



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós Graduação em Química

**PRIMoRDiA: *Software* para Cálculos de
Reatividade e Análise de Estrutura Eletrônica para
Biomoléculas**

Igor Barden Grillo

Orientador: Prof. Dr. Gerd Bruno da Rocha

Co-orientador: Prof. Dr. Gabriel Aires Urquiza de Carvalho

João Pessoa

- 10 de Fevereiro, 2023 -



Universidade Federal da Paraíba
Centro de Ciências Exatas e da Natureza
Departamento de Química
Programa de Pós Graduação em Química

PRIMoRDiA: *Software* para Cálculos de Reatividade e Análise de Estrutura Eletrônica para Biomoléculas

Igor Barden Grillo

Tese apresentada ao Programa de Pós-graduação em Química da Universidade Federal da Paraíba, como requisito parcial para a obtenção do título de Doutor em Ciências, Área de Concentração: Físico-Química.

Orientador: Prof. Dr. Gerd Bruno da Rocha

Co-orientador: Prof. Dr. Gabriel Aires Urquiza de Carvalho

João Pessoa

- 10 de Fevereiro, 2023 -

Catálogo na publicação
Seção de Catalogação e Classificação

G859p Grillo, Ígor Barden.

Primordia : software para cálculos de reatividade e análise de estrutura eletrônica para biomoléculas / Ígor Barden Grillo. - João Pessoa, 2023.
160 f. : il.

Orientação: Gerd Bruno da Rocha.

Coorientação: Gabriel Aires Urquiza de Carvalho.

Tese (Doutorado) - UFPB/CCEN.

1. Químico-quânticos. 2. Métodos semiempíricos. 3. Reações enzimáticas. 4. Sistemas biológicos. I. Rocha, Gerd Bruno da. II. Carvarlho, Gabriel Aires Urquiza de. III. Título.

UFPB/BC

CDU 544.18(043)

PRIMoRDiA: *Software* para Cálculos de Reatividade e Análise de Estrutura Eletrônica para Biomoléculas.

Tese de Doutorado apresentada pelo aluno Igor Barden Grillo e aprovada pela banca examinadora em 10 de fevereiro de 2023.



Prof. Dr. Gerd Bruno da Rocha

DQ/UFPB

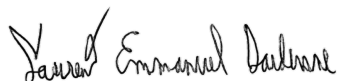
Orientador/Presidente



Prof. Dr. Gabriel Aires Urquiza de Carvalho

DQ/UFPB

Co-Orientador



Prof. Dr. Laurent Emmanuel Dardenne

COMOD/LNCC

Examinador



Profa. Dra. Kaline Rabelo Coutinho

DFG/USP-SP

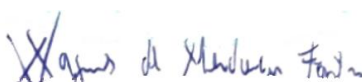
Examinadora



Profa. Dra. Karen Cacilda Weber

DQ/UFPB

Examinadora



Prof. Dr. Wagner de Mendonça Fasutino

DQ/UFPB

Examinador

Agradecimentos

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão da bolsa de fomento à pesquisa. À Universidade Federal da Paraíba, ao Departamento de Química da UFPB, e ao Programa de Pós-Graduação em Química da UFPB pela estrutura necessária para o desenvolvimento deste trabalho. Assim como ao Centro Nacional de Processamento de Alto Desempenho em São Paulo (CENAPAD-SP) e ao Laboratório Nacional de Computação Científica (LNCC), pelo fornecimento de infraestrutura computacional.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), que, através do projeto de pesquisa de "Bioinformática Estrutural de Proteínas: Modelos e Algoritmos e Aplicações Biotecnológicas"(Edital Biologia Computacional 51/2013, processo AUXPE 1375/2014), financiou importantes participações em congressos e visitas técnicas vitais ao desenvolvimento dessa tese.

Agradeço também ao meu orientador Prof. Dr. Gerd Bruno da Rocha, que me recebeu de braços abertos ao Laboratório de Química Quântica e Computacional (LQQC/UFPB) e me permitiu desenvolver concepções para essa tese; ao meu co-orientador Prof. Dr. Gabriel Aires Urquiza de Carvalho, que me ajudou desde o mestrado a idealizar o programa e suas possíveis aplicações; ao Prof. Dr. José Fernando Bachega, que por motivos externos, não foi creditado nesse trabalho como o meu segundo orientador, todavia foi um importante colaborador, atuando de forma plena nas primeiras simulações de reações enzimáticas. Aos meus colegas de laboratório (LQQC/UFPB) e ao corpo professores e administrativo do Departamento de Química da UFPB.

Qualquer lista de agradecimentos possível tornaria-se injusta, afinal, não conseguiria citar todos. Porém, deixo aqui o meu agradecimento a essas pessoas especiais: ao meu

"pai"científico, Prof. Dr. Marcus Seferin, ao Prof. Dr. Tiziano Dalla Rosa e família, Prof. André Arigony, Dr. Wagner Menezes e minha mãe, Niura Freitas Barden. Também é necessário reconhecer a ajuda com revisão do texto, dada por Mabel Abreu, e o seu suporte nessa reta final de doutorado.

Por fim, dedico esse trabalho à minha avó Odette Freitas Barden.

Muito Obrigado!

“Esse gancho que tens no braço não o inventaste tu, foi preciso que alguém tivesse a necessidade e a ideia, que sem aquela esta não ocorre, juntasse o couro e o ferro, e também estes navios que vês no rio, houve um tempo em que não tiveram velas, e outro foi o da invenção dos remos, outro o do leme, e, assim como homem, bicho da terra, se tornou marinheiro por necessidade, por necessidade se fará voador.”

– José Saramago

Resumo

Interações a nível molecular são fundamentais aos processos biológicos e entender o seu papel corretamente passa pela utilização do tratamento teórico mais apropriado, que se dá através do uso de métodos de Química Quântica. A hipótese central desta tese é sobre a capacidade de recuperar informações moleculares importantes de processos biológicos através do cálculo de descritores moleculares químico-quânticos de biomoléculas. Para isso foi desenvolvido o software PRIMoRDiA, com ferramentas específicas para lidar com as particularidades da estrutura eletrônica desses sistemas moleculares com muitos átomos, geração de *scripts* para a automatização de análises estatísticas e visualização dessas quantidades teóricas, assim como, uma implementação eficiente do código para lidar com grande volume de dados produzidos por pacotes de Química Quântica. Como teste de usabilidade dos descritores modificados, os papéis de reatividade dos resíduos de complexos enzimáticos foram estudados com o PRIMoRDiA. Para teste de aplicação, os descritores tiveram sucesso na caracterização teórica de caminhos de reações enzimáticas, onde os resultados e suas interpretações são explorados nessa tese. Foi possível observar que os descritores quânticos podem recuperar informações importantes de simulações em processos biológicos a partir de métodos de estrutura eletrônica com menor exigência computacional, como os semiempíricos. Além disso, o programa já contribuiu e vem sendo útil em diversos trabalhos, ao qual foi aplicado para estudos de interação proteína-ligante e para análise da dependência da reatividade com diferentes conformações da *Main* Protesase (MPro) do vírus SARS-CoV-2.

Palavras-Chaves: PRIMoRDiA, Descritores Moleculares Químico-Quânticos, CDFT, Sistemas Biológicos, Métodos Semiempíricos, Reações Enzimáticas.

Abstract

Interactions at the molecular level are fundamental in biological processes. Understanding the correct role played by these involves the most appropriate theoretical treatment, which takes place through quantum chemistry methods. The central hypothesis of this thesis is about the ability to recover important molecular information from biological processes through the calculation of chemical-quantum molecular descriptors of biomolecules. For this, the PRIMoRDiA software was developed, with specific tools to deal with the particularities of the electronic structure of these molecular systems with many atoms, generation of scripts for the automation of statistical analyzes and visualization of these theoretical quantities, as well as an efficient implementation of code to handle the large volume of data produced by quantum chemistry packages. As an usability test of the modified descriptors, the reactivity roles of residues from the active sites of enzyme complexes were studied with PRIMoRDiA. As an application, the descriptors were successful in the theoretical characterization of enzymatic reaction paths, where the results and their interpretations are explored in this thesis. It was possible to observe that quantum descriptors can recover important information from simulations in biological processes from electronic structure methods with less computational demand, such as semiempirical ones. In addition, the program has already contributed and has been useful in several works, in which it was applied to studies of protein-ligand interaction and for analysis of the dependence of reactivity with different conformations of the Main Protease (MPro) of the SARS-CoV-2 virus.

Key-words: PRIMoRDiA, Quantum-Chemistry Molecular Descriptors, CDFT, Biological Systems, Semiempirical Methods, Enzymatic Reactions.

Sumário

Agradecimentos	3
1 Introdução	5
2 Referencial Teórico	8
2.1 Métodos Computacionais Aplicados a Biomoléculas	9
2.1.1 Estrutura Eletrônica de Macromoléculas Biológicas	13
2.2 Teorias de Reatividade Química	16
2.2.1 Teoria do Funcional da Densidade Conceitual	18
2.2.1.1 Descritores Globais	20
2.2.1.2 Descritores Locais	24
2.2.1.3 O Problema da Dureza Local	31
2.2.2 Descritores Moleculares para Sistemas Biológicos	36
2.2.2.1 Aplicação para Fragmentos	37
2.2.2.2 Tentativa de DFT Conceitual Biológico	39
2.2.2.3 Descritores Modificados para Macromoléculas	41
3 Métodos	47
3.1 Protocolos de Dinâmica Molecular	49
3.2 Métodos Semiempíricos	51
3.3 Simulações com Métodos Híbridos de Mecânica Quântica/Mecânica Molecular	53
4 Desenvolvimento e Implementação do Software	58
4.1 Implementação	59
4.2 Usabilidade e Distribuição	63

4.3	Descritores Implementados	66
4.3.1	Descritores de Banda	73
4.4	Comparação com Outros Softwares de Cálculos de Descritores	77
4.5	Considerações Finais	80
5	Teste e Validação para Sistemas Biológicos	81
5.1	Montagem do Experimento Computacional	81
5.2	Resultados	85
5.3	Considerações Finais	95
6	Aplicação: Caracterização Teórica de Reações Enzimáticas	97
6.1	Sistemas Enzimáticos	98
6.2	Simulação dos Caminhos de Reação	100
6.3	Análise de Reatividade	104
6.3.1	Triosefosfato Isomerase	104
6.3.2	Haloalcano Dehalogenase	108
6.3.3	Adenosina Quínase	111
6.3.4	Chiquimato Desidrogenase	116
6.4	Considerações Finais	120
7	Contribuições do Software	121
7.1	Aplicações em Dissertações e Teses	121
7.2	Aplicações em Publicações em Periódicos	122
7.3	Minicursos Ministrados	127
7.4	Considerações Finais	127
8	Conclusões	128

“Aut inveniam viam aut faciam”

–Anibal de Cartago

Interações a nível molecular são fundamentais aos processos biológicos, desde reações de auto-catálise do RNA, dando origem à vida,^[1] até a formação dos tecidos orgânicos.^[2] Portanto, o estudo dos mecanismos biológicos, sobretudo as suas descrições teóricas, invariavelmente passa pela consideração da física e química de biosistemas.^[3] Indo mais além, os fenômenos quânticos são fundamentais na explicação de vários desses processos, como no caso da absorção de luz durante a fotossíntese, magneto-recepção, e em nossos mecanismos sensoriais: olfato, visão e paladar.^[4]

As enzimas são um exemplo de estruturas proteicas com impressionantes funções biológicas, sendo responsáveis por catalisar todas as reações químicas necessárias à vida em condições amenas.^[5,6] A sua ação é normalmente o objeto de estudo de novos fármacos,^[7] sendo, geralmente, a principal via de inibição de metabolismo de micro-organismos.^[8,9] Com alta especificidade e taxa de conversão, as enzimas são os modelos ideais de catalisadores, ganhando cada vez mais espaço nas aplicações industriais.^[10]

Especificamente para esses sistemas, os modelos teóricos e os métodos de simulação computacional são tão importantes quanto as técnicas experimentais. Por exemplo, a Dinâmica Molecular (DM) é largamente empregada para explicar como sistemas de proteína e DNA/RNA se comportam em ambiente fisiológico.^[11–13] Entretanto, apenas a partir da aplicação de métodos que consideram a estrutura de níveis de energia eletrônicos, oriundos da natureza quântica dos sistemas moleculares, é que se pode simular processos de quebra e formação de ligações químicas, polarização, transferência de elétrons e de carga, e a contabilização de outros efeitos.^[14,15]

Na prática, devido a questões de recursos computacionais, para estudos de simulação de reações enzimáticas são combinados os métodos de mecânica molecular, considerando campos de força clássicos (física Newtoniana), e métodos de química quântica.^[16] Esses protocolos de simulação são conhecidos como métodos Híbridos de Mecânica Quântica/-Mecânica Molecular, em inglês *Quantum Mechanics/Molecular Mechanics* (QM/MM), os quais permitem a simulação dos caminhos de reação, determinação das estruturas de estado de transição e cálculos energéticos, que servem para predições e/ou confirmações de dados cinéticos e termodinâmicos desses sistemas biomoleculares.^[17,18]

Esses métodos de simulação não somente provam dados energéticos e estruturais, mas também produzem o cálculo da estrutura eletrônica. Entre outras palavras, a função de onda resolvida, ou densidade eletrônica de onde, teoricamente, é possível extrair qualquer propriedade observável do sistema.^[19] Contudo, na grande maioria dos casos, essa informação é descartada. Para pequenas moléculas, a análise minuciosa dos níveis de energia é significativamente mais praticada, onde os valores de energia e a distribuição espacial dos orbitais moleculares são matematicamente processados para a produção de descritores de reatividade e de propriedades eletrônicas, sendo amplamente aproveitados para explicação e predição de reações.^[20,21]

Esses descritores fazem parte de um rico campo de teorias de reatividade, que evoluiu para a união de conceitos químicos com escalas quantitativas extraídas de um rigoroso tratamento matemático do DFT (*Density Functional Theory*).^[22] Mais especificamente, Parr e Yang deram início ao que é conhecido como Teoria da Densidade Funcional Conceitual (CDFT, do inglês *Conceptual Density Functional Theory*),^[23] reunindo dentro do mesmo arcabouço teórico as teorias de Ácidos e Bases Duros e Moles de Pearson,^[24] Eletronegatividade de Muliken^[25] e os índices de superdeslocalizabilidade de Fukui.^[26,27]

Para o estudo de reações de química orgânica, esses descritores vêm sendo largamente empregados desde os anos 1990 com a popularização de métodos de DFT e aumento de recursos computacionais.^[28] A origem do efeito da estrutura das enzimas na aceleração das reações ainda é alvo de vários estudos teóricos,^[29-31] e os descritores oferecem ferramentas para investigar as forças químicas que tornam a reação possível.^[32] Por isso, a expectativa do uso dessas ferramentas teóricas para sistemas biológicos existiu desde que as aplicações dos métodos de Química Quântica para esses sistemas se tornaram cada vez mais frequentes.

Ainda assim, o número de trabalhos que aplicam descritores quânticos para macromoléculas é ínfimo quando comparado ao que hoje é publicado para todos os outros tipos de sistemas. Isso porque a estrutura eletrônica desses sistemas têm características particulares,

como a ampla degenerescência nos níveis de energia próximos aos orbitais de fronteira,^[33] o qual prejudica o cálculo dos descritores locais usando a aproximação de orbitais congelados.^[34] Já para o método de diferenças finitas, o grande número de elétrons que esses sistemas comportam influenciam na devida perturbação de carga e na convergência dos estados eletricamente carregados.^[35]

Além disso, a falta de *softwares* bem estabelecidos para o cálculo desses descritores dificulta que diferentes grupos de pesquisa acessem as mesmas quantidades teóricas e com o mesmo tipo de aproximação para diferentes sistemas.^[36] Geralmente, os arquivos de saída de programas de química computacional para estruturas de proteínas inteiras são na escala de *gigabytes* e vários programas de pós-processamento de cálculos quânticos acabam apresentando grande dificuldade de carregar sistemas com 300 átomos ou mais.

Em um trabalho recente do nosso grupo de pesquisa,^[34] oriundo da dissertação de mestrado de Grillo, I.B.,^[37] ao qual o presente projeto de tese deu continuidade, nós mostramos que é possível utilizar métodos semiempíricos para obter a estrutura eletrônica necessária para o cálculo dos descritores de reatividade para um conjunto de polipeptídeos. Isso foi possível através de dois métodos de combinação dos orbitais de fronteira, que foi desenvolvido nesse trabalho.^[34,37]

O próximo passo, que foi realizado nessa tese, foi tornar essas ferramentas teóricas e computacionais, disponíveis por meio do desenvolvimento do PRIMoRDIA,^[38] em um programa que lida eficientemente com grande volume de dados, implementa todos os métodos de aproximação, representação e tipos de descritores, em conjunto com esses métodos computacionais adaptados para grandes moléculas.

O programa conta com versões estáveis disponíveis em repositório online, com guia de usuário e tutoriais (<https://github.com/igorChem/PRIMoRDIA1.0v>). Durante o desenvolvimento do trabalho, o PRIMoRDIA foi utilizado para a caracterização de sítio ativo em complexos enzima-substrato^[38] e para a avaliação da superfície de energia de reações enzimáticas.^[39,40] Além disso, o programa já contribuiu e vem sendo útil em diversos trabalhos acadêmicos, e aplicado em estudos de interação proteína-ligante^[41,42] e para análise da dependência da reatividade com diferentes conformações da MPro do vírus da SARS-CoV-II.^[43] Portanto, nesse trabalho, será abordado o desenvolvimento do *software*, suas funcionalidades e aplicações.

“Regretter un peu moins, à espérer un peu moins et à aimer un peu plus”

–André Comte-Sponville

Neste capítulo serão tratados os tópicos que dão a fundamentação teórica necessária para o desenvolvimento dos métodos e interpretação dos resultados apresentados nesse trabalho. Na primeira parte, a simulação computacional de sistemas biológicos é tratada em caráter de revisão bibliográfica, trazendo na parte introdutória informações fundamentais sobre a natureza desses sistemas, e, por conseguinte, o papel de métodos de modelagem molecular aplicados para eles. Na segunda parte, a base teórica dos descritores quânticos é desenvolvida, desde a fundamentação matemática da Teoria Funcional da Densidade Conceitual até sua aplicação para estudo de sistemas biológicos.

2.1 Métodos Computacionais Aplicados a Biomoléculas

Quando é mencionado modelagem de biomoléculas, isso inclui os grandes sistemas moleculares relevantes para os processos biológicos, como as proteínas, e seus complexos com moléculas orgânicas, e fragmentos de fitas de DNA/RNA. Esses sistemas são compostos de centenas de aminoácidos, no caso de proteínas, que podem variar em, aproximadamente, vinte tipos.^[44] Na Figura 2.1A e B são mostradas as representações moleculares das ligações peptídicas e da formação do copolímero.

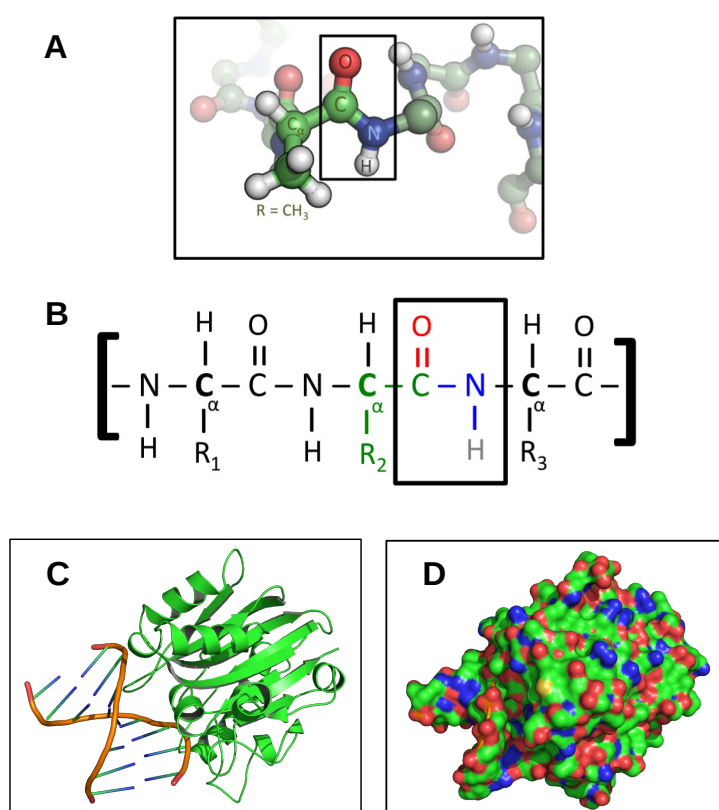


Figura 2.1: Estrutura e representações de polímeros biológicos. A) Ligação peptídica mostrada em detalhe para a representação tridimensional dos átomos. Fonte: (https://www.tutorialspoint.com/biology_part2/biology_proteins.html) B) Ligação peptídica mostrada em detalhe na formação do copolímero biológico. Fonte: (<https://byjus.com/chemistry/food-chemistry/>) C) Representação *cartoon* da estrutura terciária da 1-deoxyribonuclease (código no PDB:1DNK). Fonte: Imagem do autor. D) Representação da superfície dos átomos, usando raio de Van Der Walls, para a 1-deoxyribonuclease. Fonte: Imagem do autor.

Essas estruturas macromoleculares estão presentes em todos os organismos vivos e

também em vírus, tendo estruturas e funções únicas que dependem de como as cadeias se enovelam em ambiente aquoso, o que por sua vez é dependente dessa sequência de aminoácidos.^[45] Na Figura 2.1 são mostradas duas representações gráficas dessas cadeias já enoveladas, conhecida como *cartoon* na Figura 2.1:C), e superfície determinada pelo raio de Van Der Waals dos átomos na Figura 2.1:D). As proteínas assumem as mais diversas funções nos organismos vivos, como de transporte, sinalização, estrutural e catálise de reações químicas necessárias à vida.^[46]

Essa última é desempenhada pelas enzimas, que agem através de suas interações com substratos para aumentar a taxa de conversão em produtos, diminuindo a energia de ativação da reação química, como ilustrado na Figura 2.2.^[5,6] Isso permite que processos químicos nos organismos vivos ocorram em uma escala de tempo que possibilite sua manutenção e reprodução. Das características que mais se destacam nas enzimas em relação a catalisadores sintéticos, estão sua alta especificidade e capacidade de catalítica em condições de temperatura amenas.^[3]

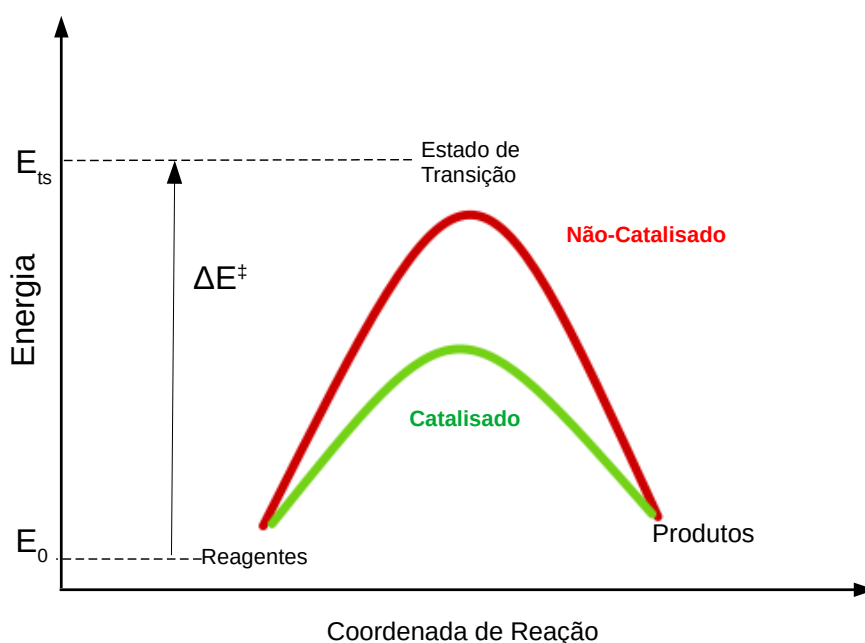


Figura 2.2: Esquema ilustrando o efeito das enzimas no abaixamento de energia de ativação de reações químicas. Fonte: Imagem do autor.

Devido a essas características, as enzimas são em grande maioria os alvos de drogas terapêuticas. É comum que fármacos utilizados no combate de microorganismos bacterianos ajam inibindo a enzima que catalisa a reação de formação de membrana celular impedindo sua reprodução e causando sua eliminação.^[47] Entretanto, devido a alta especificidade das enzimas pelo seus substratos naturais, o desenvolvimento de drogas exige um volume

significativo de informações sobre a estrutura e interações enzima-substrato, e do próprio mecanismo de catálise promovido pela enzima.^[48]

Desde a determinação da estrutura tridimensional da lisoenzima da clara do ovo em 1965,^[49] técnicas experimentais vêm sendo amplamente utilizadas para o estudo desses sistemas. Difração de raios X e ressonância magnética nuclear são as principais técnicas aplicadas para resolução estrutural e acompanhamento de estruturas intermediárias dos ciclos catalíticos,^[50] assim como estudos cinéticos que permitem a determinação de constantes de velocidade e propriedades termodinâmicas.^[51] Ademais, os mecanismos podem ser investigados através de estudos de mutações, os quais revelam o papel de aminoácidos específicos que definem o sítio ativo da enzima.^[52]

Entretanto, há diversas limitações encontradas nas abordagens experimentais, fazendo com que tentativas de determinação de mecanismos propostos por via experimental tenham se mostrado falhas.^[53,54] Além disso, a falta de teorias que expliquem sistematicamente a ação das enzimas no aprimoramento das constantes cinéticas de reação, evidencia a necessidade de complementar os experimentos com métodos de física e química computacional.^[55]

Dentre essas limitações estão: i) a determinação errônea das etapas limitantes das reações, devido ao negligenciamento da dinâmica dessas estruturas em solução, incluindo grandes mudanças conformacionais as quais, muitas vezes, são as etapas mais demoradas; ii) determinação de estruturas de estados de transição e intermediários das reações, devido ao tempo de vida extremamente curto dessas estruturas; iii) limitações experimentais, onde não é contabilizado o efeito que as mutações podem introduzir na estrutura terciária e geometria do sítio ativo, que pode alterar a interação com os substratos.^[56]

Hoje em dia, métodos de modelagem molecular são tão usuais para estudar esses sistemas quanto métodos experimentais. Simulações de dinâmica molecular, por exemplo, podem determinar uma série de informações termodinâmicas, mecânicas, e de interação com outras moléculas nas requeridas condições de temperatura, pressão, pH e composição do ambiente químico.^[11-13,57] Entretanto, esses métodos de amostragem são geralmente utilizados com campos de força clássicos, isto é, consideram os átomos como esferas rígidas, ligados entre si como um sistema de massa-mola com cargas elétricas fixas. Para considerar efeitos de polarização e quebra e formação de ligações químicas, e outras interações de origem eletrônica e de spin, é necessário considerar a natureza quântica dos sistemas.^[14,15] Para isso, é utilizado métodos de Química Quântica, onde o sistema é descrito pela função de onda ou densidade eletrônica, do qual podem ser retiradas todas as informações observáveis através da aplicação de operadores matemáticos.^[19]

No entanto, o processo de resolução da função de onda, através do método de campo auto-consistente, depende de processos de diagonalização de matrizes e cálculos de integrais multidimensionais, no qual o custo computacional cresce exponencialmente com o número de elétrons ou funções que descrevem os níveis de energia correspondentes.^[22] Para sistemas biológicos, que geralmente contêm um número de átomos muitas vezes maior que moléculas comuns na química orgânica/química fina, faz com que esses cálculos de estruturas inteiras se tornem bastante custosos.^[58] Por isso, os cálculos de estrutura eletrônica de sistemas biológicos frequentemente são feitos usando duas estratégias: aplicação de métodos de baixo custo computacional, tais como os semiempíricos, onde várias dessas integrais que compõem as matrizes são substituídas por valores calibrados para reproduzir propriedades experimentais;^[58] ou aplicação dos cálculos somente para fragmentos de interesse do sistema, considerando ou não o resto do sistema por meio de interações coulômbicas se valendo de métodos híbridos de mecânica quântica/mecânica molecular.^[16]

Particularmente, os métodos semiempíricos tiveram, e ainda têm, grande importância para a investigação da estrutura eletrônica de sistemas biológicos. Estudos recentes vêm demonstrando a sua utilidade e a vitalidade desses métodos para grandes moléculas, como os estudos de parametrização específica para transferência de elétrons em ácidos nucleicos;^[59,60] aplicação em protocolos de cálculos de energia livre para reações químicas através do método *umbrella sampling*;^[12,61] interação proteína-ligante para o desenvolvimento de inibidores;^[62–64] simulação de reações químicas para sistemas biológicos em estado excitado;^[4] descrição acurada de interações não-covalentes com aplicações em bioquímica;^[65] buscas conformacionais de carboidratos;^[66] aplicação para o problema biológico clássico do enovelamento de proteínas^[67]

Já os métodos híbridos de QM/MM são praticamente o estado da arte na simulação de caminhos de reação enzimáticos,^[68] o qual resultou no prêmio Nobel de Química concedido pela Academia Sueca de Ciências, em 2013, para Karplus, Levitt e Warshel, devido a sua grande contribuição no desenvolvimento de métodos multiescala para simulação de sistemas complexos.^[69] As principais contribuições desses protocolos computacionais híbridos são: a proposição de mecanismos de reação enzimática através da estabilização eletrostática do estado de transição dado pelo sítio ativo da proteína;^[70] a contabilização do efeito de tensionamento do substrato pela enzima,^[71] que geralmente reduz somente em 10% a energia de ativação em relação a estabilização eletrostática já mencionada.^[72]

A partir dessas observações, o estudo dos mecanismos com essas técnicas vem sendo utilizado para o teste de duas hipóteses mais populares na literatura para explicar esse

abaixamento de energia de ativação: a desestabilização do substrato, aproximando a sua correspondente estrutura de transição; e a estabilização de cargas da estrutura de transição pelo sítio ativo. De acordo com a primeira hipótese citada anteriormente, a enzima, por meio de seu sítio ativo, condiciona as estruturas dos reagentes a assumirem o que é chamado em inglês de "*Near Attack Conformation*",^[71] que em tradução livre é Conformação de Quase Ataque. Dessa forma o poder catalítico das enzimas poderia ser explicado através da restrição das conformações possíveis dos substratos, tensionando a estrutura e provocando sua evolução para o estado de transição.^[73]

Outras aplicações relevantes que se abrem com o uso de métodos de química quântica para enzimas estão bem representadas nas simulações de QM/MM da enzima citocromica P450cam, enzima bacteriana que catalisa a hidroxilação estereoespecífica da molécula de canfora:^[74] Estimativa dos efeitos da estrutura da proteína na distribuição da densidade de spin na porfirina e dos ligantes de enxofre através de ligações de hidrogênio dos resíduos do sítio ativo; predição de propriedades espectroscópicas; ocupação de orbitais em estados excitados; energia de ligação de hidrogênio que mantém o substrato no sítio ativo; efeitos de mutação na rede de ligação de hidrogênios, estimando sua importância no abaixamento da energia de ativação e portanto capacidade catalítica.

As contribuições da dinâmica das proteínas é relativamente baixa, mas existe e é contabilizada pelo coeficiente de transmissão.^[4] Assim como efeitos de tunelamento quântico que podem ser significativos, já que grande parte das reações dos substratos com os aminoácidos durante a catálise são transferências de prótons,^[75] ao qual, dependendo da distância entre o doador e o aceptor, apresentam barreiras de energia pequenas o suficiente para permitir o tunelamento.^[76] O processo de transferência de prótons é vital no funcionamento das enzimas e há hipótese de que elas evoluíram para serem mais ativas em temperatura fisiológica, onde o movimento térmico é ideal para a uma taxa de tunelamento apropriada para as reações que elas catalisam.^[77]

2.1.1 Estrutura Eletrônica de Macromoléculas Biológicas

Apesar das dificuldades da aplicação de métodos de química quântica para sistemas biológicos, é possível citar algumas análises que provém desses cálculos para o entendimento das características particulares da sua estrutura eletrônica. Devido a essa composição particular desses polímeros biológicos, várias de suas partes possuem propriedades eletrônica similares, contribuindo para a formação dos orbitais moleculares praticamente da mesma

forma. Isso faz com que grandes moléculas biológicas tenham uma alta degenerescência nos níveis de energia de fronteira, os chamados HOMO (*Highest energy Occupied Molecular Orbital*. Pt.: *orbital ocupado de maior energia*) e LUMO (*Lowest energy Unoccupied Molecular Orbital*. Pt.: *orbital virtual/desocupado de menor energia*).^[33] Essa alta degenerescência dificulta a convergência dos cálculos de SCF (*Self Consistent Field*. Pt.: *Campo Auto-Consistente*),^[78] e quando acoplado ao grande número de átomos desses sistemas, os tempos de simulação de química quântica se tornam impraticáveis. Portanto, é mais comum encontrar cálculos de propriedades termodinâmicas e cinéticas considerando somente um fragmento do sistema.^[79]

Ohno e Colaboradores^[80] aplicaram o Hamiltoniano semiempírico AM1^[81] em combinação com o MOZYME^[82] para estimar os efeitos do solvente na estrutura eletrônica das proteínas. Ainda, o método de solvente implícito COSMO^[83] foi aplicado também. Nesse estudo, eles concluíram que os efeitos de solvente na estrutura da Ribonuclease T1 (PDB:4GSP) muda significativamente o ordenamento dos orbitais moleculares e seus valores de energia. A consequência mais importante foi que os orbitais moleculares de fronteira mudam sua localização para os resíduos do sítio ativo, mostrando a influência do ambiente aquoso na descrição da reatividade desses sistemas.

Um estudo seguinte, realizado por Fukushima e colaboradores,^[33] expandiu significativamente a análise de estrutura eletrônica de enzimas, aplicando o mesmo protocolo computacional ao último estudo citado; porém, dessa vez, para 112 proteínas selecionadas, de forma a representar todas os tipos de catálise enzimática catalogadas no PDB na época. Uma das principais conclusões desse estudo foi que de fato os orbitais moleculares de fronteira estarem relacionados com a reatividade dos sistemas biológicos, e não somente os orbitais HOMO e LUMO. Devido ao número de átomos que podem ser relevantes para o sítio ativo, conforme o tamanho das enzimas aumenta, esses sistemas podem apresentar um grande número de orbitais moleculares importantes para a sua reatividade, estando relacionado a sua importância com o quão similar é o seu valor de energia em relação ao HOMO e LUMO.

Esse valor de energia vai determinar o tamanho da banda de orbitais moleculares de fronteira a ser considerada e vai estar relacionada com a especificidade da enzima, sendo as mais específicas apresentando uma reatividade mais localizada e, portanto, uma banda de orbitais moleculares menor. Segundo o estudo do Fukushima, essa banda de energia pode variar de 1 eV até 5 eV de diferença entre o valor de HOMO, para os orbitais moleculares ocupados, e LUMO para os virtuais. Esse estudo também mostrou a importância do cálculo

de estrutura eletrônica considerando todos os átomos da proteína, o efeito do solvente e a capacidade dos métodos semiempíricos de explorar as propriedades eletrônicas de grandes sistemas.

2.2 Teorias de Reatividade Química

A racionalização da reatividade das substâncias a nível molecular é o centro da ciência química. Através das leis da termodinâmica e equações da cinética de reação química, é possível determinar se uma reação vai acontecer ou não, a proporção de formação dos produtos, parâmetros ótimos e sua velocidade. No entanto, essas teorias não explicam o que leva uma substância com determinado arranjo molecular a reagir com outra. Leis ou modelos teóricos que pudessem descrever as forças que regem os fenômenos a nível molecular seriam de grande utilidade na previsão e desenho de novos processos químicos. Nesse contexto, as teorias de reatividade são o conjunto de regras que permitem relacionar uma estrutura molecular com a reatividade, capacidade desse sistema doar ou receber elétrons e propriedades relacionadas com regiões e/ou átomos específicos.^[84]

A teoria de ácido e base de Lewis foi umas das primeiras tentativas de teoria de reatividade, que depois da formulação da mecânica quântica começou a utilizar da análise da simetria e ocupação de orbitais moleculares localizados para prever reações entre bases, doadores de pares de elétrons, ácidos e receptores.^[85] No geral, as reações químicas consistem na quebra e formação de ligações entre os átomos, sendo que sua configuração nuclear tende a não mudar, variando somente a quantidade e distribuição de seus elétrons. Portanto, muitas reações eram racionalizadas somente pela análise da diferença de eletronegatividade dos compostos, popularizando alguns métodos de cálculo como os de Mulliken^[86] e Sanderson,^[87] fazendo com que começasse a ser mais utilizada uma escala de reatividade global atribuída as estruturas moleculares sem necessidade prévia de determinar qual seria o outro reagente.

Com essa mesma motivação, Pearson iniciou os estudos de princípios que pudessem determinar com melhor acurácia qual ácido teria mais afinidade com qual base, visto que a teoria de Lewis não promove ferramentas, se não a própria medida experimental, para determinar uma escala de reatividade.^[88] Através de cruzamento de dados experimentais com propriedades atômicas e moleculares, Pearson definiu de forma geral que ácidos duros tendem a reagir mais com bases duras e ácidos moles com bases moles. Na concepção dessa teoria, o que define um ácido e base duros são as características moleculares relacionadas a sua pouca polarizabilidade, como sistemas com alta densidade de carga, razão carga líquida por volume, que portanto tendem a apresentar uma densidade eletrônica mais rígida, o que representa baixa tendência para transferência de carga. Já o que define um ácido e base moles é exatamente o oposto, sistemas pouco carregados, ou com baixa densidade de carga,

mais polarizáveis e apresentando transferência de carga da base para o ácido.

Por sua vez, Fukui utilizou-se de métodos de química quântica semiempíricos, mais precisamente o modelo de Hückel, estimando somente os elétrons π de moléculas orgânicas com ligações insaturadas^[89,90] em conjunto com análise da perturbação da energia, para definir regras de reatividade. O principal resultado desses estudos foram os índices de superdeslocalizabilidade, determinando que a densidade eletrônica tende a se deslocar do HOMO do sistema doador de elétrons para o LUMO do aceptor.^[91] Entretanto, os primeiros índices desenvolvidos por Fukui ignoram a contribuição de interação eletrostática, que foi incluída no tratamento de perturbação polieletrônico feito por Klopman e colaboradores.^[92]

Através da análise das parcelas eletrostáticas e de transferência de elétrons, Klopman e colaboradores definiram duas classes de reações químicas, as controladas por carga e as controladas por orbital. A estabilização da energia em reações controladas por orbital é determinada pela transferência de elétrons do orbital molecular de mais alta energia do doador, o mesmo que o HOMO, para o orbital molecular virtual do aceptor de menor energia, LUMO, corroborando o que Fukui já tinha desenvolvido. Esse processo de transferência ocorre dessa forma pois são os orbitais moleculares que apresentam menor penalidade energética, que para o HOMO é o Potencial de Ionização (PI) e para o LUMO é a Afinidade Eletrônica (AE), por isso as energias desses níveis eletrônicos podem ser utilizados como aproximação para o PI e o AE.

Nessas reações controladas pelos orbitais, a estabilização de energia é maximizada quando a energia do HOMO do doador de elétrons é a mais próxima possível da energia do LUMO do aceptor, assim como a sobreposição dos mesmos orbitais, estimada pelo produto dos coeficientes dos orbitais atômicos e da matriz de sobreposição. Essas últimas informações são relevantes pois a análise de orbitais localizados proposta por Lewis tem sua lógica em parte preservada, já que a necessidade de maximizar a sobreposição dos orbitais moleculares significa nada menos que a região espacial em que os átomos vão se mover para formar a ligação química.

Já para as reações controladas por carga, a estabilização de energia depende basicamente da aproximação de cargas contrárias, mais comuns em reações químicas envolvendo íons, e o balanceamento das repulsões entre os elétrons e núcleos. Baseado nesses princípios, foi possível determinar quais características as moléculas devem apresentar para ter uma maior probabilidade de reagir entre si. Klopman então associou essas informações com o princípio de Pearson de ABDM,^[92] entendendo que os ácidos e bases duros reagem controlados por cargas, e que os ácidos e bases moles reagem controlados por orbitais. Portanto, a reação

entre um ácido duro e uma base mole, por exemplo, não gera uma estabilização significativa nem na parcela eletrostática, pois o ácido duro não encontra pontos de carga concentrada na base mole, e nem uma estabilização no termo de transferência de carga, já que a energia do orbital HOMO da base mole é muito distante do LUMO do ácido duro.

Conforme foram se desenvolvendo e ganhando popularidade, as teorias de reatividade de Pearson, Fukui e Klopman foram convergindo e se complementando. No entanto, essas teorias dependiam de modelos teóricos defasados, e a contabilização da reatividade de uma molécula ainda dependendo do outro reagente. Em 1978, Parr e Yang fizeram a primeira publicação do que começou a ser chamado de *Conceptual Density Functional Theory* (CDFT) (tradução: Teoria do Funcional da Densidade Conceitual),^[25] definindo escalas quantitativas para conceitos químicos com uma fundamentação matemática sólida. Com isso todas as teorias de reatividade citadas até aqui podem ser colocadas sob o guarda-chuva da CDFT,^[23] e, portanto, onde todos os descritores quânticos mais relevantes são definidos.

Nas próximas subseções será apresentado a base matemática dos descritores de reatividade da CDFT, assim como suas interpretações, aplicação desses descritores para estudo de caminhos de reação para moléculas pequenas, e aplicação desses descritores para estudar sistemas grandes relevantes a processos biológicos.

2.2.1 Teoria do Funcional da Densidade Conceitual

Os descritores de reatividade são quantidades globais ou funções locais que indicam o quanto a energia ou a densidade eletrônica variam durante algum processo de interação química.^[93] Portanto, essas quantidades são taxas, mais precisamente derivadas da energia eletrônica e de suas expansões em séries de Taylor.^[94] Dentro da DFT, a energia eletrônica E é função da variação do número de elétrons N e do potencial externo v , sendo escrita em forma de uma diferencial total (Equação 2.1), que por sua vez é conhecida como a equação fundamental do DFT. Isso porque através dela é definida as condições variacionais do sistema que permitem a minimização da energia eletrônica e a determinação da densidade eletrônica, $\rho(r)$, do estado de referência.

$$dE = \mu dN + \int \rho(r) dv(r) dr \quad (2.1)$$

Por sua vez, a função de energia $E[\rho]$ é uma funcional da densidade eletrônica ρ , dada pela Equação 2.2, que é dividida em duas contribuições: a funcional de energia eletrônica total $F[\rho]$ e a interação da densidade eletrônica com o potencial externo. As funcionais são

um tipo específico de função que toma como argumento outras funções. É nisso que o DFT se estabelece, tendo todas as suas funções que entram nos métodos variacionais para a determinação do estado de referência sendo funcionais da densidade eletrônica, que, por sua vez, é uma função das coordenadas tridimensionais r .

$$E[\rho] = F[\rho] + \int \rho(r)v(r)dr \quad (2.2)$$

A funcional de energia eletrônica total $F[\rho]$ é composta por outras funcionais que representam os termos de energia relacionado aos elétrons do sistema, como a funcional de energia cinética $T[\rho]$, funcional da energia de repulsão elétron-elétron $J[\rho]$, e a funcional de energia de troca-e-correlação $E_{xc}[\rho]$, dado na Equação 2.3.

$$F[\rho] = T[\rho] + J[\rho] + E_{xc}[\rho] \quad (2.3)$$

Portanto, a densidade eletrônica é a quantidade central, dada a distribuição dos elétrons no sistema e se integrado em todo o espaço tridimensional, tem de resultar no número total de elétrons N , como mostrado na Equação 2.4. Na DFT, a partir da densidade eletrônica do estado de referência, é possível se extrair todas as propriedades do sistema, e, portanto, as variações na densidade eletrônica são consideradas como respostas do sistema a perturbações que levam a reorganização da densidade eletrônica.

$$N = \int \rho(r)dr \quad (2.4)$$

Para determinar a densidade eletrônica do estado de referência, a Equação 2.2 deve ser minimizada, ou seja, deve ser encontrada a densidade eletrônica que zera a diferencial total dada na Equação 2.1. Esse processo é dado pela busca do multiplicador de Lagrange μ , definido na Equação 2.5.

$$\mu = \left(\frac{\partial E}{\partial N} \right) = v(r) + \frac{\delta F}{\delta \rho(r)} \quad (2.5)$$

Como definido na Equação 2.5, μ é a derivada da energia em relação ao número de elétrons, se tornando o primeiro descritor de reatividade, já que representa a estabilização energética quando o sistema tem um incremento/decréscimo no número de elétrons, ou seja, processos de interação/reação química, mudança no estado de oxidação e entre outros.^[25] Esse descritor foi nomeado de potencial químico eletrônico (PQE), pela analogia com a quantidade da termodinâmica clássica, potencial químico molecular, já que quando dois

sistemas se aproximam há a transferência de elétrons do sistema com maior μ para o de menor até que a diferença zere.^[95]

Nas próximas subseções desse capítulo serão exploradas as deduções matemáticas e interpretações teóricas dos descritores, diferenciando eles entre os globais e locais. Logo após, as aplicações dos descritores derivados na CDFT para macromoléculas biológicas serão alvo de revisão e análise.

