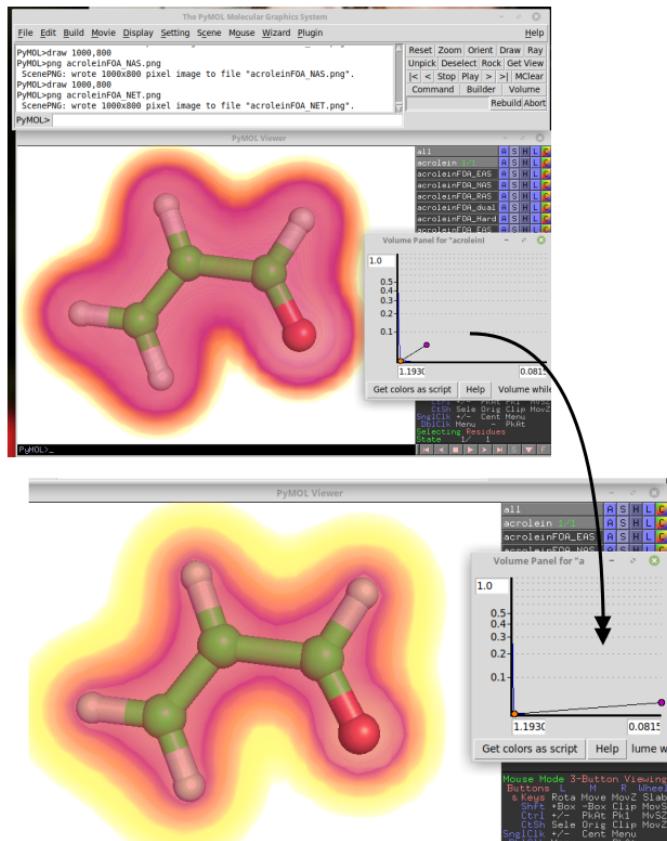


Figura 9: Renderização do descriptor de dureza local.

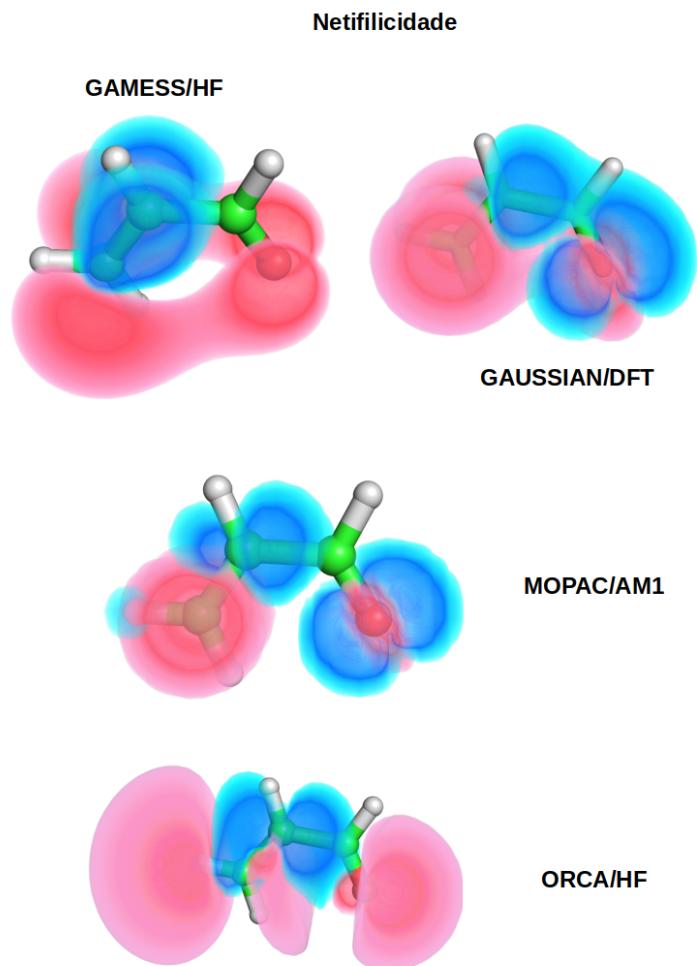


Figura 10: Descriptor de Netfilicidade para a acroleína usando diferentes métodos de estrutura eletrônica trabalhados nesse tutorial.

Portanto, nesse tutorial podemos ver também a influência dos métodos de estrutura eletrônica na obtenção dos descriptores, que é significativa mas tende a ser bem menor que o efeito provocado nos resultados energéticos, fazendo com que na maioria das vezes os descriptores possam resultar nos mesmos padrões mesmo utilizando métodos quânticos menos custosos.

2 Tutorial 2: Aproximação de Diferenças Finitas

Esse segundo tutorial demonstra o cálculo dos descritores no PRIMoRDiA com a aproximação de diferenças finitas, usando a molécula de benzeno como sistema modelo.

Há pouca diferença entre esse tutorial e o primeiro na parte de execução de comandos e ferramentas de análise de resultados. Sobre o que é necessário para o cálculo, o método de diferenças finitas requer dois arquivos extras de estrutura eletrônica, para os estados carregados positivamente e negativamente em relação ao estado eletrônico que se queira calcular as propriedades.

2.1 Contextualização

Em geral, o método de diferenças finitas é uma ferramenta amplamente utilizada para o cálculo de derivadas de forma numérica, ou seja, substituindo um método analítico. Nesse caso mais específico, o cálculo dessas propriedades eletrônicas definidas como derivadas em relação ao número de elétrons, o método de diferenças finitas é um caminho quase intuitivo, já que não é possível derivar uma função não continua.

Por isso o método de diferenças finitas é uma forma tradicional de calcular os descritores e é implementado no PRIMoRDiA. Também, devido ao fato de calcular as propriedades utilizando a informação da diferença das densidades eletrônicas completas, os descritores locais calculados por esse método tendem a ser mais acurados em respeito a efeitos dos orbitais moleculares que não só o HOMO e LUMO. No entanto, a necessidade de dois single points (cálculo de energia) a mais e a obtenção da densidade eletrônica total faz com que esse método acabe se tornando computacionalmente mais custoso.

Nesse tutorial vamos usar a molécula de benzeno como exemplo para os descritores com esse método de aproximação. A molécula de benzeno não é apresenta pontos de reatividade específicos devido a sua simetria, mas é possível visualizar propriedades relacionadas com aromaticidade e a geometria do anel.

2.2 Preparação dos Arquivos e Execução do Programa

Todos os arquivos necessários para realizar o tutorial estão no repositório junto ao código, no arquivo comprimido de nome `Tutorials_Files`. Nesse exemplo vamos utilizar a estrutura eletrônica para a molécula de benzeno em três estados de carga, calculados com o método semi-empírico AM1 no pacote de química computacional MOPAC. Observe que no input para execução desse tutorial, mostrado na caixa de texto Listing 3, os nomes dos arquivos de saída do MOPAC tem a extensão `".out"`.

Listing 3: Input editado para execução do tutorial 2

```
#RT normal
#PR pymols extrard
2 benzene.out benzene_cat.out benzene_an.out true 30 1 mopac
```

Esses arquivos não contém a estrutura eletrônica completa para os sistemas, no entanto o PRIMoRDiA entende que na mesma pasta que esses arquivos estejam também os correspondentes com a extensão ".mgf". Esses arquivos possuem as matrizes necessárias para operações de álgebra linear para obter corretamente a densidade eletrônica total com os dados que saem do MOPAC.

Para grandes moléculas, o MOPAC não consegue gerar esse arquivo, já que ele é muito caro de se escrever para estruturas com muitos átomos. Nesse caso o cálculo de diferenças finitas tem que ser feito usando os arquivos do tipo ".aux", que a partir deles o PRIMoRDiA gera uma densidade eletrônica total baseada nos orbitais moleculares, mas não garante a sua integração exata para o número de elétrons do sistema.

O cálculo dos descritores usando diferenças finitas para grandes moléculas é de fato um tanto incomum, devido ao custo computacional de se obter três estruturas eletrônicas e os cálculos de densidade total para toda a extensão do sistema. Entretanto, com as modificações apropriadas dos descritores usando a aproximação de Orbitais Congelados, FOA, não é necessário recorrer a diferenças finitas para estudar proteínas, fragmentos de DNA e/ou outras moléculas relevantes aos processos biológicos.

Indo para a execução do PRIMoRDiA em si, é bem simples e direto, uma vez com o input e os arquivos necessários na pasta basta rodar o seguinte comando no seu terminal

```
computer@user:$ /path/to/PRIMoRDiA/PRIMoRDiA_1.25v -f primordia.input
```

Exatamente da mesma forma como foi mostrada no tutorial anterior. Depois da execução desse comando, o programa deve finalizar sem nenhum erro e gerar diversos arquivos com os resultados na pasta onde o programa rodou.

2.3 Análises dos Resultados

Como no tutorial anterior, nos vamos abrir os arquivos contendo o resumo dos descritores globais e locais, e finalizar a análise visualizando os descritores locais volumétricos usando o software Pymol.

Na Figura 11, mostramos o arquivo benzene_FD.GRD aberto em um editor de texto, mostrando os valores obtidos para os descritores globais e outras propriedades eletrônicas que são extraídas pelo

PRIMoRDiA. A grande diferença para o primeiro tutorial é a presença dos valores de energia eletrônica para os estados catiônico e aniónico. Além disso, o valor de *gap* aqui, continua sendo a diferença de energia do HOMO e LUMO para a estrutura eletrônica do estado de carga neutro, ou seja, as informações em relação a potencial de ionização e afinidade eletrônica não são redundantes nesse caso.

```

1 Calculus method: Finite differences approximation
2 name: benzene
3 HOMO_E LUMO_E T_Energy Energy_CAT Energy_AN IP EA ECP Hardness Softness Electrophilicity GAP N_MAX HOF
4 -9.80500 0.20390 0.00000 -3237.85948 -3247.47964 -9.46727 -0.15289 -4.81008 4.65719 0.21472 2.48399 10.00890 2.06566 23.20334 |

```

Figura 11: Arquivo de saída do PRIMoRDiA para os descritores globais calculados para a molécula de Benzeno com o método de diferenças finitas.

Avançando para a análise dos descritores locais condensados, o arquivo "benzeneFD.lrd" é mostrado na Figura 12 aberto em um editor de texto. Como já vimos antes, esse arquivo é formatado de forma a ser facilmente copiado e colado em um software de planilhas, como mostrado na Figura 13.

```

1 benzene
2 n atom nucleophilicity electrophilicity radicality netphilicity hardness Vee hardness_lcp Fukui_pot_left Fukui_pot_right Fukui_pot_zero soft
3 1 C -0.051500 -0.060300 -0.055900 -0.008800 0.291247 2.152071 0.566609 0.573317 0.569963 -0.001890 -0.002577 -0.021859 0.000000 -0.147000 2.
4 2 C 0.168800 0.161400 0.165100 -0.007400 0.283744 2.461019 0.393545 0.398093 0.395819 -0.001589 0.007612 -0.018382 0.000000 -0.147000 3.2577
5 3 C 0.210500 0.222500 0.216500 0.012000 0.286553 2.315726 0.360641 0.349851 0.355246 0.002577 0.009982 0.029808 0.000000 -0.147000 3.092126
6 4 C -0.051600 -0.060200 -0.055900 -0.008600 0.291465 2.058173 0.566717 0.573314 0.570016 -0.001847 -0.002577 -0.021362 0.000000 -0.147000 2.
7 5 C 0.168800 0.161200 0.165000 -0.007600 0.284857 1.999467 0.393433 0.398172 0.395802 -0.001632 0.007607 -0.018878 0.000000 -0.147000 2.6631
8 6 C 0.210200 0.222400 0.216300 0.012200 0.287434 1.948558 0.360826 0.349759 0.355293 0.002620 0.009973 0.030305 0.000000 -0.147000 2.618937
9 7 H 0.063700 0.067100 0.065400 0.063400 0.277020 0.327542 0.308489 0.302755 0.305622 0.000730 0.003015 0.008446 0.000000 0.147000 0.454877 -
10 8 H 0.055300 0.056200 0.055750 0.000900 0.287474 0.330961 0.405863 0.401814 0.403839 0.000193 0.002570 0.002236 0.000000 0.147000 0.454944 -
11 9 H 0.053400 0.053200 0.053300 -0.000200 0.284322 0.331713 0.424469 0.429192 0.426830 -0.000043 0.002457 -0.000497 0.000000 0.147000 0.45493
12 10 H 0.063700 0.067200 0.065450 0.003500 0.273633 0.327542 0.308459 0.302787 0.305623 0.000752 0.003018 0.008694 0.000000 0.147000 0.454879
13 11 H 0.055300 0.056200 0.055750 0.000900 0.270863 0.330962 0.405809 0.401694 0.403751 0.000193 0.002570 0.002236 0.000000 0.147000 0.454946
14 12 H 0.053500 0.053200 0.053350 -0.000300 0.271063 0.331704 0.424203 0.428991 0.426597 -0.000064 0.002460 -0.000745 0.000000 0.147000 0.4549
15 |

```

Figura 12: Arquivo de saída do PRIMoRDiA para os descritores locais condensados para os átomos da molécula de benzeno, obtidos com o método de diferenças finitas.

	A	B	C	D	E	F	G	H	I	J	K	L
1	benzene											
2	n	atom	nucleophilicity	electrophilicity	radicality	netphilicity	hardness_Vee	hardness_lcp	Fukui_pot_left	Fukui_pot_right	Fukui_pot_zero	softness_du
3	1	C	-0.0515	-0.0603	-0.0559	-0.0088	0.291247	2.152071	0.566609	0.573317	0.569963	-0.00189
4	2	C	0.1688	0.1614	0.1651	-0.0074	0.283744	2.461019	0.393545	0.398093	0.395819	-0.001589
5	3	C	0.2105	0.2225	0.2165	0.012	0.286553	2.315726	0.360641	0.349851	0.355246	0.002577
6	4	C	-0.0516	-0.0602	-0.0559	-0.0086	0.291465	2.058173	0.566717	0.573314	0.570016	-0.001847
7	5	C	0.1688	0.1612	0.165	-0.0076	0.284857	1.999467	0.393433	0.398172	0.395802	-0.001632
8	6	C	0.2102	0.2224	0.2163	0.0122	0.287434	1.948558	0.360826	0.349759	0.355293	0.00262
9	7	H	0.0637	0.0671	0.0654	0.0034	0.27702	0.327542	0.308489	0.302755	0.305622	0.000752
10	8	H	0.0553	0.0562	0.05575	0.0009	0.287474	0.330961	0.405863	0.401814	0.403839	0.000193
11	9	H	0.0534	0.0532	0.0533	-0.0002	0.284322	0.331713	0.424469	0.429192	0.42683	-4.3E-05
12	10	H	0.0637	0.0672	0.06545	0.0035	0.273633	0.327542	0.308459	0.302787	0.305623	0.000752
13	11	H	0.0553	0.0562	0.05575	0.0009	0.270863	0.330962	0.405809	0.401694	0.403751	0.000193
14	12	H	0.0535	0.0532	0.05335	-0.0003	0.271063	0.331704	0.424203	0.428991	0.426597	-6.4E-05
15												

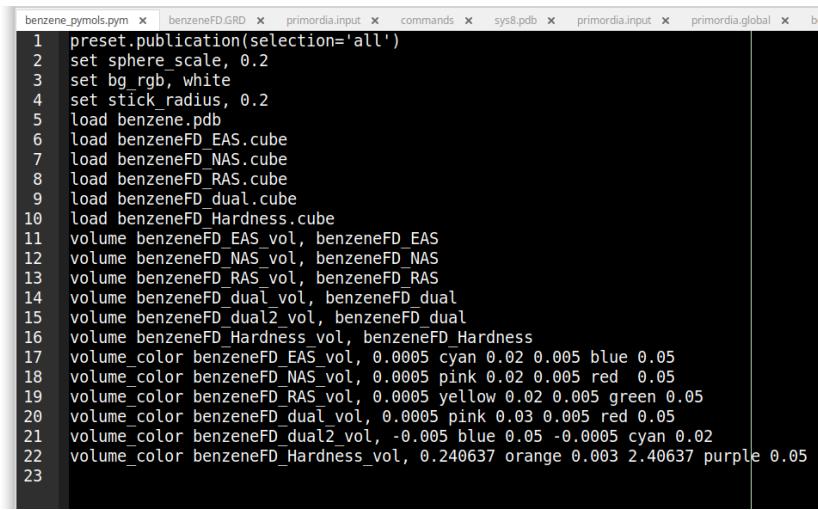
Figura 13: Planilha eletrônica com os dados dos descritores locais condensados para os átomos da molécula de benzeno obtidos com o método de diferenças finitas.

Nós podemos observar a partir dos dados dispostos na Figura 13, os diversos descritores locais obtidos com o cálculo do PRIMoR-DiA, assim como propriedades extraídas do arquivo de saída dos programas de química computacional, como nesse caso a coluna de *charge*, mostrando as cargas parciais para os átomos do sistema de referência. O descritor chamado de *Fukushima* é a localização dos orbitais de fronteira em cada átomo, e portanto não é alterado pelo métodos de diferenças finitas.

Observando os descritores de nucleofilicidade e eletrofilicidade, podemos perceber que há valores negativos associados a alguns átomos. Quando esses valores são de pequena grandeza comparados com os átomos mais reativos não há o que se preocupar quanto a qualidade dos cálculos, pois é uma característica do método de diferenças finitas indicar regiões onde a densidade eletrônica diminui frente a um processo de ganho de carga, ou aumentar frente a um processo de perda de carga. Essas variações podem ocorrer, desde que pequenas, e devem ser interpretadas com cuidado dentro do contexto do processo modelado.

