

Bruno Camargo Braga

**Mutação na tríade catalítica da IsPETase: Um estudo com simulações de dinâmica molecular**

Campo Grande

2023

Bruno Camargo Braga

**Mutação na tríade catalítica da IsPETase: Um estudo  
com simulações de dinâmica molecular**

Tese apresentada ao Instituto de Física da  
Universidade Federal de Mato Grosso do Sul,  
sendo parte dos requisitos para a obtenção do  
título de Doutor em Ciência dos Materiais.

Universidade Federal de Mato Grosso do Sul - UFMS

Instituto de Física - INFI

Programa de Pós-Graduação em Ciência dos Materiais - PGCM

Orientador: Prof. Dr. Marcos Serrou do Amaral

Campo Grande

2023

*Este trabalho é dedicado aos meus pais.*

# Agradecimentos

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) por financiar este trabalho.

Ao professor doutor Marcos Serrou do Amaral, por me inspirar em mudar da área que atuei no mestrado para esta de modelagem molecular.

Ao técnico do laboratório de modelagem molecular do Instituto de Física - UFMS, Rodrigo Fioravante, pelas discussões e ajuda com relação aos diversos pacotes computacionais utilizados durante meu período de pós-graduação.

Ao meu amigo Marcio Fernandes da Silva, por me ajudar a entender os processos de simulação molecular usando as ferramentas *AMBER/AMBERtools*. Essas discussões me ajudaram a projetar e implementar o programa de computador *ASM.py*.

Aos professores da Faculdade de Computação, dos Institutos de Física, Matemática e Química da Universidade Federal de Mato Grosso do Sul pela formação acadêmica.

Aos meus pais por todo suporte e incentivo.

*“Trust in the LORD with all your heart  
and lean not on your own understanding;  
in all your ways acknowledge him,  
and he will make your paths straight.”*

*(Holy Bible, New International Version, 2000, Proverbs, 3:5-6)*

# Resumo

O lixo plástico causa graves problemas ambientais em todo o mundo, e a reciclagem biológica é um dos métodos mais ecologicamente amigáveis para lidar com isso. A descoberta da *Ideonella sakaiensis* PET hidrolase (IsPETase) atraiu o interesse da comunidade científica. Sua natureza mesófila é uma característica interessante para novas perspectivas na pesquisa de degradação do politereftalato de etileno (PET). Aqui, usamos uma abordagem de Dinâmica Molecular (DM) e discutimos a possibilidade de variações da tríade catalítica canônica em pH 7 e 300 K. Preparamos três variantes e realizamos um protocolo de *Docking* molecular comparativo do BHET, – do inglês *bis(2-hydroxyethyl) terephthalate* – nas três variantes e na enzima nativa. Realizamos simulações de DM com foco na distância nucleofílica e nos estados de protonação dos resíduos escolhidos em nossas mutações. Durante este trabalho, notamos a possibilidade de automatizar o uso dos pacotes computacionais *AMBER/AMBERtools*. Um programa denominado ASM.py, do termo inglês *AMBER Simulation Manager*, foi desenvolvido para esse fim e registrado sob a licença GNU GPL no Instituto Nacional da Propriedade Industrial (INPI). Em todas as simulações, o ligante permaneceu ligado ao sítio ativo. A energia de ligação total nas variantes diminuiu, mas sempre permaneceu favorável para interação. Ao analisar a contribuição da interação de cada resíduo, as simulações com a nativa e a D206E apresentaram uma ligação mais estável com basicamente o mesmo padrão energético. No entanto, as variantes com lisina apresentaram redução nas contribuições desfavoráveis – enquanto a D206E/H237K apresentou um comportamento intermediário se comparado as mutantes simples. A nova distribuição de carga e conformação das cadeias laterais para ambas as variantes com lisina mantiveram a distância nucleofílica, enquanto empurrava o ligante para longe da fenda de oxiânion. As mutantes de um resíduo estudadas aqui não apresentaram resultados promissores. No entanto, o comportamento peculiar da D206E/H237K provoca curiosidade sobre quais consequências podem surgir em dados empíricos. Acreditamos que nossa discussão pode beneficiar esforços futuros quanto à eficiência de degradação do PET em condições ambientais.

**Palavras-chave:** Reciclagem Biológica. IsPETase. AMBER. Docking Molecular. Dinâmica Molecular.

# Abstract

Plastic waste causes serious environmental problems worldwide, and biological recycling is one of the most eco-friendly methods to deal with it. The discovery of *Ideonella sakaiensis* PET hydrolase (IsPETase) attracted the interest of the scientific community. Its mesophilic nature is an interesting feature for new perspectives in polyethylene terephthalate (PET) degradation research. Here, we use a Molecular Dynamics approach and discuss the possibility of variations of the canonical catalytic triad at pH 7 and 300 K. We prepare three variants and perform a comparative Molecular Docking protocol of BHET, – *bis(2-hydroxyethyl) terephthalate* – with these three variants and the wildtype. We performed molecular dynamics simulations focusing on the nucleophilic distance and protonation states of the chosen residues in our mutations. During this work, we noticed the possibility of automating the use of the computational packages *AMBER/AMBERtools*. A program called ASM.py, from the term *AMBER Simulation Manager*, was developed for this purpose and registered under the license GNU GPL at the National Institute of Industrial Property (INPI). In all simulations, the ligand remained bound to the active site. The total binding energy in the variants decreased but always remained favorable for interaction. When analyzing the interaction contribution of each residue, the simulations with the wildtype and D206E showed a more stable connection with basically the same energy pattern. However, the variants with lysine showed a reduction in the unfavorable contributions – while the D206E/H237K showed an intermediate behavior when compared to the single mutants. The new charge distribution and side chains conformation for both lysine variants maintained the nucleophilic distance while pushing the ligand away from the oxyanion gap. The one-residue mutants studied here did not show promising results. However, the peculiar behavior of D206E/H237K provokes curiosity about what consequences might arise in empirical data. We believe our discussion may benefit future efforts regarding PET degradation efficiency under environmental conditions.

**Keywords:** Biological Recycling. IsPETase. AMBER. Molecular Docking. Molecular Dynamics.

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
<b>2</b>	<b>OBJETIVOS</b>	<b>19</b>
2.1	Objetivo geral	19
2.2	Objetivo específico	19
<b>3</b>	<b>METODOLOGIA</b>	<b>20</b>
3.1	Arquivo de coordenadas da IsPETase	20
3.2	O programa ASM.py	21
3.3	Notação	21
3.4	Variantes da tríade catalítica canônica	22
3.5	Simulação de Dinâmica Molecular	23
3.5.1	Critérios adotados nas dinâmicas moleculares	23
3.6	Docking molecular	24
3.7	Cálculos de decomposição de energia (MMPBSA)	25
<b>4</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>26</b>
4.1	Gerenciador de dados ASM.py	26
4.1.1	Dinâmica molecular utilizando o AMBER	26
4.1.2	Automação do uso das ferramentas AMBER	28
4.1.3	Validação do ASM.py	31
4.1.3.1	IsPETase nativa	32
4.1.3.2	IsPETase <sup>D206E</sup>	33
4.2	Variações da tríade catalítica canônica da IsPETase e interação com o Monômero Central do PET	35
4.2.1	Docking molecular	35
4.2.2	Simulações de dinâmica molecular	37
4.2.3	Decomposição da energia de interação	39
<b>5</b>	<b>CONCLUSÕES</b>	<b>43</b>
<b>A</b>	<b>TEORIA BASE DE DINÂMICA MOLECULAR</b>	<b>46</b>
A.1	Mecânica molecular: potenciais de energia	46
A.2	Potencial de Lennard-Jones	47
A.3	Campos de força	48
A.3.1	Campos de força implementados no AMBER	49
A.4	Efeitos de solvatação	49

A.4.1	Equação de Poisson-Boltzmann . . . . .	50
A.4.2	Modelos para água - TIP3P e TIP4P . . . . .	51
A.4.3	Modelo Generalizado de Solvatação de Born . . . . .	52
<b>A.5</b>	<b>Dinâmica molecular a <i>pH</i> constante . . . . .</b>	<b>54</b>
A.5.1	Método de Mongan - dinâmica a <i>pH</i> constante com solvente implícito . . . . .	55
A.5.2	Método de Swails - dinâmica a <i>pH</i> constante com solvente Explícito . . . . .	56
A.5.2.1	Fluxograma do método de Swails . . . . .	57
<b>A.6</b>	<b>Método de Metropolis-Hastings: Uma abordagem Monte Carlo com cadeia de Markov . . . . .</b>	<b>58</b>
A.6.1	Exemplo de amostragem Metropolis-Hastings usando Python . . . . .	59
<b>A.7</b>	<b>MMPBSA . . . . .</b>	<b>63</b>
A.7.1	Decomposição da Energia Livre . . . . .	64
A.7.2	Cálculos de entropia . . . . .	64
<b>A.8</b>	<b>Arquivo PDB . . . . .</b>	<b>65</b>
<b>A.9</b>	<b>Ferramentas AMBER . . . . .</b>	<b>65</b>
<b>B</b>	<b>CONCEITOS BÁSICOS SOBRE ENZIMAS E O MODELO DE GE . . . . .</b>	<b>67</b>
B.1	<i>α</i> -aminoácidos . . . . .	67
B.2	A estrutura primária dos polipeptídeos . . . . .	69
B.3	Estruturas secundárias encontradas em proteínas . . . . .	69
B.4	A estrutura terciária dobrada das proteínas . . . . .	71
B.5	Função catalítica . . . . .	71
B.6	Um modelo termodinâmico para catálise . . . . .	71
B.7	Catálise ácida e básica em reações enzimáticas . . . . .	72
B.8	Catálise nucleofílica em reações enzimáticas . . . . .	74
B.9	Modelo de Ge: o mecanismo de reação da IsPETase . . . . .	75
<b>C</b>	<b>MATERIAL SUPLEMENTAR . . . . .</b>	<b>79</b>
C.1	<i>Docking</i> molecular: Descrição detalhada . . . . .	79
C.2	Decomposição da energia (MMPBSA) . . . . .	81
	<b>REFERÊNCIAS . . . . .</b>	<b>85</b>

# Lista de figuras

Figura 1 – Descarte indevido de PET. Foto de Stijn Dijkstra. . . . .	15
Figura 2 – Visão geral do ciclo de reciclagem plástica. Imagem adaptada do site <a href="#">School of Global Environmental Sustainability</a> . . . . .	16
Figura 3 – Esquema da degradação do PET pelas enzimas da <i>Ideonella sakaiensis</i> . Adaptado do artigo (Austin, 2018). . . . .	17
Figura 4 – Estrutura molecular da IsPETase (PDB ID 6EQE). (A) Representação 3D da proteína destacando os principais resíduos do sítio ativo. (B) Conformação das cadeias laterais dos resíduos destacados em A. . . . .	20
Figura 5 – Representação esquemática da automação realizada pelo programa ASM.py. . . . .	21
Figura 6 – Ilustração esquemática do modelo de quatro estados para o mecanismo de reação da IsPETase, proposto por Ge. Adaptado do artigo (Ge, 2021). . . . .	22
Figura 7 – Modelo entidade-relacionamento para as dinâmicas moleculares: Atomística clássica (DM) e CpHMD. Modelo construído a partir dos processos de dinâmica pelos pacotes <i>AMBER/AMBERtools</i> . . . . .	26
Figura 8 – Fluxograma sobre a preparação de dados do sistema, para simulações com os pacotes <i>AMBER/AMBERtools</i> . . . . .	27
Figura 9 – Modelo entidade-relacionamento para a implementação do ASM.py. . . . .	29
Figura 10 – Primeiros trinta aminoácidos da IsPETase nativa, indicando a região terminal-N. . . . .	31
Figura 11 – Análise comparativa entre as dinâmicas NativaS1 e NativaS2, referente ao: (A) Desvio quadrático médio do “backbone”; e ao (B) Raio de giro da enzima. . . . .	32
Figura 12 – Análise comparativa da estabilidade estrutural da IsPETase nas dinâmicas NativaS1 e NativaS2. (A) Flutuação quadrática média por resíduo do “backbone”. (B) Visão da proteína, destacando os resíduos com picos no RMSF. . . . .	33
Figura 13 – Análise comparativa entre as dinâmicas da IsPETase e de sua variante IsPETase <sup>D206E</sup> . (A) Desvio quadrático médio do “backbone”; e (B) Raio de giro das proteínas. . . . .	34
Figura 14 – Flutuação quadrática média por resíduo do “backbone” na IsPETase e em sua variante IsPETase <sup>D206E</sup> . . . . .	34

Figura 15 – Ligante BHET (destacado em azul) encaixado em cada variante IsPETase: (A) Nativa, (B) D206E, (C) D206E/H237K e (D) H237K. Neste esquema 2D, as linhas tracejadas representam ligações de hidrogênio e os rótulos de resíduos verdes são os aminoácidos com contatos hidrofóbicos com o ligante. . . . .	36
Figura 16 – Distribuição de frequência normalizada das distâncias nucleofílicas $r_{hc}$ . Destacado em azul a região de interesse. Os dados para cada sistema Variante-BHET foram obtidos da etapa de produção em pH 7, 300 K e duração de 200 ns. . . . .	37
Figura 17 – Configuração dos estados de protonação para os resíduos 206 e 237 das variantes da IsPETase. A presença dos hidrogênios marcados em vermelho depende do estado atual do resíduo. . . . .	39
Figura 18 – Os aminoácidos que contribuem para a interação IsPETase-BHET em pH 7 e 300 K. As linhas verticais destacam a tríade catalítica SER 160, ASP 206, HIS 237 e suas respectivas mutações. . . . .	40
Figura 19 – Visão comparativa do ligante BHET no sítio ativo da IsPETase, entre a nativa e as variantes com a lisina. As imagens foram tiradas da fase de produção da dinâmica molecular de cada sistema. . . . .	42
Figura 20 – Isomerismo conformacional. . . . .	47
Figura 21 – Exemplo de potencial de Lennard-Jones. . . . .	47
Figura 22 – Classe <i>Sampling</i> (“Amostragem”) escrita de maneira a criar um gráfico de amostragem aleatória, quando um objeto é instanciado. . . . .	60
Figura 23 – Amostragem aleatória usual. . . . .	60
Figura 24 – Definição da função distribuição de probabilidade. . . . .	60
Figura 25 – Proposta para a nova amostra usando a propriedade da cadeia de Markov. . . . .	61
Figura 26 – Método de amostragem de Metropolis. . . . .	61
Figura 27 – Geração dos gráficos: Amostra vs Índice da amostragem. Escrito de maneira a apresentar as amostragem aleatória usual comparada a de Metropolis. . . . .	62
Figura 28 – Comparação entre a amostragem aleatória comum (acima) e a de Metropolis (abaixo), ambas extraindo trezentos pontos. . . . .	62
Figura 29 – Ciclos termodinâmicos para cálculos de energia livre de ligação para um complexo proteína-ligante. Os sistemas solvatados estão representados em caixas azuis, enquanto os sistemas na fase gasosa estão em caixas brancas. As energias livres calculadas diretamente são mostradas em preto, enquanto a energia livre de interesse é mostrada em vermelho. Adaptado do artigo (Miller, 2012). . . . .	63
Figura 30 – Estrutura geral de $\alpha$ -aminoácidos L e D. . . . .	67
Figura 31 – As cadeias laterais dos vinte aminoácidos encontrados em proteínas. . . . .	68

Figura 32 – N-terminal de uma cadeia polipeptídica. . . . .	69
Figura 33 – Estrutura de uma $\alpha$ -hélice, com as posições dos carbonos alfa indicadas por pontos. . . . .	70
Figura 34 – Estrutura das folhas beta. . . . .	70
Figura 35 – Figura esquemática da (a) enzima e o substrato e (b) enzima, substrato e cofator. . . . .	72
Figura 36 – Perfis de energia livre para (a) uma reação catalisada por ácido e (b) uma reação catalisada por enzima que converte o substrato S no produto P. Adaptado do livro (Bugg, 2004). . . . .	73
Figura 37 – Catálise ácida. Adaptado do livro (Bugg, 2004). . . . .	73
Figura 38 – Catálise básica. Retirado do livro (Bugg, 2004). . . . .	74
Figura 39 – Faixa de valores de $pK_a$ observados para cadeias laterais de aminoácidos em sítios ativos de enzimas. Adaptado do livro (Bugg, 2004). . . . .	74
Figura 40 – Representação 3D do local de confinamento do substrato (BHET). Os aminoácidos principais são: a tríade catalítica S-H-D (em laranja), a cavidade de oxiânion formada por Y87 e M161 (em azul) e a interação $\pi - \pi$ de W185 com o anel de benzeno do BHET (em rosa). Adaptado do artigo (Ge, 2021). . . . .	76
Figura 41 – Esboço do espaço de busca e região de amostragem do ponto central. . . . .	79
Figura 42 – Esquema das duas possibilidades para o ataque nucleofílico nos complexos Enzima-BHET. . . . .	80
Figura 43 – Arquivo de conformação do ligante gerado ao utilizar o <i>AutoDock Vina</i> . . . . .	81
Figura 44 – Algoritmo de ordenação e lista resultante. Imagens geradas a partir do programa <i>AvR_DockingPlot.py</i> . . . . .	82
Figura 45 – Distância nucleofílica durante a etapa de produção, mantendo pH 7 e 300 K. Comparação entre a DM do complexo IsPETase <sup>D206E/H237K</sup> -PET e de uma DM replicata. Destacado em vermelho a região mais estável de ambas dinâmicas. . . . .	83
Figura 46 – Decomposição por resíduo da energia de interação Enzima-BHET, relativa a faixa de 75 a 100 ns destacada na Fig. 45. Comparação entre a DM da variante IsPETase <sup>D206E/H237K</sup> e de uma DM replicata. . . . .	84
Figura 47 – Visão da conformação dos resíduos do sítio ativo da variante IsPETase <sup>D206E/H237K</sup> , no estado a 100 ns da produção. . . . .	84

# Lista de tabelas

Tabela 1	– Primeiras cinco conformações de encaixe da lista ordenada baseada em $s_{v+r}$ para IsPETase e suas variantes. . . . .	35
Tabela 2	– Estados de protonação $S_i$ dos resíduos de acoplamento da IsPETase e suas variantes. O valor entre parênteses indica a permanência de cada estado no decorrer dos 200 ns de simulação. . . . .	38
Tabela 3	– Valores médios para a distância relativa ao ataque nucleofílico, a frequência que esta distância permaneceu inferior a 5 Å, e a energia de interação, calculadas nos complexos IsPETase-BHET. . . . .	40
Tabela 4	– Adaptado de (Mongan, 2004): Negativo do logaritmo das constantes de dissociação de ácido para as cadeias laterais nas moléculas de aspartato, glutamato, histidina, lisina e tirosina. . . . .	56
Tabela 5	– Cadeias laterais de aminoácidos usadas para catálise nucleofílica em reações enzimáticas. Adaptado do livro (Bugg, 2004) . . . . .	75
Tabela 6	– Valores médios relativos a decomposição de energia por resíduo e seus desvios padrões, em $kcal/mol$ . Resíduos com as contribuições mais relevantes dos complexos IsPETase-BHET. . . . .	82
Tabela 7	– Valores médios relativos a decomposição de energia por resíduo e seus desvios padrões, em $kcal/mol$ . Valores calculados utilizando os dados de 75 a 100 ns de dinâmica com a variante $D206E/H237K$ e sua replicata. . . . .	83

# Lista de Abreviaturas e Siglas

**ASM.py** AMBER Simulation Manager

**BHET** Bis(2-Hydroxyethyl) Terephthalate

**CAPES** Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil

**CpHMD** Constant pH Molecular Dynamics

**DM** Dinâmica Molecular

**DMSGB** Dinâmica Molecular com Solvatação Generalizada de Born

**EG** Ethylene glycol

**EPB** Equação de Poisson-Boltzmann (Eq. 11)

**FDP** Função de densidade de probabilidade

**FsC** *Fusarium solani pisi* cutinase

**GAFF** General Amber Force Field

**GB** Generalizado de Born

**GNU GPL** GNU General Public License

**GPU** Graphics Processing Unit

**HMR** Hydrogen Mass Repartitioning

**ILE** Isoleucina

**INPI** Instituto Nacional da Propriedade Industrial

**IsPETase** Ideonella Sakaiensis PET hidrolase

**LEU** Leucina

**MC** Monte Carlo

**MCCM** Monte Carlo com Cadeia de Markov

**MET** Metionina

**MMPBSA** Molecular Mechanics Poisson-Boltzmann Surface Area

**MPI** Message Passing Interface

**MTP** Multiple Trajectory Protocol

**NativaS1** Dados de produção com a IsPETase Nativa e Simulação de 200 ns.

**NativaS2** Dados de produção com a IsPETase Nativa e Simulação de 200 ns (replicata).

**NMR** Nuclear Magnetic Resonance

**PDB** Protein Data Bank

**PET** Politereftalato de Etileno

**PHE** Fenilalanina

**QM/MM** Quantum Mechanics/Molecular Mechanics

$r_{hc}$  Distância nucleofílica, entre uma das carbonilas do BHET e o oxigênio hidroxila da SER 160 das variantes da IsPETase.

**RMSD** Root Mean Square Deviation

**RMSF** Root Mean Square Flutuation

**SDF** Structure Data File

**STP** Single Trajectory Protocol

**TPA** Terephthalic acid

**TRP** Triptofano

**UFMS** Universidade Federal de Mato Grosso do Sul

# 1 Introdução

Um dos polímeros sintéticos mais aplicados industrialmente, o Politereftalato de Etileno (PET), é composto de ligações do tipo éster entre ácido tereftálico (TPA) e etilenoglicol (EG). Ele é um material muito utilizado na produção de garrafas plásticas e tecidos devido a sua transparência, maleabilidade e resistência a processos naturais de degradação. Sua resistência pode ser uma qualidade para a indústria, mas é um defeito em relação ao seu descarte (veja a Fig. 1). Estima-se que a quantidade total de resíduos plásticos corresponda à produção do material nas centenas de milhões de toneladas, com uma estimativa de um mínimo de 4,8 milhões de toneladas entrando nos oceanos em 2010 [1].



Figura 1 – Descarte indevido de PET. Foto de Stijn Dijkstra.

Existem vários métodos para lidar com esse problema do plástico, e podemos classifica-los em reciclagem primária, mecânica, química e biológica (a Fig. 2 apresenta uma representação esquemática do ciclo de reciclagem). A reciclagem primária utiliza resíduos limpos e não contaminados ou mistura-os com material virgem [2]. Na reciclagem mecânica, o polímero é separado de contaminantes e reprocessado. Nessa abordagem, as propriedades do produto se deterioram a cada ciclo de reciclagem. A reciclagem química envolve a transformação das cadeias poliméricas em unidades monoméricas ou sua quebra aleatória em fragmentos de cadeia com a formação associada de produtos gasosos. Os tratamentos de reciclagem química podem ser considerados os mais eficientes, com quase 100% de reaproveitamento do material, mas esses podem gerar resíduos tóxicos ou exigir condições de temperatura e pressão muito altas [3].

Quanto a reciclagem biológica, as opções adequadas de plástico a este processo

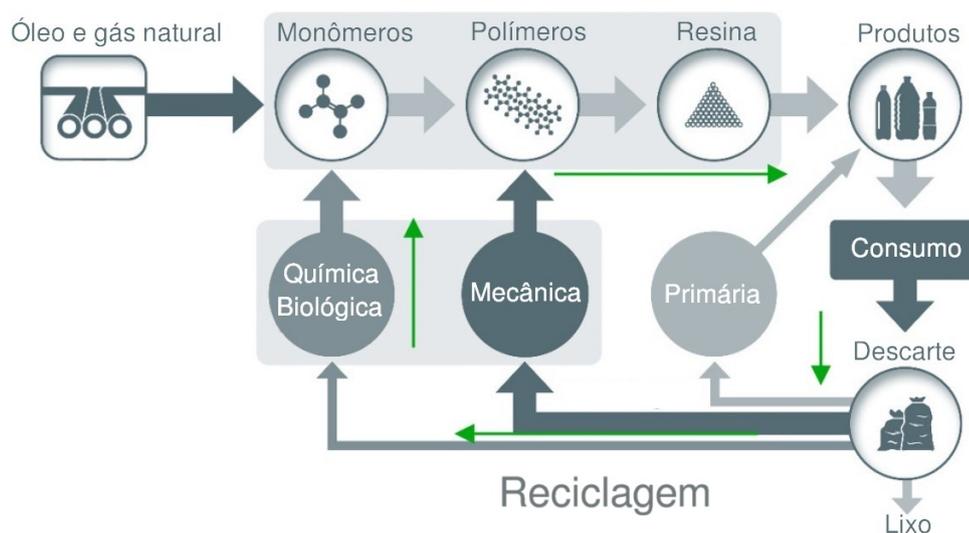


Figura 2 – Visão geral do ciclo de reciclagem plástica. Imagem adaptada do site [School of Global Environmental Sustainability](#).

são o PET e o poliuretano, pois possuem estruturas químicas hidrolisáveis. No entanto, o conteúdo cristalino rígido do PET reduz sua acessibilidade à hidrólise enzimática. Curiosamente, contudo, em temperaturas superiores a 70 °C (ou cerca de dez graus mais baixas na água), o polímero muda de um estado semelhante ao vidro rígido para uma forma elástica e torna-se significativamente mais suscetível a ataques enzimáticos [4]. No entanto, Yoshida et al. descobriram a *Ideonella sakaiensis*, uma bactéria mesófila que pode crescer em um filme PET amorfo a 30 °C [5]. Os autores identificaram as proteínas responsáveis pela degradação: a IsPETase (como foi posteriormente denominada pela comunidade científica) e a MHETase. Como ilustrado na Fig. 3, a IsPETase quebra o polímero nos fragmentos Bis(2-Hydroxyethyl) Terephthalate (BHET), Mono(2-hydroxyethyl) Terephthalate (MHET) e TPA – a IsPETase hidrolisa o BHET gerando MHET sem posterior decomposição. Em seguida a MHETase converte o MHET em TPA e EG. Eles determinaram que a atividade da IsPETase era ótima em pH 9 e com a temperatura em torno de 40 °C, em contraste com as PETases termofílicas previamente conhecidas. Na IsPETase, a tríade catalítica é composta pelos resíduos SER 160, ASP 206 e a HIS 237, com um sistema semelhante a outras  $\alpha/\beta$ -hidrolases. Em comum com a maioria das lipases, os resíduos catalíticos residem em laços, com a serina nucleofílica ocupando uma posição conhecida como *nucleophilic elbow* (o nucleófilo reside em um laço entre uma  $\alpha$ -hélice e uma  $\beta$ -folha)[6].

Muitos pesquisadores argumentam que a IsPETase tem a maior eficiência entre as hidrolases de PET, mas muitos outros refutam esse argumento. Por exemplo, Wei et al. compararam a perda de massa de um filme de PET amorfo devido à hidrólise enzimática pela IsPETase e a LC-cutinase, uma contraparte termofílica homóloga [7]. Utilizaram a IsPETase a 30 °C e a LC-cutinase a 70 °C, os resultados de Wei indicam uma perda de massa do polímero mais de 40 vezes maior alcançada pela LC-cutinase após 24 horas de incubação. Tournier et al. compararam a IsPETase e a *Fusarium solani pisi* cutinase (FsC)

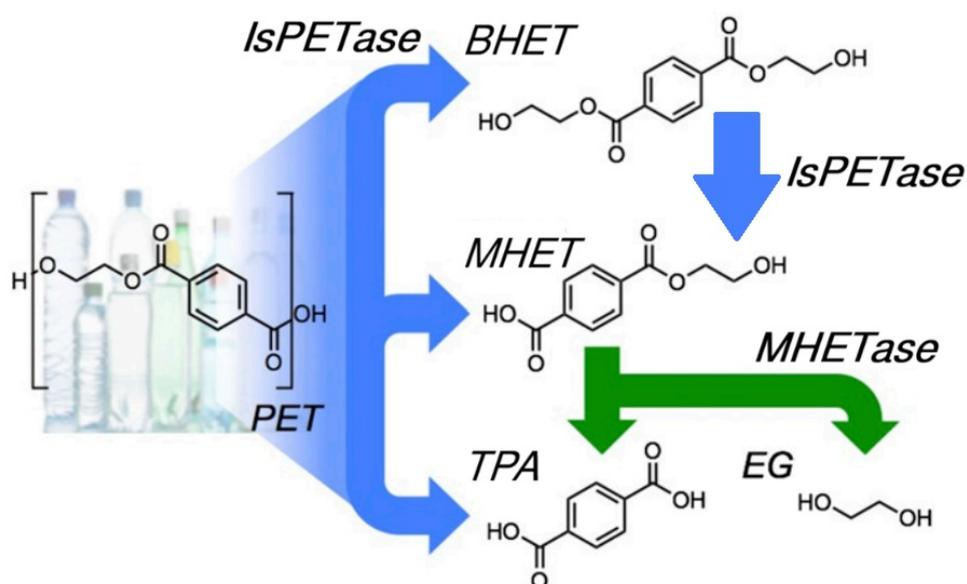


Figura 3 – Esquema da degradação do PET pelas enzimas da *Ideonella sakaiensis*. Adaptado do artigo (Austin, 2018).

a 40 °C e pH 9, indicando atividade enzimática cerca de dez vezes maior na FsC [8]. Em suas respectivas condições ótimas de atividade, a *IsPETase* é inferior a outras enzimas bem conhecidas. No entanto, sua natureza mesófila pode ser vantajosa para abordagens mais ecológicas.

As serino proteases compreendem quase um terço de todas as proteases conhecidas identificadas até o momento. As serino proteases mais difundidas na natureza possuem a tríade catalítica SER/HIS/ASP, assim como a *IsPETase*. Entretanto, existe uma grande variedade de configurações de sítios ativos, onde a serina pode ser substituída por um resíduo de treonina; a histidina pode ser substituída pelo ácido glutâmico ou a lisina; e o ácido aspártico pode ser substituído pelo ácido glutâmico, um resíduo de histidina, ou em alguns casos completamente eliminado [9]. Os diferentes arranjos permitem que as proteases funcionem em diferentes ambientes, como funcionando de maneira ideal em pH mais baixo. Assim, com inspiração nas possibilidades atípicas das serino proteases, buscamos fornecer informações com um protocolo de dinâmica molecular (DM) sobre variações da tríade canônica da *IsPETase*. Os métodos de docking molecular e DM têm sido amplamente aplicados pela comunidade científica para investigar os mecanismos de ligação de ligantes com enzimas [10, 11, 12].