2.2.1.1 Descritores Globais

O PQE é um descritor global, ou seja, o valor dele deve ser o mesmo em todo o sistema molecular considerado. Isso gera diversas implicações que vamos explorar na subseção relacionada aos descritores locais. Porém, de fato, como o PQE é um indicador de equilíbrio no que se refere a transferência de elétrons, se houvesse partes do sistema com valores diferentes então se esperaria que também ocorresse um fluxo de elétrons de uma região para outra. Portanto, essa é a principal propriedade relacionada aos descritores globais, que no CDFT são quantidades sempre relacionadas com a variação da energia do sistema em processos de transferência de elétrons. Para a obtenção do PQE é utilizado o método de diferenças finitas, devido a descontinuidade no número de elétrons, sendo necessário calcular a energia do sistema carregado negativamente e positivamente, como dado na Equação 2.6.

$$\mu = \left(\frac{\partial E}{\partial N} \right)_v = \frac{E_{N-1} - E_{N+1}}{2} \quad (2.6)$$

Os cálculos em outros estados de carga são realizados na geometria molecular fixada do estado de referência, obedecendo a restrição de potencial externo constante. Esse descritor acaba também por ser uma média entre o PI e a AE, que representam, respectivamente, a energia necessária para extrair um elétron e a energia liberada quando o sistema recebe um elétron. O PI e o AE podem ser calculados pela diferença de energia final e inicial desses processos, ou como demonstrado pelo teorema de Koopmans,^[96] o PI é aproximado pela energia do HOMO (Equação 2.7) e a AE pela energia do LUMO (Equação 2.8).

$$PI = E_{N-1} - E_N = -E_{HOMO} \quad (2.7)$$

$$AE = E_N - E_{N+1} = -E_{LUMO} \quad (2.8)$$

Logo, o PQE pode ser definido em função das energias desses orbitais também, como mostrado na Equação 2.9, podendo ser obtido diretamente com um único cálculo de estrutura eletrônica para a configuração molecular de referência.

$$\mu = \frac{E_{N-1} - E_{N+1}}{2} = \frac{E_{HOMO} + E_{LUMO}}{2} \quad (2.9)$$

Para a expansão de Taylor truncada em segunda ordem, a variação de energia em respeito ao número de elétrons no sistema é dada pela Equação 2.10, onde surge a segunda derivada da energia em respeito ao número de elétrons, definida na Equação 2.11, que também é a derivada do PQE. Essa quantidade representa a resistência do sistema de transferir elétrons e, por isso, foi associada ao conceito de dureza química η definido por Pearson.^[85]

$$\Delta E = \mu \Delta N + \frac{1}{2} \eta \Delta N^2 \quad (2.10)$$

$$\eta = \frac{\partial \mu}{\partial N} = \left(\frac{\partial^2 E}{\partial N^2} \right)_v \quad (2.11)$$

Para o sistema em equilíbrio, e, portanto, sem transferência líquida de elétrons para ocorrer, com $\Delta N \rightarrow 0$, a relação do PQE com a dureza é dada pela Equação 2.12, definindo o cálculo de η para diferenças finitas e aproximação de Koopmans na Equação 2.13. Aproveitando a definição da dureza química, a moleza química é dada pelo seu recíproco, resultando na derivada segunda da variação do número de elétrons em respeito a energia, ou a derivada do número de elétrons em respeito ao PQE (Equação 2.14).

$$\mu = -\frac{1}{2} \eta \quad (2.12)$$

$$\eta = E_{N-1} + E_{N+1} - 2E_N = E_{LUMO} - E_{HOMO} \quad (2.13)$$

$$S = \left(\frac{\partial^2 N}{\partial E^2} \right)_v = \left(\frac{\partial N}{\partial \mu} \right)_v = \frac{1}{\eta} \quad (2.14)$$

Na Figura 2.3 é mostrado um diagrama genérico da energia do sistema em função da variação do número de elétrons, ilustrando os estados de carga utilizados para o cálculo dos descritores globais usando o método de diferenças finitas. Como mostrado na Figura 2.3, a remoção de um elétron é acompanhada por um acréscimo de energia ao sistema, e, quando o sistema recebe um elétron, ele tende a liberar energia. De forma geral a primeira energia

é maior que última citada. Então, moléculas duras tendem a apresentar uma energia de retirada desse primeiro elétron maior que moléculas mais moles, indicando que o sistema oferece uma maior resistência a variação no número de elétrons, maior dureza, assim como se houver a transferência desse elétrons a variação positiva de energia vai ser grande e dada pelo PQE. Já o valor de energia do estado $n + 1$, a AE, tende a influenciar pouco no PQE mas para a dureza pode indicar que o sistema altera significativamente o PQE em relação ao número de elétrons.

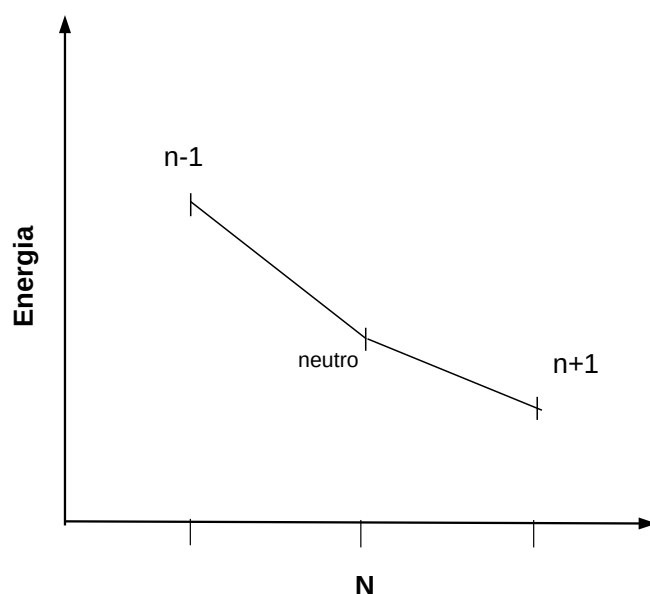


Figura 2.3: Esquema genérico mostrando a variação de energia do sistema conforme o número de elétrons N varia, enfatizando os três estados de carga necessários para o cálculo de diferenças finitas. Fonte: Imagem do autor.

A análise desses descritores a partir das energias dos orbitais permite inferências da reatividade em função da estrutura de níveis de energia eletrônica. A Figura 2.4 ilustra de forma genérica a relação dos valores de energia dos orbitais moleculares de fronteira com o PQE e a dureza química. A principal característica de moléculas moles é a baixa diferença de energia eletrônica dos nível de energia ocupado de mais alta energia, o HOMO, e o virtual de mais baixa energia, o LUMO. As moléculas moles são mais polarizáveis, o que significa que sua densidade eletrônica é mais deformável, e, portanto, os elétrons acessam outros níveis com menor penalidade energética. Isso também significa que um elétron no HOMO tem mais facilidade de ser promovido para o LUMO, o que pode ocorrer somente com a energia provida pelas flutuações térmicas, ou pela incidência de uma radiação de menor número de onda. Logo, uma molécula mole tende a mudar seu estado eletrônico

mais facilmente se tornando mais reativa.^[97]

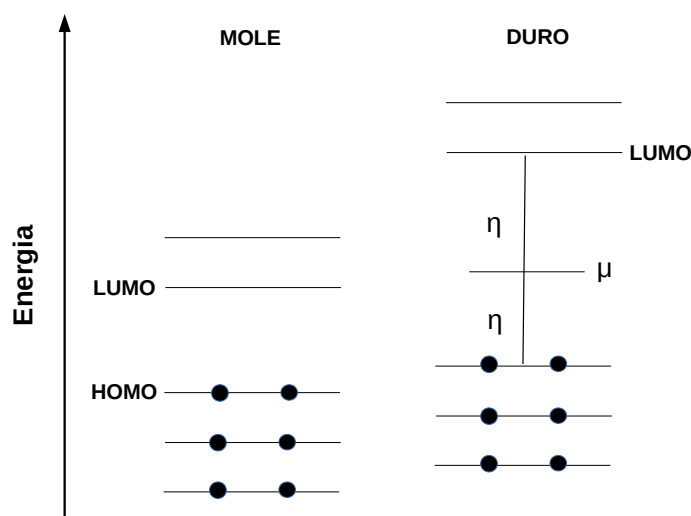


Figura 2.4: Esquema genérico ilustrando a diferença de energia entre orbitais ocupados e virtuais nos casos de sistemas considerados moles e duros. Fonte: Imagem do autor.

Já para moléculas mais duras a diferença HOMO-LUMO é bem maior, inviabilizando a formação de um estado eletrônico excitado e a transferência de elétrons do HOMO de uma molécula dura para o LUMO da outra molécula dura. Por isso é dito que as interações duro-duro são controladas por carga, já que a transferência de elétrons é mínima e vai ocorrer principalmente interações eletrostáticas. E a partir da Figura 2.4 também podemos ver que há uma grande distância da energia do HOMO de uma molécula mole para o LUMO de uma molécula dura, o que acaba por ser explicado o princípio de Pearson de ácidos e bases duros e moles.

Até aqui foram definidos os descritores globais principais que podem ser extraídos de derivadas de primeira e segunda ordem da energia. A partir dessas quantidades é possível definir uma escala quantitativa para moléculas sem predefinir os reagentes. Também é possível definir quantidades de transferência de elétrons, ignorando os efeitos eletrostáticos e da variação do potencial externo. Como por exemplo, o sistema quando em contato com um doador ideal de elétrons, de potencial químico igual a zero, se satura de elétrons e apresenta a devida estabilização de energia dada na Equação 2.15, definindo a eletrofilicidade total ω . E quando o sistema chega no equilíbrio, com $\Delta E \rightarrow 0$, a variação máxima no número de elétrons é dada por Equação 2.16.

$$-\Delta E = \omega = -\frac{\mu^2}{\eta} \quad (2.15)$$

$$\Delta N_{max} = -\frac{\mu}{\eta} \quad (2.16)$$

As aplicações desses descritores globais são amplas e foram impulsionadas pela a popularização dos cálculos de DFT no início dos anos 90. Citando alguns exemplos: correlação com basicidade de aminas;^[98] predição de várias reações orgânicas e ranqueamento de reatividade de espécies radicalares;^[97] para reações entre moléculas e íons.^[99] Já a partir do início do presente século, as aplicações dos descritores globais se concentraram bastante na racionalização de reações de Diels-Alder;^[100–103] força ácida e catálise ácida.^[104] Domingo e colaboradores são os principais pesquisadores na área da racionalização dessas reações, que por natureza são polares e, desse modo, são exploradas de forma muito eficiente usando os descritores de transferências de carga.^[21]

2.2.1.2 Descritores Locais

Os descritores locais são funções de posição, ou seja, tomam como argumento um conjunto de coordenadas no espaço tridimensional, e, por isso, geralmente servem para explicar regioselectividade de reações químicas. Esses descritores são baseados na distribuição da densidade eletrônica quando há variação no número de elétrons e/ou no potencial externo, causando a quebra e formação de ligações químicas. A partir da Equação 2.1 é extraída a derivada da densidade eletrônica em respeito ao número de elétrons para o potencial externo constante, que no CDFT foi batizada em homenagem a K. Fukui, representado pelo símbolo $f(r)$ (Equação 2.17).^[27]

$$f(r) = \left(\frac{\partial \rho(r)}{\partial N} \right)_v \quad (2.17)$$

Como definido para as derivadas em respeito ao número de elétrons, as funções de Fukui também são calculadas com métodos de diferenças finitas, sendo divididas em dois tipos, uma para cada lado da aproximação correspondendo a diferentes tipos de reatividade. A função de Fukui à esquerda $f^-(r)$, representa as regiões do sistema em que a densidade eletrônica se concentra quando o há perda elétrons, ou seja, de onde essa densidade eletrônica preferencialmente irá sair, dando a susceptibilidade ao ataque eletrofílico (SAE). Esse descritor é calculado no método de diferenças finitas pela diferença da densidade eletrônica

do estado de referência com N elétrons $\rho(r)_N$ pela densidade eletrônica com um elétron a menos $\rho(r)_{N-1}$ (Equação 2.18).

$$f^-(r) = \rho(r)_N - \rho(r)_{N-1} \quad (2.18)$$

Já a função de Fukui à direita, $f^+(r)$, representa as regiões para onde a densidade flui se o sistema ganhar elétrons, ou seja, onde a nova densidade eletrônica vai se concentrar, dando a susceptibilidade ao ataque nucleofílico. O cálculo no método de diferentes finitas é dado pela diferença da densidade eletrônica do estado com $N + 1$ elétrons, $\rho(r)_{N+1}$, pela densidade eletrônica do estado de referência, $\rho(r)_N$ (Equação 2.19).

$$f^+(r) = \rho(r)_{N+1} - \rho(r)_N \quad (2.19)$$

A média entre $f^+(r)$ e $f^-(r)$ é a $f^0(r)$, definida em Equação 2.20, originalmente pensada em representar a susceptibilidade ao ataque radicalar, mas que na prática apresenta a reatividade média do sistema.^[105] Para o cálculo desses descritores locais, usando o método de diferenças finitas, é necessário três cálculos de estrutura eletrônica, sendo dois desses considerando o dobro do número de orbitais moleculares, com conjunto de orbitais alfa e beta, também conhecidos como cálculos de camada aberta onde a multiplicidade do sistema geralmente é maior que um.^[106]

$$f^0(r) = \frac{\rho(r)_{N+1} - \rho(r)_{N-1}}{2} \quad (2.20)$$

Para os descritores globais, a partir do teorema de Koopmans, é possível obtê-los usando somente as energias dos orbitais moleculares do sistema de referência, precisando somente de um cálculo de estrutura eletrônica, que na maioria das vezes são de camada fechada. Para as funções de Fukui, também é possível aproximar as derivadas pelos orbitais moleculares de fronteira, como mostrado na Equação 2.21 para $f^-(r)$. Nessa equação, a derivada da densidade eletrônica de um sistema com M níveis eletrônicos, cada um representado por uma função de amplitude de probabilidade $|\phi(r)|$, conhecidos como orbitais de Kohn-Sham (KS) (equivalentes aos orbitais moleculares), para um decréscimo infinitesimal δ no número de elétrons N .^[107]

$$f^-(r) = \lim_{\delta \rightarrow 0^-} \frac{\partial \rho_{M+\delta}(r)}{\partial N} = |\phi_M(r)|^2 + \sum_{i=1}^{M-1} \frac{\partial}{\partial N} |\phi_i(r)|^2 \quad (2.21)$$

Lembrando que a densidade eletrônica é calculada como o somatório do modulo quadrado dos orbitais de KS ocupados $|\phi(r)_{occ}|$, mostrado na Equação 2.23.^[22] Por sua vez, os

orbitais KS são definidos pela combinação linear dos orbitais atômicos $\chi(r)$, onde os coeficientes c_i saem da solução do método variacional para encontrar a densidade eletrônica do estado de referência.

$$\rho(r) = \sum_i^{occ} |\phi(r)_i|^2 \quad (2.22)$$

$$\phi(r)_i = \sum_i c_i \chi(r) \quad (2.23)$$

Esse método de cálculo é conhecido como Aproximação de Orbital Congelado (AOC), que desconsidera a segunda parcela da Equação 2.21, que é a contribuição dos orbitais moleculares internos. Por isso a $f^-(r)$ nessa aproximação fica definida como a densidade do orbital ocupado de maior energia $\phi_M(r)$, mais comumente conhecido como o HOMO (Equação 2.24), ficando equivalente a aproximação para os descritores globais. A aproximação é possível porque a contribuição desses orbitais internos é muito pequena, como já mostrado por Fukui em seus índices de superdeslocalizabilidade^[108] e como já entendido em consenso na química, onde os elétrons mais importantes para reação química são os elétrons da camada de valência.^[109]

$$f^-(r) = |\phi_{HOMO}(r)|^2 \quad (2.24)$$

Para a $f^+(r)$, a derivada é dada na Equação 2.25, e, segundo a aproximação, esse descritor local é dado pela densidade do orbital ϕ_{M+1} , que é equivalente ao LUMO (Equação 2.26). A $f^0(r)$ na AOC pode ser obtido pela média das densidade do HOMO e LUMO, no entanto, a representação da susceptibilidade ao ataque radical tende a ser mais fiel quando utilizada a densidade do orbital SOMO (do inglês: *Single Occupied Molecular Orbitais*) (Equação 2.27).

$$f^+(r) = \lim_{\delta \rightarrow 0^+} \frac{\partial \rho_{M+\delta}(r)}{\partial N} = |\phi_{M+1}(r)|^2 + \sum_{i=1}^M \frac{\partial}{\partial N} |\phi_i(r)|^2 \quad (2.25)$$

$$f^+(r) = |\phi_{LUMO}(r)|^2 \quad (2.26)$$

$$f^0(r) = |\phi_{SOMO}(r)|^2 \quad (2.27)$$

A segunda derivada da densidade eletrônica em respeito ao número de elétrons $f^{(2)}$ também é uma função local comumente utilizada.^[110] Através das relações de Maxwell, se pode verificar que essa derivada é igual à derivada funcional da dureza química em

respeito ao potencial externo com o número de elétrons constante,^[111] como mostrado na Equação 2.28. Essa derivada exprime a taxa com que o sistema evolui para a sua estabilidade conforme o potencial externo muda, ou seja, a evolução da geometria molecular para o estado de referência. Igualmente, é como a função de Fukui varia durante a variação do número de elétrons do sistema.^[112]

$$f^{(2)}(r) = \left(\frac{\partial^2 \rho(r)}{\partial N^2} \right)_v = \left(\frac{\partial f(r)}{\partial N} \right)_v = \left(\frac{\delta \eta}{\delta v(r)} \right)_N \quad (2.28)$$

O método de cálculo prático dessa derivada é dado pela Equação 2.29, que acaba sendo a reatividade efetiva entre eletrofilicidade e nucleofilicidade, normalmente chamado de descritor Dual, por representar ao mesmo tempo as regiões mais propensas a participar de transferências de elétrons. Além disso, esse descritor acaba sendo citado como mais acurado que as outras funções de Fukui, devido ao cancelamento de erros das aproximações dada pela subtração.^[110]

$$\Delta f^\pm = f^+(r) - f^-(r) \quad (2.29)$$

Apesar da definição teórica da densidade eletrônica se estender por todo o espaço tridimensional, na prática essas funções locais são obtidas para uma grade de pontos em uma espaço delimitado, representada por uma caixa centrada no sistema com volume suficiente para englobar praticamente todos os valores significativos. Cada ponto no espaço r em que a propriedade vai ser avaliada é chamada de *voxel*, correspondente tridimensional do pixel, formando uma grade de valores escalares dessa propriedade nesse espaço cúbico.

Essa é uma das formas de representar os descritores locais, no qual essa informação é utilizada para gerar gráficos de iso-superfícies (superfícies formadas entre *voxels* que apresentam mesma faixa de valor), ou pela renderização de várias dessas superfícies, chamada de volume. Para fins práticos, essa representação vai ser chamada de volumétrica durante esse trabalho. Isso é importante pois há outras formas de se representar os descritores locais, como a representação condensada por átomos,^[113] onde o descritor $\Gamma(r)$ no espaço r é integrado em uma base comum Ω para determinado átomo k , como mostrado na Equação 2.30. Na Figura 2.6:A é mostrado o exemplo de representação volumétrica para o polipeptídeo TRP-cage, e na Figura 2.6:B o exemplo de representação condensada para a molécula de 4-aminopyrimidin-2(1H).^[114]

$$\Gamma_k = \int_{\Omega k} \Gamma(r)_k dr \quad (2.30)$$

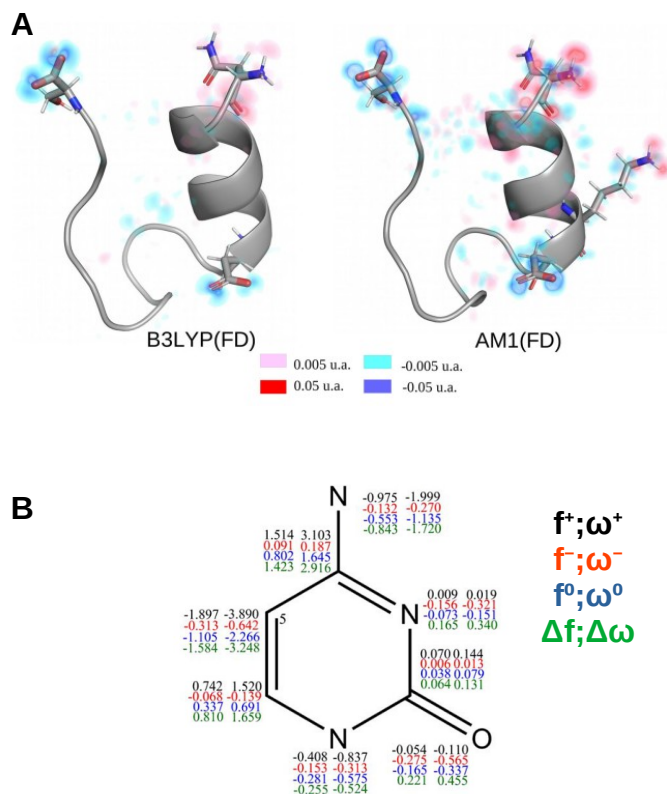


Figura 2.5: A) Imagem retirada do trabalho de Grillo e colaboradores,^[34] B) Imagem retirada do trabalho de Jerbi e colaboradores^[114]

No cálculo dessas propriedades, usando método de diferenças finitas, a densidade eletrônica de cada estado de carga é condensada para cada átomo em forma de carga parcial. Nessa abordagem, a condensação é geralmente realizada por meio de análises populacionais, sendo a mais comum a de Mulliken.^[115] Já para a AOC, os orbitais de fronteira são os que devem ser condensados para os átomos, e isso é feito de forma a particionar esses orbitais moleculares por contribuição dos orbitais atômicos AO de índice ν e sua sobreposição com os outros orbitais μ do mesmo átomo k , contabilizando como o produto cruzado ponderado pela matriz de sobreposição, $S_{\nu\mu}$. Esse processo pode ser realizado para a condensação da densidade de probabilidade para qualquer orbital molecular ψ_k^i , como mostrado na Equação 2.31.

$$|\psi_k^i|^2 = \sum_{\nu \in k}^{AO} |C_{\nu i}|^2 + \sum_{\mu \notin \nu}^{AO} |C_{\nu i} C_{\mu i}| S_{\mu\nu} \quad (2.31)$$

Especificamente para as funções de Fukui, em $f^-(r)$ a condensação é realizada para o orbital HOMO (Equação 2.32), e em $f^+(r)$ para o LUMO (Equação 2.33). Esse processo de condensação promove a possibilidade de analisar as susceptibilidades aos ataques que o

sistema pode sofrer em função dos centros atômicos, perdendo parte da informação em relação a orientação espacial, mas ganhando na interpretação de quais átomos apresentam maior probabilidade de participar do processo químico investigado.

$$f_k^- = |\psi_k^{HOMO}|^2 = \sum_{v \in k} |C_{vHOMO}|^2 + \sum_{\mu \notin v} |C_{vHOMO}C_{\mu HOMO}|S_{\mu v} \quad (2.32)$$

$$f_k^+ = |\psi_k^{LUMO}|^2 = \sum_{v \in k} |C_{vLUMO}|^2 + \sum_{\mu \notin v} |C_{vLUMO}C_{\mu LUMO}|S_{\mu v} \quad (2.33)$$

As funções de Fukui indicam os locais na molécula onde há a maior propensão de haver ataques eletrofílicos ou nucleofílicos, basicamente reações onde há uma transferência de elétrons significativa, e portanto, de caráter mole-mole, resultando em ligações de maior característica covalente do que iônica.^[116] No entanto, quando comparada a reatividade local de diferentes moléculas, a função de Fukui não está apta a identificar se a região da molécula *A* mais suscetível a ataque eletrofílico, por exemplo, seja mais reativa do que a região de mesma característica na molécula *B*. Então, é dito que as funções de Fukui são úteis para identificar a reatividade de caráter mole-mole, interações controlados por orbitais, e de tipo intramolecular, isto é, comparando diferentes regiões da mesma molécula.^[117]

Para as comparações intermoleculares existe o descritor de moleza local, definido pela troca na derivada do número de elétrons pela densidade eletrônica, que é a própria versão local de *N*. Na Equação 2.34 é mostrada a definição desse descritor local, que devido a identidades já conhecidas do CDFT, resulta no produto da função de Fukui com a moleza global.^[118]

$$s(r) = \left(\frac{\partial \rho(r)}{\partial \mu} \right)_v = \left(\frac{\partial N}{\partial \mu} \right)_v \left(\frac{\partial \rho(r)}{\partial N} \right)_v = Sf(r) \quad (2.34)$$

Portanto, a moleza local é a distribuição da moleza global ponderada pela função de Fukui, sendo máxima onde a função de Fukui também é. O descritor de moleza local então é redundante para racionalização da regioselectividade intramolecular, mas é útil para comparar regiões de diferentes moléculas. Por causa da definição matemática, a moleza local acaba tendo suas equações de cálculo baseadas na função de Fukui, apresentando a moleza local para susceptibilidade ao ataque eletrofílico $s^-(r)$ (Equação 2.35), susceptibilidade ao ataque nucleofílico $s^+(r)$ (Equação 2.36) e multifílico $s^\pm(r)$ (Equação 2.37).^[119]

$$s^-(r) = Sf^-(r) \quad (2.35)$$

$$s^+(r) = Sf^+(r) \quad (2.36)$$

$$s^\pm(r) = Sf^\pm(r) \quad (2.37)$$

Nessa mesma ideia da multiplicação de um descritor global por um local, a eletrofilicidade local é definida de forma semelhante. Como mostrado na equação Equação 2.38 representando a estabilização de energia local quando o sistema doa elétrons $\omega^-(r)$, quando recebe $\omega^+(r)$ (Equação 2.39), e a para os dois processos ao mesmo tempo $\omega^\pm(r)$ (Equação 2.40), definindo assim o que é chamado de descritor multifílico.^[120]

$$\omega^-(r) = \omega f^-(r) \quad (2.38)$$

$$\omega^+(r) = \omega^+ f(r) \quad (2.39)$$

$$\Delta\omega(r) = \omega f^\pm(r) \quad (2.40)$$

Esses descritores locais são amplamente encontrados na literatura para explicar a regioselectividade de reações que tem significativa transferência de elétrons. Dos exemplos de aplicação, é possível citar: basicidade de aminas;^[121] reações de substituição aromática;^[122–124] reações de Diels-Alder;^[45,102,125] ataques nucleofílicos em compostos carbonílicos;^[119] correlação com pKa e capacidade de inibição de corrosão de aminoácidos;^[126] racionalização de interações entre bases de DNA;^[114,127] aplicação para sólidos, superfície e catalise;^[128–131] na parte de química medicinal, dando suporte na busca de novos inibidores usando os descritores como função de escolha,^[132] capacidade antioxidante,^[133] atividade biológica^[134,135] e correlação com toxicidade.^[136,137] As aplicações específicas para sistemas biológicos e acompanhamento de reação química estão separadas em outra seção desse capítulo, e serão trazidas com mais detalhes.

Na Figura 2.6, é mostrado como o descritor dual $f^{(2)}$ é utilizado para mostrar simultaneamente as regiões susceptíveis a ataques eletrofílicos (vermelho) e a ataques nucleofílicos (verde), para reações de transferência de prótons acopladas com transferência de elétrons. Já na Figura 2.7 é exemplificado como a eletrofilicidade local ω_k , e seu inverso a nucleofilicidade local N_k , mostram quais os átomos de carbono mais reativos que controlam a formação do produto cinético, aquele que apresenta a barreira de energia livre de ativação menor.^[138]

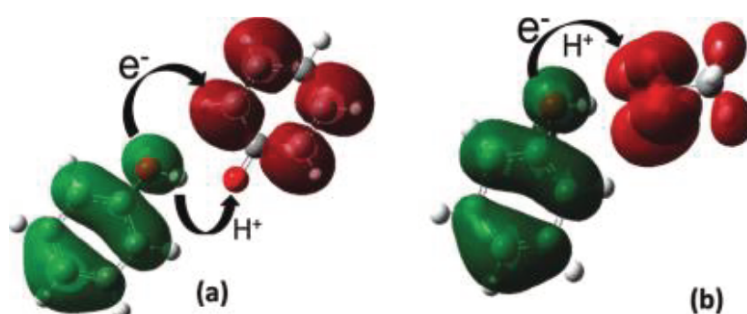


Figura 2.6: Uso de descritores Dual para estudar processos de transferência de prótons acopladas com transferência de elétrons. Regiões em verde são as nucleofílicas e vermelhas eletrofílicas A) reação: $ArOH + ArO^{\bullet}$ B) reação: $ArOH + CH_3OO^{\bullet}$. Imagem retirada do trabalho de Liu e colaboradores^[138]

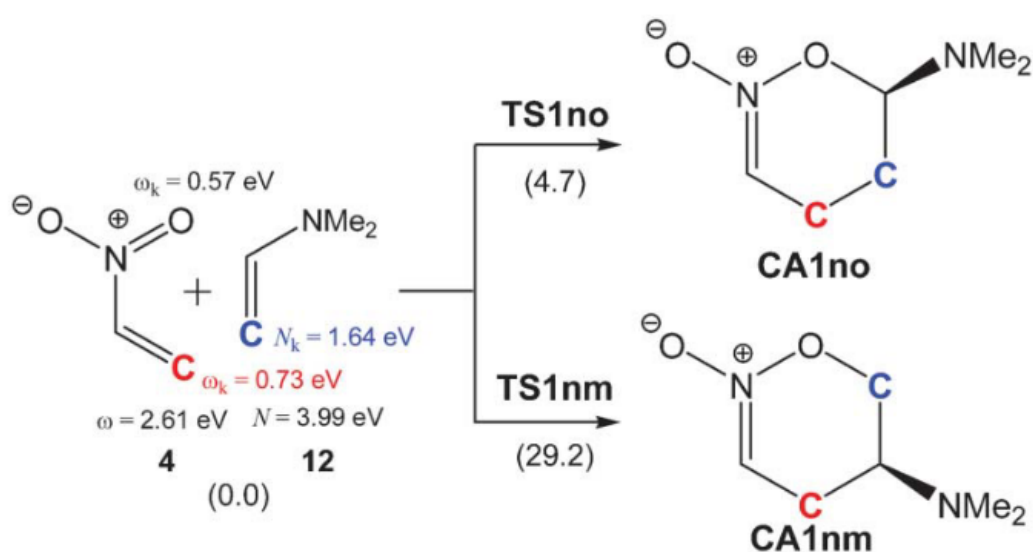


Figura 2.7: Seleção do produto cinético através da reação entre os carbonos com maior eletrofilicidade local ω_k (molécula 4 desse estudo) e com a nucleofilicidade N_k (molécula 12). Entre parênteses na imagem está a energia de ativação em kcal/mol. Fonte: Imagem retirada do trabalho de Domingo e colaboradores^[125]

2.2.1.3 O Problema da Dureza Local

Todos os descritores locais definidos até aqui estão ligados a efeitos que governam a reatividade dos sistemas pela transferência de carga e sobreposição de orbitais, resumidamente,

interações mole-mole. No entanto, há também as reações controladas por carga, ou seja, aquelas em que os efeitos eletrostáticos dão origem a força que movem os átomos em direção ao estado de transição.^[92] Há também aqueles processos onde a aproximação dos substratos é governada por essas interações duro-duro, que tem um maior alcance, e quando há a aproximação dos átomos, a sobreposição e reorganização dos orbitais moleculares domina como força principal da reação.

Logo, é importante também um descritor local que descreva a regioselectividade de processos duro-duro. O próprio Potencial Eletrostático Molecular (PEM) e as cargas parciais dos átomos são formas de representar localmente o efeito das interações eletrostáticas. No entanto, dentro do arcabouço teórico do DFT a quantidade lógica seria a dureza local $\eta(r)$, definida como o inverso da moleza local $s(r)$ na Equação 2.41.^[139]

$$\eta(r) = \left(\frac{\partial \mu}{\partial \rho(r)} \right)_v \quad (2.41)$$

O grande problema dessa definição é que ela não integra para a quantidade global, como é o caso da moleza local cuja integral em todo o espaço r resulta na moleza global. Como mostrado na Equação 2.42, a dureza local só integra para dureza global se for em conjunto com o produto da função de Fukui, por causa da identidade mostrada na Equação 2.43.

$$\eta = \int \eta(r) f(r) dr \quad (2.42)$$

$$\partial \rho(r) = f(r) dN \quad (2.43)$$

Isso faz com que um descritor, que é de imensa importância, tenha sua definição matemática ambígua, mostrando uma falha no que deveria ser o ponto mais forte do CDFT, que é seu rigoroso tratamento teórico-matemático. Apesar disso, há várias definições para dureza local, chamadas de equações de trabalho, que seriam formas de aplicar localmente essa quantidade.^[140] Dentre as soluções possíveis, há aquelas que se preocupam mais com a capacidade de explicação do caráter duro-duro dos sistemas estudados e outras com uma definição matemática que respeite os requerimentos da CDFT.^[141]

Esse é o conhecido "problema da dureza local", que, em publicação recente reunindo os principais nomes do CDFT para uma revisão e direcionamento dos próximos passos na área,^[142] foi dado de certa forma um ultimato ao problema da dureza local, onde se exige a definição de um só método de cálculo e que se respeite as identidades matemáticas estabelecidas na teoria. Por hora, nessa Tese, vão ser abordadas algumas das definições de equações

de trabalho para dureza local que foram testadas para sistemas biológicos nos estudos de caso contidos nesse trabalho e também implementado no software PRIMoRDiA.^[38]

A primeira estratégia consiste em desconsiderar a definição pura da dureza local, utilizando uma função auxiliar no qual a integral do produto entre ela e a definição de dureza local resulte no valor global. Como já dito anteriormente, isso pode ser realizado com o produto da função de Fukui, mas outras funções locais que sejam normalizadas podem ser utilizadas também.^[143]

Uma definição dentro das equações do CDFT para a dureza local foi desenvolvida por Berkowitz e colaboradores,^[139] considerando o tratamento de fluido clássico para um sistema multi-eletrônico das equações de densidade eletrônica, mostrado na

$$\eta(r) = \left(\frac{\partial \mu}{\partial \rho(r)} \right)_v = \frac{1}{2N} \int \frac{\delta^2 F[\rho]}{\delta \rho(r) \delta \rho(r')} \rho(r') dr' \quad (2.44)$$

A partir dessa definição, podemos retirar uma função não local da densidade eletrônica, que é um função da posição r que depende do valor em outro ponto do sistema r' , que nesse caso é chamado de dureza de núcleo, definida na Equação 2.45. Descritores não locais são derivadas de maior ordem da funcional de energia em respeito à posição, e são propriedades que representam efeitos perturbativos, onde o valor em r é acoplado com o valor em r' , representando o efeito de outros pontos do sistema na posição que se deseja avaliar.

$$\eta(r, r') = - \frac{\delta^2 F[\rho]}{\delta \rho(r) \delta \rho(r')} \quad (2.45)$$

Para definir, então, uma função local para a dureza, precisamos entrar na definição de $F[\rho]$, que é subdivida em três funcionais da densidade (Equação 2.2), logo podemos assumir que a dureza local seja particionada para cada dessas funcionais (Equação 2.46). No entanto, para regiões da superfície molecular, as funcionais de energia cinética e de troca-e-correlação apresentam um decaimento abrupto, pelas aproximações do tratamento estatístico de Tomas-Fermi-Dirac das funcionais de densidade.

$$\eta_F(r) = \eta_T(r) + \eta_J(r) - \eta_K(r) \quad (2.46)$$

Logo, a dureza local pode ser aproximada como na Equação 2.47, aproximando $F[\rho]$ como a funcional de interação elétron-elétron $J[\rho]$, mostrada na Equação 2.48. Essa equação de trabalho é a primeira definição que começou a ser utilizada, onde é somado todos os valores de densidade eletrônica em todos os pontos do sistema que não sejam o r , e

ponderado pela distância entre r e r' . Essa definição é de fato bastante similar ao potencial eletrostático molecular Φ (Equação 2.49) quando ignorado as cargas Z dos núcleos dos átomos A .

$$\eta(r) = \frac{1}{2N} \int \frac{\delta^2 J[\rho]}{\delta \rho(r) \delta \rho(r')} \rho(r') dr' = \frac{1}{2N} \int \frac{1}{|r - r'|} \rho(r') dr' \quad (2.47)$$

$$J[\rho] = \frac{1}{2} \int \int \frac{\rho(r) \rho(r')}{|r - r'|} dr dr' \quad (2.48)$$

$$\Phi = \sum_A \frac{Z_A}{|r - R_A|} - \int \frac{\rho(r')}{|r - r'|} dr' \quad (2.49)$$

Por esse fato, muitos estudos começaram a utilizar o próprio potencial eletrostático molecular, ou as cargas em excesso nos átomos, como forma de dureza local. De fato, essas formas de estimar o potencial de diferentes regiões do espaço de uma molécula de participar em interações duro-duro tem relação com os efeitos físicos esperados. No entanto, nem na definição proposta na Equação 2.49 nem no uso de outras propriedades ligadas a carga, não apresentam nenhuma relação com a quantidade global da dureza química.

Outra equação de trabalho surge da derivação da definição anterior (Equação 2.48), substituindo a densidade eletrônica pela função de Fukui, e eliminando o fator de normalização em relação ao número de elétrons ($1/N$), isso por que a função de Fukui já é normalizada. Essa aproximação dá origem ao potencial de Fukui e é a própria dureza no r correspondente ao núcleo dos átomos.^[144] A equação do potencial de Fukui fica como mostrado na Equação 2.50, sendo que, nessa definição, a função de Fukui é encarada como a distribuição da densidade eletrônica de fronteira, ou seja a mais reativa.

$$v(r)^{f^-} = \int \frac{f^-(r')}{|r - r'|} dr' \quad (2.50)$$

A definição mais usual do potencial de Fukui se dá com a função de Fukui esquerda f^- , podendo ser tanto a densidade de probabilidade do HOMO, quanto a definição de diferenças finitas. A princípio, essa função de Fukui é a mais apropriada para aproximar a densidade eletrônica, já que é relacionada a um orbital ocupado. No entanto, é possível utilizar as outras funções de Fukui também, resultando em equações de trabalho que podem ser mais apropriadas para avaliar a reatividade de um sistema específico, como mostrado na Equação 2.51 e na Equação 2.52.

$$v(r)^{f^+} = \int \frac{f^+(r')}{|r - r'|} dr' \quad (2.51)$$

$$v(r)f^0 = \int \frac{f^0(r')}{|r-r'|} dr' \quad (2.52)$$

Um terceiro tipo de equação de trabalho foi definida por Gal e colaboradores,^[145] com a intenção tanto de representar os efeitos de natureza duro-duro locais como também de possibilitar a integração da quantidade, resultando na dureza química global. Para isso, foi necessário desenvolver uma versão local do PQE, que de fato é contraintuitivo com a própria natureza da analogia com a quantidade termodinâmica clássica. No entanto, essa parte foi desconsiderada, de forma que, usando as propriedades matemáticas bem específicas das funcionais, mostrada na Equação 2.53, onde o potencial químico eletrônico acaba sendo escrito como uma integral sobre o espaço de uma outra quantidade local, nesse caso a densidade eletrônica, como mostrado na Equação 2.54. Logo, retirando a integral na Equação 2.54, o PQE se torna uma função de r , isto é, uma função local, que pode ser escrita como na Equação 2.55.

$$\frac{\delta E[\rho]}{\delta \rho(r)} = \frac{1}{N} \int \rho(r') \frac{\delta E[\rho]}{\delta \rho(r')} dr' \quad (2.53)$$

$$\mu = \frac{1}{N} \int \rho(r) \frac{\delta E[\rho]}{\delta \rho(r)} dr \quad (2.54)$$

$$\mu(r) = \frac{\rho(r)}{N} \frac{\delta E_v[\rho]}{\delta \rho(r)} = \frac{\rho(r)}{N} \mu \quad (2.55)$$

Basicamente, o $\mu(r)$ é uma distribuição da densidade eletrônica normalizada, se tornando uma função que serve para determinar a abundância local do PQE. Com essa definição, a derivada da Equação 2.54 pode ser encontrada em relação ao número de elétrons, resultando na Equação 2.56.

$$\eta(r) = \left(\rho(HOMO) - \frac{\rho(r)}{N} \right) \frac{\mu}{2N} + \frac{\rho(r)}{N} \eta \quad (2.56)$$

O último tipo de equação de trabalho que será abordado aqui, e talvez o mais simples, sendo basicamente a distribuição da quantidade global em função da função de Fukui, como mostrado na Equação 2.57.^[146] Essa definição é de fácil cálculo, integra para dureza global e ainda corrobora com a intuição química. Isso porque é esperado de um nucleófilo duro que sua energia de ionização seja alta, se mostrando mais resistente a transferência de elétrons mas que sua densidade eletrônica mais reativa seja o ponto de interação com um eletrófilo duro. No entanto, essa definição acaba sendo redundante em apresentar as mesmas regiões de um sistema molecular tanto para interações duro-duro quanto mole-mole.^[116]

$$\eta(r) = PI f^-(r) - AE f^+(r) \quad (2.57)$$

Grande parte do que pode ser encontrado na literatura sobre a dureza local é o desenvolvimento matemático de equações de trabalho e da busca de uma definição não ambígua em relação ao CDFT. As aplicações são mais encontradas como testes para átomos individuais e/ou íons. Ozen e colaboradores utilizaram a carga parcial,^[147] PEM e a definição de dureza local dada na Equação 2.46 como equações de trabalho para correlacionar com a energia de ligação de hidrogênio em moléculas orgânicas. Nesse estudo foi mostrado que as ligações de hidrogênio correspondendo em até 15 kcal/mol são consideradas dominadas praticamente por interações eletrostáticas e correlacionam bem com as aproximações de dureza local citadas.

2.2.2 Descritores Moleculares para Sistemas Biológicos

Os métodos híbridos de mecânica quântica/mecânica molecular tornaram disponíveis a obtenção de propriedades cinéticas e termodinâmicas de forma acurada para sistemas enzimáticos, caminhos de reação e determinação de estruturas importantes, como já discorrido anteriormente nesse capítulo. Entretanto, informações valiosas que podem ser extraídas da estrutura eletrônica resolvida são negligenciadas pela grande maioria desses estudos.^[34] Depois de custosos procedimentos computacionais para resolver a estrutura eletrônica de sistemas enzimáticos, as trajetórias geradas são racionalizadas somente com análises de distância e ligações de hidrogênio, informações que já são exploradas quando usados potenciais clássicos em simulações moleculares.

Os raros estudos, mas ainda relevantes, que exploraram a estrutura eletrônica de sistemas biológicos serão apresentados a seguir, assim como os desenvolvimentos mais recentes protagonizados pelo nosso grupo de pesquisa. Essa seção é dividida da seguinte forma: i) aplicação para fragmentos de sistemas biológicos, revisando os primeiros estudos que consideraram a estrutura eletrônica nas suas análises; ii) tentativa de um CDFT biológico, no qual é revisado trabalhos do grupo do Geerlings, um dos principais pesquisadores do CDFT, onde os descritores são aplicados para reações enzimáticas; iii) Estratégias de cálculo dos descritores para macromoléculas; iv) descritores de banda, desenvolvidos pelo nosso grupo.

2.2.2.1 Aplicação para Fragmentos

O primeiro trabalho que podemos citar sobre aplicação dos descritores do CDFT para sistemas de interesse biológico foi o de Li e colaboradores.^[148] Eles utilizaram a função de Fukui para identificar a conformação mais reativa do substrato para a enzima fosfato 5-enolpiruvilxiquimato-3 sintase (EPSP). Nesse estudo, os átomos dos resíduos da enzima não foram considerados nos cálculos quânticos, somente os do substrato fosfato 5-enolpiruvato. Esse substrato participa em mais dois outros ciclos catalíticos onde o fosfato também não é o grupo mais reativo. Isso é um exemplo de como o ambiente enzimático pode controlar a seletividade e somente a intuição química não é o suficiente para explicar a reatividade desses sistemas complexos. Nesse estudo, os autores concluíram que os ângulos de rotação dos grupos em relação a ligação dupla carbono-carbono modificavam a reatividade do substrato, e que, portanto, esse poderia ser um dos efeitos que o ambiente enzimático teria como influenciar na reação.

Em 1998, Huang e colaboradores incluíram os átomos de quatro aminoácidos nos cálculos quânticos em conjunto com moléculas potenciais fármacos anti-HIV.^[149] O estudo foi conduzido de maneira a investigar a reatividade dos ligantes nas regiões chamadas de "Zn-fingers", onde alguns resíduos da proteína interagem com íons de zinco mantendo a estabilidade estrutural da enzima nucleocapsídio HIV-1 (NCp7). Os descritores mostraram que uma cisteína em específico apresentava a maior nucleofilicidade, podendo formar ligações covalentes nesse sítio e, por conseguinte, destabilizando a estrutura terciária da enzima e a inibindo.

Alber e colaboradores^[150] usaram a diferença de densidade eletrônica entre os complexos, enzima e substrato, para a enzima timidina quinase, muito importante na recuperação de timidina no vírus da herpes. Os cálculos de DFT foram conduzidos somente com os átomos dos aminoácidos do sítio ativo, tirosina-127, metionina-128 e o substrato timidina, variando estados de carga e protonação. A partir dos resultados foi descartado interações de sobreposição de orbitais, mostrando que a interação é puramente hidrofóbica e eletrostática. No entanto, a diferença de densidade eletrônica introduzida pelo substrato, ou seja, a do complexo menos a densidade eletrônica obtida somente para os resíduos, não é uma definição de descritor de reatividade do CDFT. A análise das funções de Fukui, e principalmente moleza local, já que havia dois sistemas comparados com átomos diferentes, poderia ser empregada para realmente descartar a hipótese de interações mole-mole. As interações apontadas foram de natureza eletrostática somente, mas nenhuma análise com

cargas parciais, potencial eletrostático ou algum método de dureza local foi utilizado.

Já no trabalho de Mignon e colaboradores,^[151] publicado em 2002, as cargas parciais de Mulliken foram utilizadas para estimar a nucleofilicidade na catálise realizada pela ribonuclease, onde o oxigênio do grupo hidroxila da ribose é ativado através da transferência de próton de um aminoácido básico para a quebra da guanosina. De fato, a transferência de prótons é um processo duro-duro e nesse estudo foi aplicado os princípios de Pearson ABDM para racionalizar o processo. No entanto, as cargas parciais geralmente são escritas nos arquivos de saída de cálculos de química quântica e não necessitam de nenhum pós-processamento matemático para serem utilizados nas análises teóricas. Com quase duas décadas depois das primeiras formulações de dureza local, nesse estudo foi utilizado a quantidade local mais simples para racionalizar esse tipo de interação.

Rivas e colaboradores,^[152] em 2004, usaram a moleza de grupo para descrever a regioespecificidade da reação de transferência de hidreto em um fragmento de isoalloxazine-nicotinamid. Lee e colaboradores^[153] usaram os descritores para investigar sítios de ligação com zinco em enzimas, em 2008. Os próximos trabalhos utilizando CDFT em contexto biológico só apareceram novamente em 2017, tirando os de Roos e colaboradores que foi separada para uma análise particular na próxima subseção, onde Frau e colaboradores,^[154,155] usaram os descritores em fragmentos de peptídeos para determinar sítios de glicação preferencial. Rabi e colaboradores^[156] usaram as funções de Fukui condensadas para racionalizar as interações no sítio ativo da epóxido hidrolase.

Recentemente, Oller e colaboradores^[157] apresentaram uma aplicação do CDFT utilizando cálculos das funções de Fukui considerando temperatura finita. Essa abordagem tem como objetivo de incluir os efeitos térmicos nos níveis de energia eletrônico através da contabilização de várias configurações do sistema, representando os micro-estados.^[158] Nesse trabalho,^[157] foi estudado rotações de alguns resíduos do sítio ativo da enzima crotonil-CoA carboxilase/redutase com NADPH. As diferentes configurações estruturais foram obtidas de simulações de dinâmica molecular MQ/MM com o método semi-empírico DFTB3, para cada ponto no caminho de reação gerado. Eles conseguiram utilizar os descritores para mostrar que o ambiente enzimático aumenta a propensão dos processos de transferência de carga por meio da exclusão da interação dos substratos com o solvente. No entanto, as flutuações térmicas tiveram pouco efeito nos valores dos descritores. De fato, as contribuições térmicas para os níveis de energia eletrônica só começam a ter relevância quando consideradas temperaturas a partir de 1000K, o que para o caso de sistemas enzimáticos não é relevante na grande maioria das vezes.

2.2.2.2 Tentativa de DFT Conceitual Biológico

Foi separado aqui em uma seção específica os trabalhos de Roos e colaboradores devido ao seu pretendido impacto na área de cálculo de descritores de reatividade para enzimas. Outra informação relevante sobre esses trabalhos é que o autor correspondente é Paul Geerlings, um dos mais proeminentes pesquisadores vivos na área do DFT conceitual. Esses trabalhos se iniciaram em 2003, uma série de estudos aplicando descritores para um sistema enzimático específico, arsenato redutase. No primeiro deles, Roos e colaboradores^[159] usaram a moleza química para elencar a reatividade de uma série de complexos de arsenatos ($H_3AsO_4/H_2AsO_4^-/HASO_4^{2-}/AsO_4^{3-}$) e fosfatos ($H_3PO_4/H_2PO_4^-/HPO_4^{2-}/PO_4^{3-}$), com o tiolato CH_3S^- , que vem da da estrutura da enzima. Esses cálculos foram realizados para os sistemas isolados no vácuo com métodos de DFT.