Ainda sobre esses dois descritores, podemos ver que os átomos que apresentam alta nucleofilicidade também apresentam alta eletrofilicidade. Por ser simétrico, o benzeno tende a apresentar padrões de reatividade bem distribuídos nos átomos de carbono e geralmente guiados pela geometria do anel. Portanto a análise desses descritores numéricos por átomo pode acabar se tornando confusa para esse tipo de sistemas, e por isso é mais interessante avaliar os descritores locais na representação volumétrica.

Assim como em todos os tutoriais nesse documento, nós vamos utilizar o software gráfico Pymol, se aproveitando do script escrito pelo PRIMoRDiA, que na pasta deve ter o nome de *benzene_pymols.pym*, que pode ser visualizado aberto em editor de texto na Figura 14.



```

1 preset.publication(selection='all')
2 set sphere_scale, 0.2
3 set bg_rgb, white
4 set stick_radius, 0.2
5 load benzene.pdb
6 load benzeneFD_EAS.cube
7 load benzeneFD_NAS.cube
8 load benzeneFD_RAS.cube
9 load benzeneFD_dual.cube
10 load benzeneFD_Hardness.cube
11 volume benzeneFD_EAS.vol, benzeneFD_EAS
12 volume benzeneFD_NAS.vol, benzeneFD_NAS
13 volume benzeneFD_RAS.vol, benzeneFD_RAS
14 volume benzeneFD_dual.vol, benzeneFD_dual
15 volume benzeneFD_dual2.vol, benzeneFD_dual
16 volume benzeneFD_Hardness.vol, benzeneFD_Hardness
17 volume_color benzeneFD_EAS.vol, 0.0005 cyan 0.02 0.005 blue 0.05
18 volume_color benzeneFD_NAS.vol, 0.0005 pink 0.02 0.005 red 0.05
19 volume_color benzeneFD_RAS.vol, 0.0005 yellow 0.02 0.005 green 0.05
20 volume_color benzeneFD_dual.vol, 0.0005 pink 0.03 0.005 red 0.05
21 volume_color benzeneFD_dual2.vol, -0.005 blue 0.05 -0.0005 cyan 0.02
22 volume_color benzeneFD_Hardness.vol, 0.240637 orange 0.003 2.40637 purple 0.05
23

```

Figura 14: Editor de texto com o arquivo de script escrito pelo PRIMoRDiA para automatizar a visualização dos descritores volumétricos no Pymol.

Através de seu terminal, com o caminho correspondente ao diretório/pasta dos arquivos do tutorial, abra a janela do Pymol, e no seu terminal execute o seguinte comando

`@benzene_pymols.pym`

A partir desse comando, o Pymol vai carregar os arquivos .cube e um de PDB também gerado pelo PRIMoRDiA, todos de uma vez. A janela de ficar como a mostrada na Figura 15.

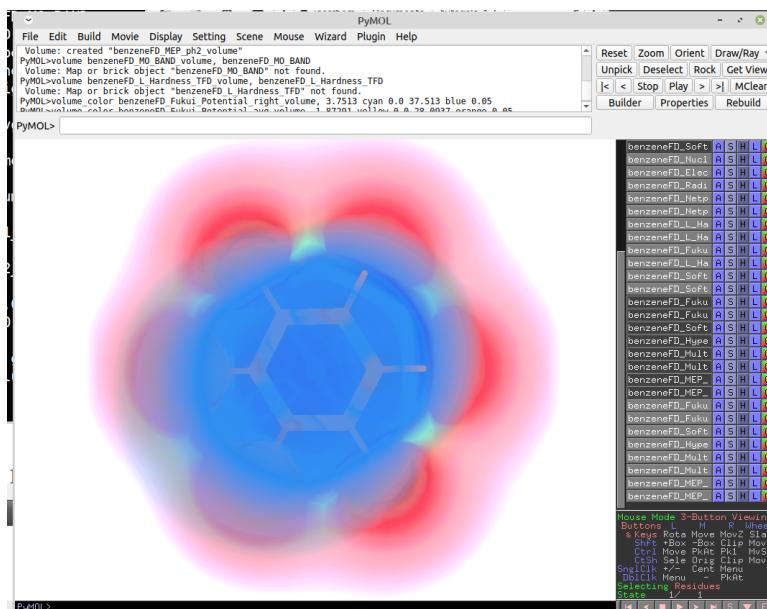


Figura 15: Software Pymol com os dados do tutorial carregados a partir do script gerado no PRIMoRDiA.

Na parte lateral da janela estão listados os objetos carregados, e os que estão mais iluminados são aqueles que estão ativos e aparecendo na tela. Selecione somente um e para salvar a imagem dessa representação em alta resolução execute o seguinte comando

```
draw 3000,2800,antialias=2
```

Depois de alguns segundos, ou minutos dependendo do poder computacional e de placas gráficas disponíveis, essa imagem aparece em destaque em um fundo preto, como na Figura 16. A imagem foi renderizada, para salva-lá definitivamente no disco, use o seguinte comando

```
png benzene_nuc.png
```

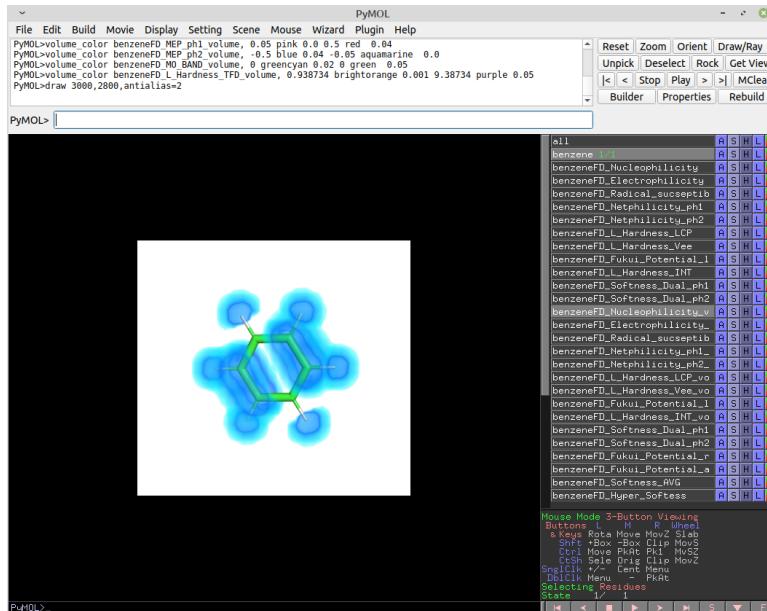


Figura 16: Imagem do descritor volumétrico de nucleofilicidade renderizada no software Pymol.

Podemos notar que esse descritor apresenta uma distribuição bem simétrica para todo a molécula, mas não seguindo o padrão circular, dando preferência a quatro átomos de carbono do anel em detrimento dos outros dois.

Um descritor interessante de se observar é o de dureza química local, que no PRIMoRDiA pode ser obtido com até sete diferentes equações de trabalho, que são completamente explicadas no guia do usuário. O PRIMoRDiA não coloca no script todos os descritores locais possíveis, até para não sobrecarregar o software gráfico, pois dependendo dos recursos computacionais disponíveis o mesmo pode travar.

Também, para todos os descritores, o PRIMoRDiA sugere uma

paleta de cores, então geralmente não há necessidade de ajustar manualmente os valores e opacidades das iso-superfícies, como já foi mostrado em tutoriais antigos. De qualquer forma, é indicada que os tutoriais do Pymol sejam realizados para a produção de imagens mais customizadas.

Voltando para os descritores de dureza local, na Figura 17 nos mostramos a renderização desse descritor calculado com a equação de trabalho que se baseia na aproximação do potencial químico local, abreviada de LCP. Já na figura Figura 18, é mostrada a dureza local estimada com a aproximação da parte de interação inter-eletrônica do potencial eletrostático molecular. Outra propriedade local que representa as interações de tipo duro-duro, aquelas controladas por-carga como explicado no guia do usuário, é o próprio potencial eletrostático molecular mostrado na Figura 19.

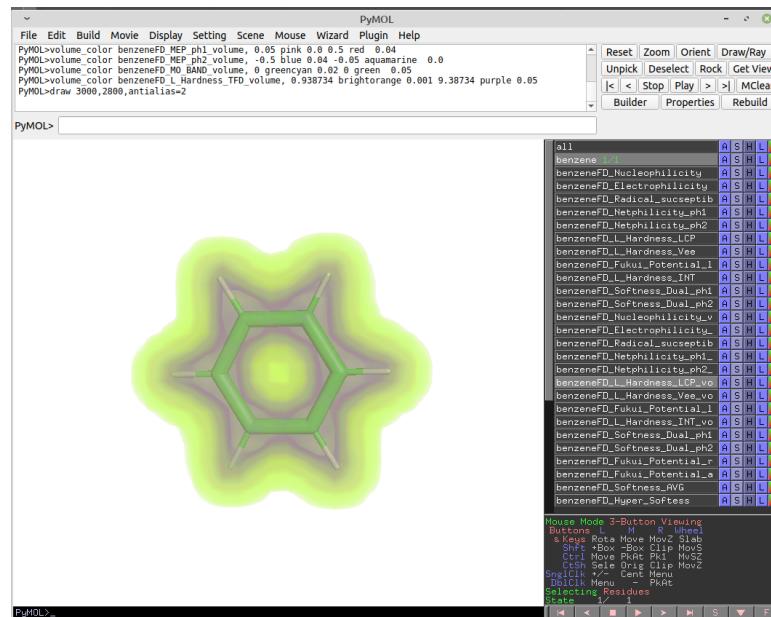


Figura 17: Dureza local usando, a aproximação de potencial químico local, renderizada no Pymol.

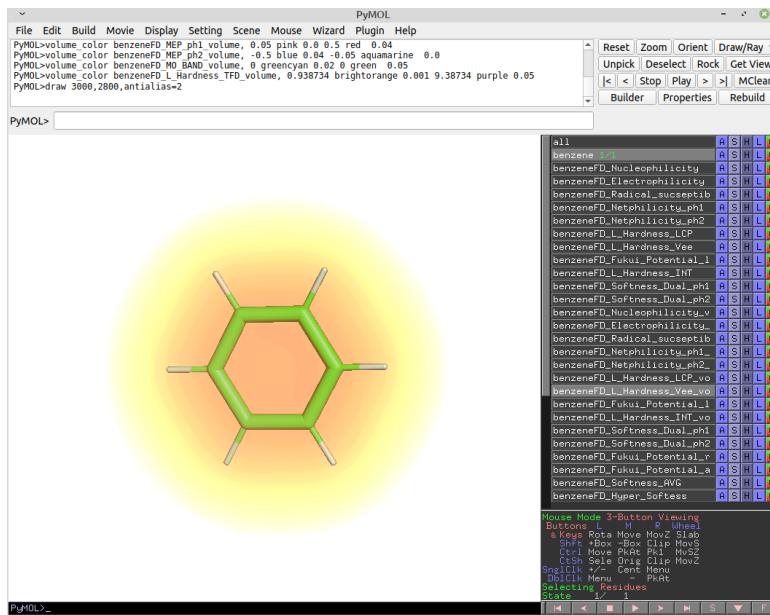


Figura 18: Dureza local usando, a aproximação de interação inter-eletrônica do potencial eletrostático molecular, renderizada no Pymol.

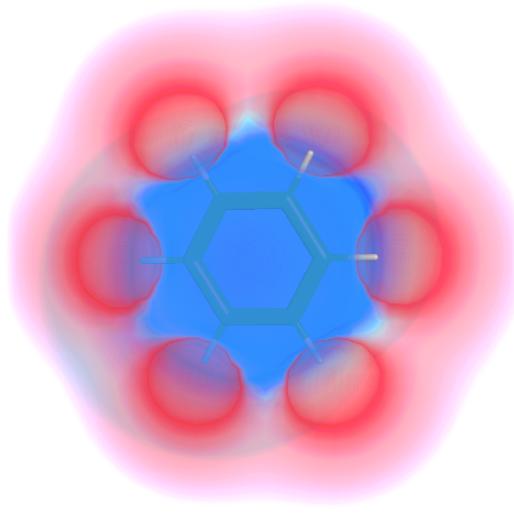


Figura 19: Potencial eletrostático molecular estimado a partir da cargas parciais. Imagem renderizada e salva em png.

A partir desses descritores, podemos ver como as principais interações com a densidade eletrônica surgem a partir do centro do anel aromático.

3 Tutorial 3: Descritores de Banda

Assim como na aproximação de orbitais congelados, nesse tutorial vamos explorar dois métodos de combinação de orbitais de fronteira para o ajuste dos descritores para macromoléculas. Para os cálculos, vamos utilizar um polipeptídeo bastante comum em testes, a TRP-cage, que teve a estrutura eletrônica obtida usando métodos semi-empíricos implementados no MOPAC, incluindo o método de escalonamento linear MOZYME.

Assuntos abordados nesse tutorial:

- Cálculo dos descritores usando nossos métodos de banda dos orbitais de fronteira.
- Uso do pacote gráfico Pymol para visualização e geração de imagens de alta qualidade para os descritores condensados.
- Demonstração dos descritores condensados por resíduos
- Representação volumétrica para esses descritores, em especial para regiões limitadas definidas pelo usuário
- Análise de densidade de estados de energia

3.1 Contextualização

O PRIMoRDiA é um programa que foi desenvolvido desde o início para o cálculo de descritores de reatividade para grandes moléculas com relevância para sistemas biológicos, mais especificamente proteínas e fragmentos de DNA/RNA. Essas moléculas são biopolímeros e possuem uma estrutura eletrônica peculiar, com vários orbitais moleculares com energias similares as do HOMO e LUMO e com distribuição espacial relevante para processos químicos que envolvam esses sistemas.

3.2 Preparação dos Arquivos e Execução do Programa

Os arquivos necessários para realizar o tutorial estão na pasta comprimida `Tutorials_Files` deste repositório. Os arquivos correspondem a outputs programa `mopac` de cálculos feitos com método semi-empírico PM7, um com método de escalonamento linear `mozyme` `2jof_PM7_lmo.aux`, e sem esse método `2jof_PM7.aux`. Para cada um desses arquivos tem um arquivo de PDB, que contém as informações relevantes ao biopolímero que vai ser utilizado pelo PRIMoRDiA para escrever descritores e representações especiais. Esses PDBs são baseados no que pode ser encontrado pelo código `2JOF`. A representação em cartoon desse polipeptídeo se encontra abaixo na Figura 20.

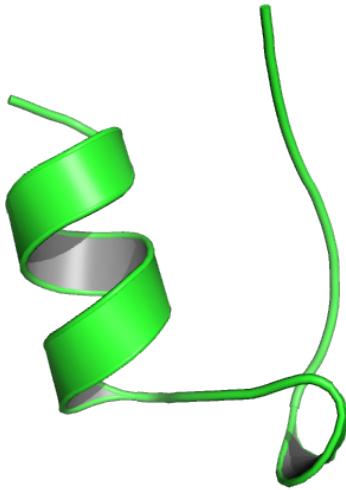


Figura 20: Representação cartoon para o polipeptídeo TRP-cage.

Para facilitar a produção de input a partir desses arquivos vamos usar o seguinte comando

```
computer@user:$ path/to/PRIMoRDiA_1.25v -input -op 3 -p mopac
```

Um arquivo de chamado de primordia.input é gerado na pasta, igual ao mostrado na ???. A primeira keyword encontrada no arquivo é "#RT normal"para indicar que o Run Type (tipo de cálculo) é do tipo comum esperado para a execução do PRIMoRDiA. O keyword "#PR"indica os que as próximas entradas na linha são parâmetros especiais, que no caso, por hora, é "eband"e "1", indicando que o limite para considerar orbitais moleculares em relação HOMO-LUMO é de 1eV. As duas linhas que começam com o inteiro "3", indicam a opção de cálculo do PRIMoRDiA especial para macromoléculas.

Listing 4: Input editado para execução do tutorial 3

```
#RT normal
#PR eband 1 extrard pymols
3 2jof_PM7.aux true 40 0 2jof_PM7.pdb mopac 0 0 0 0 10 EW
3 2jof_PM7_1mo.aux true 40 0 2jof_PM7_1mo.pdb mopac 0 0 0 0 EW
```

Esses argumentos são detalhadamente explicados no guia do usuário, mas também são descritos um por um na lista abaixo

- "3": Indica a opção de cálculo dos descriptores para o PRIMoRDiA
- "2jof.aux": Nome do arquivo de output que contém a informação de estrutura eletrônica

- "true": Keyword para opção de dureza local para ser calculada na versão volumétrica
- "40": 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 ser utilizados pelo método BD
- "2jof_PM7.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
- "EW": Keyword para sinalizar o método de cálculo de descritores para macromoléculas ponderado por energia

Para a produção de arquivos cube, com dados volumétricos, é possível limitar uma área de cálculo para obter uma definição maior com menor custo computacional. Para isso é necessário definir um caixa onde o grid vai ser calculado, onde o centro e o tamanho é dado no arquivo de input, como foi indicado acima.

O PRIMORDiA tem implementado dois métodos para combinar esse orbitais moleculares:

1. Band Density (BD: sigla e Keyword utilizada dentro do programa)
2. Energy weighted (EW: sigla e Keyword utilizada dentro do programa)

O primeiro é inspirado no cálculo de densidade eletrônica, com a particularidade de considerar somente uma seleção de orbitais moleculares a partir do HOMO, para os ocupados, e LUMO para os virtuais. O segundo é uma combinação que é ponderada pela diferença de energia dos orbitais moleculares dos orbitais de fronteira HOMO-LUMO. Há um limite para considerar esses orbitais que é definido por um valor de corte de energia definido pelo usuário.