No capítulo 3 desta tese de doutorado será apresentado o sistema estudado e o procedimento usado quanto as dinâmicas moleculares realizadas. As discussões dos resultados serão feitas no capítulo 4 e a conclusão no 5. Nos Apêndices A e B, serão apresentados brevemente alguns modelos e conceitos importantes para a compreensão dos processos de dinâmica molecular, além de alguns métodos computacionais.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## 2 Objetivos

### 2.1 Objetivo geral

Estudo da interação do PET com a enzima IsPETase, por meio das técnicas de *Docking* e Dinâmica Molecular.

### 2.2 Objetivo específico

Propor variantes da IsPETase buscando melhorar a interação com o PET a pH 7 e 300 K.

## 3 Metodologia

### 3.1 Arquivo de coordenadas da IsPETase

A estrutura molecular que utilizamos da IsPETase foi determinada por Austin et al. [6] por meio de cristalografia de raios X a  $pH$  6 e 290 K. Esta estrutura pode ser obtida como um arquivo tipo *Protein Data Bank (PDB)* pela identificação [PDB:6EQE](#) e possui ótima resolução<sup>1</sup>, com valor de 0,92 Å – veja a visualização 3D na Fig. 4 produzida com o uso do programa UCSF Chimera [13].

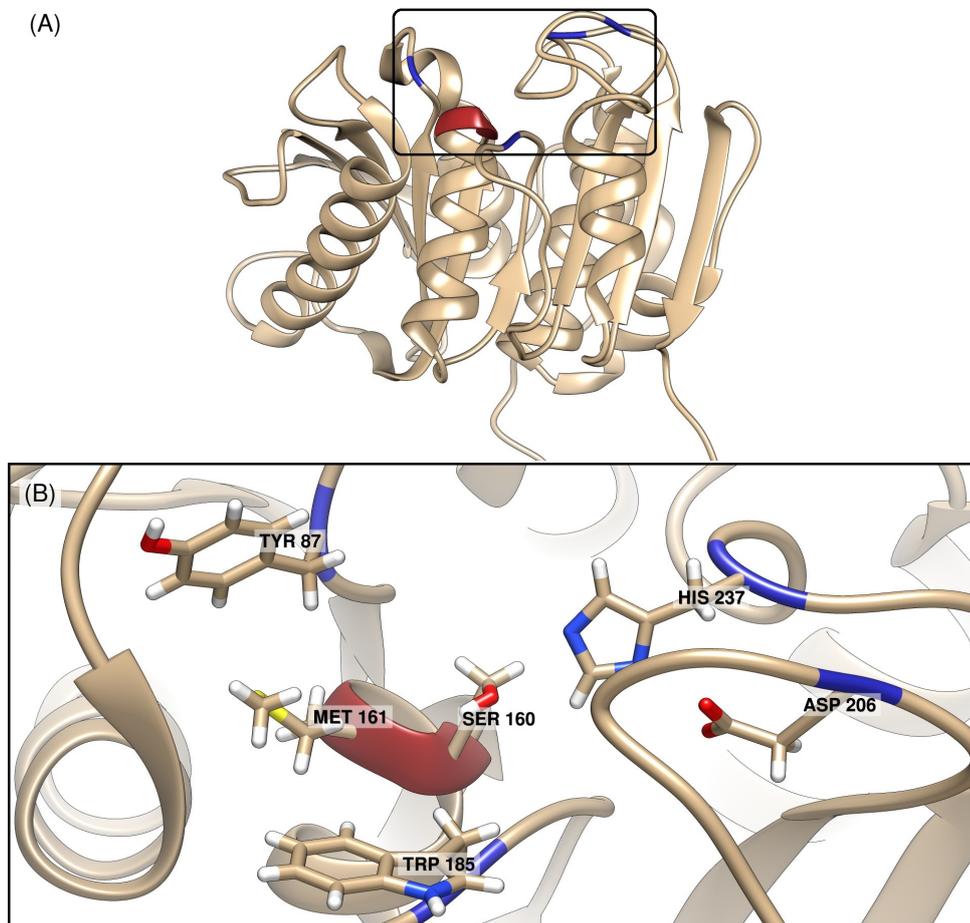


Figura 4 – Estrutura molecular da IsPETase (PDB ID 6EQE). (A) Representação 3D da proteína destacando os principais resíduos do sítio ativo. (B) Conformação das cadeias laterais dos resíduos destacados em A.

<sup>1</sup> Veja o apêndice A.8.

## 3.2 O programa ASM.py

Neste trabalho de doutorado todas as simulações moleculares foram realizadas usando o ASM.py (Fig. 5), um gerenciador de dados criado durante o doutorado e descrito na seção 4.1.2. Trata-se de um intermédio que agiliza os processos de dinâmica em proteínas por meio dos pacotes computacionais *AMBER/AMBERtools*. Uma vantagem desse programa está na facilidade de reprodução dos resultados obtidos neste trabalho. O código fonte do programa foi registrado com licença GNU GPL e pode ser encontrado no [GitHub](#).

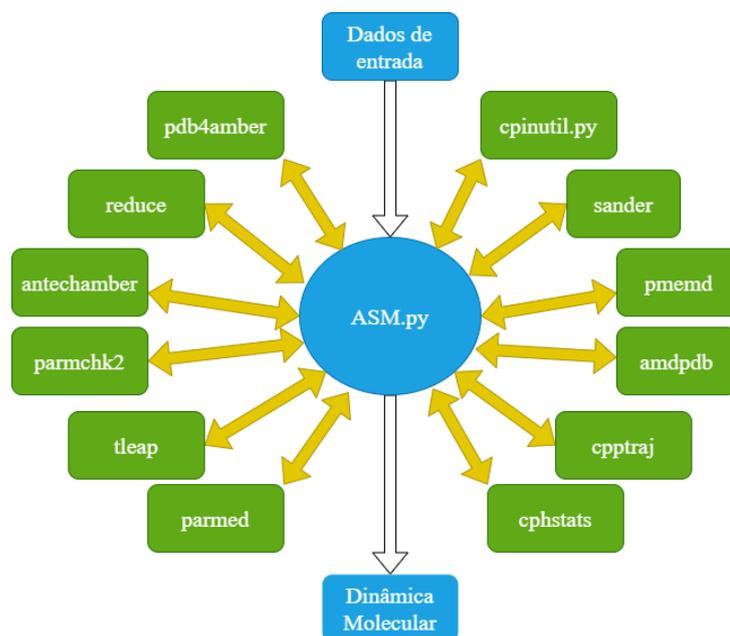


Figura 5 – Representação esquemática da automação realizada pelo programa ASM.py.

## 3.3 Notação

Para a determinação estrutural da IsPETase (referente ao *6EQE*), a sequência de peptídeo inicial (1 – 28) é removida para a produção do domínio central da proteína. Um domínio de proteína é uma região da cadeia polipeptídica que se auto-estabiliza e se dobra independentemente do resto, formando uma estrutura tridimensional compacta. Muitas proteínas consistem em vários domínios. Em geral, os domínios variam de tamanho entre cerca de 50 e 250 aminoácidos. Como consequência dessa numeração a partir do 29, as ferramentas do AMBER renumeram os resíduos do sistema na etapa de preparação (pré minimização) e as “IDs” são reduzidas em 28 unidades cada. Para facilitar a discussão, será usado a ID do arquivo *6EQE* na nomeação da mutação e discutido quando necessário com a nova numeração. A notação adotada para identificação de uma mutação seguirá o

formato usado na literatura:



A exemplo, a mutação da histidina da tríade por uma alanina seria nomeada como IsPETase<sup>H237A</sup>.

### 3.4 Variantes da tríade catalítica canônica

Para melhorar a reação enzimática da IsPETase, devemos interferir no mecanismo de reação da enzima sem causar a interrupção de sua atividade. Como mencionado no final da Sec. 1, em configurações não canônicas para serino proteases, a serina pode ser substituída por uma treonina, a histidina pode ser substituída por ácido glutâmico ou lisina e o ácido aspártico pode ser substituído por ácido glutâmico ou uma histidina. Quanto a nossas propostas de variante, não alteramos o nucleófilo da IsPETase (SER 160) e focamos nos resíduos de acoplamento (ASP 206 e HIS 237). O impacto mais significativo esperado no sítio ativo da IsPETase foi devido à remoção do imidazol da histidina, por isso estudamos as variações: D206E, H237K e D206E/H237K. De acordo com o modelo de Ge (cuja representação esquemática está ilustrada na Fig. 6) para o mecanismo de reação da IsPETase [14], os resultados de cálculos mecânico quânticos de Vyas et al. [15], e muitos outros na literatura, as distâncias entre o nucleófilo e o eletrófilo que são favoráveis a ataques nucleofílicos são próximas a 4 Å. Nesta tese, denotaremos essa distância nucleofílica (entre uma carbonila do BHET e o oxigênio da hidroxila da SER 160) como  $r_{hc}$ .

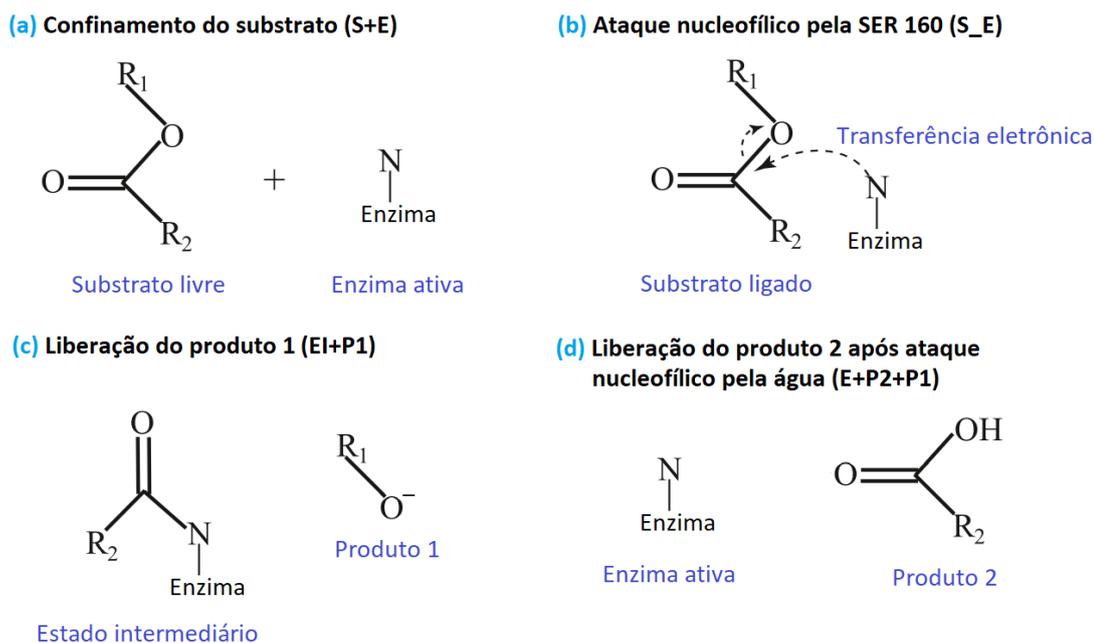


Figura 6 – Ilustração esquemática do modelo de quatro estados para o mecanismo de reação da IsPETase, proposto por Ge. Adaptado do artigo (Ge, 2021).

## 3.5 Simulação de Dinâmica Molecular

A simulação de dinâmica molecular é uma técnica de amostragem para sistemas clássicos de muitos corpos, pois com ela não é explorado todo o espaço de fase, além de depender de conformações aleatórias como estado inicial [16]. Ela possui uma abordagem semelhante a experimentos reais, com etapas de preparação e medição. No lugar de preparar uma amostra do material de estudo é selecionado um sistema modelo consistindo de  $N$  partículas. Para medir uma grandeza observável em simulações de dinâmica molecular, é necessário expressá-la em função das posições e momentos das partículas no sistema. Por exemplo: uma definição conveniente de temperatura em um sistema clássico de muitos corpos, faz uso da equipartição de energia sobre todos os graus de liberdade que entram quadraticamente no Hamiltoniano do sistema. Em particular a energia cinética média por grau de liberdade, pode ser calculada como  $\langle \frac{1}{2}mv^2 \rangle = \frac{1}{2}k_B T$ , onde  $k_B$  é a constante de Boltzmann e  $T$  a temperatura do sistema. Na prática, a energia cinética total medida seria dividida pelo número de graus de liberdade  $n_{lib}$ . À medida que a energia cinética total de um sistema flutua, o mesmo ocorre com a temperatura instantânea. As flutuações relativas serão da ordem  $1/\sqrt{n_{lib}}$ . Como  $n_{lib}$  é normalmente da ordem de  $10^2$  a  $10^3$ , as flutuações estatísticas da temperatura são da ordem de 5% a 10% [16].

### 3.5.1 Critérios adotados nas dinâmicas moleculares

Os parâmetros gerais de dinâmica utilizados, são os mesmos para as simulações com a proteína isolada do PET e do complexo enzima-substrato pós *docking* molecular. As dinâmicas moleculares realizadas seguiram cinco etapas: a preparação do sistema onde são, por exemplo, implementadas as mutações, os campos de força e o solvente; e estágios de minimização, aquecimento, termalização e produção, como descritos abaixo:

1. *Minimização*: Uma vez que a geometria pode não corresponder aos mínimos reais no campo de força usado e pode também resultar em conflitos e sobreposições de átomos em outros resíduos, é sempre uma boa prática minimizar a energia do sistema antes de iniciar uma dinâmica molecular. A falha em minimização pode levar a instabilidades durante a dinâmica.
2. *Aquecimento*: É realizado um aquecimento gradual do sistema e após atingida a temperatura final, mantém-se o estado com pouca flutuação de temperatura até o fim da dinâmica. Para essa regulação foi utilizado o termostato de Langevin [17, 18].
3. *Termalização*: Estágio popularmente conhecido pelo termo equivocado “equilíbrio”. Mantém o sistema sob condições restritivas semelhantes as do aquecimento, mas usualmente em outro *ensemble* estatístico. O objetivo deste estágio é alcançar um estado no qual as propriedades de interesse do sistema permaneçam estacionárias.

4. *Produção*: Etapa similar a termalização, porém sem as restrições na estrutura usualmente aplicadas na termalização. Esta etapa pode durar tanto quanto se queira, podendo alcançar desde algumas centenas de nanossegundos até alguns milissegundos de simulação (pode levar meses de processamento dependendo da máquina utilizada).

Realizamos todas as simulações de dinâmica molecular usando o módulo *pmemd* do AMBER 18 [19]. Cada sistema foi colocado em uma caixa octaédrica, com água TIP3P [20] até as limitações da caixa, 12 Å de distância dos átomos do soluto e neutralizada com íons  $Cl^-$  e  $Na^+$ . Definimos o raio de corte para interações não ligadas em 12 Å, de modo a ficar compatível com o contorno da caixa octaédrica e restringimos o comprimento das ligações que incluem hidrogênio com o algoritmo SHAKE [21]. Usamos 4 fs para o passo de tempo, enquanto utilizamos o método *Hydrogen Mass Repartitioning (HMR)* [22]. A interação proteína-átomo foi descrita pelo campo de força ff14SB [23], enquanto o ligante foi descrito usando o *General Amber Force Field (GAFF)* [24]. Como nosso foco era estudar a possibilidade de variantes mais eficientes para a tríade canônica da IsPETase em pH 7 e 300 K, utilizamos o método *Constant pH Molecular Dynamics (CpHMD)* [25] para permitir transições dos estados de protonação/desprotonação - de acordo com o critério de Metropolis [26] - dos resíduos de acoplamento ASP 206 e HIS 237 na enzima nativa e em suas respectivas mutações na IsPETase<sup>D206E</sup>, IsPETase<sup>D206E/H237K</sup> e IsPETase<sup>H237K</sup>.

Todas as simulações de dinâmica molecular foram realizadas em replicatas. Essas replicatas são simulações com os mesmos parâmetros de entrada da original, e que são feitos os mesmos procedimentos desde a preparação até a etapa de produção. Elas são importantes pois o método de dinâmica molecular é baseado em uma amostragem aleatória (veja o Apêndice A), e portanto o sistema pode explorar conformações distintas nas replicatas. Em cada simulação, a minimização começa com 2000 iterações do método *steepest descent* e prossegue por mais 8000 iterações no método *conjugate gradient*. Nós favorecemos o segundo método por seu baixo custo computacional – veja a referência [27] que apresenta uma comparação entre eles. O aquecimento de cada replicata foi realizado com 4 ns de duração, no *ensemble* NVT, de 10K a 300K com uma taxa média de 0,48 K/ps. A termalização dura 40 ns a pressão constante de 1 bar (*ensemble* NPT). E finalmente, a etapa de produção de cada replicata teve duração de 200 ns no *ensemble* NVT mantendo pH 7. Extraordinariamente, nos testes de validação do programa ASM.py, na etapa de produção utilizamos dois valores de pH (7 e 9).

## 3.6 Docking molecular

A molécula do BHET foi utilizada como o substrato para estudar a interação PET-IsPETase. Esta escolha foi devida a duas razões: a IsPETase também interage com o BHET; além de que esta molécula foi utilizada para estudar a IsPETase pela comunidade

científica, possivelmente pela maior facilidade com relação ao método de *Docking* usando um substrato pequeno. Este monômero pode ser encontrado no formato *Structure Data File (SDF)* no site do PDB pelo identificador [C8X](#). Realizamos a ligação molecular do BHET com cada enzima usando o AutoDock Vina 1.1.2 [28]. O modelo enzimático foi devidamente mutado, com AMBERtools [19], antes da parametrização exigida pelo Vina. As conformações de encaixe consideradas para as simulações DM foram aquelas com as menores energias de afinidade (escores Vina), que também apresentaram distâncias favoráveis para ataques nucleofílicos.

Realizamos 400 execuções usando o AutoDock Vina para cada sistema variante-BHET. Em cada execução, o tamanho do espaço de busca foi definido como  $20 \text{ \AA} \times 20 \text{ \AA} \times 20 \text{ \AA}$ , e foi centralizado na posição  $(X_i, Y_i, Z_i)$ . As coordenadas do centro foram amostradas aleatoriamente em cada execução seguindo a regra:

$$\begin{aligned} X_i &= \text{random}(OG_x^{S160} - 5 \text{ \AA}, OG_x^{S160} + 5 \text{ \AA}), \\ Y_i &= \text{random}(OG_y^{S160} - 5 \text{ \AA}, OG_y^{S160} + 5 \text{ \AA}), \\ Z_i &= \text{random}(OG_z^{S160} - 5 \text{ \AA}, OG_z^{S160} + 5 \text{ \AA}). \end{aligned}$$

O método  $\text{random}(A, B)$  gera um número aleatório entre A e B. Para isso, utilizamos o módulo *random* da biblioteca padrão do Python 3 [29]. O termo  $OG_{x,y,z}^{S160}$  refere-se às coordenadas do oxigênio do grupo hidroxila da serina 160, que serve como nucleófilo para cada variante. Definimos o nível de exaustividade do programa para 10 em cada execução. Como resultado, obtivemos 3563, 3562, 3534 e 3511 conformações de encaixe do BHET na enzima nativa e nas variantes D206E, H237K e D206E/H237K, respectivamente.

### 3.7 Cálculos de decomposição de energia (MMPBSA)

Os *ensembles* de estado ligado e não ligado foram extraídos usando o protocolo *Multiple Trajectory Protocol (MTP)* e utilizamos o método de decomposição por resíduo. Quanto aos cálculos da contribuição vibracional da entropia utilizamos o método *nmode*, onde foram feitos até 10 000 passos de minimização utilizando 800 instâncias por dinâmica molecular. Em média os cálculos *Molecular Mechanics Poisson-Boltzmann Surface Area (MMPBSA)* feitos usando toda a etapa de produção (200 ns) demoraram cerca de 30 horas por dinâmica molecular. Veja o Apêndice A.7.

## 4 Resultados e discussões

Nesta tese os resultados são divididos em duas partes: uma descrição do programa de computador registrado, e os resultados obtidos das dinâmicas com o monômero do PET. Na primeira parte discutimos brevemente sobre os processos básicos de uso das ferramentas do AMBER para a realização de dinâmicas moleculares (com viés para a implementação do ASM.py), a automação de tais processos e a validação da implementação criada. Na segunda apresento os resultados de *docking* e dinâmica moleculares. Não é discutido a implementação detalhando as funções do programa, mas forneço o código fonte por meio do link [GitHub](#).

### 4.1 Gerenciador de dados ASM.py

#### 4.1.1 Dinâmica molecular utilizando o AMBER

Para compreensão geral de uma simulação de dinâmica molecular, veja o modelo entidade-relacionamento [30] da Fig. 7 que apresenta os atributos de maior relevância nas simulações de DM e *Constant pH Molecular Dynamics (CpHMD)*, utilizando os pacotes *AMBER/AMBERtools*. Para realizar as dinâmicas moleculares precisamos de uma descrição

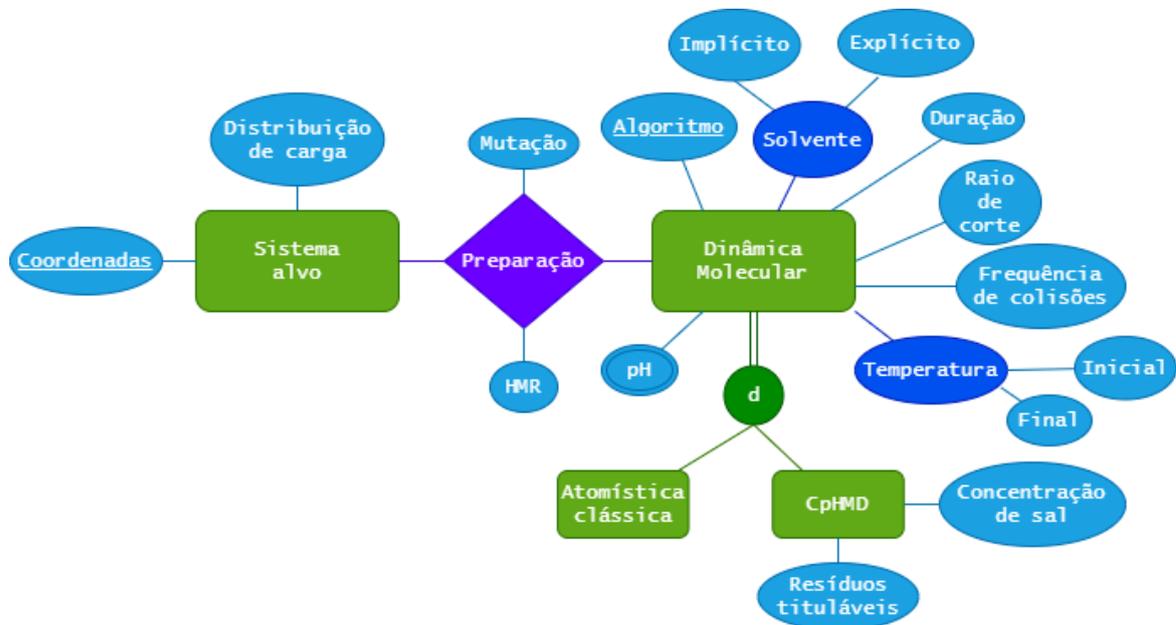


Figura 7 – Modelo entidade-relacionamento para as dinâmicas moleculares: Atomística clássica (DM) e CpHMD. Modelo construído a partir dos processos de dinâmica pelos pacotes *AMBER/AMBERtools*.

detalhada do modelo virtual de nosso sistema alvo. Quando disponível em bancos de dados

internacionais (por exemplo o [RCSB](#)), o modelo compartilhado descreve as coordenadas de cada átomo do sistema, além de algumas informações relevantes do experimento utilizadas para inferência do modelo. Um formato amplamente utilizado na literatura para guardar essas informações é o PDB. Porém, como tantos outros pacotes, o *AMBER/AMBERtools* requer algumas especificidades em seus arquivos, de maneira que é necessário toda uma preparação dos dados do sistema antes que os arquivos necessários estejam prontos. Por meio de cálculos de minimização de energia, os pacotes computacionais sugerem uma distribuição de carga e em seguida os parâmetros necessários para construção de campos de força que representam o sistema. Veja o apêndice A para uma breve descrição dos campos de força utilizados.

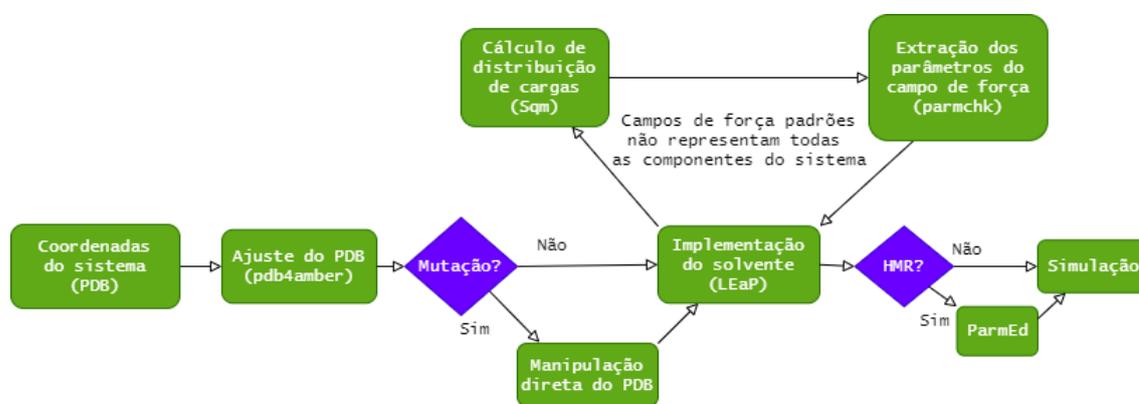


Figura 8 – Fluxograma sobre a preparação de dados do sistema, para simulações com os pacotes *AMBER/AMBERtools*.

A Fig.8 apresenta um fluxograma do processo de preparação dos dados de entrada, para os pacotes *AMBER/AMBERtools*. É necessário ajustar o arquivo de coordenadas do sistema, obtido de bancos de dados internacionais ou construído por meio de editores moleculares, pois alguns nomes de resíduos foram modificados nas bibliotecas do AMBER. Além disso, arquivos de banco de dados normalmente possuem informações “redundantes”<sup>1</sup>. Do PDB ajustado, é realizada ou não uma mutação (edição das coordenadas de um resíduo para substituí-lo por outro) e então são verificadas todas as componentes do arquivo, para ter certeza que os campos de força padrão do AMBER os representam corretamente. Caso as informações das bibliotecas não sejam suficiente, em casos simples, basta calcular a distribuição de cargas do sistema e extrair disto os parâmetros de campo de força a serem usados. Existem situações em que o sistema alvo possui uma espécie que não foi parametrizada e adicionada nas bibliotecas das ferramentas do AMBER, devido a não ocorrência desta espécie em sistemas proteicos comumente estudados por dinâmica molecular. Nessas situações deve-se parametrizar a espécie com ferramentas específicas (como o *Metal Center Parameter Builder MCPB.py* [31]). Uma situação exemplo seria a presença do magnésio na clorofila.

<sup>1</sup> Mais de um tipo de coordenada para certos átomos, para descrever melhor os resultados experimentais.

Com os devidos campos de força, pode-se dar continuidade com a implementação do modelo de solvente (explícito ou implícito). Nesta etapa, nos casos de dinâmica da enzima com o ligante, são acrescentadas as coordenadas do ligante (obtidas pelo uso do AutoDock Vina) antes de definir as limitações volumétricas respectivas ao solvente. A informação do solvente deve então ser levada aos pacotes posteriormente usados, para a escolha adequada dos algoritmos de dinâmica molecular. O último passo de preparação se dá pela decisão de aplicar ou não o método de repartição de massa do hidrogênio HMR [22] (que foi utilizado em todas as dinâmicas realizadas neste trabalho). Uma vez que o sistema tenha sido devidamente preparado, pode-se dar início as seguintes etapas de simulação molecular: minimização, aquecimento, termalização e produção (descritas na sec. 3.5.1).

#### 4.1.2 Automação do uso das ferramentas AMBER

Como descrito previamente, é necessário amplo conhecimento das funcionalidades de múltiplas ferramentas do *AMBER/AMBERtools* e do processo de simulação para iniciar uma dinâmica molecular. Como consequência a esses e outros fatores, é demasiadamente laborioso aprender e/ou revisar os métodos corretos de uso desses pacotes computacionais. Embora necessário para pesquisadores da área, uma automação das funções básicas dos pacotes ajudaria ambos novatos e veteranos a focar na ciência, e não em detalhes computacionais irrelevantes oriundos da inexperiência com tais ferramentas. Devido à falta de um programa que faria um intermédio entre as ferramentas e o usuário, decidiu-se desenvolver um gerenciador de dados, objetivando:

1. Agilizar dinâmicas moleculares futuras deste trabalho; e
2. Facilitar o acesso a tais simulações para novos usuários do *AMBER/AMBERtools*.

O gerenciador<sup>2</sup> foi denominado ASM.py. Até a data de defesa da tese, o foco do programa é de suporte a pesquisa com a IsPETase. Como consequência, ele está restrito (temporariamente) aos seguintes tipos de dinâmica: atomística clássica e a *pH* constante.

Na Fig. 9, é apresentado um modelo entidade-relacionamento do ASM.py. O código foi escrito na linguagem de programação Python [29], versão 3.6, estruturando-o com orientação a objetos. A implementação desse foi dividida nas classes *Amber\_par*, *Amber\_run* e *Amber\_mutation*. Baseado nas informações fornecidas pelo usuário, em formato resumido e simplificado, *Amber\_par* detalha a metodologia usada nas dinâmicas criando arquivos necessários para uso de ferramentas específicas do AMBER. A comunicação via “prompt” de comando com as ferramentas do AMBER, é realizada por executáveis criados pela implementação da *Amber\_run*. A *Amber\_mutation* como o nome sugere é responsável pela edição dos arquivos de coordenadas, baseado na modificação desejada.

<sup>2</sup> Cujá versão mais recente pode ser encontrada no [GitHub](#).



Figura 9 – Modelo entidade-relacionamento para a implementação do ASM.py.

Essas classes estão em uma relação hierárquica de maneira que *Amber\_mutation* herda todos os atributos e métodos de sua classe pai (*Amber\_run*) e a última herda tudo da classe *Amber\_par*. Esse método de implementação facilita a correção de problemas e atualização do código fonte. Na Fig. 9, pode-se observar os atributos da classe *Amber\_par* divididos nos grupos:

1. *Solvente*: O uso de solvente implícito ou explícito é uma informação chave para qualquer simulação. Caso escolhido o último, deve-se informar qual dentre as opções: água, metanol, clorofórmio, N-metil acetamida ou ureia.
2. *Análise*: Guarda informações dos arquivos de análise que serão criados.
3. *Sistema*: Guarda as informações básicas do sistema como as coordenadas de cada

átomo e a distribuição de cargas.