Esse estudo caracterizou a interação do substrato com o enxofre da cisteína, aminoácido presente no sítio ativo que realiza um ataque nucleofílico no centro metálico, como interação mole-mole, isto é, interação controlada pela deslocalização dos orbitais moleculares de fronteira, com transferência de carga e formação de ligação covalente. Esse é um exemplo de que as interações entre enzima e substrato não são somente eletrostáticas e esse tipo de interação jamais seria capturado por métodos de mecânica molecular apenas. No entanto, a inclusão dos efeitos do resto da estrutura da enzima só ocorreu no segundo estudo realizado por Roos e colaboradores,^[160] através de esquema híbrido QM/MM, usando o método ONIOM para otimizar os complexos antes dos cálculos de estrutura eletrônica para a obtenção dos descritores. Além disso, nesse segundo estudo, eles analisaram as ativações da nucleofilicidade do tiolato e da eletrofilicidade do átomo de arsênio (no segundo estudo não foram analisados os fosfatos), assim como os efeitos de um íon de potássio presente no sítio ativo.

No terceiro estudo do grupo da Roos,^[161] foi estudado os mesmos complexos enzimáticos, dessa vez focando no desenvolvimento, aplicação e validação experimental de descritores de nucleofugabilidade. A reação analisada é a da formação do complexo enzima-arsenato, depois ocorre a redução por uma molécula de água, e logo em seguida o arsenato reduzido com o OH^- se dissociando. Nesse trabalho, eles usaram as propriedades eletrônicas para medir como o ambiente enzimático modula a saída desse grupo após o ataque nucleofílico. Eles concluíram que o ambiente enzimático estabiliza a saída do grupo hidroxila, onde um resíduo de serina estabiliza a formação do complexo por ligação covalente.

Nesses três trabalhos citados, nós podemos observar que o grupo aplicou os mesmos

protocolos computacionais de DFT e formas de cálculo dos descritores que já era utilizado para pequenas moléculas. Também, os sistemas e tipos de reações enzimáticas não foram exploradas, focando somente em um sistema específico.

Em 2009, o grupo publicou um trabalho de revisão intitulado de: "*Enzymatic Catalysis: the Emerging Role of Conceptual Density Functional Theory*",^[32] que teve como principal objetivo revisar os trabalhos que utilizaram CDFT nos anos anteriores e de alguma forma estabelecer como aplicar, criando o que seria o CDFT biológico. No entanto, esse foi o último trabalho da Roos em conjunto com o Paul Geerlings, que por sua vez também cessou suas contribuições para os sistemas biológicos.

A partir das análises desses trabalhos, é possível perceber que sempre houve o movimento em direção ao tratamento da estrutura eletrônica desses sistemas, e que os mesmos ganharam força com a popularização de métodos computacionalmente mais eficientes de DFT e com o aumento da aplicação de protocolos de cálculos QM/MM. No entanto, não houve uma preocupação desses estudos de adaptar o cálculo desses descritores para as particularidades dos sistemas biológicos, ou de incluir mais átomos para entender o efeito que o resto da estrutura tem nos níveis de energia. Outrossim, o que é possível notar a falta de uma aplicação consistente de métodos de dureza local, um descritor que em tese é o mais adequado para descrever interações não-covalentes e de caráter majoritariamente eletrostático, que tem um papel fundamental em todos os processos químicos que ocorrem com sistemas moleculares de interesse biológico.

Ademais, podemos observar que há um espaçamento temporal grande entre os artigos publicados e um baixo número de aplicações para sistemas biológicos. Enquanto isso, o CDFT continua sendo amplamente aplicado para pequenas moléculas e com vários desenvolvimentos teóricos, como funções não locais, descritores específicos para estados excitados, cálculos para temperatura finita, derivadas da energia eletrônica de maior ordem e etc. De fato, é interessante analisar todo o desenvolvimento do campo através de dois trabalhos de revisão liderados por Paul Geerlings, um de 2003 e outro de 2020. O primeiro, um trabalho publicado na revista *Chemical Reviews*,^[23] onde foi citado mais de 800 trabalhos, incluindo as contribuições realizadas para sistemas biológicos até aquele ano. Já o trabalho publicado em 2020, o objetivo foi o de discutir os desenvolvimentos e perspectivas do campo, assim como também estabelecer regras para a criação de novos descritores. Nesse último trabalho, não houve uma contextualização da aplicação do CDFT aplicado para sistemas biológicos, incluindo os trabalhos do próprio Paul Geerlings.

Recentemente, em abril de 2022, foi publicado um nova edição de um livro que propõe

uma revisão geral de toda literatura do CDFT, incluindo os novos desenvolvimentos teóricos, aplicações e implementações de cálculos em softwares.^[162] Dentre seus 32 capítulos, nenhum foi dedicado para aplicações específicas a sistemas biológicos. Entretanto, um novo artigo de revisão, publicado em agosto de 2022 na revista *Pharmaceutics*,^[163] escrito somente por Paul Geerlings, focou nos desenvolvimentismos do CDFT para sistemas biológicos com a intenção de mostrar o seu potencial para o desenvolvimento de novos agentes terapêuticos. Nessa revisão, os trabalhos de Roos que já vimos aqui são citados, assim como alguns trabalhos no nosso grupo.

O que não fica claro aqui é para que lado os principais desenvolvedores CDFT pretendem ir em relação aos sistemas biológicos e o chamado DFT conceitual biológico. De fato, é um campo incipiente e que tinha cessado suas aparições na literatura, mas que nos últimos anos parece estar tomando um novo folego. Os trabalhos de Flores e colaboradores,^[164–166] têm investido esforços em desenvolver uma versão do CDFT para o estudo de peptídeos, mais especificamente peptídeos-circulares de origem de organismos marinhos, para uso como possíveis fármacos. Eles empregaram descritores globais e locais do CDFT em conjunto com ferramentas de bioinformática e métodos estatísticos baseados em aprendizagem de máquina. Para esse conjunto de técnicas, os autores chamaram de peptídeologia computacional (CP). Nesses estudos, foram utilizados métodos de DFT para toda a estrutura das moléculas, mas que na verdade não passam em número de átomos dos sistemas que já citamos aqui em forma de fragmento.

2.2.2.3 Descritores Modificados para Macromoléculas

Em paralelo ao desenvolvimento de um "CDFT biológico", outros grupos tentaram aplicar descritores de química quântica para estruturas proteicas considerando todos os seus átomos. Esses cálculos têm como característica a combinação de métodos semiempíricos com algoritmos de escalonamento linear para que esses cálculos sejam realizados em um tempo computacional razoável. Por considerar um número muito maior de átomos nos cálculos de estrutura eletrônica, esses grupos de pesquisa focaram em meios de circundar algumas dificuldades na obtenção dos descritores de reatividade. Como já demonstrado por Khandogin e colaboradores,^[35] a perturbação padrão no número de elétrons executada no método de diferenças finitas praticamente não altera estrutura eletrônica de macromoléculas a ponto de indicar claramente os locais mais reativos pela função de Fukui.

Por sua vez, a função de Fukui é uma quantidade central na descrição da reatividade

local dentro do CDFT, sendo necessária até em algumas definições de dureza local. Portanto, Khandogin e colaboradores,^[35,167] tentaram modificar a forma com que é calculada esses descritores, aplicando diferentes perturbações de carga para o cálculo de diferenças finitas. A função de Fukui ficou como mostrada na Equação 2.58, onde o N é um inteiro representando quantos elétrons serão tirados ou inseridos no sistema.

$$f^{\pm}(r) = \pm \frac{1}{\Delta N} [\rho(r)_{N\pm 1} - \rho(r)_N] \quad (2.58)$$

A definição do valor de ΔN é dependente do tamanho do sistema e da natureza de sua reatividade. No estudo de Khandogin e colaboradores,^[35,167] foi testado três valores de ΔN : 1, 5 e 10, para a proteína de envelope do vírus do HIV complexada com um fragmento de RNA. Além do desenvolvimento e aplicação de descritores específicos para macromoléculas, esse estudo apresenta algumas particularidades na configuração dos cálculos de química computacional. Primeiramente, os descritores foram calculados para todas as 25 estruturas disponíveis para esse complexo por experimento de RMN, usando o Hamiltoniano PM3 com método de escalonamento linear *Divide-and-Conquer* (D&C). Essas estruturas foram otimizadas com algoritmo *steepest descent*, as dezenove com o menor calor de formação foram selecionadas para novos cálculos de estrutura eletrônica, dessa vez com raio de corte de interações eletrostáticos mais alargado. Todos esses cálculos foram realizados com o método de solvente implícito COSMO.^[83]

Os autores chegaram a conclusão que o melhor valor de ΔN foi o de 5, sendo que o menor testado, 1, apresentou grandes flutuações nos valores dos descritores, dada as várias estruturas. Já para o maior valor, de 10, eles verificaram que todos os átomos do sitio ativo e de seus arredores estavam mostrando valores de reatividade significativos. Para o melhor cenário de cálculo determinado pelos autores, eles encontraram uma reatividade similar a outros estudos que aplicaram análises para essa proteína de envelope do vírus do HIV, mas com novas informações sobre o papel de certos resíduos que antes não tinham como ser verificados, pois boa parte da estrutura do sistema não era considerada. Além disso, com esse protocolo computacional, eles tiveram acesso a todos os níveis de energia, podendo realizar uma análise de densidade de estados. Através dessa análise, eles verificaram que a presença do fragmento de RNA no complexo altera significativamente a energia dos orbitais moleculares. Outro mérito desse estudo está na parte da visualização dos resultados, que, com o poder gráfico da época, conseguiu exprimir visualmente a reatividade local desses sistemas, como é mostrada na Figura 2.8 a dureza local calculada para o complexo.

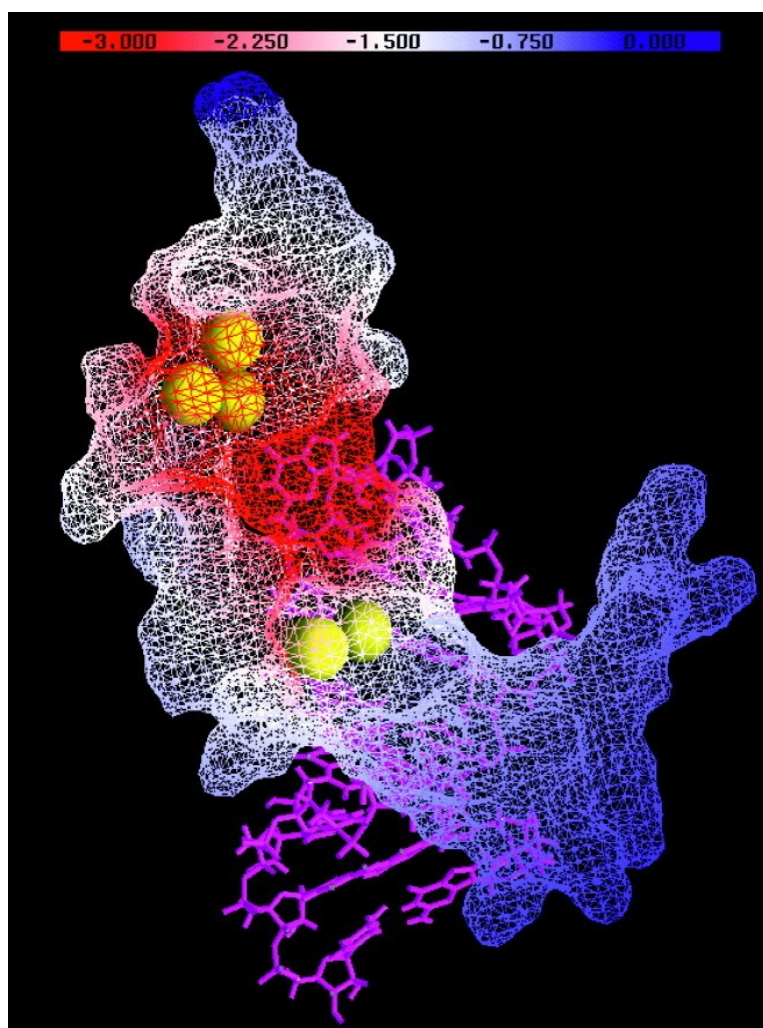


Figura 2.8: Mapa de dureza local, calculada com a aproximação do Potencial de Fukui para a proteína de envelope do vírus do HIV complexada com um fragmento de RNA. Imagem retirada do trabalho de Khandongin e colaboradores.^[167]

Os descritores modificados dessa forma só foram aparecer novamente em 2010 no trabalho de Faver e Merz,^[168] onde eles investigaram a utilidade dos princípios da teoria de ácidos e bases duros e moles para alguns tipos de simulações associadas com sistemas biológicos e processos biofísicos, como por exemplo: Ancoragem molecular, classificando quantitativamente complexos proteína-ligante; detecção de sítio ativo de enzimas; enovelamento de proteínas, usando as funções de Fukui para classificar as estruturas mais próximas da nativa. Essencialmente, nesse estudo, só foi utilizada as funções de Fukui condensadas, mostrando que essas quantidades têm potencial de explicação desses processos. Todos esses cálculos realizados por Faver and Merz utilizaram o Hamiltoniano semiempírico AM1 e o algoritmo D&C em conjunto com o modelo de solvente implícito COSMO.

Faver e Merz também investigaram sobre a escolha ótima do valor de perturbação ΔN

no cálculo dos descritores, no qual nesse último estudo foi achado que o mesmo acaba sendo muito particular de cada sistema. Isso torna essa abordagem de modificação dos descritores difícil de difundir e aplicar, pois não há uma sistematização, necessitando de cálculos de exploração, o que aumentaria bastante o custo computacional total para a avaliação da reatividade de uma macromolécula. Além disso, cada cálculo desses necessita ser feito para três estados de carga, sendo que esses cálculos com o sistema carregado positivamente e negativamente geralmente oferecem mais desafios de convergência, se somando ao problema de tempo computacional. Portanto, o método de diferenças finitas combinado com esse estilo de modificação de descritores para macromoléculas acaba por ser uma abordagem difícil de sistematizar e escalonar.

A alternativa ao método de diferenças finitas é uso dos orbitais moleculares, e como já citado nesse trabalho, Fukushima e colaboradores^[33] tiveram sucesso em demonstrar que a reatividade de enzimas estão ligados a esses níveis de energia, mas que não podem ser limitados somente análise do HOMO e LUMO, como é feito geralmente para moléculas pequenas. De fato, existe alguns trabalhos que atacam essa questão, devido a significativa degenerescência desses orbitais de fronteira para moléculas aromáticas. Uma alternativa simples é a combinação desses orbitais através de uma média aritmética dos níveis considerados importantes.^[145] Já no método FERMO (*Frontier Effective Reactivity Molecular Orbital*), do grupo de Silva e colaboradores,^[169] os orbitais são escolhidos através de uma inspeção visual da sua distribuição espacial, sendo escolhidos aqueles que têm maior localização entre os átomos nos quais se espera que participem da reação que está em análise. Também, no grupo de Pino-Rios e colaboradores,^[170] os orbitais moleculares ocupados são combinados de forma que os mais próximos do HOMO tenham maior relevância.

Todos esses estudos citados no parágrafo anterior trazem visões úteis de como os descritores poderiam ser modificados para grandes biomoléculas. Todas essas estratégias foram aplicadas para sistemas diferentes de proteínas, sem uma forma sistemática de escolha dos níveis de energia. Principalmente a questão de uma análise visual de todos os níveis de energia para um sistema biológico seria exaustivo e não produtivo.

Em 2020 foi publicado pelo nosso grupo o estudo resultante da dissertação de mestrado predecessora desse trabalho.^[34] Justamente nele foram testadas algumas hipóteses sobre o cálculo dos descritores para macromoléculas, como por exemplo: se métodos semiempíricos se comparam aos de DFT para obtenção da estrutura eletrônica de peptídeos e cálculo das funções de Fukui; como os métodos de orbitais congelados se comparam com os de diferenças finitas para um conjunto de peptídeos considerando diversos Hamiltonianos

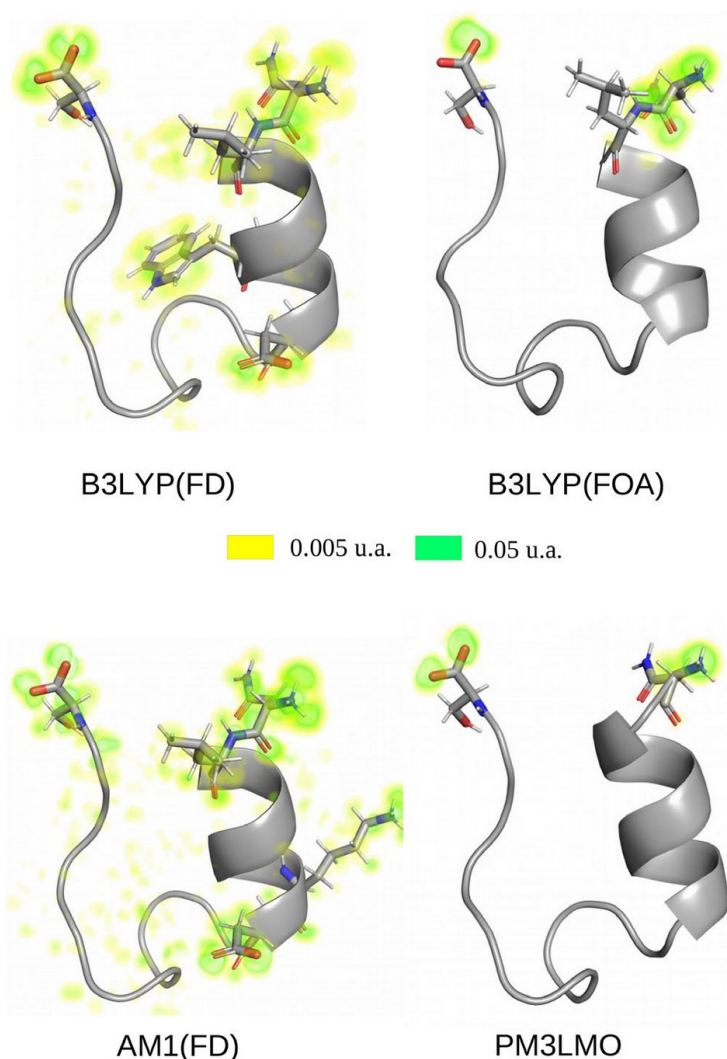


Figura 2.9: Susceptibilidade ao ataque radical para a TRP-cage usando quatro protocolos diferentes de cálculo de descritores e estrutura eletrônica. Fonte: Imagem adaptada do trabalho de Grillo e colaboradores^[34]

semiempíricos; e finalmente como combinar os orbitais moleculares para obter a melhor estimativa das funções de Fukui.

Na Figura 2.9 é mostrado alguns dos resultados para o peptídeo TRP-cage, onde é possível observar as variações nos padrões da função de Fukui média ($f(r)^0$), obtidas com os diferentes protocolos de cálculo de estrutura eletrônica avaliados. Além do mais, é possível notar que, para o método de diferenças finitas, há mais sítios reativos que para os outros métodos que utilizaram dos orbitais moleculares. Isso ocorre pelo fato de que o método de diferenças finitas considera a densidade eletrônica total dos diferentes estados de carga, o que faz com que outras regiões da molécula que não estão localizadas nos orbitais HOMO e LUMO também possam aparecer como as mais reativas.

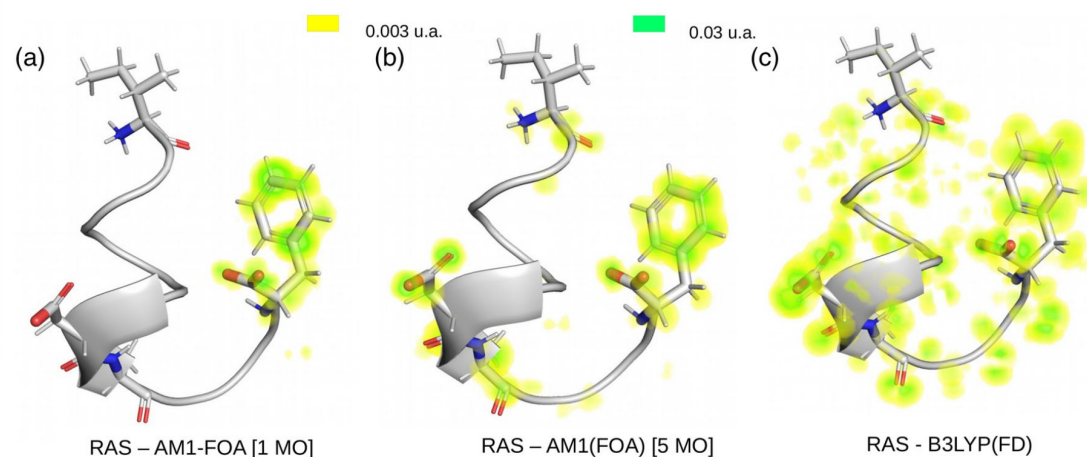


Figura 2.10: Função de Fukui Média para o peptídeo 1AQG usando diferentes protocolos de obtenção (a) orbitais congelados a partir da estrutura eletrônica do Hamiltoniano AM1; (b) combinação de 5 orbitais moleculares de fronteira ocupados e 5 virtuais a partir da estrutura eletrônica do Hamiltoniano AM1; (c) Usando diferenças finitas a partir da estrutura eletrônica calculada com B3LYP. Figura Retirada do Trabalho e Grillo e colaboradores^[34]

Apesar de o método de diferenças finitas ser mais completo na descrição da reatividade local, nesse trabalho foi concluído que é possível combinar orbitais moleculares para complementar as outras áreas e minimizar o erro em relação ao protocolo computacional de referência, que é o que emprega diferenças finitas a partir da estrutura eletrônica dos cálculos de B3LYP. O resultado da combinação de 5 orbitais moleculares ocupados e 5 virtuais, com as energias mais próximas do HOMO e LUMO respectivamente, é mostrada na Figura 2.10. Com esses resultados foi possível guiar os avanços para métodos mais robustos e sistemáticos na obtenção dos descritores para macromoléculas, sem descartar a agilidade dos métodos semiempíricos e algoritmos de escalonamento linear.

“As regras foram feitas para serem observadas”

–Ivo Vedana

Devido ao fato desse trabalho ser resultado de desenvolvimentos de diferentes publicações, diversos métodos de simulação e protocolos computacionais foram empregados para diferentes sistemas nos estudos de caso apresentados. Portanto, nessa seção encontra-se um detalhamento básico do uso de simulações, algoritmos e protocolos de cálculo comuns a todas as partes do trabalho. A metodologia específica de cada parte é dada em seção correspondente, como esquematizado na Figura 3.1.

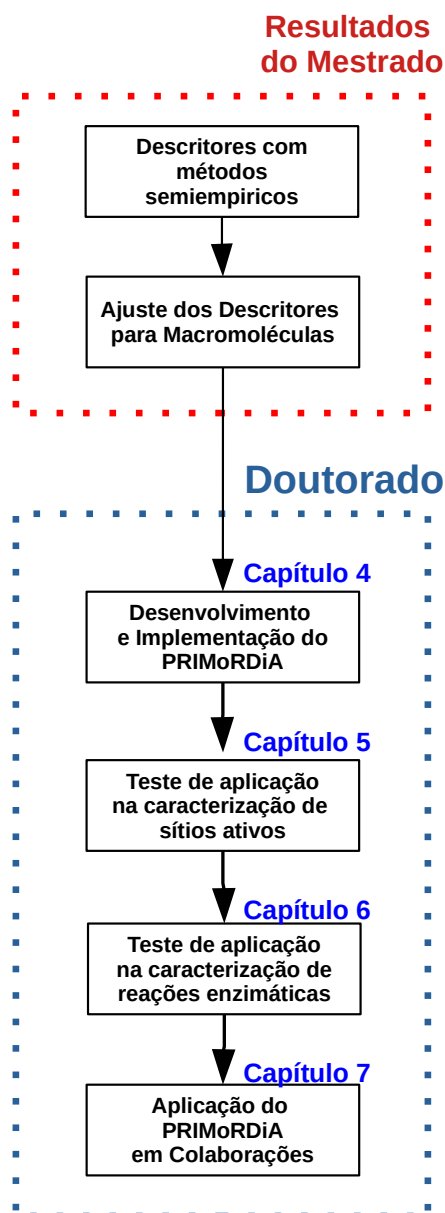


Figura 3.1: Fluxograma representando as partes em que o presente trabalho foi conduzido.

Na Tabela 3.1, são listadas todas as estruturas de proteínas utilizadas nos estudos de caso durante esse trabalho, com o código do banco de dados Protein Data Bank (PDB),^[171] a resolução experimental em Angstroms e a indicação de qual aplicação e capítulo ela é utilizada.

Tabela 3.1: Estruturas moleculares utilizadas durante esse trabalho

PDB(Å)	Sistema	Aplicação nesse estudo
1A0I(2,6)	DNA-ligase	Caracterização de sítios ativo(Cap 5.)
1DNK(2,3)	Deoxiribonuclease I	Caracterização de sítios ativo(Cap 5.)
1QFE(2,1)	3-desidroquinato desidratase	Caracterização de sítios ativo(Cap 5.)
1XC8(1,95)	DNA-formamidopyrimidine glicosilase	Caracterização de sítios ativo(Cap 5.)
1DQS(1,8)	3-desidroquinato sintase	Caracterização de sítios ativo(Cap 5.)
1BTL(1,8)	Beta-lactamase	Caracterização de sítios ativo(Cap 5.)
2CY0(1,9)	Chiquimato desidrogenase	Caracterização de sítios ativo(Cap 5.)
1TPH(1,8)	Triosefosfato Isomerase	Caracterização de sítio ativo(Cap 5.)
2CY0(2,16)	Protease Principal SARS-CoV-II	Análise de reatividade(Cap 5.)
7TIM(1,9)	Triosefosfato isomerase	Caracterização de caminho de reação(Cap 6.)
2DHC(2,3)	Haloalcano dehalogenase	Caracterização de caminho de reação(Cap 6.)
5KB5(1,8)	Adenosina quínase	Caracterização de caminho de reação(Cap 6.)
4P4G(1,70)	Chiquimato desidrogenase	Caracterização de caminho de reação(Cap 6.)

3.1 Protocolos de Dinâmica Molecular

Nesse trabalho, todas as simulações de dinâmica molecular foram realizadas com campos de força de mecânica molecular, i.e., usando modelos baseados em física Newtoniana. O objetivo da aplicação desses métodos é o de amostrar estruturas representativas, considerando

que esses polímeros biológicos, como enzimas e/ou fitas de DNA, se encontram estáveis em ambiente aquoso, em uma dada pressão e temperatura.

De forma geral, para todos os sistemas que passaram por no mínimo uma simulação de equilíbrio por dinâmica molecular, certas etapas foram respeitadas: i) obtenção do arquivo de estrutura cristalográfica de maior resolução possível no banco de dados do PDB, e que ao mesmo tempo tenha o máximo de todas as espécies moleculares de interesse, como moléculas orgânicas e íons; ii) a edição dessas estruturas se necessário, isso porque muitas vezes a enzima é complexada com um inibidor e não com o substrato natural; iii) adição dos átomos de hidrogênio e configuração do estado de protonação dos resíduos, feitos todos no Tleap do AmberTools;^[172] iv) parametrização das moléculas orgânicas com o campo de força GAFF^[173] usando o Antechamber do Ambertools;^[172] v) parametrização das estruturas proteicas e/ou compostas de bases nitrogenadas comuns a DNA/RNA, usando o campo de força amberff99sb;^[174] vi) solvatação por moléculas de água do modelo TIP3P^[175] e adição de íons de cloreto e/ou sódio até a neutralização da carga formal do sistema; vii) arquivos de coordenadas e topologia são escritos com o Tleap.^[172]

Após essa etapa de preparação das estruturas, elas já estão aptas a serem utilizadas para cálculos de estrutura eletrônica, considerando a geometria molecular assim como foi determinada pelos experimentos de cristalografia. Para os protocolos de simulação utilizados no capítulo 5, as estruturas preparadas com os protocolos descritos até aqui serão rotuladas com "L1". A nomenclatura escolhida foi a mesma utilizada na publicação no periódico,^[38] e, portanto, usamos o "L" de *level* para sinalizar o nível de tratamento das estruturas.

As próximas etapas de simulação, de minimização, termalização, pressurização e equilíbrio, foram rodadas no software GROMACS.^[176] Para isso, a biblioteca PARMED^[177] foi utilizada para transcrever os arquivos de topologia do formato do Amber para o GROMACS. A minimização de energia é feita com o algoritmo *Steepest Descent*, e foram usadas condições periódicas de contorno em todas as dimensões, sendo 5000 o número máximo de etapas de otimização, com parâmetro de tolerância de força máxima de 1000 kJ/mol/nm e 0.01kJ de etapa de variação de energia. O resultado dessa minimização de energia produziu as geometrias correspondentes ao nível L2.

A partir das estruturas obtidas para L2, um protocolo de equilíbrio por simulação de dinâmica molecular foi aplicado para gerar as estruturas correspondentes ao nível L3: simulação de 100ps no ensemble NVT para a termalização do sistema; simulação de 100ps NPT para a pressurização e ajuste da densidade do solvente; 10ns de produção. Para essas

simulações foi utilizado a temperatura de 310,15K, algoritmo LINCS^[178] para a restrição das vibrações das ligações com átomos de hidrogênio e 2fs para etapa de integração.

Nas simulações de termalização com NVT, os átomos do complexo ficaram com os movimentos restringidos, permitindo somente o movimento dos átomos do solvente para a sua acomodação ao redor da proteína e determinação das velocidades iniciais dos átomos. Com a temperatura, pressão e densidade ajustadas, as simulações sem restrição nenhuma foram realizadas com tempo de simulação de 10ns. Da etapa de produção, as coordenadas foram salvas a cada 10ps, resultando em 1000 *frames* que foram utilizados para análise de clusterização em função do raio de giro e RMSD da estrutura, obtendo assim para cada sistema a estrutura mais provável durante a simulação de produção.

3.2 Métodos Semiempíricos

Em várias partes desse texto é mencionado o uso de métodos semiempíricos para a obtenção da estrutura eletrônica de sistemas moleculares. Esses métodos obedecem o mesmo formalismo do processo iterativo desenvolvido nas equações de Hartree-Fock-Roothaan.^[179] A diferença se dá no cálculo das integrais de repulsão elétron-elétron, que por sua vez são substituídos por parâmetros previamente treinados para reproduzir propriedades experimentais ou calculadas de um dado conjunto de sistemas.^[180] Essa é a aproximação mais impactante do formalismo matemático dos métodos semiempíricos, conhecida como NDDO, do inglês *Neglect of Diatomic Differential Overlap*. Essa estratégia reduz drasticamente o custo computacional do cálculo de estrutura eletrônica, mas introduz diversos erros na estimativa de propriedades importantes, além de geralmente não prover resultados acurados para sistemas significativamente diferente dos utilizados para treinar os parâmetros.

Outras aproximações realizadas em relação aos métodos de *ab initio* são: a restrição dos orbitais internos, onde os elétrons que não são os de valência não entram nos cálculos de SCF; restrição do número de funções de bases, limitando o potencial aumento do custo computacional com a consideração de várias funções para representação dos orbitais atômicos.^[180] Com o aumento do poder computacional, outros formalismos surgiram reduzindo as aproximações feitas no NDDO, como o MNDO,^[181] onde alguns outros tipos de orbitais são permitidos, inclusão de parâmetros específicos para átomos, aproximação para integrais de dois elétrons e etc.

A reparametrização utilizando conjuntos mais diversos de moléculas e a consideração de outros efeitos com funções empíricas, como o de dispersão e ligação de hidrogênio,

resultaram em uma família de métodos. Com isso, temos os seguintes métodos: o próprio MNDO;^[181] Austin Model 1,^[181] o AM1 que é baseado no MNDO e muda somente na parametrização das interações de repulsão núcleo-núcleo; AM1/d^[182] e AM1dPhot,^[183] que são derivados do AM1 parametrizadas para lidar com correções de ligações de hidrogênio e transferências de fosfatos respectivamente; Recife Model 1,^[184] o RM1 que é uma reparametrização do AM1; os métodos parametrizados, PM3,^[185] PM6^[186] e PM7,^[187] onde a maioria dos parâmetros é independente do tipo de átomo. Os métodos da família OMx são os que apresentam correção de ortogonalidade, adicionando explicitamente interações na matriz de Fock, como de penetração de repulsão de Pauli e interações entre elétrons de valência e elétrons internos.^[188] Além desses existe também o DFTB, que utiliza o formalismo do DFT,^[189] mas com substituição do cálculo das funções de densidade por funções empíricas a partir de raios de corte de interação.

Além dessa estratégia de acumulo de aproximações, para a contabilização de sistemas grandes pode ser utilizado esquemas de linearização da complexidade computacional através da fragmentação do sistema para a diminuição das matrizes. Duas técnicas mencionadas na Tese são o *Divide and Conquer* (D&C)^[190] e o MOZYME,^[191] que geralmente são as escolhidas para serem combinadas com os métodos semiempíricos. O método D&C é fundamentado na divisão do cálculo da matriz densidade do sistema completo em várias matrizes menores referente à sub-sistemas, fazendo com que se reduza o efeito do aumento exponencial do custo computacional relacionado ao tamanho das matrizes. Já no caso do MOZYME, se usa um esquema de localização de orbitais moleculares considerando a estrutura de Lewis do sistema como chute inicial, fazendo com que várias parcelas de blocos não diagonais das matrizes a serem construídas possam ser consideradas nulas e portanto sejam ignoradas no momento do cálculo.

Nas simulações computacionais desse trabalho, as estruturas eletrônicas dos sistemas foram obtidas com métodos semiempíricos usando o software MOPAC,^[192] usando algoritmo MOZYME, com o método de solvente implícito COSMO.^[83] O raio de corte empregado para a contabilização das interações eletrostáticas no MOPAC foi de 9.0 Å, o raio de van der Waals da sonda eletrostática do COSMO foi de 1,3Å e a constante dielétrica foi de 78,4 eV, correspondente a um ambiente aquoso. Os Hamiltonianos usados serão especificados na parte do texto que relatar os resultados, mas os considerados foram: AM1, RM1, PM3, PM6 e PM7.

3.3 Simulações com Métodos Híbridos de Mecânica Quântica/Mecânica Molecular

O método híbrido é como podemos chamar a forma de obtenção de energia do sistema quando calculada por uma combinação de modelos de energia, que, especificamente nas simulações realizadas nessa Tese, foram os de Mecânica Quântica e de Mecânica Molecular. O modelo teórico e os protocolos computacionais utilizados para o cálculo de energia nos dois métodos são totalmente diferentes, mas os dois estimam a soma total das energias cinética e potencial, podendo ser representada pelo operador Hamiltoniano. Portanto, para calcular a energia de um sistema com os métodos híbridos é aplicado a uma parte do sistema um Hamiltoniano clássico e para outra parte um Hamiltoniano quântico.^[16] Para fins de simplificação, é interessante denominar essas partes como regiões, e, portanto, os cálculos híbridos QM/MM devem ter pelo menos uma região de cada. Na Figura 3.2, é mostrada a representação dessa interação.

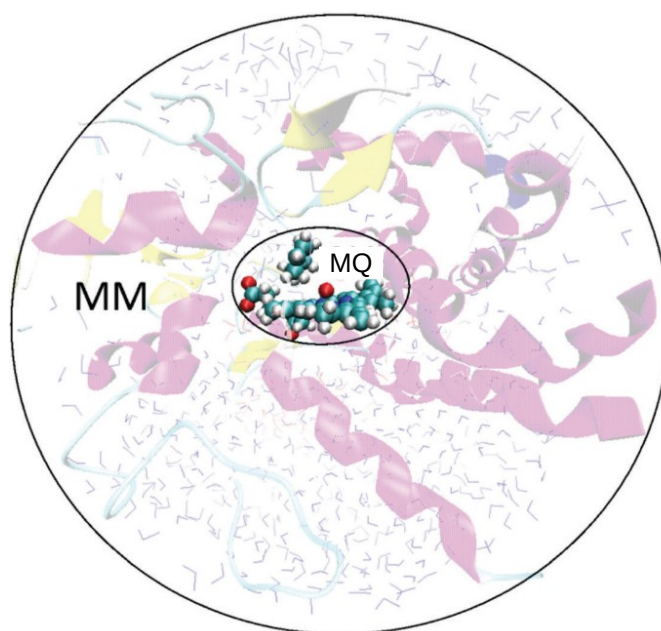


Figura 3.2: Representação gráfica da separação de regiões do sistema no que diz respeito ao modelo de energia utilizado. Fonte: Imagem retirada do trabalho de Lonsdale e colaboradores^[193]

No entanto, não faria sentido separar o sistema dessa forma se não fosse possível estimar o efeito que uma parte causa na outra, ainda mais que há a possibilidade de ligações químicas entre essas duas regiões. Portanto, existe um terceiro Hamiltoniano, que é o de Potencial Híbrido, que estima o quanto de energia do sistema advém da interação

entre as regiões. Logo, o Hamiltoniano total se torna uma soma dos Hamiltonianos das diferentes regiões com o Hamiltoniano Híbrido, como mostrado na Equação 3.1. Além disso, também deve ser decidido como as cargas atômicas da região MM irão interagir com os elétrons da região QM, podendo ser uma interação clássica, como já é considerada com os núcleos, gerando um parcela de energia potencial correspondente; ou essas cargas podem ser acopladas no Hamiltoniano quântico, o que é chamado de método de polarização, onde a região MM acaba afetando diretamente a estrutura eletrônica resolvida.

$$\hat{H} = \hat{H}_{QM} + \hat{H}_{MM} + \hat{H}_{QM/MM} \quad (3.1)$$

Somando a isso, a forma como as ligações químicas entre as regiões serão tratadas e como a energia de interação será calculada determina os principais aspectos dos cálculos utilizando métodos híbridos. A grande maioria das implementações utiliza o método aditivo, como o mostrado na Equação 3.1. Sendo a outra alternativa o método subtrativo onde os átomos da região QM são calculados em conjunto com o potencial MM para subtrair essa parcela de energia da total, como mostrado na Equação 3.2, na qual as regiões são definidas nos parênteses.

$$E = E_{(QM)}^{QM} + E_{(QM+MM)}^{MM} - E_{(QM)}^{MM} \quad (3.2)$$

Para a simulação de sistemas biológicos, inevitavelmente a fronteira QM/MM passa por ligações químicas covalentes, fazendo com que esses átomos fiquem com a valência incorreta na região QM e com a carga MM potencialmente alterada. Para evitar a introdução de erros significativos, a quebra dessa ligação deve ser realizada com um distância significativa dos átomos mais reativos do sistema, no caso da simulação de uma reação química, e essa ligação de preferência deve ser uma onde os átomos envolvidos sejam de eletronegatividade igual.^[17] No caso de biosistemas, as ligações de corte são escolhidas entre átomos de carbono, e geralmente suas valências são completadas pela adição de um átomo de hidrogênio ou pela inclusão de um orbital para representar a ligação covalente, mas que não será considerada no SCF, isto é, não será otimizado, sendo portanto um orbital congelado durante esses cálculos. Igualmente, pode-se parametrizar átomos especiais que vão emular o potencial exercido na fronteira, chamados de *linking atoms*.^[194] Na Figura 3.3 é mostrado a representação de duas dessas estratégias para o tratamento de ligações químicas na fronteira das regiões, sendo a primeira alternativa a de *linking atoms* e a segunda de orbital congelado.

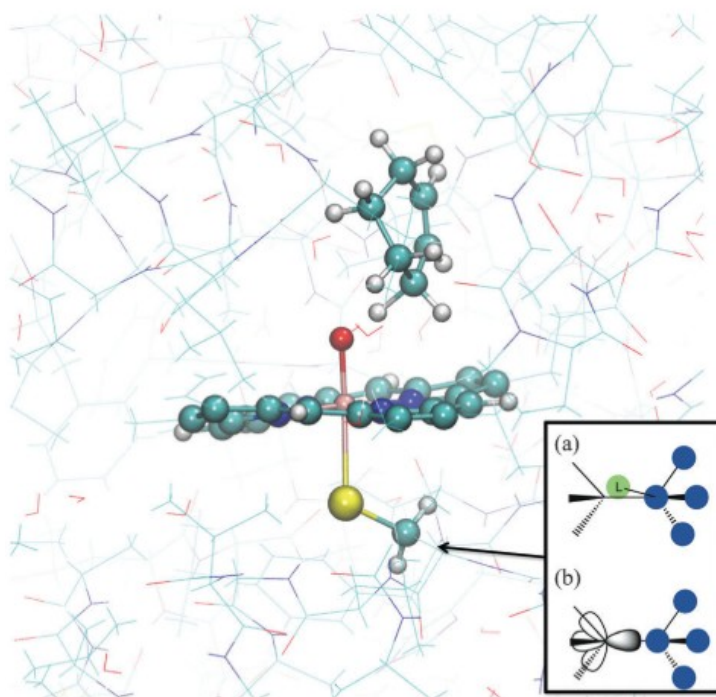


Figura 3.3: Representação gráfica dos esquemas de tratamento das ligações químicas na fronteira da região considerada com método de mecânica quântica e mecânica molecular. A) *Linking Atom*; B) Orbital congelado. Figura retirada do trabalho de Lonsdale e colaboradores^[193]

Por mais que esses métodos acelerem bastante a obtenção da energia total do sistema, sistemas biológicos acabam por conter um número de átomos significativamente grande, o que muitas vezes ainda inclui os átomos da caixa de solvatação. Para simulações de otimização de geometria ou dinâmica molecular usando potencial híbrido, geralmente é feito mais algumas aproximações, como um recorte do sistema e a imposição de restrição de movimento para grupos de átomos que não são relevantes para os processos simulados. Isso ocorreu para os sistemas enzimáticos considerados no estudo de aplicações do PRIMoRDIA^[39,40] é mostrado o exemplo de partição do sistema para o complexo enzimático da Adenosina Quínase, onde é considerado dentro da região quântica todos os átomos pertencentes a resíduos dentro de um raio de 7 Å do centro da reação, a região MM móvel até 20 Å, e uma camada de átomos fixos até 25 Å. Todos os átomos além desse limite foram excluídos, no qual a sua grande maioria eram moléculas de solvente. Esse processo foi realizado para todos os sistemas enzimáticos considerados no capítulo 6.

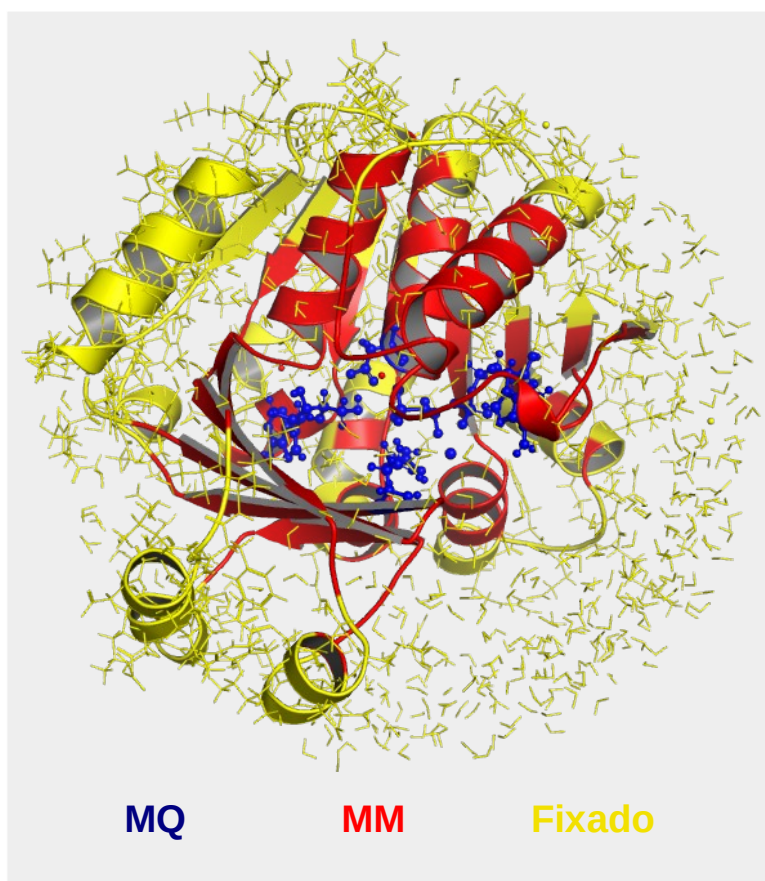


Figura 3.4: Subdivisão das regiões com diferentes tratamentos de nível teórico e de restrição de movimento para o sistema enzimático da adenosina quinase. QM = região tratada pelos Hamiltoniano quântico; MM = região tratada com campo de força clássico; Fonte: Grillo e colaboradores.^[40]

Depois de os sistemas serem preparados com os protocolos de dinâmica molecular já descritos, os arquivos de campo de força e coordenadas foram carregados no software EasyHybrid,^[195] interface gráfica para as bibliotecas de simulação molecular para Python pDynamo.^[196] Dentro deste software, o sistema tem sua geometria otimizada utilizando somente o modelo MM advindo do campo de força clássico, com o método *steepest descent*, até o sistema atingir 1,0 Å de variação média quadrática estrutural (RMS: *root mean square deviation*). Depois, o sistema passa pelo processo de corte descrito no parágrafo anterior, e tem a sua geometria otimizada mais uma vez, mas dessa vez considerando um critério mais rígido de 0,1Å de RMS. Depois, é definida a região QM, e executada mais uma otimização de geometria, só que dessa vez considerando o potencial híbrido QM/MM com o mesmo critério de parada de 0,1 Å.

Logo após a última otimização da geometria, as coordenadas estão consideradas prontas para os métodos de varredura relaxada, isto é, as coordenadas dos átomos da coordenada de

reação são atualizadas para varrer pontos ao longo da superfície de energia potencial. Para cada ponto a geometria é otimizada (relaxada) e a energia é calculada nesse ponto. Para esses cálculos, as coordenadas de reação (CR) foram definidas como a distância combinada entre os átomos que participam da reação, como mostrado genericamente na Equação 3.3. Na maioria dos casos considerados nesse estudo, o átomo B vai se dissociar do átomo A se ligando ao átomo C, portanto, nos primeiros passos da simulação a coordenada de reação possuirá valores negativos, já que a distância entre B e C é maior que a de A e B. No caso específico da Haloalcano Dehalogenase, é monitorada a associação do átomo de cloro ao carbono C2 do substrato, que está ligado ao átomo C1, logo é medido a distância entre Cl-C2 e C2-Cl.

$$CR = d(A, B) - d(B, C) \quad (3.3)$$

Essa coordenada é definida na biblioteca do pDynamo, aplicando pesos de correção para retirar o efeito causado pela diferença de massas entre os átomos, assim como um potencial de restrição harmônico, que fará com que para cada passo a distância nova imposta se mantenha. Quanto maior o valor da constante de força utilizada nesse potencial, maior será o custo energético para que os átomos da coordenada de reação se movam durante a otimização. Se a geometria em que o estado se encontrar for muito desfavorável energeticamente, isso vai se refletir no valor de energia total desse ponto da superfície de energia potencial. Logo, constantes de força muito altas podem fazer com que o sistema relaxe em menor grau, podendo resultar em valores de energia muito altos. Já constantes de força baixas podem fazer com que o sistema tenda a voltar para as coordenada originais, sendo bastante difícil de varrer valores igualmente espaçados da coordenada de reação. Para as simulações realizadas nesse estudo, foi utilizada a constante de $4000 \text{ kJmol}^{-1} \text{Å}^{-2}$.