Vamos editar o arquivo de input para que a primeira entrada rode com o método BD e a segunda rode com uma caixa definida para os descritores volumétricos. Nesse exemplo vamos definir o centro da caixa nas coordenadas do nitrogênio da cadeia lateral do resíduo

TRP_6, informações retiradas do arquivo de PDB mostrado na Figura 21.

72	ATOM	71	HE21	GLN	A	5	21.493	-0.087	-3.479	1.00	20.00
73	ATOM	72	HE22	GLN	A	5	21.487	-0.274	-5.238	1.00	20.00
74	ATOM	73	N	TRP	A	6	26.657	1.026	-0.296	1.00	20.00
75	ATOM	74	CA	TRP	A	6	27.939	0.765	0.370	1.00	20.00
76	ATOM	75	C	TRP	A	6	27.748	0.150	1.763	1.00	20.00
77	ATOM	76	O	TRP	A	6	28.319	-0.901	2.056	1.00	20.00
78	ATOM	77	CB	TRP	A	6	28.769	2.060	0.434	1.00	20.00
79	ATOM	78	CG	TRP	A	6	30.176	1.872	0.902	1.00	20.00
80	ATOM	79	CD1	TRP	A	6	31.263	1.567	0.110	1.00	20.00
81	ATOM	80	CD2	TRP	A	6	30.744	2.042	2.239	1.00	20.00
82	ATOM	81	CE2	TRP	A	6	32.119	1.664	2.192	1.00	20.00
83	ATOM	82	CE3	TRP	A	6	30.248	2.491	3.482	1.00	20.00
84	ATOM	83	NE1	TRP	A	6	32.335	1.325	0.876	1.00	20.00
85	ATOM	84	CZ2	TRP	A	6	32.957	1.710	3.316	1.00	20.00
86	ATOM	85	CZ3	TRP	A	6	31.084	2.567	4.614	1.00	20.00
87	ATOM	86	CH2	TRP	A	6	32.434	2.181	4.532	1.00	20.00
88	ATOM	87	H	TRP	A	6	26.531	1.932	-0.742	1.00	20.00
89	ATOM	88	HA	TRP	A	6	28.488	0.036	-0.224	1.00	20.00
90	ATOM	89	HB2	TRP	A	6	28.799	2.512	-0.558	1.00	20.00
91	ATOM	90	HB3	TRP	A	6	28.292	2.767	1.107	1.00	20.00
92	ATOM	91	HD1	TRP	A	6	31.132	1.354	-0.957	1.00	20.00
93	ATOM	92	HE1	TRP	A	6	33.220	0.985	0.488	1.00	20.00
94	ATOM	93	HE3	TRP	A	6	29.219	2.809	3.550	1.00	20.00
95	ATOM	94	HZ2	TRP	A	6	33.993	1.416	3.234	1.00	20.00

Figura 21: Arquivo de PDB da TRP-cage aberta em editor de texto.

Vamos modificar manualmente o input com esses dados de coordenada e tamanho de caixa 10 angstroms, com o input ficando como mostrado na caixa de texto Listing 5.

Listing 5: Input editado para execução do tutorial 3

```
#RT normal
#PR eband 1 extrard pymols
3 2jof_PM7.aux true 40 0 2jof_PM7.pdb mopac 32.335 1.325 0.876 10 EW
3 2jof_PM7_lmo.aux true 40 10 2jof_PM7_lmo.pdb mopac 0 0 0 BD
```

Finalmente, com esse input podemos colocar o programa para rodar com o seguinte cálculo

```
computer@user:$ path/to/PRIMORDIA_1.25v -f primordia.input
```

O programa deve finalizar a sua execução sem erros e produzir vários arquivos no diretório, com extensão ".cube", ".pym", ".lrd" e etc. Vamos explicar os arquivos com os resultados de descritores e como analisá-los na próxima seção desse tutorial.

3.3 Análises dos Resultados

Dos resultados produzidos temos os descritores globais, locais em representação condensada (arquivos ".lrd"), locais em representação volumétrica (arquivos ".cube") e locais condensados escritos em PDB (nas pastas "RD_PDB"). Para macromoléculas é interessante observar os descritores globais mais por uma questão de acompanhamento da normalidade do cálculo quântico executado, pois os processos nesses sistemas tendem a ser somente locais, tendo

pouca relevância a propensão de uma proteína inteira de transferir ou receber elétrons. Por isso não vamos analisar do arquivo "primordia.global".

Para esse lançamento estável do programa, PRIMoRDiA 1.25v, a principal modificação é a consolidação dos scripts de automatização de análise e produção de resultados com qualidade de publicação, usando o pacote gráfico Pymol e o pacote estatístico R. Por isso, nos próximos passos do tutorial, vamos focar na produção de imagens a partir desses scripts.

3.3.1 Descritores Condensados

Descritores locais são funções matemáticas dependente de coordenadas espaciais, descrevendo as tendências de um certo ponto no espaço molecular. No caso dos descritores condensados, esses valores são particionados para cada centro atômico, simplificando a interpretação dos resultados, e para o caso de grandes sistemas, tornando a visualização mais clara, já que a representação volumétrica para sistemas de proteínas e afins pode ficar bem poluída.

No entanto, esse tipo de representação pode ser bem complicada de trabalhar, já que é uma lista de valores por átomos, ou seja, uma tabela que pode ser significativamente grande para macromoléculas. Pensando em facilitar a análise dessas resultados, o PRIMoRDiA escreve um arquivo de PDB para cada descritor calculado com os valores no campo do b-factor, e assim podemos usar o pacote gráfico Pymol para visualizar os descritores colorindo os átomos com uma paleta customizada. Para operar o Pymol mais automaticamente possível, o PRIMoRDiA escreve um arquivo de script com o sufixo "`_pymols_pdb.pym`", com um exemplo de sua estrutura mostrado na Figura 22.

```

1 |preset.publication(selection='all')
2 set sphere_scale, 0.2
3 set bg_rgb, white
4 set stick_radius, 0.18
5 load 2jof_PM7_PDB_RD/2jof_PM7_nucleophilicity.pdb
6 load 2jof_PM7_PDB_RD/2jof_PM7_electrophilicity.pdb
7 load 2jof_PM7_PDB_RD/2jof_PM7_radicality.pdb
8 load 2jof_PM7_PDB_RD/2jof_PM7_netphilicity.pdb
9 load 2jof_PM7_PDB_RD/2jof_PM7_hardness_Vee.pdb
10 load 2jof_PM7_PDB_RD/2jof_PM7_hardness_lcp.pdb
11 load 2jof_PM7_PDB_RD/2jof_PM7_fukui_pot_left.pdb
12 load 2jof_PM7_PDB_RD/2jof_PM7_fukui_pot_right.pdb
13 load 2jof_PM7_PDB_RD/2jof_PM7_fukui_pot_zero.pdb
14 load 2jof_PM7_PDB_RD/2jof_PM7_softness_dual.pdb
15 load 2jof_PM7_PDB_RD/2jof_PM7_hyper_softness.pdb
16 load 2jof_PM7_PDB_RD/2jof_PM7_fukushima.pdb
17 load 2jof_PM7_PDB_RD/2jof_PM7_mep.pdb
18 load 2jof_PM7_PDB_RD/2jof_PM7_hardness_TFD.pdb
19 load 2jof_PM7_PDB_RD/2jof_PM7_softness_avg.pdb
20 load 2jof_PM7_PDB_RD/2jof_PM7_hardness_int.pdb
21 load 2jof_PM7_PDB_RD/2jof_PM7_multiphilicity.pdb
22 load 2jof_PM7_PDB_RD/2jof_PM7_charge.pdb
23 load 2jof_PM7_PDB_RD/2jof_PM7_electron_density.pdb
24 spectrum b, blue_white_red, minimum=-0.3, maximum=0.3
25 spectrum b, white_yellow_orange_red_black, minimum=0.1, maximum=0.5
26 spectrum b, white_cyan_blue, minimum=0, maximum=0.5
27 spectrum b, white_pink_red, minimum=0, maximum=0.1
28

```

Figura 22: Script de automação da visualização dos descritores locais condensados no Pymol escrita no PRIMoRDiA e aberta em editor de texto.

Esse arquivo tem comandos para executar no Pymol modificando parâmetros de visualização e para carregar os PDBs com os descritores. Além disso ele já executa algumas paletas de cores, sendo que uma vai sobrepondo a outra, mas servem para o usuário pegar e escolher uma ou modificar seus parâmetros e cores. Abra o seu Pymol, execute o comando mostrado a baixo

`@2jof_PM7_pymols_pdb.pym`

Depois da execução desse comando, a janela do pymol deve abrir e carregar todos os arquivos de PDB e executar todos os comandos que configuram a paleta de cores baseada nos valores dos descritores. O último comando de "spectrum b..." executada deve ser o que fica aparecendo, como na Figura 23.

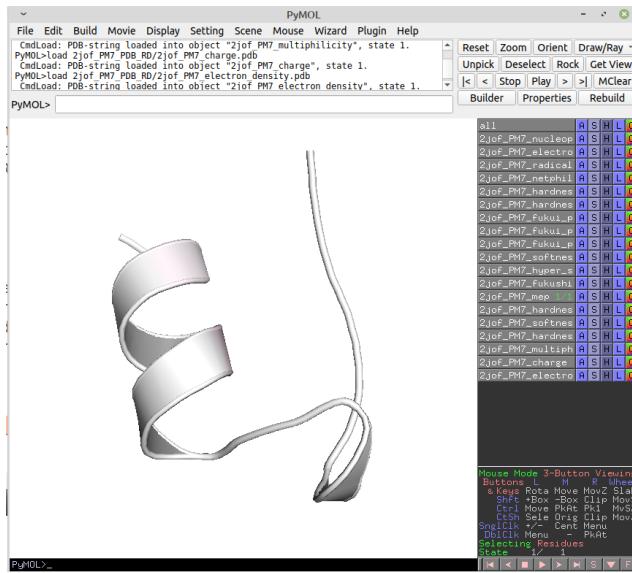


Figura 23: Descritores locais condensados para a estrutura da TRP-cage carregados no Pymol.

Na próxima imagem nós mostramos o Pymol com somente um dos objetos ativos, o 2jof_PM7_nucleophilicity, e mostrando o comando a ser executado no terminal, que foi selecionado/copiado a partir de um daqueles que foi executado quando o script foi chamado.

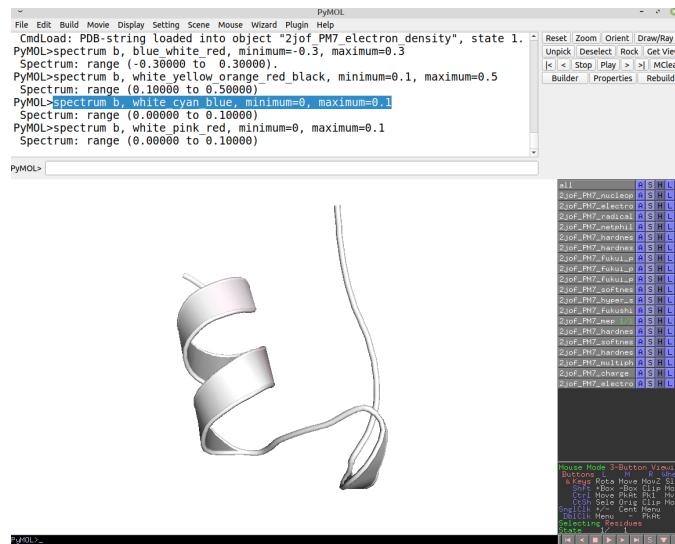


Figura 24: Janela do Pymol mostrando o comando que já tinha sido executado.

Depois de executar esse comando no Pymol o resultado é o mostrado Figura 25, onde os átomos de algumas cadeias laterais estão selecionados, já que somente a estrutura secundária está em evidência, com o próximo comando no Pymol os átomos selecionados aparecem em representação de bastões.

```
show sticks, sele
```

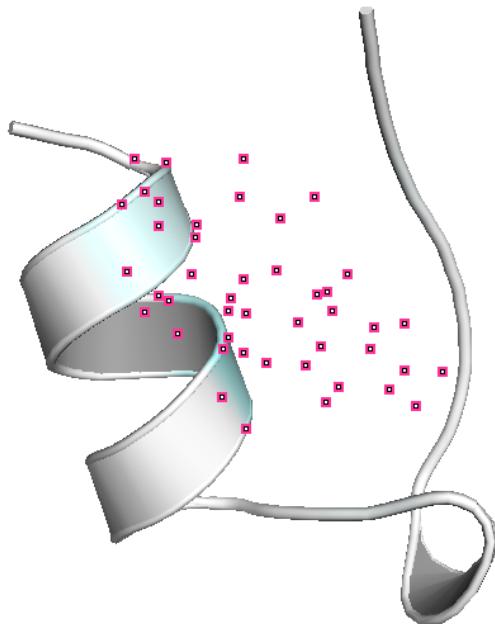


Figura 25: Estrutura no Pymol com duas cadeias laterais selecionadas.

Os átomos e ligações entre os mesmos serão mostrados na representação de sticks (bastões), resultando na figura Figura 26, mostrando a estrutura pintada baseada nos valores de suscetibilidade a um ataque electrofílico, o mesmo que nucleofilicidade. Onde a cor azul está mais escuro é onde os valores são maiores, é possível alterar esses valores de máximo e mínimo bem como as cores também. Na figura abaixo é possível observar que o comando "ray" foi executado no Pymol. Escreva o mesmo e execute, depois de alguns segundos aparece um fundo de transparência, significando que a imagem já foi renderizada e pode ser salva com o comando "png", como já usado para os descritores volumétricos nos tutoriais anteriores.

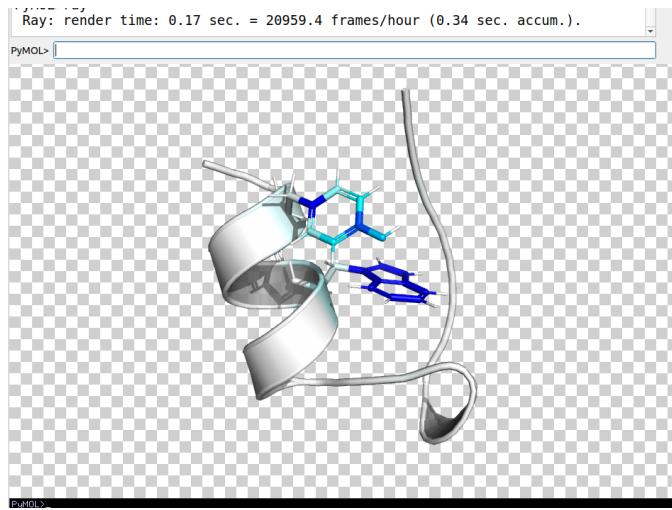


Figura 26: Representação do descritor local condensada nucleofilicidade para a estrutura da TRP-cage, com fundo de transparência, renderizada e pronta para ser salva em arquivo.

Esse processo de colorir os átomos, renderizar e salvar essas imagens pode ser repetido para todos os outros descritores. Para os descritores de dureza local é interessante utilizar outros modos de *ray tracing*. Para próxima imagem mostrada, foi ativado no Pymol o objeto 2jof_PM7_hardness_B, e executado os seguintes comandos no Pymol

```
set ray_trace_mode=1
show sticks
set spectrum b,white_yellow_orange_red_black,minimum=0.1,maximum=0.2
ray
png 2jof_hardneess.png
```

A imagem salva em disco deve ser similar a mostrada na Figura 27, mostrando a dureza local para a estrutura da TRP-cage.

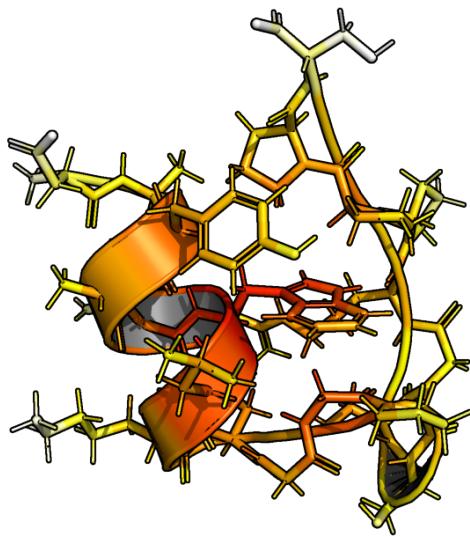


Figura 27: Imagem salva para a dureza local em representação condensada gerada no Pymol.