4. *Simulação*: Guarda as informações sobre o tipo de dinâmica (DM ou CpHMD), o raio de corte para interações não ligadas (potencial de Lennard-Jones) e a duração de cada etapa da simulação.
5. *Processamento rápido*: O usuário pode escolher se usará cálculos em paralelo, via *Message Passing Interface (MPI)* nos processadores ou se usará uma *Graphics Processing Unit (GPU)* para agilizar o tempo de processamento. O método HMR sempre é utilizado.
6. *Mutação*: Guarda as informações quanto as posições e *ids* dos aminoácidos escolhidos do sistema original e qual método de mutação é utilizado entre: escolha por semelhança ou definida pelo usuário.

A classe *Amber\_run* utiliza da função “*system*” do python para se comunicar com as ferramentas do *AMBER/AMBERtools*, criando arquivos de execução *shell* quando necessário. A classe *Amber\_mutation* possui a maior complexidade, devido a manipulação de diversos conjuntos de cadeias de caracteres, para a alteração do(s) resíduo(s) escolhidos na mutação. Pode-se observar os atributos da classe *Amber\_mutation* divididos nos grupos:

1. *Mutação enviesada*: Guarda informação sobre o tipo e qual aminoácido foi sorteado na mutação “aleatória” – restrito aos aminoácidos do sítio ativo.
2. *Mutação não enviesada*: Guarda as informações dos aminoácidos escolhidos e suas mutações.
3. *Notação PDB*: Guarda as identidades dos átomos de cada aminoácido, definida no site [RCSB](#). Informação utilizada para verificar se a existência de átomos específicos faz sentido após a renomeação do aminoácido (algumas estruturas laterais possuem similaridade).
4. *Rastreamento de coordenadas*: Procura as novas *ids* dos aminoácidos escolhidos para mutação<sup>3</sup> após possíveis renumerações de resíduos devidas a formatação do PDB original. Também é responsável por criar (ou atualizar) uma lista com as posições dos átomos de cada aminoácido envolvido na mutação.
5. *Verificação de erros*: Verifica se ocorreu alguma anomalia que impeça a continuação de algum processo da mutação ou da simulação após completar a etapa de preparação. Está implementado o tratamento das principais fontes de erro deste processo.

<sup>3</sup> Utiliza os valores das coordenadas dos átomos dos aminoácidos escolhidos para comparação.

6. *Modificação do sistema*: Método de alteração das coordenadas e topologia para realização da mutação.

### 4.1.3 Validação do ASM.py

As análises de estabilidade estrutural apresentadas nesta tese são: “*Root Mean Square Fluctuation (RMSF)*”, “*Root Mean Square Deviation (RMSD)*” e raio de giro. Para essas, a coleta de dados em cada dinâmica molecular foi restringida apenas aos átomos do “*backbone*” da proteína. Este termo se refere a cadeia polipeptídica do sistema desconsiderando as cadeias laterais de cada aminoácido, sendo assim a “*espinha dorsal*” que mantém a estrutura proteica ligada. Essas análises de estabilidade possuem as seguintes funcionalidades:

1. *Root Mean Square Fluctuation (RMSF)*: Indica a flutuação média em relação a cada resíduo do sistema. De modo que, regiões de laço (fora de estruturas secundárias estáveis) apresentam os maiores picos de flutuação da proteína estudada.
2. *Root Mean Square Deviation (RMSD)*: Indica a flutuação média do “*backbone*” do sistema em relação a duração da dinâmica molecular.
3. Raio de giro: Devido a forma arredondada da enzima, este parâmetro indica o (des)enovelamento das cadeias polipeptídicas.

Quanto aos gráficos aqui apresentados, foi utilizado o último estado do sistema na etapa de termalização das dinâmicas como um estado referência, para verificação do comportamento da proteína durante a etapa de produção. Devido a isso, é perceptível uma pequena flutuação no começo dos gráficos de RMSD, que se da pela mudança de regime entre etapas (de *ensemble* à pressão constante NPT para *ensemble* a volume constante NVT). Quando necessário será discutido as análises que não referenciam o último estado da termalização. As regiões terminais da enzima não estão tridimensionalmente em estruturas secundárias estáveis (veja a Fig. 10), além disso, por serem terminais, é esperado alta flutuação dos resíduos. Para evitar ruído desnecessário nas análises RMSF, foi ignorado as regiões terminas.

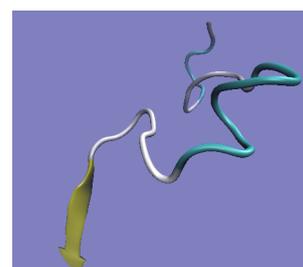


Figura 10 – Primeiros trinta aminoácidos da IsPETase nativa, indicando a região terminal-N.

Nesta seção são apresentados os resultados de uma dinâmica e uma replicata desta para a IsPETase nativa, e uma dinâmica para a mutante IsPETase<sup>D206E</sup>. Os dados da

primeira dinâmica molecular, com a IsPETase nativa, serão referenciados por NativaS1 e sua replicata por NativaS2.

#### 4.1.3.1 IsPETase nativa

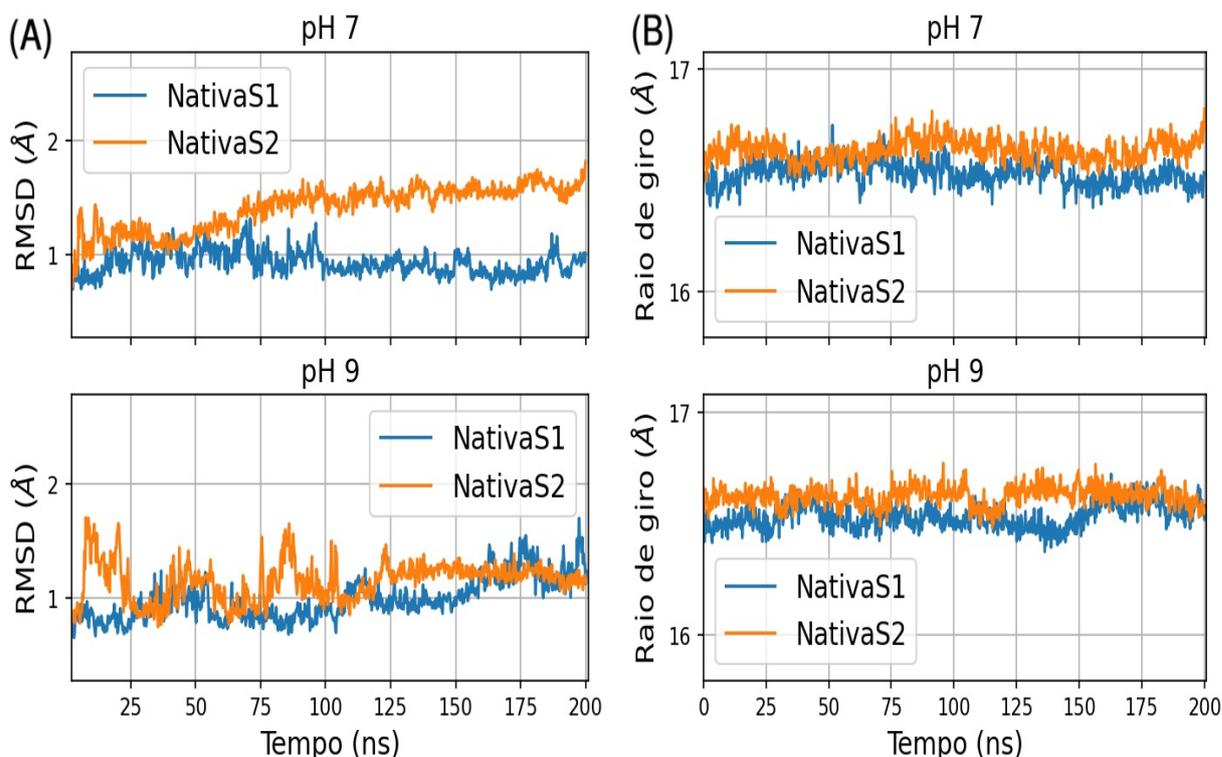


Figura 11 – Análise comparativa entre as dinâmicas NativaS1 e NativaS2, referente ao: (A) Desvio quadrático médio do “backbone”; e ao (B) Raio de giro da enzima.

Aqui estão apresentados os resultados das análises estruturais (Figs. 11 e 12) para dois valores de  $pH$  (7 e 9), comparando as simulações NativaS1 e NativaS2. No gráfico de RMSD (Fig. 11-A) observamos nos primeiros 100  $ns$  a enzima explorando as conformações possíveis, e então estabilizando no restante da dinâmica. No gráfico do raio de giro (Fig. 11-B), verificamos que a menos de uma oscilação percentual baixa – e esperada – o “backbone” permanece inalterado durante toda a dinâmica em ambos casos. Dito isso, estabelecemos as estruturas logo após os primeiros 100  $ns$  como as referências de cada dinâmica e utilizamos os a faixa de 100 – 200  $ns$  para a análise de RMSF.

Na Fig. 12-A vemos claramente regiões de alta estabilidade que se repetem em ambas dinâmicas. Após verificar a sequência de coordenadas salvas durante as dinâmicas, por um programa de visualização molecular, podemos observar que os picos de maiores flutuações ocorrem de fato nos laços da proteína. A numeração de alguns resíduos foi destacada na Fig. 12-B de modo a relacioná-la com o gráfico de RMSF. Podemos dizer que a proteína se mantém estável nesses valores de  $pH$  e temperatura (300  $K$ ), portanto essa metodologia prevê corretamente o comportamento estrutural da enzima, pois está de acordo com resultados experimentais já reportados (a exemplo da referência [5]).

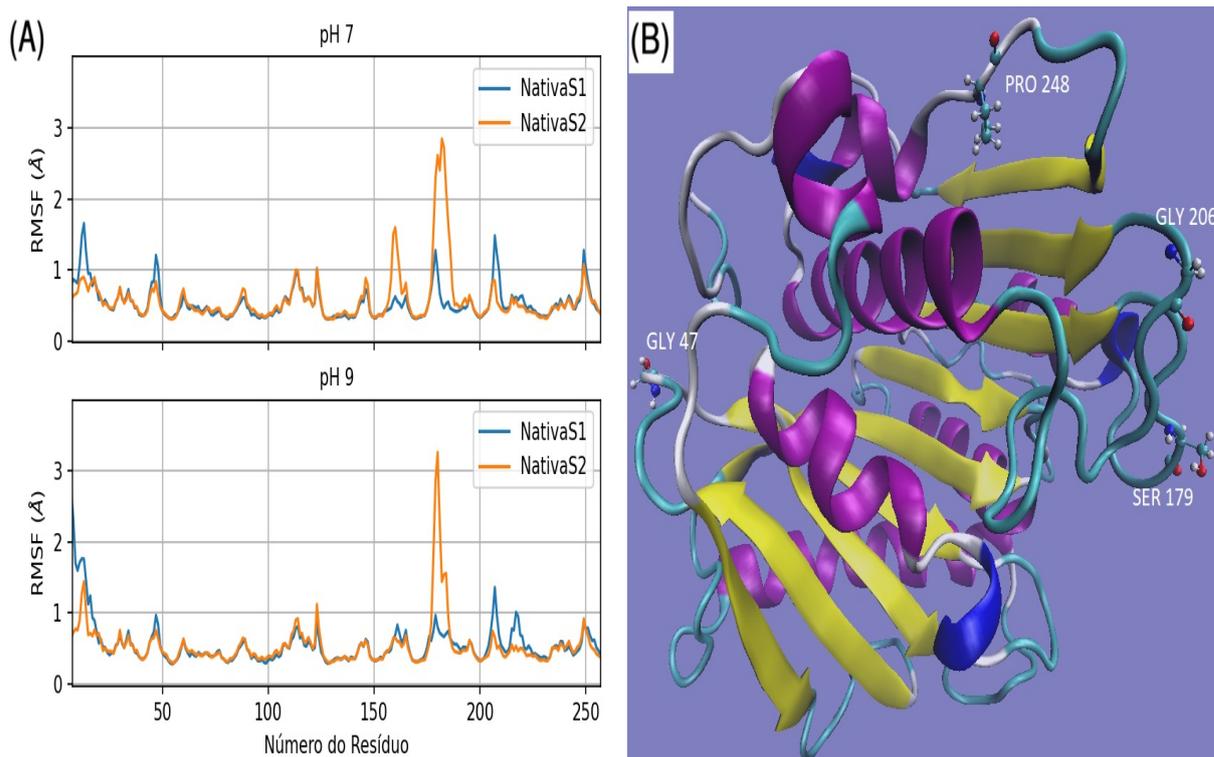


Figura 12 – Análise comparativa da estabilidade estrutural da IsPETase nas dinâmicas NativaS1 e NativaS2. (A) Flutuação quadrática média por resíduo do “backbone”. (B) Visão da proteína, destacando os resíduos com picos no RMSF.

#### 4.1.3.2 IsPETase<sup>D206E</sup>

Nesta seção são apresentadas as análises estruturais da variante D206E, comparando-as com os valores médios dos dados da NativaS1 e NativaS2 – referidas aqui apenas por “Nativa”. A variante D206E apresenta pouca variação nos dados de RMSD e raio de giro, enquanto manteve a mesma faixa de valores que Nativa (Fig. 13).

Como já discutido na Sec. 3.3, o número de identificação dos resíduos é reduzido em 28 unidades. Portanto, a mutação do aspartato (D) 206 pelo glutamato (E) é observada no gráfico de RMSF pelo pico de flutuação no resíduo 178 (Fig. 14) – o ponto mais alto de intensidade se refere ao resíduo 179, o resíduo 178 está cerca de 0,4 Å abaixo do pico. Como podemos observar das três análises, a variante está explorando conformações distintas da nativa. Porém de modo geral, ambas as dinâmicas apresentam comportamento indistinguível uma da outra, mantendo a estabilidade estrutural. O quanto esta mutação interfere no comportamento da enzima só será perceptível ao verificarmos a interação enzima-substrato, que é o escopo da próxima seção.

Deste exemplo e de testes com mutações múltiplas, verificamos que o método adotado para mutação não quebra erroneamente a proteína e alterações mínimas do sistema mantêm um comportamento semelhante ao original. Quanto aos testes com alterações de múltiplas componentes do sistema, foi julgada desnecessária a descrição na

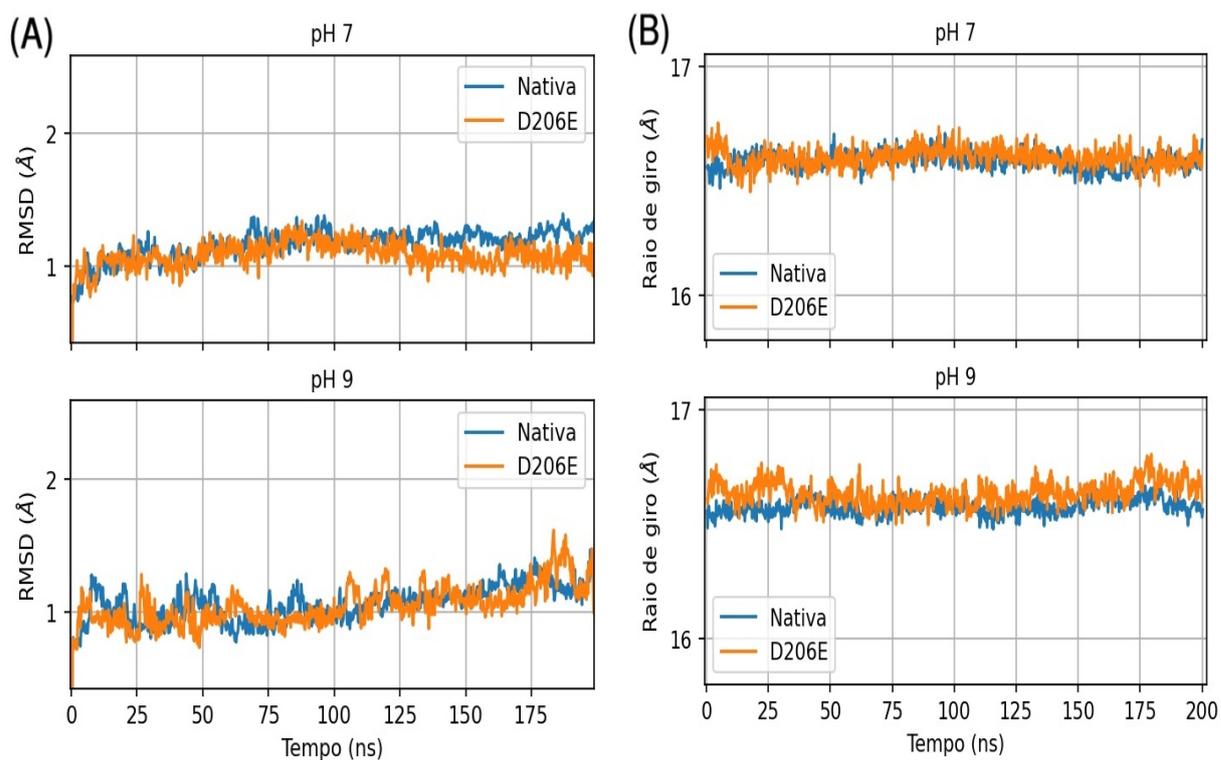


Figura 13 – Análise comparativa entre as dinâmicas da IsPETase e de sua variante IsPETase<sup>D206E</sup>. (A) Desvio quadrático médio do “backbone”; e (B) Raio de giro das proteínas.

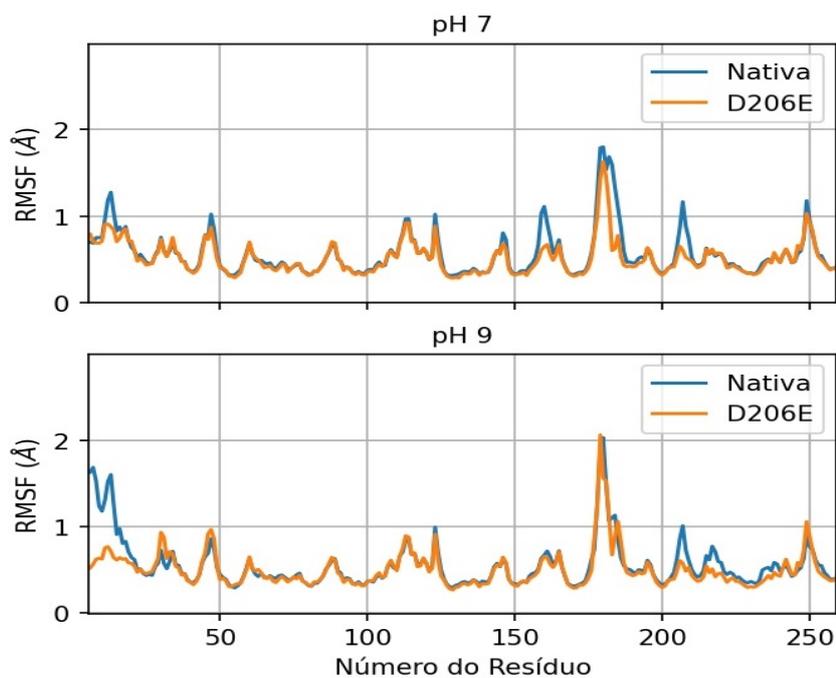


Figura 14 – Flutuação quadrática média por resíduo do “backbone” na IsPETase e em sua variante IsPETase<sup>D206E</sup>.

tese, pois são referentes a simples depuração de casos teste para o programa. Dito isto, validamos o uso do ASM.py para seguir com o estudo da IsPETase em dinâmicas pós encaixe do BHET.

## 4.2 Variações da tríade catalítica canônica da IsPETase e interação com o Monômero Central do PET

A discussão nesta seção está organizada da seguinte maneira: a preparação de nossos quatro complexos – com *docking* molecular na enzima nativa e as variantes D206E, D206E/ H237K e H237K; a dinâmica molecular com foco na distância nucleofílica  $r_{hc}$  e nos estados de protonação das posições dos resíduos escolhidos para as mutações; além de uma análise da energia de ligação com base na etapa de produção de cada dinâmica molecular.

### 4.2.1 Docking molecular

Tabela 1 – Primeiras cinco conformações de encaixe da lista ordenada baseada em  $s_{v+r}$  para IsPETase e suas variantes.

IsPETase	Conformações do BHET	$r_{hc}$ (Å)	Escore Vina (kcal/mol)	$s_{v+r}$
Nativa	<b>WT<sub>0</sub></b>	<b>2,93</b>	<b>-5,80</b>	<b>-2,87</b>
	WT <sub>1</sub>	2,90	-5,70	-2,80
	WT <sub>2</sub>	3,06	-5,80	-2,74
	WT <sub>3</sub>	2,97	-5,70	-2,73
	WT <sub>4</sub>	2,97	-5,70	-2,73
D206E	<b>DE<sub>0</sub></b>	<b>3,01</b>	<b>-5,80</b>	<b>-2,79</b>
	DE <sub>1</sub>	3,02	-5,80	-2,78
	DE <sub>2</sub>	2,93	-5,70	-2,77
	DE <sub>3</sub>	2,95	-5,70	-2,75
	DE <sub>4</sub>	2,95	-5,70	-2,75
D206E/H237K	<b>DEHK<sub>0</sub></b>	<b>3,20</b>	<b>-5,60</b>	<b>-2,40</b>
	DEHK <sub>1</sub>	3,21	-5,60	-2,39
	DEHK <sub>2</sub>	3,22	-5,60	-2,38
	DEHK <sub>3</sub>	3,13	-5,50	-2,37
	DEHK <sub>4</sub>	3,15	-5,50	-2,35
H237K	<b>HK<sub>0</sub></b>	<b>3,21</b>	<b>-5,60</b>	<b>-2,39</b>
	HK <sub>1</sub>	3,21	-5,60	-2,39
	HK <sub>2</sub>	3,21	-5,60	-2,39
	HK <sub>3</sub>	3,22	-5,60	-2,38
	HK <sub>4</sub>	3,23	-5,60	-2,37

Conforme descrito na Sec. 3.6 preparamos centenas de regiões para o encaixe do BHET ao redor da serina nucleofílica em cada variante da IsPETase. Com esta preparação,

obtivemos milhares de conformações de encaixe em cada complexo. Podemos verificar rapidamente todas essas com um script simples em qualquer linguagem de programação. Para julgar cada uma das propostas e identificar as melhores, analisamos dois parâmetros: o escore vina e os valores de  $r_{hc}$ . Quanto menor o escore vina, melhor o ajuste, e como era de interesse configurações com  $r_{hc} \sim 4 \text{ \AA}$ , dada a distribuição dos dados, definimos um fator de minimização adimensional como

$$s_{v+r} = |r_{hc}| - |\text{escore vina}|.$$

Procurar um bom modelo de encaixe seria o mesmo que procurar o menor valor possível de  $s_{v+r}$ . No entanto, para evitar conformações indesejáveis, como um com  $|r_{hc}| \gg 4$  e  $|\text{escore vina}| \ll -5,5$  ou um com  $|r_{hc}| < 4$  e  $|\text{escore vina}| > -5,5$ , precisávamos ordenar nossa lista de conformações em vez de procurar apenas um com  $s_{v+r}$  pequeno. Optamos pelo algoritmo *quick sort* por ser simples de escrever e ter um tempo de processamento ótimo, em média. Após a etapa de ordenação, poderíamos facilmente descartar quaisquer valores atípicos e escolher um modelo desejável entre as primeiras opções de nossa lista (veja a Tab. 1). Para uma descrição mais detalhada desta seção, veja o apêndice C.1.

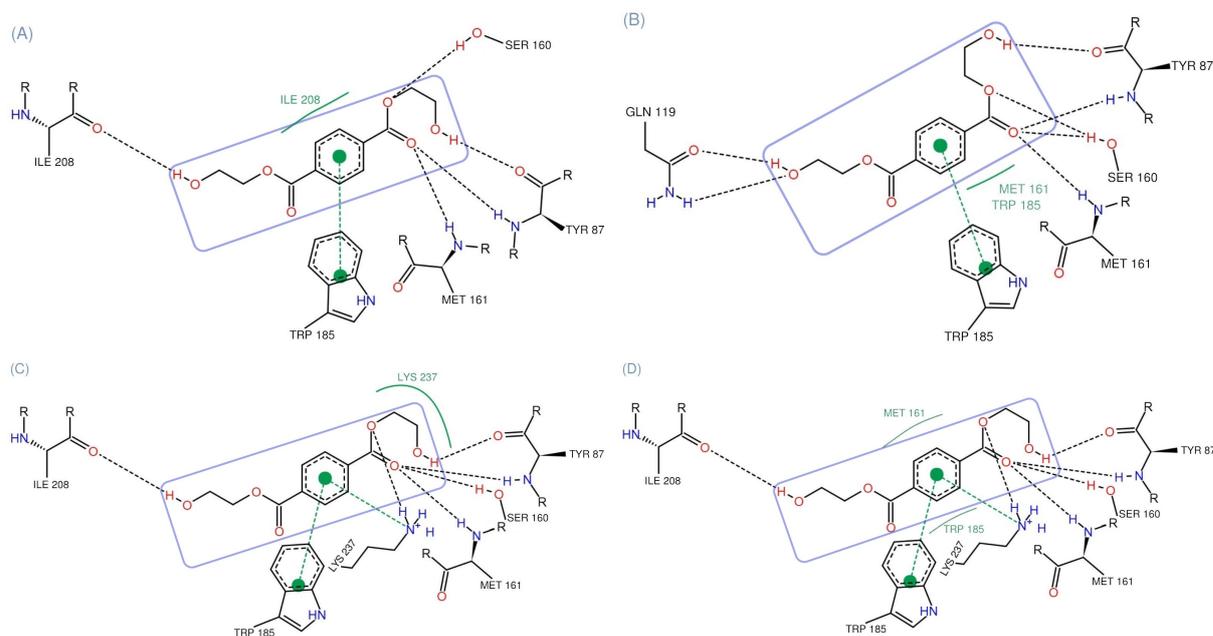


Figura 15 – Ligante BHET (destacado em azul) encaixado em cada variante IsPETase: (A) Nativa, (B) D206E, (C) D206E/H237K e (D) H237K. Neste esquema 2D, as linhas tracejadas representam ligações de hidrogênio e os rótulos de resíduos verdes são os aminoácidos com contatos hidrofóbicos com o ligante.

Na Fig. 15, apresentamos um esquema bidimensional do BHET encaixado com as conformações  $WT_0$ ,  $DE_0$ ,  $DEHK_0$  e  $HK_0$  nas respectivas formas variantes da IsPETase, usando o Programa PoseView [32]. Como podemos ver, o BHET está ligado a cada variante em estreita proximidade com os resíduos da fenda de oxiânion (TYR 87 e MET 161) enquanto em uma interação aromática-aromática com o TRP 185. A figura indica que as

variantes com mutação de apenas um resíduo (Fig. 15-B e Fig. 15-D) favorecem o contato hidrofóbico do BHET com o MET 161 e o TRP 185. No entanto, conformações semelhantes foram encontradas nos resultados de *docking* com a nativa – como a conformação  $WT_2$ , que tem a mesma afinidade (escore vina) que o  $DE_0$ . Porém, a conformação na Fig. 15-C, onde o BHET mantém contato mais próximo com o resíduo 237 ao invés do MET e TRP, foi observada apenas na variante D206E/H237K.

#### 4.2.2 Simulações de dinâmica molecular

As simulações de dinâmica molecular aqui foram realizadas após o *docking* molecular discutido previamente, usando as conformações  $WT_0$ ,  $DE_0$ ,  $DEHK_0$  e  $HK_0$  da Tab. 1. Não exploramos a paisagem conformacional antes de encaixar o ligante nas variantes, mas fizemos algumas replicatas das dinâmicas pós-*docking*. Em cada uma das dinâmicas replicadas, mantivemos todos os parâmetros de entrada e refizemos as etapas pós-*docking*: minimização pós-solvatação, aquecimento, termalização e produção. As replicatas apresentaram os mesmos comportamentos em média para cada variante, portanto vamos nos concentrar em uma dinâmica molecular por variante. Na Fig. 16, apresentamos o compor-

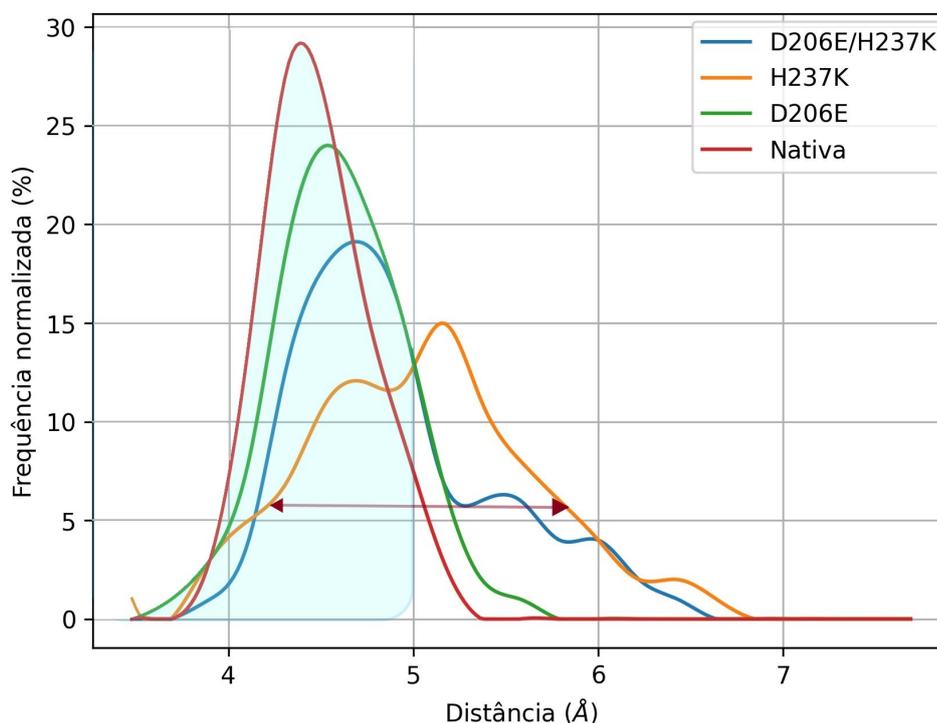


Figura 16 – Distribuição de frequência normalizada das distâncias nucleofílicas  $r_{hc}$ . Destacado em azul a região de interesse. Os dados para cada sistema Variante-BHET foram obtidos da etapa de produção em pH 7, 300 K e duração de 200 ns.

tamento do ligante no decorrer das simulações, focando nas distâncias nucleofílicas  $r_{hc}$ . Em geral, com todas as variantes, o BHET permaneceu ligado ao sítio ativo, e como indicado pela região azul os complexos variante-BHET permaneceram com maior frequência em

configurações favoráveis a ataques nucleofílicos. A nativa e a variante IsPETase<sup>D206E</sup> mantiveram uma ligação mais estável, enquanto que a IsPETase<sup>H237K</sup> apresentou uma distribuição de frequência mais larga, que indica maior mobilidade do BHET na região — possivelmente devida a interação mais fraca com os resíduos do sítio ativo.