Desenvolvimento e Implementação do Software

“Para a mente absurda, a razão é inútil, e não há nada além da razão”

–Albert Camus

Uma das questões que dificulta o desenvolvimento e aplicação de descritores para sistemas biológicos é a falta de um código que tenha implementado os métodos mais importantes, assim como as aproximações e representações já definidas na literatura. Isso fica mais claro quando comparados os métodos e quantidade aplicados para sistemas pequenos em relação aos biopolímeros. Na revisão já apresentada, vimos que há uma vasta literatura preocupada com o desenvolvimento de equações de trabalho para dureza local, as quais na maioria das vezes são aplicadas para sistemas de teste moléculas contendo não mais que 20 átomos, enquanto que para sistemas biológicos as últimas avaliações de interações duro-duro foram realizados através da simples análise de cargas parciais.

Na sua grande maioria, os descritores locais do CDFT requerem a densidade eletrônica de diferentes estados de carga ou densidade dos orbitais de fronteira. Para moléculas pequenas, essas quantidades podem ser facilmente obtidas com programas de pós-processamento para pacotes de química computacional, como o Avogadro,^[197] Gabedit,^[198] MOLDEN^[199] e vários outros.^[200–202] No entanto, esses programas não são projetados para carregar arquivos que contenham a estrutura eletrônica de macromoléculas, geralmente apresentando falha na sua execução quando carregam informações de sistemas com mais de 100 átomos.

Tanto para a área de simulação de dinâmica molecular quanto a de métodos de obtenção de estrutura eletrônica há diversos pacotes/programas disponíveis, com vários níveis de acesso, e a maioria com pelo menos licença acadêmica e possuindo seu código aberto. Dentre esses softwares populares, como por exemplo GROMACS,^[176] NAMD^[203] para simulações de dinâmica molecular, e GAMESS,^[204] ORCA,^[205] GAUSSIAN^[206] e MOPAC^[192] para

cálculos quânticos, há a implementação dos principais métodos de simulação, algoritmos e protocolos computacionais. Isso permite que diferentes grupos ao redor do mundo desenvolvam trabalhos para diferentes sistemas moleculares e que sejam testados pelos seus pares e com protocolos aprimorados, modificando parâmetros e métodos de simulação.

Todas essas características não são encontradas para a análise da função de onda de sistemas biológicos grandes, revelando uma lacuna no campo de modelagem molecular. Dentro desse contexto, foi criado o software PRIMoRDiA, que é um acrônimo para **PRIMoRDiA Macromolecular Reactivity Descriptors Access**. A origem do projeto ocorreu na dissertação de mestrado^[37] como um conjunto de códigos na linguagem Python, que é própria para *scripts* e pós-processamentos pouco complexos e de uso rápido.^[207] O código evoluiu de forma a utilizar uma linguagem compilada de alta performance, C++.^[208]

4.1 Implementação

Para essa Tese, o projeto evoluiu para o desenvolvimento de um software que seja robusto o suficiente para a distribuição à comunidade científica, com grande gama de descritores e automatização de análises estatísticas para trajetórias de caminhos de reação e de dinâmica molecular. Para cumprir isso, o código do PRIMoRDiA continuou sendo escrito em C++, usando amplamente as bibliotecas padrões e práticas de programação mais seguras e eficientes, como por exemplo a semântica de movimento e passagem de argumentos por referência ao invés do uso de ponteiros. Para o aumento de performance, as diretivas do openMP são utilizadas para tornar o programa paralelo, usando o máximo possível da capacidade dos computadores de processamento multi-core.^[208]

O código foi desenvolvido para possuir algumas funcionalidades específicas, como um leitor de arquivos de saída de cálculos de estrutura eletrônica otimizado para grandes sistemas, que geralmente apresentam tamanhos na ordem de *gigabytes*. Assim como, a implementação dos principais descritores de reatividade definidos da C-DFT, métodos de cálculo, aproximações e representações, incluindo os métodos de cálculo desenvolvidos pelo nosso grupo e ajustados para macromoléculas.

Ademais, produzir todas as informações úteis sobre as propriedades eletrônicas, com formato especial para proteínas, *scripts* de automação para análise de resultados e visualização em pacotes gráficos.

Esse fluxo de informação é mostrado na Figura 4.1, a partir do arquivo de saída de um pacote de química computacional. Além disso, o software conta com funcionalidades para

a automatização de análise de trajetórias de dinâmica molecular e de reação química. O PRIMoRDiA, atualmente na sua versão estável 1.25 (Novembro/2022),^[209] suporta arquivos de saída de quatro pacotes de química computacionais: GAUSSIAN, GAMESS, ORCA e MOPAC, podendo calcular os descritores a partir diversos tipos de métodos de estrutura eletrônica.

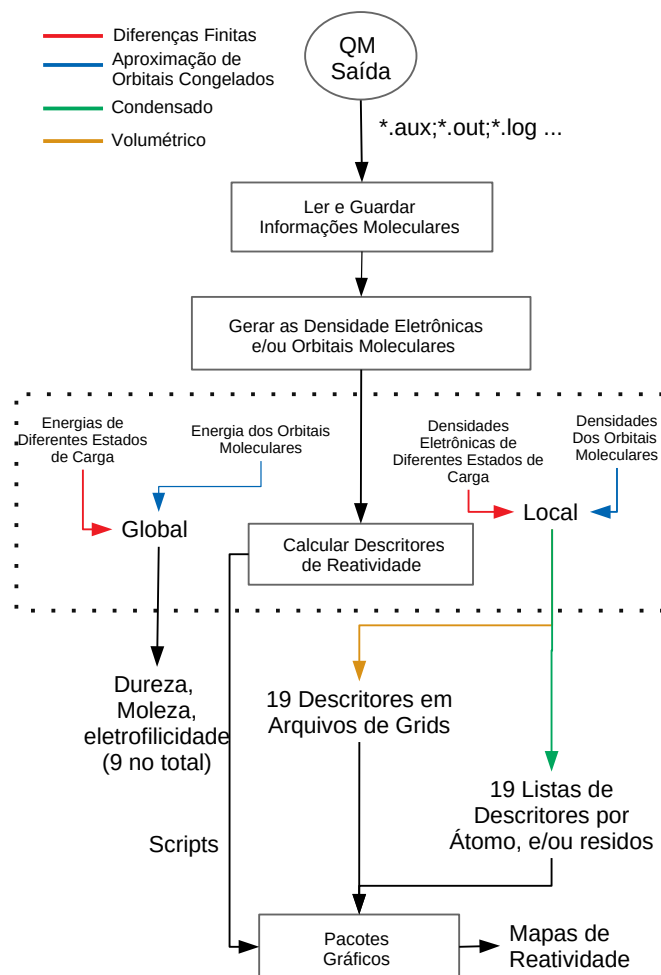


Figura 4.1: Fluxograma representando o processamento de informações realizado pelo programa PRIMoRDiA e os dados que são produzidos.

Como uma demonstração da capacidade de aceleração por uso de diretivas openMP, é mostrado na Figura 4.2 a relação tempo de computação com número de processadores para o cálculo de densidade eletrônica da molécula de fulereno. Essa funcionalidade pode ser utilizada no programa de forma individual, mas geralmente é requerida no cálculo dos descritores por meio da aproximação de diferenças finitas. É um dos procedimentos mais custosos do programa, tendo a exigência computacional dada por uma função cúbica com a resolução da grade.

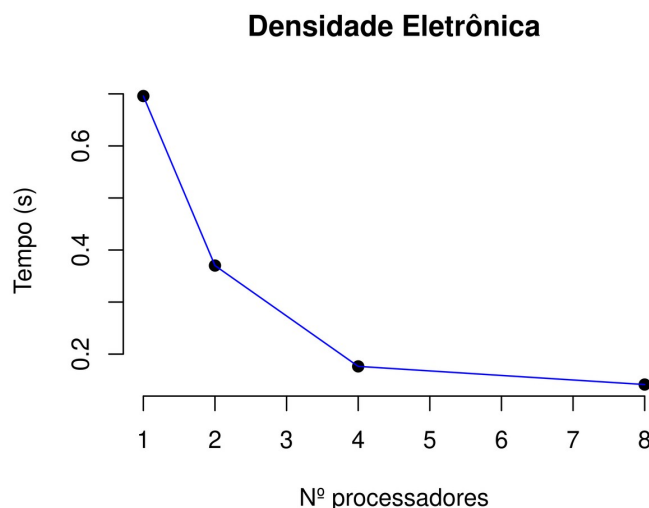


Figura 4.2: Relação do tempo de cálculo para a densidade eletrônica em função do número de processadores. Fonte: O autor.

Os cálculos de densidade eletrônica requisitados no PRIMoRDiA foram realizados a partir da estrutura eletrônica do fullereno, molécula composta de 60 átomos de carbono, calculada pelo método semiempírico PM7. Nessas primeiras avaliações, mostradas na Figura 4.2, foi requerido grade de resolução de quarenta pontos, gerando um campo escalar contendo 64000 *voxels*. Nesses resultados, mostramos que o uso de dois processadores corta em metade o tempo de cálculo em relação ao uso de uma só, e o cálculo acelera com o uso de quatro processadores, atingindo a melhor performance para o caso analisado.

Na Figura 4.3 é mostrado o mesmo tipo de cálculo de densidade eletrônica, mas dessa vez fixando o número de processadores em 8 e variando o tamanho da resolução da grade de pontos. O principal resultado de análise de performance com implementação do código para cálculos de densidade eletrônica é a conservação da dependência exponencial com o aumento da resolução, mas com uma dependência linear com o número de *voxels*, ou seja, o custo não está aumentando mais do que o ideal esperado para o algoritmo.

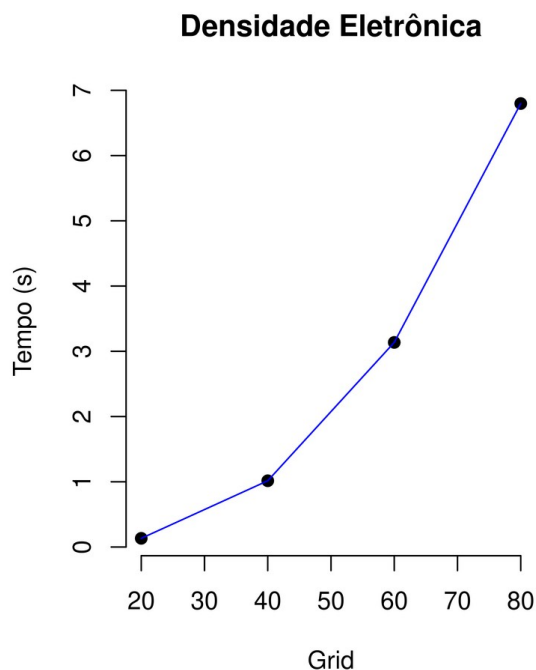


Figura 4.3: Relação do tempo de cálculo para a densidade eletrônica em função do número de pontos da grade requeridos. Fonte: O autor.

As densidades calculadas com diferentes resoluções geram arquivos do tipo de formatação padrão ".cube", que podem ser carregados e renderizados em softwares gráficos. Na Figura 4.4 é mostrado a densidade eletrônica para a molécula de fulereno renderizada no software Pymol, com dois valores de iso-superfície. De fato, volumes renderizados com resolução abaixo de 40x40x40 voxels podem resultar em imagens para esses campos escalares com acabamento mais grosseiro. Como é possível ver nessa imagem, a depender da área onde essa grade vai se espalhar, a resolução de 40 tende a ser o suficiente para produção de imagens de alta qualidade.

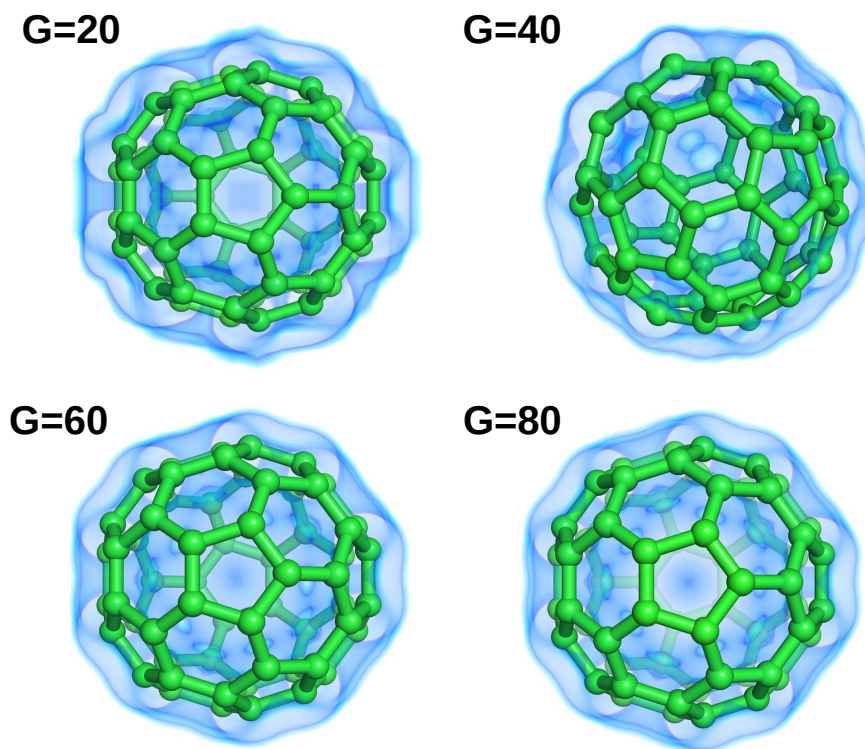


Figura 4.4: Densidades eletrônicas calculados com diferentes tamanhos de grade para a molécula de fulereno. Fonte: O autor.

4.2 Usabilidade e Distribuição

O PRIMoRDiA é um software de código livre sob a licença pública Mozilla 2.0. No repositório do Git Hub^[38] é disponibilizado versões estáveis do programa, com todos os arquivos fonte, dados para teste e binários já compilados. Nós indicamos fortemente a compilação do programa usando o *CMake*. O programa foi desenvolvido para sistemas operacionais Linux, visando usuários acostumados a trabalhar na área de modelagem molecular. No entanto, para fins didáticos e de divulgação, existe um tutorial ensinando como usar os recursos gratuitos do Google para computação em nuvem, o Google colabs. Para baixar os dados do repositório é possível usar o seguinte comando do git:

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

Ainda, 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 têm a garantia de estarem com todas as suas funcionalidades devidamente testadas. O 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>.

A compilação com o CMake é muito simples de se realizar, com poucos comandos é gerado um binário do PRIMoRDiA otimizado para a sua máquina. A desvantagem da compilação em relação a utilização dos binários prontos é a necessidade de instalação de algumas bibliotecas exigidas pelo PRIMoRDiA. Às vezes, o usuário também não tem os compiladores e o próprio CMake para o processo de compilação. Em distribuições Linux mais populares baseadas em Debian, como por exemplo Ubuntu e Mint, isso pode ser facilmente resolvido instalando as bibliotecas a partir do repositório. Por padrão, o PRIMoRDiA utiliza o g++-8 e a biblioteca Eigen 3. Para assegurar que sua máquina tenha o necessário para compilar o PRIMoRDiA, execute os comandos indicados no próximo bloco.

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

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

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

O processo de compilação deve ocorrer sem problemas e gerar um arquivo binário chamado "PRIMoRDIA_1.25v ", caso a versão 1.25 seja a escolhida (mais recomendado). No caso da impossibilidade de compilação, um arquivo binário para a última versão estável é fornecido no repositório. O mesmo processo de colocar no caminho pode ser realizado. Problemas de falta de alguma biblioteca compartilhada podem acontecer, assim como a performance máxima não ser alcançada. Entretanto, encorajamos reportar qualquer problema com a execução do binário. Depois da instalação do programa no seu sistema, é possível utilizar o programa como mostrado a seguir

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

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

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

```
Starting PRIMoRDiA software!
Calculations starting at: Fri Nov 11 16:16:28 2022

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

Total execution time of PRIMoRDiA program: 0.000414134 seconds
```

No repositório do programa, são disponibilizados arquivos com todas as informações

necessárias para utilizar e interpretar os resultados do programa, que são os Guias do Usuário, com versão em português e inglês. Igualmente, tutoriais completos estão disponíveis, cobrindo praticamente todas as funcionalidades do programa, incluindo instruções de programas auxiliares para a geração de imagens em qualidade de publicação e análises estatísticas.

4.3 Descritores Implementados

Na Tabela 4.1, estão citados todas as propriedades eletrônicas e descritores de reatividade nos arquivos de saída do PRIMoRDIA. Esses nomes já estão atualizados para a versão 1.25, visto que não tem mais sentido utilizar qualquer versão anterior a essa. Na coluna de observação são dados detalhes sobre a obtenção desses descritores e sua interpretação. Como é possível observar, alguns descritores são "crus", extraídos diretamente do processo de leitura do arquivo de saída de pacotes de química computacional, como energias eletrônicas e calor de formação, essa última sendo facilmente obtida nos cálculos com método semiempírico no pacote MOPAC.

Tabela 4.1: Descritores globais calculados/extraídos pelo PRIMoRDiA.

Descritor	Definição Matemática	Observação
Energia eletrônica	-	Extraído do <i>output</i> do cálculo de estrutura eletrônica do estado de referência
Energia eletrônica do cátion	-	Extraído do <i>output</i> do cálculo de estrutura eletrônica do estado de carregado positivamente
Energia eletrônica do anion	-	Extraído do <i>output</i> do cálculo de estrutura eletrônica do estado de carregado negativamente
Calor de Formação	-	Calor de formação calculado por método do MOPAC.
Potencial de Ionização	$E_{N-1} - E_N$	Energia necessária para extrair um elétron. Negativo do HOMO na aproximação de orbitais congelados.
Afinidade Eletrônica	$E_N - E_{N+1}$	Energia liberada na absorção de um elétron. Negativo do LUMO na aproximação de orbitais congelados.
Potencial Químico Eletrônico (μ)	$\left(\frac{\partial E}{\partial N}\right)_v$	Tendência de doar elétrons negativo da eletronegatividade de Muliken
Dureza (η)	$\left(\frac{\partial^2 E}{\partial N^2}\right)_v$	Resistência em variar o número de elétrons. Relacionado com a estabilidade química.
Moleza (S)	$1/\eta$	Facilidade de variar o número de elétrons. Recíproco da dureza.
Eletrofilicidade total(ω)	$\frac{\mu^2 S}{2}$	Energia de estabilização do sistema na absorção do máximo de elétrons.
Diferença HOMO-LUMO	$E_{LU} - E_{HO}$	Diferença de energia dos orbitais moleculares fronteira.
Número máximo de elétrons recebíveis	$-\frac{\omega}{\eta}$	Máximo de elétrons que o sistema pode receber de um doador ideal.

Na Tabela 4.2 estão listados todos os descritores locais obtidos pelo programa desde a sua primeira versão estável. Na Tabela 4.3 estão listados os descritores locais disponíveis na versão condensada, também desde a primeira versão. Na Tabela 4.4 estão tabelados os novos descritores locais volumétricos e por fim na Tabela 4.5 os novos descritores locais em representação condensada, disponíveis a partir da versão 1.25.

Tabela 4.2: Descritores locais calculados no PRIMoRDiA desde a versão 1.0v e escritos em representação volumétrica, em função da coordenada tridimensional r .

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

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

Tabela 4.3: Descritores locais calculados pelo PRIMoRDiA 1.0v e escritos em representação condensada.

Descritor	Definição Matemática	Observação
Densidade eletrônica ($\rho(k)$)		Integrado na base do átomo
Carga parcial ($q(k)$)	-	Extraído do arquivo de saída do pacote de química computacional
Susceptibilidade a ataque eletrofílico ($f^-(k)$)	$\lim_{\partial N \rightarrow 0^-} \left(\frac{\partial \rho(k)}{\partial N} \right)_v$	-
Susceptibilidade a ataque nucleofílico ($f^+(k)$)	$\lim_{\partial N \rightarrow 0^+} \left(\frac{\partial \rho(k)}{\partial N} \right)_v$	-
Susceptibilidade a ataque radical ($f^0(k)$)	$\lim_{\partial N \rightarrow 0} \left(\frac{\partial \rho(k)}{\partial N} \right)_v$	-
Descritor dual ($f^\pm(k)$)	$\lim_{\partial N \rightarrow 0} \left(\frac{\partial^2 \rho(k)}{\partial N^2} \right)_v$	-
Dureza local (potencial elétron-elétron) ($h(k)^{vee}$)	$\frac{1}{2N} \sum_{\alpha \neq k}^{atoms} \frac{\rho(\alpha)}{ r_k - r_\alpha }$	-
Dureza local (potencial químico) ($h(k)^{LCP}$)	$\left(f^-(k) - \frac{\rho(k)}{N} \right) \frac{\mu}{2N} + \frac{\rho(k)}{N} \eta$	-
Dureza local (potencial Fukui) ($v_f^-(k)$)	$\sum_{\alpha \neq k}^{atoms} \frac{f^-(\alpha)}{ r_k - r_\alpha }$	-
Moleza local dual ($s^\pm(k)$)	$S f^\pm(k)$	-
Hiper moleza local ($s^{(2)}(k)$)	$S^{(2)} f^\pm(k)$	-
Multifilicidade local ($\omega^\pm(k)$)	$\Delta \omega f^\pm(k)$	-
Localização de banda de fronteira	-	Se os métodos de banda forem requeridos

Na versão 1.25 do PRIMoRDiA, as funções de Fukui foram renomeadas para condizer com o padrão de reatividade, independente do tipo de protocolo utilizado para o seu

cálculo. A função de Fukui esquerda $f^-(r)$ é chamada então de nucleofilicidade, e a função de Fukui direita $f^+(r)$ como eletrofilicidade. Na primeira versão do programa, assim como nos trabalhos iniciais publicados pelo nosso grupo, a $f^-(r)$ era chamada de susceptibilidade a ataque eletrofílico, a $f^+(r)$ de susceptibilidade a ataque radical, o $f^\pm(k)$ de descritor dual e a $f^0(r)$ como susceptibilidade a ataque radical. De certa forma, essas nomenclaturas são mais apropriadas, pois servem para indicar as regiões onde a densidade eletrônica vai se localizar quando os processos indicados ocorrerem.

Tabela 4.4: Descritores locais adicionados no PRIMoRDiA a partir da versão 1.25 escritos em representação volumétrica.

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

No entanto, esses nomes acabam causando confusão, pois eles indicam uma propriedade em um ponto no espaço molecular que ocorre em referência a um processo desencadeado por um sistema hipotético. Para o descritor dual, foi escolhida o nome de Netfilicidade,

vindo do "*net*" do inglês e se referindo a reatividade líquida, efetiva, entre eletrofilicidade e nucleofilicidade.

Tabela 4.5: Descritores locais escritos em representação condensada e que foram adicionados no PRIMoRDiA a partir da versão 1.25.

Descritor	Definição matemática
Potencial Fukui nucleofílico ($\nu_f^+(k)$)	$\sum_{\alpha \neq k}^{atoms} \frac{f^+(\alpha)}{ r_k - r_\alpha }$
Potencial Fukui radical ($\nu_f^0(k)$)	$\sum_{\alpha \neq k}^{atoms} \frac{f^0(\alpha)}{ r_k - r_\alpha }$
Potencial molecular eletrostático ($\Phi(r)$)	$\sum_{\alpha \neq k} \frac{Z_\alpha}{ r_k - R_\alpha } - \int \frac{\rho(r)}{ r - R_\alpha } dr$
Moleza local média ($s^\pm(k)$)	$Sf^0(k)$
Dureza local (TFD) ($h(k)^{TFD}$)	$h(k)_T + h(k)_{EX} + h(k)_{EE}$

4.3.1 Descritores de Banda

Como já discutido na revisão realizada no referencial teórico desse trabalho, os polímeros biológicos possuem estruturas eletrônicas com propriedades particulares e que necessitam ser levadas em consideração no momento do cálculo dos descritores. A utilização de estratégias voltadas para o método de diferenças finitas está fora de questão devido ao custo computacional associado e possíveis problemas de convergência para grandes sistemas moleculares. Por isso, foi desenvolvido e implementado no PRIMoRDiA dois métodos de combinação de orbitais moleculares. O primeiro chamado de densidade de banda, que no software está codificado com a sigla BD, do inglês, *band density*, que estima as funções de Fukui como uma combinação linear de um dado número de orbitais moleculares a partir do HOMO e LUMO. Nessa definição, a função de Fukui esquerda $f(r)^-$ fica definida como mostrada na Equação 4.1, onde o parâmetro b é o número de orbitais a serem combinadas com energia menor que o HOMO. A mesma coisa acontece para a função de Fukui a direita $f(r)^+$, definida na Equação 4.2, considerando os orbitais virtuais de maior energia que o

LUMO.

$$f_{BD}^- = \sum_{i=b}^{HOMO} |\psi_i|^2 \quad (4.1)$$

$$f_{BD}^+ = \sum_{i=LUMO}^b |\psi_i|^2 \quad (4.2)$$

O nome desse método, densidade de banda, foi escolhido pela similaridade de como é obtida a densidade eletrônica, a qual é realizada da mesma forma, mas que, por sua vez, considera todos os orbitais moleculares ocupados. Esse método é mais dependente do conhecimento da natureza do sistema, de como as energias dos orbitais moleculares se distribuem, e por isso o PRIMoRDiA fornece de forma automatizada a análise de densidade de estados. Na prática, esse método acaba sendo mais útil quando é utilizado algoritmos de escalonamento linear baseados em estratégias de localização de orbitais, que geralmente resulta em orbitais moleculares com energias mais distantes entre si. O segundo método é baseado numa combinação linear ponderada por pesos de Boltzmann, e é codificada no software como "EW", do inglês *Energy Weighted*, significando "ponderado por energia". Nessa abordagem, os orbitais moleculares ocupados com maior energia mais próxima do HOMO contribuem mais para a função de Fukui esquerda $f(r)^-$, como mostrado na Equação 4.3. Para a função de Fukui direita ocorre da mesma forma, para os orbitais moleculares virtuais em relação ao LUMO, como mostrado na Equação 4.4.

$$f_{EW}^- = \sum_{i=b}^{HOMO} e^{-|E_i - E_{HOMO}|} |\psi_i|^2 \quad (4.3)$$

$$f_{EW}^+ = \sum_{i=LUMO}^b e^{-|E_i - E_{LUMO}|} |\psi_i|^2 \quad (4.4)$$

Controles de energia também são implementadas, permitindo somente a contabilização dos orbitais moleculares com uma diferença de energia de no máximo 1 eV para os orbitais moleculares de fronteira. Esse valor é o padrão no software e pode ser alterado pelo usuário no arquivo de *input*, sendo 1 eV indicado por Fukushima e colaboradores^[33] como valor mínimo para contabilização da banda de orbitais mais reativo, sendo o limite máximo de 5 eV. Com isso, o segundo método, EW, é apropriado para a exploração de sistemas no qual a reatividade é bem desconhecida, sendo possível incluir uma banda larga de orbitais, isso por que os orbitais com energia muito diferente das do orbital de fronteira terão uma contribuição praticamente nula para os descritores.

Essas modificações nas funções de Fukui afetam diretamente todos os outros descritores locais que usam essas funções nas suas definições, mostradas nas tabelas na seção anterior desse capítulo. Para exemplificar as modificações realizadas nas funções de Fukui, foi calculado o descritor de netfilicidade para o polipeptídeo muito comum utilizado em testes, TRP-cage, a partir do resultado de cálculo de estrutura eletrônica semiempírico PM7 em conjunto com o método de escalonamento linear MOZYME e de solvente implícito COSMO. Foram feitos vários cálculos para esse mesmo resultado variando apenas os parâmetros dos orbitais incluídos, como a quantidade para o método de *Band Density*, e a energia da largura de banda no caso do método *Energy Weighted*.

Na Figura 4.5 é mostrado a renderização desses descritores calculados com o método BD, variando o número de orbitais moleculares ocupados e virtuais com as energias logo próximas ao HOMO e LUMO, respectivamente. É possível observar o aumento do número de regiões onde o descritor é renderizado, mostrando cada vez mais regiões reativas conforme o aumento da banda de orbitais utilizada.

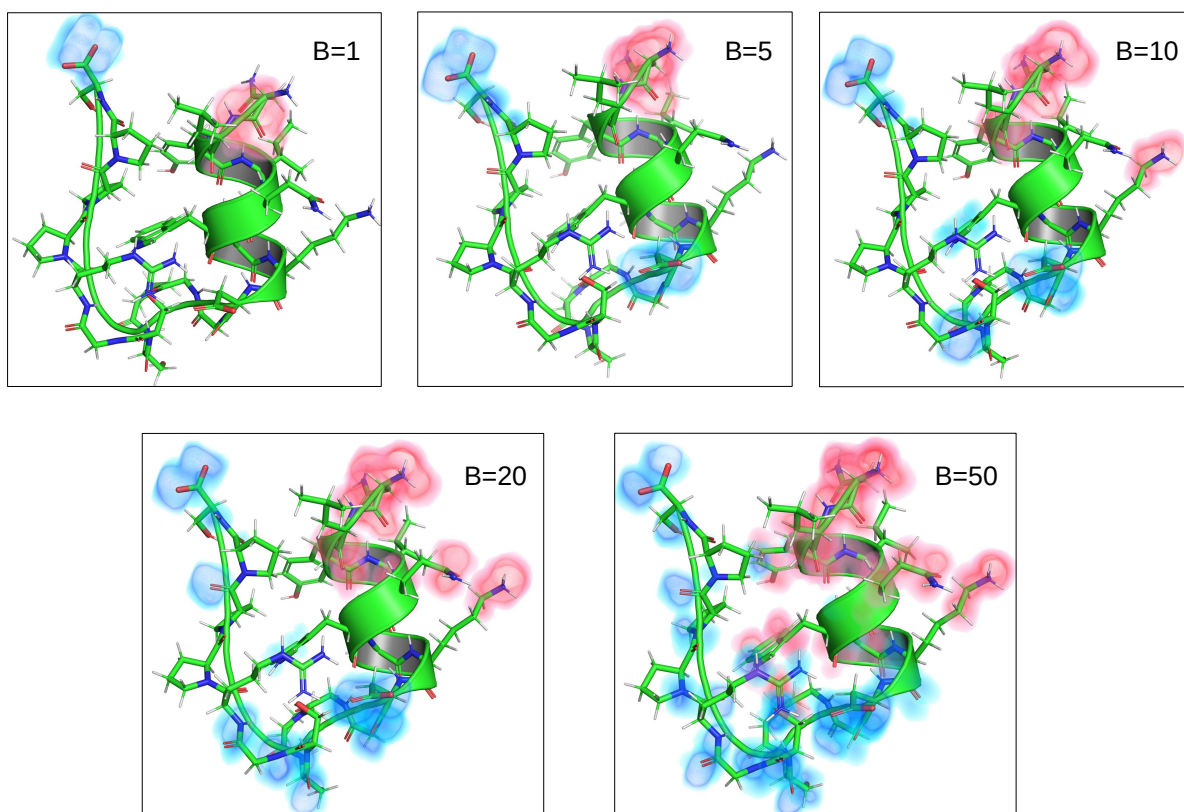


Figura 4.5: Descritores de netfilicidade calculados com diferentes tamanhos de banda para o número de orbitais moleculares. Fonte: O autor.

Para os cálculos usando o método EW é mostrado na Figura 4.6 os resultados para a

tolerância de energia para inclusão dos orbitais na banda de 1 eV e 5 eV. Utilizando esse método para essa estrutura, é possível observar que não há significativa diferença, pois esses orbitais com energia bastante distante dos orbitais de fronteira contribuíram muito pouco para a localização do descritor de nucleofilicidade. O uso do método de escalonamento linear com orbitais localizados geralmente produz orbitais moleculares com energias mais espaçadas do que quando é utilizado os orbitais canônicos, causando o comportamento observado na Figura 4.6.

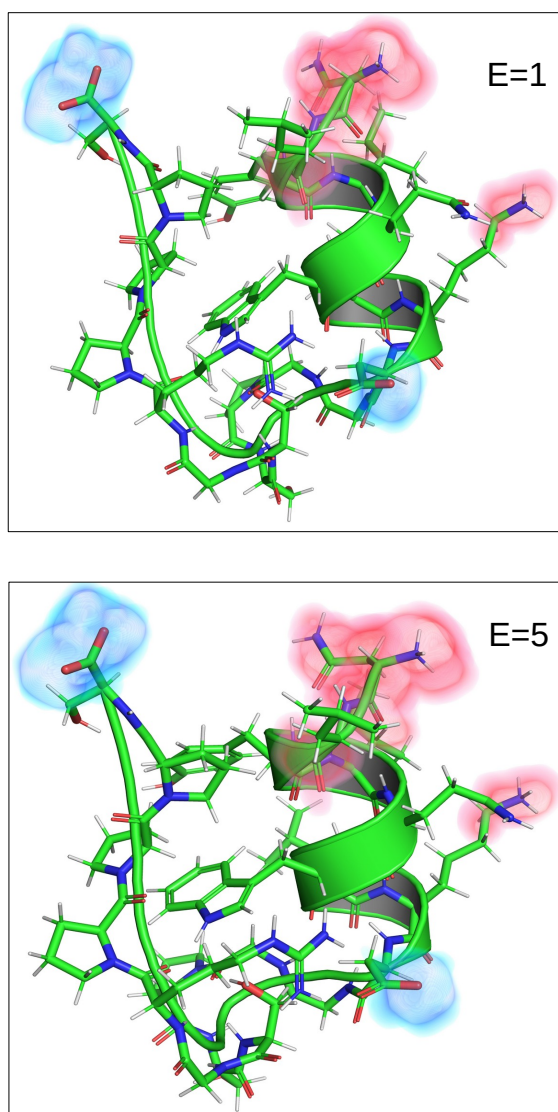


Figura 4.6: Descritores de netfilicidade calculados com diferentes faixas de energia para a escolha de orbitais moleculares. Fonte: O autor.

4.4 Comparação com Outros Softwares de Cálculos de Descritores

De fato, o PRIMoRDiA não é o primeiro e nem o único software dedicado a calcular propriedades eletrônicas para sistemas químicos, sendo sua grande diferença a especialização para lidar com grandes sistemas. Com o objetivo é disseminar o máximo de métodos e aplicações para biomoléculas, é interessante a comparação com outros códigos. Primeiramente, não serão considerados softwares que contêm funcionalidades de pós processamento, pois, mesmo podendo gerar as quantidades básicas para os cálculos dos descritores, eles ainda requerem etapas extras de processamento desses dados. Assim os programas utilizados para comparação aqui serão: UCA-Fukui,^[200] PyGlobal,^[201] MultiWFN^[210] e Chemtools.^[211]

Os dois primeiros foram desenvolvidos unicamente para calcular descritores de reatividade ligados ao CDFT, mas para um número limitado dos mesmos. O PyGlobal, como já dado a entender no seu nome, só calcula quantidades globais, sendo um software especializado para prover tabelamento dessas quantidades para um número grande de cálculos e o seu posterior tratamento estatístico. Já o UCA-Fukui, calcula as funções de Fukui com os dois métodos de aproximação e também um tipo de análise de orbitais localizados. Já o Chemtools e o MultiWFN são softwares que oferecem ampla gama de ferramentas para análise de estrutura eletrônica, com contabilização de interações não covalentes (*Non-Covalent-Interaction*) e análises de topologia da densidade eletrônica, assim como uma lista de descritores globais e locais do CDFT.

Na Tabela 4.6 é mostrado de forma geral os tipos de descritores implementados em cada um dos softwares comparados. O MultiWFN é um dos mais completos, e o que mais se assemelha com o PRIMoRDiA por ter funções de Fukui adaptadas para sistemas de alta degenerescência, mas que não se propagam para outros descritores, além de ter uma preocupação com a performance computacional para cálculos exigentes como de campos escalares de densidade eletrônica. O Chemtools, por sua vez, é uma biblioteca para Python de código livre, que tem como objetivo a produção de gráficos para análise visual de descritores relacionados com as primeiras funções de Fukui, com nenhuma demonstração de aplicação para grandes moléculas. O que chama mais atenção é a falta de implementação de descritores ligados com a definição do CDFT da dureza local, enquanto que para o PRIMoRDiA, até o momento, estão implementados sete equações de trabalho. Além disso, para o que o PRIMoRDiA se propõem, que é o tratamento específico de grandes sistemas,

não há nenhum software que apresente implementações comparáveis.

Tabela 4.6: Descritores implementados nos programas comparados.

Descritor	PRIMoRDiA	MultiWfn	UCA-Fukui	Chemtools	PyGlobal
Descritores Globais	Sim	Diferenças finitas	Sim	Sim	Sim
Funções de Fukui	Sim	Sim	Sim	Sim	Não
Funções de Fukui para Sistemas Degenerados	Sim	Sim	Não	Não	Não
Descritores Macromoleculares	Sim	Não	Não	Não	Não
Equações de Trabalho Dureza Local	Sete	Zero	Zero	Zero	Zero
Moleza Local	Sim	Sim	Não	Não	Não
Condensados para átomos	Sim	Não	Sim	Sim	Não
Condensados para resíduos	Sim	Não	Não	Não	Não

Os descritores de moleza local só se encontram no PRIMoRDiA e no MultiWFn, e, com exceção do PyGlobal, todos os pacotes que calculam descritores locais produzem esses resultados também em representação condensada para átomos. No entanto, somente o PRIMoRDiA possui uma implementação de descritor condensada para resíduos de polímeros biológicos. Na Tabela 4.7, é mostrado funcionalidades mais recorrentes desses programas que não estão relacionados com os tipos de descritores implementados.

A obtenção da densidade eletrônica total e dos orbitais moleculares pode ser feita com esses softwares, com exceção do UCA-Fukui e do PyGlobal. Além disso, o ChemTools fornece uma interface com outras bibliotecas de visualização desses resultados, assim como o MultiWFn fornece uma interface gráfica que mostra os descritores e essas outras propriedades eletrônicas. No caso do PRIMoRDiA, o programa não dispõe de ferramentas gráficas próprias, mas todos os resultados de descritores locais podem ser visualizados no software Pymol, processo facilitado pela produção de *scripts* durante os cálculos feito no PRIMoRDiA.

Tabela 4.7: Comparação das funcionalidades entre os programas de pós-processamento da densidade eletrônica baseados em CDFT.

Funcionalidade	PRIMoRDiA	MultiWfn	UCA-Fukui	Chemtools	PyGlobal
Densidade Eletrônica	Sim	Sim	Não	Sim	Não
Orbitais Moleculares	Sim	Sim	Não	Sim	Não
Interações não covalentes	Não	Sim	No	Sim	Não
Interface Gráfica	Não	Opcional	Sim	Não	Não
Cálculos em batelada	Sim	Sim	Não	Por <i>scripting</i>	Sim
Statistical Analysis Automation Scripts	Sim	Não	Não	Não	Sim
<i>Plots</i> dos Descritores locais	<i>Scripts</i> para o Pymol	Sim	Não	Sim	Não

Ademais, inspirado nas funcionalidades do PyGlobal, o PRIMoRDiA escreve *scripts* para serem executados no pacote R para a automação de análises estatísticas para trajetórias de reação e de técnicas de amostragem estrutural, como a de dinâmica molecular. Para que essa funcionalidade seja possível, é natural esperar que o PRIMoRDiA faça cálculos para vários sistemas por cada execução do programa. Isso também é possível de realizar no MultiWFn, PyGlobal e no ChemTools, mas somente com o usuário automatizando com um *script* em Python.

Finalmente, podemos citar mais dois softwares para fechar essa seção de comparação: o GPUAM^[212] e o NCIPLoT4,^[213,214] que são programas especializados para lidar com grandes sistemas de forma eficiente. No entanto, eles são focados na descrição de efeitos de interações não covalentes e/ou análises topológicas da densidade eletrônica.

4.5 Considerações Finais

Nesse capítulo, foram apresentados os principais desafios de implementação de um novo software para lidar com o grande volume de dados provenientes de cálculos de mecânica quântica para grandes sistemas moleculares, e a apresentação de todos os descritores disponíveis, assim como aqueles modificados para lidar com as principais particularidades dessas macromoléculas. Até onde foi verificado, o PRIMoRDIA é o único software a preencher o nicho de análise de estrutura eletrônica e reatividade para grandes moléculas, e ainda provê o maior conjunto de tipos de descritores, compreendendo também os métodos de dureza local que geralmente são ignorados nessas implementações.

Teste e Validação para Sistemas Biológicos

"Fate is just the weight of circumstances"

–Neil Pert

Como teste de aplicação dos descritores modificados para macromoléculas foi projetado um experimento computacional para testar a sua validade frente a dados de reatividade já bem estabelecidos. O teste consiste em usar alguns dos principais descritores que explicam interações duro-duro e mole-mole para predizer o papel de átomos dos substratos e de resíduos catalíticos.

5.1 Montagem do Experimento Computacional

Para isso, os sistemas de teste foram selecionados da base de dados MACIE,^[6] que contém a informação de quase mil reações enzimáticas, com descrição de mecanismos, papéis exercidos pelos aminoácidos do sítio ativo e informações gerais. Essas informações são compilações da literatura,^[6] considerando artigos que resolveram os mecanismos através de experimentos ou de modelagem molecular, que por sua vez também é correlacionada com dados experimentais de constantes cinéticas e métodos de determinação estrutural.

Dentre todas as 962 entradas nesse banco de dados, 686 tem o mecanismo de reação completamente descrito. A partir desses sistemas, foram selecionados 8 complexos enzimáticos considerando diversos fatores: ter o substrato no sítio ativo, ou parte dele para que possa ser editado e parametrizado; ter uma resolução estrutural do PDB de até 3.0 Å, já que longas simulações de dinâmica molecular (50-100ns) para a avaliação mais completa o possível de conformações não seria praticável para todos os sistemas; uma diversidade no

questo classe enzimática, sendo que, das 6 possíveis, foi conseguido selecionar 5 diferentes; uma diversidade nos mecanismo de catálise, contando ataques nucleofílicos, transferência de prótons e hidretos, íons metálicos no sítio ativo, água catalítica e presença de cofatores. Cumprindo esses requisitos citados acima, resultou nos sistemas com as informações descritas na tabela Tabela 5.1

Tabela 5.1: Sistemas enzimáticos escolhidos para teste de aplicação do PRIMoRDiA.

Nome	Substrato	Co-fator	Classe Enzimática	MACIE código
DNA-ligase	fDNA	ATP	Ligase	202
Deoxiribonuclease I	fDNA	-	Hidrolase	41
3-desidroquinato desidratase	3-desidroquinato	-	Liase	54
DNA-formamidopyrimidine glicosilase	7-metil-guanina in DNA	-	Hidrolase	659
3-desidroquinato sintase	3-deoxi-7-O-fosfonato-beta-D-treo-hept-6-ulopiranosonato	NAD	Liase	59
Beta-lactamase	Penicilina	-	Hidrolase	2
Chiquimato desidrogenase	chiquímato	NADP	Oxidoreductase	775
Triosefosfato Isomerase	glicerol fosfato	-	Isomerase	324

A partir desses sistemas foram realizados cálculos de energia usando o Hamiltoniano semiempírico PM7 combinado com o MOZYME para obter a estrutura eletrônica necessária ao cálculo dos descritores. As estruturas tridimensionais foram preparadas utilizando três níveis de protocolos de simulação com o nível de complexidade e tempo computacional crescente. Isso porque os métodos de preparação de estruturas na área de modelagem molecular por vezes utilizam rotinas de equilibrção que exigem tempo computacional maior que o cálculo quântico e do PRIMoRDiA, se tornando o gargalo da avaliação de

reatividade desses sistemas.

Portanto, é também do interesse do estudo testar a qualidade da obtenção dos descritores por meio de estruturas eletrônicas provenientes de geometrias, que, por sua vez, tem origem em diferentes protocolos de preparação. Na Figura 5.1 as etapas de preparação e cálculo realizadas para os sistemas de teste estão dispostas em um fluxograma. Cada protocolo está devidamente descrito na seção de métodos desse trabalho.

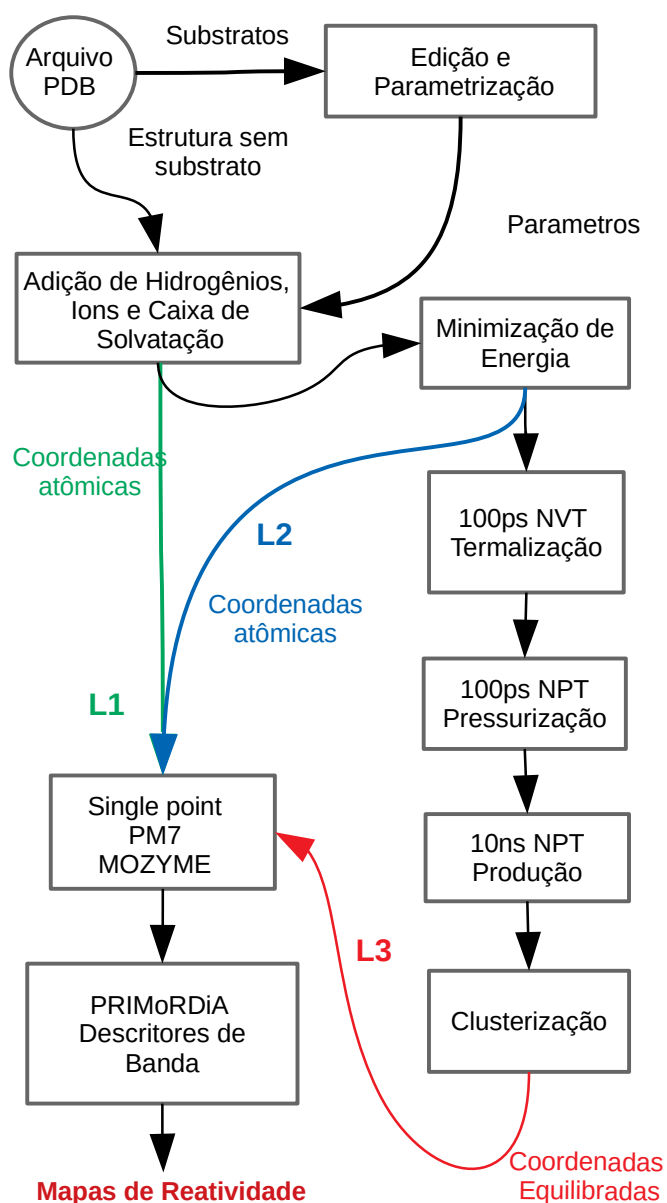


Figura 5.1: Fluxograma resumindo os métodos de modelagem empregados para cada processo de preparação das coordenadas tridimensionais para o teste de aplicação do PRIMoR-DiA.

Portanto, 24 cálculos de energia foram feitos para as coordenadas obtidas por cada métodos de tratamento, usando o método de solvente implícito COSMO. A partir dos

arquivos de saída do MOPAC, o PRIMoRDiA foi utilizado para o cálculo dos descritores usando o método de combinação de orbitais moleculares *Energy Weighted* com largura de banda de 3.0 eV. Os descritores analisados para cada sítio ativo serão o nucleofilicidade local, eletrofilicidade local e dureza local com a equação de trabalho baseada na funcional de interação elétron-elétron (Equação 2.48). A nucleofilicidade e a eletrofilicidade serão respectivamente, calculados usando as definições dadas na Equação 4.2 e Equação 2.58, com os valores condensados para os átomos, e, através da definição dos resíduos, esse valor por átomo vai ser somado e atribuído aos mesmos.

Como esses descritores calculados com esse ajuste para macromoléculas dependem das características de estrutura eletrônica de cada sistema, eles não são normalizados (na versão do programa disponível na época que esses cálculos foram realizados). Portanto, para saber se os valores são significativamente altos ou baixos, também foi avaliada a média e o máximo desses descritores para todo o sistema.