3.3.2 Representação Volumétrica

Os descritores com representação volumétrica seguem a mesma lógica de visualização que os outros dois últimos tutoriais. A diferença, no caso dos cálculos específicos para macromoléculas, através dos métodos BD e EW, é na seleção e combinação dos orbitais moleculares. Para automatizar a visualização dos descritores básicos em representação volumétrica, o PRIMoRDiA escreve um script semelhante ao mostrado na Figura 28, que é para a entrada `2jof_PM7_lmo.aux`. Para a outra entrada, `2jof_PM7.aux`, também foi escrito um script desses, que depois de ser executado abre as estruturas e renderiza os cubos no Pymol como na Figura 29.

```

preset.publication(selection='all')
set sphere_scale, 0.2
set bg_rgb, white
set stick_radius, 0.18
load 2jof_PMT_lmo.pdb
load 2jof_PMT_lmoFOA_HOMO_ph1.cube
load 2jof_PMT_lmoFOA_HOMO_ph2.cube
load 2jof_PMT_lmoFOA_LUMO_ph1.cube
load 2jof_PMT_lmoFOA_LUMO_ph2.cube
load 2jof_PMT_lmoFOA_Nucleophilicity.cube
load 2jof_PMT_lmoFOA_Electrophilicity.cube
load 2jof_PMT_lmoFOA_Radical_susceptibility.cube
load 2jof_PMT_lmoFOA_Netphilicity_ph1.cube
load 2jof_PMT_lmoFOA_Netphilicity_ph2.cube
load 2jof_PMT_lmoFOA_L_Hardness_LCP.cube
load 2jof_PMT_lmoFOA_L_Hardness_Vee.cube
load 2jof_PMT_lmoFOA_Fukui_Potential_left.cube
load 2jof_PMT_lmoFOA_L_Hardness_INT.cube
load 2jof_PMT_lmoFOA_Softness_Dual_ph1.cube
load 2jof_PMT_lmoFOA_Softness_Dual_ph2.cube
volume 2jof_PMT_lmoFOA_HOMO_ph1.volume, 2jof_PMT_lmoFOA_HOMO_ph1
volume 2jof_PMT_lmoFOA_HOMO_ph2.volume, 2jof_PMT_lmoFOA_HOMO_ph2
volume 2jof_PMT_lmoFOA_LUMO_ph1.volume, 2jof_PMT_lmoFOA_LUMO_ph1
volume 2jof_PMT_lmoFOA_LUMO_ph2.volume, 2jof_PMT_lmoFOA_LUMO_ph2
volume 2jof_PMT_lmoFOA_Nucleophilicity.volume, 2jof_PMT_lmoFOA_Nucleophilicity
volume 2jof_PMT_lmoFOA_Electrophilicity.volume, 2jof_PMT_lmoFOA_Electrophilicity
volume 2jof_PMT_lmoFOA_Radical_susceptibility.volume, 2jof_PMT_lmoFOA_Radical_susceptibility
volume 2jof_PMT_lmoFOA_Netphilicity_ph1.volume, 2jof_PMT_lmoFOA_Netphilicity_ph1
volume 2jof_PMT_lmoFOA_Netphilicity_ph2.volume, 2jof_PMT_lmoFOA_Netphilicity_ph2
volume 2jof_PMT_lmoFOA_L_Hardness_LCP.volume, 2jof_PMT_lmoFOA_L_Hardness_LCP
volume 2jof_PMT_lmoFOA_L_Hardness_Vee.volume, 2jof_PMT_lmoFOA_L_Hardness_Vee
volume 2jof_PMT_lmoFOA_Fukui_Potential_left.volume, 2jof_PMT_lmoFOA_Fukui_Potential_left
volume 2jof_PMT_lmoFOA_L_Hardness_INT.volume, 2jof_PMT_lmoFOA_L_Hardness_INT
volume 2jof_PMT_lmoFOA_Softness_Dual_ph1.volume, 2jof_PMT_lmoFOA_Softness_Dual_ph1
volume 2jof_PMT_lmoFOA_Softness_Dual_ph2.volume, 2jof_PMT_lmoFOA_Softness_Dual_ph2

```

Figura 28: Parte do script de automação da visualização dos descritores locais volumétricos no Pymol escrita no PRIMoRDiA e aberta em editor de texto.

Na Figura 29 já nos adiantamos e selecionamos só as fases do descrito de netfilicidade, lembrando que todos os descritores que possuem valores negativos e positivos são escritos em dois arquivos idênticos, mas com nomes diferentes, para que o Pymol renderize separadamente. Também, na mesma imagem é possível ver que já tem algumas cadeias laterais já selecionada, que são aqueles que é possível perceber que os volumes estão renderizados sobre seus átomos.

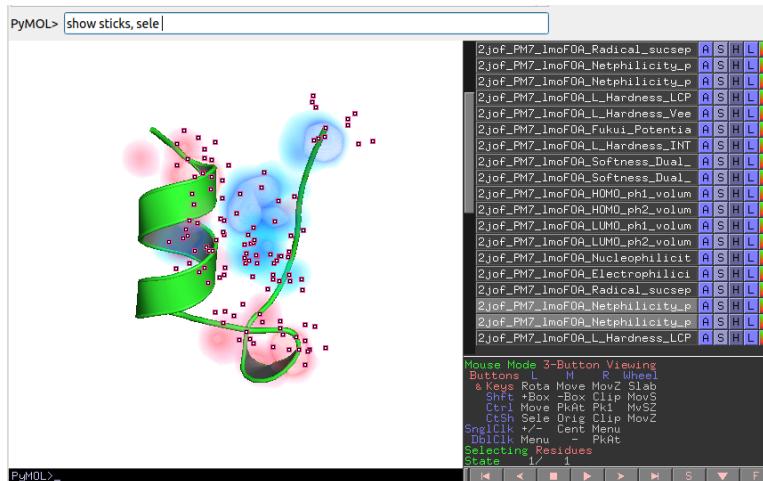


Figura 29: Janela do Pymol mostrando os descritores locais volumétricos carregadas pelo script gerado no PRIMoRDiA para a estrutura da TRP-cage com alguns resíduos selecionados.

Na área de comando do Pymol mostrado na Figura 29 aparece o comando para que a representação de bastões seja ativada para

essas cadeias selecionadas. Depois desse comando ser executados os bastões aparecem como mostrado na Figura 30. O comando de *draw* pode ser utilizado para renderizar e de depois salvar a imagem em disco com o comando *png*, como já exemplificado nos tutoriais 1 e 2.

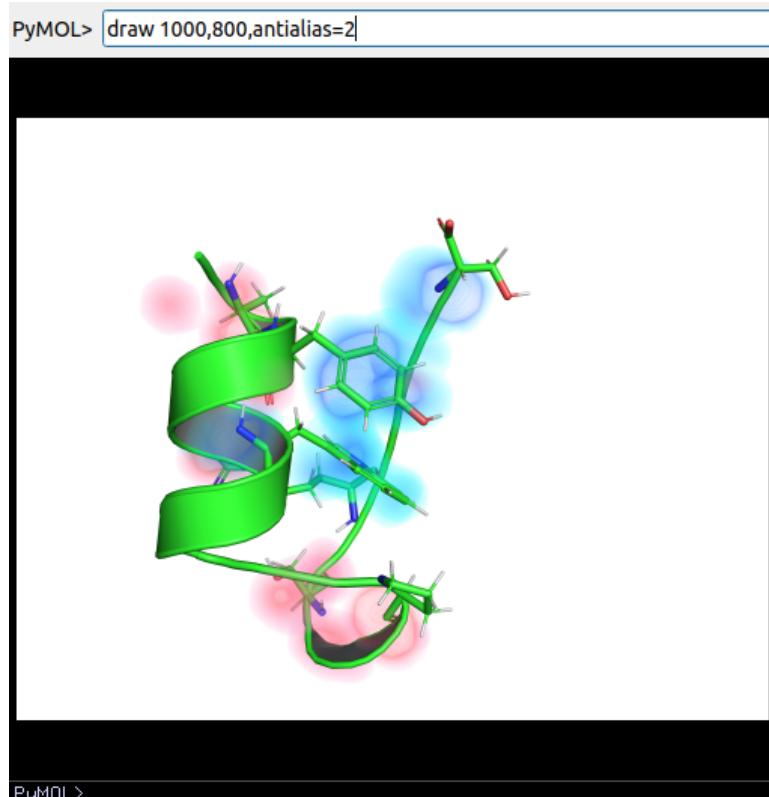


Figura 30: Janela do Pymol mostrando os descritores locais volumétricos carregadas pelo script gerado no PRIMoRDiA para a estrutura da TRP-cage com alguns resíduos mostrados em representação de bastões.

Agora vamos visualizar os resultados dos descritores calculados para a caixa definida no arquivo de input, executando o script do Pymol corresponde para essa entrada. O resultado deve ser como o apresentado na Figura 31, carregando e renderizando vários descritores para a região limitada.

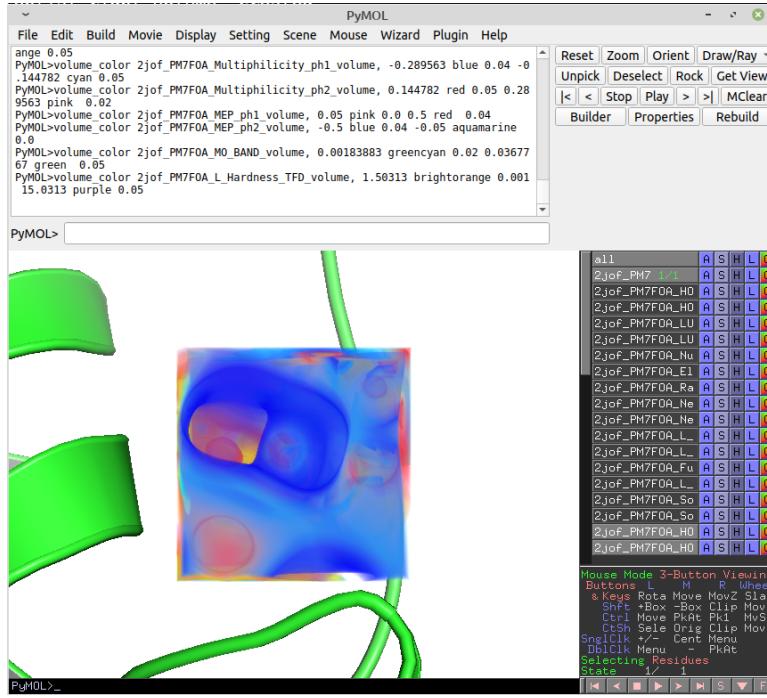


Figura 31: Janela do Pymol mostrando os descritores locais volumétricos carregadas pelo script gerado no PRIMoRDiA para um região cúbica específica da TRP-cage.

Se desativarmos todos os objetos de volume, ficamos só com os objetos de cube carregados, que no visualizador apareceram marcando as arestas dos limites do cube. Na Figura 32, mostramos esse limite em vermelho, com uma cor diferente para chamar atenção.

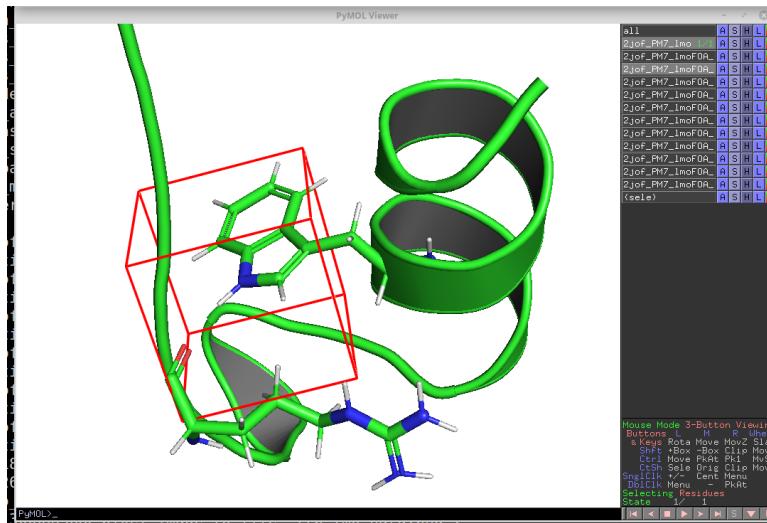


Figura 32: .

Em relação ao outros descritores volumétricos, os definidos dentro desses espaços limitados pelo usuário, não apresentam nenhuma diferença prática. A questão aqui é que com sistemas com um

grande número de átomos é possível visualizar os descritores para regiões específicas sem perder resolução e que possam ser calculados em um tempo praticável. Nas próximas figuras estão apresentados alguns descritores renderizados para esse região delimitada e já salvas em qualidade de publicação dentro do Pymol. Na Figura 33, o descritor de dureza local utilizando a aproximação baseada no potencial químico local. Na Figura 33, mostramos esse volume renderizado com uma paleta diferente, realçando melhor as regiões que apresentam valores maiores que outros.

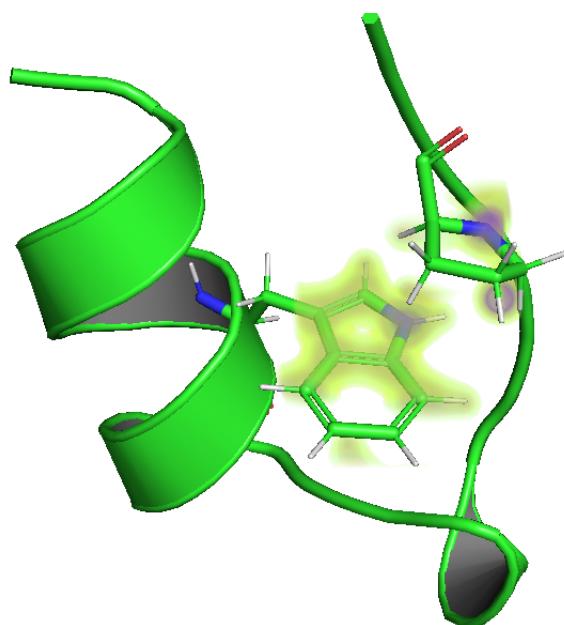


Figura 33: Exemplo de imagem do descritor renderizada para a região delimitada pelo usuário. Dureza local baseada na aproximação do potencial químico local.

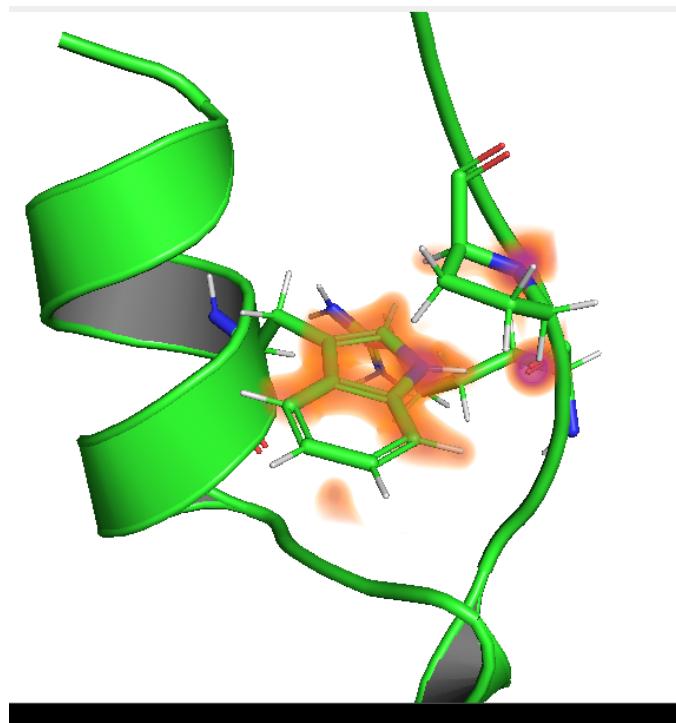


Figura 34: Exemplo de imagem do descriptor renderizada para a região delimitada pelo usuário. Dureza local baseada na aproximação do potencial químico local.

Na Figura 35, a última imagem é ampliada e a distância entre o hidrogênio da cadeia lateral e oxigênio do resíduo da cadeia principal é evidenciada, mostrando que nessa região onde claramente há uma ligação de hidrogênio também apresenta os maiores valores do descriptor de dureza local em questão.

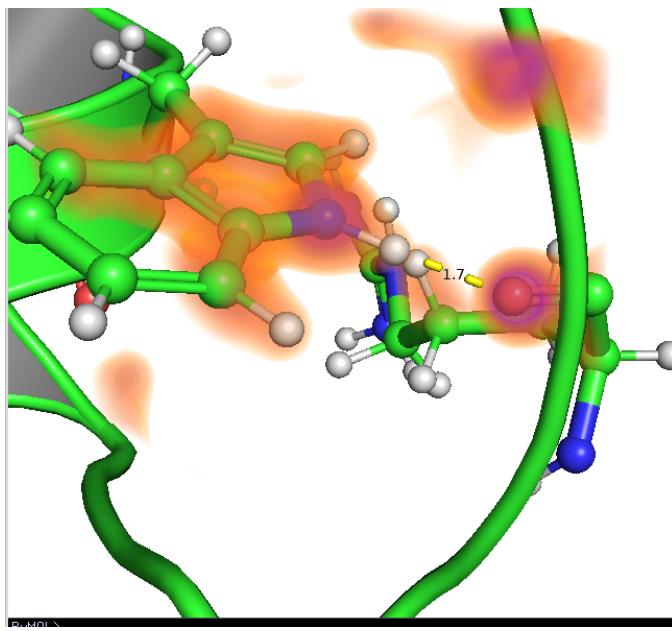


Figura 35: Exemplo de imagem do descriptor renderizada para a região delimitada pelo usuário. Imagem ampliada para a região de interação de por ligação de hidrogênio, evidenciando a distância entre os átomos. Dureza local baseada na aproximação do potencial químico local.