Tabela 2 – Estados de protonação  $S_i$  dos resíduos de acoplamento da IsPETase e suas variantes. O valor entre parênteses indica a permanência de cada estado no decorrer dos 200 ns de simulação.

IsPETase	Resíduo 206	Resíduo 237
Nativa	$S_0(100,0\%)$	$S_0(0,6\%), S_1(51,9\%), S_2(47,5\%)$
D206E	$S_0(97,4\%), S_1(1,7\%), S_3(0,9\%)$	$S_0(28,5\%), S_1(63,5\%), S_2(8,0\%)$
D206E/H237K	$S_0(82,1\%), S_1(7,0\%), S_2(0,2\%), S_3(10,2\%), S_4(0,5\%)$	$S_0(94,0\%), S_1(6,0\%),$
H237K	$S_0(100,0\%)$	$S_0(95,4\%), S_1(4,6\%)$

Na Tab. 2, apresentamos as preferências em permanência e as mudanças nos estados de protonação — respectivas aos resíduos 206 e 237 — que ocorreram durante a simulação de cada sistema. O  $i$ -ésimo estado de cada resíduo foi escrito na notação abreviada  $S_i$  para facilitar a leitura da tabela. A representação tridimensional de cada estado de protonação é apresentada na Fig. 17. Pode-se notar que, em todos os casos, os resíduos apresentaram preferências por um estado de protonação específico. Na nativa o aspartato da tríade catalítica permaneceu desprotonado e sem indicação de mudança de estado ao longo da produção. A histidina também permaneceu desprotonada, mas com igual probabilidade de estar no estado em que o hidrogênio está ligado ao nitrogênio ND1 ou ao nitrogênio NE2. Na variante D206E, podemos ver que o resíduo 206 sofre uma pequena queda na permanência no estado desprotonado, mas permanece neste estado por 97,4% da MD. A histidina, por outro lado, permaneceu protonada em 28,5% da MD (um aumento considerável em relação aos 0,6% da simulação com a nativa). Na mutante dupla D206E/H237K, vemos que o estado desprotonado do resíduo 206 permanece preferencial, mas menos estável (com uma queda na permanência para 82,1%) e, ao contrário da nativa e da D206E, na mutante dupla o resíduo 237 permanece preferencialmente protonado (94,0%).

Parece haver uma pequena interferência no estado de protonação do resíduo 206 devido à mutação no resíduo 237 e vice-versa. Na nativa, é igualmente provável que a histidina esteja desprotonada no estado 1 ou 2, enquanto que na IsPETase<sup>D206E</sup> é mais provável que a histidina esteja no estado 1. Podemos ver um comportamento semelhante entre as variantes D206E e D206E/H237K, onde a mutação no resíduo 237 torna o glutamato ligeiramente mais propenso a estar protonado. Essa interferência possivelmente se deve à curta distância entre esses resíduos. Na simulação com a H237K, podemos ver o

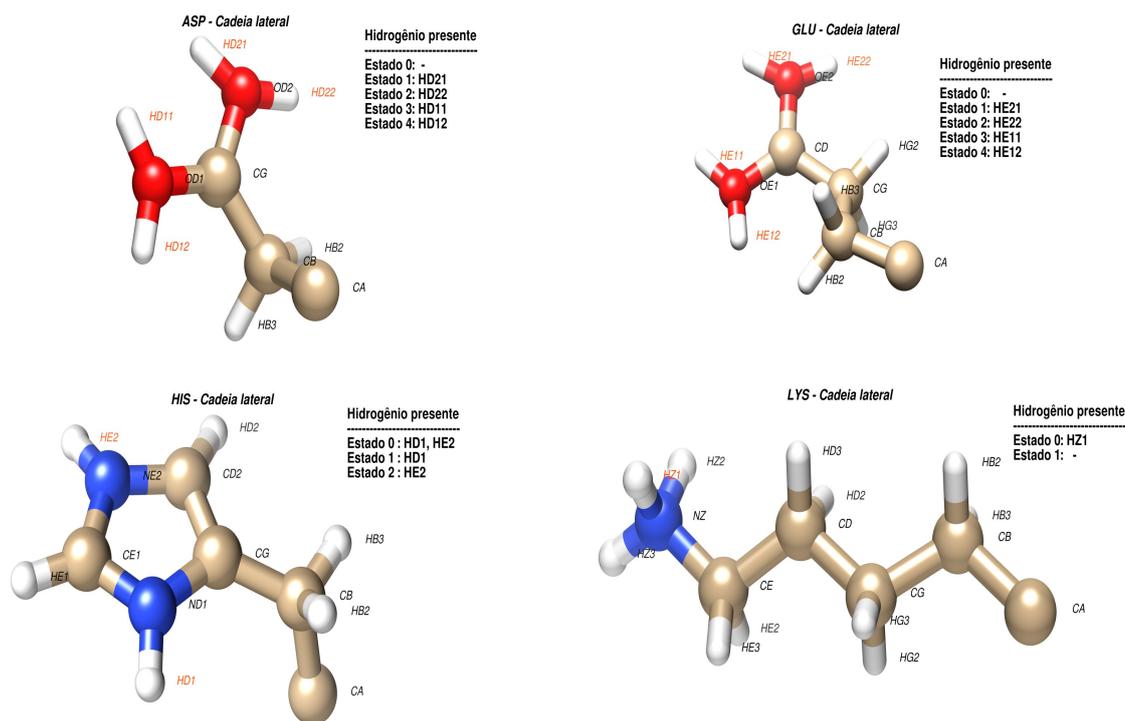


Figura 17 – Configuração dos estados de protonação para os resíduos 206 e 237 das variantes da IsPETase. A presença dos hidrogênios marcados em vermelho depende do estado atual do resíduo.

ASP 206 com o mesmo comportamento apresentado na nativa e a LYS 237 com o mesmo comportamento que em nossa mutante dupla, curiosamente, a troca no resíduo 206 de glutamato para aspartato não afetou o estado de protonação da LYS 237, como ocorrido nos outros casos.

### 4.2.3 Decomposição da energia de interação

Nesta seção, discutiremos a interação dos aminoácidos chave dos quatro complexos discutidos anteriormente. Aqui vamos nos referir à diferença na estabilidade apresentada pela Fig. 16 usando duas métricas: o valor médio da distância nucleofílica  $\overline{r_{hc}}$  e a frequência que  $r_{hc}$  permaneceu inferior a 5 Å, denotada aqui por  $f_{<5}$ . Esta frequência será dada como um valor percentual relativo a duração de 200 ns das simulações. Pela Tab. 3 poderíamos supor uma relação direta entre o  $\Delta G_{Lig}$  e as métricas  $\overline{r_{hc}}$  e  $f_{<5}$ . Contudo, a ordem de grandeza para as diferenças em  $\overline{r_{hc}}$  entre os complexos é insignificante. Então, a única associação plausível que poderíamos fazer é

$$\overline{r_{hc}} \sim 5 \leftrightarrow \Delta G_{Lig} < 0.$$

Por outro lado, a relação onde interações mais intensas ( $\Delta G_{Lig} \ll 0$ ) mantém valores maiores de  $f_{<5}$  era esperada, pois é conhecido da literatura que o ataque nucleofílico é mais favorável a distâncias próximas de 4 Å. Note que, relativo a simulação com a nativa,

Tabela 3 – Valores médios para a distância relativa ao ataque nucleofílico, a frequência que esta distância permaneceu inferior a 5 Å, e a energia de interação, calculadas nos complexos IsPETase-BHET.

IsPETase	$\bar{r}_{hc}$ (Å)	$f_{<5}$ (%)	$\Delta G_{Lig}$ (kcal/mol)
Nativa	4,50	96	-21,15 ( $\pm 3,28$ )
<i>D206E</i>	4,62	91	-18,21 ( $\pm 5,12$ )
<i>D206E/H237K</i>	4,91	70	-11,28 ( $\pm 7,23$ )
<i>H237K</i>	5,05	51	-6,48 ( $\pm 4,89$ )

ambas mutantes simples mostraram um declínio nos valores de  $f_{<5}$ . Curiosamente, ao invés da mutante dupla perder ainda mais intensidade na interação IsPETase-BHET, ela assume uma posição intermediária entre as mutantes simples.

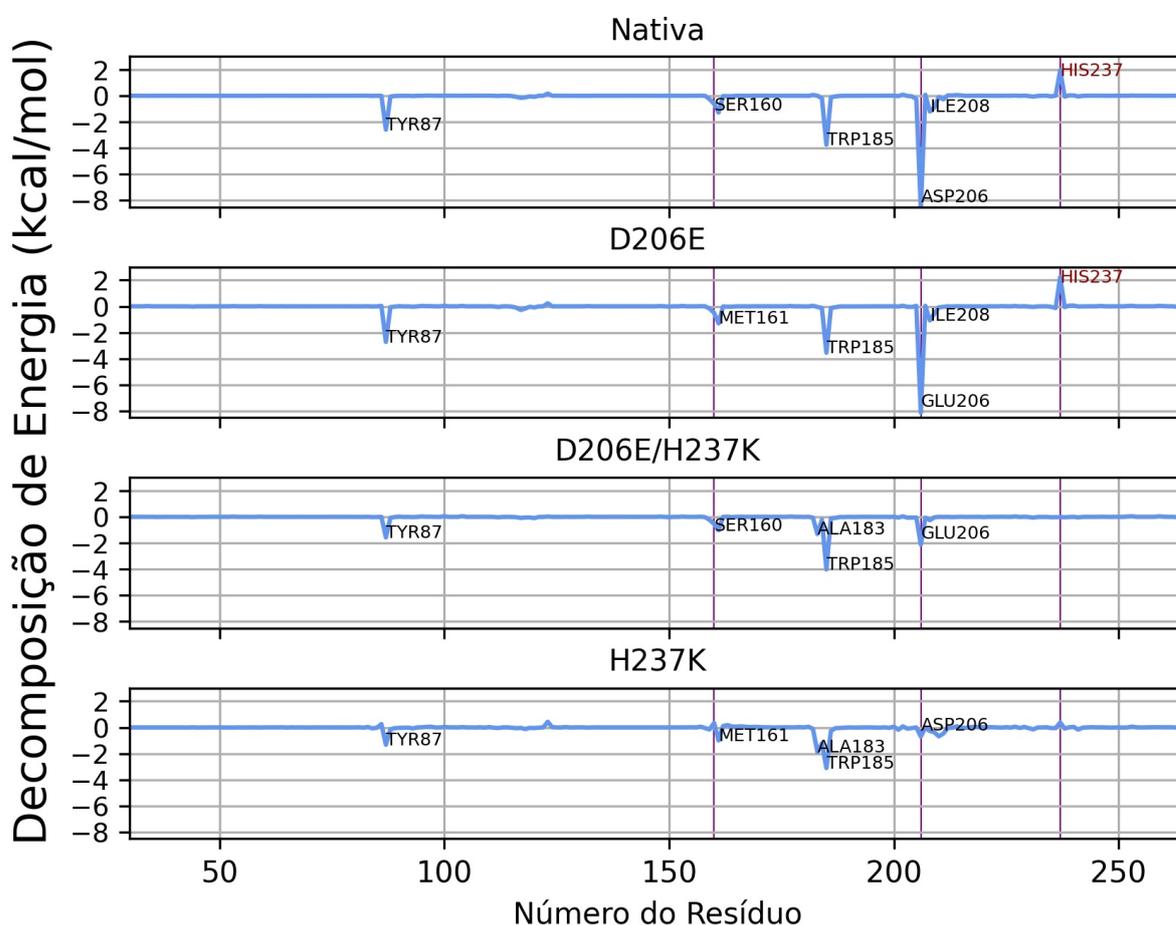


Figura 18 – Os aminoácidos que contribuem para a interação IsPETase-BHET em pH 7 e 300 K. As linhas verticais destacam a tríade catalítica SER 160, ASP 206, HIS 237 e suas respectivas mutações.

A Fig. 18 apresenta a análise de decomposição da energia de ligação total para nossos quatro complexos, fornecendo a contribuição de cada aminoácido para a ligação. Esta análise mostra as três regiões de interação discutidas no modelo de Ge [14] – a fenda de oxiânion (TYR 87 e MET 161), a tríade catalítica (SER 160, ASP 206 e HIS

237) e o TRP 185, que tem uma interação aromática-aromática com o anel benzênico do BHET. De acordo com a literatura, as serino proteases com a tríade catalítica canônica SER/HIS/ASP possuem a serina como nucleófilo, a histidina como ácido ou base geral, e o aspartato com a função de ajudar a orientar a histidina e neutralizar a carga que ela pode desenvolver na transição de estados. Estudos de mutagênese indicam que a maior contribuição para a catálise geralmente vem do aspartato [33]. Como podemos ver pela decomposição no sistema Nativa-BHET (figura superior 18), o aminoácido com a maior contribuição favorável para a interação é o aspartato da tríade, como esperado. Podemos notar também que a histidina aparece com uma contribuição desfavorável. Ao comparar a nativa com nossas variantes, a D206E mantém o mesmo padrão de energia. Por outro lado, tanto a D206E/H237K quanto a H237K apresentaram uma redução de intensidade nos vales de energia, principalmente no resíduo 206. Porém, as mutantes com lisina também apresentaram redução na contribuição desfavorável do resíduo 237. A H237K ainda possui alguns picos desfavoráveis com pequenas intensidades, mas curiosamente a D206E/H237K não apresenta nenhum desfavorecimento à interação enzima-substrato.

Da discussão na Sec. 4.2.2 sobre os estados de protonação dos resíduos 206 e 237, percebemos uma diferença de carga entre os sistemas nativa/D206E e os com as mutantes de lisina. Isto é devido ao próton extra na cadeia lateral da lisina no estado 0. Uma vez que observamos uma redução de intensidade nas contribuições totais de ambas as mutantes de lisina, ou a mudança conformacional devido a diferentes cadeias laterais (do imidazol para a cadeia alifática), a nova distribuição de carga ou ambas afetaram negativamente a região. A substituição D206E sozinha basicamente não faz nada para a interação IsPETase-BHET. No entanto, quando combinada com a adição da lisina na posição 237, há uma interferência favorável – do vale de energia quase inexistente em 206 na variante H237K para o padrão apresentado pela mutante dupla, onde não há contribuições desfavoráveis e ao mesmo tempo mostrando um aumento de intensidade nos vales de 185 e 206.

Na Fig. 19, apresentamos uma visualização comparativa da conformação dos sítios ativos da IsPETase nativa e das variantes com inserção da lisina na posição 237, utilizando o software UCSF Chimera [13]. Na nativa (Fig. 19-A), o ligante se estabiliza no centro do sítio ativo permitindo proximidade com a serina nucleofílica e a fenda de oxiânion (MET e TYR). Na Fig. 19-B, com a inserção de carga devido à lisina, podemos ver uma mudança na conformação dos resíduos 206 e 237, onde a lisina se estabilizou próxima ao aspartato enquanto o bloqueava do ligante - isso pode explicar porque na variante H237K o resíduo 206 perde a energia de ligação (como mostra a Fig. 18). Com essa mudança, também vemos como o ligante se afasta da fenda de oxiânion. Na Fig. 19-C, onde também temos a mutação do resíduo 206 para glutamato, vemos a reabertura da conformação das cadeias laterais para os resíduos 206 e 237, permitindo que o ligante volte a interagir com o resíduo 206 (indicado pelo poço de energia na Fig. 18).

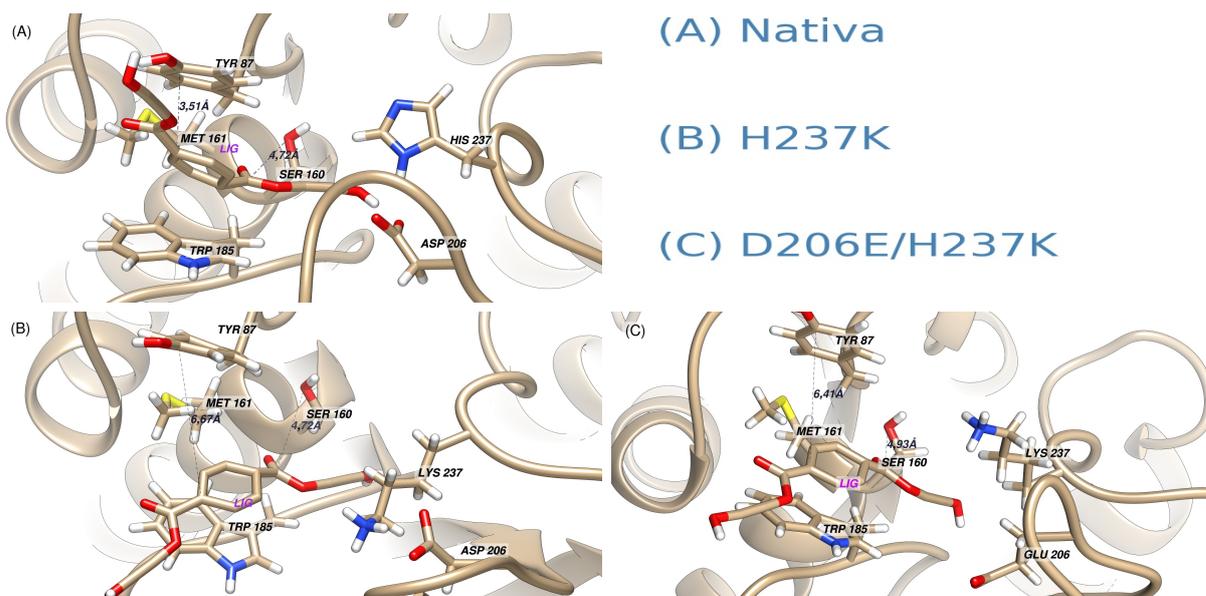


Figura 19 – Visão comparativa do ligante BHET no sítio ativo da IsPETase, entre a nativa e as variantes com a lisina. As imagens foram tiradas da fase de produção da dinâmica molecular de cada sistema.

Mesmo que na Fig. 19 não possamos visualizar o comportamento médio de nossos complexos, mas apenas a conformação de um único instante, com a discussão da Fig. 18 e Tab. 3 como suporte, pode-se concluir que a nova distribuição de carga e conformação das cadeias laterais estão mantendo a distância nucleofílica e empurrando o BHET para longe da fenda de oxidação. Os únicos fatores para a diminuição da energia de interação são o distanciamento da fenda de oxidação e a conformação das cadeias laterais em 206 e 237. No apêndice C.2, apresentamos uma conformação mais favorável dos resíduos onde o BHET estabiliza cerca de um angstrom mais próximo a tirosina da fenda de oxidação, o cálculo do MMPBSA da dinâmica neste apêndice apresentou  $|\Delta G_{Lig}| = (-16,28 \pm 8,00)$  kcal/mol.

A contribuição desfavorável da histidina pode estar correlacionada com a baixa atividade da IsPETase em pH 7 e 300 K - suas condições ótimas de atividade estão em torno de pH 9 e 313 K. Entre as variantes estudadas, tanto a D206E quanto a H237K não apresentaram resultados promissores quanto à interação da enzima com este monômero plástico. No entanto, o comportamento peculiar da mutante dupla D206E/H237K pode ser utilizado para pesquisas futuras sobre a degradação do PET nessas condições, além de provocar curiosidade sobre quais consequências poderiam surgir em resultados empíricos.

## 5 Conclusões

Após o estudo dos processos de dinâmica molecular e de uso dos pacotes computacionais *AMBER/AMBERtools*, notou-se a possibilidade de automatização dos procedimentos necessários para a preparação do sistema e execução de cada estágio de uma simulação de Dinâmica Molecular. Visto isso, foi implementado um programa (escrito em Python) para ser um gerenciador de dados, sendo um intermediário entre as ferramentas *AMBER/AMBERtools* e o usuário. Esse programa, denominado *ASM.py*, foi então utilizado para realizar todas as simulações deste trabalho. Com uso dessa nova ferramenta, a metodologia adotada foi resumida pela execução do programa (facilitando a replicação dos resultados obtidos). O programa *ASM.py* foi registrado com licença GNU GPL no INPI sob número BR512021001880-0. Tanto o *ASM.py* quanto outras ferramentas focadas em *Docking* e em análises com auxílio gráfico estão disponíveis na [minha página do GitHub](#).

Após a validação da implementação do *ASM.py*, preparamos três variantes (D206E, H237K e D206E/H237K) e realizamos um protocolo de *Docking* Molecular comparativo nas três e na enzima nativa. Para encontrar boas conformações de encaixe do ligante (BHET) em cada enzima (IsPETase e as três variantes propostas), utilizamos a ideia da busca em largura. Ao invés de usar um nível altíssimo de exaustividade no programa de *Docking*, para buscar a melhor conformação de um caminho específico do espaço de fase, buscamos amostrar centenas de estados iniciais distintos ao modificar a superfície de contato das enzimas com nosso ligante. Com essas centenas de estados iniciais utilizamos um nível baixo para a exaustividade – o valor 8 é o padrão para sistemas simples, o valor que utilizamos foi de 10. Assim, obtivemos boas conformações e prosseguimos para as DMs.

Analisando a distribuição de frequência da distância nucleofílica em cada sistema, verificamos que embora todos os sistemas mantiveram o ligante no sítio ativo, uma das variantes (H237K) indicava instabilidade do ligante na região. Com isso realizamos o protocolo MMPBSA para analisar a interação enzima-substrato utilizando a decomposição por resíduo. Nossas descobertas indicaram uma pequena perda em  $|\Delta G_{Lig}|$  para a mutante simples D206E, mas no geral a variante manteve o mesmo padrão de energia da nativa. Já para a mutante H237K, observamos uma redução de intensidade na contribuição energética de todos os resíduos chave. No entanto, apesar de ambas as mutantes individuais apresentarem um declínio na energia de interação, a combinação delas (D206E/H237K) acabou interferindo favoravelmente. Os resultados de *Docking* e DM dos complexos variante-BHET presentes nesta tese foram publicados no artigo [34].

As discussões apresentadas aqui podem beneficiar esforços futuros quanto a reci-

clagem biológica do PET em condições ambientais. Dentre as possibilidades de trabalhos seguintes a este estão: estudos utilizando a técnica *Quantum mechanics/Molecular Mechanics* (QM/MM), para habilitar a mudança de cargas parciais durante a DM, de modo a verificar se o ataque nucleofílico continua ativando em nossa variante dupla (D206E/H237K); analisar outras variantes da forma canônica como por exemplo alterando o nucleófilo ou adicionando um segundo nucleófilo no sítio ativo; ou estudos experimentais de nossa mutante dupla visando medir se ocorreu alteração no tempo ativo da enzima e em caso positivo verificar o novo rendimento.



REPÚBLICA FEDERATIVA DO BRASIL  
MINISTÉRIO DA ECONOMIA  
**INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL**  
DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

## Certificado de Registro de Programa de Computador

Processo Nº: **BR512021001880-0**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 22/02/2021, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

**Título:** ASM (Amber Simulation Manager)

**Data de publicação:** 22/02/2021

**Data de criação:** 22/02/2021

**Titular(es):** UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL

**Autor(es):** BRUNO CAMARGO BRAGA; MARCOS SERROU DO AMARA

**Linguagem:** PYTHON; SHELL SCRIPT

**Campo de aplicação:** FQ-05; FQ-07; FQ-08; FQ-13

**Tipo de programa:** AT-01; FA-01; GI-01; SO-07; SO-09

**Algoritmo hash:** SHA-512

**Resumo digital hash:**

124474A2D5FF92A0342BDBC5DC61F804A1B5BEB55C1D62029297F14922C88A0D0A9B27A312F972F5D1CFBCEB  
42CE34D762981DDE6BAC6D51AB91B004F5A20938

**Expedido em:** 17/08/2021

**Aprovado por:**

Carlos Alexandre Fernandes Silva

Chefe da DIPTO

# A Teoria Base de Dinâmica Molecular

## A.1 Mecânica molecular: potenciais de energia

A energia potencial de uma molécula é geralmente calculada a partir da expressão:

$$E_{tot} = E_{str} + E_{tor} + E_{ang} + E_{oop} + E_{ster} + E_{elet} + E_{nb} - E_{solv} \quad (1)$$

onde  $E_{str}$  é a energia necessária para esticar ou comprimir uma ligação,  $E_{tor}$  é a energia para rotação em torno de uma ligação,  $E_{ang}$  é a energia necessária para comprimir ou expandir um ângulo de ligação,  $E_{oop}$  é a energia necessária para deformar um sistema da planaridade,  $E_{ster}$  é a energia gerada pela repulsão estérica<sup>1</sup> entre grupos ou átomos próximos,  $E_{elet}$  é o potencial eletrostático entre as cargas dentro da molécula,  $E_{solv}$  é a energia de solvatação e  $E_{nb}$  engloba todas as outras interações intermoleculares<sup>2</sup> referentes aos termos não ligados<sup>3</sup> entre moléculas não polares [35].

As funções de energia potencial  $E_{str}$ ,  $E_{ang}$ ,  $E_{oop}$  e  $E_{ster}$  são geralmente estimadas por uma função simples derivada da Lei de Hooke da forma:

$$E = K \cdot (R - R_0)^2 \quad (2)$$

onde  $R - R_0$  é a mudança no comprimento, largura, ou deslocamento fora do plano de uma ligação, ângulo ou átomo. A função  $E_{tor}$  relaciona a energia de quatro átomos ligados sequencialmente ao seu ângulo diedral<sup>4</sup> (de torção) e é geralmente baseada na preferência observada por conformações alternadas sobre eclipsadas (veja a Fig. 20), especialmente sempre que existe a possibilidade de uma rotação “livre” sobre uma ligação. Geralmente está relacionado ao ângulo de torção  $w$  pela expressão:

$$E_{tor} = K \cdot [1 + (-1)^J \cdot \cos(J \cdot w)], \text{ para } J = 1, 2 \text{ e } 3 \quad (3)$$

As formas de  $E_{elet}$  e  $E_{nb}$  não são tão padronizadas quanto dos outros potenciais listados acima e variam de uma implementação da mecânica molecular para outra. É particularmente importante ressaltar que as funções de energia potencial usadas na mecânica molecular são selecionadas empiricamente e suas formas não referem-se aos

<sup>1</sup> Repulsão estérica refere-se ao arranjo dos átomos na molécula. Esse efeito surge do fato que cada átomo dentro de uma molécula ocupa uma certa quantidade de espaço. Se os átomos forem colocados muito próximos, há um custo associado em energia devido à sobreposição de nuvens eletrônicas (repulsão de Pauli ou Born), e isso pode afetar a forma preferencial da molécula e reatividade.

<sup>2</sup> Ligações intermoleculares são aquelas que mantêm átomos em diferentes moléculas unidos.

<sup>3</sup> Representam forças de van der Waals e eletrostáticas.

<sup>4</sup> Ângulo formado no encontro de dois planos.

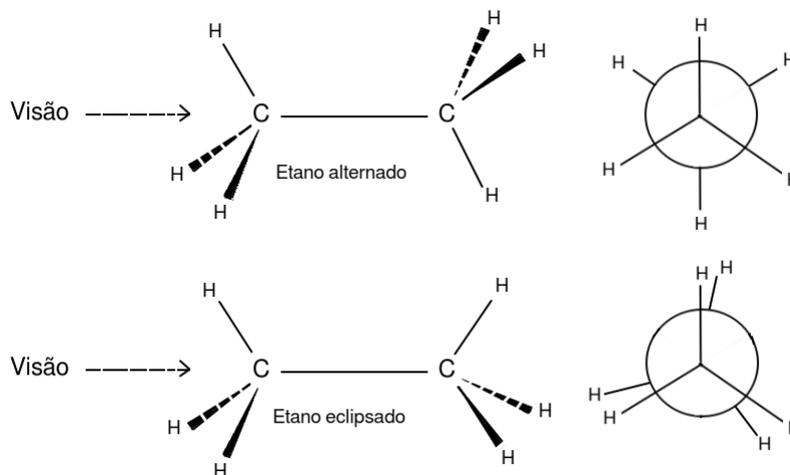


Figura 20 – Isomerismo conformacional.

fundamentos científicos do movimento atômico ou molecular, ou efeitos estereoeletrônicos<sup>5</sup>.

## A.2 Potencial de Lennard-Jones

A formulação padrão do potencial de Lennard-Jones 6-12  $V_{i,j}$  [36] entre dois átomos  $i$  e  $j$  está apresentada na Eq. 4,

$$V_{i,j} = 4\epsilon_{i,j} \left[ \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^{12} - \left( \frac{\sigma_{i,j}}{r_{i,j}} \right)^6 \right] \quad (4)$$

onde  $r_{i,j}$  é a distância que separa os dois átomos,  $\epsilon_{i,j}$  é a profundidade do poço de potencial para a interação dos átomos  $i$  e  $j$ , e  $\sigma_{i,j}$  é a distância onde o potencial é exatamente zero, ou seja, onde a ‘repulsão’ começa para os dois átomos. Ambos  $\epsilon_{i,j}$  e  $\sigma_{i,j}$  são específicos para o par de átomos.

Outra formulação possível de  $V_{i,j}$ , relativa ao conceito de raios de van der Waals<sup>6</sup>, é apresentada pela Eq. 5.

$$V_{i,j} = \epsilon_{i,j} \left[ \left( \frac{R_{min}}{r_{i,j}} \right)^{12} - 2 \left( \frac{R_{min}}{r_{i,j}} \right)^6 \right] \quad (5)$$

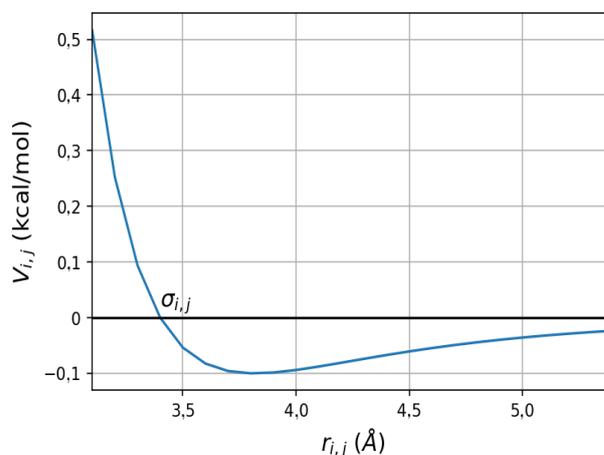


Figura 21 – Exemplo de potencial de Lennard-Jones.

<sup>5</sup> Um efeito na geometria molecular, reatividade ou propriedades físicas devido às relações espaciais na estrutura eletrônica das moléculas, em particular a interação entre orbitais atômicos ou moleculares.