Para utilizar esses valores para prever o papel/reactividade do sítio ativo das enzimas, é necessário estabelecer critérios e entender como essas informações estão anotadas no MA-CIE. No banco de dados, os aminoácidos da enzima podem ser separados como reativos ou estabilizantes. Os reativos são categorizados nos tipos: nucleófilos, eletrófilos, doadores ou aceptores de prótons. Os estabilizantes são anotados como: interação eletrostática, doadores/aceptores de ligação de hidrogênio. Para os substratos não é atribuído um papel em específico, mas são os participantes diretos na reação e, portanto, podem se encaixar nos mesmos papéis que os resíduos reativos.

Os átomos que participam diretamente da reação são classificados na mesma lógica dos substratos, considerando se é esperado que a interação seja do tipo mole-mole e duro-duro. Para hidretos e prótons, íons pequenos e carregados, é esperado que as interações governadas por carga dominem, e, portanto para esses átomos é esperado que o seu valor de dureza local esteja bem próximo do valor máximo de dureza local para todos os átomos do sistema. No entanto, para diferenciar hidretos de prótons é necessário introduzir outro critério. Como hidretos são íons e em tese carregam um par de elétrons, é esperado que eles possuam densidade eletrônica correspondente a banda de orbitais de fronteira, ou seja, apresentar um valor não-nulo do descritor de nucleofilicidade.

Para os átomos/resíduos classificados como nucleófilos, é esperado valores significativos de nucleofilicidade, e para os classificados como eletrófilos valores significativos de eletrofilicidade. Para os resíduos considerados como estabilizantes, é esperado que tenham valores altos de dureza local, já que a maior parte da interação é de caráter duro-duro, mas

também pode ser considerada uma predição certa se apresentarem valor significativo de eletrofilicidade/nucleofilicidade dependendo do papel de acceptor/doador de ligação de hidrogênio.

5.2 Resultados

Primeiramente, serão apresentados alguns resultados relativos aos protocolos de preparação das estruturas e dos cálculos quânticos. Os resultados das simulações de dinâmica molecular, realizadas para gerar as coordenadas referentes aos nível de tratamento 3, estão resumidos na Tabela 5.2 e na Figura 5.2. A parte da produção, ou seja, da coleta de dados da simulação de dinâmica molecular, foi de 10 ns, com estruturas sendo salvas a cada 10ps e portanto gerando 1000 estruturas para as estatísticas. Esses dados foram utilizados para estabelecer com certa segurança de que, durante essas simulações, as estruturas atingissem uma estabilidade estrutural e também prover critérios de escolha das coordenadas mais representativas. Na Tabela 5.2 é mostrado o ponto da trajetória onde foi escolhido as coordenadas de cada sistema, na qual é possível notar que todas estão na parte final da simulação, com valores de RMSD e raio de giro bem próximos aos valores médios para toda a simulação.

Tabela 5.2: Resultados das simulações da análise de trajetória da dinâmica molecular e dados de clusterização.

Sistem	Média RMSD	Média Raio de Giro	Ponto escolhido	RMSD do Ponto	Raio de giro do ponto
1a0i	5,8886	4,2698	998	5,9421	4,2736
1dnk	5,2507	3,7948	995	5,3088	3,8005
1qfe	4,8794	3,5364	999	4,9238	3,5507
1xc8	5,7814	4,1833	997	5,8370	4,1797
1dqs	5,6571	4,1132	998	5,6981	4,1126
1btl	5,6935	4,1264	999	5,75554	4,1266
2cy0	5,0307	3,6473	996	5,0650	3,6534
1tph	5,1105	3,6898	1000	5,1532	3,6878

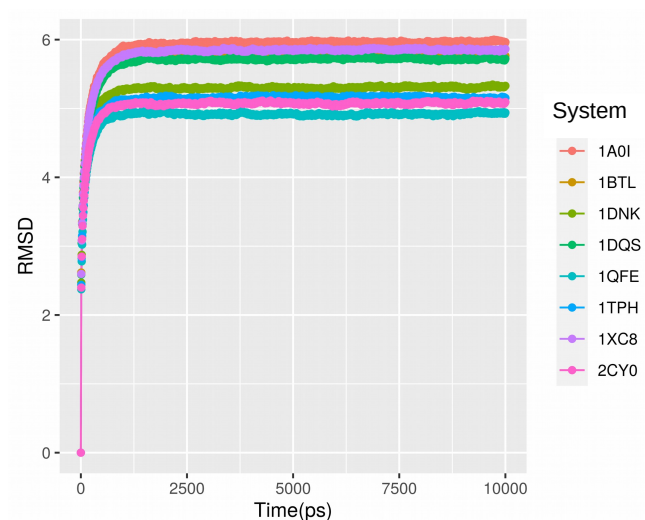


Figura 5.2: Variação do valor de RMSD para cada estrutura durante a simulação de dinâmica molecular.

Na Tabela 5.3 são mostrados os resultados dos cálculos de estrutura eletrônica feitos para os oito complexos em cada nível de tratamento para a produção das coordenadas iniciais. Os descritores globais geralmente não tem relevância para a explicação da reatividade de grandes sistemas, no entanto, são bons indicadores da qualidade do cálculo e podem ser combinados com descritores locais distribuindo alguma propriedade eletrônica global por região. Na Figura 5.3 são mostrados gráficos de barra resumindo alguns dos dados obtidos, organizando-os por nível de tratamento para melhor comparação dos valores. É possível observar, por exemplo, que o calor de formação tende a ser ligeiramente menor nas estruturas que passaram pelo menos por uma minimização de energia, nível de tratamento dois, com exceção do sistema codificado como 1DNK, a Deoxiribonuclease I, que apresenta um valor de calor de formação significativamente menor para o segundo nível de tratamento em relação aos outros dois e também aos outros sistemas.

Tabela 5.3: Resultados dos cálculos de estrutura eletrônica semiempírico.

Sistema	Calor de Formação ($kcal.mol^{-1}$)	Potencial de Ionização (eV)	Afinidade Eletrônica (eV)	Potencial químico eletrônico (eV)	Dureza (eV)	Moleza (eV)	Nível de Tratamento
2CY0	-28539,6	4,986	4,743	-4,8645	0,243	4,115	1
1A0I	-30726,0	6,961	4,431	-5,6960	2,530	0,3952	1
1QFE	-19533,3	8,311	0,232	-4,2715	8,079	0,1237	1
1DNK	-22872,7	5,832	4,795	-5,3135	1,037	0,9643	1
1DQS	-24764,9	8,291	5,518	-6,9045	2,773	0,3606	1
1XC8	-31271,8	4,682	1,526	-3,1040	3,156	0,3168	1
1BTL	-12630,6	7,532	1,444	-4,4880	6,088	0,1642	1
1TPH	-20574,4	8,518	0,324	-4,4210	8,194	0,1220	1
2CY0	-21063,9	8,727	1,882	-5,3045	6,845	0,1460	2
1A0I	-33976,1	7,652	0,278	-3,9650	7,374	0,1356	2
1QFE	-21173,5	8,677	-0,021	-4,3280	8,698	0,1149	2
1DNK	-83695,2	8,872	-0,089	-4,3915	8,961	0,1115	2
1DQS	-32321,1	8,320	0,293	-4,3065	8,027	0,1245	2
1XC8	-32870,5	4,924	3,194	-4,0590	1,73	0,5780	2
1BTL	-29230,1	4,338	0,127	-2,2325	4,211	0,2374	2
1TPH	-22139,5	8,690	0,311	-4,5005	8,379	0,1193	2
2CY0	-21074,4	8,183	1,604	-4,8935	6,579	0,1519	3
1A0I	-35702	7,076	0,364	-3,72	6,712	0,1489	3
1QFE	-20800,9	8,499	0,485	-4,4920	8,014	0,1247	3
1DNK	-26347,3	8,778	0,154	-4,4660	8,624	0,1159	3
1DQS	-32589,1	7,638	0,267	-3,9525	7,371	0,1356	3
1XC8	-31853,8	5,390	0,440	-2,9150	4,950	0,2020	3
1BTL	-32638,6	8,573	1,744	-5,1585	6,829	0,1464	3
1TPH	-30712,6	8,345	0,623	-4,4840	7,722	0,1295	3

Em geral, as coordenadas dos sistemas que passaram por minimização de energia ou equilíbrio, apresentaram propriedades eletrônicas que indicam uma maior estabilidade e menor reatividade, assim como uma aumento significativo na distância HOMO e LUMO e

valores de afinidade eletrônica menores. O valor do Potencial de Ionização também tende a estabilizar em um valor entre 8 e 9 eV para todos os sistemas, depois das etapas de minimização e equilíbrio.

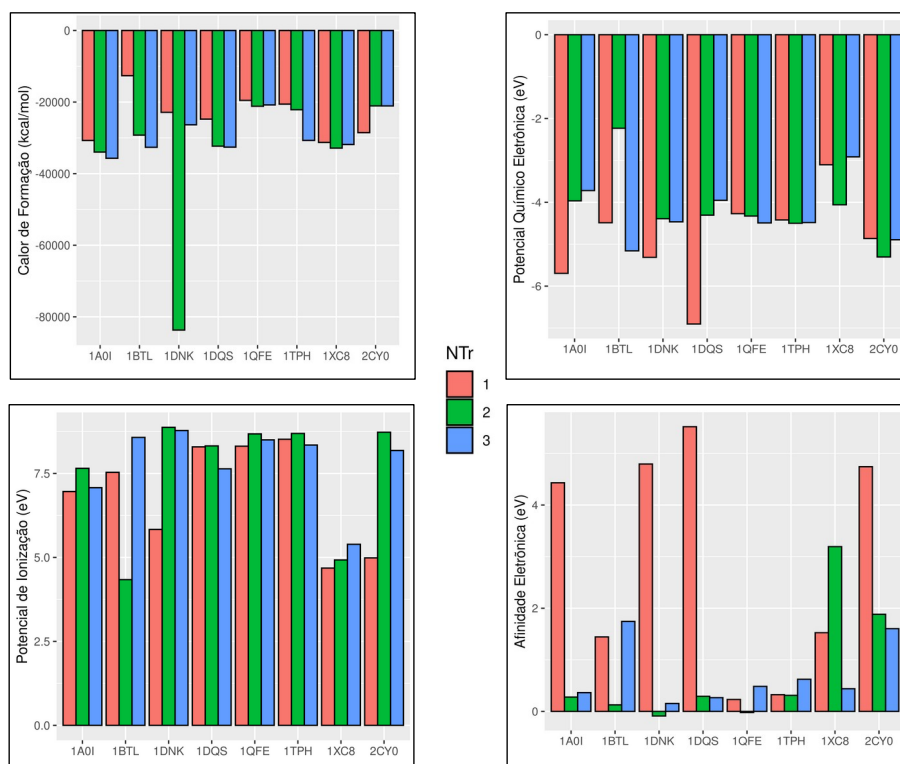


Figura 5.3: Resumo dos descritores globais/propriedades eletrônicas para os sistemas calculados a partir de diferentes métodos de obtenção das coordenadas iniciais.

Depois de realizada a preparação das estruturas e os cálculos de estrutura eletrônica, os resultados do software MOPAC foram alimentados no PRIMoRDiA, produzindo os descritores. Segundo os critérios já explicados na seção anterior, os papéis dos objetos analisados dos sítios ativos foram comparados com os seus papéis dados pela análise dos valores dos descritores, sendo então classificada como correta ou incorreta a predição do comportamento de reatividade. Nas próximas tabelas estão apresentados os resultados para o sistema correspondente ao complexo da DNA-Liase (1A0I), como exemplo. Todas as outras tabelas estão disponíveis no material suplementar do trabalho publicado sobre o PRIMoRDiA.^[38]

Na Tabela 5.4 são mostrados os resultados para as coordenadas preparadas com primeiro nível de tratamento, na Tabela 5.5 com o segundo e na Tabela 5.6 com o terceiro nível. É importante notar que os valores para os objetos que são substratos ou resíduos, os valores para todos dos átomos pertencentes são somados. Na Tabela 5.4 o átomo de fósforo

(PA-ATP) do substrato, que é o centro atacado pelo nitrogênio nucleofílico da lisina-33, foi erroneamente classificado pelos descritores para as coordenadas obtidas com o primeiro nível de tratamento, já que o valor do descritor de eletrofilicidade foi nulo. Já o substrato em si, apresentou valores significativos de eletrofilicidade, indicando que há centros disponíveis para receber essa densidade eletrônica da lisina, que, por sua vez, apresenta valores significativos de nucleofilicidade. Os outros resíduos de lisina também estão classificados corretamente segundo os critérios definidos nesse trabalho, devido ao valor somado de dureza local que supera em mais que o dobro a média para os resíduos de todo o sistema.

Tabela 5.4: Descritores de Reatividade e resultados para o teste de predição dos papéis átomos e resíduos no sistema da DNA-Ligase, com as coordenadas preparadas com o primeiro nível de tratamento.

Elemento	Tipo	Papel	Nucleofilicidade	Eletrofilicidade	Dureza	Resultado da previsão
ATP	substrato	eletrofílico	3,9318	1,5372	237,7153	correta
PA-ATP	átomo	eletrofílico	0,0359	0,0002	5,3925	incorreta
Lys33	resíduo	nucleofílico	0,0992	0,0000	117,5530	correto
NZ-L33	átomo	nucleofílico	0,0676	0,0000	5,5208	correto
Lys239	resíduo	estabilizante eletrostático	0,0134	0,0000	118,9906	correta
Lys238	resíduo	estabilizante eletrostático	0,0209	0,0000	117,9048	correta
Médio	estatística atômica	-	0,0040	0,0003	4,4976	-
Máximo	estatística atômica	-	1,6782	0,8979	5,7504	-
Média	estatística por resíduo	-	0,0592	0,0044	66,1294	-

Na Tabela 5.5 são mostrados os resultados para o segundo nível de tratamento, mostrando já que a alteração estrutural modificou a estrutura eletrônica a ponto do átomo reativo do substrato apresentar valor de eletrofilicidade, resultando em todas as classificações realizadas pelos descritores como corretas. Já as coordenadas extraídas do processo

de equilíbrio resultou novamente em uma estrutura eletrônica que volta a mostrar esse mesmo átomo como não eletrofílico (Tabela 5.6). A causa desse erro pode ser também na dificuldade de localizar orbitais virtuais na molécula de ATP, já que é uma molécula com átomos com alta eletronegatividade e com carga formal negativa. O interessante é perceber a invariabilidade do descritor de dureza local, que manteve valores similares independente das coordenadas utilizadas no cálculo.

Tabela 5.5: Descritores de Reatividade e resultados para o teste de predição dos papéis átomos e resíduos no sistema da DNA-Ligase, com as coordenadas preparadas com o segundo nível de tratamento.

Elemento	Tipo	Papel	Nucleofili- cidade	Eletrofi- licidade	Dureza	Resultado da previsão
ATP	substrato	eletrofílico	1,1067	1,3679	239,1333	correta
PA-ATP	átomo	eletrofílico	0,0144	0,0277	5,4289	correta
Lys33	resíduo	nucleofílico	1,0498	0,1513	117,4125	correta
NZ-L33	átomo	nucleofílico	0,9396	0,0000	5,5134	correta
Lys239	resíduo	estabilizante eletrostático	0,0551	0,0000	119,7894	correta
Lys238	resíduo	estabilizante eletrostático	0,0514	0,0000	116,1209	correta
Média	estatística atômica	-	0,0037	0,0414	4,4834	-
Máximo	estatística atômica	-	0,9396	2,8797	5,7585	-
Médio	estatística por resíduo	-	0,0541	0,6075	65,8499	-

Tabela 5.6: Descritores de Reatividade e resultados para o teste de predição dos papéis átomos e resíduos no sistema da DNA-Ligase, com as coordenadas preparadas com o terceiro nível de tratamento.

Elemento	Tipo	Papel	Nucleofi- licidade	Eletrofi- licidade	Dureza	Resultado da previsão
ATP	Substrato	eletrofílico	0,6176	0,0000	240,2251	incorreto
PA-ATP	Substrato	eletrofílico	0,0064	0,0000	5,5492	incorreto
Lys33	Resíduo	nucleofílico	0,0942	0,9987	116,1576	correto
NZ-L33	Átomo	nucleofílico	0,0001	0,9478	5,5253	incorreto
Lys239	Resíduo	estabilizante eletrostático	0,0209	0,0000	117,9048	correta
Lys238	Resíduo	estabilizante eletrostático	0,0217	0,6184	117,5661	correta
Média	estatística atômica	-	0,0074	0,0002	4,4218	-
Máximo	estatística atômica	-	0,9592	0,9478	5,7185	-
Média	estatística por resíduo	-	0,0922	0,0027	54,9889	-

Os resultados das predições foram compilados e estão resumidos nos gráfico de barras da Figura 5.4, separando os resultados da predição por tipo de papel e nível de tratamento das estruturas. O tipo de papel que apresentou maior dificuldade para a predição pelos descritores foi o de caráter mole-mole, eletrofilicidade/nucleofilicidade, com 77,77% de acerto para o nível de tratamento 1, e 72,22% para os níveis de tratamento 2 e 3. Isso demonstra que é necessário um estudo mais minucioso da estrutura eletrônica quando esse tipo de interação mole-mole está presente na análise que será feita com o uso dos descritores obtidos com o PRIMoRDiA.

Exatamente esses descritores, que advém das funções de Fukui modificadas, são mais sensíveis as escolhas dos parâmetros energéticos escolhidos no *input* do programa. Igualmente, a localização dos orbitais e suas energias são significativamente influenciados pelo tipo de método utilizado para obter a estrutura eletrônica, como os algoritmos de escalonamento linear que utilizam esquemas de localização de orbitais moleculares e imprimem

certos vieses de energia na resolução da estrutura eletrônica.

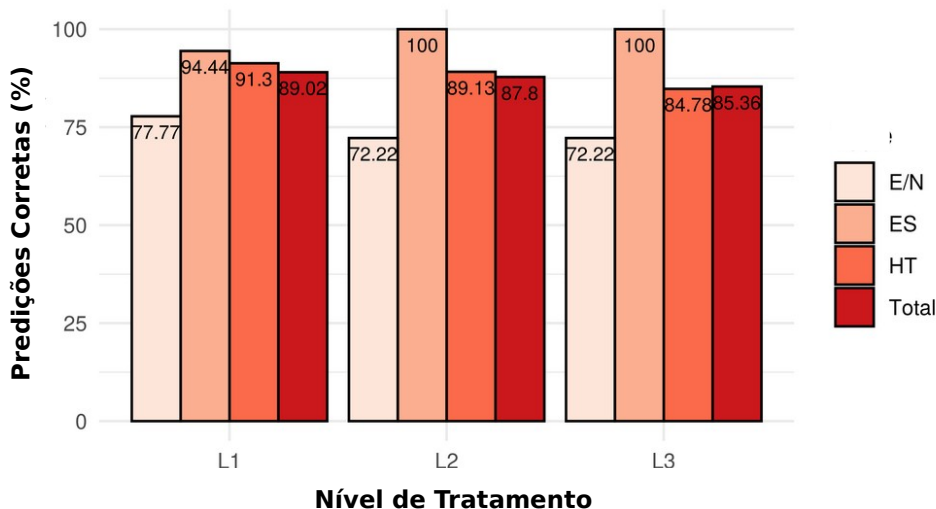


Figura 5.4: Resultados das predições feitas pelos descritores dos papéis catalogados para os elementos do sítios ativos dos complexos enzimáticos. E (eletrofilicidade) N (Nucleofilicidade); ES (Estabilização Eletrostática); HT (transferência de hidrogênio).

Já para os efeitos de estabilização eletrostática, houve uma melhora nas taxas de predição em relação aos do primeiro tratamento, que apresenta um valor de 94.44%, que para os níveis de tratamentos 2 e 3 é 100% de acerto. No entanto, para os outros tipos de papéis de reatividade, a capacidade de predição dos descritores pioram progressivamente com a complexidade do tratamento das estruturas, apesar da diferença não ser mais que 5% quando comparada os valores totais. O tipo de papel que apresentou maior dificuldade de predição pelos descritores foi o de eletrófilo/nucleófilo, que foi causado basicamente por cofatores carregados negativamente, como NAD e ATP, nos quais apresentaram ser mais nucleófilos do que eletrófilos, mesmo sendo esperado que eles recebessem um hidreto carregando um par de elétrons.

Na Figura 5.5 são mostrados os mapas de reatividade para a estrutura Deoxiribonuclease I, onde há também reações de transferência de hidrogênio, nos quais os papéis anotados na base de dados foram confirmados pelos descritores com maior sucesso quando utilizada as coordenadas do último nível de tratamento. A primeira etapa do mecanismo da reação enzimática desempenhada por esse complexo é a ativação de uma molécula de água perto da ligação entre a citosina-13(DC13) e a 14(DC14) da dupla fita que vai ser clivada, que fica propensa a realizar um ataque nucleofílico no átomo de fósforo P da DC14.

Na Figura 5.5:B é possível observar a propensão do oxigênio da água em realizar o

ataque nucleofílico, assim como o nitrogênio da histidina-252 que vai ativar a água. Na Figura 5.5:C é mostrado como os átomos mais duros são os do centro reativo, indicando onde é a região de maior concentração de interações eletrostáticas e propensas a interações duro-duro, como reações de transferência de prótons.

O que é importante notar nesse caso específico é que o primeiro protocolo aplicado pode ter retirado moléculas de água do experimento de cristalografia e ter recolocado outras em posições não favoráveis a apresentar uma reatividade significativa. Somente o último protocolo de preparação de estruturas apresenta etapas de equilíbrio do solvente, o que pode ter influenciado nos resultados incorretos para os níveis de tratamento 1 e 2.

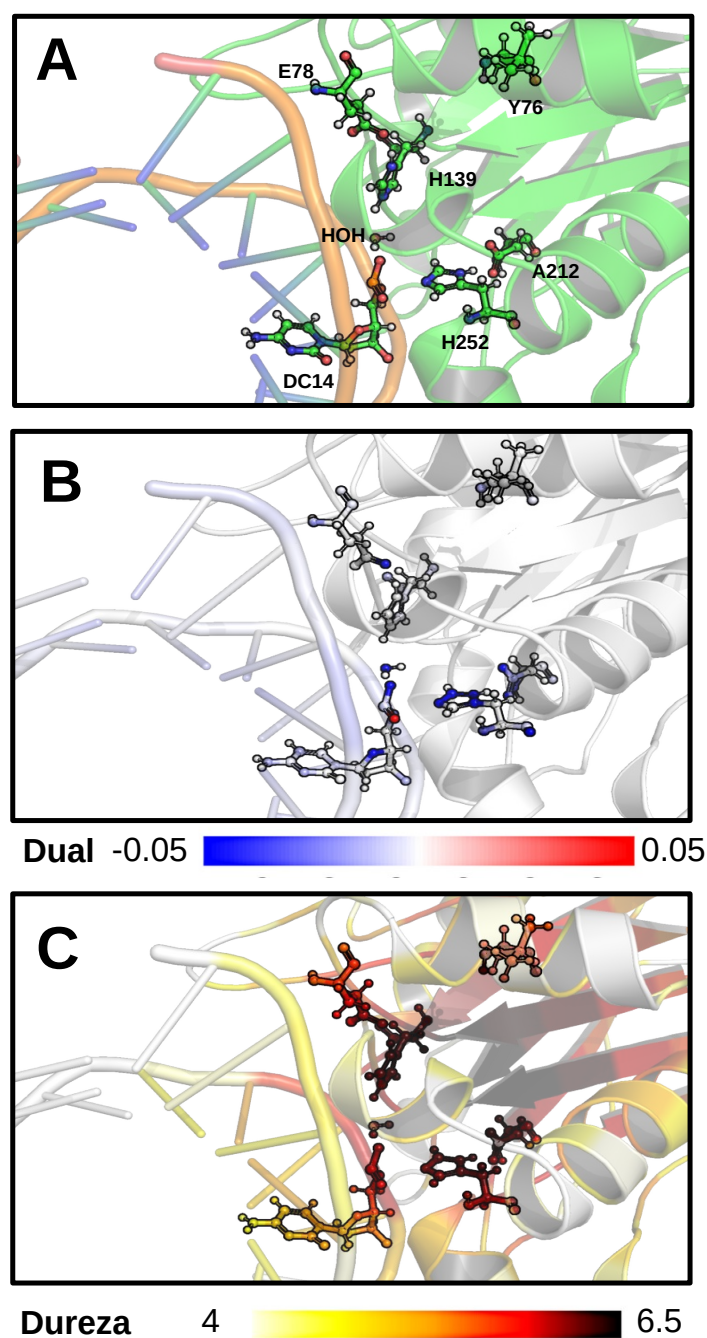


Figura 5.5: Mapas de reatividade para a deoxirribonuclease I (1DNK) calculados a partir das coordenadas obtidas com o terceiro nível de tratamento. (A) Descrição do sítio ativo; (B) Descritor de Netfilicidade; (C) Dureza local.

Constatando que um nível de tratamento de estruturas mais simples possível tem a capacidade de prover coordenadas razoáveis para os cálculos de reatividade, e frente ao problema da pandemia de coronavírus, foram realizados cálculos para uma das primeiras estruturas da protease principal do SARS-CoV-II complexado com um potencial inibidor. O objetivo desses cálculos era prover dados oriundos da estrutura eletrônica para o entendimento de possíveis mecanismos de ligação para inibidores, como no caso do complexado na

estrutura cristalográfica 6LU7, que apresenta o N3, uma molécula desenvolvida com ajuda de métodos computacionais que já é conhecida por conseguir inibir proteases principais de outros coronavírus.

Na Figura 5.6 são mostrados os mapas de reatividade focando no sítio ativo com e sem o potencial inibidor que se liga covalentemente a cys-145, que de fato é um dos resíduos catalíticos da enzima, central para a reprodução viral. Os descritores mostram que o átomo de enxofre da cys-145 que se liga covalentemente com o inibidor N3, se mantém ativado para realizar ataques eletrofílicos. A presença do inibidor altera significativamente a estrutura eletrônica do sistema, tornando o muito mais eletrofílico que nucleofílico, diminuindo a reatividade da histidina-41 da díade catalítica. Os mapas de dureza local mostrados na Figura 5.6C e Figura 5.6D, mostram que as histidina-163 e histidina-164 são os principais pontos de estabilização eletrostática, que podem ser responsáveis por estabilizar tanto o substrato natural quando potenciais inibidores.

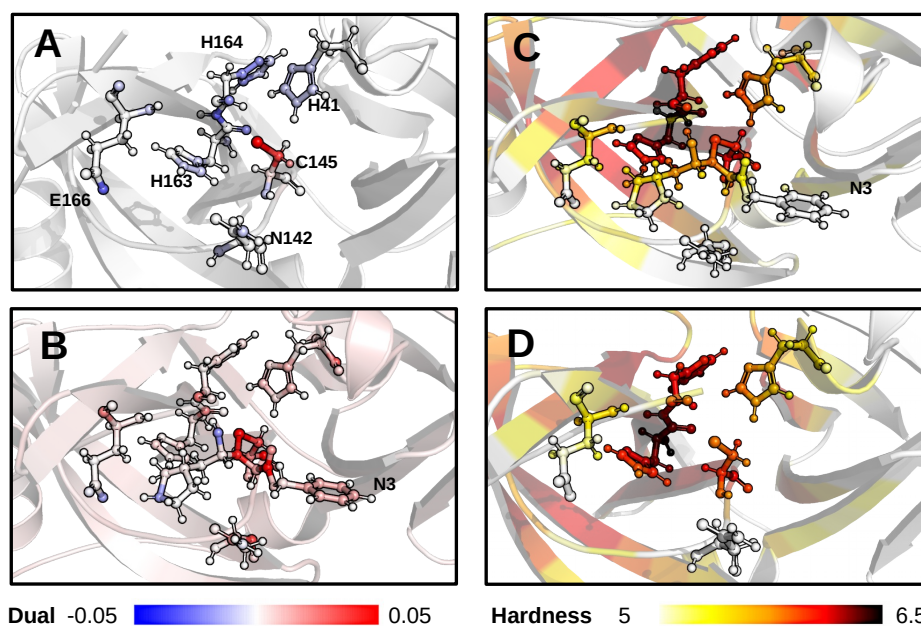


Figura 5.6: Descritores de reatividade para a protease principal da SARS-CoV-II (A) Netfilicidade; (B) Netfilicidade do complexo com o potencial inibidor; (C) Dureza Local; (D) Dureza local para o complexo com o potencial inibidor. Fonte: Grillo e colaboradores. ^[38]

5.3 Considerações Finais

Nesse capítulo foi executado um experimento computacional para testar a capacidades dos descritores, implementados exclusivamente no PRIMoRDiA, de se relacionar com dados

bem consolidados para sistemas biológicos cujo a principal função é a catálise, ou seja, a reatividade química. A combinação de protocolos de preparação de estruturas de baixa exigência computacional com métodos de estrutura eletrônica semiempírico se mostraram suficientes para obtenção de informações de reatividade de resíduos até átomos individuais. Esse tipo de abordagem torna escalonável a análise de conjuntos grandes de estruturas onde a reatividade é pouco conhecida, podendo ser utilizada para determinação de coordenadas iniciais de reação ou em conjecturas de mecanismos. As análises gráficas feitas com imagens produzidas no Pymol são facilitadas por *scripts* gerados no próprio PRIMoRDIA, com exemplos descritos nos tutoriais do programa, criando uma linguagem visual com paletas de cores propostas para cada descritor.

Aplicação: Caracterização Teórica de Reações Enzimáticas

“A ordem surge do caos”

–Ilya Prigogine

A aplicação dos descritores para a racionalização de trajetórias de reações enzimáticas foi desenvolvida em dois trabalhos diferentes que já foram publicados.^[39,40] O primeiro deles apresenta três sistemas enzimáticos, dois com uma coordenada de reação e um com duas. No segundo, a aplicação é voltada apenas para um sistema enzimático com duas coordenadas, onde é calculado também a energia livre da reação que é relacionada com os descritores. Portanto, os descritores já foram aplicados para a caracterização teórica de quatro sistemas enzimáticos.

6.1 Sistemas Enzimáticos

Os sistemas enzimáticos foram os seguintes: triosefosfato isomerase (TIM), haloalcano dehalogenase (HDH), Adenosina Quínase (AQ) e chiquimato desidrogenase (CDH). As etapas dos ciclos catalíticos que foram modelados estão na Figura 6.1.

A TIM catalisa a interconversão do fosfato de dihidróxido-acetona (DHAP) para o fosfato gliceraldeído (GAP). A primeira etapa do mecanismo de catálise é mostrada na Figura 6.1:A, onde o glutamato-165 captura o próton H1 da molécula de DHAP. Na segunda etapa do mecanismo a histidina-96 transfere um próton de volta para o C2 do DHAP completando a isomerização e a formação do GAP. A HIS96 na primeira etapa é importante por que aumenta a acidez do substrato e facilita a extração do primeiro próton e a asparagina-11 cumpre um papel de estabilização eletrostática.^[215] A TIM é uma das enzimas mais caracterizadas que existe, incluindo métodos teóricos,^[216] e é de grande importância para o caminho glicolítico.^[217]

A HDH é uma importante enzima para biocatálise em aplicações industriais, catalisando a hidroxilação de alcanos halogenados com cloreto ou brometos, produzindo álcoois primários.^[218] A primeira etapa do mecanismo foi simulada, mostrada na Figura 6.1:B, que é a dissociação do cloro do dicloroetano (DCE) através do ataque nucleofílico do aspartato-124 no C2 do DCE. No caso desse estudo, a estrutura cristalográfica disponível está com o substrato ligado covalentemente ao ASP124 e o íon cloreto já dissociado. Por esse motivo, a coordenada de reação considerada foi simulada no sentido invertida.

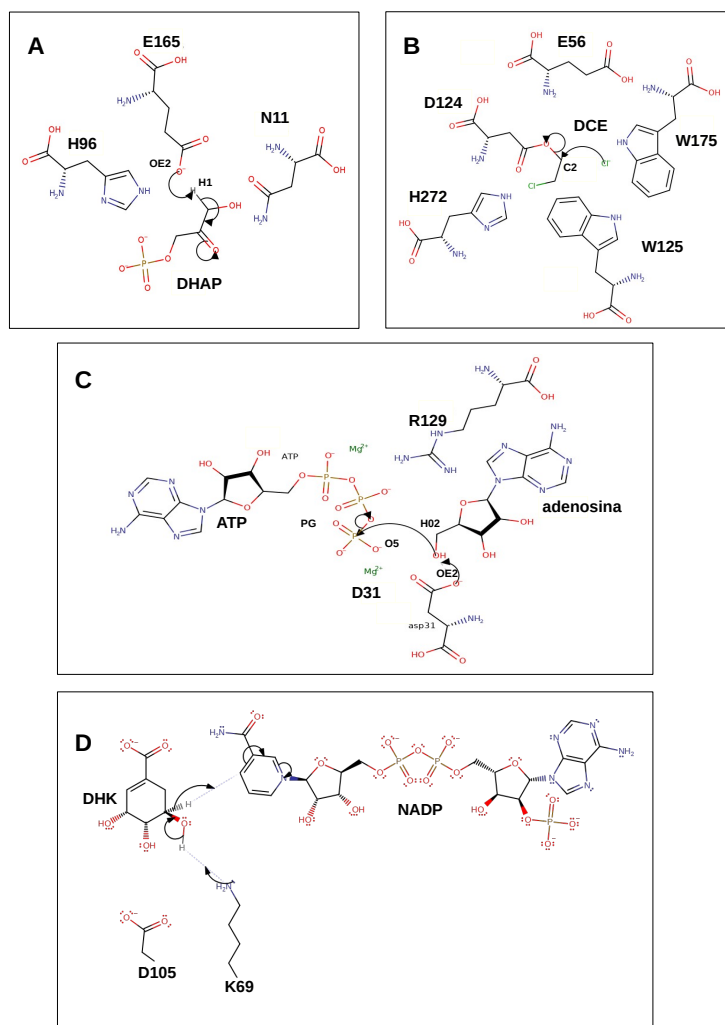


Figura 6.1: Reações enzimáticas modeladas nesse estudo. A) primeira etapa da reação de isomerização dihidróxido-acetona catalisada pela triosefosfato isomerase; B) dissociação do cloro do dicloroetano (DCE) através do ataque nucleofílico do aspartato-124 no C2 do DCE, catalisada pela haloalcano dehalogenase; C) transferência de um grupo fosforila de uma trifosfato adenosina (ATP) para uma molécula de adenosina catalisada pela adenosina quinase; D) redução do oxigênio do grupo carboxila do chiquímato pela captura do próton e a transferência de um hidreto do carbono ligado a hidroxila para o cofator NADP. Fonte: Imagens retiradas do banco de dados *Mechanism and Catalytic Site Atlas*.^[6]

A AQ é uma enzima importante para a regulação do metabolismo de purinas, sendo uma enzima central para o metabolismo do protozoário parasitário *Toxoplasma Gondii*,^[219] o que por sua vez o torna alvo recorrente de estudos de desenvolvimento de fármacos.^[220] Essa enzima catalisa a transferência de um grupo fosforila de uma trifosfato adenosina (ATP) para uma molécula de adenosina, formando uma monofosfato adenina (AMP). Nesse estudo foram monitorados duas coordenadas de reação para essa enzima, a da transferência do

grupo fosforila e a extração de um próton da adenosina por uma aspartato da enzima, que ativa o substrato, como mostrado na Figura 6.1:C. Mesmo com esse mecanismo proposto, ainda não é certo se a ativação da adenosina ocorre de forma simultânea ou por etapas em relação a transferência do grupo fosforila, e não foram encontrados estudos de modelagem teórica dessa reação até o momento.

A última enzima utilizada nesse estudo, a CDH, catalisa uma das etapas do caminho do chiquimato, importante substrato de microrganismos e plantas para a síntese de diversos aminoácidos essenciais, não sendo presente em humanos, logo possuindo alto potencial de alvo para desenvolvimento de fármacos.^[47] Em específico, a reação de catálise é a redução do oxigênio de carboxila pela captura do próton e a transferência de um hidreto do carbono ligado a hidroxila para o cofator NADP. Na simulação realizada nesse estudo foram utilizadas essas duas coordenadas de reação, partindo dos produtos para os reagentes, ou seja, a transferência do hidreto e do próton para o chiquimato.

6.2 Simulação dos Caminhos de Reação

Genericamente, os métodos de simulação empregados para todos esses sistemas estão disposto no fluxograma da Figura 6.2. A parametrização e as simulações de dinâmica molecular utilizadas estão descritas na seção de métodos dessa Tese. Para o trabalho que usou os três sistemas enzimáticos, a preparação do sítio ativo foi mais conservadora em relação a variação conformacional do sistema, buscando preservar as informações experimentais obtidas em cristalografia.

Portanto, para esses sistemas, as coordenadas e topologias foram carregadas no software EasyHybrid,^[195] onde foi conduzido uma minimização de energia usando mecânica molecular, seguido de uma otimização de geometria com o potencial QM/MM. Já para a CDH foi realizado um protocolo de equilíbrio usando simulação de dinâmica molecular no GROMACS: 100 ps de termalização em NVT; 100 ps de pressurização em NPT; e 5ns de produção NPT, salvando as coordenadas a cada 20 ps, gerando 2500 estruturas para análise subsequente. A estrutura utilizada para os cálculos QM/MM é retirada da análise de estrutura que tem as distâncias mais prováveis entre os átomos das coordenadas de reação.

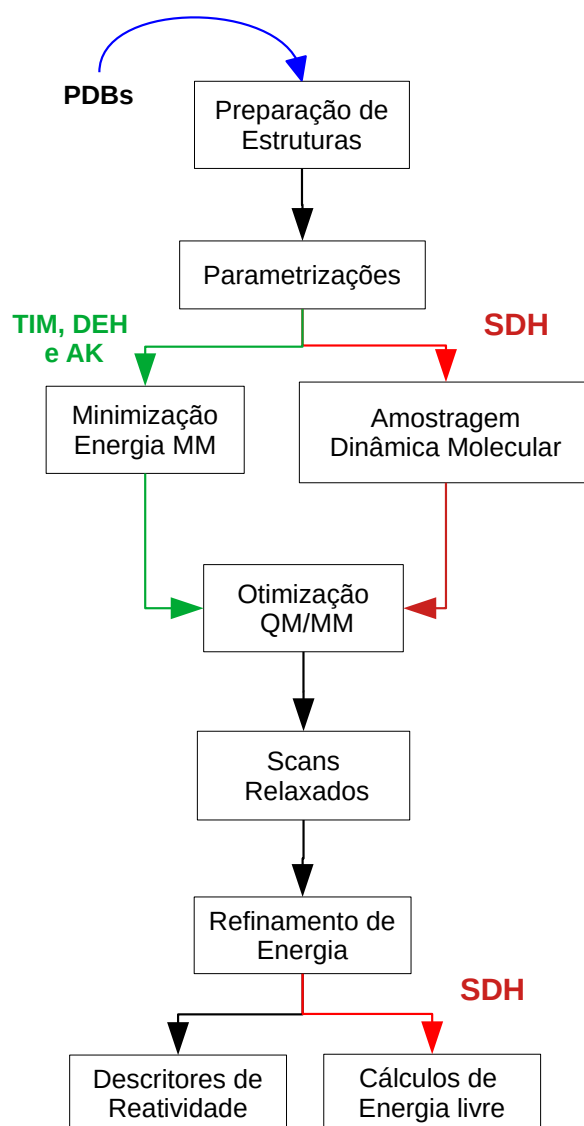


Figura 6.2: Fluxograma resumindo os métodos computacionais utilizados nesses estudos com as diferentes enzimas. Fonte: Imagem do autor.

Os caminhos de reação para cada um desses sistemas foram gerados através de simulações de QM/MM de *scans* relaxados. Para o primeiro grupo de enzimas, os métodos de estrutura eletrônica empregados na região QM são os hamiltonianos semiempírico: AM1, RM1, PM3, PM6 e PM7. Já para a CDH os métodos foram: AM1, RM1, PM3, PM6, PM7, AM1/d-PhoT, B3LYP/3-21G e B3LYP/6-31+G*.

Na Figura 6.3 é mostrado o resultado dos perfis de energia para as simulações QM/MM para os três sistemas simulados no primeiro estudo.^[40] Para a TIM e HDH nas quais somente uma coordenada de reação foi monitorada em cada, foi possível colocar os perfis de energia calculados para todos os métodos de energia utilizados. No caso da AQ, foi escolhido somente o melhor Hamiltoniano, no quesitos de valor da barreira energética e

razoabilidade do caminho de reação descrito na superfície de energia.

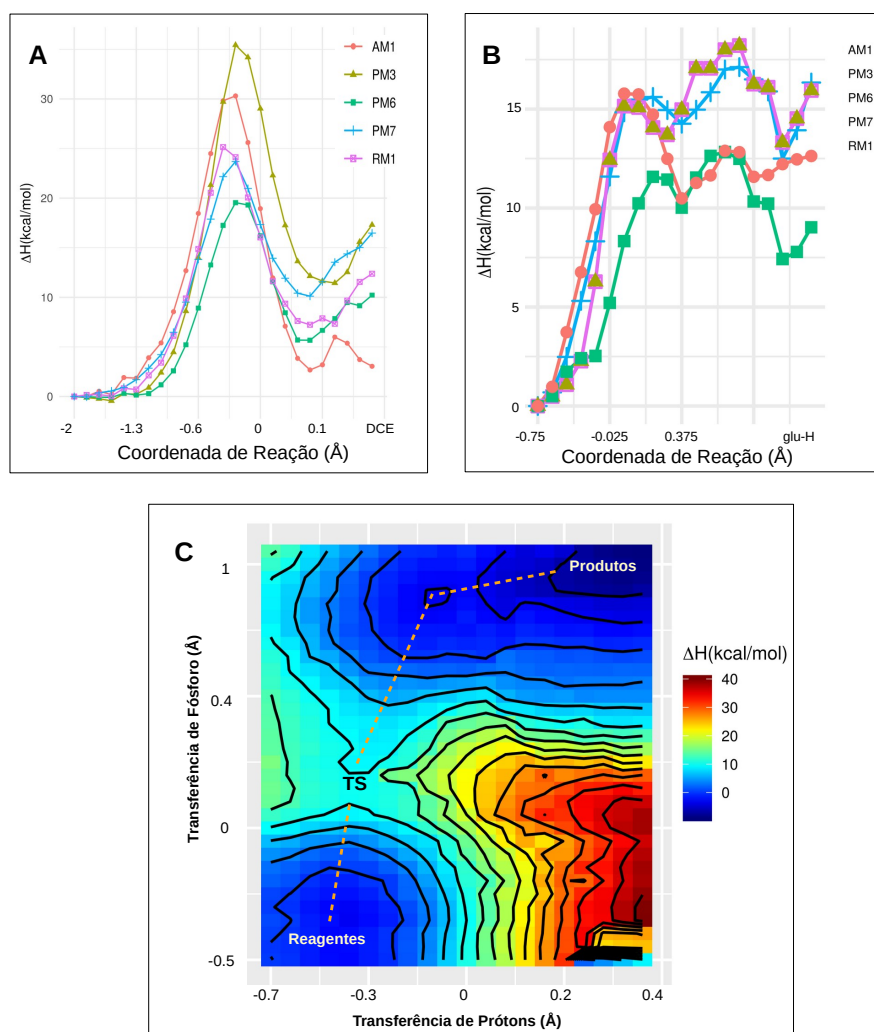


Figura 6.3: Superfície de energia potencial para os três primeiros sistemas enzimáticos simulados A) Reação catalisada pela haloalcano dehalogenase; B) Reação catalisada pela triosefosfato-isomerase; C) Reação catalisada pela adenosina quínase com método semiempírico PM7. Fonte: imagem adaptada das figuras do trabalho de Grillo e colaboradores.^[40]

Na simulação reversa da coordenada de reação da dissociação do cloreto na reação da HDH, foi encontrada uma barreira de 24 kcal/mol e uma variação de entalpia de aproximadamente 10 kcal/mol, o que corrobora o que foi encontrado em outros estudos computacionais que já tinham correlacionado com informações experimentais do trabalho de Kennes e colaboradores.^[221] Com exceção do PM6, todos os outros métodos mostraram uma barreira de em torno 15 kcal/mol para a abstração do próton na reação da TIM, concordando com a barreira experimental de 13 kcal/mol. Já para a AQ, ainda não havia nenhum estudo

de determinação de mecanismo, seja experimental ou computacional, e nossa colaboração nesse artigo mostrou padrões semelhantes a reação catalisada pela enzima mevalonato quinase.^[222] Usando a variação do calor de formação obtido com cálculos de PM7 foi indicado um estado de transição com uma barreira de energia de 10,22 kcal/mol e uma calor de reação de -8,86 kcal/mol.

Na Figura 6.4 é apresentado as principais estruturas da superfície de energia potencial da CDH e o mapa de contorno da energia potencial em função da transferência do próton e do hidreto, apresentando uma barreira de aproximadamente 25 kcal/mol (100 kJ/mol). Nas duas reações que foram simuladas com duas coordenadas, os mapas de energia potencial demonstraram reações de duas etapas. Para a AQ, primeiro ocorre a transferência do grupo fosforila, seguido da abstração do próton da adenosina. E no caso da CDH, primeiro ocorre a transferência do hidreto seguido da transferência do próton. Como já mencionado, vamos usar as informações da energia de superfície potencial para explorar o que pode ser estudado com os descritores, o detalhamento e discussão dessas simulações estão nos trabalhos publicados pelo nosso grupo.^[39,40]

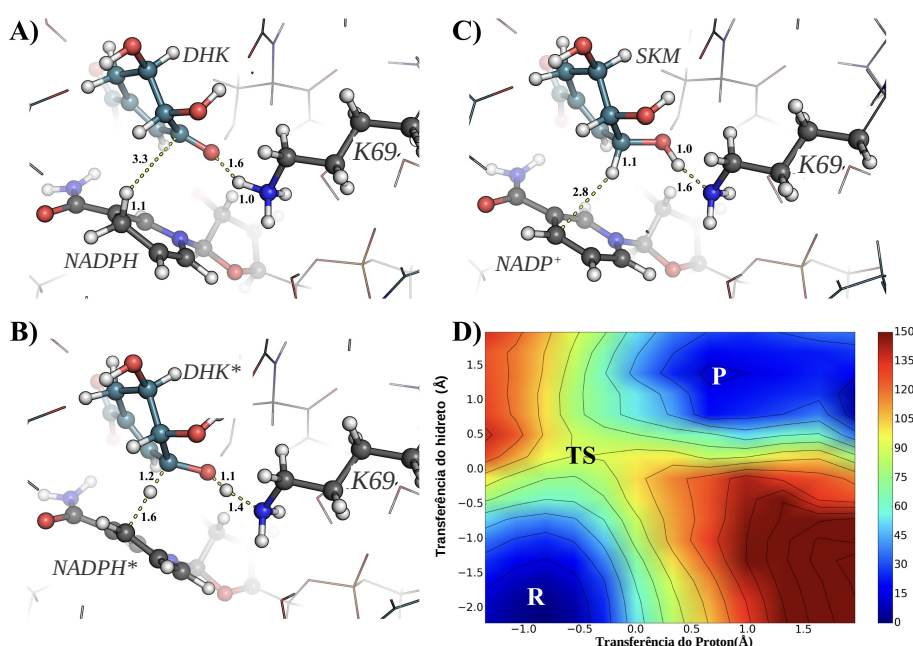


Figura 6.4: Estruturas chave e superfície de energia potencial para as duas coordenadas da reação catalisada pela chiquímato desidrogenase calculadas com potencial B3LYP/6-31+G*/MM. A) Complexo enzima-substrato na geometria inicial (P) da simulação das coordenadas de reação; B) Estrutura referente ao estado de transição correspondente (TS) a superfície de energia potencial; C) Estrutura referente ao produto da reação (P); D) superfície de energia potencial em kJ/mol.