3.4 Representação Especial Para Resíduos e Densidade de Estados

Para macromoléculas biológicas há diversas análises focadas em propriedades e estatísticas em função dos monômeros que os compõem. Na grande maioria, esses sistemas são proteínas que são compostas de fragmentos de amino-ácidos, chamados de resíduos de amino-ácidos. Da mesma forma que já apresentamos os descritores condensados para os átomos, é interessante analisar esses valores para cada um desses fragmentos do sistema. PRIMoRDiA utiliza as informações do PDB para somar os descritores dos átomos por resíduo e escrever em um arquivo formatado, como na Figura 36.

```

1 #res Nucleophilicity Electrophilicity Radicals Hardness_Vee Hardness_LCP Fukui_pot_left Fukui_pot_right Fukui_pot_zero softn
2 1ASP 0.000047 0.196202 0.098124 0.196155 1.572429 1.246886 3.831282 4.933386 4.382334 -0.048172 0.005918 -0.443072 0.693494 -0.073080 43.48
3 2ALA 0.001107 0.195596 0.098352 0.194490 1.332675 0.833347 3.695541 4.433898 4.064720 -0.047763 0.005932 -0.439310 0.692051 0.064300 29.061
4 3TYR 0.660673 0.587201 0.623472 0.073472 2.941187 1.770429 12.965459 10.502487 11.373973 0.018043 0.037629 0.165957 2.509603 0.022630 62.0
5 4ALA 0.000527 0.086474 0.043501 0.085948 1.364701 0.826317 3.610692 4.180204 3.895448 -0.021107 0.002623 -0.194137 0.305985 -0.013820 28.81
6 5GLN 0.001350 0.022433 0.011891 0.021083 2.307037 1.433682 5.614325 6.609253 6.111789 -0.005178 0.000717 -0.047622 0.080175 -0.006120 49.99
7 6TRP 2.325697 0.384173 1.354935 -1.941524 3.811355 1.953870 21.420384 12.473799 16.928882 0.476799 0.081715 4.385480 2.886220 -0.001400 69.
8 7LEU 0.001024 0.074981 0.038003 0.073957 2.697134 1.398316 9.588930 8.048403 8.181666 -0.018162 0.002292 -0.167052 0.265689 -0.028690 48.76
9 8LYS 0.000008 0.084905 0.042456 0.084896 2.724491 1.424280 6.479512 7.440169 6.959840 -0.020849 0.002561 -0.191762 0.300096 0.978680 49.667
10 9ASP 0.000004 0.109826 0.054915 0.109822 1.719865 1.363234 4.209618 5.265296 4.737457 -0.026970 0.003312 -0.248064 0.388178 -0.927156 45.44
11 10GLY 0.000168 0.099295 0.049732 0.099127 0.994898 0.630461 2.698183 3.119944 2.909064 -0.024344 0.002999 -0.223907 0.351063 -0.041440 21.9
12 11GLY 0.001816 0.074693 0.038254 0.072878 1.103087 0.601506 4.380177 3.545539 3.962856 -0.017897 0.002307 -0.164615 0.265193 0.089660 20.97
13 12PRO 0.0001395 0.111834 0.056614 0.110439 3.788980 1.114722 6.485771 5.592681 6.039226 -0.027122 0.003414 -0.249459 0.396187 -0.019270 38.8
14 13SER 0.000006 0.147275 0.073641 0.147261 1.371789 0.993846 3.558035 4.428792 3.993414 -0.036166 0.004441 -0.332648 0.520539 -0.037140 34.6
15 14SER 0.000001 0.181425 0.090713 0.181424 1.546528 1.007916 4.110005 5.188317 4.649161 -0.044554 0.005471 -0.409798 0.641238 -0.039740 35.1
16 15GLY 0.000000 0.094519 0.047250 0.094518 0.889732 0.641020 2.528992 2.992615 2.760803 -0.023212 0.002850 -0.213496 0.334070 0.027620 22.35
17 16ARG 0.000451 0.481992 0.241222 0.481541 3.371486 1.748639 10.318387 11.524630 10.921508 -0.118257 0.014548 -1.087696 1.703870 0.945780 60
18 17PRO 0.000529 0.027119 0.053180 1.714908 1.065170 5.489696 4.814029 5.151863 -0.013060 0.001636 -0.120122 0.190180 0.017260 37.14
19 18PRO 0.004618 0.010071 0.007345 0.005453 1.931782 1.104373 8.498590 5.456805 6.977697 -0.001339 0.000443 -0.012317 0.038631 0.018800 38.51
20 19PRO 0.009572 0.0063030 0.001801 0.002457 1.935335 1.107264 6.870249 5.435734 6.152992 -0.006604 0.000109 -0.005551 0.011083 -0.033250 38.6
21 20SER 0.000006 0.000364 0.000185 0.000358 1.249890 1.147167 3.736653 3.080400 3.408526 -0.000088 0.000011 -0.000808 0.001290 -0.942470 40.0
22 Residues_Average 0.150000 0.150000 0.150000 -0.000000 1.918465 1.167622 6.504524 5.951498 6.228011 -0.000000 0.000946 -0.000000 0.628742 0.0
23

```

Figura 36: Arquivo com os descritores locais condensados por resíduo (fragmento de amino-ácido) para a estrutura da TRP-cage aberto em editor de texto.

Esses dados podem ser facilmente selecionados/copiados e transferidos para uma tabela de excel/libreoffice, como é exemplificado na Figura 37. Para uma dada trajetória, de dinâmica molecular ou coordenada de reação, valores para resíduos específicos podem ser reunidos e analisados durante esses processos. Mais sobre isso é exemplificado no tutorial 5.

#res	A	B	C	D	E	F	G	H	I	J	K
	Nucleophilicity	Electrophilicity	Radicality	Netphilicity	Hardness	V ₀	Hardness_L ₀	Fukui_pot_I ₀	Fukui_pot_R ₀	Fukui_pot_z ₀	softness_dualhyp
1											
2	1ASP	4.7E-05	0.196202	0.098124	0.196155	1.572429	1.246886	3.831282	4.933386	4.382334	-0.048172
3	2ALA	0.001107	0.195596	0.098352	0.19449	1.332675	0.833347	3.695541	4.433898	4.06472	-0.047763
4	3TYR	0.660673	0.587201	0.623937	-0.073472	2.941187	1.770429	12.965459	10.502487	11.733973	0.018043
5	4ALA	0.000527	0.086474	0.043501	0.085948	1.364701	0.826317	3.610692	4.180204	3.895448	-0.021107
6	5GLN	0.00135	0.024233	0.011891	0.021083	2.307037	1.433682	5.614325	6.609253	6.111789	-0.005178
7	6TRP	2.325697	0.384173	1.354935	-1.941524	3.811355	1.95387	21.420384	12.437379	16.928882	0.476799
8	7LEU	0.001024	0.074981	0.038003	0.073957	2.697134	1.398316	9.58893	8.048403	8.818666	-0.018162
9	8LYS	8E-06	0.084905	0.042456	0.084896	2.724491	1.424228	6.479512	7.440169	6.95984	-0.020849
10	9ASP	4E-06	0.109826	0.054915	0.109822	1.719865	1.303234	4.209618	5.265296	4.737457	-0.02697
11	10GLY	0.000168	0.099295	0.049732	0.099127	0.994898	0.630461	2.698183	3.119944	2.909064	-0.024344
12	11GLY	0.001816	0.074693	0.038254	0.072878	1.103087	0.601506	4.380177	3.545539	3.962858	-0.017897
13	12PRO	0.001395	0.111834	0.056614	0.110439	1.78898	1.114722	6.485771	5.592681	6.039226	-0.027122
14	13SER	6E-06	0.147275	0.073641	0.147269	1.371789	0.993846	3.558035	4.428792	3.993414	-0.036166
15	14SER	1E-06	0.181425	0.090713	0.181424	1.546528	1.007916	4.110005	5.188317	4.649161	-0.044554
16	15GLY	0	0.094519	0.047259	0.094518	0.889732	0.64102	2.528992	2.992615	2.760803	-0.023212
17	16ARG	0.000451	0.481992	0.241222	0.481541	3.371486	1.748639	10.318387	11.52463	10.921508	-0.118257
18	17PRO	0.000529	0.053709	0.027119	0.05318	1.714908	1.06517	5.489696	4.814029	5.151863	-0.01306
19	18PRO	0.004618	0.010071	0.007345	0.005453	1.931782	1.104373	8.49859	5.456805	6.977697	-0.001339
20	19PRO	0.000572	0.00303	0.001801	0.002457	1.935335	1.107264	6.870249	5.435734	6.152992	-0.000604
21	20SER	6E-06	0.000364	0.000185	0.000358	1.24989	1.147167	3.736653	3.0804	3.408526	-8.8E-05
22	Residues Av ^b	0.15	0.15	0.15	0	1.918465	1.167622	6.504524	5.951498	6.228011	0

Figura 37: Planilha eletrônica com os descritores locais condensados por resíduo (fragmento de amino-ácido) para a estrutura da TRP-cage.

Através da análise desses dados numéricos, é possível indicar a reatividade de sistemas biológicos pelo seus resíduos, o que é muito útil na hora de definir domínios importantes para enzimas, como sitio ativo, sitio de ligação, sitio alostérico e etc. Além disso, essas tabelas de dados são ricas em informações para algoritmos de aprendizagem de máquina, com propriedades eletrônicas calculadas por métodos quânticos para monômeros dos sistemas.

Uma das funcionalidades do PRIMoRDiA é escrever uma tabela de energia e um script para ser executado em R para a produção de um gráfico de densidade de estados de energia. A análise desses dados é interessante para biopolímeros, devido a alta concentração de níveis de energia com a energia similar ao do HOMO e LUMO. Esse script é produzido quando a keyword "dos" é utilizada no campo de parâmetros do input e pode ser útil para avaliar o número de orbitais máximos utilizados ou a valor de corte de energia para contabilização dos orbitais moleculares.

Na pasta desse tutorial, depois que o input foi executado, deve ter arquivos com final "DOS.R" e "DOS". Para gerar uma imagem automaticamente com o script é só executar o comando

```
computer@user:$ Rscript 2jof_PM7_lmo_DOS.R
```

E na pasta deve aparecer uma imagem como a figura abaixo

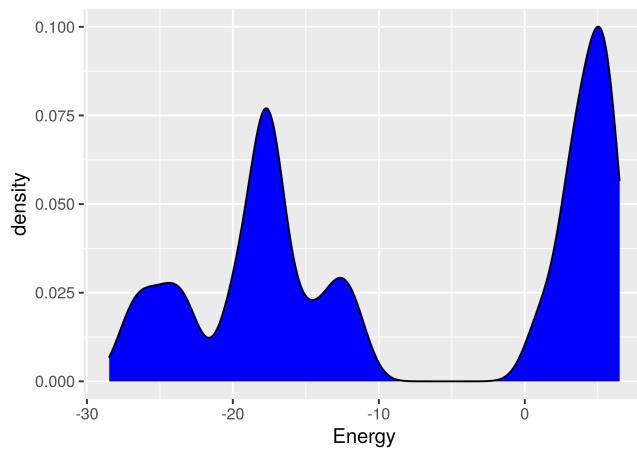


Figura 38: Gráfico de densidade de estados de energia para os orbitais moleculares da TRP-Cage calculados com PM7 junto com algoritmo MOZYME.

E o comando abaixo cria a figura da densidade de estados para a TRP calculada normalmente com o PM7

```
computer@user:$ Rscript 2jof_PM7_DOS.R
```

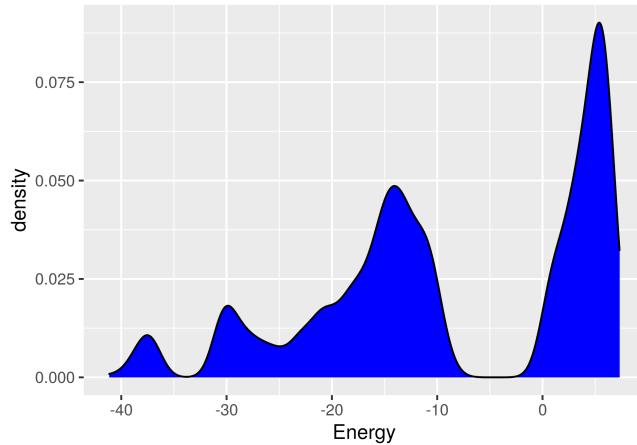


Figura 39: Gráfico de densidade de estados de energia para os orbitais moleculares da TRP-Cage calculados com PM7.

Nós podemos observar que os valores de energia dos orbitais calculados com o algoritmo MOZYME são mais concentrados que os que não utilizaram essa estratégia de aceleração dos cálculos. Isso faz com que o gap HOMO-LUMO fique maior nesse primeiro caso, o que acaba alterando bastante na hora de escolher os orbitais moleculares para combinar na hora de calcular os descritores. Portanto, o ideal para sistemas novos é sempre que possível fazer a análise desse gráfico que facilmente pode ser produzido com a ajuda do PRIMoRDiA.

4 Tutorial 4: Análise de Força Ácida

Esse tutorial utiliza a mesma aproximação e lógica de interpretação dos resultados do Tutorial 1, mas focado em algo de maior interesse do dia-a-dia da química. Como nesse primeiro. Vamos demonstrar como utilizar o programa para obter os descritores de reatividade usando as informações dos orbitais moleculares de fronteira, opção 1 de cálculo do PRIMoRDiA, partindo de cálculos semi-empíricos no MOPAC, para quatro aminoácidos, gerando descritores de reatividade que podem ser utilizados para explicar os locais preferenciais de protonação.

4.1 Contextualização

Aminoácidos são moléculas orgânicas anfóteras, isto é, podem agir como ácidos e bases, devido aos grupos amina e ácido carboxílico. Além disso, alguns tem cadeias laterais que são considerados como ácidos ou bases, e quando estão na composição de uma proteína essa acidez/basicidade vai ser relevante para diversos processos e propriedades.

A força ácida é medida pelo pKa, que é o negativo do log da constante de força ácida. De forma resumida, quando o pH está menor que o pKa de um grupo da molécula esse grupo tende a ficar protonado, em caso contrário sem esse próton. Ou seja, quanto menor o pKa mais ácido e quanto maior mais básico. No entanto, as medidas de pKa mudam com o ambiente químico e a sua obtenção de forma teórica apresenta diversos desafios quando os sistemas envolvidos são proteínas.

Os descritores podem ser úteis para identificar locais nas moléculas para sítios preferências para transferência de elétrons. Mais especificativamente, a dureza local é o descritor ideal para transferência de prótons, que é um partícula pequena carregada e portanto dominado por interações de natureza eletrostática. Nesse tutorial, vamos calcular os descritores para uma cisteína, ácido aspártico, ácido glutâmico e uma lisina, e cruzar com dados de pKa e potencial de ionização.

4.2 Arquivos Necessários e Execução

Todos os arquivos necessários para a execução se encontram na pasta comprimida fornecida junto ao programa, sendo eles os aquivos que contém as informações de saída do programa MOPAC para os quatro amino-ácidos estudados. Assim também como um arquivo de input com as mesmas linhas de texto que mostradas abaixo.

Listing 6: Input editado para execução do tutorial 3

```
#RT normal
#PR pymols
1 asp.aux true 40 mopac
1 cys.aux true 40 mopac
1 glu.aux true 40 mopac
1 lys.aux true 40 mopac
```

Como já mostrado nos tutoriais anteriores, depois de ter o input na pasta onde estão os arquivos é só rodar o programa com a flag `-f`. Depois disso os arquivos com os resultados devem ser criados na pasta.

4.3 Análise dos Resultados

Os resultados para esse tutorial são bem parecidos com os obtidos no tutorial 1, sendo que utilizamos a mesma opção de cálculo para geração dos descritores com aproximação de orbitais congelados. Na Figura 40 é mostrada o arquivo de texto com os resultados dos descritores globais, na Figura 41 os mesmos dados transpostos para um planilha do Libre Office.

```
primordia.global × aspFOA_EAS.cube × primordia.input × commands ×
1 Molecule Energy Ecat Ean HOF IP EA ECP Hardness Softness Electrophilicity nMax gap
2 asp -7998.93000000 0.00000000 0.00000000 -174.71400000 4.86000000 -3.10000000 -0.88000000 3.98000000
3 cys -5694.33000000 0.00000000 0.00000000 -38.59950000 8.83400000 0.49400000 -4.66400000 4.17000000 0
4 glu -9264.45000000 0.00000000 0.00000000 -157.49800000 4.03300000 -2.06700000 -0.98300000 3.05000000
5 lys -9454.76000000 0.00000000 0.00000000 -61.88870000 8.41100000 0.62000000 -4.51550000 3.89550000 0
6
```

Figura 40: Arquivo de texto com os resultados dos descritores globais para os amino-ácidos calculados.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Molecule	Energy	Ecat	Ean	HOF	IP	EA	ECP	Hardness	Softness	Electrophilicity	nMax	gap
2	asp	-7998.93	0	0	-174.714	4.86	-3.1	-0.88	3.98	0.25125628	0.09728643	0.44221106	7.96
3	cys	-5694.33	0	0	-38.5995	8.834	0.494	-4.664	4.17	0.23980815	2.60826091	2.23693046	8.34
4	glu	-9264.45	0	0	-157.498	4.033	-2.067	-0.983	3.05	0.32786885	0.15840803	0.64459016	6.1
5	lys	-9454.76	0	0	-61.8887	8.411	0.62	-4.5155	3.8955	0.25670646	2.61708898	2.31831601	7.791

Figura 41: Planilha com os resultados dos descritores globais para os amino-ácidos calculados.

O que pode ser extraído desses resultados é que para o mesmo método de estrutura eletrônica as moléculas de lisina e ácido aspártico tem dureza química bem próximas, a cisteína com uma dureza bem maior que todos e o ácido glutâmico com a menor. De fato, o ácido glutâmico tem dois oxigênios que geralmente em ambientes enzimáticos agem como aceptores de prótons e/ou participando em ataques nucleofílicos. Entretanto, por mais que a cisteína se apresente como mais dura nesses resultados, é o amino-ácido com a maior capacidade de receber elétrons junto com a lisina.

Já para os descritores locais, para cada molécula calculada foi gerado um arquivo ".lrd", como já explicado em tutoriais anteriores como

explora-los e analisa-los em planilhas. Os dados dos descritores locais mais apropriados para avaliar a força ácida vão ser reunidas posteriormente para os átomos aceptores/receptores de prótons.