<sup>6</sup> Raio de uma esfera sólida imaginária empregada para representar um átomo.

Nesse caso,  $R_{min}$  é a soma dos raios de van der Waals,  $R_i + R_j$  dos átomos  $i$  e  $j$ , a distância de contato na qual o potencial está no mínimo, ou seja, em um valor de  $-\epsilon$ . Um exemplo de representação gráfica do potencial de Lennard-Jones é apresentado na Fig. 21.

Combinando as Eqs. 4 e 5 gera a relação a seguir entre  $\sigma$  e  $R_{min}$ :

$$R_{min} = 2^{1/6}\sigma \quad (6)$$

Em campos de força, a notação “A, B” do potencial de Lennard-Jones é comumente usada:

$$V_{i,j} = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \quad (7)$$

onde  $A_{i,j}$  e  $B_{i,j}$  são parâmetros específicos para pares de átomos  $i$  e  $j$ . O significado de  $A_{i,j}$  e  $B_{i,j}$  é facilmente deduzido:

$$A = 4\epsilon\sigma^{12} \text{ e } B = 4\epsilon\sigma^6 \quad (8)$$

### A.3 Campos de força

O modelo apresentado por Cornell et. al. [37], (Eq. 9) descreve respectivamente as ligações químicas, os ângulos por uma expressão harmônica diagonal simples, energias diédricas como um conjunto simples de parâmetros (muitas vezes apenas especificados pelos dois átomos centrais), a interação Van Der Waals por um potencial de Lennard-Jones e interações eletrostáticas por uma interação Coulombiana de cargas pontuais centradas no átomo. As interações eletrostáticas e de Van der Waals são calculadas apenas entre átomos em moléculas diferentes ou para átomos na mesma molécula separados por pelo menos três ligações. Essas interações não ligadas separadas por exatamente três ligações<sup>7</sup> são reduzidas pela aplicação de um fator de escala.

$$E_{total} = \sum_{\text{ligações}} k_r (r - r_{eq})^2 + \sum_{\text{ângulos}} k_\theta (\theta - \theta_{eq})^2 + \sum_{\text{diedral}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (9)$$

A suposição é que essa representação simples das energias de ligação e ângulo é adequada para modelar a maioria dos sistemas não tensionados. O objetivo deste campo de força é modelar com precisão energias conformacionais e interações intermoleculares envolvendo proteínas, ácidos nucleicos e outras moléculas com grupos funcionais relacionados que são de interesse em química orgânica e biológica. Os valores  $r_{eq}$ ,  $\theta_{eq}$ ,  $k_r$  e  $k_\theta$  são usados com valores iniciais e ajustados conforme necessário para reproduzir frequências

<sup>7</sup> Interações 1-4.

experimentais dos modos normais. Esses valores são inicialmente derivados pelo ajuste aos dados de frequência estrutural e vibracional em pequenos fragmentos moleculares que constituem as proteínas e os ácidos nucleicos. Por exemplo, em fragmentos complexos como as bases de ácido nucleico, os valores  $r_{eq}$  e  $\theta_{eq}$  são retirados de dados estruturais de raios-X, o valor  $k_r$  é determinado por interpolação linear entre valores de ligação simples e dupla usando as distâncias de ligação observadas, e o valor de  $k_\theta$  é obtido da análise vibracional de um átomo  $sp^2$  simples contendo fragmentos como benzeno e N-metiactamida (NMA).

Os parâmetros diédricos são otimizados na molécula mais simples possível e, em seguida, aplicados a moléculas maiores e mais complexas. Esta abordagem está em contraste com a empregada por muitos desenvolvedores de campo de força, onde os parâmetros são otimizados para melhor reproduzir as energias conformacionais de um grande número de moléculas. Uma vantagem da abordagem de Cornell [37] é a falta de dependência dos parâmetros resultantes nas moléculas particulares escolhidas para o conjunto de teste.

### A.3.1 Campos de força implementados no AMBER

Os pacotes computacionais do AMBER [19] são projetados para funcionar com vários tipos simples de campos de força. A parametrização tradicional usa cargas parciais fixas, centradas nos átomos. Modificações menos comumente usadas adicionam dipolos polarizáveis aos átomos, de modo que a descrição da carga depende do ambiente; tais potenciais são chamados de “polarizáveis” ou “não aditivos” [38]. A forma mais simples dos campos de força do AMBER (negligenciando o solvente implícito ou os termos de polarização) usa o Hamiltoniano abaixo:

$$E_{total} = \sum_{\text{ligações}} k_b (r - r_0)^2 + \sum_{\text{ângulos}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{diedrais}} V_n [1 + \cos(n\phi - \gamma)] + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] \quad (10)$$

Nesta equação, os termos  $k_b$ ,  $r_0$ ,  $k_\theta$ ,  $\theta_0$ ,  $V_n$ ,  $\gamma$ ,  $A_{ij}$  e  $B_{ij}$  devem ser especificados nos arquivos de parâmetros para os vários campos de força do AMBER. A diferença para o modelo de Cornell (Eq. 9) é a aplicação do fator multiplicativo no somatório dos ângulos diedrais. É modificado a ordem que o fator é utilizado para reduzir o tempo de execução.

## A.4 Efeitos de solvatação

A solvatação é um dos problemas não resolvidos remanescentes na mecânica molecular. A maioria das implementações contorna a dificuldade assumindo que o modelo está ou na fase gasosa, quando todos os efeitos da solvatação são ignorados, ou em um meio dipolar homogêneo, quando as características topográficas frequentemente assimétricas também são ignoradas. Nesta modelagem assume-se que a molécula está permeada no

meio<sup>8</sup> uniformemente. Mesmo pequenas moléculas de solvente globalmente simétricas só podem interagir com a superfície externa de uma molécula compacta e não com seu interior. Portanto, a solvatação não será um fator em muitas das interações intramoleculares<sup>9</sup> de moléculas compactas. Na realidade, o tratamento eficaz da solvatação deve abranger simultaneamente as complexidades topográficas das interações solvente-soluto e solvente-solvente [35, 39, 40]. A maioria das determinações experimentais de estrutura ou conformação são feitas na fase sólida ou em solução. É conhecido na literatura que a distribuição observada dos estados conformacionais de uma molécula é drasticamente afetada pela solvatação e que a natureza do solvente é de grande importância. Na fase sólida, a ocorrência de mais de uma conformação de uma molécula na célula unitária, embora não generalizada, é bem conhecida e essas diferentes conformações da mesma molécula geralmente têm comprimentos e ângulos de ligação ligeiramente diferentes.

Nesta seção são apresentados os modelos de solvatação mais utilizados nos pacotes *AMBER/AMBERtools*, com uma breve introdução aos modelos de água TIP3P e TIP4P.

#### A.4.1 Equação de Poisson-Boltzmann

A Equação de Poisson-Boltzmann (Eq. 11) (EPB) é dada por [39]:

$$\nabla \cdot [\epsilon(\mathbf{r})\nabla \cdot \phi(\mathbf{r})] - \epsilon(\mathbf{r})\kappa(\mathbf{r})^2 \sinh[\phi(\mathbf{r})] + \frac{4\pi\rho(\mathbf{r})}{k_B T} = 0 \quad (11)$$

onde  $\phi(\mathbf{r})$  é o potencial eletrostático em unidades de  $k_B T/q$  ( $k_B$  é a constante de Boltzmann,  $T$  é a temperatura absoluta e  $q$  é a carga um próton),  $\epsilon$  é a constante dielétrica e  $\rho$  é a densidade de carga fixa (em unidades de carga de prótons).  $\kappa^2 = 1/\lambda^2 = 8\pi q^2 I / ek_B T$ , onde  $\lambda$  é o comprimento de Debye e  $I$  é a força iônica da solução em massa. O segundo termo da Eq. 11 leva em conta os efeitos do sal e está ausente quando nenhum íon móvel está presente no sistema ( $\kappa = 0$ ). Sob esta condição, a Eq. 11 se reduz à equação de Poisson, que por sua vez se reduz à lei de Coulomb quando a constante dielétrica é uniforme em todo o espaço. No entanto, como a água é mais facilmente polarizada por um campo elétrico do que a maioria dos solutos, pelo menos dois valores de  $\epsilon$  são necessários para capturar a física básica das moléculas polares em solução aquosa. O efeito de ter duas constantes dielétricas é expresso na Eq. 11 através da derivada de  $\epsilon(\mathbf{r})$  no primeiro termo, que é diferente de zero apenas onde  $\epsilon(\mathbf{r})$  varia. Esta região corresponde à superfície molecular, onde há uma “descontinuidade” dielétrica entre o soluto  $\epsilon$ -baixo e o solvente  $\epsilon$ -alto.

A superfície molecular é definida como a superfície de contato formada entre o envelope de van der Waals<sup>10</sup> da molécula e as moléculas do solvente. Todas as regiões dentro da superfície são atribuídas a um valor baixo de  $\epsilon$  ( $\sim 2$  a  $4$ ), enquanto as regiões

<sup>8</sup> Solvente.

<sup>9</sup> São as ligações fortes que mantêm os átomos de uma determinada molécula unidos.

<sup>10</sup> Uma superfície baseada nos raios de corte de van der Waals para átomos individuais.

externas são atribuídas ao  $\epsilon$  do solvente (para água,  $\epsilon \sim 80$ ). A distribuição de carga da molécula é geralmente representada na forma de cargas pontuais localizadas em núcleos atômicos. Como as soluções analíticas para a equação de Poisson estão disponíveis apenas para objetos geométricos simples, todas as aplicações da EPB para moléculas em solução exigiam uma suposição simplificadora para a forma do soluto e sua distribuição de carga. Por exemplo, pequenos solutos e proteínas foram tratados como esferas, DNA como cilindros e membranas como planos. No entanto, as soluções numéricas para a EPB tornam possível descrever a forma do soluto em detalhes atômicos, mantendo uma descrição “contínua” simplificada apenas do solvente.

#### A.4.2 Modelos para água - TIP3P e TIP4P

As funções potenciais consideradas para modelos de água geralmente envolvem um monômero de água rígido que é representado por três, quatro ou cinco locais de interação. As interações Coulombianas entre todos os pares intermoleculares de cargas junto com um único termo de Lennard-Jones entre os oxigênios determinam a energia de dimerização<sup>11</sup>  $\epsilon_{mn}$  para os monômeros  $m$  e  $n$  conforme dado pela Eq. 12.

$$\epsilon_{mn} = \sum_i^m \sum_j^n \left[ \frac{A}{r_{OO}^{12}} - \frac{B}{r_{OO}^6} + \frac{q_i q_j e^2}{r_{ij}} \right] \quad (12)$$

Os parâmetros ( $A$  e  $B$ ) são escolhidos para produzir resultados estruturais e energéticos razoáveis na fase gasosa para complexos de água e álcoois. Na década de 80, Impey et. al. [20] reparametrizou esse modelo de 3 pontos para melhor reproduzir a energia e a densidade da água líquida. Essa parametrização é conhecida como potencial TIP3P. Modelos de 4 pontos (TIP4P) também foram investigados. Nesses casos, a carga negativa é movida para fora do oxigênio e em direção aos hidrogênios em um ponto (M) na bissetriz do ângulo  $HOH$ . A Eq. 12 ainda se aplica com um pequeno aumento na complexidade: Dez distâncias são necessárias para avaliar a função em vez de nove para o modelo de 3 pontos. Em comparação com o experimento, os erros nas densidades para as funções potenciais apresentados por Jorgensen et. al. são TIP3P (2%) e TIP4P (0%) [20].

Uma descrição importante da estrutura em um líquido é fornecida pelas funções de distribuição radial  $g_{\alpha\beta}(r)$  e funções de estrutura parcial  $a_{\alpha\beta}(\kappa)$  que podem ser determinadas por experimentos de difração. As duas funções estão relacionadas pela Eq. 13 onde  $j_0(z) = \frac{\sin z}{z}$ ,  $\rho$  é a densidade aparente e  $\kappa = (4\pi/\lambda)\sin\theta$  é a transferência de momento para espalhamento elástico de radiação com comprimento de onda  $\lambda$  em um ângulo  $2\theta$ .

$$a_{\alpha\beta}(\kappa) = 4\pi\rho \int_0^\infty [g_{\alpha\beta}(r) - 1] r^2 j_0(\kappa r) dr \quad (13)$$

<sup>11</sup> É um processo onde duas moléculas de composição química semelhante se unem para formar um único polímero conhecido como dímero.

As três funções de estrutura parcial para água ( $a_{OO}$ ,  $a_{OH}$ ,  $a_{HH}$ ) podem ser obtidas a partir de experimentos de difração de nêutrons em pelo menos três misturas isotópicas diferentes<sup>12</sup>.

### A.4.3 Modelo Generalizado de Solvatação de Born

Em uma simulação de dinâmica, muitas vezes o principal interesse reside no comportamento do soluto ao invés do solvente, pelo menos em nível atômico. Para manter os efeitos do solvente no equilíbrio conformacional do soluto sem o custo computacional associado a milhares de átomos de solvente, uma descrição implícita de solvatação poderia ser usada para substituir a água e os contraíons explicitamente modelados. Nessas simulações, os efeitos de um solvente de alto dielétrico são aproximados por um modelo eletrostático contínuo<sup>13</sup>. O modelo *Generalizado de Born (GB)* [40] usando a aproximação emparelhada de Hawkins [41] que foi implementado nos pacotes *AMBER/AMBERtools* inclui um termo Debye-Huckel para levar em conta efeitos do sal em baixas concentrações:

$$\Delta G^{GB} = -\frac{1}{2} \left( 1 - \frac{e^{\kappa f_{GB}}}{\epsilon} \right) \sum_{i,j} \frac{q_i q_j}{f_{GB}} \quad (14)$$

onde  $q_i$  e  $q_j$  são cargas parciais atômicas,  $\epsilon$  é a constante dielétrica do solvente,  $\kappa$  é o parâmetro de Debye-Huckel e a soma dupla percorre todos os pares de átomos.  $f_{GB}$  é uma função que interpola entre um raio de Born efetivo  $\alpha_i$ , quando a distância  $r_{ij}$  entre os átomos é curta e pelo  $r_{ij}$  a grandes distâncias:

$$f_{GB} = \left[ r_{ij}^2 + \alpha_i \alpha_j e^{-\frac{r_{ij}^2}{4\alpha_i \alpha_j}} \right]^{1/2} \quad (15)$$

O raio de Born efetivo  $\alpha_i$  descreve o quão profundamente enterrada uma carga está no meio de baixo dielétrico (como uma proteína ou ácido nucleico). Ele depende não apenas do raio intrínseco  $\rho_i$  do átomo  $i$ , mas também das posições relativas e raios intrínsecos de todos os outros átomos nas moléculas:

$$\alpha_i^{-1} = \rho_i^{-1} - \sum_{j \neq i} g_{ij} \quad (16)$$

onde  $g_{ij}$  é dado por

$$g_{ij} = \frac{1}{2} \left[ \frac{1}{L_{ij}} - \frac{1}{U_{ij}} + \frac{r_{ij}}{4} \left( \frac{1}{U_{ij}^2} - \frac{1}{L_{ij}^2} \right) + \frac{1}{2r_{ij}} \ln \frac{L_{ij}}{U_{ij}} + \frac{\rho_j^2}{4r_{ij}} \left( \frac{1}{L_{ij}^2} - \frac{1}{U_{ij}^2} \right) \right] \quad (17)$$

<sup>12</sup> Resultados de Monte Carlo (MC) e experimentais são comparados nas figs. 1-12 do artigo [20].

<sup>13</sup> O solvente aquoso é representado como um contínuo de dielétrico 80. A solução da EPB fornece o campo elétrico em todo o espaço, incluindo o campo de reação do solvente, para uma distribuição de carga de soluto e contorno dielétrico fixos [38, 39].

com

$$L_{ij} = \begin{cases} 1, & \text{se } r_{ij} + \rho_j \leq \rho_i, \\ \rho_i, & \text{se } r_{ij} - \rho_j \leq \rho_i < r_{ij} + \rho_j, \\ r_{ij} - \rho_j, & \text{se } \rho_i \leq r_{ij} - \rho_j \end{cases} \quad (18)$$

e

$$U_{ij} = \begin{cases} 1, & \text{se } r_{ij} + \rho_j \leq \rho_i, \\ r_{ij} - \rho_j, & \text{se } \rho_i < r_{ij} + \rho_j \end{cases} \quad (19)$$

O raio intrínseco  $\rho_i$  está relacionado ao raio atômico  $R_i$  pela Eq. 20. Os fatores de sobreposição  $S_i$  correspondem aproximadamente a sobreposição simultânea de três ou mais átomos. Um “offset” geral é incluído para ajustar as magnitudes das energias de solvatação para coincidir com os cálculos de Poisson-Boltzmann.

$$\rho_i = S_i (R_i + b_{offset}) \quad (20)$$

Além do aumento inerente na velocidade da simulação, certos processos também podem ocorrer mais rapidamente nas simulações GB em relação às simulações com solvente explícito. A amostragem aprimorada e convergência rápida são provavelmente devidas à ausência de forças de atrito do solvente e à falta de necessidade de reorganização das moléculas de água em resposta às mudanças conformacionais do soluto. Embora isso possa servir como uma vantagem para impulsionar a convergência, as simulações de dinâmica molecular usando o modelo GB não descrevem mais a escala de tempo ou a cinética desses processos corretamente<sup>14</sup>. Devido à falta de uma caixa de solvente predefinida, as simulações GB também são independentes da forma do sistema. As simulações GB podem, portanto, ser usadas para estudar várias conformações de uma molécula que não podem ser tratadas por simulações de solvente explícitas. Uma vantagem adicional dos modelos contínuos reside na média implícita sobre os graus de liberdade do solvente, de modo que essa média não precisa ser feita por uma amostragem das configurações do soluto explicitamente. Isso simplifica muito os cálculos dos parâmetros termodinâmicos e pode ser usado para examinar as diferenças de energia livre entre as diferentes conformações de uma molécula ou para obter energias livres de ligação. Nesses tipos de cálculos termodinâmicos, “imagens instantâneas” do soluto são extraídas de simulações com solvente explícito, e cálculos separados (GB ou EPB) são realizados para obter as energias livres de solvatação eletrostática. As energias livres de solvatação são então adicionadas às energias da fase gasosa, juntamente com as estimativas da entropia, para chegar aos valores finais de energia livre. A realização das simulações de DM usando GB eliminaria o cálculo separado das energias de solvatação, introduzindo consistência no campo de força usado para simulação e para cálculo termodinâmico.

<sup>14</sup> Para contornar esse problema, os aspectos de fricção da solvatação poderiam ser incluídos por um modelo de Langevin

## A.5 Dinâmica molecular a $pH$ constante

O  $pH$  da solução frequentemente tem um impacto dramático nos sistemas biomoleculares. Isto é devido a mudanças no estado de protonação predominante de grupos tituláveis (principalmente cadeias laterais de certos aminoácidos e terminações de cadeias de peptídeos) conforme o  $pH$  do solvente varia [42, 25]. O estado de protonação de um grupo titulável é determinado pelo  $pH$  do solvente, e a acidez relativa do grupo, medida por seu  $pK_a$ . O  $pK_a$  instantâneo de um determinado grupo é influenciado por seu ambiente eletrostático, que é determinado pela conformação da proteína e pelo estado de protonação de outros grupos tituláveis. O estado de protonação, por sua vez, tem forte efeito na conformação da proteína, principalmente devido às diferenças de carga entre os diferentes estados de protonação. Os estados nativos de algumas proteínas são estáveis apenas em uma faixa estreita de  $pH$  e podem desnaturar completamente em ambientes de  $pH$  extremos.

Como o comportamento biomolecular pode depender fortemente dos estados de protonação de vários resíduos tituláveis, modelos computacionais precisos projetados para tratar tais sistemas devem de alguma forma levar em conta os efeitos do  $pH$ . Embora a abordagem tradicional de atribuir um estado de protonação fixo para cada resíduo titulável no início da simulação ainda seja a abordagem mais comum, vários métodos foram desenvolvidos na tentativa de tratar os efeitos do  $pH$  em biomoléculas de forma mais quantitativa. De particular interesse, a técnica de dinâmica molecular a  $pH$  constante (CpHMD) é um método que aproveita a capacidade da dinâmica molecular atomística clássica, para amostrar o espaço conformacional ao mesmo tempo em que é feito a amostragem dos estados de protonação disponíveis – de acordo com o *ensemble* semi gran-canônico [43]. Dentro da miríade de métodos CpHMD disponíveis, existem duas abordagens fundamentalmente distintas:

1. *Estados contínuos de protonação*: uma “coordenada de titulação” contínua que descreve uma “partícula de titulação” fictícia é introduzida em cada local protonável que é propagado como parte da dinâmica atomística clássica, de acordo com uma força dependente do  $pH$  que atua sobre esta partícula.
2. *Estados discretos de protonação*: usa a dinâmica molecular para propagar as coordenadas espaciais, enquanto ocasionalmente interrompe a dinâmica para tentar mudanças nos estados de protonação dos resíduos tituláveis usando um critério Monte Carlo com Cadeia de Markov (MCCM), veja o Apêndice A.6.

Modelos para solvente implícito - neste caso GB e EPB - obtêm uma média geral dos graus de liberdade do solvente, incorporando instantaneamente os efeitos do relaxamento do solvente em torno das mudanças discretas do estado de protonação. Portanto, os passos

MC que buscam mudanças dos estados de protonação possuem uma probabilidade razoável de sucesso quando o  $pH$  da solução é definido próximo ao  $pK_a$  do grupo titulável. Na presença de moléculas de solvente explícito, no entanto, a orientação do solvente em torno de qualquer resíduo titulável exposto ao solvente resistirá a qualquer mudança de estado de protonação. Em média, a distribuição do solvente tende a resistir às mudanças do estado de protonação, impondo uma barreira da ordem de  $100 \text{ kcal/mol}$  conforme estimado por Swails et. al. [25], tornando difícil a titulação com estados de protonação discretos diretamente em solvente explícito.

### A.5.1 Método de Mongan - dinâmica a $pH$ constante com solvente implícito

O método proposto por Mongan et. al. [42] emprega uma Dinâmica Molecular com Solvatação Generalizada de Born (DMSGb), usando uma amostragem de Monte Carlo periódica dos estados de protonação. Entre cada etapa de Monte Carlo, o sistema evolui de acordo com a DMSGb padrão. Em cada etapa de Monte Carlo, uma região de titulação e um novo estado de protonação para aquele local são escolhidos aleatoriamente. Uma energia livre de transição para a protonação ou desprotonação é calculada de acordo com

$$\Delta G = k_B T (pH - pK_{a,ref}) \ln 10 + \Delta G_{elet} - \Delta G_{elet,ref} \quad (21)$$

onde  $k_B$  é a constante de Boltzmann,  $T$  é a temperatura,  $pH$  é do solvente especificado,  $pK_{a,ref}$  é o  $pK_a$  do composto de referência apropriado (ver Tab. 4<sup>15</sup>),  $\Delta G_{elet}$  é a componente eletrostática da energia livre calculada para o grupo titulável na proteína e  $\Delta G_{elet,ref}$  é a componente eletrostática da energia livre de transição para o composto de referência. Esta equação é baseada em uma divisão da energia livre de transição total em porções eletrostáticas e não eletrostáticas. A energia livre de transição não eletrostática compreende todas as contribuições de energia livre não contabilizadas na eletrostática GB, incluindo a energia livre de ligação da mecânica quântica e a energia livre de solvatação de prótons. A energia livre de transição não eletrostática pode ser assumida como tendo aproximadamente o mesmo valor, independente do ambiente eletrostático. Sob essa suposição, um composto de referência com  $pK_a$  conhecido pode ser introduzido para cancelar a porção não eletrostática da energia livre de transição, resultando na Eq. 21. A porção eletrostática da energia livre de transição ( $\Delta G_{elet}$ ) é calculada tomando a diferença entre o potencial calculado para o estado de protonação atual e o potencial calculado para o estado proposto; como não há necessidade de equilíbrio com solvente, isso é feito em uma única etapa. A Eq. 21 pode então ser usada para calcular a energia livre de transição total. A energia livre de transição total  $\Delta G$  é usada como base para a aplicação do critério de Metrópolis<sup>16</sup> para determinar se a transição será aceita. Se a transição for aceita, a DM continua com o grupo titulável

<sup>15</sup> A His- $\delta$  representa a histidina com o nitrogênio  $\delta$ , do imidazol, protonado. E a His- $\epsilon$  com o nitrogênio  $\epsilon$ , do imidazol, protonado.

<sup>16</sup> O algoritmo Metropolis-Hastings é um método MCMC (veja o Apêndice A.6).

no novo estado de protonação; caso contrário, a DM continua sem nenhuma alteração para o estado de protonação. Sob esse modelo, a carga total na molécula é geralmente diferente de zero e muda quando um grupo titulável muda o estado de protonação. Como a solvatação GB não emprega condições de contorno periódicas e a energia livre associada à introdução e solvatação de uma carga está incluída na parte não eletrostática da energia livre de transição considerada acima, a mudança da carga total não apresenta um grande problema.

Tabela 4 – Adaptado de (Mongan, 2004): Negativo do logaritmo das constantes de dissociação de ácido para as cadeias laterais nas moléculas de aspartato, glutamato, histidina, lisina e tirosina.

Símbolo	Nome	$pK_{a,ref}$
Asp	Aspartato	4,0
Glu	Glutamato	4,4
His- $\delta$	Histidina- $\delta$	6,5
His- $\epsilon$	Histidina- $\epsilon$	7,1
Lys	Lisina	10,4
Tyr	Tirosina	9,6

Computacionalmente, o tempo para avaliar uma etapa de Monte Carlo é menor do que o necessário para uma etapa de dinâmica, portanto, a CpHMD usando esta abordagem é apenas ligeiramente mais lenta do que a DMSGb com estado de protonação constante tradicional.

### A.5.2 Método de Swails - dinâmica a $pH$ constante com solvente Explícito

No método de Swails et. al. para CpHMD [25], estruturas são amostradas a partir do *ensemble* semi gran-canônico [43], cuja função de distribuição de probabilidade é dada por

$$\rho(\mathbf{q}, \mathbf{p}, \mathbf{n}) = \frac{e^{\beta\mu^*n - \beta H(\mathbf{q}, \mathbf{p}, \mathbf{n})}}{\sum_{\mathbf{n}'} \int d\mathbf{p}' d\mathbf{q}' e^{\beta\mu^*n - \beta H(\mathbf{q}, \mathbf{p}, \mathbf{n})}} \quad (22)$$

onde  $\beta = 1/k_B T$ ,  $\mu^*$  é o potencial químico do hidrônio<sup>17</sup> (diretamente relacionado ao  $pH$  da solução),  $q$  é a coordenada generalizada das partículas do sistema,  $p$  é o momento conjugado e  $n$  é o número total de prótons tituláveis presentes naquele estado. Quando em negrito,  $n$  se refere ao vetor de estado de protonação, especificando não apenas o número total de prótons presentes, mas em quais sítios tituláveis esses prótons estão localizados. O denominador na Eq. 22 é a função de partição do conjunto semi gran-canônico.

A dinâmica atomística clássica, que amostra da função de probabilidade condicional das posições e momentos com estados de protonação fixos  $\rho(\mathbf{q}, \mathbf{p}|\mathbf{n})$ , usada em conjunto com os passos MC nos estados de protonação, que amostra da função de probabilidade

<sup>17</sup> Água protonada.

condicional do vetor de estado de protonação em uma conformação de proteína fixa  $\rho(\mathbf{n}|\mathbf{q}, \mathbf{p})$ , amostra apropriadamente da função de distribuição de probabilidade desejada  $\rho(\mathbf{q}, \mathbf{p}, \mathbf{n})$ .

Em solvente explícito,  $\rho(\mathbf{n}|\mathbf{q}, \mathbf{p})$  é difícil de amostrar diretamente, uma vez que a orientação do solvente é relaxada em relação ao vetor de estado de protonação atual. As coordenadas (e momentos) do sistema podem ser separadas em graus de liberdade de soluto e solvente. A amostragem do estado de protonação é então realizada de acordo com a probabilidade condicional

$$\rho' = \rho(\mathbf{p}_{solv}, \mathbf{q}_{solv}, \mathbf{n} | \mathbf{p}_{soluto}, \mathbf{q}_{soluto}) \quad (23)$$

onde  $q_{solv}$  e  $p_{solv}$  são distribuições de solventes relaxadas de posições e momentos em torno do vetor de estado de protonação,  $\mathbf{n}$ . A função de distribuição  $\rho'$  na Eq. 23 pode ser aproximada usando modelos contínuos, como as EPB ou GB, evitando assim cálculos de relaxamento de solvente caros associados a cada tentativa de mudança de estado de protonação. O relaxamento na dinâmica deve ser realizado após as alterações do estado de protonação, para gerar a distribuição de solvente relaxada e não correlacionada exigida pela Eq. 23.

#### A.5.2.1 Fluxograma do método de Swails

O método CpHMD de Swails et. al. pode ser dividido em três etapas que se repetem ao longo de toda dinâmica:

1. A DM em solvente explícito é realizada usando um conjunto constante de estados de protonação (fornecido no começo da simulação). Em algum ponto a DM é interrompida, o solvente (incluindo quaisquer íons não estruturais) é removido e o potencial é alternado para um modelo GB disponível.
2. Após a interrupção, mudanças no estado de protonação para cada resíduo titulável são propostas uma vez em ordem aleatória, escolhendo entre os estados de protonação disponíveis desse resíduo, excluindo o estado atualmente ocupado. A diferença de energia eletrostática entre os estados de protonação propostos e o atual, até mesmo a decisão de aceitar ou não os estados propostos (critério de Metropolis, veja Apêndice A.6), são calculadas da mesma forma que na implementação original da GB. Se a mudança de estado for aceita, o estado atual é modificado de forma adequada, e o próximo resíduo (escolhido aleatoriamente sem substituição) é titulado com este novo estado. Para cada resíduo que é titulado, há uma chance de 25% de que uma titulação multissítio ocorrerá com um resíduo vizinho; isto é, a proposta envolverá mudanças no estado de protonação tanto do resíduo quanto de seu vizinho<sup>18</sup>. A

<sup>18</sup> Dois resíduos tituláveis são “vizinhos” se quaisquer dois átomos de hidrogênio titulados estiverem a 2 Å um do outro.

inclusão de saltos de estado de protonação multissítio é importante para sistemas que possuem resíduos tituláveis em interação próxima. Sem esses movimentos em vários locais, as transferências de prótons entre resíduos tituláveis adjacentes (envolvidos em uma ligação de hidrogênio) nunca ocorreriam devido à alta penalidade de interromper a iteração, pela adição de outro próton ou remoção do próton envolvido na ligação de hidrogênio.