6.3 Análise de Reatividade

As reações dos sistemas enzimáticos TIM, HDH e AQ seguiram um protocolo de simulações diferente do CDH, pois foram desenvolvidos em publicações diferentes. Para o primeiro grupo, as quatro equações de trabalho para a dureza local foram calculadas, assim como descritores de interação de pares, que são descritores para os átomos que estão formando a nova ligação química do processo simulado. Já para o CDH, só uma equação de trabalho foi explorada, pois no primeiro trabalho com as três enzimas já tinha sido feito uma avaliação desses métodos. Ainda no trabalho com a CDH, foi realizada uma análise da variação dos descritores nos resíduos do sítio ativo para a trajetória de reação. Com a trajetória gerada para cada reação estudada, o número de átomos na região QM é significativamente aumentado para refinamento de energia usando vários métodos de estrutura eletrônica. A partir desses cálculos o sítio ativo foi analisado com os descritores de reatividade calculados pelo PRIMoRDiA para o caminho de reação, usando o método EW para as funções de Fukui com banda de 3 eV.

Ainda para o trabalho com os três sistemas enzimáticos, foi utilizado um descritor de interação por pares, que para o descritor de dureza local se mostrou útil para a análise da coordenada de reação. A definição da dureza de pares atômicos (DPA) é dada pela Equação 6.1, sendo como o produto da dureza local H_{α}^{σ} para o átomo α e a dureza local H_{β}^{σ} do átomo β , obtidos com a equação de trabalho σ , ponderado pela distância $R_{\alpha\beta}$ entre os dois átomos.

$$DPA_{\alpha\beta}^{\sigma} = \frac{H_{\alpha}^{\sigma} H_{\beta}^{\sigma}}{R_{\alpha\beta}} \quad (6.1)$$

6.3.1 Triosefosfato Isomerase

Na Figura 6.5, é mostrada a variação do descritor de dureza de pares atômicos para os átomos de oxigênio OE2 do GLU165 e do hidrogênio H1 do DHAP, que foram calculados a partir de resultados de cinco tipos de Hamiltonianos semiempíricos e usando quatro equações de trabalho para dureza local. Para essa trajetória, a única equação de trabalho utilizada para calcular o DPA que não mostrou um perfil de reação razoável foi a baseada na distribuição de Fukui (Figura 6.5:D). É possível observar que, para o DPA baseado na interação elétron-elétron, o potencial eletrostático molecular (PEM) e o potencial químico local apresentam um valor máximo perto do estado de transição. Como seria esperado para uma reação naturalmente governada por interações duro-duro, a transferência de próton

estudada nesse caso não conseguiu ser descrita por uma equação de trabalho de dureza com alta dependência da função de Fukui.

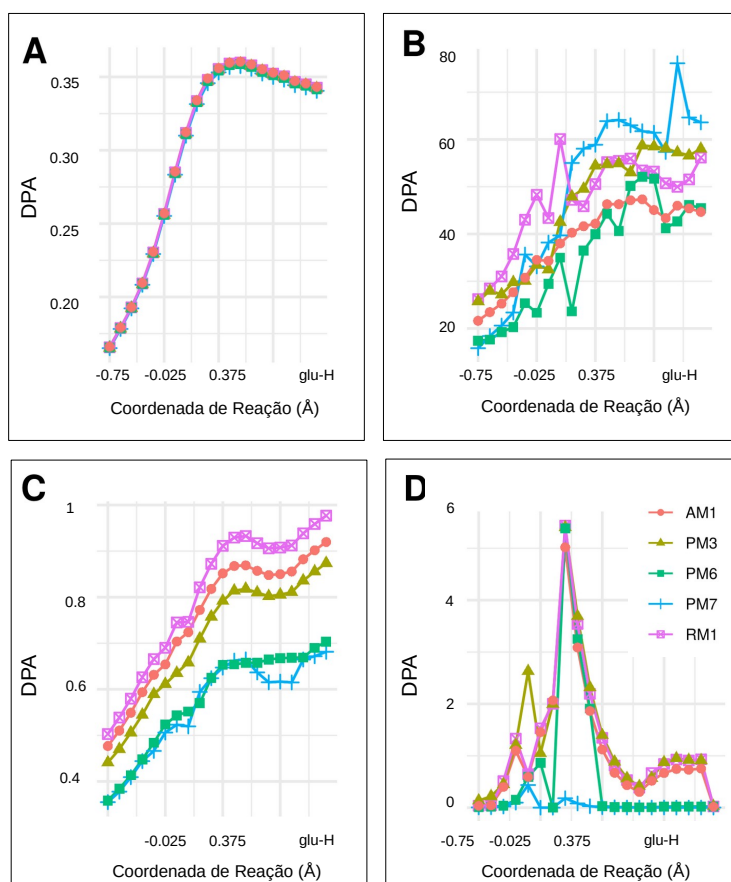


Figura 6.5: Descritor de Dureza de Pares Atômicos (DPA) entre os átomos OE2 e H1 diretamente envolvidos na reação catalisada pela triosefosfato-isomerase, calculado a partir da estrutura eletrônica de diferentes Hamiltonianos semiempíricos e equações de trabalho para dureza local: A) Dureza local estimada usando a aproximação de interação elétron-elétron do potencial eletrostático molecular; B) potencial de Fukui; C) potencial químico local; D) distribuição pela função de Fukui. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

Entretanto, como é esperado numa transferência de carga do glutamato para o substrato, para a captura do próton, podemos observar alguns padrões na nucleofilicidade do oxigênio OE2(GLU165) e da eletrofilicidade do H1(DHAP), como mostrado na Figura 6.6. Para todos os Hamiltonianos, a nucleofilicidade do OE2(GLU165) apresenta máximo valor encontrado no sítio ativo da TIM, apresentando uma queda abrupta quando a coordenada de reação evolui para o estado de transição. Já o H1(DHAP), inicialmente apresenta um valor significativamente baixo de eletrofilicidade, aumentando até chegar ao máximo para

o estado de transição e depois cai abruptamente.

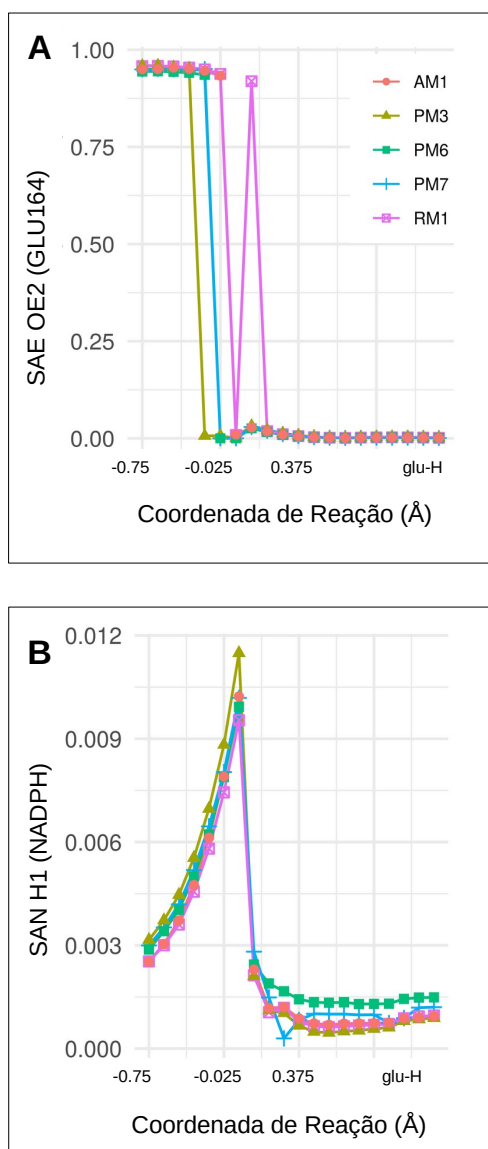


Figura 6.6: Descritores para os átomos da reação catalisada pela triosefosfato isomerase. A) Susceptibilidade ao ataque eletrofílico (SAE), mesmo que nucleofilicidade, para o átomo de oxigênio (OE2) do glutamato; B) Susceptibilidade ao ataque nucleofílico (SAN), mesmo que eletrofilicidade, para o átomo de hidrogênio que é capturado do substrato. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

Na Figura 6.7 são mostrados os descritores de dureza local (aproximação elétron-elétron do PEM) e descritor dual (diferença entre f^+ e f^-), para o sítio ativo da TIM nas estruturas consideradas de maior interesse da coordenada de reação. É possível observar que os maiores valores de dureza local se concentram nos átomos do substrato, e para o GLU165 o oxigênio OE2 apresenta maior dureza local que o OE1, ou seja, já sendo indicado pelo

descriptor como o átomo mais provável para participar em uma reação de caráter duro-duro.

Essa diferença entre os átomos OE1 e OE2 aumenta no estado de transição e depois diminui nos produtos. Os átomos do substrato se mantêm como os mais duros, assim como o próton que já foi transferido e o próton da HIS96, que vai ser transferido na próxima etapa do mecanismo de inversão de isomeria da molécula do DHAP. Na Figura 6.7E e F é mostrado o descriptor dual para a geometria inicial e para a geometria final respectivamente. Para os valores negativos em azul estão coloridos os átomos mais nucleófilos, como os dois oxigênios do grupo carboxílico do GLU165.

Esse comportamento para esses átomos se reverte na geometria dos produtos, aparecendo mais como eletrófilos. Já o substrato na geometria final vira um forte nucleófilo na região da dupla ligação formada quando o próton foi retirado e o oxigênio ligado ao C2 do substrato está ativado para um ataque nucleofílico, condizendo corretamente com o próximo passo do mecanismo que é a captura do próton do HIS96.

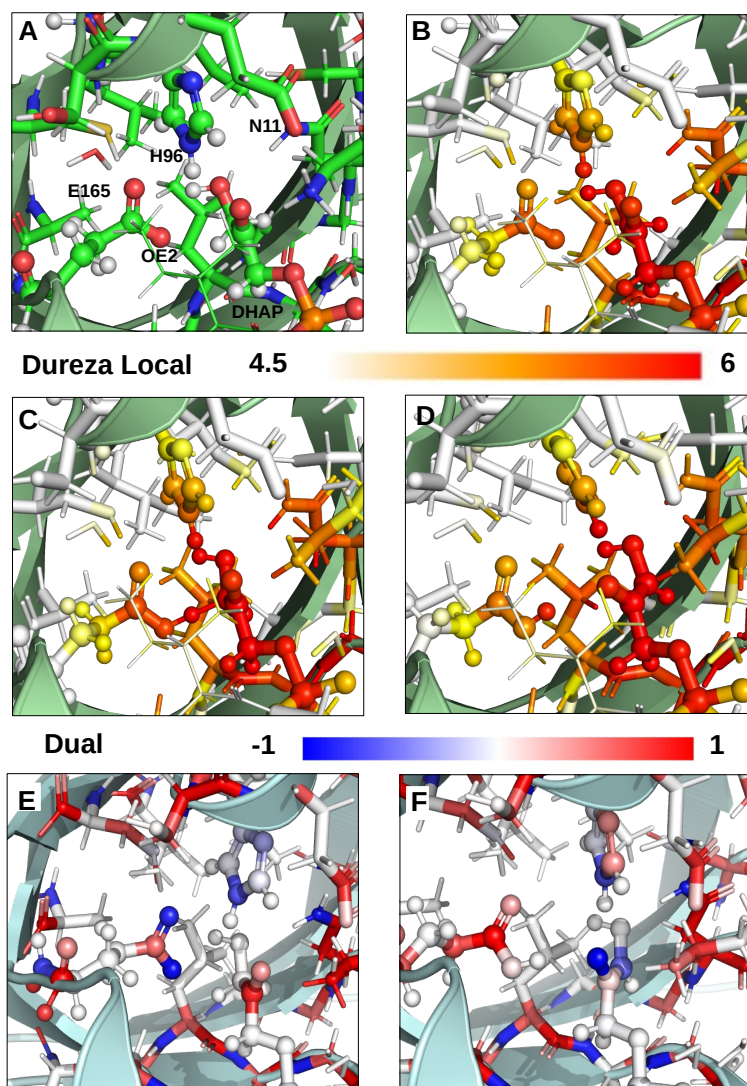


Figura 6.7: Mapas de reatividade para as principais estruturas da coordenada de reação catalisada pela triosefosfato-isomerase calculados a partir da estrutura eletrônica obtida com Hamiltoniano AM1. A dureza local avaliada nesses imagens foi a calculada com o método de aproximação das interações elétron-elétron do potencial eletrostático molecular: A) Descrição do sítio ativo para a geometria inicial; B) dureza local para a geometria inicial. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

6.3.2 Haloalcano Dehalogenase

Dos resultados dos descritores para a HDH, na Figura 6.8 é apresentado os perfis de DPA entre o átomo C2(DCE) e o Cl^- . É possível observar que existe um aumento da interação duro-duro durante a ligação entre esses átomos, com exceção da equação de trabalho baseada na distribuição da função de Fukui (Figura 6.8:D). No entanto, não ficou claro por esse perfil de DPA uma mudança que caracterizasse um estado de transição, podendo o

aumento do valor se dar somente pelo denominador da distância no cálculo da DPA.

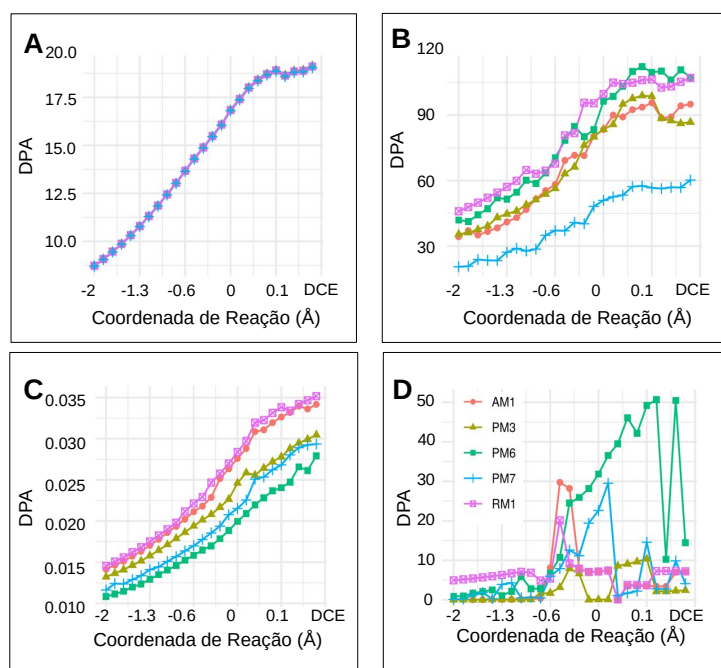


Figura 6.8: Descritor de Dureza de Pares Atômicos (DPA) entre os átomos C2 e Cl diretamente envolvidos na reação catalisada pela haloalcano dehalogenase, calculado a partir da estrutura eletrônica de diferentes Hamiltonianos semi-empíricos e equações de trabalho para dureza local: A) Dureza local estimada usando a aproximação de interação elétron-elétron do potencial eletrostático molecular; B) potencial de Fukui; C) potencial químico local; D) distribuição pela função de Fukui. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

Na Figura 6.9 é mostrada a variação da dureza local nos átomos Cl e C2(DCE), calculados com a equação de trabalho baseada na interação elétron-elétron do PEM. Nesses perfis de dureza por átomo é possível ver que um máximo é atingido no estado de transição, que é um pouco antes do 0 Å da coordenada de reação. No estado final, a dureza final dos dois átomos é menor que a inicial. O aumento das interações duro-duro em direção ao estado de transição pode ter relação com o fato da quebra de ligação covalente entre o oxigênio do aspartato e o C2, que permaneceu próximo participando das interações eletrostáticas conforme o cloreto também se aproximava.

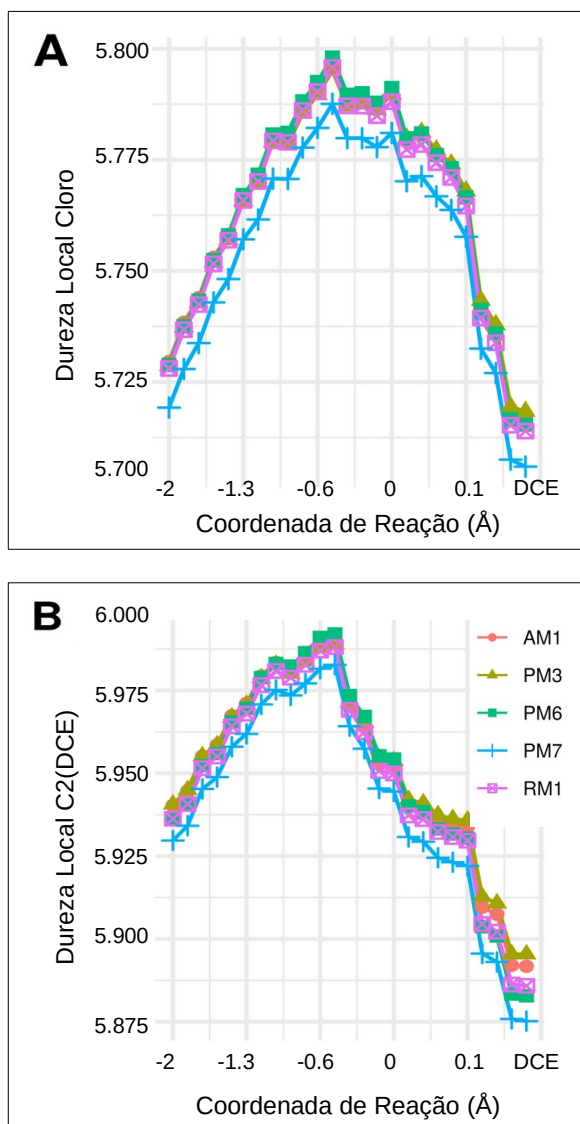


Figura 6.9: Variação do descritores de dureza local, obtido com a aproximação de interação elétron-elétron do potencial eletrostático molecular, na reação catalisada pela haloalcano dehalogenase: A) para o átomo de cloro; B) para átomo C2 do substrato. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

Na Figura 6.10, o sítio ativo da HDH e a dureza local para as estruturas chave na coordenada de reação são mostradas. É possível observar uma leve diminuição da dureza local nos átomos do aspartato no estado final, e os resíduos glutamato-65 e triptofano-125 apresentando também os maiores valores de dureza nos átomos da sua cadeia lateral, indicando uma interação eletrostática importante no sítio ativo, como já apontado em outras análises para a estabilização do cloreto.

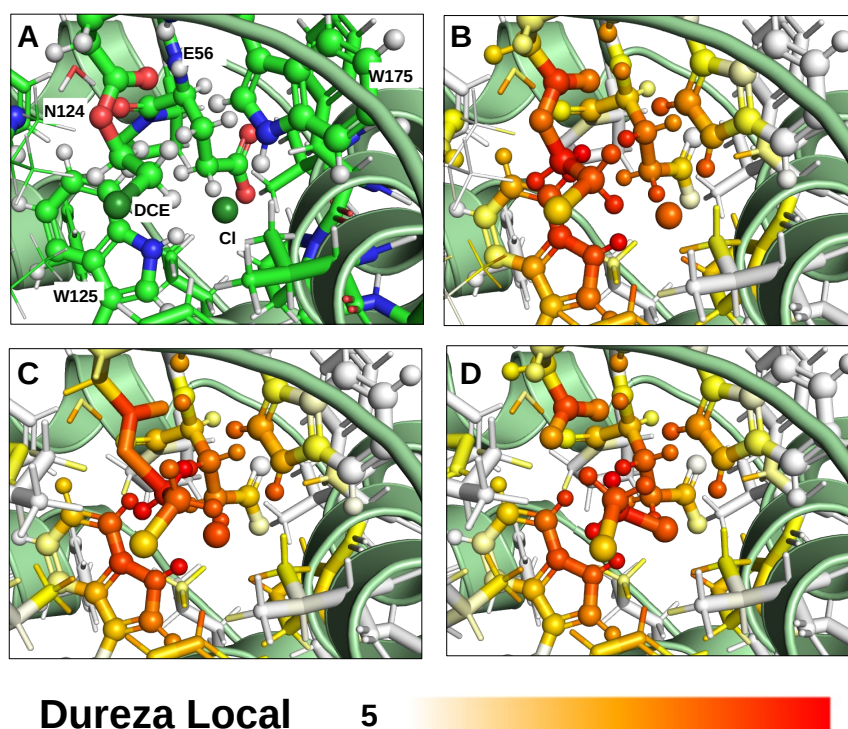


Figura 6.10: Mapas de reatividade para as principais estruturas da coordenada de reação catalisada pela haloalcano dehalogenase calculados a partir da estrutura eletrônica obtida com Hamiltoniano RM1. A dureza local avaliada nessas imagens foi a calculada com o método de aproximação das interações elétron-elétron do potencial eletrostático molecular: A) Descrição do sítio ativo para a geometria inicial; B) dureza local para a geometria inicial. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

6.3.3 Adenosina Quínase

Na Figura 6.11, é mostrada a descrição do sítio ativo da AQ para as geometrias iniciais (A), estado de transição (B) e produtos (C). Na geometria inicial, os átomos de fósforo PA do ATP está a 3,2 Å do oxigênio da molécula de adenosina, que está ligado a um próton, o qual, por sua vez, está a uma distância de 1,9 Å do aspartato-31. As coordenadas otimizadas do estado de transição e dos produtos foram determinadas pela superfície de energia potencial mostrada na Figura 6.3, que tiveram os mapas de reatividade produzidos e analisados. O mecanismo proposto foi, até onde foi encontrado na literatura, a primeira proposta para a catálise realizada pela AQ, que pelo perfil energético e análise estrutural, foi definido que a transferência do grupo fosforila ocorre primeiro, seguido da transferência do próton para o aspartato-31. Análogo a superfície de energia potencial mostrada na Figura 6.3.

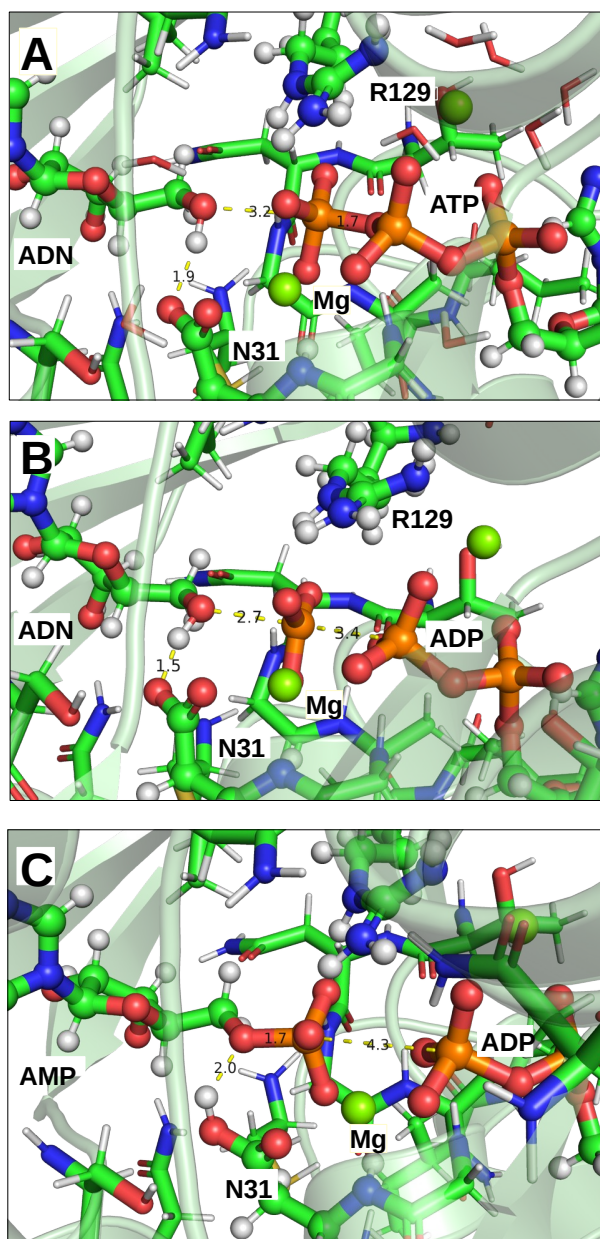


Figura 6.11: Descrição do sítio ativo e geometrias das principais estruturas da trajetória da reação catalisada pela adenosina quinase. A) geometria inicial; B) estado de transição; C) produtos. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

Análogo ao como foi mostrado, a superfície de energia potencial para as duas coordenadas de reação na Figura 6.3, os valores de DPA para as quatro equações de trabalho utilizadas nesse estudo estão apresentada na Figura 6.12, considerando os átomos de oxigênio da adenosina e o fósforo gama do ATP. Com exceção da quarta equação de trabalho, baseado na distribuição de Fukui, as interações duro-duro entre os átomos que vão formar a ligação tendem a aumentar na geometria dos produtos, indicando que há uma ganho de

interações duro-duro na direção da formação dessa ligação.

Para a terceira equação de trabalho utilizada, usando o potencial de Fukui mostrado na Figura 6.12:C, observou-se que a mesma foi a que mais se assemelhou ao padrão de variação de energia potencial, tendo um aumento das interações duro-duro maior na direção da transferência do grupo fosforila e com um leve decréscimo nas estruturas de estado de transição. Esses detalhes nos revelou que nesse ponto mais reativo, as interações duro-duro diminuem ligeiramente, provavelmente pelo fato de ser um tipo de ligação de caráter mais covalente, com alta transferência de carga.

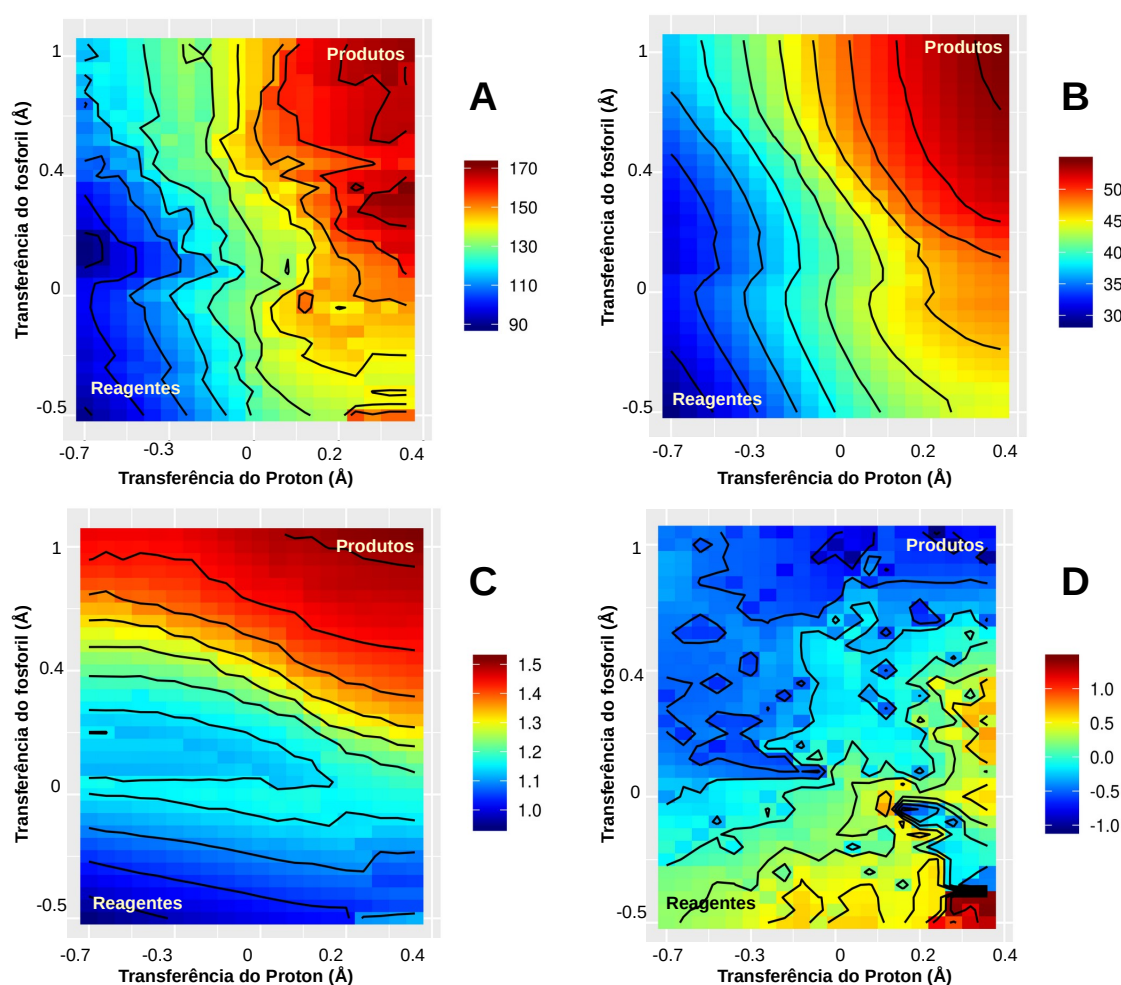


Figura 6.12: Descritor de Dureza de Pares Atômicos (DPA) entre os átomos O3 e PG diretamente envolvidos na reação catalisada pela adenosina quinase, calculado a partir da estrutura eletrônica do Hamiltoniano semi-empírico PM7 com diferentes equações de trabalho para dureza local: A) aproximação de interação elétron-elétron do potencial eletrostático molecular; B) potencial de Fukui; C) potencial químico local; D) distribuição pela função de Fukui. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

Na Figura 6.13, os descritores de netfilicidade e de dureza local são apresentados para as estruturas chave da trajetória simulada para a AQ. Nos mapas de dureza local, Figura 6.13:A, C e E, o grupo guanidínio da arginina-129, e o cátion Mg^{+2} apresentam os maiores valores desse descritor. São elementos que são considerados importantes na estabilização da carga negativa do grupo fosforila que é transferido, estando em uma distância do átomo de fósforo mais próximo de aproximadamente de 2Å e, portanto, com esse papel de estabilização eletrostática identificado pelo descritor utilizado.

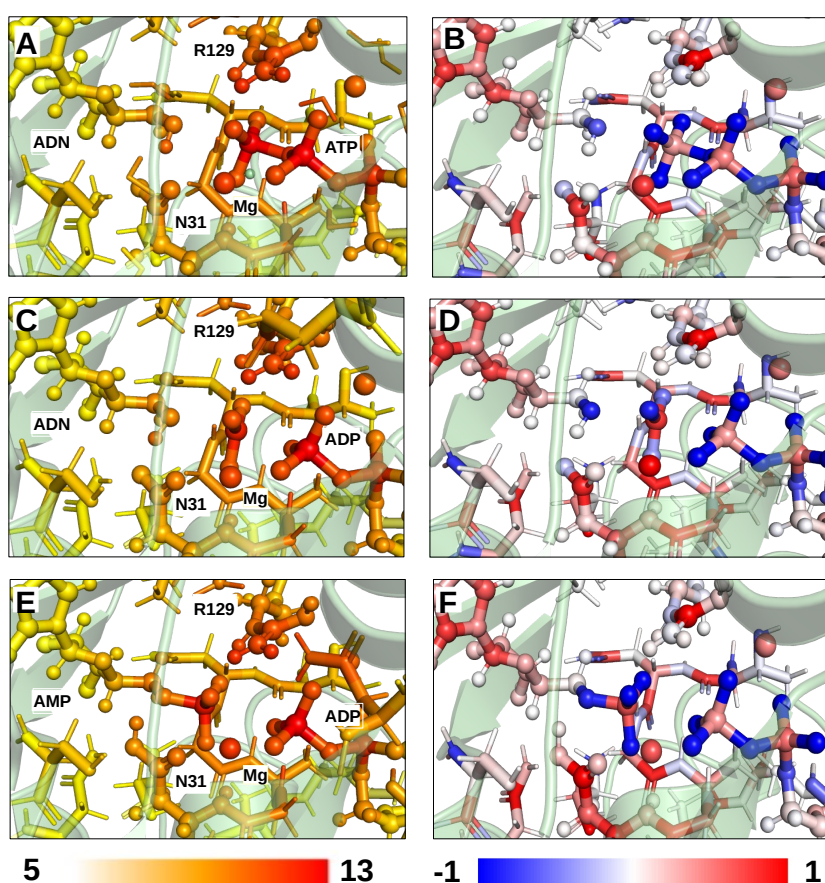


Figura 6.13: Mapas de reatividade para as principais estruturas da reação catalisada pela adenosina quinase, com a estrutura eletrônica calculada com o Hamiltoniano PM7. A) dureza local para geometria inicial; B) netfilicidade para geometria inicial; C) dureza local para o estado de transição; D) netfilicidade para o estado de transição; E) dureza local para o produto, F) netfilicidade para o produto. A dureza local calculada com a aproximação do potencial de Fukui. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

Para a netfilicidade, calculada pela diferença entre a eletrofilicidade e nucleofilicidade nos átomos, a arginina-129 e o Mg^{+2} apresentam tendência de receber transferência de

densidade eletrônica. Da mesma maneira, o oxigênio do grupo fosforila mais próximo desses grupos apresenta um valor mais negativo do descritor dual do que os outros dois oxigênios mais distantes no estado de transição (Figura 6.13:D), indicando que os grupos carregados positivamente deslocalizam a carga do grupo fosforila e o estabilizam, sendo a hipótese mais provável do efeito da enzima no abaixamento da barreira de energia dessa reação. Somando a isso, o descritor ainda mostra como a nucleofilicidade do oxigênio da ribose aumenta no estado de transição, assim como a eletrofilicidade do fósforo do grupo transferido.

Enquanto isso, a propensão do oxigênio do aspartato-31 de realizar um ataque nucleofílico para capturar o próton da ribose se mantém baixa mesmo no estado de transição, indicando que é mais provável que ocorra a transferência do grupo fosforila primeiro para a ribose, ocasionando a formação do AMP, antes da captura desse próton. Para confirmar isso, o valor de eletrofilicidade nesse oxigênio do aspartato foi monitorado para as duas coordenadas da reação, como mostrado na Figura 6.14, onde o valor do descritor aumenta de forma substancial na direção da transferência do grupo fosforila, indicando que sua reatividade atinge o máximo apenas depois do estado de transição determinado pelos cálculos de energia potencial, e depois se desativa novamente nos produtos, já que o próton já foi capturado.

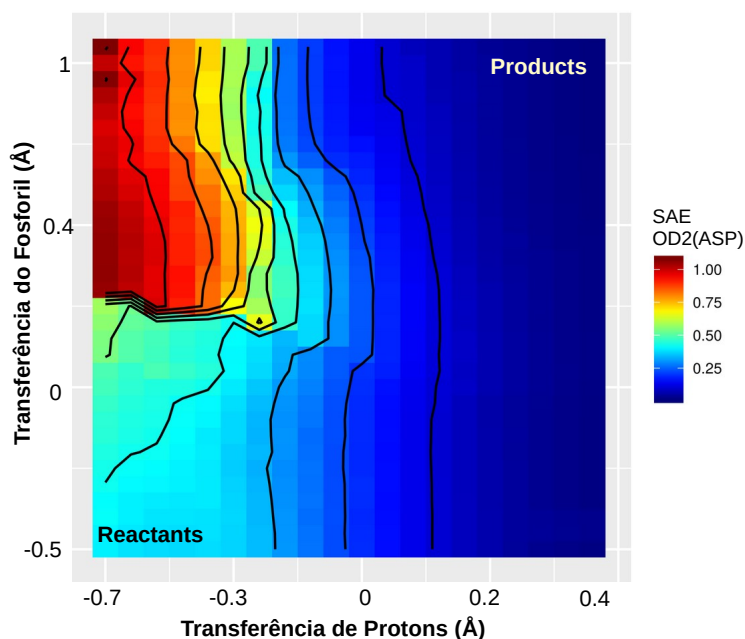


Figura 6.14: Variação dos descritores de susceptibilidade ao ataque eletrofílico, mesmo que nucleofilicidade, em função das coordenadas de reação catalisada pela adenosina quinase obtidos com a estrutura eletrônica calculada com Hamiltoniano PM7. Fonte: Imagem com legendas traduzidas retiradas do trabalho de Grillo e colaboradores.^[40]

6.3.4 Chiquimato Desidrogenase

A análise da superfície de energia potencial resultou em um mecanismo para a catálise da CDH, onde o hidreto é transferido primeiro seguido da transferência do próton, também sendo a primeira simulação teórica desse caminho de reação encontrado na literatura, que geralmente foca em outras etapas do ciclo da chiquimato. Para ver se os descritores corroboram com o mecanismo proposto, é analisada a variação de dureza local para o átomo de hidrogênio que é transferido como um hidreto em função das duas coordenadas de reação analisadas, como mostrado na Figura 6.15:A.

A dureza desse átomo aumenta conforme ele é transferido do NADP para o C3 do substrato, até chegar ao máximo no estado de transição. Na Figura 6.15:B é mostrada a nucleofilicidade do mesmo átomo, mostrando um valor significativamente alto no estado de transição. Na Figura 6.15:C é possível observar a nucleofilicidade do O3 do substrato que abstrai o próton da lisina-69, que também só apresenta valores significativos depois da transferência do hidreto. A partir desses mapas de contorno dos descritores nesses átomos, podemos afirmar que os descritores também indicam que o mecanismo se dá primeiro com a transferência do hidreto e depois o O3(DHK) se ativa para capturar o próton.

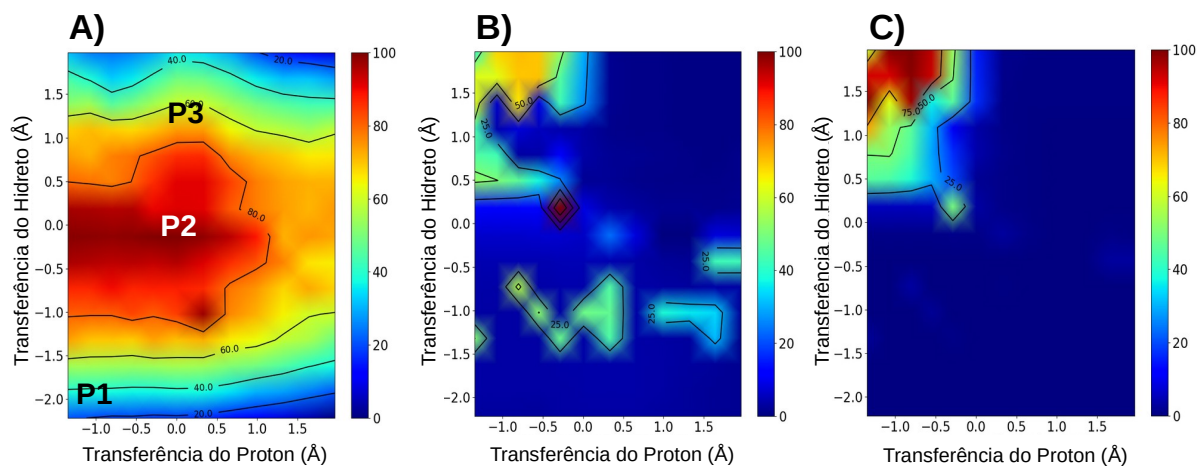


Figura 6.15: Mapas de contorno dos descritores de reatividade para as coordenadas de reatividade da reação catalisada pela chiquimato desidrogenase: A) Dureza local para o hidreto H42; B) Nucleofilicidade do hidreto H42; C) Nucleofilicidade do O3(DHK). Fonte: Imagem traduzida do trabalho de Grillo e colaboradores.^[39]

Na Figura 6.16, a dureza local é mostrada para o sítio ativo da CDH, na geometria inicial e do estado de transição. É possível observar que os átomos com os maiores valores nessas duas geometrias são aqueles que participam diretamente na reação, incluindo o hidreto e o próton transferidos. A principal variação encontrada foi a do incremento na dureza local da treonina-65 no estado de transição, indicando que esse resíduo pode ser importante na estabilização dessa estrutura através de interações duro-duro.

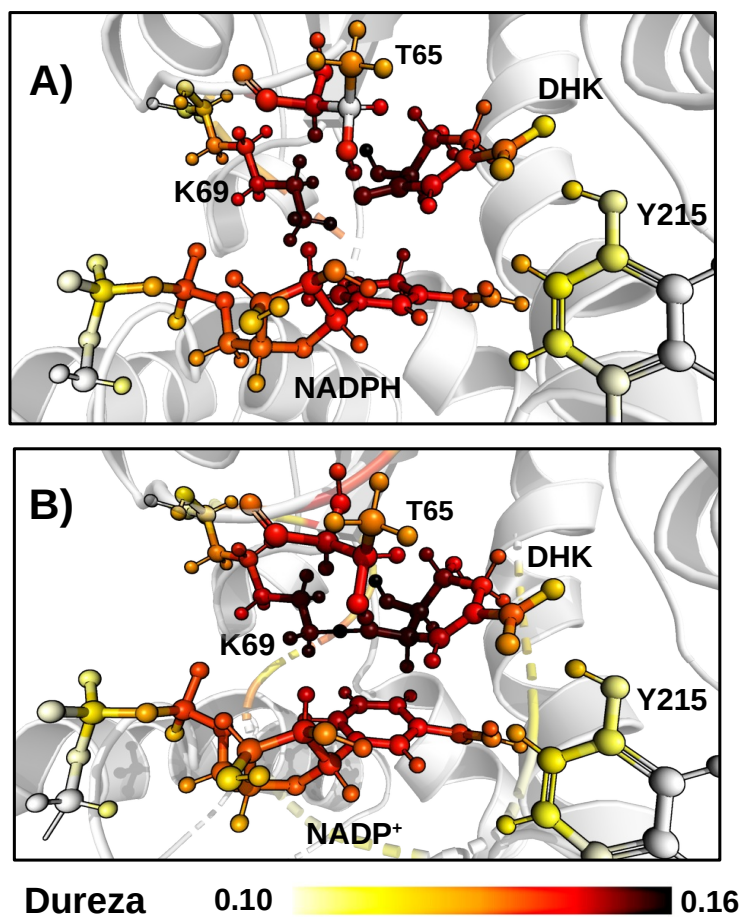


Figura 6.16: Mapas de reatividade de dureza local para a reação catalisada pela chiquímato desidrogenase: A) geometria inicial; B) geometria do estado de transição. Imagem traduzida do trabalho de Grillo e colaboradores.^[39]

Para analisar as propensões de ataques nucleofílicos e eletrofílicos e a transferência de carga, os mapas de reatividade na Figura 6.17 são avaliados. Primeiramente, é importante lembrar que esse é um processo esperado por ser governado em sua maior parte por interações duro-duro, e que esses mapas não necessariamente vão indicar a reatividade em todos os átomos que estão participando da reação. Nas coordenadas tomadas como iniciais, os átomos do anel de nicotamida apresentam os maiores valores de eletrofilicidade mesmo possuindo carga formal ligeiramente negativa.

Geralmente a transferência de um hidreto é atrelado a uma transferência de um par de elétrons em conjunto, o que não é indicado aqui, o anel de nicotamida está propenso a receber carga e não doar. No estado de transição, é possível observar que as cargas parciais de fato diminuem, mas em pequeno valor. Na estrutura do estado de transição, é possível observar que o hidreto foi praticamente todo transferido para a molécula de chiquimato, que

de fato aumenta a sua nucleofilicidade significativamente. Também é interessante notar que o resíduo de aspartato-105 apresenta ao longo de toda reação valores de eletrofilicidade, mesmo tendo os seus oxigênios da carboxila carregados negativamente. Isto é mais um indício que a análise de cargas parciais é limitada no que se refere ao perfil de reatividade de sistemas complexos, como são os enzimáticos.

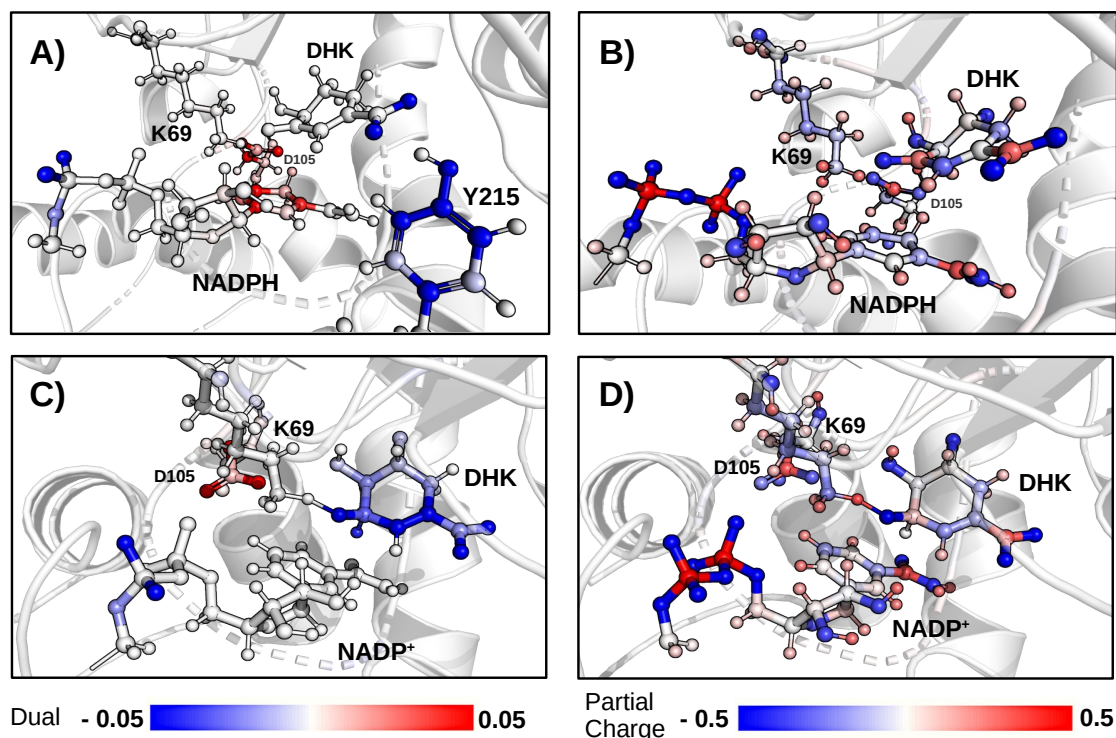


Figura 6.17: Mapas de reatividade de netfilicidade e carga parcial para a reação catalisada pela chiquímato desidrogenase: A) netfilicidade para geometria inicial, mesmo que Dual; B) carga parcial; C) netfilicidade para estado de transição; D) carga parcial para o estado de transição. Imagem traduzida do trabalho de Grillo e colaboradores.^[39]

Na Figura 6.18, são analisados os valores médios e as variações dos descritores nos resíduos do sítio ativo da CDH nas coordenadas de reação. Também é mostrado a média geral para todos dos resíduos do sítio ativo (AVG), para comparação. A treonina apresenta um valor médio de dureza e nucleofilicidade maior que a média geral e com pouca variabilidade. Nos dois resíduos onde há a transferência de átomos, NADPH e DHK, a variabilidade dos valores dos descritores é significativamente alto. O ácido aspártico-105 apresenta valores mínimos de eletrofilicidade, indicando um efeito de aumento da acidez da lisina-69, isto é, ele tende a deslocalizar a densidade eletrônica facilitando a saída do próton na direção contrária.