Para ajudar de forma qualitativa na análise, podemos gerar os descritores volumétricos no pacote gráfico Pymol. Isso pode ser facilmente realizado usando os scripts gerados pelo PRIMoRDiA, como já vimos em outros tutoriais. Abra o Pymol e execute the ".pymols" script no terminal do programa. O resultado deve ser o parecido com o mostrado na Figura 42, com os objetos de cube carregados, os de volume gerados e selecionados.

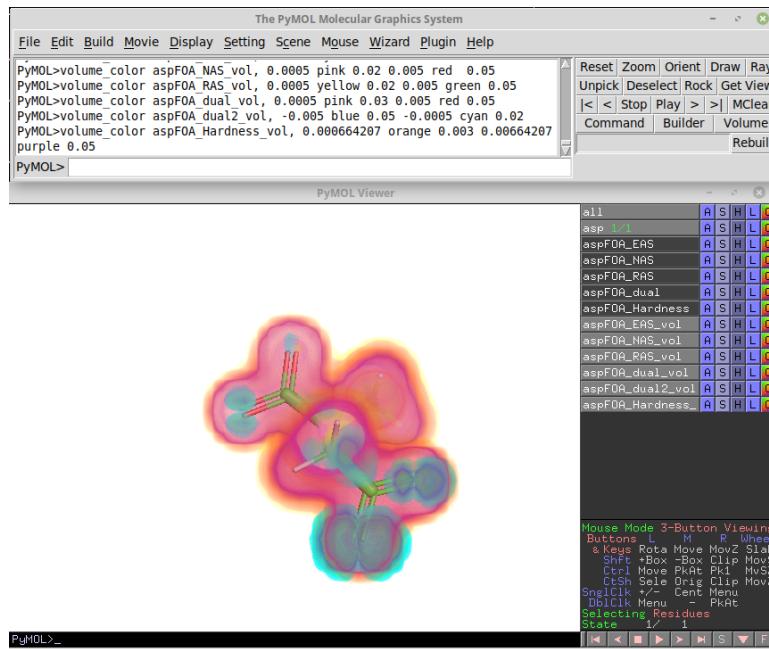


Figura 42: Descritores locais básicos em representação volumétrica para o ácido aspártico, renderizados no programa Pymol.

Na Figura 43 o descritor de Netfilicidade, ou as duas fases do descritor dual, é mostrado para o ácido aspártico. Com o comando "Draw", no Pymol é possível renderizar a imagem para salva-lá na resolução desejada e em alta qualidade, resultado na Figura 44. Nessa imagem é possível observar as superfícies renderizadas na cor azul estão localizadas em átomos de oxigênio, e representa a tendência desses locais de receber um ataque eletrolítico, e portanto são nucleófilos, e em rosa e vermelho locais onde tendem a receber um ataque nucleofílico e portanto são eletrófilos. Para o ácido aspártico, quase todos os oxigênios de ácido carboxílico apresentam ser nucleófilos, mas com clara preferência para um dos oxigênios da cadeia lateral. Enquanto isso, no grupo amina apresenta ser um eletrófilo, com uma disposição esférica do descritor ao redor do grupo, sinalizando que qualquer próton pode ser retirado para que o par de elétrons desse nitrogênio carregado positivamente seja

restaurado.

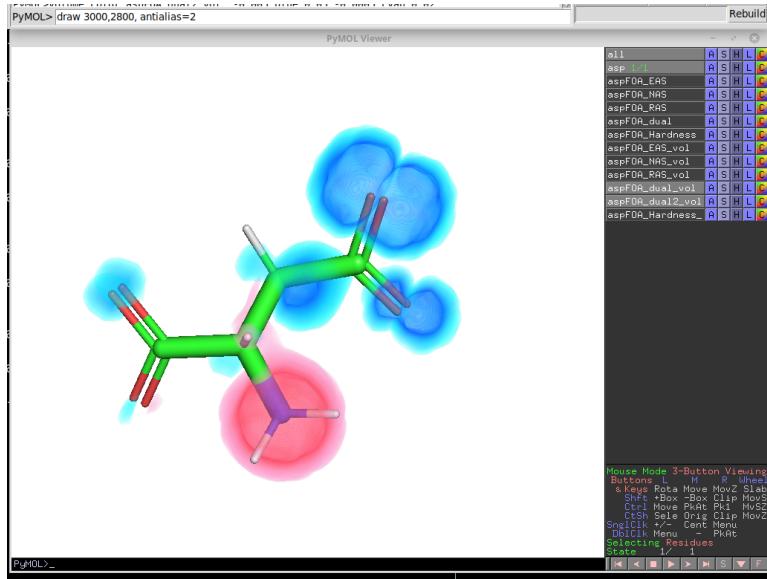


Figura 43: Descritor de Netfilicidade(dual) mostrado no Pymol para a molécula de ácido aspártico, com o comando no terminal para renderizar a imagem em alta qualidade.

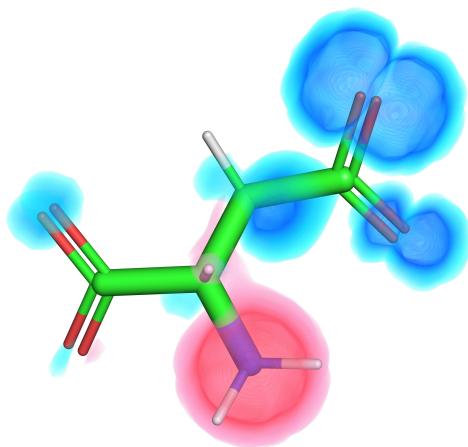


Figura 44: Imagem salve em disco para o descritor de Netfilicidade(dual) para a molécula de ácido aspártico.

Passando para a análise do descritor de dureza, é só desativar os objetos de volume anteriores no Pymol e ativamos o último, "aspFOA_Hardness_vol", uma representação volumétrica do descritor deve aparecer como mostrado na Figura 45

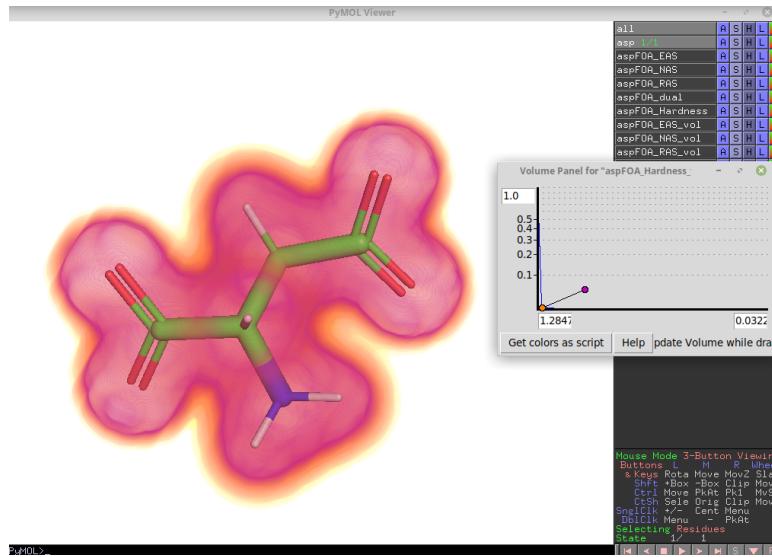


Figura 45: Descritor de dureza local para o ácido aspártico, mostrando o painel de volume do Pymol onde é possível criar a ajustar as iso-superfícies.

A janela que aparece ao lado da Figura 46 é aberta colocando em "volume" no painel ao lado da área de execução de comandos do Pymol. Clicando e movimentando os pontos nessa janela é possível mudar as cores, os valores das iso-superfícies e suas opacidades. Também é possível adicionar mais iso-superfícies. Na figura abaixo é mostrado uma modificação feita para o descritores de dureza local.

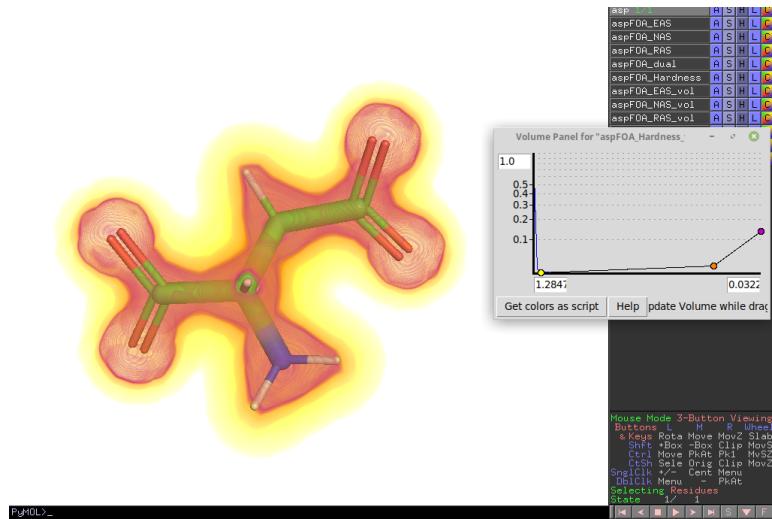


Figura 46: Descritor de dureza local para o ácido aspártico, mostrando o painel de volume do Pymol com novos valores sugeridos para as iso-superfícies para uma melhor visualização.

Repetindo o processo de renderização e de salvar as imagens para os outros sistemas, nós obtemos os descritores para quatro os sistemas, como mostrada nas próximas duas figuras: Netfilicidade na Figura 47 e de dureza na Figura 48.

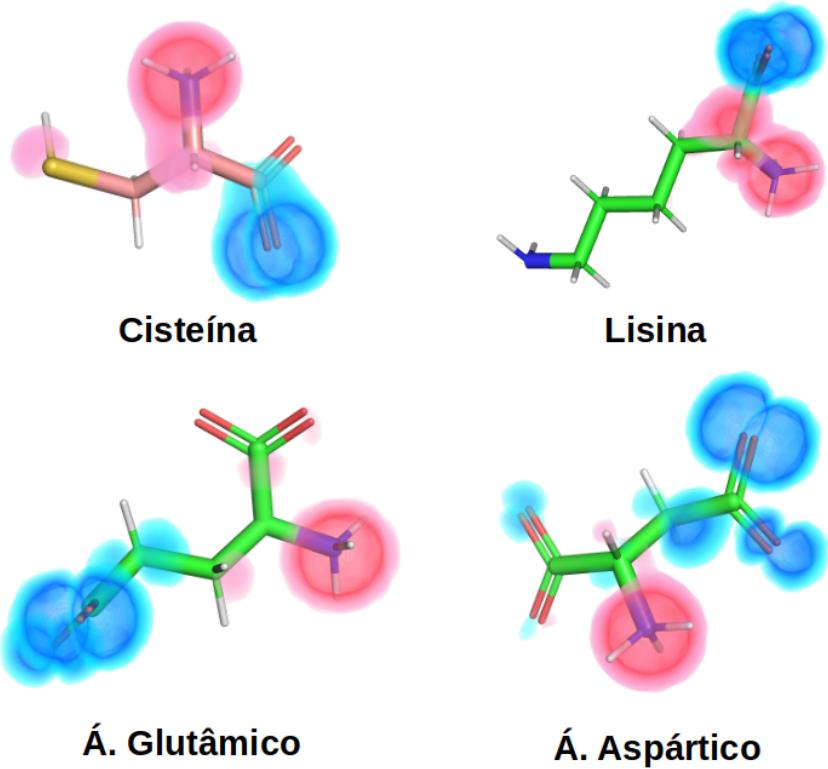


Figura 47: Descriptor local de Netfilicidade para os quatro amino-ácidos analisados no tutorial.

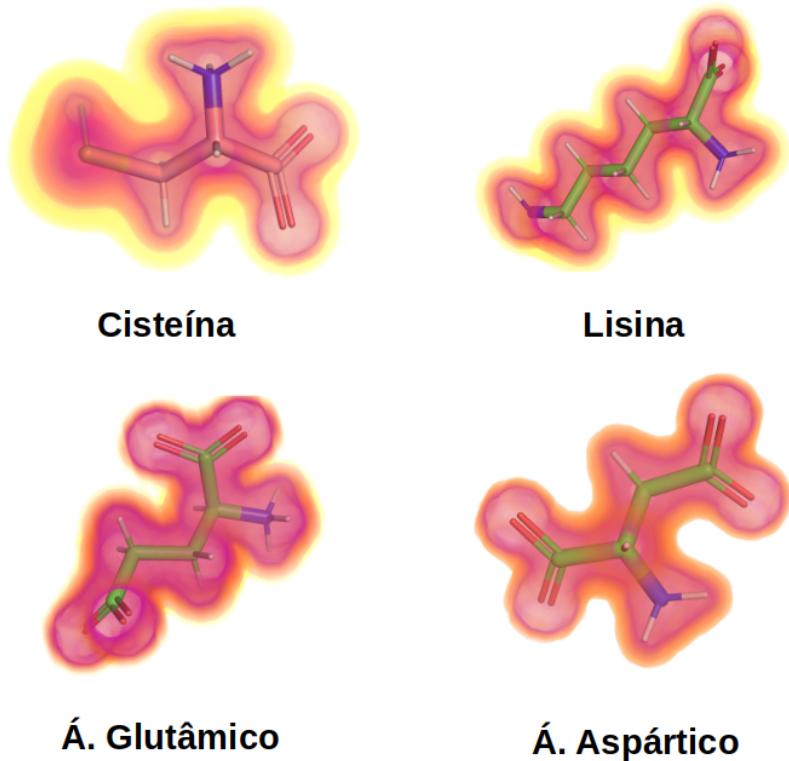


Figura 48: Descritor de dureza local para os quatro amino-ácidos analisados no tutorial, usando o método de potencial químico local.

	-CO2H	-NH2	sidechain	pI
Aspartic Acid	1.99	9.90	3.90	2.98
Cysteine	1.92	10.70	8.37	5.15
Glutamic Acid	2.10	9.47	4.07	3.08
Lysine	2.16	9.06	10.54	9.47

Figura 49: Tabela com os resultados de descritores locais para os grupos aceptores/doadores de prótons da cadeia principal de cada um dos quatro amino-ácidos.

	Nucleofilicidade		Eletrofilicidade	Dureza			
	O1	O2	N1	N1	O1	O2	
Aspartic Acid	0.0524	0.8118	0.3003	2.6830	3.6419	3.1563	
Cysteine	0.3138	0.6704	0.2332	2.9926	3.7783	3.5915	
Glutamic Acid	0.0633	0.0011	0.3071	1.7035	2.2454	1.5802	
Lysine	0.3193	0.6697	0.2687	1.6476	1.9888	2.1211	
	carga parcial			global			
	N1	O1	O2	ECP	Hardness	IP	EA
Aspartic Acid	-0.0728	-0.5939	-0.6224	-0.88	3.98	4.86	-3.1
Cysteine	-0.0516	-0.4869	-0.5251	-4.664	4.17	8.834	0.494
Glutamic Acid	-0.0727	-0.5556	-0.5533	-0.983	3.05	4.033	-2.067
Lysine	-0.0738	-0.4896	-0.5603	-4.5155	3.8955	8.411	0.62

Figura 50: Tabela com os resultados de descritores locais para os átomos aceptores/doadores de prótons da cadeia lateral de cada um dos quatro amino-ácidos. Também uma tabela complementar com os descritores globais, cargas parciais e o valor conhecido de Potencial de Ionização.

Juntando as informações que podemos observar dos descritores condensados e volumétricos que essas quantidades teóricas conseguem localizar as porções das moléculas onde transferências de prótons são mais prováveis de ocorrer. O descritor de dureza química é o mais apropriado para esse tipo de processo, transferência de pequenos grupos carregados, e mostrou uma relação com o pKa dos grupos amina, com a cisteína mostrando a maior dureza no nitrogênio e também o maior pKa. Obviamente, esse pequeno estudo está repleto de limitações e só tem intenção de mostrar uma relação prática dos resultados obtidos com o PRIMoRDiA e um problema de modelagem molecular comum.

5 Tutorial 5: Análise de Reação Enzimática

Nesse tutorial vai ser mostrado um exemplo prático de como o PRIMoRDiA pode ser explorado para o estudo de sistemas biológicos e seus principais processos. A grande força do PRIMoRDiA está nos descritores modificados para macromoléculas que disponibiliza e na automatização do processamento dessas propriedades eletrônicas. Vamos entrar com a estrutura eletrônica de um fragmento de uma sistema enzimático para uma pequena trajetória de reação e sairemos com um resumo estatístico da evolução dos descritores nesse processo.

5.1 Contextualização

Reações enzimáticas são reações químicas catalisadas por proteínas, que fazem com que transformações químicas ocorram em um tempo hábil para o funcionamento do metabolismo de organismos vivos. Portanto, as enzimas são componentes importantíssimos para a manutenção da vida, participando desde a replicação de DNA até a decomposição do açúcar.

O estudo do mecanismo dessas reações é muito importante para o desenvolvimento de drogas, entendimento de metabolismo, mecanismos de doenças e replicação de micro-organismos e aplicações industriais dessas enzimas. Em um dos estudos do nosso grupo, foi com o mecanismo da reação catalisada pela triose-fosfato-isomerase (TIM), mais especificamente a etapa mostrada no esquema abaixo na Figura 51.