3. Se qualquer uma das tentativas de mudança do estado de protonação for aceita, o soluto é congelado enquanto a DM é realizada no solvente (e quaisquer íons) para relaxar a distribuição do solvente em torno dos novos estados. A duração desse relaxamento  $\tau_{rlx}$  é um parâmetro ajustável do método. Depois que o relaxamento é concluído, as velocidades dos átomos do soluto são restauradas aos seus valores anteriores ao relaxamento e a dinâmica é resumida.

## A.6 Método de Metropolis-Hastings: Uma abordagem Monte Carlo com cadeia de Markov

Esta seção da tese tem o intuito de auxiliar a compreensão do método CpHMD, aos leitores mais focados na implementação das ferramentas computacionais de dinâmica molecular. Esta seção foi colocada aqui pela importância da técnica nas abordagens utilizadas pelas ferramentas do AMBER, e devido a dificuldade de encontrar referências com um linguajar voltado ao público leigo.

Monte Carlo é a prática de estimar as propriedades de uma distribuição de dados examinando amostras aleatórias. Por exemplo, em vez de encontrar a média de uma distribuição normal calculando-a diretamente a partir das equações da distribuição, uma abordagem de Monte Carlo seria extrair um grande número de amostras aleatórias e calcular a média amostral. O benefício da abordagem de Monte-Carlo está na maior facilidade em calcular a média de uma grande amostra do que em determinar a média diretamente das equações da distribuição. Este benefício é mais pronunciado quando as amostras aleatórias são fáceis de extrair ou quando as equações da distribuição são difíceis de trabalhar.

No método de Monte Carlo com Cadeia de Markov (MCCM), a ideia é que as amostras aleatórias são geradas por um processo sequencial especial. Cada amostra aleatória é usada como ponto de partida para gerar a próxima. Uma propriedade especial dessa cadeia é que, embora cada nova amostra dependa da anterior, as novas amostras não dependem de nenhuma anterior à anterior [26]. MCCM é particularmente útil na inferência bayesiana por causa do foco nas distribuições posteriores, que muitas vezes são difíceis de trabalhar por meio de uma verificação analítica. Nestes casos, o método MCCM

permite ao usuário aproximar aspectos de distribuições posteriores que não podem ser calculados diretamente (por exemplo, amostras aleatórias das distribuições posteriores, médias das posteriores, etc.). A inferência bayesiana usa as informações fornecidas pelos dados observados sobre um (conjunto de) parâmetro(s), formalmente a probabilidade, para atualizar um estado anterior de noções sobre um (conjunto de) parâmetro(s) para se tornar um estado posterior de noções sobre um (conjunto de) parâmetro(s). Formalmente, a regra de Bayes é definida como:

$$p(\mu|D) \propto p(D|\mu) \cdot p(\mu) \quad (24)$$

onde  $\mu$  indica um (conjunto de) parâmetro(s) de interesse,  $D$  os dados,  $p(\mu|D)$  a probabilidade de  $\mu$  a cerca dos dados (a probabilidade posterior),  $p(D|\mu)$  a verossimilhança ou probabilidade dos dados  $\mu$  e  $p(\mu)$  indica a probabilidade anterior de  $\mu$ . Nessa equação o símbolo  $\propto$  significa “é proporcional a”.

A forma como os dados são usados para atualizar a noção anterior é examinando a probabilidade dos dados, partindo de um determinado (conjunto de) valor(es) do(s) parâmetro(s) de interesse. Com o objetivo de avaliar essa probabilidade para cada combinação de valores de parâmetros. Quando uma expressão analítica para essa probabilidade está disponível, ela pode ser combinada com a anterior para derivar a posterior analiticamente. Frequentemente, na prática, não se tem acesso a tal expressão analítica. Na inferência bayesiana, esse problema é mais frequentemente resolvido via MCMC: extraindo uma sequência de amostras a partir da posterior e examinando sua média, intervalo e assim por diante.

### A.6.1 Exemplo de amostragem Metropolis-Hastings usando Python

Nesta seção, será usado um espaço de amostragem trivial apenas para esclarecer o critério de Metropolis. Essa abordagem será comparada à amostragem aleatória usual das simulações de Monte Carlo.

Nas Figs. 22-27 é apresentado um script básico em Python (versão 3.6 ou superior) [29] para gerar amostras utilizando os conceitos apresentados previamente, a biblioteca *matplotlib* [44] para geração de gráficos e a biblioteca padrão *random* para a implementação pseudoaleatória. Nas Figs. 22 e 23 está apresentado um método usual de amostragem aleatória, onde os pontos extraídos são números reais no intervalo  $(0, 20)$ <sup>19</sup>.

Nas Figs. 24-27 estão descritas as definições de uma classe filha da “Sampling”. Para facilitar a compreensão do que foi implementado, os métodos serão discutidos fora de ordem. Caso seja de interesse do leitor replicar a execução do código, basta se atentar a numeração das linhas apresentada nas figuras.

<sup>19</sup> Valores *default* podem ou não ser usados.

```

1 import random
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 class Sampling:
6     def __init__(self, a=0, b=20, fator=10.):
7         self.a = a
8         self.b = b
9         self.nome = "Amostragem aleatória"
10
11         #desta maneira aumenta o numero de pontos, sem alterar o intervalo
12         X = [x/fator for x in range(int(self.a*fator),int(self.b*fator))]
13         Y = self.samples(X)
14
15         Ind = np.zeros(len(X))
16         for i in range(len(X)):
17             Ind[i] = int(X[i]*fator)
18
19         self.plot(Ind,Y)

```

Figura 22 – Classe *Sampling* (“Amostragem”) escrita de maneira a criar um gráfico de amostragem aleatória, quando um objeto é instanciado.

```

29     def samples(self, x_r = [1,2,3,4]):
30         random.seed()
31         samp = np.zeros(len(x_r))
32
33         for j in range(len(x_r)):
34             samp[j] = random.uniform(self.a,self.b)
35
36         return samp

```

Figura 23 – Amostragem aleatória usual.

Na Fig. 24, é chamado o método *norm* da biblioteca *Scipy* [45] que representa a função normal de distribuição de probabilidade. Este método é usado no código para amostrar ruídos com facilidade. A definição da linha 40, diz respeito a função de densidade de probabilidade (FDP) usada. Nesse caso a FDP é a distribuição normal e poderia ter usado a própria *norm* declarada na linha 38, mas foi escrito desta maneira para facilitar a troca da FDP.

```

38 from scipy.stats import norm
39
40 def pgauss(esperanca, desv_pad):
41     return lambda x: (np.e**(-( ((x-esperanca)/desv_pad)**2 ))/2)

```

Figura 24 – Definição da função distribuição de probabilidade.

Na Fig. 25, é apresentado um método simples de proposta para a próxima amostra da cadeia de Markov. Como pode ser visto na linha 93, a proposta é definida como a soma entre a amostra anterior (devido a propriedade da cadeia) e um ruído (representando uma pequena flutuação do valor atual). Esta proposta é feita em “loop”, até a verificação escrita nas linhas 94 – 100 garantirem que a amostra pertença ao intervalo de interesse.

Na Fig. 26 é descrito a amostragem MCCM para o exemplo apresentado neste

```

85     def proposta(self, anterior, ruído):
86         t = True
87         Ab = False
88         if self.a > self.b:
89             Ab = True
90         while t:
91             '''norm.rvs faz uma amostragem de tamanho 'size' numa dist.
92             normal centrada em 'loc' com desvio 'scale'. '''
93             novo = anterior + norm.rvs(loc=0, scale=ruído, size=1)
94             if Ab:
95                 #se 'novo' estiver no intervalo, retorne ele
96                 if novo >=self.b and novo <=self.a:
97                     t = False
98             else:
99                 if novo >=self.a and novo <=self.b:
100                    t = False
101
102         return novo

```

Figura 25 – Proposta para a nova amostra usando a propriedade da cadeia de Markov.

apêndice. Nas linhas 74 – 81 é feito um “loop” para gerar todas as amostras. Na linha 75 é gerado uma proposta seguindo o método da Fig. 25. Na linha 77 é calculado a razão de probabilidades entre a amostra proposta e a atual (neste exemplo usando a distribuição normal definida na Fig. 24). Nas linhas 78 – 81 está descrito o critério de Metropolis, se a razão de probabilidades for maior que uma probabilidade aleatória é aceito a proposta. O “bias” dessa amostragem faz com que propostas próximas a média da distribuição sejam mais facilmente aceitas.

```

60     def samples(self, x_r = [1,2,3,4]):
61         self.nome = "Amostragem MCMC"
62         #random.seed()
63         samp = np.zeros(len(x_r))
64         #ruído muito pequeno gera uma amostragem ruim
65         ruído = abs(self.a -self.b)/5.
66
67         #valor inicial
68         samp[0] = 1.
69         #definindo a pdf no centro do intervalo
70         esperanca = (self.a + self.b)/2.
71         desv_pad = abs(self.a -self.b)/4.
72         pdf = pgauss(esperanca, desv_pad)
73
74         for i in range(1, len(x_r)):
75             prop = self.proposta(samp[i-1], ruído)
76             #critério de Metropolis
77             razao_pdf = pdf(prop)/pdf(samp[i-1])
78             if razao_pdf > random.uniform(0,1):
79                 samp[i] = prop
80             else:
81                 samp[i] = samp[i-1]
82
83         return samp

```

Figura 26 – Método de amostragem de Metropolis.

Na Fig. 27, é apresentado um método que compara o gráfico criado com a amostra-

gem aleatória usual, implementado na classe pai (Fig. 23), com o criado pelo método de Metropolis (Fig. 26). Na Fig.28, estão apresentados os resultados da execução do código

```
43 class MCMC_sampling(Sampling):
44     def plot(self, X, Y):
45         Ypai = Sampling.samples(self,X)
46         fig, axs = plt.subplots(2, sharex=True, sharey=True)
47         axs[0].plot(X,Ypai)
48         axs[0].set_title("Amostragem aleatória")
49         axs[1].plot(X,Y)
50         axs[1].set_title("Amostragem de Metropolis")
51         for ax in axs.flat:
52             ax.set(xlabel='Indice', ylabel='Amostra')
53
54         #como sao os mesmo eixos, o comando abaixo esconde as redundancias
55         for ax in axs.flat:
56             ax.label_outer()
57
58         plt.show()
```

Figura 27 – Geração dos gráficos: Amostra vs Índice da amostragem. Escrito de maneira a apresentar as amostragem aleatória usual comparada a de Metropolis.

com trezentas amostras.

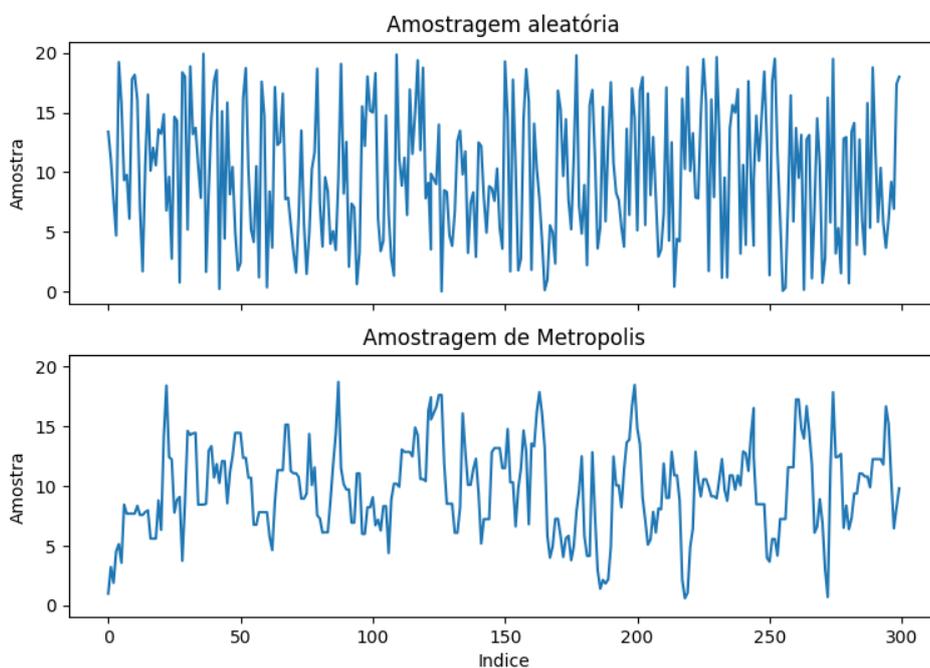


Figura 28 – Comparação entre a amostragem aleatória comum (acima) e a de Metropolis (abaixo), ambas extraindo trezentos pontos.

## A.7 MMPBSA

Um esquema para os cálculos de energia livre de ligação [46] em um complexo receptor-ligante (ligado não covalentemente) é apresentado como ciclos termodinâmicos na Fig. 29. As energias livres de ligação são calculadas subtraindo as energias livres do receptor não ligado e do ligante da energia livre do complexo ligado, como descrito na Eq. 25:

$$\Delta G_{binding,solvated} = \Delta G_{complex,solvated} - (\Delta G_{receptor,solvated} + \Delta G_{ligand,solvated}) \quad (25)$$

A variação de energia livre associada a cada termo do lado direito é estimada de acordo com a Eq. 26:

$$\Delta G_{solvated} = E_{gas} + \Delta G_{solvation} - TS_{solute} \quad (26)$$

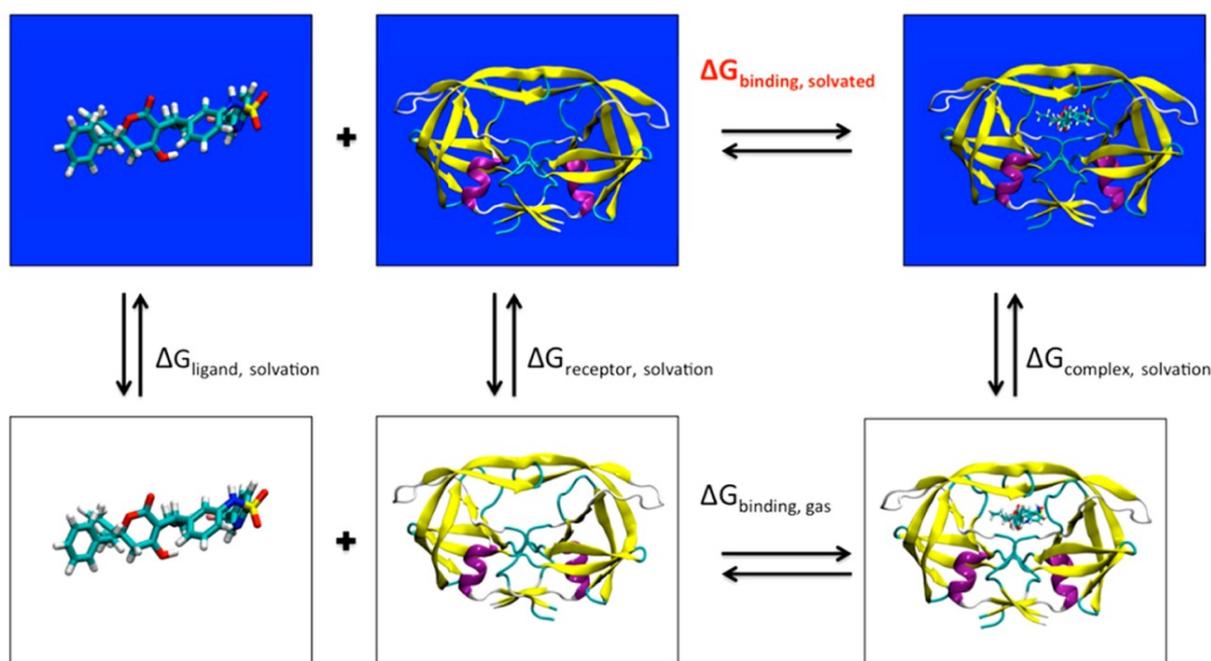


Figura 29 – Ciclos termodinâmicos para cálculos de energia livre de ligação para um complexo proteína-ligante. Os sistemas solvatados estão representados em caixas azuis, enquanto os sistemas na fase gasosa estão em caixas brancas. As energias livres calculadas diretamente são mostradas em preto, enquanto a energia livre de interesse é mostrada em vermelho. Adaptado do artigo (Miller, 2012).

As energias da fase gasosa ( $E_{gas}$ ) são frequentemente oriundas dos campos de força, enquanto as energias livres de solvatação ( $\Delta G_{solvation}$ ) são calculadas usando um modelo de solvente implícito, e a contribuição entrópica (S) é estimada usando aproximações conhecidas. A energia livre de solvatação é ainda decomposta como a soma das contribuições eletrostática e apolar. Vários modelos de solventes implícitos estão disponíveis para calcular as energias livres de solvatação, incluindo o modelo generalizado de Born e o de Poisson-Boltzmann.

As energias descritas nas equações acima são energias de ponto único do sistema. No entanto, na prática, essas energias são estimadas de acordo com as médias de um conjunto de estruturas representativas. Por exemplo, expressar a Eq. 26 em termos de médias resulta em:

$$\Delta G_{solvated} \simeq \langle E_{gas} \rangle + \langle \Delta G_{solvation} \rangle - T \langle S_{solute} \rangle \quad (27)$$

$$= \frac{1}{N} \sum_{i=1}^N [E_{i,gas} + \Delta G_{i,solvation} - T S_{i,solute}] \quad (28)$$

onde  $i$  é o índice de uma instância específica e  $N$  é o número total de instâncias analisados.

Existem duas abordagens para gerar os conjuntos necessários para o estado ligado e não ligado dos cálculos de energia de ligação – todos os conjuntos podem ser extraídos de uma única trajetória de dinâmica molecular ou Monte Carlo do complexo ligado, ou as trajetórias podem ser geradas para cada estado usando simulações separadas. Essas abordagens são chamadas de protocolo *Single Trajectory Protocol (STP)* e *Multiple Trajectory Protocol (MTP)* respectivamente, e cada um tem vantagens e desvantagens distintas. O STP é menos oneroso computacionalmente do que o MTP porque apenas uma única trajetória é necessária para gerar todos os três conjuntos. Além disso, os termos potenciais internos se cancelam exatamente no STP, porque as conformações nos *ensembles* ligados e não ligados são as mesmas, levando a flutuações mais baixas e convergência mais fácil na energia livre de ligação. O STP é apropriado se os conjuntos de receptor e ligante forem comparáveis nos estados ligado e não ligado. No entanto, as conformações que populam os *ensembles* não ligados normalmente adotam configurações tensas quando extraídas do *ensemble* de estado ligado, superestabilizando assim a ligação, em comparação com o MTP.

### A.7.1 Decomposição da Energia Livre

O programa AMBER fornece vários esquemas para decompor energias livres, calculadas em contribuições de resíduos específicos usando modelos de solventes implícitos generalizado de Born ou de Poisson-Boltzmann. As interações podem ser decompostas para cada resíduo, incluindo apenas aquelas interações nas quais um dos átomos do resíduo está envolvido – um esquema denominado de decomposição por resíduo. Alternativamente, as interações podem ser decompostas por pares de resíduos específicos, incluindo apenas aquelas interações nas quais um átomo de cada um dos resíduos analisados está participando – um esquema chamado decomposição de pares.

### A.7.2 Cálculos de entropia

Os modelos de solventes implícitos usados para calcular as energias livres de ligação geralmente negligenciam algumas contribuições para a entropia do soluto. Se assumirmos

que os sistemas biológicos obedecem a um modelo de rotor rígido, podemos calcular as entropias translacional e rotacional usando fórmulas padrão da mecânica estatística e podemos aproximar a contribuição da entropia vibracional usando a aproximação *quase-harmônica* ou o método *nmode* (veja as referências [19, 46] para mais detalhes). Tipicamente, os cálculos *nmode* são computacionalmente exigentes para grandes sistemas, porque eles requerem a minimização de cada instância e a realização de diagonalizações de matriz para obter as frequências vibracionais. Por causa da diagonalização, os cálculos de modo normal escalam aproximadamente como  $(3N)^3$ , onde  $N$  é o número de átomos no sistema. Enquanto a abordagem *quase-harmônica* é menos onerosa computacionalmente, um grande número de instâncias ou membros do *ensemble* são normalmente necessários para extrapolar o limite assintótico da entropia total para cada *ensemble*, o que aumenta o custo computacional da simulação original.

## A.8 Arquivo PDB

O PDB de uma estrutura é uma instância do sistema<sup>20</sup> que foi modelada por um cristalógrafo, utilizando técnicas como por exemplo a difração de raio X. Uma característica importante de considerar ao utilizar um PDB é a sua resolução. Essa quantidade é uma medida do nível de detalhe presente no padrão de difração e no mapa de densidade eletrônica posteriormente calculado. Estruturas de resolução com valores de 1 Å ou menos, são altamente ordenadas e cada átomo do sistema pode facilmente ser visto pelo mapa de densidade eletrônica. Estruturas de resolução com valores de 3 Å ou superior, mostram apenas os contornos básicos da cadeia da proteína e a estrutura atômica deve ser inferida por meio de outras análises<sup>21</sup>.

## A.9 Ferramentas AMBER

A seguir estão descritos brevemente os principais pacotes *AMBER/AMBERtools* utilizados neste trabalho:

1. *pdb4amber*: Analisa os arquivos PDB e os limpa para uso posterior, especialmente com os programas LEaP do AMBER.
2. *sqm*: Este é o código extraído das partes de *Quantum Mechanics/Molecular Mechanics (QM/MM)* do *sander*, mas é limitado a cálculos quânticos. Atualmente é limitado a cálculos de ponto único e minimizações de energia (otimizações de geometria). Ele suporta uma grande variedade de hamiltonianos semi-empíricos.

<sup>20</sup> Registro do estado do sistema, em determinado ponto no tempo.

<sup>21</sup> Veja o guia pelo site [rcsb.org](http://rcsb.org) para um exemplo com quatro resoluções e maiores detalhes sobre esta propriedade.

3. *LEaP*: LEaP é o nome genérico dado aos programas tLeap e xLeap. Esses dois programas compartilham uma linguagem de comando comum, mas o programa xleap foi aprimorado com a adição de uma interface gráfica de usuário X-windows. O LEaP pode ser usado para preparar dados para os programas AMBER de mecânica molecular e é a ferramenta básica para construir arquivos de campo de força.
4. *ParmEd*: É um editor de arquivo de topologia escrito em Python que permite o controle de alto nível do arquivo de campo de força primário no Amber: o arquivo prmtop. O ParmEd modifica o arquivo de topologia produzindo um novo, que funciona com os programas sander, pmemd e NAB e fornece opções indisponíveis de outra forma.
5. *Sander/Pmemd*: É um módulo Amber que realiza minimização de energia, dinâmica molecular e refinamentos de *Nuclear Magnetic Resonance (NMR)*. Algumas das funcionalidades do sander estão disponíveis com melhor desempenho computacional no módulo pmemd disponibilizado apenas na versão paga do AMBER.

## B Conceitos Básicos sobre Enzimas e o Modelo de Ge

Enzimas pertencem a uma família bioquímica de macromoléculas conhecidas como proteínas. A característica comum de todas as proteínas é a estrutura polipeptídica, composta por uma sequência linear de  $\alpha$ -aminoácidos unidos por ligações amida. Os conceitos aqui presentes foram resumidos do livro “*Introduction to Enzyme and Coenzyme Chemistry*” de 2004 de T. Bugg [47]. Neste apêndice será apresentado brevemente conceitos fundamentais sobre enzimas, desde as estruturas dos aminoácidos até a estrutura terciária e uma rápida discussão sobre a função catalítica. Ao final do apêndice, discutimos brevemente um modelo para o mecanismo de reação da IsPETase.

### B.1 $\alpha$ -aminoácidos

As proteínas são compostas de uma família de vinte unidades estruturais,  $\alpha$ -aminoácidos, que são moléculas quirais, ou seja, são distinguíveis de suas imagens espelhadas<sup>1</sup>. A estrutura geral de um  $\alpha$ -aminoácido é apresentada na Fig. 30. Cada  $\alpha$ -aminoácido pode ser encontrado como o isômero L ou D, dependendo da configuração no átomo de carbono alfa (exceto para glicina onde R=H). Todas as proteínas são compostas apenas de L-aminoácidos, consequentemente as enzimas são moléculas inerentemente quirais. Os D-aminoácidos são raros em sistemas biológicos, embora sejam encontrados em vários produtos naturais e notavelmente na camada de peptidoglicano das paredes celulares bacterianas.

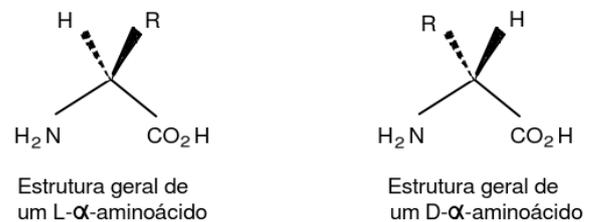
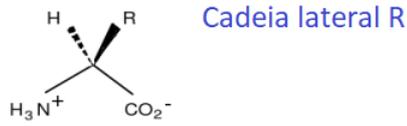


Figura 30 – Estrutura geral de  $\alpha$ -aminoácidos L e D.

Os  $\alpha$ -aminoácidos usados para formar as proteínas somam apenas 20 (veja a Fig. 31), onde as diferenças entre estes vinte estão na natureza da cadeia lateral R. Várias cadeias laterais são de caráter hidrofóbico, por exemplo, o tioéter da Metionina (MET); as cadeias laterais alifáticas ramificadas da Valina (VAL), Leucina (LEU) e Isoleucina (ILE); e as cadeias laterais aromáticas de Fenilalanina (PHE) e Triptofano (TRP). O restante das cadeias laterais de aminoácidos são de caráter hidrofílico [47]. A natureza da cadeia lateral confere certas propriedades físicas e químicas ao aminoácido correspondente e à cadeia

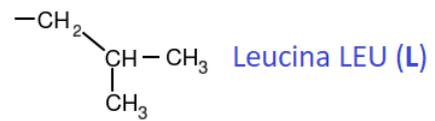
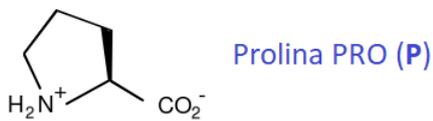
<sup>1</sup> A imagem espelhada de um objeto aquiral, como uma esfera, não pode ser distinguida do objeto.

### Estrutura Geral



### Cadeia Lateral

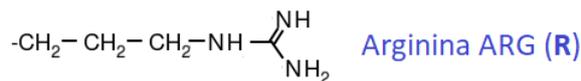
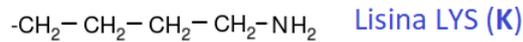
#### Alifática



#### Ácida



#### Básica



#### Polar



#### Aromática

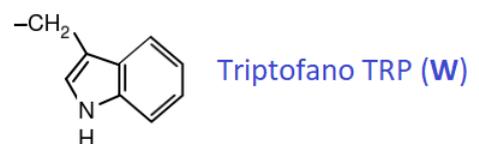
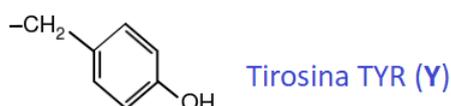


Figura 31 – As cadeias laterais dos vinte aminoácidos encontrados em proteínas.

polipeptídica na qual está localizado.

## B.2 A estrutura primária dos polipeptídeos

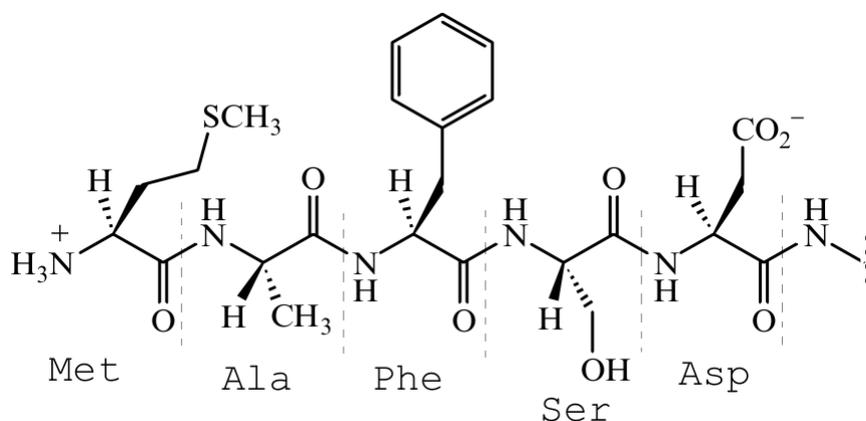


Figura 32 – N-terminal de uma cadeia polipeptídica.

Para formar a cadeia polipeptídica encontrada nas proteínas, cada aminoácido é ligado ao próximo por meio de uma ligação amida, formando uma sequência linear de centenas a milhares de aminoácidos – esta é a estrutura primária da proteína. Uma porção da extremidade amino-terminal (ou N-terminal) de um polipeptídeo é apresentada na Fig. 32, junto com as representações abreviadas para esta sequência polipeptídica. As sequências de aminoácidos contêm todas as informações para conferir a estrutura tridimensional das proteínas em geral e a atividade catalítica das enzimas em particular.

## B.3 Estruturas secundárias encontradas em proteínas

Quando a sequência polipeptídica linear da proteína é formada dentro das células pelos ribossomos, a cadeia polipeptídica se dobra espontaneamente para formar a estrutura tridimensional única e estável da proteína. Fatores que podem ser importantes no processo de dobra são:

1. Empacotamento das cadeias laterais de aminoácidos hidrofóbicos e exclusão de água (solvente);
2. Formação de interações não covalentes específicas;
3. Formação de estruturas secundárias.

Estrutura secundária é o termo dado às regiões locais (de 10 – 20 aminoácidos) de estruturas tridimensionais estáveis e ordenadas mantidas juntas por ligações de hidrogênio, ou seja, ligações não covalentes entre hidrogênios ácidos (O–H, N–H) e pares solitários.