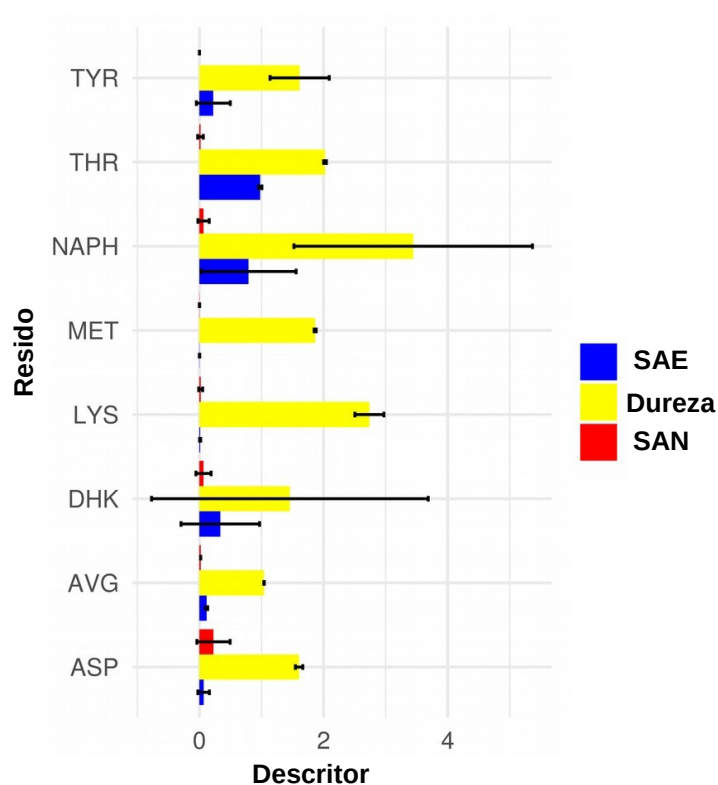


Figura 6.18: Média e desvio padrão (barras pretas) dos descritores de nucleofilicidade (SAE), eletrofilicidade (SAN), e dureza local para todas as geometrias do scan relaxado nas duas coordenadas de reação. AVG = média dos valores para todos os resíduos. Imagem traduzida do trabalho de Grillo e colaboradores.^[39]

6.4 Considerações Finais

Através da análise dos mecanismos simulados de quatro sistemas enzimáticos, foi possível observar a capacidade dos descritores de indicar os principais estruturas dessas trajetórias, apresentando perfis similares a energia potencial, com menor variabilidade ao método de estrutura eletrônica utilizado e adicionando uma dimensão à descrição desses processos. Os estudos de reações enzimáticas geralmente só contam com a observação das distâncias entre os átomos e avaliação de carga parcial, que, por sua vez, se mostra insuficiente, pois esses valores mudam pouco e para poucos átomos no sítio ativo durante a reação. Portanto, os descritores calculados com o PRIMoRDiA se mostraram eficientes no estudo de elucidação de mecanismos, testados naqueles onde já se tem uma caracterização sólida e aplicados em caminhos de reação até então não descritos na literatura.

Contribuições do Software

“Hoje eu sou trovão. Sou claridade no céu, eu sou fogo no chão”

–Lucas Silveira

Como já mostrado em outros capítulos dessa Tese, conforme o programa foi sendo desenvolvido ele foi sendo aplicado em vários trabalhos de modelagem molecular de sistemas biológicos, incluindo dissertações de mestrado, teses de doutorado e publicações em periódicos da área. Assim como a presença do nosso grupo em minicursos ministrados em escolas de modelagem molecular, como a da UNESP-Araraquara e do Laboratório Nacional de Computação Científica (LNCC).

7.1 Aplicações em Dissertações e Teses

O software foi utilizado na dissertação de mestrado de Pereira,^[223] que executou um estudo teórico sobre o mecanismo de ação da subunidade A tóxica da Ricina, uma enzima produzida na planta de mamoneira (*ricinus comunins*) que é extremamente tóxica. Os descritores foram aplicados de forma a gerar mapas de reatividade do sistema complexado com um fragmento que simula a parte do ribossomo clivada pela enzima, sendo útil para caracterizar a reatividade do sítio ativo e apontar o mecanismo mais provável. Além disso, foi investigado o ataque nucleofílico da adenina do alvo no resíduo arginina-180 e a propensão do glutamato-177 de ativar a água catalítica por meio da abstração de um de seus prótons. Também se verificou, através da dureza local, possíveis interações estabilizantes entre os resíduos de tirosina-80 e tirosina-123.

Na tese de doutorado do Rocha,^[224] a assinatura da estrutura eletrônica nos processos

de enovelamento de proteínas foi estudado usando os descritores calculados com o PRIMoRDiA, mostrando que há uma tendência na localização de densidade eletrônica e um padrão mais claro ainda na variação de dureza local nos principais resíduos que compõem o núcleo hidrofóbico desses polipeptídeos. O software permitiu a análise eficiente da trajetória de enovelamento de três polipeptídeos, mostrando como há momentos críticos para o enovelamento, onde a mudança estrutural está associada com uma variação brusca nas propriedades eletrônicas.

Na tese de doutorado de Silva,^[225] o software foi útil na análise da relação da estrutura eletrônica com a presença de nós geométricos em proteínas. Analisando o conjunto de estruturas geradas por amostragem de dinâmica molecular em proteínas homólogas, em relação a função e estrutura terciária, foi detectado que a proteína com nó apresenta propriedades eletrônicas e de reatividade estatisticamente diferentes, podendo ter como fonte desse efeito essa particularidade estrutural.

Na dissertação do Farias,^[226] os descritores foram utilizados em combinação com protocolo de redesenho de sítios ativos, o Rosetta, para distinguir estruturas com potencial catalítico. Já na dissertação da Sadaf Rani,^[227] o PRIMoRDiA foi empregado para a caracterização teórica do sítio ativo do sistema enzimático gluco-6-fosfato desidrogenase, ajudando a confirmar o papel do resíduo catalítico que extrai o próton do gluco-6-fosfato.

7.2 Aplicações em Publicações em Periódicos

Quanto a artigos publicados, em ordem cronológica, no trabalho de Grillo e colaboradores,^[34] os descritores foram calculados para estimar o efeito nas propriedades eletrônicas no sítio ativo da enzima HIV-1 protease devido a ligação com inibidores, que por sua vez já tinham dados experimentais de energia livre de ligação. Na Figura 7.1, apresentamos alguns resultados dessas correlações, mostrando que a variação do descritor de nucleofilicidade conseguiram classificar corretamente a grande maioria dos inibidores (Figura 7.1:A). Ainda, a interação média de cada resíduo de amino-acido no sítio ativo é mostrada na Figura 7.1:C).

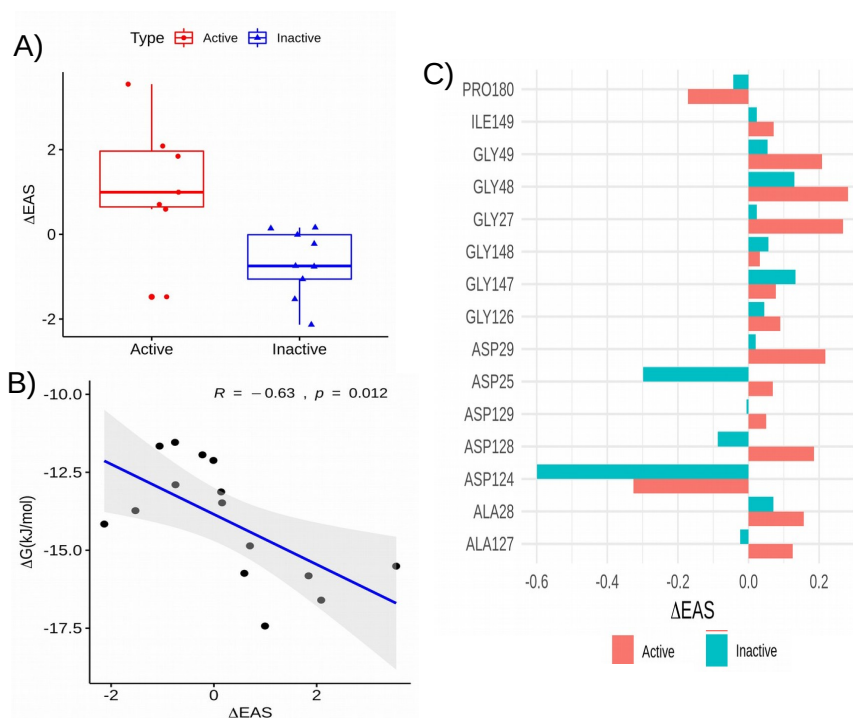


Figura 7.1: Aplicações dos descritores de reatividade para complexos da HIV-1 protease: A) ranqueamento entre os ligantes ativos e inativos usando a variação do descritor de nucleofilicidade; B) Correlação entre a variação do descritor de nucleofilicidade com os valores experimentais de energia livre de ligação; C) Contribuição média de cada resíduo de amino-ácido do sítio ativo para a interação com os inibidores no descritor de nucleofilicidade. Fonte: Imagem adaptadas de figuras retiradas do trabalho de Grillo e colaboradores.^[34]

Como é possível observar, os inibidores inativos causam uma maior variação negativa no descritor de nucleofilicidade nos resíduos de sítio ativo, diminuindo a reatividade dos aspartatos, que normalmente são resíduos que realizam ataques nucleofílicos em complexos enzimáticos. Já os ligantes ativos mostraram um aumento significativo na nucleofilicidade dos resíduos de glicina, o que pode indicar que esses pontos do sítio ativo aumentam a concentração de densidade eletrônica que contribuem para as interações estabilizantes com os inibidores. Para o inibidor mais ativo, ou seja, com a variação de energia livre de ligação mais negativa, é mostrada na Figura 7.2^[34] o descritor de netfilicidade do complexo. Os resíduos dos átomos com carbonos em amarelo na Figura 7.2:A, são dos resíduos alanina-28, aspartato-29, e glicina-48, onde a nucleofilicidade aumentou e está perto do grupo molecular do inibidor que apresenta seus átomos com valores significativos de eletrofilicidade, indicando a possível interação por complementariedade de densidade eletrônica.

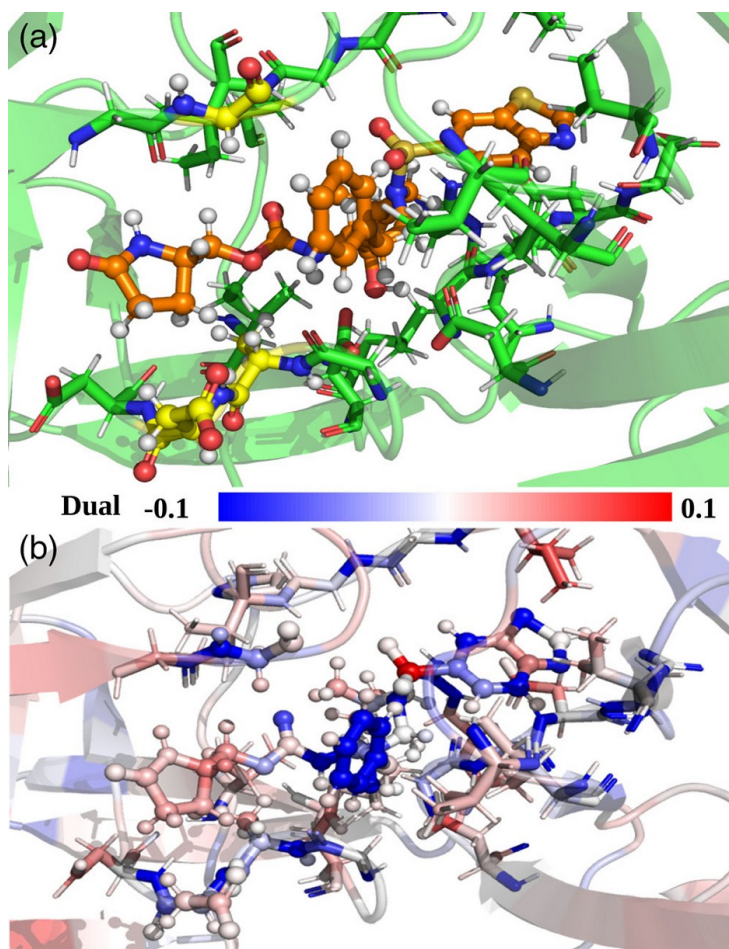


Figura 7.2: Descritor de netfilicidade para o sítio ativo complexado com o inibidor com maior afinidade: A) Átomos do sítio ativo; B) descritor de netfilicidade. Fonte: Imagem retirada do trabalho de Grillo e colaboradores.^[34]

Com o mesmo objetivo de explicar a origem das interações entre possíveis inibidores e enzimas, no estudo de Rocha e colaboradores,^[41] o PRIMoRDiA foi utilizado para calcular os descritores para complexos entre a subunidade tóxica da Ricina (RTA) e seis inibidores com valor de energia livre de ligação estimados experimentalmente. Na Figura 7.3, são mostrados os mapas de reatividade para o descritor de dureza local para complexo de maior poder inibidor, codificado como 19M, para o com menor, codificado como JP3. Nesse trabalho, se chegou a conclusão que os melhores inibidores mostraram as maiores interações duro-duro, principalmente no grupo da pterina, que é o grupo químico mais parecido com o substrato natural.

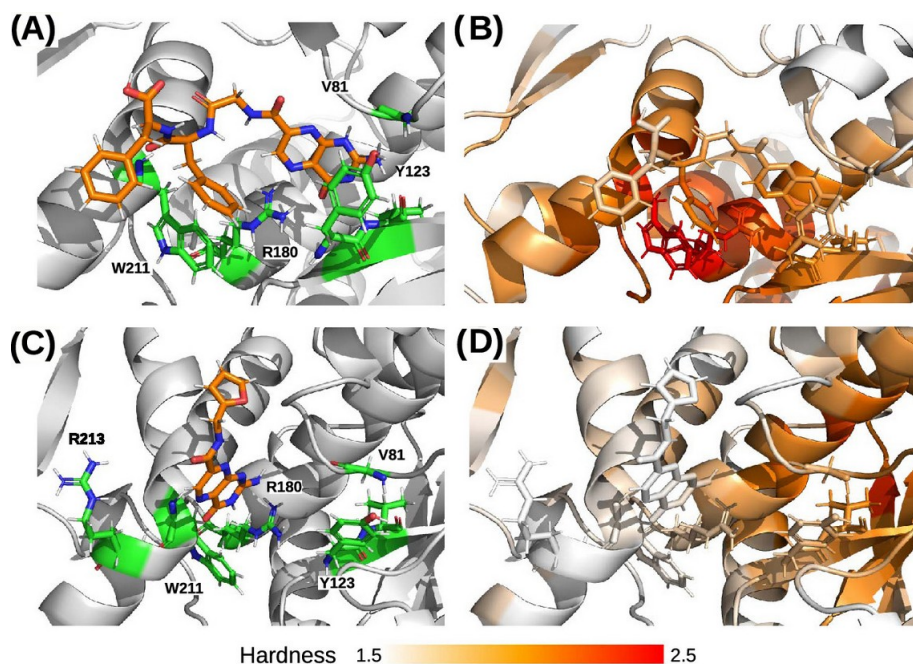


Figura 7.3: Dureza local para os complexos da sub-unidade tóxica da Ricina (RTA): A) Sítio de ligação descrito para o complexo com o 19M (laranja) e resíduos próximos em verde; B) Dureza local para o complexo RTA–19M; C) Sítio de ligação descrito para o complexo com o JP3 (laranja) e resíduos próximos em verde; D) Dureza local para o complexo RTA–19M. Fonte: Figura traduzida do trabalho de Rocha e colaboradores.^[41]

Teve mais um estudo onde o PRIMoRDiA foi utilizado para indicar possíveis pontos de interação com potenciais fármacos sintetizados, mais especificamente derivados de bishidrazona, com a enzima α -glucosidase no trabalho de Kaya e colaboradores.^[42] Os resultados apontam como algumas variações na composição química dessas moléculas podem alterar na reatividade e possível interação com o sítio ativo. No entanto, esses ainda são resultados preliminares.

Além dessas aplicações para explicar interações proteína-ligante, o PRIMoRDiA foi utilizado para analisar trajetórias de dinâmica molecular para criar correlações entre os descritores quânticos com a abertura do sítio ativo das proteases principais dos vírus SARS-CoV e SARS-CoV-2. Os descritores básicos de nefilicidade e dureza local foram utilizados para monitorar a reatividade nos resíduos catalíticos, mostrando que há uma diferença significativa entre as proteases dos dois vírus, como mostrado na Figura 7.4.

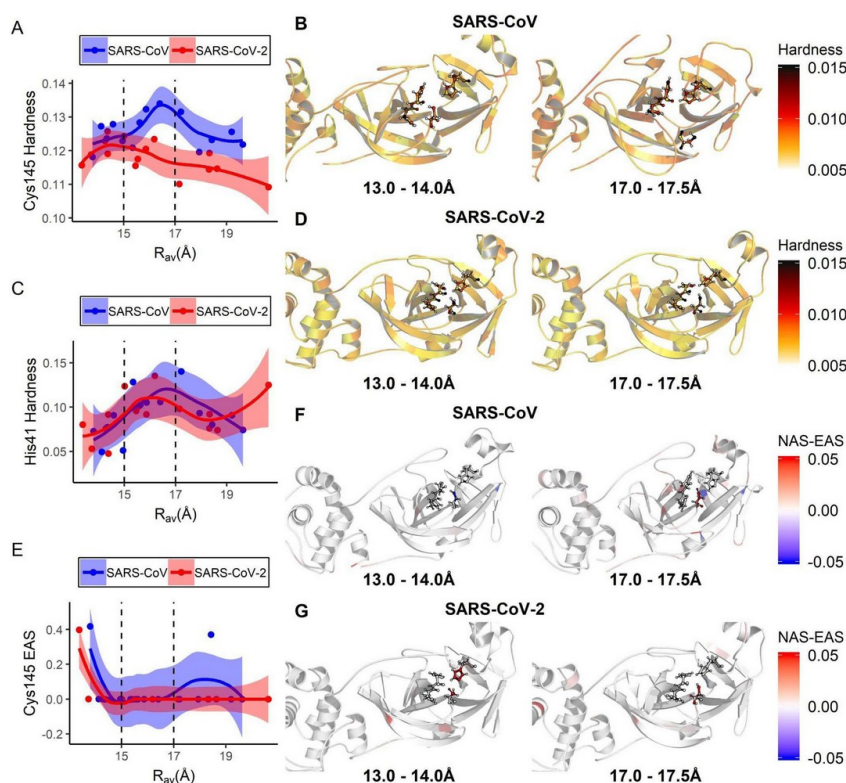


Figura 7.4: Variação dos descritores em relação a variação estrutural das proteases principais (MPro) dos vírus SARS-CoV and SARS-CoV-2: A) Média móvel do valor de dureza local para o resíduo catalítico cisteína-145 em relação ao raio de abertura do sítio ativo ($R_{av}()$); B) Mapas de dureza local para a MPro da SARS-CoV, como mínimo e máximo raios de abertura; C) Média móvel do valor de dureza local para o resíduo catalítico histidina-41 em relação ao $R_{av}()$; D) Mapas de dureza local para a MPro da SARS-CoV-2, como mínimo e máximo raios de abertura; E) Média móvel do valor de nucleofilicidade para o resíduo catalítico cisteína-145 em relação ao $R_{av}()$; F) O mesmo que mostrado na B) mas com o descritor de nucleofilicidade; G) O mesmo que a D) mas com o descritor de nucleofilicidade. Fonte: Figura retirada do trabalho Rocha e colaboradores.^[43]

Dessas análises, foi concluído que a protease principal da primeira cepa do vírus modela a reatividade da cisteína para a transferência de próton, descrito pela dureza local, em uma janela estreita de valores de abertura do sítio ativo e que não condiz com os mesmos valores de abertura observadas para a reatividade por transferência de carga, descrito através das interações mole-mole pelo descritor de nucleofilicidade.

7.3 Minicursos Ministrados

O primeiro minicurso foi na escola de modelagem molecular de sistemas biológicos da Universidade Estadual Paulista (UNESP), que ocorreu inteiramente de forma virtual em janeiro de 2021. A terceira aula do minicurso "Modelando sistemas moleculares e sistemas biológicos com métodos de química quântica" foi inteiramente sobre o PRIMoRDiA, com tutoriais da primeira versão do programa. O minicurso se encontra disponível no canal do YouTube da Escola de Modelagem Molecular da Unesp.^[228]

A versão 1.2 do software foi apresentada na décima edição da Escola de Modelagem Molecular de Sistemas Biológicos (EMMSB), realizada a cada dois anos no Laboratório Nacional de Computação Científica, durante o minicurso 'Métodos Quânticos Semiempírico' inteiramente disponível online no canal do YouTube do EMMSB.^[229] Nesse último curso o software foi apresentado em duas aulas online de duas horas cada, mostrando teoria, uso geral e tutoriais específicos para grandes sistemas e reações enzimáticas.

7.4 Considerações Finais

Neste capítulo foram mostrados exemplos da concretização do último objetivo desse trabalho de Tese, que é disponibilizar a ferramenta criada a ponto de que estudos acadêmicos e científicos possam ser realizados por diversos grupos de pesquisa. Se pode observar o começo da difusão dessa abordagem computacional em casos comuns na modelagem molecular de sistemas biológicos e de uma linguagem visual utilizada para representar as reatividades, principalmente dos descritores de dureza local e de eletrofilicidade/nucleofilicidade local, cumprindo assim a descrição básica dos efeitos duro-duro e mole-mole. Além disso, o nosso grupo de pesquisa tem realizado cursos para a comunidade científica nacional, mostrando o uso da ferramenta e suas possíveis aplicações.

Conclusões

“Per aspera ad astra”

– Provérbio romano

Nesse trabalho foi mostrado o desenvolvimento do software PRIMoRDiA, detalhes de sua implementação, teste dos descritores quânticos modificados para sistemas biológicos, aplicação na caracterização teórica de simulações de reações enzimáticas e as contribuições do software em trabalhos acadêmicos já publicados.

O software lançado, disponível gratuitamente para a comunidade científica, se mostrou eficiente no tratamento de grandes volumes de dados provenientes de cálculos de estrutura eletrônica de grandes sistemas biológicos. Os descritores modificados para lidar com as particularidades da estrutura eletrônica desses sistemas se mostraram como uma alternativa de alto desempenho para a análise de reatividade quando usado métodos quânticos semi-empíricos de baixo custo. Para a análise de sítios ativos de complexos enzimáticos, a aplicação do PRIMoRDiA se mostrou eficiente para estruturas que passaram por protocolos de preparação menos exigentes computacionalmente. Mais especificamente, para a caracterização de caminhos de reações enzimáticas, esses métodos se comparam com varreduras de superfície de energia potencial.

Por fim, o objetivo de criar uma ferramenta que disponibilize todas essas funcionalidades para a comunidade científica vem sendo realizado, com vários exemplos de usos do software e publicações por diversos grupos de pesquisa. Como perceptivas, é esperado que o PRIMoRDiA ganhe mais amplitude de aplicações na literatura e que o software continue sendo aprimorado, dando suporte a cada vez mais tipos de pacotes de química computacional e que sejam implementados novos descritores.

Referências Bibliográficas

- [1] DE JONG, W. A.; BYLASKA, E.; GOVIND, N.; JANSSEN, C. L.; KOWALSKI, K.; MÜLLER, T.; NIELSEN, I. M.; VAN DAM, H. J.; VERYAZOV, V.; LINDH, R. Utilizing high performance computing for chemistry: parallel computational chemistry. *Physical Chemistry Chemical Physics*, v. 12, n. 26, p. 6896–6920, 2010.
- [2] FENG, X. Chemical and biochemical basis of cell-bone matrix interaction in health and disease. *Current chemical biology*, v. 3, n. 2, p. 189–196, 2009.
- [3] HOLLIDAY, G. L.; ALMONACID, D. E.; MITCHELL, J. B.; THORNTON, J. M. The chemistry of protein catalysis. *Journal of molecular biology*, v. 372, n. 5, p. 1261–1277, 2007.
- [4] BRUNK, E.; ROTH LISBERGER, U. Mixed quantum mechanical/molecular mechanical molecular dynamics simulations of biological systems in ground and electronically excited states. *Chemical reviews*, v. 115, n. 12, p. 6217–6263, 2015.
- [5] KNOWLES, J. R.; ALBERY, W. J. Perfection in enzyme catalysis: the energetics of triosephosphate isomerase. *Accounts of chemical research*, v. 10, n. 4, p. 105–111, 1977.
- [6] HOLLIDAY, G. L.; ALMONACID, D. E.; BARTLETT, G. J.; O'BOYLE, N. M.; TORRANCE, J. W.; MURRAY-RUST, P.; MITCHELL, J. B.; THORNTON, J. M. Macie (mechanism, annotation and classification in enzymes): novel tools for searching catalytic mechanisms. *Nucleic acids research*, v. 35, n. suppl_1, p. D515–D520, 2007.
- [7] WEBER, I. T.; AGNISWAMY, J.; FU, G.; SHEN, C.-H.; HARRISON, R. W. Reaction intermediates discovered in crystal structures of enzymes. *Advances in protein chemistry and structural biology*, v. 87, p. 57–86, 2012.

- [8] SCHÖNBRUNN, E.; ESCHENBURG, S.; SHUTTLEWORTH, W. A.; SCHLOSS, J. V.; AMRHEIN, N.; EVANS, J. N.; KABSCH, W. Interaction of the herbicide glyphosate with its target enzyme 5-enolpyruvylshikimate 3-phosphate synthase in atomic detail. *Proceedings of the National Academy of Sciences*, v. 98, n. 4, p. 1376–1380, 2001.
- [9] STEINRÜCKEN, H.; AMRHEIN, N. The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimate acid-3-phosphate synthase. *Biochemical and biophysical research communications*, v. 94, n. 4, p. 1207–1212, 1980.
- [10] FENG, Y.; WANG, Y.; CHU, H.; FAN, Y.; CAO, X.; LIU, Y.; LI, G.; XUE, S. Stereoselective catalysis controlled by a native leucine or variant isoleucine wing-gatekeeper in 2-haloacid dehalogenase. *FEBS letters*, v. 593, n. 3, p. 308–318, 2019.
- [11] KARPLUS, M.; KURIYAN, J. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences*, v. 102, n. 19, p. 6679–6685, 2005.
- [12] SWIDEREK, K.; TUÑÓN, I.; MARTÍ, S.; MOLINER, V. Protein conformational landscapes and catalysis. influence of active site conformations in the reaction catalyzed by l-lactate dehydrogenase. *ACS catalysis*, v. 5, n. 2, p. 1172–1185, 2015.
- [13] AGARWAL, P. K.; GEIST, A.; GORIN, A. Protein dynamics and enzymatic catalysis: Investigating the peptidyl- prolyl cis- trans isomerization activity of cyclophilin a. *Biochemistry*, v. 43, n. 33, p. 10605–10618, 2004.
- [14] LIU, S.; RONG, C.; LU, T. Electronic forces as descriptors of nucleophilic and electrophilic regioselectivity and stereoselectivity. *Physical Chemistry Chemical Physics*, v. 19, n. 2, p. 1496–1503, 2017.
- [15] EHRLICH, S.; GÖLLER, A. H.; GRIMME, S. Towards full quantum-mechanics-based protein–ligand binding affinities. *ChemPhysChem*, v. 18, n. 8, p. 898–905, 2017.
- [16] WARSHEL, A.; LEVITT, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of molecular biology*, v. 103, n. 2, p. 227–249, 1976.
- [17] FIELD, M. J.; BASH, P. A.; KARPLUS, M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, v. 11, n. 6, p. 700–733, 1990.

- [18] EURENIUS, K. P.; CHATFIELD, D. C.; BROOKS, B. R.; HODOSCEK, M. Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. i. theoretical considerations. *International journal of quantum chemistry*, v. 60, n. 6, p. 1189–1200, 1996.
- [19] DIRAC, P. A. M. Quantum mechanics of many-electron systems. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, v. 123, n. 792, p. 714–733, 1929.
- [20] DOMINGO, L. R.; PÉREZ, P. Global and local reactivity indices for electrophilic/nucleophilic free radicals. *Organic & Biomolecular Chemistry*, v. 11, n. 26, p. 4350–4358, 2013.
- [21] DOMINGO, L. R.; RÍOS-GUTIÉRREZ, M.; PÉREZ, P. Applications of the conceptual density functional theory indices to organic chemistry reactivity. *Molecules*, v. 21, n. 6, p. 748, 2016.
- [22] KOHN, W.; BECKE, A. D.; PARR, R. G. Density functional theory of electronic structure. *The Journal of Physical Chemistry*, v. 100, n. 31, p. 12974–12980, 1996.
- [23] GEERLINGS, P.; DE PROFT, F.; LANGENAEKER, W. Conceptual density functional theory. *Chemical reviews*, v. 103, n. 5, p. 1793–1874, 2003.
- [24] PEARSON, R. G. Recent advances in the concept of hard and soft acids and bases. *Journal of Chemical Education*, v. 64, n. 7, p. 561, 1987.
- [25] PARR, R. G.; DONNELLY, R. A.; LEVY, M.; PALKE, W. E. Electronegativity: the density functional viewpoint. *The Journal of Chemical Physics*, v. 68, n. 8, p. 3801–3807, 1978.
- [26] FUKUI, K. The role of frontier orbitals in chemical reactions. In: *Frontier Orbitals and Reaction Paths: Selected Papers of Kenichi Fukui*. World Scientific, 1997. p. 150–170.
- [27] PARR, R. G.; YANG, W. Density functional approach to the frontier-electron theory of chemical reactivity. *Journal of the American Chemical Society*, Washington, v. 106, n. 14, p. 4049–4050, 1984.
- [28] GEERLINGS, P.; DE PROFT, F. Hsab principle: Applications of its global and local forms in organic chemistry. *International Journal of Quantum Chemistry*, v. 80, n. 2, p. 227–235, 2000.

- [29] LONSDALE, R.; HARVEY, J. N.; MULHOLLAND, A. J. A practical guide to modelling enzyme-catalysed reactions. *Chemical Society Reviews*, v. 41, n. 8, p. 3025–3038, 2012.
- [30] RAMOS, M. J.; FERNANDES, P. A. Computational enzymatic catalysis. *Accounts of chemical research*, v. 41, n. 6, p. 689–698, 2008.
- [31] GARCIA-VILOCA, M.; GAO, J.; KARPLUS, M.; TRUHLAR, D. G. How enzymes work: analysis by modern rate theory and computer simulations. *Science*, Washington, v. 303, n. 5655, p. 186–195, 2004.
- [32] ROOS, G.; GEERLINGS, P.; MESSENS, J. Enzymatic catalysis: the emerging role of conceptual density functional theory. *The Journal of Physical Chemistry B*, v. 113, n. 41, p. 13465–13475, 2009.
- [33] FUKUSHIMA, K.; WADA, M.; SAKURAI, M. An insight into the general relationship between the three dimensional structures of enzymes and their electronic wave functions: Implication for the prediction of functional sites of enzymes. *Proteins: Structure, Function, and Bioinformatics*, v. 71, n. 4, p. 1940–1954, 2008.
- [34] GRILLO, I. B.; URQUIZA-CARVALHO, G. A.; CHAVES, E. J. F.; ROCHA, G. B. Semi-empirical methods do fukui functions: Unlocking a modeling framework for biosystems. *Journal of Computational Chemistry*, v. 41, n. 9, p. 862–873, 2020.
- [35] KHANDOGIN, J.; MUSIER-FORSYTH, K.; YORK, D. M. Insights into the regioselectivity and rna-binding affinity of hiv-1 nucleocapsid protein from linear-scaling quantum methods. *Journal of molecular biology*, v. 330, n. 5, p. 993–1004, 2003.
- [36] WALTERS, W. P. Code sharing in the open science era. *Journal of Chemical Information and Modeling*, v. 60, n. 10, p. 4417–4420, 2020.
- [37] GRILLO, I. B. et al. Estudo da computação de descritores de reatividade de um sistema polipeptídico usando hamiltonianos semi-empírico. 2019.
- [38] GRILLO, I. B.; URQUIZA-CARVALHO, G. A.; ROCHA, G. B. Primordia: A software to explore reactivity and electronic structure in large biomolecules. *Journal of Chemical Information and Modeling*, v. 60, n. 12, p. 5885–5890, 2020.
- [39] GRILLO, I. B.; BACHEGA, J. F. R.; TIMMERS, L. F. S.; CACERES, R. A.; DE SOUZA, O. N.; FIELD, M. J.; ROCHA, G. B. Theoretical characterization of the shikimate

- 5-dehydrogenase reaction from mycobacterium tuberculosis by hybrid qc/mm simulations and quantum chemical descriptors. *Journal of Molecular Modeling*, v. 26, n. 11, p. 1–12, 2020.
- [40] GRILLO, I. B.; URQUIZA-CARVALHO, G. A.; BACHEGA, J. F. R.; ROCHA, G. B. Elucidating enzymatic catalysis using fast quantum chemical descriptors. *Journal of Chemical Information and Modeling*, v. 60, n. 2, p. 578–591, 2020.
- [41] ROCHA-SANTOS, A.; CHAVES, E. J. F.; GRILLO, I. B.; DE FREITAS, A. S.; ARAÚJO, D. A. M.; ROCHA, G. B. Thermochemical and quantum descriptor calculations for gaining insight into ricin toxin a (rta) inhibitors. *ACS omega*, v. 6, n. 13, p. 8764–8777, 2021.
- [42] KAYA, Y.; ERÇAĞ, A.; SERDAROĞLU, G.; KAYA, S.; GRILLO, I. B.; ROCHA, G. B. Synthesis, spectroscopic characterization, dft calculations, and molecular docking studies of new unsymmetric bishydrazone derivatives. *Journal of Molecular Structure*, v. 1244, p. 131224, 2021.
- [43] ROCHA, R. E.; CHAVES, E. J.; FISCHER, P. H.; COSTA, L. S.; GRILLO, I. B.; DA CRUZ, L. E.; GUEDES, F. C.; DA SILVEIRA, C. H.; SCOTTI, M. T.; CAMARGO, A. D. et al. A higher flexibility at the sars-cov-2 main protease active site compared to sars-cov and its potentialities for new inhibitor virtual screening targeting multi-conformers. *Journal of Biomolecular Structure and Dynamics*, p. 1–21, 2021.
- [44] RAJENDRAN, A.; ENDO, M.; SUGIYAMA, H. Structural and functional analysis of proteins by high-speed atomic force microscopy. *Advances in protein chemistry and structural biology*, v. 87, p. 5–55, 2012.
- [45] DUNKER, A. K.; BROWN, C. J.; LAWSON, J. D.; IAKOUCHEVA, L. M.; OBRADOVIĆ, Z. Intrinsic disorder and protein function. *Biochemistry*, v. 41, n. 21, p. 6573–6582, 2002.
- [46] CREIGHTON, T. E. Protein folding. *Biochemical journal*, v. 270, n. 1, p. 1, 1990.
- [47] HERMANN, J. C.; HENSEN, C.; RIDDER, L.; MULHOLLAND, A. J.; HÖLTJE, H.-D. Mechanisms of antibiotic resistance: Qm/mm modeling of the acylation reaction of a class a β -lactamase with benzylpenicillin. *Journal of the American Chemical Society*, Washington, v. 127, n. 12, p. 4454–4465, 2005.

- [48] SCHRAMM, V. L. Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annual review of biochemistry*, v. 80, p. 703–732, 2011.
- [49] PHILLIPS, D. C. The hen egg-white lysozyme molecule. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, p. 483–495, 1967.
- [50] ANDERSON, K. S.; JOHNSON, K. A. Kinetic and structural analysis of enzyme intermediates: lessons from epsp synthase. *Chemical Reviews*, v. 90, n. 7, p. 1131–1149, 1990.
- [51] ARMENTA, S.; BLANCO, M. Ion mobility spectrometry: A valuable tool for kinetic studies in enzymology. *Analytica chimica acta*, v. 685, n. 1, p. 1–8, 2011.
- [52] PERACCHI, A. Enzyme catalysis: removing chemically ‘essential’ residues by site-directed mutagenesis. *Trends in biochemical sciences*, v. 26, n. 8, p. 497–503, 2001.
- [53] VAN DER KAMP, M. W.; MULHOLLAND, A. J. Computational enzymology: insight into biological catalysts from modelling. *Natural product reports*, v. 25, n. 6, p. 1001–1014, 2008.
- [54] BOWMAN, A. L.; GRANT, I. M.; MULHOLLAND, A. J. Qm/mm simulations predict a covalent intermediate in the hen egg white lysozyme reaction with its natural substrate. *Chemical communications*, , n. 37, p. 4425–4427, 2008.
- [55] KLINMAN, J. P. An integrated model for enzyme catalysis emerges from studies of hydrogen tunneling. *Chemical physics letters*, v. 471, n. 4-6, p. 179–193, 2009.
- [56] MONARD, G.; MERZ, K. M. Combined quantum mechanical/molecular mechanical methodologies applied to biomolecular systems. *Accounts of chemical research*, v. 32, n. 10, p. 904–911, 1999.
- [57] SCALVINI, L.; GHIDINI, A.; LODOLA, A.; CALLEGARI, D.; RIVARA, S.; PIOMELLI, D.; MOR, M. N-acyl ethanolamine acid amidase (naaa): Mechanism of palmitoyl ethanolamide hydrolysis revealed by mechanistic simulations. *ACS Catalysis*, v. 10, n. 20, p. 11797–11813, 2020.
- [58] GORDON, M. S.; FEDOROV, D. G.; PRUITT, S. R.; SLIPCHENKO, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chemical reviews*, v. 112, n. 1, p. 632–672, 2012.

- [59] HUANG, M.; GIESE, T. J.; YORK, D. M. Nucleic acid reactivity: Challenges for next-generation semiempirical quantum models. *Journal of computational chemistry*, v. 36, n. 18, p. 1370–1389, 2015.
- [60] VOITYYUK, A. A. Assessment of semiempirical methods for the computation of charge transfer in dna π -stacks. *Chemical physics letters*, v. 427, n. 1-3, p. 177–180, 2006.
- [61] MLYNSKY, V.; BANAS, P.; SPONER, J.; VAN DER KAMP, M. W.; MULHOLLAND, A. J.; OTYEPKA, M. Comparison of ab initio, dft, and semiempirical qm/mm approaches for description of catalytic mechanism of hairpin ribozyme. *Journal of Chemical Theory and Computation*, v. 10, n. 4, p. 1608–1622, 2014.
- [62] PINHEIRO, P. D. S. M.; RODRIGUES, D. A.; SANT'ANNA, C. M. R.; FRAGA, C. A. M. Modeling zinc-oxygen coordination in histone deacetylase: a comparison of semiempirical methods performance. *International Journal of Quantum Chemistry*, v. 118, n. 21, p. e25720, 2018.
- [63] THIRIOT, E.; MONARD, G. Combining a genetic algorithm with a linear scaling semiempirical method for protein–ligand docking. *Journal of Molecular Structure: THEOCHEM*, v. 898, n. 1-3, p. 31–41, 2009.
- [64] CHAVES, E. J. F.; GOMES DA CRUZ, L. E.; PADILHA, I. Q. M.; SILVEIRA, C. H.; ARAUJO, D. A. M.; ROCHA, G. B. Discovery of rta ricin subunit inhibitors: a computational study using pm7 quantum chemical method and steered molecular dynamics. *Journal of Biomolecular Structure and Dynamics*, v. 40, n. 12, p. 5427–5445, 2022.
- [65] CHRISTENSEN, A. S.; KUBAR, T.; CUI, Q.; ELSTNER, M. Semiempirical quantum mechanical methods for noncovalent interactions for chemical and biochemical applications. *Chemical Reviews*, v. 116, n. 9, p. 5301–5337, 2016.
- [66] BARNETT, C. B.; NAIDOO, K. J. Ring puckering: a metric for evaluating the accuracy of am1, pm3, pm3carb-1, and scc-dftb carbohydrate qm/mm simulations. *The Journal of Physical Chemistry B*, v. 114, n. 51, p. 17142–17154, 2010.
- [67] URQUIZA-CARVALHO, G. A.; FRAGOSO, W. D.; ROCHA, G. B. Assessment of semiempirical enthalpy of formation in solution as an effective energy function to discriminate native-like structures in protein decoy sets. *Journal of computational chemistry*, v. 37, n. 21, p. 1962–1972, 2016.

- [68] SENN, H. M.; THIEL, W. Qm/mm methods for biomolecular systems. *Angewandte Chemie International Edition*, v. 48, n. 7, p. 1198–1229, 2009.
- [69] KARPLUS, M.; LEVITT, M.; WARSHEL, A. The nobel prize in chemistry 2013. *Nobel Media AB 2014*, 2013.
- [70] CLAEYSSSENS, F.; RANAGHAN, K. E.; MANBY, F. R.; HARVEY, J. N.; MULHOLLAND, A. J. Multiple high-level qm/mm reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization. *Chemical communications*, , n. 40, p. 5068–5070, 2005.
- [71] ZHANG, X.; ZHANG, X.; BRUICE, T. C. A definitive mechanism for chorismate mutase. *Biochemistry*, v. 44, n. 31, p. 10443–10448, 2005.
- [72] CLAEYSSSENS, F.; RANAGHAN, K. E.; MANBY, F. R.; HARVEY, J. N.; MULHOLLAND, A. J. Multiple high-level qm/mm reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization. *Chemical communications*, , n. 40, p. 5068–5070, 2005.
- [73] HUR, S.; BRUICE, T. C. The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proceedings of the National Academy of Sciences*, v. 100, n. 21, p. 12015–12020, 2003.
- [74] ALTUN, A.; SHAIK, S.; THIEL, W. Systematic qm/mm investigation of factors that affect the cytochrome p450-catalyzed hydrogen abstraction of camphor. *Journal of computational chemistry*, v. 27, n. 12, p. 1324–1337, 2006.
- [75] TRESADERN, G.; WANG, H.; FAULDER, P. F.; BURTON, N. A.; HILLIER, I. H. Extreme tunnelling in methylamine dehydrogenase revealed by hybrid qm/mm calculations: potential energy surface profile for methylamine and ethanolamine substrates and kinetic isotope effect values. *Molecular Physics*, v. 101, n. 17, p. 2775–2784, 2003.
- [76] MASGRAU, L.; ROUJEINIKOVA, A.; JOHANNISSEN, L. O.; HOTH, P.; BASRAN, J.; RANAGHAN, K. E.; MULHOLLAND, A. J.; SUTCLIFFE, M. J.; SCRUTTON, N. S.; LEYS, D. Atomic description of an enzyme reaction dominated by proton tunneling. *Science*, Washington, v. 312, n. 5771, p. 237–241, 2006.
- [77] PU, J.; GAO, J.; TRUHLAR, D. G. Multidimensional tunneling, recrossing, and the

- transmission coefficient for enzymatic reactions. *Chemical Reviews*, v. 106, n. 8, p. 3140–3169, 2006.
- [78] EHRENREICH, H.; COHEN, M. H. Self-consistent field approach to the many-electron problem. *Physical Review*, New York, v. 115, n. 4, p. 786, 1959.
- [79] GORDON, M. S. *Fragmentation: Toward accurate calculations on complex molecular systems*. John Wiley & Sons, 2017.
- [80] OHNO, K.; KAMIYA, N.; ASAKAWA, N.; INOUE, Y.; SAKURAI, M. Effects of hydration on the electronic structure of an enzyme: implications for the catalytic function. *Journal of the American Chemical Society*, Washington, v. 123, n. 33, p. 8161–8162, 2001.
- [81] DEWAR, M. J.; ZOEBISCH, E. G.; HEALY, E. F.; STEWART, J. J. Development and use of quantum mechanical molecular models. 76. am1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, Washington, v. 107, n. 13, p. 3902–3909, 1985.
- [82] STEWART, J. J. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations. *International Journal of Quantum Chemistry*, v. 58, n. 2, p. 133–146, 1996.
- [83] KLAMT, A. The cosmo and cosmo-rs solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, v. 1, n. 5, p. 699–709, 2011.
- [84] COHEN, M. H.; WASSERMAN, A. On the foundations of chemical reactivity theory. *The Journal of Physical Chemistry A*, v. 111, n. 11, p. 2229–2242, 2007.
- [85] PARR, R. G.; PEARSON, R. G. Absolute hardness: companion parameter to absolute electronegativity. *Journal of the American chemical society*, v. 105, n. 26, p. 7512–7516, 1983.
- [86] MULLIKEN, R. S. A new electroaffinity scale; together with data on valence states and on valence ionization potentials and electron affinities. *The Journal of Chemical Physics*, v. 2, n. 11, p. 782–793, 1934.
- [87] GORDY, W.; THOMAS, W. O. Electronegativities of the elements. *The Journal of Chemical Physics*, v. 24, n. 2, p. 439–444, 1956.

- [88] PEARSON, R. G. Hard and soft acids and bases. *Journal of the American Chemical Society*, v. 85, n. 22, p. 3533–3539, 1963.
- [89] FUKUI, K.; YONEZAWA, T.; NAGATA, C.; SHINGU, H. Molecular orbital theory of orientation in aromatic, heteroaromatic, and other conjugated molecules. *The Journal of Chemical Physics*, v. 22, n. 8, p. 1433–1442, 1954.
- [90] FUKUI, K.; YONEZAWA, T.; NAGATA, C. Mo-theoretical approach to the mechanism of charge transfer in the process of aromatic substitutions. In: *Frontier Orbitals And Reaction Paths: Selected Papers of Kenichi Fukui*. World Scientific, 1997. p. 54–66.
- [91] FUKUI, K.; FUJIMOTO, H. An mo-theoretical interpretation of the nature of chemical reactions.: I. partitioning analysis of the interaction energy. In: *Frontier Orbitals and Reaction Paths: Selected Papers of Kenichi Fukui*. World Scientific, 1997. p. 257–265.
- [92] KLOPMAN, G. Chemical reactivity and the concept of charge-and frontier-controlled reactions. *Journal of the American Chemical Society*, Washington, v. 90, n. 2, p. 223–234, 1968.
- [93] CHERMETTE, H. Chemical reactivity indexes in density functional theory. *Journal of Computational Chemistry*, v. 20, p. 129–154, 1999.
- [94] CÁRDENAS, C.; AYERS, P. W.; CEDILLO, A. Reactivity indicators for degenerate states in the density-functional theoretic chemical reactivity theory. *The Journal of chemical physics*, v. 134, n. 17, p. 174103, 2011.
- [95] MARCH, N.; PUCCI, R. Chemical potential related to total energy in isoelectronic sequences of positive ions. *The Journal of Chemical Physics*, v. 78, n. 5, p. 2480–2484, 1983.
- [96] MANNE, R.; ÅBERG, T. Koopmans' theorem for inner-shell ionization. *Chemical Physics Letters*, Amsterdam, v. 7, n. 2, p. 282–284, 1970.
- [97] PEARSON, R. G. Absolute electronegativity and hardness: applications to organic chemistry. *The Journal of Organic Chemistry*, v. 54, n. 6, p. 1423–1430, 1989.
- [98] YANG, W.; MORTIER, W. J. The use of global and local molecular parameters for the analysis of the gas-phase basicity of amines. *Journal of the American Chemical Society*, Washington, v. 108, n. 19, p. 5708–5711, 1986.