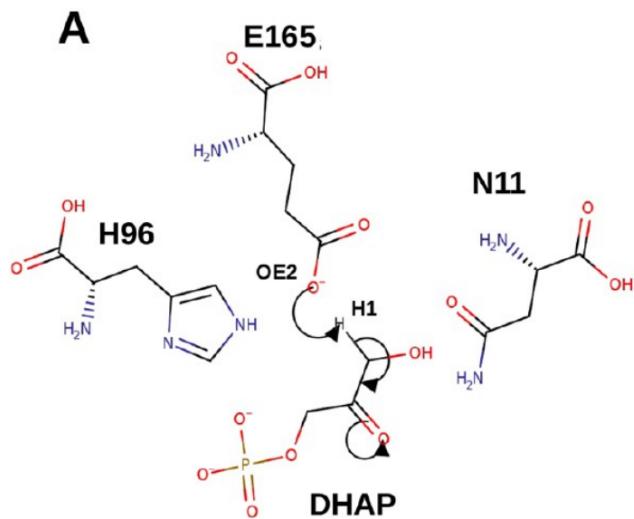


Figura 51: Representação do mecanismo de catálise enzimática estudado nesse tutorial.

Nós utilizamos esse enzima para testar a eficácia dos descritores em caracterizar teoricamente essa reação, descrevendo as principais

forças, interações e efeitos do ambiente. A TIM é uma das enzimas mais caracterizadas na literatura, é uma enzima central no caminho glicolítico da interconversão do 3-fosfato D-gliceroldeído (GAP em inglês) para o fosfato dihidroxiacetona, com grande eficiência.

Nesse tutorial nós vamos calcular os descritores para esse sistema em uma trajetória de reação obtida por simulação de QM/MM.

5.2 Arquivos necessários e Execução

Os arquivos necessários para os cálculos desse tutorial estão na pasta . São arquivos de ".aux" e ".pdb", para as etapas do caminho de reação. Esse caminho de reação engloba somente 110 átomos, feito dessa forma para obter arquivos mais leves para o tutorial. A trajetória do scan foi obtida como explicado no tutorial do easyhybrid.

Para fazer as análises especiais para a reação enzimática nós vamos outro método de execução, que é sinalizado no PRIMoRDiA pela keyword "reaction" depois de "#RT" na primeira linha, ou seja, o Run Type passa para o modo de reação. Na pasta já tem o arquivo de input, que agora possui vários elementos importantes de citar, e está mostrado na figura abaixo

É possível perceber que tem varias outras linhas com keywords antes não usadas e somente uma entrada de arquivo ".aux" para rodar com a opção 3 de cálculo. Como esse é um cálculo para reação química, o PRIMoRDiA entende aquela entrada como valendo para todos os auxs na pasta que tenham o sufixo sinalizado, nesse caso no nosso input foi o sufixo "sys", rodando de "sys0.aux" até "sys14.aux", obedecendo o parâmetro "dimX".

Vamos analisar os outros parâmetros do input

Listing 7: Input editado para execução do tutorial 3

```
#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
```

Segunda linha

- Segunda linha
 - #PR: flag para sinalizar que serão dados parâmetros para o PRIMoRDiA
 - eband: keyword sinalizando que o próximo parâmetro é um valor de energia de corte para considerar os orbitais moleculares

- Rscript: Keyword importante par gerar os scripts de análise dos descritores durante a trajetória de reação
 - pymols: keyword para o output de scripts para o Pymol
- Terceira linha
 - #Reaction: flag para sinalizar que vai ser passado parâmetros para análise de coordenada de reação
 - RC1: keyword sinalizando que vais ser passado o número dos átomos que compõem a coordenada de reação
 - 98: número do átomo de carbono (C1) do GAP
 - 107:número do átomo de hidrogênio (H1) ligado ao DHAP
 - 51: número do átomo de oxigênio (OE2) do GLU
- Quarta Linha
 - #Reaction: flag para sinalizar que vai ser passado parâmetros para análise de coordenada de reação
 - dimX : keyword para sinalizar que o próximo valor na linha da o número de etapas da trajetória
 - 15 : número de etapas da trajetória
- Quinta Linha
 - #TRJ: flag para sinalizar que vai ser passado parâmetros tratar com o conjunto de frames que compõem a trajetória de reação
 - residues : keyword indicando que os próximos valores inteiros passados nessa linha são o número dos resíduos no PDB que se deseja analisar para a dada trajetória.
 - 2: número no PDB da histidina catalítica
 - 4: número no PDB do Glutamato catalítico
 - 7: número no PDB do resíduo representando o substrato.
- Sexta Linha
 - 3 : opção de cálculo do PRIMORDiA para macromoléculas
 - sys.aux: nome base para os arquivos, antes do ponto o prefixo sem o número que indica o frame, e depois do ponto a extensão do arquivo
 - none : keyword sinalizando que nenhum métodos de dureza local foi pedido para cálculo da representação volumétrica, o que já não seria feito já que o valor de grid é zero
 - 0 : finura do grid para representação volumétrica, sempre deixar em zero até uma nova atualização do software

- 10 : número máximo de orbitais moleculares de fronteira para ser utilizado se o método de combinação de orbitais for BD
- sys.pdb: nome base para os arquivos de pdb, mesma lógica para os arquivos que contém a estrutura eletrónica
- mopac : nome do programa de onde foi calculada a estrutura eletrónica
- 0 : parâmetro não relevante com grid 0
- 0 : parâmetro não relevante com grid 0
- 0 : parâmetro não relevante com grid 0
- 0 : parâmetro não relevante com grid 0
- EW : Keyword para o método de combinação de orbitais moleculares

5.3 Análise dos Resultados

Para analisar todos os descritores e outras propriedades de energia, o PRIMoRDiA cria um script para executar no R que por sua vez produz gráficos dessas quantidades variando em função das frames. Depois da execução deve aparecer arquivos com sufixo ".R", um deles com nome "sys0_reaction_analysis.R" e outro "sys0_residuos_analysis.R". Vamos executar o primeiro com o Rscript, como no comando abaixo

```
computer@user:$ Rscript sys0\reaction\analysis.R
```

Depois desse comando, o R deve produzir diversos arquivos de imagem com propriedades globais e locais para cada átomo da coordenada de reação. Na figura abaixo é mostrado um desses arquivos com as propriedades globais.

```

1 #RT reaction
2 #PR eband 2 Rscript pymols
3 #Reaction RC1 98 107 51
4 #Reaction dimX 15
5 #TRJ residues 2 4 7
6 3 sys.aux none 0 10 sys.pdb mopac 0 0 0 0 EW
7

```

Figura 52: Arquivo contendo os descritores globais gerado pelo PRIMoRDiA.

No primeiro gráfico da Figura 53 HOF é o calor de formação (heat of formation) que o Mopac calcula, e é uma das melhores formas de variação de energia em uma reação química simulada com método semiempírico. Lembrando que todas essas propriedades são em referência ao primeiro valor, ou seja, cada ponto uma variação ao estado inicial. Nessas simulações uma estrutura que normalmente

se deseja determinar é a estrutura de transição, que é determinada pelo ponto máximo de variação de energia, nesse caso na penúltima etapa. No entanto, nesse nosso cálculo com PM6, a área quântica foi bem menor do que a utilizada na nossa publicação e o HOF não foi um bom indicador do estado de transição. Já a energia eletrônica mostra um primeiro máximo no ponto zero da coordenada de reação, que é o momento que o próton está a mesma distância do carbono e do oxigênio. Em geral, os outros descritores globais indicam um aumento da moleza química (softness) e consequente diminuição da dureza (hardness), aumento do potencial químico e da eletrofilicidade total indicam que o sistema tende a receber transferência de elétrons.

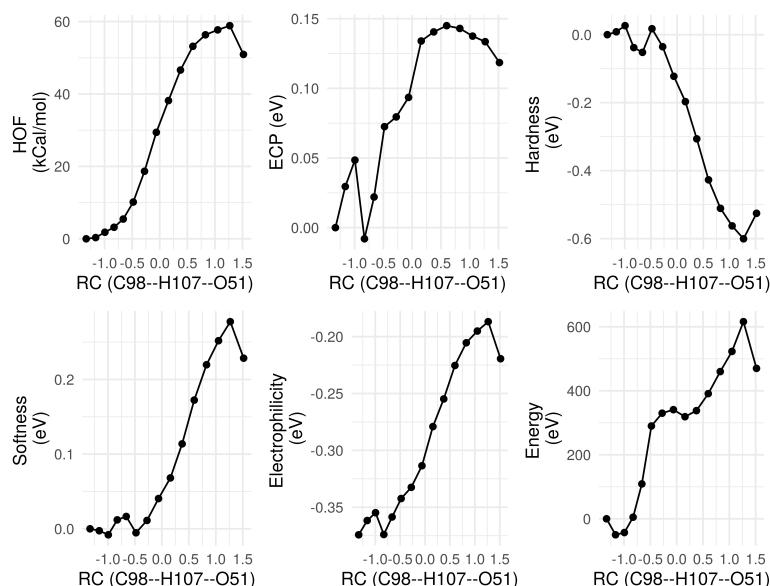


Figura 53: Descritores globais calculados para cada etapa da trajetória de reação.

Na Figura 54 a baixo podemos conferir a variação de energia para mesma reação calculada com diversos métodos, publicado em um artigo na Journal of Chemical Information Modeling.

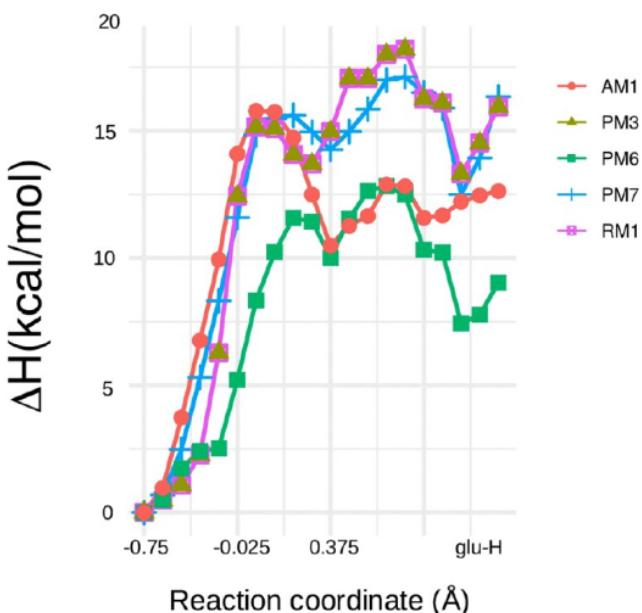


Figure 3. Variation of the enthalpy of formation for the simulated reaction coordinate geometries of the TIM catalysis first elementary step, the GLU165 residue protonation (glu-H) through the H1 extraction from DHAP.

Figura 54: Variação de energia calculada para essa reação em estudo teórico publicado por nosso grupo. DOI:10.1021/acs.jcim.9b00860

Na análise que foi publicada, boa parte dos métodos apresenta um estado de transição perto de -0.025 da coordenada de reação, quase no ponto de igual distância da transferência do próton, depois há uma vale, denotando um estado intermediário e depois mais um aumento de energia até outro vale no qual o produto deve se acomodar. Como visto na imagem acima, o PM6 é o método mais discordante em relação ao calor de formação.

Na Figura 55, nós mostramos os descriptores para o átomo de oxigênio do resíduo GLU. No script gerado pelo PRIMoRDiA nós analisamos 9 propriedade por átomo da coordenada de reação. Para o oxigênio é interessante ressaltar algumas, como o decréscimo na densidade eletrônica, atingindo um mínimo no estado de transição e depois retomando o valor nos produtos. A carga parcial aumenta também, indicando que há uma transferência de elétrons do oxigênio para o carbono, e que depois o sistema em volta faz uma "devolução" de carga para o oxigênio, mas que no final ainda fica com menos, já que agora o grupo carboxílico está protonado e portanto neutro. Já para a dureza local esse átomo apresenta comportamentos diferentes para cada tipo de método de cálculo, para interpretar o que cada um significa é importante entender cada modelo. Brevemente, a dureza baseada no LCP tem mais correção com o descriptor global de dureza, indicando que esse tinha a propensão de interagir por

processo duro-duro e depois foi diminuindo esse comportamento conforme foi transferindo carga.

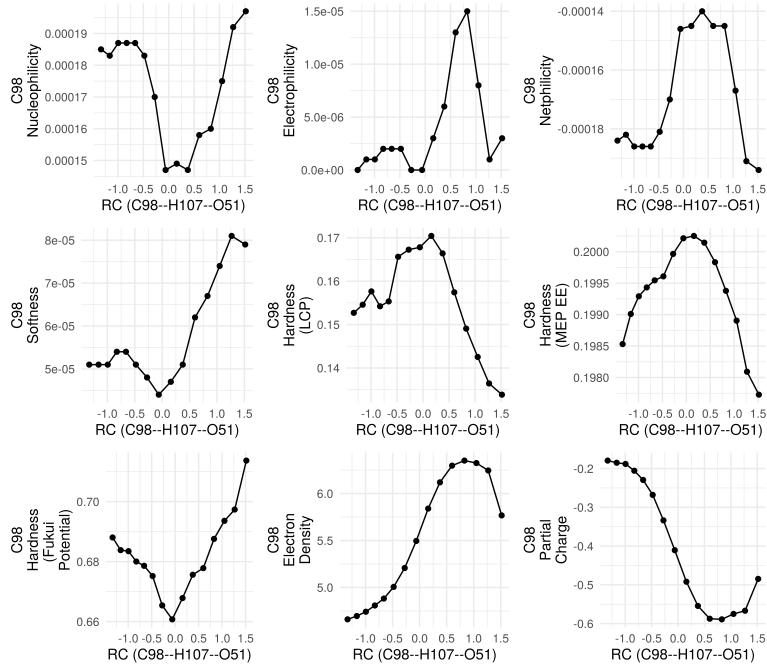


Figura 55: Descritores locais calculados para o átomo O2E do grupo carboxílico do Glutamato catalítico para a trajetória de reação.

Os descritores para o carbono C1 estão na Figura 56. Tem bastante coisa interessante para analisar nesse caso. Primeiro, como que para dois métodos de dureza o valor máximo condiz com o estado de transição e se assemelha bastante com o perfil de energia da reação. Já que a interação dominante para transferência de prótons é a duro-duro, o carbono assume o valor máximo para esse descritor, em dois métodos mepEE e LCP. O método de dureza baseado no potencial de Fukui da a força exercida nos núcleos em respeito da variação do número de elétrons, o que tem uma interpretação mais diferente da geral sobre dureza comum.

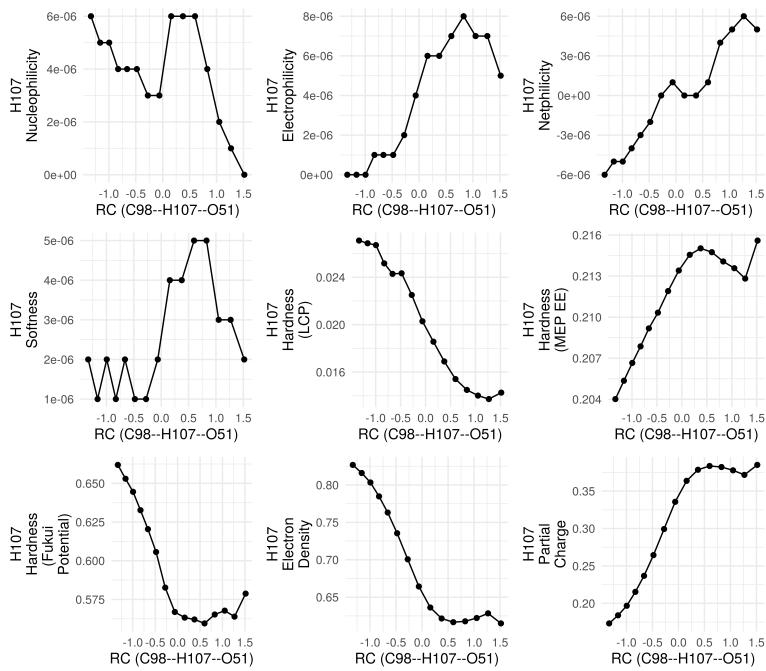


Figura 56: Descritores locais calculados para o átomo C1 do substrato para a trajetória de reação.

Os descritores para o átomo transferido, o hidrogênio do DHAP (H1), se encontram na figura de baixo. Os descritores baseados nos orbitais moleculares, como eletrofilicidade, nucleofilicidade e netfilicidade mostraram variação mas com valores muito baixos. Agora é interessante como a densidade eletrônica vai diminuindo, se tornando cada vez mais um próton, um hidrogênio carregado positivamente. A dureza local com maior ligação com as interações eletrostáticas (MEP EE), foi a que apresentou maior coerência com o processo, com um aumento das interações eletrostáticas próximo ao estado de transição, diminui e atinge o máximo nos produtos, onde a carga parcial alcança o maior valor, sendo que esse átomo tende a ser transferido pelo mesmo tipo de interação.