Existem pelo menos três formas estáveis de estrutura secundária comumente observadas em proteínas: a  $\alpha$ -hélice, a folha beta e a volta beta. A  $\alpha$ -hélice é uma estrutura helicoidal formada por uma única cadeia polipeptídica, na qual, ligações de hidrogênio são formadas entre a carbonila e o H da ligação amida quatro resíduos à frente na cadeia polipeptídica, como mostrado na Fig. 33. Nesta estrutura, cada ligação amida forma duas ligações de

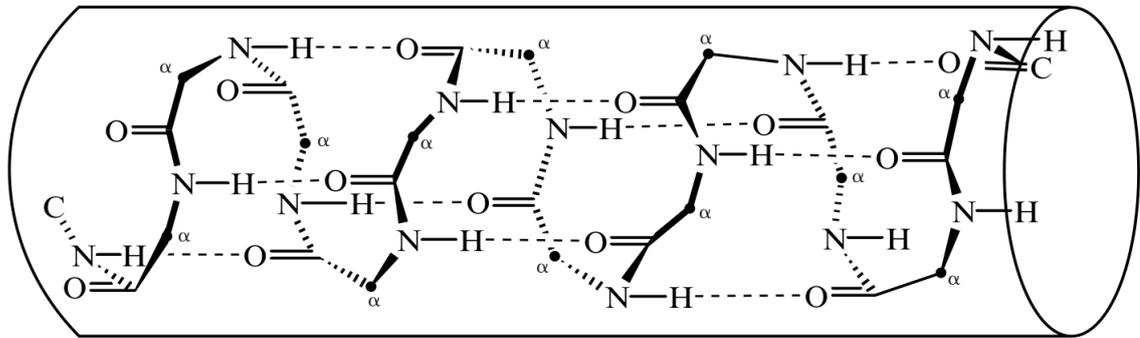


Figura 33 – Estrutura de uma  $\alpha$ -hélice, com as posições dos carbonos alfa indicadas por pontos.

hidrogênio específicas, tornando-a uma unidade estrutural muito estável. Todas as cadeias laterais de aminoácidos apontam para fora da altura da hélice, conseqüentemente as cadeias laterais de aminoácidos que estão separadas por quatro resíduos na sequência primária acabarão próximas (espacialmente). As interações entre essas cadeias laterais podem levar a outras interações favoráveis dentro da hélice ou com outras estruturas secundárias.

A folha beta é uma estrutura formada por duas ou mais fitas polipeptídicas lineares, mantidas juntas por uma série de ligações de hidrogênio entre as fitas. Existem dois tipos de estruturas de folha beta: folhas beta paralelas, nas quais as cadeias de peptídeos procedem na mesma direção amino-para-carboxila; e antiparalelo, em que as fitas procedem em direções opostas. Ambos os tipos estão ilustrados na Fig. 34. Por fim, a volta beta é uma

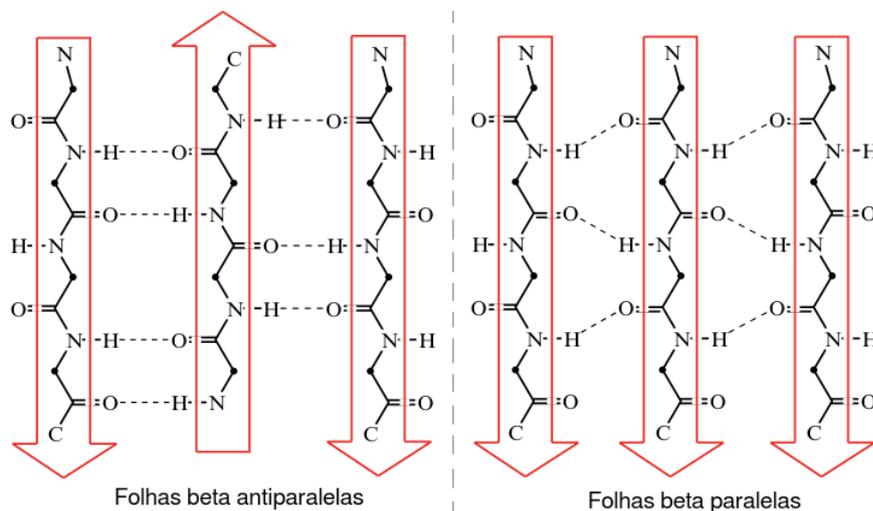


Figura 34 – Estrutura das folhas beta.

estrutura frequentemente formada no final de uma folha beta que leva a uma curva de  $180^\circ$  na direção da cadeia peptídica.

## B.4 A estrutura terciária dobrada das proteínas

A estrutura tridimensional das subunidades proteicas, conhecida como estrutura terciária, surge do empacotamento de estruturas secundárias para formar uma conformação global estável, que no caso das enzimas é cataliticamente ativa. Esse empacotamento geralmente envolve enterrar as cadeias laterais de aminoácidos hidrofóbicos no interior da proteína e posicionar as cadeias laterais de aminoácidos hidrofílicos na superfície. Embora em teoria o número de estruturas terciárias de proteínas possíveis seja virtualmente infinito, na prática as proteínas são frequentemente compostas de padrões estruturais comuns, a partir dos quais a estrutura da proteína pode ser categorizada como:

- *Proteínas alfa-helicoidais*: são constituídas apenas por  $\alpha$ -hélices que se agrupam para formar a estrutura terciária;
- *Estruturas  $\alpha/\beta$* : consistem em formações regulares de folha beta–alfa hélice–folha beta paralela;
- *Estruturas  $\beta$  anti-paralelas*: consistem em conjuntos regulares de folha beta–volta beta–folha beta anti-paralela.

As proteínas geralmente consistem em mais de uma estrutura terciária, que se encaixam para formar a “estrutura quaternária” ativa.

## B.5 Função catalítica

A diferença entre enzimas e outras proteínas é a atividade catalítica enzimática. A parte da estrutura terciária responsável pela atividade catalítica é chamada de “sítio ativo” da enzima e muitas vezes representa apenas 10 – 20% do volume total da enzima. O sítio ativo é geralmente uma fenda ou cavidade hidrofílica, contendo uma formação de cadeias laterais de aminoácidos que se ligam ao substrato e iniciam a reação enzimática, como representado na Fig. 35-a. Em alguns casos, o sítio ativo da enzima também se liga a um ou mais cofatores que auxiliam na catálise de tipos específicos de reações enzimáticas, como mostrado na Fig. 35-b.

## B.6 Um modelo termodinâmico para catálise

A catálise de uma reação química é alcançada reduzindo a energia de ativação para aquela reação, a energia de ativação sendo a diferença na energia livre entre o(s)

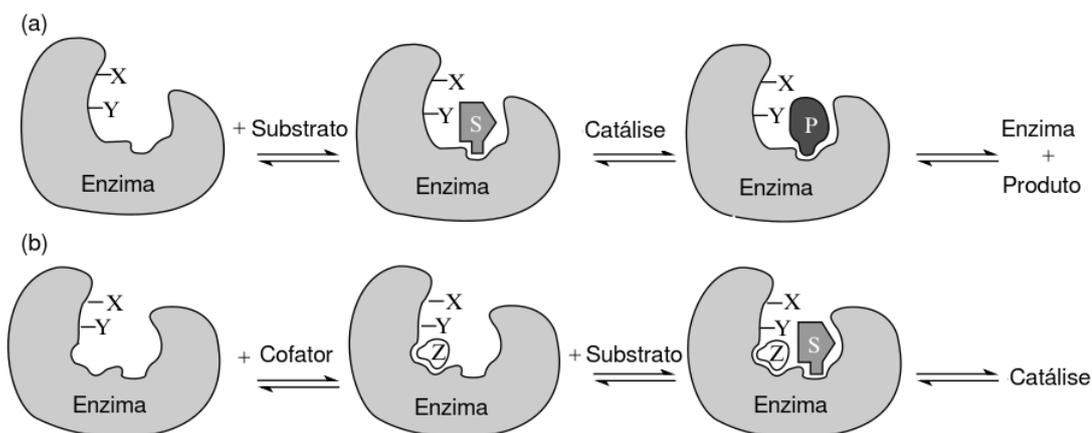


Figura 35 – Figura esquemática da (a) enzima e o substrato e (b) enzima, substrato e cofator.

reagente(s) e o estado de transição para a reação. A Fig. 36-a ilustra o perfil de energia livre da catálise por ácido de uma reação química, que converte um substrato (S) em um produto (P). Neste caso, uma espécie química intermediária  $\text{SH}^+$  é formada após a protonação de S. Se a conversão de  $\text{SH}^+$  em  $\text{PH}^+$  é mais fácil do que a conversão de S em P, então a energia de ativação para a reação será reduzida e, portanto, a reação será mais rápida. Um estado intermediário é uma espécie química estável (ou semi-estável) formada durante a reação e é, portanto, um mínimo de energia local, enquanto um estado de transição é, por definição, um máximo de energia local.

Na Fig. 36-b, um intermediário enzima-substrato (ES) é formado após a ligação do substrato, que é então convertido no complexo enzima-produto (EP) diretamente ou por meio de um ou mais intermediários adicionais. O que no caso não enzimático era uma reação intermolecular tornou-se efetivamente uma reação intramolecular. As reações intramoleculares geralmente ocorrem muito mais rapidamente e sob condições de reação muito mais suaves do que suas contrapartes intermoleculares, uma vez que os dois grupos reagentes já estão próximos um do outro.

## B.7 Catálise ácida e básica em reações enzimáticas

A catálise ácida e básica está envolvida em todos os processos enzimáticos que envolvem a transferência de prótons, existem poucas enzimas que não possuem grupos catalíticos ácidos ou básicos em seus sítios ativos. No entanto, ao contrário das reações orgânicas que podem ser realizadas sob uma ampla gama de condições de pH para se adequar à reação, as enzimas têm uma limitação estrita de que devem operar em pH fisiológico, na faixa de 5–9.

A catálise ácida ocorre quando o substrato é protonado por um resíduo catalítico como mostrado na Fig. 37. O grupo ácido do sítio ativo deve, portanto, ser protonado em

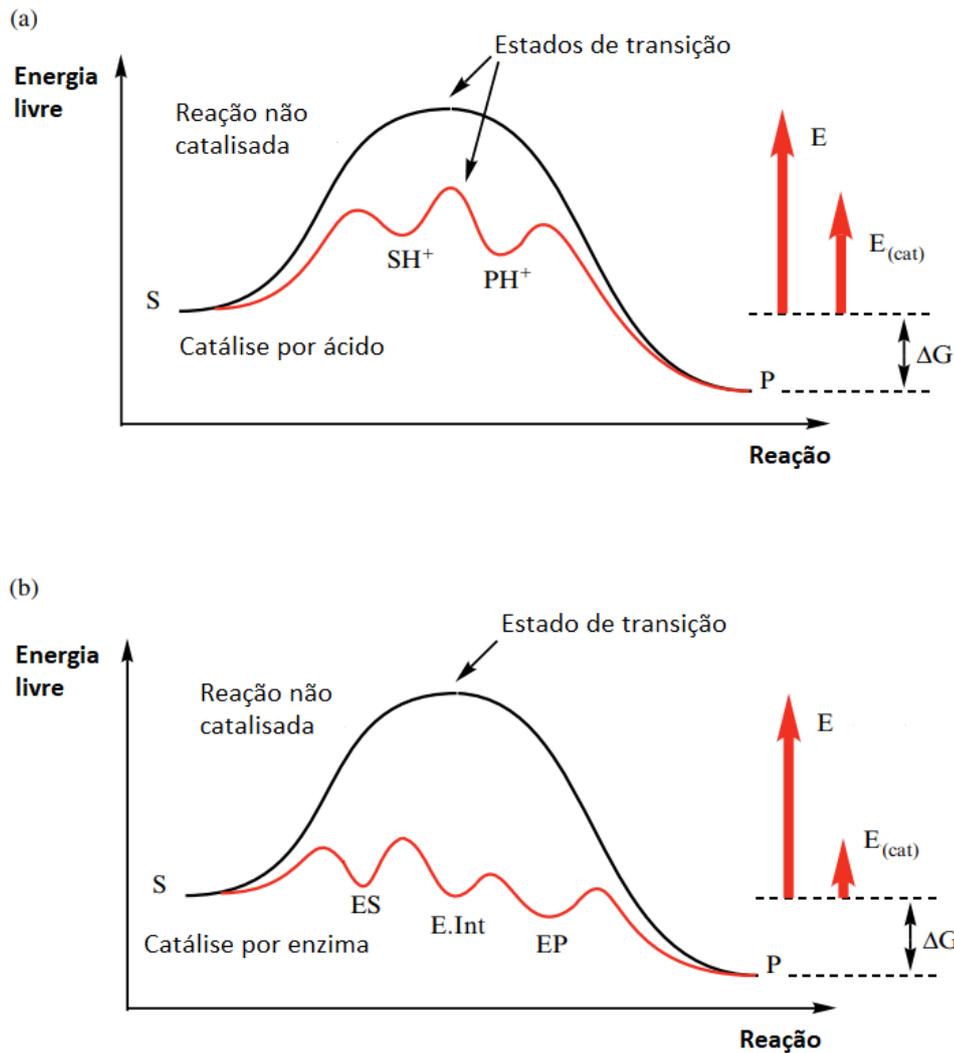


Figura 36 – Perfis de energia livre para (a) uma reação catalisada por ácido e (b) uma reação catalisada por enzima que converte o substrato S no produto P. Adaptado do livro (Bugg, 2004).

pH fisiológico, mas seu  $pK_a$  deve estar logo acima (no intervalo de 7 a 10). Se o  $pK_a$  de uma cadeia lateral fosse maior que 10, então seria termodinamicamente desfavorável para transferência do próton.

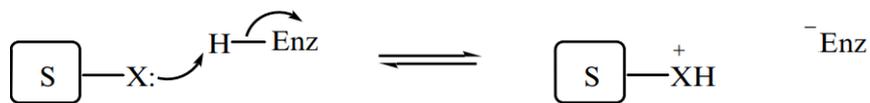


Figura 37 – Catálise ácida. Adaptado do livro (Bugg, 2004).

A catálise básica ocorre quando o substrato é desprotonado ou quando a água é desprotonada antes do ataque ao substrato (Fig. 38). Grupos básicos de sítios ativos devem, portanto, estar desprotonados em pH fisiológico, mas com valores de  $pK_a$  logo abaixo. Intervalos típicos de  $pK_a$  para cadeias laterais de aminoácidos em sítios ativos de enzimas estão mostrados na Fig. 39. Embora os valores de  $pK_a$  dados na Fig. 39 sejam os

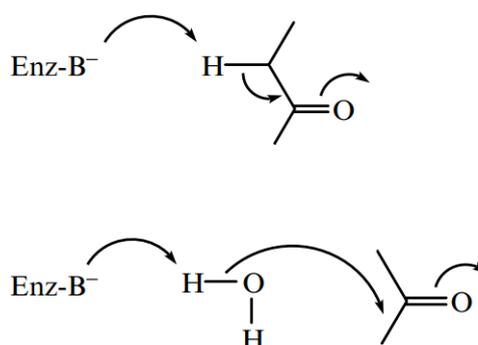


Figura 38 – Catálise básica. Retirado do livro (Bugg, 2004).

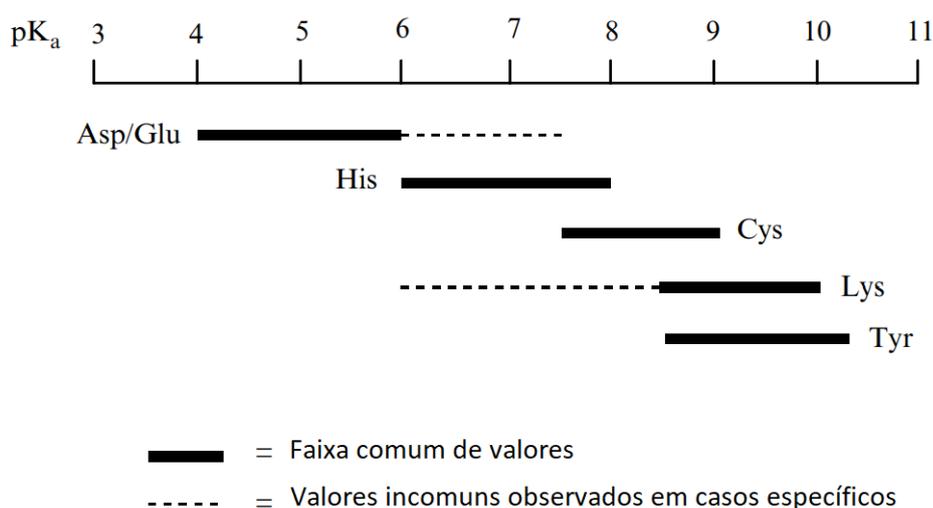


Figura 39 – Faixa de valores de  $pK_a$  observados para cadeias laterais de aminoácidos em sítios ativos de enzimas. Adaptado do livro (Bugg, 2004).

valores típicos encontrados em proteínas, em alguns casos os valores de  $pK_a$  dos grupos ácidos e básicos do sítio ativo podem ser fortemente influenciados por seus ambientes.

É importante notar que a histidina, cuja cadeia lateral contém um anel de imidazol de  $pK_a$  6–8, pode atuar como um resíduo ácido ou básico, dependendo de seu  $pK_a$  local particular, tornando-se um reagente versátil para química enzimática ácida/básica.

## B.8 Catálise nucleofílica em reações enzimáticas

A catálise nucleofílica envolve um ataque nucleofílico de um grupo do sítio ativo no substrato, formando uma ligação covalente entre a enzima e o substrato e, portanto, um intermediário covalente no mecanismo de reação. Esta é uma estratégia particularmente eficaz para reações catalisadas por enzimas por duas razões. Em primeiro lugar, uma enzima é capaz de posicionar um nucleófilo<sup>2</sup> do sítio ativo nas proximidades (com o

<sup>2</sup> Uma espécie química que doa um par de elétrons para formar uma ligação química.

alinhamento correto) e atacar o substrato, gerando uma concentração efetiva muito alta de nucleófilos. Em segundo lugar, uma vez que os sítios ativos da enzima são frequentemente excluídos das moléculas de água, é provável que um nucleófilo do sítio ativo da enzima seja dessolvatado. Um nucleófilo carregado em solução aquosa seria rodeado por várias camadas de moléculas de água, o que reduz bastante a polaridade e a eficácia do nucleófilo. No entanto, um nucleófilo dessolvatado em um sítio ativo excluído de água será um nucleófilo muito mais potente do que sua contraparte em solução.

Tabela 5 – Cadeias laterais de aminoácidos usadas para catálise nucleofílica em reações enzimáticas. Adaptado do livro (Bugg, 2004)

Aminoácido	Exemplos
Serina	Serino proteases, esterases, lipases
Treonina e histidina	Fosfotransferases
Cisteína	Proteases de cisteína, aciltransferases
Aspartato e glutamato	Epóxido hidrolase, haloalcano desalogenase
Lisina	Acetoacetato descarboxilase, aldolases classe I
Tirosina	Topoisomerases de DNA

As enzimas têm uma gama de potenciais nucleófilos à sua disposição, que são apresentados na Tab. 5. Provavelmente, o melhor nucleófilo disponível para as enzimas é o grupo tiol da cadeia lateral da cisteína. O grupo  $\epsilon$ -amino da lisina é usado em vários casos para formar ligações com grupos cetona em substratos. Outro nucleófilo disponível para as enzimas é o versátil anel imidazol da histidina. Este grupo é mais frequentemente usado para química ácido/base, mas ocasionalmente é usado como um nucleófilo em, por exemplo, reações de fosfotransferência. Finalmente, as enzimas têm nucleófilos de oxigênio disponíveis na forma de grupos hidroxila de serina, treonina e tirosina, e os grupos carboxilato de aspartato e glutamato.

## B.9 Modelo de Ge: o mecanismo de reação da IsPETase

Características estruturais importantes sobre a atividade catalítica da IsPETase estão apresentadas na Fig. 40. Estas incluem (i) a tríade catalítica S160-H237-D206 (em laranja), (ii) a cavidade de oxiânion formada por Y87 e M161 (em azul) e (iii) a interação  $\pi - \pi$  estabilizadora entre o W185 (em rosa) e o anel de benzeno do BHET adjacente a carbonila sob ataque. Com base em experimentos, Ge propôs um modelo físico do mecanismo de reação da enzima [14], para investigar computacionalmente a atividade da IsPETase e suas variantes.

O modelo contém quatro estados que estão envolvidos no processo de degradação do PET, incluindo (a) um estado com o substrato livre, (b) um estado ligado ao substrato, (c) um estado intermediário da enzima e (d) um estado com o substrato degradado, como

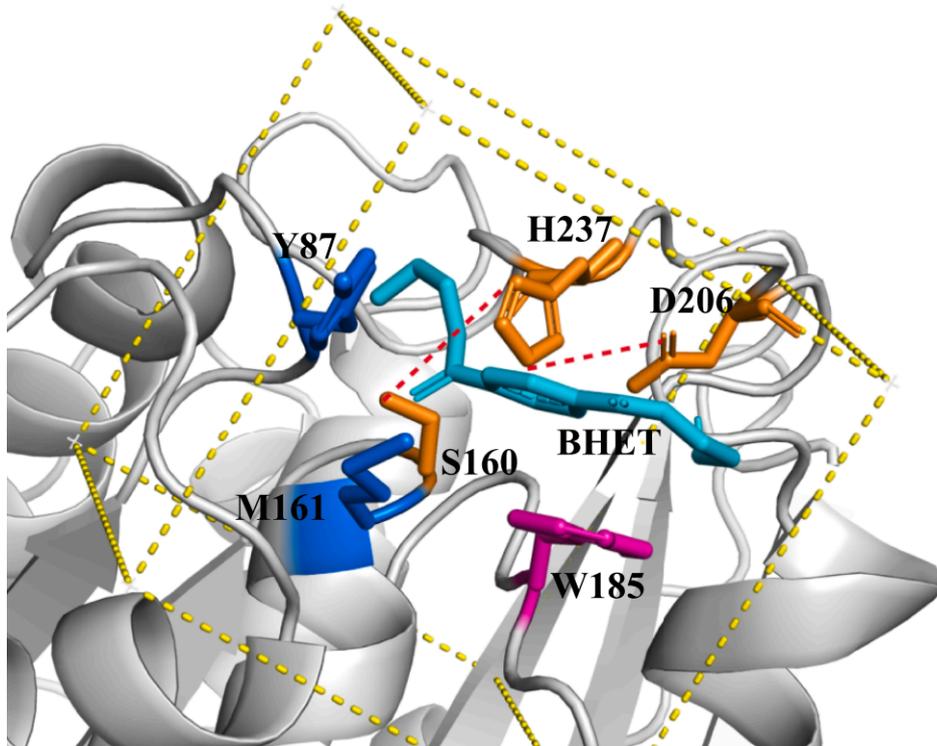
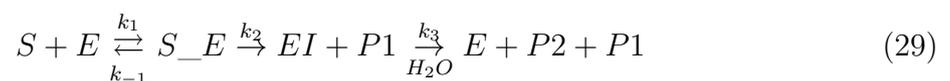


Figura 40 – Representação 3D do local de confinamento do substrato (BHET). Os aminoácidos principais são: a tríade catalítica S-H-D (em laranja), a cavidade de oxianion formada por Y87 e M161 (em azul) e a interação  $\pi - \pi$  de W185 com o anel de benzeno do BHET (em rosa). Adaptado do artigo (Ge, 2021).

mostrado na Fig. 6. Na primeira etapa, o processo de ligação/desligamento (a)  $\leftrightarrow$  (b), é termodinamicamente reversível com uma taxa de ativação  $k_1$  e uma taxa de desativação  $k_{-1}$ . A segunda etapa, o processo (b) a (c), consiste em um ataque nucleofílico pelo grupo hidroxila da S160 no grupo carbonila do substrato, formando o estado intermediário e liberando o produto 1 com uma taxa constante  $k_2$ . Na terceira etapa, o processo (c) a (d), o oxigênio da água realiza um segundo ataque nucleofílico a carbonila do intermediário, liberando a enzima e o produto 2 com uma taxa constante  $k_3$ . O mecanismo de reação da IsPETase pode ser resumido pela Eq. 29.



Ao aplicar a lei de ação de massa<sup>3</sup> à cinética da enzima na Eq. 29, obtém-se as

<sup>3</sup> A taxa de uma reação elementar é proporcional ao produto das concentrações das moléculas participantes.

seguintes equações diferenciais não lineares acopladas:

$$\frac{d[P1]}{dt} = k_2[S\_E] \quad (30)$$

$$\frac{d[P2]}{dt} = k_3[EI] \quad (31)$$

$$\frac{d[S\_E]}{dt} = k_1[S][E] - (k_{-1} + k_2)[S\_E] = 0 \quad (32)$$

$$\frac{d[EI]}{dt} = k_2[S\_E] - k_3[EI] = 0 \quad (33)$$

onde  $[S]$ ,  $[E]$ ,  $[S\_E]$ ,  $[EI]$ ,  $[P1]$  e  $[P2]$  são, respectivamente, as concentrações correspondentes para o substrato, a enzima, o estado ligado enzima-substrato, o estado intermediário, produto 1 e produto 2. Ge assumiu estados estáveis de  $[S\_E]$  e  $[EI]$ , implicando que essas concentrações permanecem constantes durante a reação, conforme as Eqs. 32 e 33. A soma de todas as formas da enzima é constante em relação ao tempo, ou seja,  $[E] + [S\_E] + [EI] = [E]_0$ . Resolvendo este conjunto de equações diferenciais, deriva-se a taxa de reação da seguinte forma:

$$\frac{d[P1]}{dt} = \frac{d[P2]}{dt} = -\frac{d[S]}{dt} = \frac{k_{cat}[E]_0[S]}{K_M + [S]} \quad (34)$$

onde  $k_{cat} = k_2k_3/(k_2 + k_3)$  e  $K_M = k_3(k_{-1} + k_2)/\{k_1(k_2 + k_3)\}$ . A Eq. 34 é conhecida como a equação de Michaelis-Menten [48] e tem uma solução de forma fechada,

$$[S](t) = K_M W \left( \frac{[S]_0}{K_M} \exp \left( \frac{-k_{cat}[E]_0 t + [S]_0}{K_M} \right) \right) \quad (35)$$

onde  $W$  é a função ômega que satisfaz  $W(x)\exp(W(x)) = x$ . Esta solução pode ser verificada prontamente substituindo-a na Eq. 34, que é válida após chegar a um estado estacionário do estado ligado enzima-substrato e o intermediário. Para a equação de Michaelis-Menten na Eq. 34, a eficiência catalítica da enzima pode ser calculada como  $k_{cat}/K_M = k_2k_1/(k_2 + k_{-1}) \approx k_2(k_1/k_{-1})$  para  $k_2 \ll k_{-1}$ .

As constantes de taxa de reação na eficiência catalítica podem ser obtidas modelando as duas primeiras etapas. Na primeira etapa, o substrato é ligado à proximidade do sítio ativo da enzima (o nucleófilo N da enzima) com uma constante de associação

$$K_a \equiv \frac{k_1}{k_{-1}} = \exp(-\Delta G/RT) \quad (36)$$

onde  $\Delta G$  é a energia livre de ligação de Gibbs (mudança de energia livre de Gibbs do estado de substrato livre para o estado ligado do substrato),  $R = 1,987 \times 10^{-3} \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$  é a constante do gás e  $T = 300\text{K}$  é a temperatura ambiente. Na segunda etapa, o nucleófilo ativado da enzima ataca o carbono da carbonila e forma um estado intermediário covalentemente ligado. Ao considerar a taxa de transferência de elétrons devido ao tunelamento quântico, a eficiência do primeiro ataque nucleofílico pode ser modelada como:

$$k_2 = \kappa \cdot \exp(-\beta r) \quad (37)$$

onde a taxa de transferência de elétrons  $\kappa$  depende da nucleofilicidade (de N),  $\beta$  é a constante de decaimento do tunelamento e  $r$  é a distância mais curta entre o nucleófilo e os carbonos de carbonila. Com base na mecânica quântica, para um elétron ligado com potencial de ionização  $U_0$ , a constante de decaimento  $\beta$  pode ser estimada como  $\sqrt{8mU_0/\hbar^2}$ , que é  $1 \text{ \AA}^{-1}$  se  $U_0$  for  $1 \text{ eV}$ . Em princípio,  $U_0$  pode ser estimado pelo *gap* de energia entre o orbital molecular ocupado mais alto (HOMO) do nucleófilo e o orbital molecular não ocupado mais baixo (LUMO) do eletrófilo.

## C Material Suplementar

### C.1 *Docking* molecular: Descrição detalhada

Como descrito na Sec. 3.6, é necessário escolher um espaço de busca para o AutoDock Vina gerar os modelos de encaixe do ligante. A Fig. 41 apresenta uma representação do espaço de busca cúbico que utilizamos. A posição exata desse cubo depende de como definimos o centro do espaço. Como o interesse era encaixar o ligante no sitio ativo com uma conformação favorável ao ataque da serina nucleofílica, escolhemos um cubo  $10 \text{ \AA} \times 10 \text{ \AA} \times 10 \text{ \AA}$  – representado na Fig. 41 como a região esverdeada – centralizado na hidroxila desse nucleófilo (SER 160), para então escolher o centro e definir o espaço de busca. Desse modo, amostrar pontos quaisquer do cubo interno várias vezes, gera superfícies de contato ligeiramente distintas. Com esta metodologia podemos gerar centenas de conformações de encaixe plausíveis utilizando níveis baixos de exaustividade no AutoDock Vina. Em relação

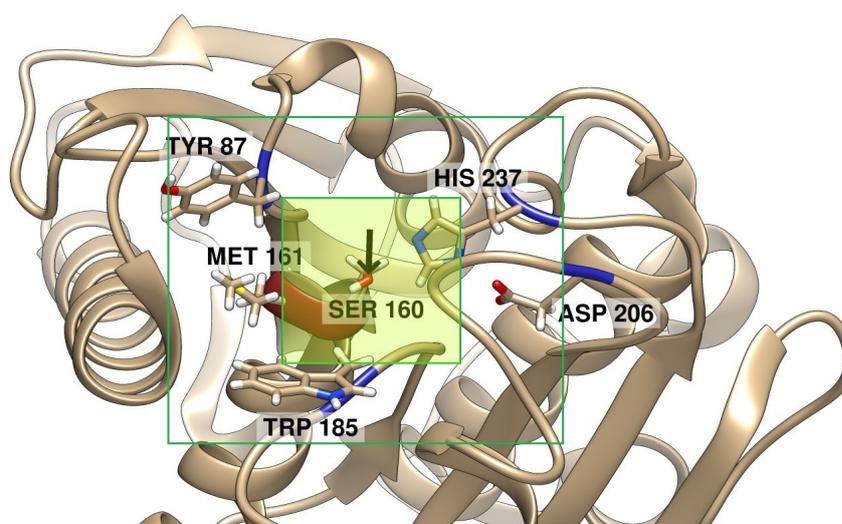


Figura 41 – Esboço do espaço de busca e região de amostragem do ponto central.

a duração deste procedimento, utilizando o processador *Intel(R) Core(TM) i5-9300H CPU @ 2.40GHz* e executando o Vina com 400 espaços de busca aleatórios, obtemos de maneira consistente cerca de 3500 conformações num período de três a quatro horas por complexo variante-BHET.