- [99] TACHIBANA, A.; KAWAUCHI, S.; NAKAMURA, K.; INABA, H. Density functional theory of chemical reactivity indices in some ion—molecule reaction systems. *International journal of quantum chemistry*, v. 57, n. 4, p. 673–682, 1996.
- [100] XIA, Y.; YIN, D.; RONG, C.; XU, Q.; YIN, D.; LIU, S. Impact of lewis acids on diels-alder reaction reactivity: A conceptual density functional theory study. *The Journal of Physical Chemistry A*, v. 112, n. 40, p. 9970–9977, 2008.
- [101] DOMINGO, L. R.; AURELL, M. J.; PÉREZ, P.; CONTRERAS, R. Quantitative characterization of the local electrophilicity of organic molecules. understanding the regioselectivity on diels- alder reactions. *The Journal of Physical Chemistry A*, v. 106, n. 29, p. 6871–6875, 2002.
- [102] DOMINGO, L. R.; SÁEZ, J. A. Understanding the mechanism of polar diels–alder reactions. *Organic & biomolecular chemistry*, v. 7, n. 17, p. 3576–3583, 2009.
- [103] MORELL, C.; AYERS, P. W.; GRAND, A.; GUTIÉRREZ-OLIVA, S.; TORO-LABBE, A. Rationalization of diels–alder reactions through the use of the dual reactivity descriptor $\delta f(r)$. *Physical Chemistry Chemical Physics*, v. 10, n. 48, p. 7239–7246, 2008.
- [104] DESHLAHRA, P.; IGLESIA, E. Reactivity descriptors in acid catalysis: acid strength, proton affinity and host–guest interactions. *Chemical Communications*, v. 56, n. 54, p. 7371–7398, 2020.
- [105] KORCHOWIEC, J.; UCHIMARU, T. The charge transfer Fukui function: Extension of the finite-difference approach to reactive systems. *Journal of Physical Chemistry A*, v. 102, n. 49, p. 10167–10172, 1998.
- [106] SLATER, J. C. A simplification of the Hartree-Fock method. *Physical Review*, New York, v. 81, n. 3, p. 385–390, 1951.
- [107] YANG, W.; PARR, R. G.; PUCCI, R. Electron density, kohn–sham frontier orbitals, and fukui functions. *The Journal of Chemical Physics*, v. 81, n. 6, p. 2862–2863, 1984.
- [108] FUKUI, K.; FUJIMOTO, H. An MO-theoretical Interpretation of the Nature of Chemical Reactions. I. Partitioning Analysis of the Nature of Chemical of the Interaction Energy. *Bull. Chem. Soc. Jap.*, v. 41, n. 1964, p. 1989–1997, 1968.

- [109] HOCKETT, P.; BISGAARD, C. Z.; CLARKIN, O. J.; STOLOW, A. Time-resolved imaging of purely valence-electron dynamics during a chemical reaction. *Nature Physics*, v. 7, n. 8, p. 612–615, 2011.
- [110] MARTÍNEZ-ARAYA, J. I. Why is the dual descriptor a more accurate local reactivity descriptor than fukui functions? *Journal of Mathematical Chemistry*, v. 53, n. 2, p. 451–465, 2015.
- [111] NALEWAJSKI, R. F.; PARR, R. G. Legendre transforms and maxwell relations in density functional theory. *The Journal of Chemical Physics*, v. 77, n. 1, p. 399–407, 1982.
- [112] MORELL, C.; GRAND, A.; TORO-LABBE, A. New dual descriptor for chemical reactivity. *The Journal of Physical Chemistry A*, v. 109, n. 1, p. 205–212, 2005.
- [113] BULTINCK, P.; FIAS, S.; VAN ALSENOY, C.; AYERS, P. W.; CARBÓ-DORCA, R. Critical thoughts on computing atom condensed fukui functions. *The Journal of chemical physics*, v. 127, n. 3, p. 034102, 2007.
- [114] JERBI, J.; SPRINGBORG, M. Reactivity descriptors for dna bases and the methylation of cytosine. *International Journal of Quantum Chemistry*, v. 118, n. 11, p. e25538, 2018.
- [115] LEE, C.; YANG, W.; PARR, R. G. Local softness and chemical reactivity in the molecules co, scn- and h2co. *Journal of Molecular Structure: Theochem*, v. 163, p. 305–313, 1988.
- [116] MELIN, J.; APARICIO, F.; SUBRAMANIAN, V.; GALVÁN, M.; CHATTARAJ, P. K. Is the fukui function a right descriptor of hard- hard interactions? *The Journal of Physical Chemistry A*, v. 108, n. 13, p. 2487–2491, 2004.
- [117] PARTHASARATHI, R.; PADMANABHAN, J.; ELANGO, M.; SUBRAMANIAN, V.; CHATTARAJ, P. Intermolecular reactivity through the generalized philicity concept. *Chemical physics letters*, v. 394, n. 4-6, p. 225–230, 2004.
- [118] BERKOWITZ, M.; PARR, R. G. Molecular hardness and softness, local hardness and softness, hardness and softness kernels, and relations among these quantities. *The Journal of chemical physics*, v. 88, n. 4, p. 2554–2557, 1988.
- [119] ROY, R.; KRISHNAMURTI, S.; GEERLINGS, P.; PAL, S. Local softness and hardness based reactivity descriptors for predicting intra-and intermolecular reactivity

- sequences: carbonyl compounds. *The Journal of Physical Chemistry A*, v. 102, n. 21, p. 3746–3755, 1998.
- [120] CHATTARAJ, P. K.; GIRI, S. Electrophilicity index within a conceptual dft framework. *Annual Reports Section "C" (Physical Chemistry)*, v. 105, p. 13–39, 2009.
- [121] LANGENAEKER, W.; DE DECKER, M.; GEERLINGS, P.; RAEYMAEKERS, P. Quantum-chemical study of the fukui function as a reactivity index: probing the acidity of bridging hydroxyls in zeolite-type model systems. *Journal of Molecular Structure: THEOCHEM*, v. 207, n. 1-2, p. 115–130, 1990.
- [122] PRATIHAR, S.; ROY, S. Nucleophilicity and site selectivity of commonly used arenes and heteroarenes. *The Journal of organic chemistry*, v. 75, n. 15, p. 4957–4963, 2010.
- [123] SHRAVANI, M.; BALAIAH, S.; SRINIVAS, K.; BHANUPRAKASH, K.; HUC, I. Unusual regioselective electrophilic substitutions in quinoline foldamers: Conceptual dft and frontier molecular orbital analysis reveal the crucial role of folding and substituents. *ChemPhysChem*, v. 13, n. 15, p. 3526–3534, 2012.
- [124] PETKOVA, N. I.; NIKOLOVA, R. D.; KOSTOV, K. L.; MINEVA, T.; VAYSSILOV, G. N. Theoretical and experimental local reactivity parameters of 3-substituted coumarin derivatives. *The Journal of Physical Chemistry A*, v. 118, n. 46, p. 11062–11073, 2014.
- [125] DOMINGO, L. R.; PÉREZ, P.; SÁEZ, J. A. Understanding the local reactivity in polar organic reactions through electrophilic and nucleophilic parr functions. *RSC advances*, v. 3, n. 5, p. 1486–1494, 2013.
- [126] FRAU, J.; GLOSSMAN-MITNIK, D. Conceptual dft descriptors of amino acids with potential corrosion inhibition properties calculated with the latest minnesota density functionals. *Frontiers in chemistry*, v. 5, p. 16, 2017.
- [127] PARTHASARATHI, R.; AMUTHA, R.; SUBRAMANIAN, V.; NAIR, B. U.; RAMASAMI, T. Bader's and reactivity descriptors' analysis of dna base pairs. *The Journal of Physical Chemistry A*, v. 108, n. 17, p. 3817–3828, 2004.
- [128] YANG, W.; PARR, R. G. Hardness, softness, and the fukui function in the electronic theory of metals and catalysis. *Proceedings of the National Academy of Sciences*, v. 82, n. 20, p. 6723–6726, 1985.

- [129] CERÓN, M. L.; CALATAYUD, M. Application of dual descriptor to understand the activity of cu/zro 2 catalysts in the water gas shift reaction. *Journal of molecular modeling*, v. 23, n. 2, p. 34, 2017.
- [130] MARTÍNEZ, J.; CRUZ, V.; RAMOS, J.; GUTIÉRREZ-OLIVA, S.; MARTÍNEZ-SALAZAR, J.; TORO-LABBÉ, A. On the nature of the active site in bis (imino) pyridyl iron, a catalyst for olefin polymerization. *The Journal of Physical Chemistry C*, v. 112, n. 13, p. 5023–5028, 2008.
- [131] CERÓN, M.; GOMEZ, T.; CALATAYUD, M.; CÁRDENAS, C. Computing the fukui function in solid-state chemistry: Application to alkaline earth oxides bulk and surfaces. *The Journal of Physical Chemistry A*, v. 124, n. 14, p. 2826–2833, 2020.
- [132] KUMAR SRIVASTAVA, H.; CHOUDHURY, C.; NARAHARI SASTRY, G. The efficacy of conceptual dft descriptors and docking scores on the qsar models of hiv protease inhibitors. *Medicinal Chemistry*, v. 8, n. 5, p. 811–825, 2012.
- [133] SANDOVAL-YAÑEZ, C.; MASCAYANO, C.; MARTÍNEZ-ARAYA, J. I. A theoretical assessment of antioxidant capacity of flavonoids by means of local hyper-softness. *Arabian journal of chemistry*, v. 11, n. 4, p. 554–563, 2018.
- [134] BECK, M. E. Do fukui function maxima relate to sites of metabolism? a critical case study. *Journal of chemical information and modeling*, v. 45, n. 2, p. 273–282, 2005.
- [135] KARTHICK, T.; TANDON, P. Computational approaches to find the active binding sites of biological targets against busulfan. *Journal of molecular modeling*, v. 22, n. 6, p. 1–9, 2016.
- [136] SAID, M. S.; NAJIM, Z. A. Theoretical approach to relate the reactivity descriptors and mulliken charges with carcinogenicity of some methylated benzo [a] anthracene. *Pakistan Journal of Analytical & Environmental Chemistry*, v. 13, n. 1, p. 8, 2012.
- [137] PARTHASARATHI, R.; SUBRAMANIAN, V.; ROY, D. R.; CHATTARAJ, P. Electrophilicity index as a possible descriptor of biological activity. *Bioorganic & medicinal chemistry*, v. 12, n. 21, p. 5533–5543, 2004.
- [138] LIU, S.; ESS, D. H.; SCHAUER, C. K. Density functional reactivity theory characterizes charge separation propensity in proton-coupled electron transfer reactions. *The Journal of Physical Chemistry A*, v. 115, n. 18, p. 4738–4742, 2011.

- [139] BERKOWITZ, M.; GHOSH, S. K.; PARR, R. G. On the concept of local hardness in chemistry. *Journal of the American Chemical Society*, Washington, v. 107, n. 24, p. 6811–6814, 1985.
- [140] CHATTARAJ, P. K.; ROY, D. R.; GEERLINGS, P.; TORRENT-SUCARRAT, M. Local hardness: A critical account. *Theo. Chem. Acc.*, v. 118, n. 5-6, p. 923–930, 2007.
- [141] GÁL, T. Why the traditional concept of local hardness does not work. *Theoretical Chemistry Accounts*, v. 131, n. 7, p. 1–14, 2012.
- [142] GEERLINGS, P.; CHAMORRO, E.; CHATTARAJ, P. K.; DE PROFT, F.; GÁZQUEZ, J. L.; LIU, S.; MORELL, C.; TORO-LABBÉ, A.; VELA, A.; AYERS, P. Conceptual density functional theory: status, prospects, issues. *Theoretical Chemistry Accounts*, v. 139, n. 2, p. 1–18, 2020.
- [143] LANGENAEKER, W.; DE PROFT, F.; GEERLINGS, P. Development of local hardness-related reactivity indices: their application in a study of the reactivity of monosubstituted benzenes within the HSAB context. *The Journal of Physical Chemistry*, v. 99, n. 17, p. 6424–6431, 1995.
- [144] CÁRDENAS, C.; TIZNADO, W.; AYERS, P. W.; FUENTEALBA, P. The Fukui potential and the capacity of charge and the global hardness of atoms. *The Journal of Physical Chemistry A*, v. 115, n. 11, p. 2325–2331, 2011.
- [145] GAL, T.; GEERLINGS, P.; DE PROFT, F.; TORRENT-SUCARRAT, M. A new approach to local hardness. *Physical Chemistry Chemical Physics*, v. 13, n. 33, p. 15003–15015, 2011.
- [146] MENESES, L.; TIZNADO, W.; CONTRERAS, R.; FUENTEALBA, P. A proposal for a new local hardness as selectivity index. *Chemical physics letters*, v. 383, n. 1-2, p. 181–187, 2004.
- [147] ÖZEN, A. S.; DE PROFT, F.; AVIYENTE, V.; GEERLINGS, P. Interpretation of hydrogen bonding in the weak and strong regions using conceptual DFT descriptors. *The Journal of Physical Chemistry A*, v. 110, n. 17, p. 5860–5868, 2006.
- [148] LI, Y.; EVANS, J. The hard-soft acid-base principle in enzymatic catalysis: Dual reactivity of phosphoenolpyruvate. *Proceedings of the National Academy of Sciences*, v. 93, n. 10, p. 4612–4616, 1996.

- [149] HUANG, M.; MAYNARD, A.; TURPIN, J. A.; GRAHAM, L.; JANINI, G. M.; COVELL, D. G.; RICE, W. G. Anti-hiv agents that selectively target retroviral nucleocapsid protein zinc fingers without affecting cellular zinc finger proteins. *Journal of medicinal chemistry*, v. 41, n. 9, p. 1371–1381, 1998.
- [150] ALBER, F.; KUONEN, O.; SCAPOZZA, L.; FOLKERS, G.; CARLONI, P. Density functional studies on herpes simplex virus type 1 thymidine kinase–substrate interactions: The role of tyr-172 and met-128 in thymine fixation. *Proteins: Structure, Function, and Bioinformatics*, v. 31, n. 4, p. 453–459, 1998.
- [151] MIGNON, P.; STEYAERT, J.; LORIS, R.; GEERLINGS, P.; LOVERIX, S. A nucleophile activation dyad in ribonucleases. 2002.
- [152] RIVAS, P.; ZAPATA-TORRES, G.; MELIN, J.; CONTRERAS, R. Probing the hydride transfer process in the lumiflavine–1-methylnicotinamide model system using group softness. *Tetrahedron*, v. 60, n. 19, p. 4189–4196, 2004.
- [153] LEE, Y.-M.; LIM, C. Physical basis of structural and catalytic zn-binding sites in proteins. *Journal of molecular biology*, v. 379, n. 3, p. 545–553, 2008.
- [154] FRAU, J.; GLOSSMAN-MITNIK, D. Comparative study of the chemical reactivity of helical peptide models for protein glycation. *Computational Chemistry*, v. 5, n. 2, p. 65–73, 2017.
- [155] FRAU, J.; RAMIS, R.; GLOSSMAN-MITNIK, D. Computational prediction of the preferred glycation sites of model helical peptides derived from human serum albumin (hsa) and lysozyme helix 4 (lh4). *Theoretical Chemistry Accounts*, v. 136, n. 4, p. 39, 2017.
- [156] RABI, S.; PATEL, A. H.; BURGER, S. K.; VERSTRAELEN, T.; AYERS, P. W. Exploring the substrate selectivity of human seh and m. tuberculosis ehb using qm/mm. *Structural Chemistry*, v. 28, n. 5, p. 1501–1511, 2017.
- [157] OLLER, J.; SAEZ, D. A.; VOHRINGER-MARTINEZ, E. Atom-condensed fukui function in condensed phases and biological systems and its application to enzymatic fixation of carbon dioxide. *The Journal of Physical Chemistry A*, v. 124, n. 5, p. 849–857, 2020.

- [158] FRANCO-PÉREZ, M.; AYERS, P. W.; GÁZQUEZ, J. L. Average electronic energy is the central quantity in conceptual chemical reactivity theory. *Theoretical Chemistry Accounts*, v. 135, n. 8, p. 1–8, 2016.
- [159] ROOS, G.; LOVERIX, S.; DE PROFT, F.; WYNS, L.; GEERLINGS, P. A computational and conceptual dft study of the reactivity of anionic compounds: implications for enzymatic catalysis. *The Journal of Physical Chemistry A*, v. 107, n. 35, p. 6828–6836, 2003.
- [160] ROOS, G.; MESSENS, J.; LOVERIX, S.; WYNS, L.; GEERLINGS, P. A computational and conceptual dft study on the michaelis complex of pi258 arsenate reductase. structural aspects and activation of the electrophile and nucleophile. *The Journal of Physical Chemistry B*, v. 108, n. 44, p. 17216–17225, 2004.
- [161] ROOS, G.; LOVERIX, S.; BROSENS, E.; VAN BELLE, K.; WYNS, L.; GEERLINGS, P.; MESSENS, J. The activation of electrophile, nucleophile and leaving group during the reaction catalysed by pi258 arsenate reductase. *ChemBioChem*, v. 7, n. 6, p. 981–989, 2006.
- [162] LIU, S. *Conceptual density functional theory: Towards a new chemical reactivity theory*. John Wiley & Sons, 2022.
- [163] GEERLINGS, P. From density functional theory to conceptual density functional theory and biosystems. *Pharmaceuticals*, v. 15, n. 9, p. 1112, 2022.
- [164] FLORES-HOLGUÍN, N.; FRAU, J.; GLOSSMAN-MITNIK, D. Virtual Prospection of Marine Cyclopeptides as Therapeutics by Means of Conceptual DFT and Computational ADMET. *Pharmaceuticals*, v. 15, n. 5, may 2022.
- [165] FLORES-HOLGUÍN, N.; FRAU, J.; GLOSSMAN-MITNIK, D. Virtual screening of marine natural compounds by means of chemoinformatics and CDFT-based computational peptidology. *Marine Drugs*, v. 18, n. 9, 2020.
- [166] FLORES-HOLGUÍN, N.; FRAU, J.; GLOSSMAN-MITNIK, D. An integrated molecular modeling protocol for drug screening based on conceptual density functional theory and chemoinformatics for the study of marine cyclopeptides. *Journal of Molecular Modeling*, v. 27, n. 11, 2021.

- [167] KHANDOGIN, J.; YORK, D. M. Quantum descriptors for biological macromolecules from linear-scaling electronic structure methods. *Proteins: Structure, Function, and Bioinformatics*, v. 56, n. 4, p. 724–737, 2004.
- [168] FAVER, J.; MERZ JR, K. M. Utility of the hard/soft acid- base principle via the fukui function in biological systems. *Journal of chemical theory and computation*, v. 6, n. 2, p. 548–559, 2010.
- [169] LA PORTA, F.; GIACOPPO, J.; RAMOS, P.; GUERREIRO, M.; RAMALHO, T. Computational insights into the role of the frontier orbital in the chemistry of tridentate ligands. *American Journal of Chemistry*, v. 2, n. 5, p. 255–262, 2012.
- [170] PINO-RIOS, R.; YAÑEZ, O.; INOSTROZA, D.; RUIZ, L.; CARDENAS, C.; FUENTE-ALBA, P.; TIZNADO, W. Proposal of a simple and effective local reactivity descriptor through a topological analysis of an orbital-weighted fukui function. *Journal of computational chemistry*, v. 38, n. 8, p. 481–488, 2017.
- [171] BERMAN, H. M. The protein data bank: a historical perspective. *Acta Crystallographica Section A*, v. 64, n. 1, p. 88–95, 2008.
- [172] SALOMON-FERRER, R.; CASE, D. A.; WALKER, R. C. An overview of the amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, v. 3, n. 2, p. 198–210, 2013.
- [173] WANG, J.; WOLF, R. M.; CALDWELL, J. W.; KOLLMAN, P. A.; CASE, D. A. Development and testing of a general amber force field. *Journal of computational chemistry*, v. 25, n. 9, p. 1157–1174, 2004.
- [174] LINDORFF-LARSEN, K.; PIANA, S.; PALMO, K.; MARAGAKIS, P.; KLEPEIS, J. L.; DROR, R. O.; SHAW, D. E. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, v. 78, n. 8, p. 1950–1958, 2010.
- [175] PRICE, D. J.; BROOKS III, C. L. A modified tip3p water potential for simulation with ewald summation. *The Journal of chemical physics*, v. 121, n. 20, p. 10096–10103, 2004.
- [176] VAN DER SPOEL, D.; LINDAHL, E.; HESS, B.; GROENHOF, G.; MARK, A. E.; BERENDSEN, H. J. Gromacs: fast, flexible, and free. *Journal of computational chemistry*, v. 26, n. 16, p. 1701–1718, 2005.

- [177] SWAILS, J.; HERNANDEZ, C.; MOBLEY, D. L.; NGUYEN, H.; WANG, L.-P.; JANOWSKI, P. Parmed. URL: <https://github.com/ParmEd/ParmEd>, 2010.
- [178] HESS, B.; BEKKER, H.; BERENDSEN, H. J.; FRAAIJE, J. G. Lincs: a linear constraint solver for molecular simulations. *Journal of computational chemistry*, v. 18, n. 12, p. 1463–1472, 1997.
- [179] THIEL, W. Semiempirical quantum-chemical methods in computational chemistry. In: *Theory and applications of computational chemistry*. Elsevier, 2005. p. 559–580.
- [180] HUSCH, T.; VAUCHER, A. C.; REIHER, M. Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation. *International Journal of Quantum Chemistry*, v. 118, 12 2018.
- [181] DEWAR, M. J.; THIEL, W. Ground states of molecules. 38. the mndo method. approximations and parameters. *Journal of the American Chemical Society*, Washington, v. 99, n. 15, p. 4899–4907, 1977.
- [182] MCNAMARA, J. P.; HILLIER, I. H. Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules. *Physical chemistry chemical physics*, v. 9, n. 19, p. 2362–2370, 2007.
- [183] NAM, K.; CUI, Q.; GAO, J.; YORK, D. M. Specific reaction parametrization of the am1/d hamiltonian for phosphoryl transfer reactions: H, o, and p atoms. *Journal of Chemical Theory and Computation*, v. 3, n. 2, p. 486–504, 2007.
- [184] ROCHA, G. B.; FREIRE, R. O.; SIMAS, A. M.; STEWART, J. J. Rm1: A reparameterization of am1 for h, c, n, o, p, s, f, cl, br, and i. *Journal of computational chemistry*, v. 27, n. 10, p. 1101–1111, 2006.
- [185] STEWART, J. J. Optimization of parameters for semiempirical methods i. method. *Journal of computational chemistry*, v. 10, n. 2, p. 209–220, 1989.
- [186] STEWART, J. J. Optimization of parameters for semiempirical methods v: Modification of nddo approximations and application to 70 elements. *Journal of Molecular modeling*, v. 13, n. 12, p. 1173–1213, 2007.

- [187] STEWART, J. J. Optimization of parameters for semiempirical methods vi: more modifications to the nddo approximations and re-optimization of parameters. *Journal of molecular modeling*, v. 19, n. 1, p. 1–32, 2013.
- [188] KOLB, M.; THIEL, W. Beyond the mndo model: Methodical considerations and numerical results. *Journal of computational chemistry*, v. 14, n. 7, p. 775–789, 1993.
- [189] XU, L.; CAI, W.; SHAO, X. Performance of the semiempirical AM1, PM3, MNDO, and tight-binding methods in comparison with DFT method for the large fullerenes C116-C120. *Journal of Molecular Structure: THEOCHEM*, v. 817, n. 1-3, p. 35–41, sep 2007.
- [190] KOBAYASHI, M.; NAKAI, H. Divide-and-conquer approaches to quantum chemistry: Theory and implementation. In: *Linear-scaling techniques in computational chemistry and physics*. Springer, 2011. p. 97–127.
- [191] TITMUSS, S. J.; CUMMINS, P. L.; BLIZNYUK, A. A.; RENDELL, A. P.; GREASY, J. E. Comparison of linear-scaling semiempirical methods and combined quantum mechanical/molecular mechanical methods applied to enzyme reactions. *Chemical Physics Letters*, Amsterdam, v. 320, n. 1-2, p. 169–176, 2000.
- [192] STEWART, J. J. Mopac: a semiempirical molecular orbital program. *Journal of computer-aided molecular design*, v. 4, n. 1, p. 1–103, 1990.
- [193] LONSDALE, R.; RANAGHAN, K. E.; MULHOLLAND, A. J. Computational enzymology. *Chemical communications*, v. 46, n. 14, p. 2354–2372, 2010.
- [194] GAO, J.; AMARA, P.; ALHAMBRA, C.; FIELD, M. J. A generalized hybrid orbital (gho) method for the treatment of boundary atoms in combined qm/mm calculations. *The Journal of Physical Chemistry A*, v. 102, n. 24, p. 4714–4721, 1998.
- [195] BACHEGA, J. F. R.; TIMMERS, L. F. S.; ASSIRATI, L.; BACHEGA, L. R.; FIELD, M. J.; WYMORE, T. Gtcdynamo: A pymol plug-in for qc/mm hybrid potential simulations. 2013.
- [196] FIELD, M. J. The pdynamo program for molecular simulations using hybrid quantum chemical and molecular mechanical potentials. *Journal of Chemical Theory and Computation*, v. 4, n. 7, p. 1151–1161, 2008.

- [197] HANWELL, M. D.; CURTIS, D. E.; LONIE, D. C.; VANDERMEERSCH, T.; ZUREK, E.; HUTCHISON, G. R. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of cheminformatics*, v. 4, n. 1, p. 1–17, 2012.
- [198] ALLOUCHE, A.-R. Gabedit—a graphical user interface for computational chemistry softwares. *Journal of computational chemistry*, v. 32, n. 1, p. 174–182, 2011.
- [199] SCHAFTENAAR, G.; NOORDIK, J. H. Molden: a pre-and post-processing program for molecular and electronic structures. *Journal of computer-aided molecular design*, v. 14, n. 2, p. 123–134, 2000.
- [200] SÁNCHEZ-MÁRQUEZ, J.; ZORRILLA, D.; SÁNCHEZ-CORONILLA, A.; DESIRÉE, M.; NAVAS, J.; FERNÁNDEZ-LORENZO, C.; ALCÁNTARA, R.; MARTÍN-CALLEJA, J. Introducing “uca-fukui” software: reactivity-index calculations. *Journal of molecular modeling*, v. 20, n. 11, p. 1–13, 2014.
- [201] NATH, S. R.; KURUP, S. S.; JOSHI, K. A. Pyglobal: A toolkit for automated compilation of dft-based descriptors, 2016.
- [202] LU, T.; CHEN, F. Multiwfn: a multifunctional wavefunction analyzer. *Journal of computational chemistry*, v. 33, n. 5, p. 580–592, 2012.
- [203] NELSON, M. T.; HUMPHREY, W.; GURSOY, A.; DALKE, A.; KALÉ, L. V.; SKEEL, R. D.; SCHULTEN, K. Namd: a parallel, object-oriented molecular dynamics program. *The International Journal of Supercomputer Applications and High Performance Computing*, v. 10, n. 4, p. 251–268, 1996.
- [204] GORDON, M. S.; SCHMIDT, M. W. Advances in electronic structure theory: Gamess a decade later. In: *Theory and applications of computational chemistry*. Elsevier, 2005. p. 1167–1189.
- [205] NEESE, F. The orca program system. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, v. 2, n. 1, p. 73–78, 2012.
- [206] FRISCH, A. et al. gaussian 09w reference. *Wallingford, USA*, 25p, 2009.
- [207] LUTZ, M. *Programming python*. "O'Reilly Media, Inc.", 2001.
- [208] STROUSTRUP, B. An overview of the c++ programming language. *Handbook of object technology*, 1999.

- [209] Primordia 1.2v lançamento de versão estável.
- [210] LU, T.; CHEN, Q. Realization of conceptual density functional theory and information-theoretic approach in multiwfn program. *Conceptual Density Functional Theory: Towards a New Chemical Reactivity Theory*, v. 2, p. 631–647, 2022.
- [211] PUJAL*, L.; TEHRANI*, A.; HEIDAR-ZADEH, F. Chemtools: Gain chemical insight from quantum chemistry calculations. *Conceptual Density Functional Theory: Towards a New Chemical Reactivity Theory*, v. 2, p. 649–661, 2022.
- [212] HERNÁNDEZ-ESPARZA, R.; VÁZQUEZ-MAYAGOITIA, Á.; SORIANO-AGUEDA, L.-A.; VARGAS, R.; GARZA, J. Gpus as boosters to analyze scalar and vector fields in quantum chemistry. *International Journal of Quantum Chemistry*, v. 119, n. 2, p. e25671, 2019.
- [213] PECCATI, F. Nciplot4 guide for biomolecules: an analysis tool for noncovalent interactions. *Journal of chemical information and modeling*, v. 60, n. 1, p. 6–10, 2020.
- [214] BOTO, R. A.; PECCATI, F.; LAPLAZA, R.; QUAN, C.; CARBONE, A.; PIQUEMAL, J.-P.; MADAY, Y.; CONTRERAS-GARCIA, J. Nciplot4: Fast, robust, and quantitative analysis of noncovalent interactions. *Journal of Chemical Theory and Computation*, v. 16, n. 7, p. 4150–4158, 2020.
- [215] HARRIS, T. K.; COLE, R. N.; COMER, F. I.; MILDVAN, A. S. Proton transfer in the mechanism of triosephosphate isomerase. *Biochemistry*, v. 37, n. 47, p. 16828–16838, 1998.
- [216] WIERENGA, R.; KAPETANIOU, E.; VENKATESAN, R. Triosephosphate isomerase: a highly evolved biocatalyst. *Cellular and molecular life sciences*, v. 67, n. 23, p. 3961–3982, 2010.
- [217] GUALLAR, V.; JACOBSON, M.; MCDERMOTT, A.; FRIESNER, R. A. Computational modeling of the catalytic reaction in triosephosphate isomerase. *Journal of molecular biology*, v. 337, n. 1, p. 227–239, 2004.
- [218] DAMBORSKÝ, J.; KUTÝ, M.; NĚMEC, M.; KOCA, J. A molecular modeling study of the catalytic mechanism of haloalkane dehalogenase: 1. quantum chemical study of the first reaction step. *Journal of chemical information and computer sciences*, v. 37, n. 3, p. 562–568, 1997.

- [219] SCHUMACHER, M. A.; SCOTT, D. M.; MATHEWS, I. I.; EALICK, S. E.; ROOS, D. S.; ULLMAN, B.; BRENNAN, R. G. Crystal structures of toxoplasma gondii adenosine kinase reveal a novel catalytic mechanism and prodrug binding. *Journal of molecular biology*, v. 296, n. 2, p. 549–567, 2000.
- [220] KOWALUK, E. A.; JARVIS, M. F. Therapeutic potential of adenosine kinase inhibitors. *Expert opinion on investigational drugs*, v. 9, n. 3, p. 551–564, 2000.
- [221] KENNES, C.; PRIES, F.; KROOSHOF, G. H.; BOKMA, E.; KINGMA, J.; JANSSEN, D. B. Replacement of tryptophan residues in haloalkane dehalogenase reduces halide binding and catalytic activity. *European journal of biochemistry*, v. 228, n. 2, p. 403–407, 1995.
- [222] MCCLORY, J.; LIN, J.-T.; TIMSON, D. J.; ZHANG, J.; HUANG, M. Catalytic mechanism of mevalonate kinase revisited, a qm/mm study. *Organic & biomolecular chemistry*, v. 17, n. 9, p. 2423–2431, 2019.
- [223] PEREIRA, L. D. S. et al. Estudo teórico do mecanismo de ação da enzima rta da ricina com um ligante análogo ao substrato natural rrna 28s. 2020.
- [224] SANTOS, A. R. et al. Estudo de interações proteína-ligante e do enovelamento de proteínas através de dinâmica molecular, modelos de estado de markov, cálculos quânticos e descritores de reatividade. 2021.
- [225] SILVA, J. C. A. D. et al. Simulação molecular de proteínas com nós geométricos. 2022.
- [226] SOUZA FARIAS, W. M. Uso de descritores químico-quânticos como etapa de pós-processamento do protocolo de redesenho de enzimas. 2022.
- [227] RANI, S. Combined quantum/classical studies of the mechanism of g6pd in the oxidative pentose phosphate pathway. 2023.
- [228] IGOR BARDEN GRILLO, G. B. R. Modelando sistemas moleculares e biológicos com métodos de química quântica - parte 3. url = <https://www.youtube.com/watch?v=r-fXI9JJtnA&t=186s>.
- [229] IGOR BARDEN GRILLO, G. B. R. Minicurso - (dia 5) métodos quânticos semiempíricos. url = <https://www.youtube.com/watch?v=itgqEWeA6nM&t=83s>.



Theoretical characterization of the shikimate 5-dehydrogenase reaction from *Mycobacterium tuberculosis* by hybrid QC/MM simulations and quantum chemical descriptors

Igor Barden Grillo¹ · José Fernando Ruggiero Bachega^{2,3,4} · Luis Fernando S. M. Timmers^{2,5} · Rafael A. Caceres⁶ · Osmar Norberto de Souza^{2,7} · Martin J. Field^{8,9} · Gerd Bruno Rocha¹

Received: 2 June 2020 / Accepted: 7 September 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

In this study, we have investigated the enzyme shikimate 5-dehydrogenase from the causative agent of tuberculosis, *Mycobacterium tuberculosis*. We have employed a mixture of computational techniques, including molecular dynamics, hybrid quantum chemical/molecular mechanical potentials, relaxed surface scans, quantum chemical descriptors and free-energy simulations, to elucidate the enzyme's reaction pathway. Overall, we find a two-step mechanism, with a single transition state, that proceeds by an energetically uphill hydride transfer, followed by an energetically downhill proton transfer. Our mechanism and calculated free energy barrier for the reaction, 64.9 kJ mol⁻¹, are in good agreement with those predicted from experiment. An analysis of quantum chemical descriptors along the reaction pathway indicated a possibly important, yet currently unreported, role of the active site threonine residue, Thr65.

Keywords *Mycobacterium tuberculosis* · Free-energy profiles · Fast quantum chemical descriptors · Quantum chemical/molecular mechanical potentials · Reaction mechanism · Shikimate 5-dehydrogenase

Introduction

Enzymes are the most important catalysts in life chemistry. These proteins can increase reaction rates by a factor up to 10¹⁹, thereby tuning the chemical reactions in metabolic pathways [42]. As a result, enzymes have substantial relevance in many areas of biochemistry, including biotechnology, nanoscience, and medicinal chemistry. The selection of a suitable protein to target is a crucial step in the drug development process of bioactive molecules, such as drugs and herbicides. The protein should either be in the

metabolic pathway that is present in the pathogen but absent in the hosts, following the rule of selective toxicity, or be sufficiently different that it can be specifically targeted.

The shikimate (SKM) pathway [19] is a seven-step enzymatic pathway that converts erythrose 4-phosphate and phosphoenolpyruvate to chorismate, a biosynthetic precursor of aromatic amino acids, vitamins B9 and K1, ubiquinone, and salicylate. This route is present in plants, fungi, bacteria, and apicomplexan parasites, but not in metazoans, such as humans beings, which makes the enzymes from this enzymatic pathway promising targets for non-toxic antimicrobials and herbicides. The widely used herbicide known as glyphosate is an effective inhibitor of 5-enolpyruvyl-shikimate-3-phosphate (EPSP) synthase, the sixth step on the SKM pathway [36, 37].

An example of an organism in which the SKM pathway has been targeted is *Mycobacterium tuberculosis*. This bacterium is the causative agent of tuberculosis which, according to the World Health Organization, is the most fatal infectious disease with over one million deaths per year [28]. Gene disruption experiments have demonstrated that the SKM pathway is essential for the viability of *M. tuberculosis* [29] and, as a result, substantial efforts have been made to develop inhibitors against its constituent

This paper belongs to Topical Collection XX-Brazilian Symposium of Theoretical Chemistry (SBQT2019)

Igor Barden Grillo and José Fernando Ruggiero Bachega are co-first authors.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00894-020-04536-9>) contains supplementary material, which is available to authorized users.

✉ Igor Barden Grillo
barden.igor@gmail.com

Extended author information available on the last page of the article.

Elucidating Enzymatic Catalysis Using Fast Quantum Chemical Descriptors

Igor Barden Grillo, Gabriel A. Urquiza-Carvalho, José Fernando Ruggiero Bachega, and Gerd Bruno Rocha*



Cite This: <https://dx.doi.org/10.1021/acs.jcim.9b00860>



Read Online

ACCESS |



Metrics & More

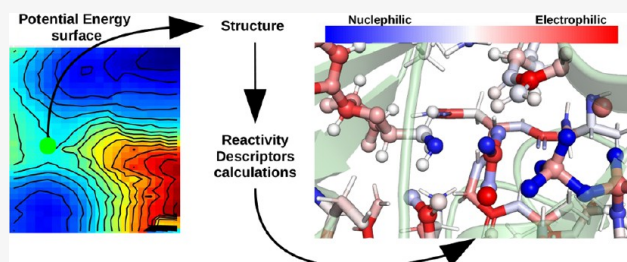


Article Recommendations



Supporting Information

ABSTRACT: In general, computational simulations of enzymatic catalysis processes are thermodynamic and structural surveys to complement experimental studies, requiring high level computational methods to match accurate energy values. In the present work, we propose the usage of reactivity descriptors, theoretical quantities calculated from the electronic structure, to characterize enzymatic catalysis outlining its reaction profile using low-level computational methods, such as semiempirical Hamiltonians. We simulate three enzymatic reactions paths, one containing two reaction coordinates and without prior computational study performed, and calculate the reactivity descriptors for all obtained structures. We observed that the active site local hardness does not change substantially, even more so for the amino-acid residues that are said to stabilize the reaction structures. This corroborates with the theory that activation energy lowering is caused by the electrostatic environment of the active sites. Also, for the quantities describing the atom electrophilicity and nucleophilicity, we observed abrupt changes along the reaction coordinates, which also shows the enzyme participation as a reactant in the catalyzed reaction. We expect that such electronic structure analysis allows the expedient proposition and/or prediction of new mechanisms, providing chemical characterization of the enzyme active sites, thus hastening the process of transforming the resolved protein three-dimensional structures in catalytic information.



INTRODUCTION

Enzymatic catalysis allows the chemistry of life to take place in a reasonable time scale.¹ Once a genome is sequenced, millions of new protein are discovered, a small part has its three-dimensional structure determined, and even fewer have their function studied in detail. A fraction of these becomes the subject of studies about enzymatic catalysis, which is central to guide mutation studies, drug design, and technological applications for other substrates.²

The experimental approach is very limited to provide a desired description of these processes, mainly due to the short-lived reaction intermediates, the number of amino-acid residues, and the specifics of how their roles in the mechanism govern the catalysis.³ Thus, the emergence of computational modeling, especially of quantum chemistry/molecular mechanics (QC/MM) hybrid methods,^{4,5} has been successfully applied for reaction path determination, transition state structure optimization, and thermodynamic/kinetic properties prediction.⁶

By employing quantum chemical calculations, the theoretical characterization of the active site becomes possible with the exploration of its electronic structure.⁷ Some tools for carrying out such exploration are reactivity descriptors (RDs), which are theoretical quantities that outline the propensity of a molecule/atoms to react.⁸ An emblematic example of the use

of such RDs in enzymatic reaction studies was their utilization for determining the most reactive angle of attack and ionization state of the phosphoenolpyruvate substrate at the 5-enolpyruvylshikimate 3-phosphate synthase active site.⁹ Additionally, it is possible to find applications for the determination of the protonation states of catalytic amino-acid residues¹⁰ and the oxidation state of arsenate ions.¹¹

The RDs cited above are within the conceptual density functional theory (CDFT) framework.⁸ This framework provides rigorous mathematical definitions for concepts like molecular hardness and electrophilic attack susceptibility, within the fundamental equations of the density functional theory,¹² thus bringing together reactivity theories such as the frontier molecular orbital¹³ and the hard and soft acid Base theory.¹⁴

The application of these quantities is very extensive for small molecules,¹⁵ being successfully employed to provide explanations in terms of generic Lewis acid/base for the majority of organic reactions, such as substitutions, eliminations, and

Special Issue: Molecular Simulation in Latin America: Coming of Age

Received: September 27, 2019

Published: January 2, 2020



ACS Publications

© XXXX American Chemical Society

A

<https://dx.doi.org/10.1021/acs.jcim.9b00860>
J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

PRIMoRDiA: A Software to Explore Reactivity and Electronic Structure in Large Biomolecules

Igor Barden Grillo, Gabriel A. Urquiza-Carvalho, and Gerd Bruno Rocha*



Cite This: <https://dx.doi.org/10.1021/acs.jcim.0c00655>



Read Online

ACCESS |



Metrics & More

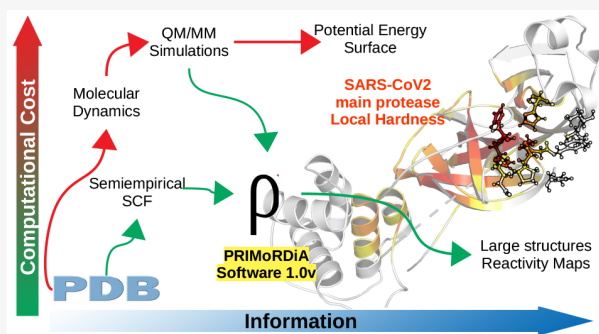


Article Recommendations



Supporting Information

ABSTRACT: Plenty of enzymes with structural data do not have their mechanism of catalysis elucidated. Reactivity descriptors, theoretical quantities generated from resolved electronic structure, provide a way to predict and rationalize chemical processes of such systems. In this Application Note, we present PRIMoRDiA (PRIMoRDiA Macromolecular Reactivity Descriptors Access), a software built to calculate the reactivity descriptors of large biosystems by employing an efficient and accurate treatment of the large output files produced by quantum chemistry packages. Here, we show the general implementation details and the software main features. Calculated descriptors were applied for a set of enzymatic systems in order to show their relevance for biological studies and the software potential for use in large scale. Also, we test PRIMoRDiA to aid in the interaction depiction between the SARS-CoV-2 main protease and a potential inhibitor.



INTRODUCTION

The increase in computational power combined with the development of fast algorithms have allowed the application of quantum chemistry methods to biological macromolecules.¹ The exploration of this new frontier requires novel computational techniques in order to handle the wealth of electronic structure information that is now available for such large systems.

Reactivity descriptors are an alternative to predict and rationalize chemical processes. These descriptors are convenient theoretical quantities generated from mathematical manipulations of converged electronic structure results, such as molecular orbital (MO) coefficients and energies.² The most employed ones are defined in the context of conceptual density functional theory (CDFT),³ which unifies different reactivity theories in a single theoretical framework.

Even though the electronic structure information has become available for large biomolecules, many enzymes with resolved 3D structures do not have their catalytic cycles chemically described. In this context, the descriptors could be employed as an *in silico* way of revealing the reactivity roles of amino acid residues and substrate atoms. However, most of the common protocols for the computation of reactivity descriptors fail for larger molecules; too often are issues related to the unique characteristics of the molecule's electronic structure.

Recently, a study carried out in our research group showed that semiempirical quantum chemistry methods display sufficient accuracy to obtain chemical reactivity information for polypeptides. In this work, we also showed that the correct

computation of the descriptors requires a window of MOs around the HOMO–LUMO gap, instead of only the frontier orbitals.⁴ Also, in another two of our research group's works, the reactivity descriptors properly adjusted for macromolecules were capable of theoretically characterizing steps in the catalytic mechanisms of some enzymes.^{5,6}

Despite such advances, the popular postprocessing computational chemistry programs are not suited for calculation of the electronic structure of macromolecules. UCA-Fukui is a specialized software for calculating common CDFT descriptors,⁷ but it is limited to parse outputs of the GAUSSIAN package and has not been tested for large molecules. PYGlobal is another application designed to do these calculations,⁸ which emphasizes automation features to produce global quantities for several structures simultaneously as well as tools to ease statistical analysis. However, PYGlobal does not calculate local descriptors, which are the most relevant for biological structures.

To fill this niche, we have developed PRIMoRDiA (PRIMoRDiA Macromolecular Reactivity Descriptors Access), a software capable of accurately and efficiently computing reactivity descriptors for macromolecules. The current version

Received: June 9, 2020



ACS Publications

© XXXX American Chemical Society

A

<https://dx.doi.org/10.1021/acs.jcim.0c00655>
J. Chem. Inf. Model. XXXX, XXX, XXX–XXX

Quantum chemical descriptors based on semiempirical methods for large biomolecules

Cite as: J. Chem. Phys. 158, 201001 (2023); doi: 10.1063/5.0132687

Submitted: 30 October 2022 • Accepted: 3 May 2023 •

Published Online: 24 May 2023



Igor B. Grillo,^{a)} Gabriel A. Urquiza-Carvalho,^{b)} and Gerd B. Rocha^{c)}

AFFILIATIONS

Departamento de Química, Centro de Ciências Exatas e da Natureza, Universidade Federal da Paraíba, João Pessoa, Brazil

Note: This paper is part of the JCP Special Topic on Modern Semiempirical Electronic Structure Methods.

^{a)}Electronic mail: barden.grillo@gmail.com

^{b)}Electronic mail: urquizagabes@gmail.com

^{c)}Author to whom correspondence should be addressed: gbr@quimica.ufpb.br. URL: <https://www.quantum-chem.pro.br>

ABSTRACT

In this Review, we reviewed the efforts to expand the applications of conceptual density functional theory reactivity descriptors and hard and soft acid and base principles for macromolecules and other strategies that focused on low-level quantum chemistry methods. Currently, recent applications are taking advantage of modifications of these descriptors using semiempirical electronic structures to explain enzymatic catalysis reactions, protein-binding processes, and structural analysis in proteins. We have explored these new solutions along with their implementations in the software PRIMoRDIA, discussing their impact on the field and its perspectives. We show the main issues in the analysis of the electronic structure of macromolecules, which are the application of the same calculation protocols used for small molecules without considering particularities in those large systems' electronic configuration. The major result of our discussions is that the use of semiempirical methods is crucial to obtain such a type of analysis, which can provide a powerful dimension of information and be part of future low-cost predictive tools. We expect semiempirical methods continue playing an important role in the quantum chemistry evaluation of large molecules. As computational resources advance, semiempirical methods might lead us to explore the electronic structure of even larger biological macromolecular entities and sets of structures representing larger timescales.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0132687>

I. INTRODUCTION

Computational modeling techniques are as important for studying biological systems as experimental techniques.^{1–3} Molecular dynamics (MD) simulations are an example of how solely theoretical analysis can reveal relevant information about proteins in their medium, interacting with other molecules in a great variety of conditions.^{4–7} Despite the mentioned applications, molecular dynamics that use molecular mechanics (MM) force fields do not account for the electronic structure and cannot model interactions such as those reliant on polarizability, charge transfers, and the rupture and formation of chemical bonds.^{8,9}

To simulate chemical reactions and increase the accuracy of simulations, methods based on quantum mechanics (QM) are needed.¹⁰ However, using these methods comes with a steep price because the computational resources required in those simulations increase exponentially with the system size, becoming impractical

to perform calculations for macromolecules, such as entire protein structures.^{11,12} Thus, approximations in the wave function, electron density, Hamiltonians, and system truncation were often sought for those systems.^{13,14}

Semiempirical Hamiltonians¹⁵ and hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) simulations¹⁶ have already enabled energetic and structural investigations of enzymatic-catalyzed chemical reactions even before the start of the 21st century.^{17,18} However, only with the possibility of using popular DFT functionals, the use of biological quantum chemistry simulations did see an upward trend.^{19,20} Indeed, DFT has played an important role in the investigation of enzymatic catalysis either in early gas-phase simulations or in increasing applications with QM/MM protocols^{21–23} and is still relevant in biophysical and biochemistry studies.²⁴ Additionally, molecular docking and affinity studies of large biological structures have become the benchmark for these protocols, significantly changing the drug discovery process.^{9,25}