Como no Tutorial 3, vamos gerar imagens dos descritores para alguns pontos da trajetória. O PRIMORDIA escreve para cada arquivo um script de Pymol para automatizar a visualização dos descritores. Abra o Pymol e execute o arquivo com o prefixo "sys0" e sufixo ".pym". Como mostrado na Figura 57 e no caixa de comando abaixo.

```
@sys0_pymols.pym
```

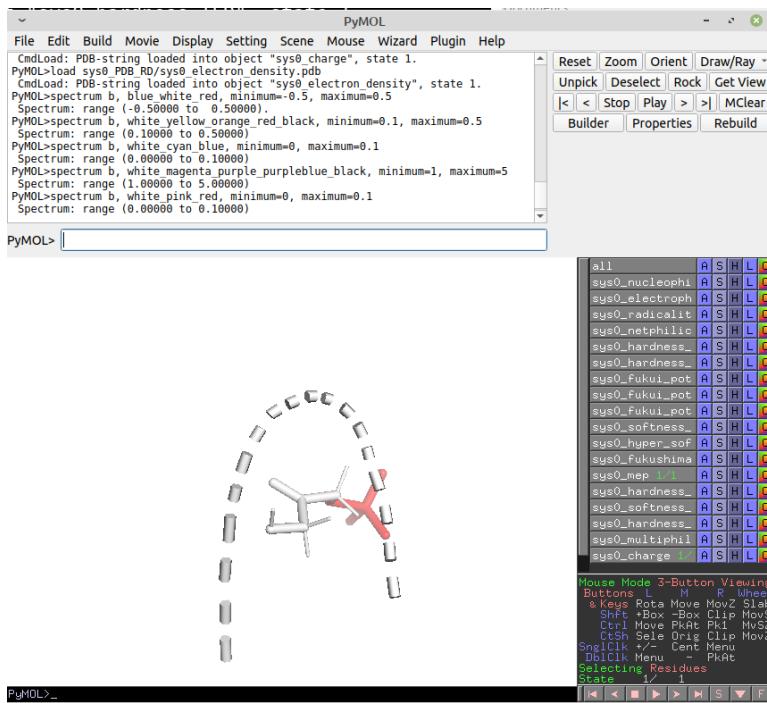


Figura 57: Janela do Pymol com os resultados dos descritores carregados e o objeto do descritor de .

As próximas figuras foram resultado do processo de renderização no pymol para o descritor de dureza local com o segundo método disponível no PRIMoRDiA. O objeto com a proteína inteira foi colorido com os carbonos em verde e escondido todas as outras representação que não fossem cartoon para não atrapalhar a visualização do objeto com o descritor. Para a palheta escolhida o comando utilizado foi

```
spectrum b, white_yellow_orange_red_black, minimum=0.1,maximum=0.2
```

Esses valores podem sofrer pequenos ajustar para melhorar o destaque dos átomo com o maior valor de dureza. Depois, com o comando "ray"seguido do "png"as imagens foram salvadas. A Figura 58 mostra a dureza local para a primeira etapa da coordenada, na Figura 59 para a sétima estrutura, que corresponde ao estado de transição, e na Figura 60 para a etapa correspondendo aos produtos.

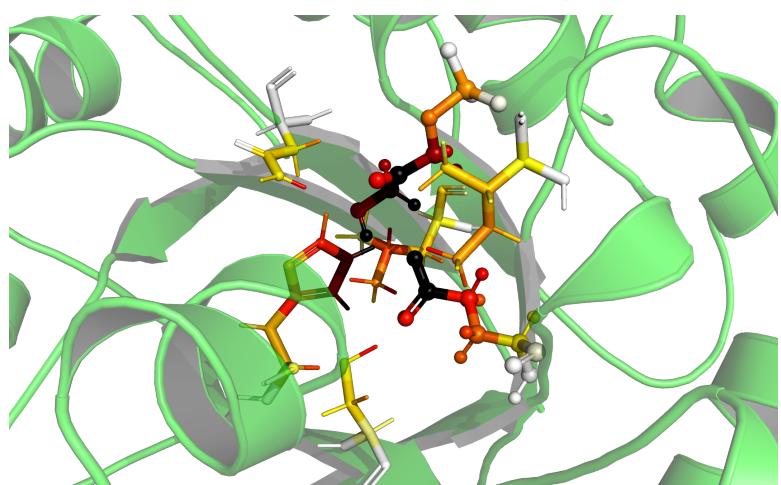


Figura 58: Dureza local para os átomos da enzima tratados com mecânica quântica para a estrutura corresponde ao primeiro passo da trajetória de reação.

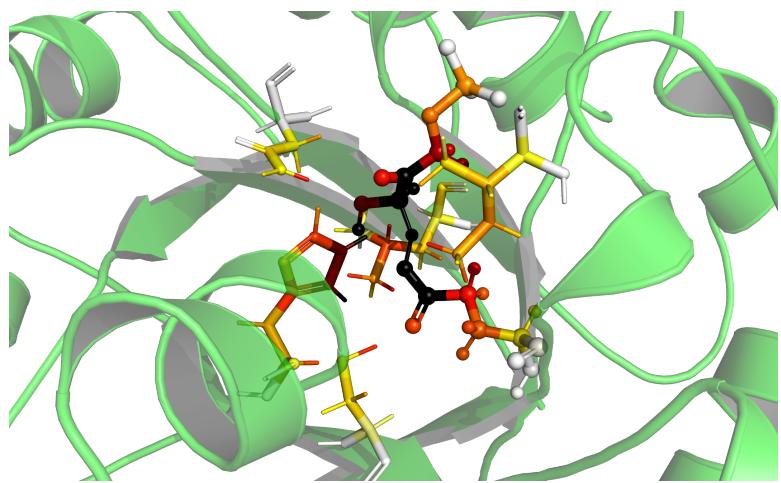


Figura 59: Dureza local para os átomos da enzima tratados com mecânica quântica para a estrutura corresponde ao estado de transição da trajetória de reação.

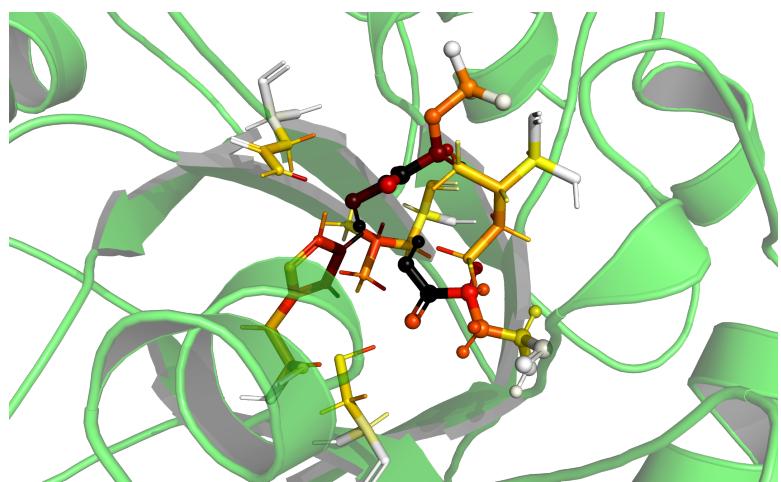


Figura 60: Dureza local para os átomos da enzima tratados com mecânica quântica para a estrutura corresponde ao estado de transição da trajetória de reação.

O PRIMoRDiA também escreve scripts especiais do R para a análise dos descritores para resíduos específicos, indicados no input, produzindo gráficos de barra com a média e o desvio padrão, como na figura abaixo para o descritor de eletrofilicidade na Figura 61, nucleofilicidade na Figura 62.

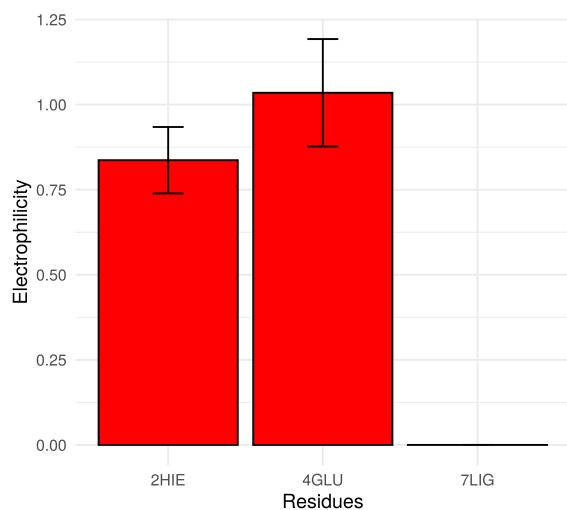


Figura 61: Gráfico de barras para a eletrofilicidade nos resíduos indicados durante a trajetória.

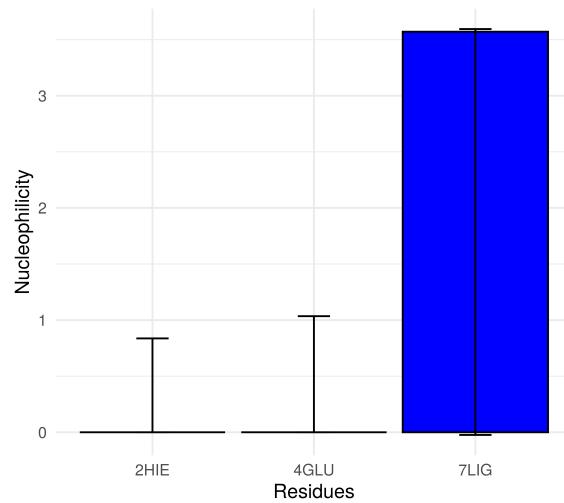


Figura 62: .

Também, é acompanhado a variação desses valores durante a trajetória com o cálculo de médias móveis, como na Figura 63 é mostrado para densidade eletrônica em cada resíduo indicado no input, para a dureza local na Figura 64 e moleza localFigura 65.

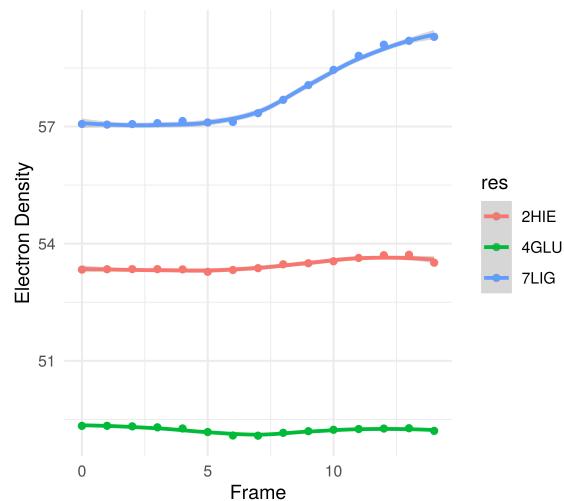


Figura 63: Médias móveis para a densidade eletrônica para os resíduos indicados no input para análise da trajetória.

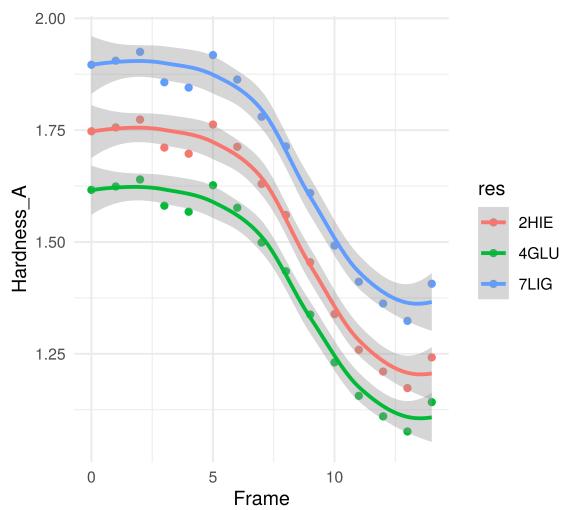


Figura 64: Médias móveis para a dureza local para os resíduos indicados no input para análise da trajetória.

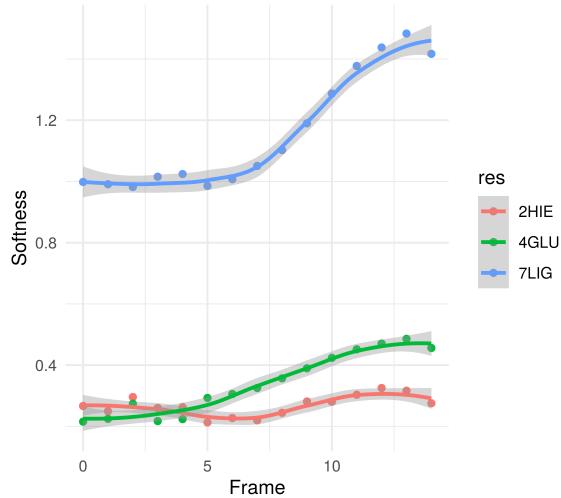


Figura 65: Médias móveis para a moleza local para os resíduos indicados no input para análise da trajetória.

Podemos ver como a densidade eletrônica e a moleza local no LIG, o substrato, aumenta. Nessa estrutura de transição formada no produto da nossa simulação, o substrato forma uma ligação dupla quando o próton é doado, e como os descritores indicam o carbono e a moléculas receberam doação de carga. O aumento da moleza pode significar que agora com a dupla formada essa molécula tem maior propensão de formar ligações covalentes através de sobreposição de orbitais moleculares e também transferir carga.

Há diversas análises que podemos fazer baseados nos descritores de reatividade, que por sua vez também são numerosos, para cada átomo envolvido e para cada resíduo. O PRIMoRDiA fornece uma

forma automática de dispor todos esses resultados da forma mais pronta para análise possível.

6 Tutorial Extra: Google Colabs

PRIMoRDiA é um software escrito para ser compilado e executado de forma simples e eficiência em ambientes Linux. No entanto, nem sempre o usuário vai ter acesso a uma máquina local com um sistema operacional baseado em Unix, e muitas vezes para fins didáticos e de teste do software o usuário não está disposto a trocar de sistema operacional. De fato, o PRIMoRDiA não pretende ser um software multi-plataforma em essência, mas com os recursos de computação em nuvem que estão disponíveis nos dias de hoje, é possível que o programa possa ser executado de qualquer dispositivo com acesso a internet.

Nesse tutorial vamos explorar o uso da plataforma Google Colabs. Para isso só é preciso ter uma conta gmail, e recursos como 12GB de ram e 100GB de storage estarão disponíveis gratuitamente. Além da questão de plataforma, também a falta de recursos locais pode ser resolvida através dessa abordagem.

Vamos exemplificar a instalação e a execução do primeiro tutorial apresentado nesse documento. O primeiro passo é com a sua conta Google entrar na plataforma colabs: <https://colab.research.google.com>. Tem tutoriais básicos lá ensinando a usar a plataforma para tarefas simples e análise de dados em Python. De fato, o Google colabs te fornece um ambiente de Python, com a mesma lógica de um notebook de Jupyter, que são formas de executar scripts interativos em Python no Browser.

No nosso diretório principal do repositório, se encontra um arquivo nomeado "Tutorial_PRIMoRDiA_Colab.ipynb". Esse arquivo pode ser aberto no Google Colab onde é possível executar todas as etapas desse tutorial. No entanto, também é possível criar seu próprio arquivo e copiar os comandos como mostrado nas próximas figuras.

Na Figura 66, vemos os comandos necessários para instalar duas bibliotecas necessárias para compilar o PRIMoRDiA, é na verdade o mesmo processo de compilação realizado em uma máquina Linux local. A grande diferença é que nesse ambiente a linguagem principal é Python3, e comandos que vão precisar rodar no "bash" (linguagem de interface entre o usuário e o sistemas baseados em Unix), precisam iniciar com um ponto de exclamação "!" ou de porcentagem "%". Basicamente para execução de binários se usa o primeiro e para comandos necessários para navegação nos arquivos se usa o último.

▼ Installation of third party libraries

```
[ ] !apt install g++-8  
!apt install libeigen3-dev
```

Figura 66: Instalando bibliotecas externa no ambiente do Google Colab.

Portanto, em uma célula de código do seu arquivo aberto no Colab, digite os mesmos comandos mostrados na Figura 66, e execute clicando no botão comumente conhecido pelo "play". Terminada essa execução, vamos clonar o repositório do PRIMoRDiA, navegar para dentro da pasta, configurar a compilação e executar ela. Esse passo a passo é mostrado na Figura 67.

Clone the Repo and Install

```
[ ] !git clone https://github.com/igorChem/PRIMoRDiA1.0v  
%cd /content/PRIMoRDiA1.0v  
!cmake .  
!make
```

Testing the compiled executable without arguments:

```
▶ ./PRIMoRDiA_1.25v
```

Figura 67: Configurando e compilando o executável.

Lembrando que no Colab, o diretório principal é o /content. Essas dados ficam na nuvem, ocupando espaço do seu Google Drive. Para acessar os dados na pasta é só acessar o menu que fica a esquerda na tela, onde também é possível selecionar os arquivos e fazer o download.

A lógica de execução do PRIMoRDiA segue igual a mostrada nos outros tutoriais. Para executá-lo não esqueça de primeiro descomprimir o diretório usando o comando "!tar -xf Tutorials_Files.tar.bz". Depois disso é possível entrar no diretório onde estão os dados para a execução do primeiro tutorial, como mostrado na Figura 68.

▼ Tutorial 1

Go to the folder where are the Tutorial Files

```
[ ] %cd /content/PRIMORDia1.0v/Tutorials_Files/Tutorial_1  
[ ] !/content/PRIMORDia1.0v/PRIMORDia_1.25v -input -op 1 -p mopac -grid 40
```

Let us see the produced input:

```
[ ] %cat < primordia.input
```

Let us use the edited input provided along the tutorial files

```
[ ] %cat < primordia_edit.input  
[ ] !/content/PRIMORDia1.0v/PRIMORDia_1.25v -f primordia_edit.input
```

List the results of the calculation

```
▶ %ls
```

Figura 68: Comandos utilizados na execução do primeiro tutorial dentro da plataforma do Google Colabs.

Depois de executado os comandos, pode demorar alguns minutos, os resultados aparecerão na mesma pasta. Para seguir fazendo a parte de análise do tutorial 1, como já descrito nesse documento, é preciso baixar os arquivos para utilizar as ferramentas gráficas na sua máquina local.

Concluindo, não há nenhuma grande diferença de execução dos tutoriais quando realizados na plataforma do Google Colabs, podendo ser uma ferramenta de grande ajuda para aprender a utilizar o software e para aulas ou cursos acadêmicos, ou até mesmo para aqueles com recursos computacionais bem limitados.