Uma vez obtidos esses dados, precisamos encontrar a melhor dentre essas milhares de conformações. Como discutido na Sec. 4.2.1, definimos o fator de minimização  $s_{v+r}$  que depende da afinidade calculada e da distância nucleofílica. Como apresentado na Fig. 42, devido a simetria do BHET temos duas possibilidades para o ataque da serina e não temos nenhuma garantia que o ligante se posicionará sempre na mesma direção, favorecendo uma das distâncias  $r_\alpha$  ou  $r_\beta$ . Por isso, precisamos verificar dois valores de

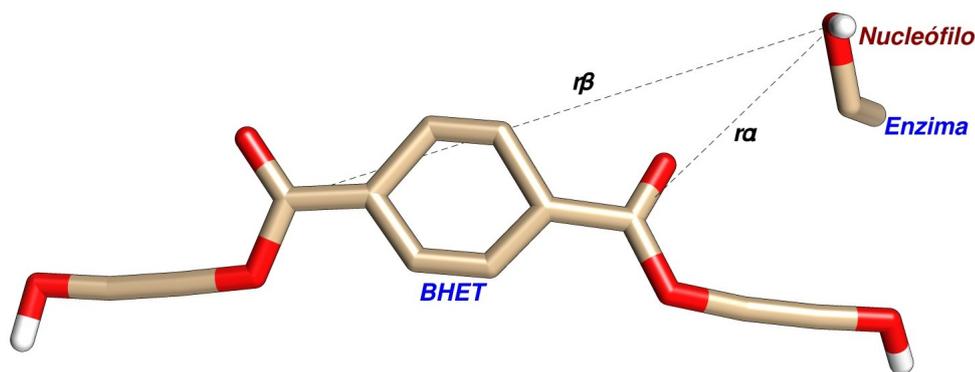


Figura 42 – Esquema das duas possibilidades para o ataque nucleofílico nos complexos Enzima-BHET.

$s_{v+r}$  para cada conformação. Na Fig. 43 é apresentado o formato do arquivo gerado pelo procedimento de *Docking*. Para analisar todos os resultados, precisamos que um programa busque corretamente os parâmetros: *escore vina*,  $r_\alpha$  e  $r_\beta$ . O *escore vina* para a afinidade é apresentado em *kcal/mol* no topo da descrição (como vemos destacado em roxo na Fig. 43). Quanto as distâncias, precisamos identificar as posições dos carbonos alvo do ligante e calcular os parâmetros a partir da posição referência<sup>1</sup> do oxigênio da hidroxila da SER 160. Para identificar os carbonos alvo basta buscar na lista de ligações (*CONNECT*) as linhas com quatro números. O primeiro número é o átomo referência e os demais indicam a quais outros ele está ligado. A exemplo dessa busca, notamos que na Fig. 43 o carbono de posição 12 na lista está ligado a dois átomos de oxigênio (13 e 14) e um carbono (9), ou seja, a posição 12 da lista *HETATM* apresenta a posição de um dos alvos.

Uma vez que um programa busque a lista de tuplas de quatro elementos (identificação da conformação, *escore vina*,  $r_\alpha$  e  $r_\beta$ ) de todos os arquivos resultantes do processo de *Docking*, basta ordenar a lista e gerar um arquivo com esta informação. Na Fig. 44, está apresentado o trecho com a definição do algoritmo de ordenação do programa *AvR\_DockingPlot.py*, que está disponível no meu diretório do [GitHub](#) e o topo do arquivo gerado com a lista ordenada. Como apresentado na figura o algoritmo escolhido foi o *quick sort* recursivo. Pode não estar aparente na implementação, mas a função *SortZeroTuple* recebe como argumento a lista com as tuplas de 4 elementos. Na Fig. 44, *t* recebe uma lista de tuplas com dois elementos (nas linhas 54, 60 e 70), entretanto, os valores *default* em Python mostram apenas o formato esperado e não o exigido. Da maneira que o programa foi implementado, as tuplas de entrada podem ter qualquer tamanho desde que o parâmetro de interesse permaneça na primeira posição.

<sup>1</sup> Obtido do PDB da enzima nativa ou variante utilizado como entrada no AutoDock Vina.

```

MODEL 1
REMARK VINA RESULT: -5.4 0.000 0.000
REMARK 10 active torsions:
REMARK status: (A for Active; T for Inactive)
REMARK 1 A between atoms: C_1 and C_3
REMARK 2 A between atoms: C_1 and O_18
REMARK 3 A between atoms: C_3 and O_17
REMARK 4 A between atoms: C_5 and C_6
REMARK 5 A between atoms: C_5 and O_14
REMARK 6 A between atoms: C_6 and O_13
REMARK 7 A between atoms: C_7 and C_10
REMARK 8 A between atoms: C_8 and O_17
REMARK 9 A between atoms: C_8 and C_12
REMARK 10 A between atoms: C_10 and O_14
COMPND IsPETaseD206H237K_C8X_782.pdbqt
AUTHOR GENERATED BY OPENBABEL 2.3.2
HETATM 1 C BET 1 -10.230 -13.976 -16.574 1.00 0.00 C
HETATM 2 C BET 1 -8.875 -13.277 -16.699 1.00 0.00 C
HETATM 3 O BET 1 -9.035 -12.040 -17.441 1.00 0.00 O
HETATM 4 C BET 1 -8.247 -11.006 -17.084 1.00 0.00 C
HETATM 5 O BET 1 -8.744 -9.989 -16.642 1.00 0.00 O
HETATM 6 C BET 1 -6.780 -11.109 -17.232 1.00 0.00 C
HETATM 7 C BET 1 -5.941 -10.342 -16.422 1.00 0.00 C
HETATM 8 C BET 1 -4.573 -10.438 -16.559 1.00 0.00 C
HETATM 9 C BET 1 -4.025 -11.303 -17.508 1.00 0.00 C
HETATM 10 C BET 1 -4.864 -12.071 -18.318 1.00 0.00 C
HETATM 11 C BET 1 -6.232 -11.975 -18.181 1.00 0.00 C
HETATM 12 C BET 1 -2.557 -11.406 -17.656 1.00 0.00 C
HETATM 13 O BET 1 -2.052 -11.346 -18.759 1.00 0.00 O
HETATM 14 O BET 1 -1.780 -11.565 -16.568 1.00 0.00 O
HETATM 15 C BET 1 -0.976 -10.428 -16.159 1.00 0.00 C
HETATM 16 C BET 1 -0.663 -10.533 -14.666 1.00 0.00 C
HETATM 17 O BET 1 0.186 -11.660 -14.436 1.00 0.00 O
HETATM 18 H BET 1 -0.146 -12.486 -14.813 1.00 0.00 H
HETATM 19 O BET 1 -10.396 -14.456 -15.238 1.00 0.00 O
HETATM 20 H BET 1 -10.565 -13.761 -14.588 1.00 0.00 H
CONNECT 1 2 19
CONNECT 2 3 1
CONNECT 3 4 2
CONNECT 4 3 6 5
CONNECT 5 4
CONNECT 6 11 4 7
CONNECT 7 6 8
CONNECT 8 9 7
CONNECT 9 10 12 8
CONNECT 10 11 9
CONNECT 11 10 6
CONNECT 12 13 9 14
CONNECT 13 12
CONNECT 14 12 15
CONNECT 15 14 16
CONNECT 16 15 17
CONNECT 17 18 16
CONNECT 18 17
CONNECT 19 1 20
CONNECT 20 19
MASTER 0 0 0 0 0 0 0 0 20 0 20 0
END
ENDMDL
MODEL 2

```

Afinidade

Carbonos  
alvo

Figura 43 – Arquivo de conformação do ligante gerado ao utilizar o *AutoDock Vina*.

## C.2 Decomposição da energia (MMPBSA)

Na Tab. 6, estão os valores médios da decomposição de energia apresentados na Fig. 18. Como o valor de interesse da mutante dupla – o vale de energia do resíduo 206 – não tem garantia de estar negativo devido ao tamanho do desvio padrão, fizemos novamente os cálculos do MMPBSA. Desta vez ao invés de utilizar toda a etapa de produção da dinâmica de 200 ns, focamos nas faixas mais estáveis quanto a distância nucleofílica das dinâmicas desta variante e de uma replicata. Como apresentado na Fig. 45, a dinâmica apresentada no corpo principal desta tese permaneceu mais estável no decorrer de toda a etapa de produção. Escolhemos as melhores faixas de 25 ns de ambas, para calcular com um oitavo do tempo (cerca de três horas e meia cada). Notamos desta vez, pela Fig. 46 e Tab. 7, que nesta faixa mais estável obtivemos garantia que na conformação da mutante dupla na Fig. 47, o vale energético do resíduo 206 permanece negativo. Esta conformação é semelhante a apresentada na Fig. 19-C, porém com a cadeia lateral do

```

53 # Quick sort for the first coordinate of my tuples
54 def swap(t=[(0,'1'),(-6,'2')], ith = 0, jth = 1):
55     if ith != jth:
56         temp = t[ith]
57         t[ith] = t[jth]
58         t[jth] = temp
59
60 def partition(t=[(0,'1'),(-6,'2')], begin= 0, end = 1):
61     pivot = t[end]
62     rpos = begin -1
63     for i in range(begin,end):
64         if t[i][0] < pivot[0]:
65             rpos += 1
66             swap(t,rpos,i)
67     swap(t,rpos+1,end)
68     return rpos+1
69
70 def SortZerothTuple(t=[(0,'1'),(-6,'2')], begin= 0, end = 1):
71     if begin < end:
72         pivot = partition(t=t,begin=begin,end=end)
73
74         a = SortZerothTuple(t=t,begin=begin,end=pivot-1)
75         b = SortZerothTuple(t=t,begin=pivot+1,end=end)

```

1	§: Minimizing factor; Coords: (affinity, nucleophilic distance)
2	0: S=-2.42; MODEL 2608; Coords=(-5.6, 3.18)
3	1: S=-2.41; MODEL 2239; Coords=(-5.6, 3.19)
4	2: S=-2.41; MODEL 2509; Coords=(-5.6, 3.19)
5	3: S=-2.40; MODEL 2932; Coords=(-5.6, 3.2)
6	4: S=-2.40; MODEL 136; Coords=(-5.6, 3.2)
7	5: S=-2.39; MODEL 1807; Coords=(-5.6, 3.21)
8	6: S=-2.39; MODEL 1843; Coords=(-5.6, 3.21)
9	7: S=-2.37; MODEL 2437; Coords=(-5.6, 3.23)
10	8: S=-2.37; MODEL 1420; Coords=(-5.6, 3.23)
11	9: S=-2.33; MODEL 883; Coords=(-5.5, 3.17)
12	10: S=-2.33; MODEL 2077; Coords=(-5.5, 3.17)
13	11: S=-2.33; MODEL 280; Coords=(-5.5, 3.17)
14	12: S=-2.33; MODEL 847; Coords=(-5.5, 3.17)
15	13: S=-2.32; MODEL 3436; Coords=(-5.5, 3.18)
16	14: S=-2.32; MODEL 901; Coords=(-5.5, 3.18)
17	15: S=-2.32; MODEL 3238; Coords=(-5.5, 3.18)
18	16: S=-2.32; MODEL 1951; Coords=(-5.5, 3.18)
19	17: S=-2.31; MODEL 640; Coords=(-5.5, 3.19)
20	18: S=-2.31; MODEL 2212; Coords=(-5.5, 3.19)
21	19: S=-2.31; MODEL 2256; Coords=(-5.5, 3.19)

Figura 44 – Algoritmo de ordenação e lista resultante. Imagens geradas a partir do programa *AvR\_DockingPlot.py*.

Tabela 6 – Valores médios relativos a decomposição de energia por resíduo e seus desvios padrões, em *kcal/mol*. Resíduos com as contribuições mais relevantes dos complexos IsPETase-BHET.

Resíduo	IsPETase	<i>D206E</i>	<i>D206E/H237K</i>	<i>H237K</i>
87 (TYR)	-2,59 ( $\pm 0,73$ )	-2,72 ( $\pm 0,89$ )	-1,57 ( $\pm 0,83$ )	-1,32 ( $\pm 1,13$ )
160 (SER)	-0,62 ( $\pm 0,54$ )	-0,50 ( $\pm 0,65$ )	-0,55 ( $\pm 0,60$ )	0,35 ( $\pm 1,00$ )
161 (MET)	-1,29 ( $\pm 0,38$ )	-1,30 ( $\pm 0,44$ )	-1,02 ( $\pm 0,37$ )	-0,99 ( $\pm 0,39$ )
185 (TRP)	-3,74 ( $\pm 0,83$ )	-3,55 ( $\pm 0,89$ )	-4,02 ( $\pm 0,81$ )	-3,10 ( $\pm 1,56$ )
206	-8,64 ( $\pm 2,50$ )	-8,13 ( $\pm 3,72$ )	-2,12 ( $\pm 3,74$ )	-0,70 ( $\pm 0,56$ )
208 (ILE)	-1,21 ( $\pm 0,44$ )	-1,08 ( $\pm 0,54$ )	-0,25 ( $\pm 0,35$ )	-0,27 ( $\pm 0,41$ )
237	1,95 ( $\pm 1,69$ )	2,24 ( $\pm 1,56$ )	-0,02 ( $\pm 1,02$ )	0,38 ( $\pm 1,30$ )

resíduo 206 rotacionada de forma que os oxigênios terminais estão mais próximos da LYS 237, e nesse caso o ligante permaneceu mais próximo da fenda de oxianion.

Pela descrição da ferramenta utilizada (*ambpdb* do AMBERtools), o estado utilizado do arquivo de trajetória para salvar as coordenadas como PDB é relativo ao estado salvo para recomeçar a dinâmica (caso ocorra alguma falha). Portanto o PDB utilizado nas Figs. 19-C e 47, são referentes aos estados a 200 ns e 100 ns, respectivamente. Pela Fig. 45, notamos que a conformação da Fig. 47 é a mais próxima do comportamento médio.

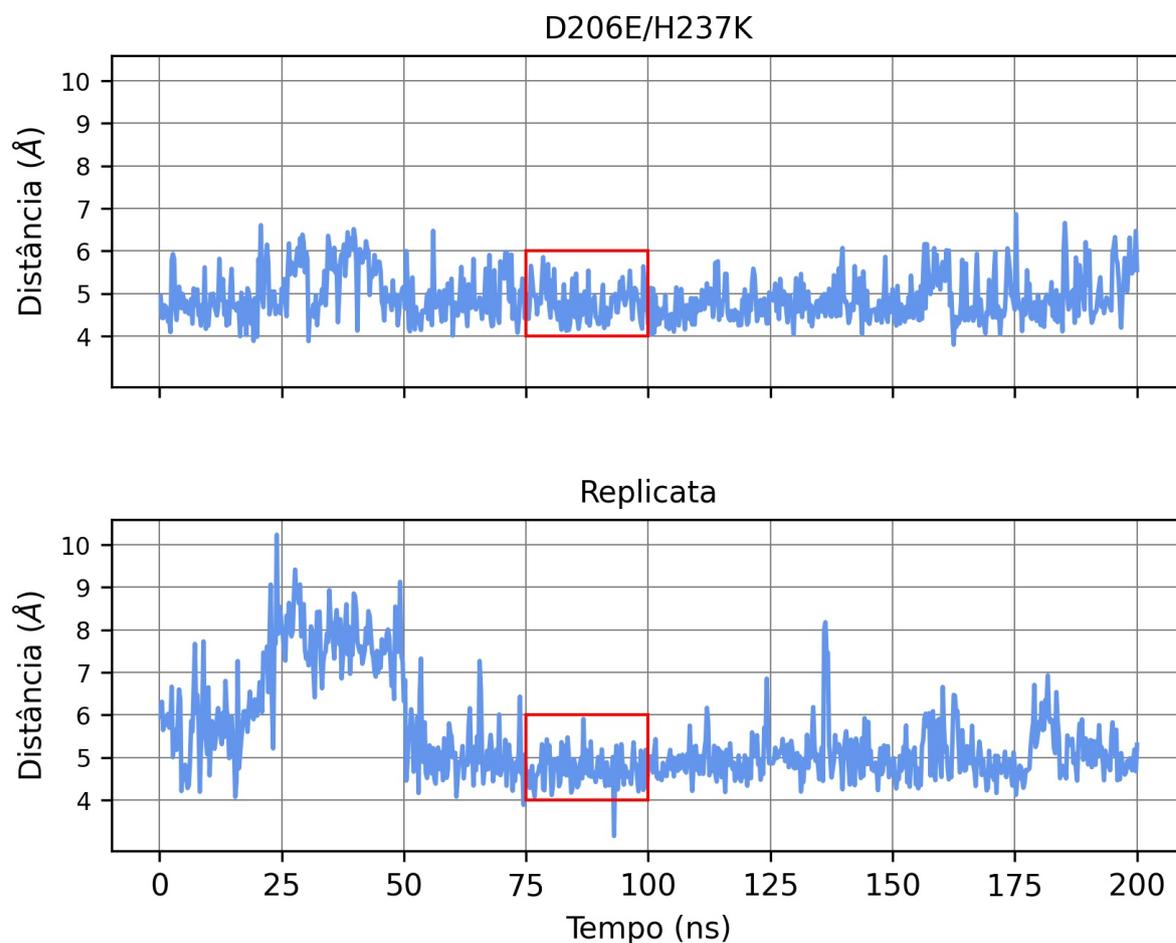


Figura 45 – Distância nucleofílica durante a etapa de produção, mantendo pH 7 e 300 K. Comparação entre a DM do complexo IsPETase<sup>D206E/H237K</sup>-PET e de uma DM replicata. Destacado em vermelho a região mais estável de ambas dinâmicas.

Tabela 7 – Valores médios relativos a decomposição de energia por resíduo e seus desvios padrões, em *kcal/mol*. Valores calculados utilizando os dados de 75 a 100 ns de dinâmica com a variante *D206E/H237K* e sua replicata.

Resíduo	<i>D206E/H237K</i>	Replicata
87 (TYR)	-1,43 ( $\pm 0,82$ )	-1,18 ( $\pm 0,78$ )
160 (SER)	-0,82 ( $\pm 0,59$ )	-0,24 ( $\pm 0,59$ )
161 (MET)	-1,03 ( $\pm 0,36$ )	-1,05 ( $\pm 0,41$ )
185 (TRP)	-3,77 ( $\pm 0,79$ )	-4,12 ( $\pm 0,78$ )
206	-4,86 ( $\pm 3,60$ )	0,11 ( $\pm 1,02$ )
208 (ILE)	-0,60 ( $\pm 0,48$ )	-0,03 ( $\pm 0,01$ )
237	-0,12 ( $\pm 0,96$ )	-0,06 ( $\pm 1,03$ )
$\Delta G_{Lig}$	-16,28 ( $\pm 8,00$ )	-8,15 ( $\pm 4,21$ )

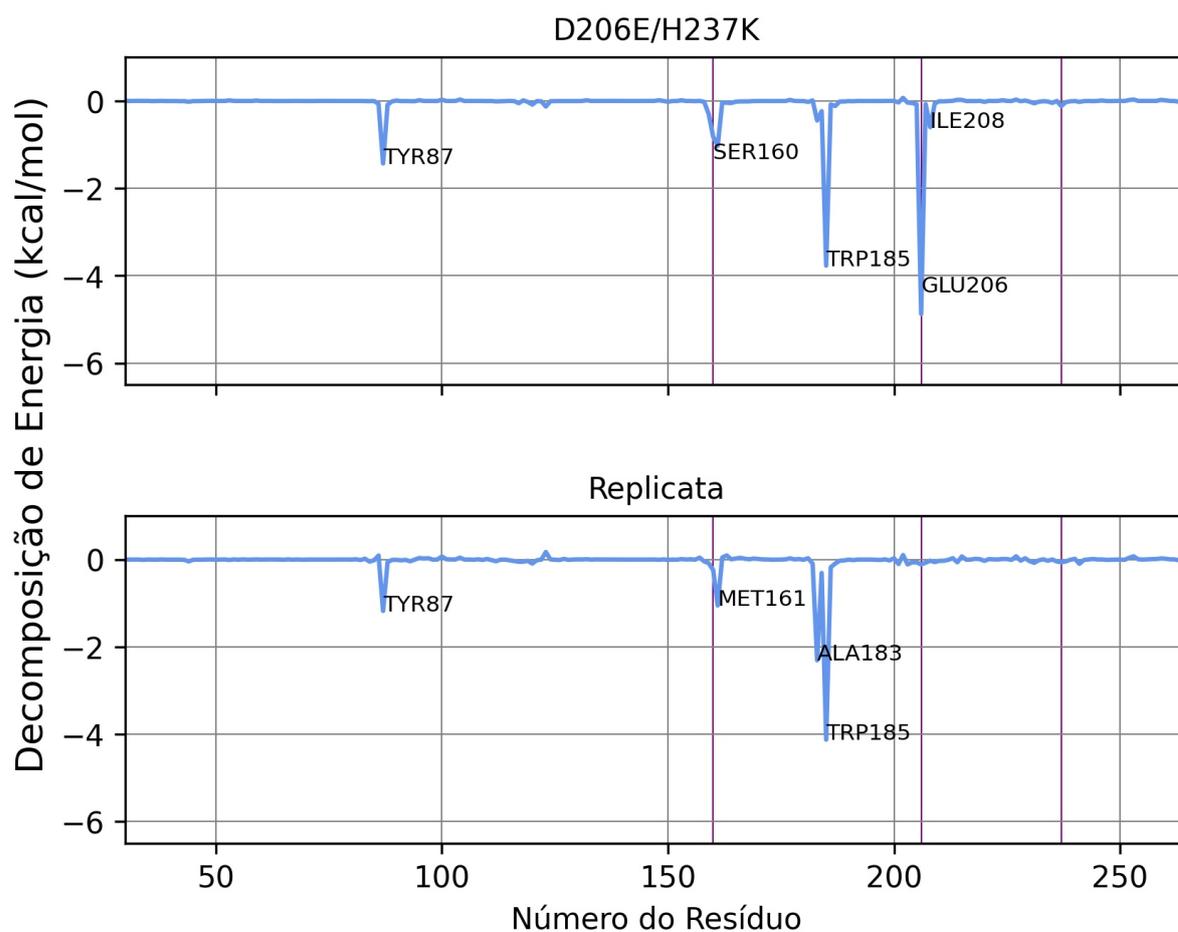


Figura 46 – Decomposição por resíduo da energia de interação Enzima-BHET, relativa a faixa de 75 a 100 ns destacada na Fig. 45. Comparação entre a DM da variante IsPETase<sup>D206E/H237K</sup> e de uma DM replicata.

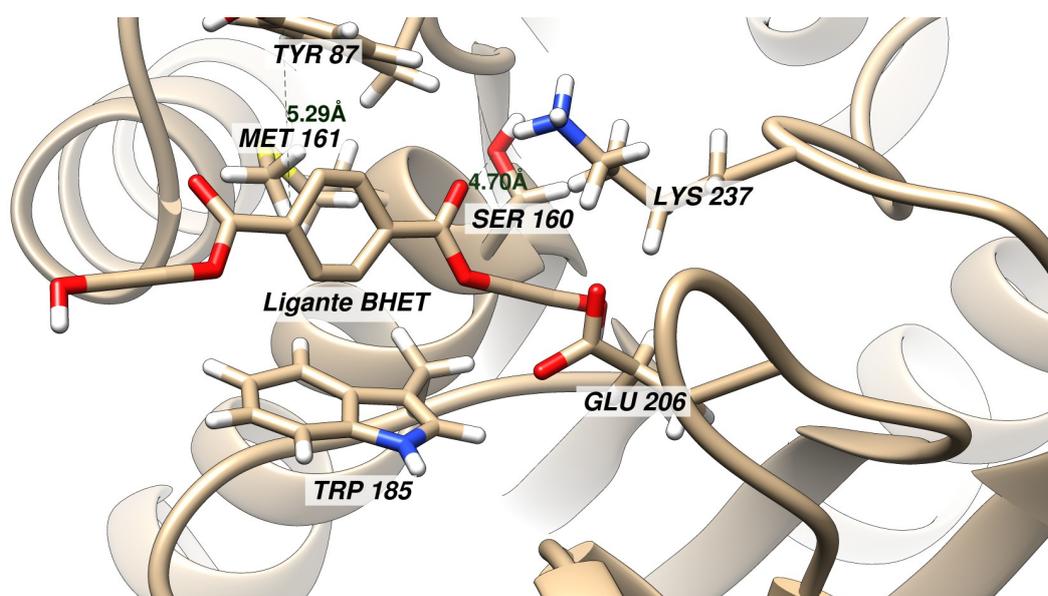


Figura 47 – Visão da conformação dos resíduos do sítio ativo da variante IsPETase<sup>D206E/H237K</sup>, no estado a 100 ns da produção.

# Referências

- 1 JAMBECK, J. R. et al. Plastic waste inputs from land into the ocean. *Science*, American Association for the Advancement of Science, v. 347, n. 6223, p. 768–771, 2015. ISSN 0036-8075. Disponível em: <<https://science.sciencemag.org/content/347/6223/768>>.
- 2 NEALE, C.; HILYARD, N.; BARBER, P. Observations on the economics of recycling industrial scrap plastic in new products. *Conservation & Recycling*, v. 6, n. 3, p. 91–105, 1983. ISSN 0361-3658. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0361365883900346>>.
- 3 SINHA, V.; PATEL, M. R.; PATEL, J. V. Pet waste management by chemical recycling: A review. *Journal of Polymers and the Environment*, v. 18, p. 8–25, 2010. ISSN 1572-8900. Disponível em: <<https://doi.org/10.1007/s10924-008-0106-7>>.
- 4 KAWAI, F. The current state of research on pet hydrolyzing enzymes available for biorecycling. *Catalysts*, v. 11, n. 2, 2021. ISSN 2073-4344. Disponível em: <<https://www.mdpi.com/2073-4344/11/2/206>>.
- 5 YOSHIDA, S. et al. A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*, American Association for the Advancement of Science, v. 351, n. 6278, p. 1196–1199, 2016. ISSN 0036-8075. Disponível em: <<https://science.sciencemag.org/content/351/6278/1196>>.
- 6 AUSTIN, H. P. et al. Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proceedings of the National Academy of Sciences*, National Academy of Sciences, v. 115, n. 19, p. E4350–E4357, 2018. ISSN 0027-8424. Disponível em: <<https://www.pnas.org/content/115/19/E4350>>.
- 7 WEI, R. et al. Conformational fitting of a flexible oligomeric substrate does not explain the enzymatic pet degradation. *Nature Communications*, v. 10, p. 5581, 2019. ISSN 2041-1723. Disponível em: <<https://doi.org/10.1038/s41467-019-13492-9>>.
- 8 TOURNIER, V. et al. An engineered pet depolymerase to break down and recycle plastic bottles. *Nature*, v. 580, p. 216–219, 2020. ISSN 1476-4687. Disponível em: <<https://doi.org/10.1038/s41586-020-2149-4>>.
- 9 EKICI, z. D.; PAETZEL, M.; DALBEY, R. E. Unconventional serine proteases: Variations on the catalytic ser/his/asp triad configuration. *Protein Science*, v. 17, n. 12, p. 2023–2037, 2008. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.035436.108>>.
- 10 WANG, J.-F. et al. Insights from modeling the 3d structure of nad(p)h-dependent d-xylose reductase of pichia stipitis and its binding interactions with nad and nadp. *Biochemical and Biophysical Research Communications*, v. 359, n. 2, p. 323–329, 2007. ISSN 0006-291X. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0006291X07010534>>.
- 11 ZENG, Q.-K. et al. Reversal of coenzyme specificity and improvement of catalytic efficiency of pichiastipitis xylose reductase by rational site-directed

- mutagenesis. *Biotechnology Letters*, v. 31, p. 1025–1029, 2009. Disponível em: <<https://doi.org/10.1007/s10529-009-9980-x>>.
- 12 EBERLE, R. J. et al. In vitro study of hesperetin and hesperidin as inhibitors of zika and chikungunya virus proteases. *PLOS ONE*, Public Library of Science, v. 16, n. 3, p. 1–28, 03 2021. Disponível em: <<https://doi.org/10.1371/journal.pone.0246319>>.
- 13 PETTERSEN, E. F. et al. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, v. 25, n. 13, p. 1605–1612, 2004. Disponível em: <<https://doi.org/10.1002/jcc.20084>>.
- 14 GE, B.; HU, G.; CHEN, C. Plastic bioconversion: Reaction mechanism of petases. *Chinese Journal of Physics*, v. 73, p. 331–339, 2021. ISSN 0577-9073. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0577907321001805>>.
- 15 VYAS, B. et al. Molecular dynamics/quantum mechanics guided designing of natural products based prodrugs of epalrestat. *Journal of Molecular Structure*, v. 1171, p. 556–563, 2018. ISSN 0022-2860. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022286018307312>>.
- 16 FRENKEL, D.; SMIT, B. *Understanding Molecular Simulation: From Algorithms to Applications*. 1st. ed. USA: Academic Press, Inc., 1996. ISBN 0122673700.
- 17 BRÜNGER, A.; BROOKS, C. L.; KARPLUS, M. Stochastic boundary conditions for molecular dynamics simulations of st2 water. *Chemical Physics Letters*, v. 105, n. 5, p. 495 – 500, 1984. ISSN 0009-2614. Disponível em: <<http://www.sciencedirect.com/science/article/pii/0009261484800986>>.
- 18 FRANK, J.; GOTTWALD, G. A. The langevin limit of the nosé-hoover-langevin thermostat. *J Stat Phys*, v. 143, p. 715–724, 2011. Disponível em: <<https://doi.org/10.1007/s10955-011-0203-1>>.
- 19 CASE, D. et al. *Amber 2018*. [S.l.]: University of California, San Francisco, 2018.
- 20 JORGENSEN, W. L. et al. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, v. 79, n. 2, p. 926–935, 1983. Disponível em: <<https://doi.org/10.1063/1.445869>>.
- 21 RYCKAERT, J.-P.; CICCOTTI, G.; BERENDSEN, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, v. 23, n. 3, p. 327–341, 1977. ISSN 0021-9991. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0021999177900985>>.
- 22 HOPKINS, C. W. et al. Long-time-step molecular dynamics through hydrogen mass repartitioning. *Journal of Chemical Theory and Computation*, v. 11, p. 1864–1874, 2015. Disponível em: <<http://dx.doi.org/10.1021/ct5010406>>.
- 23 MAIER, J. A. et al. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, v. 11, n. 8, p. 3696–3713, 2015. PMID: 26574453. Disponível em: <<https://doi.org/10.1021/acs.jctc.5b00255>>.
- 24 WANG, J. et al. Development and testing of a general amber force field. *Journal of Computational Chemistry*, v. 25, n. 9, p. 1157–1174, 2004. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20035>>.

- 25 SWAILS, J. M.; YORK, D. M.; ROITBERG, A. E. Constant ph replica exchange molecular dynamics in explicit solvent using discrete protonation states: Implementation, testing, and validation. *Journal of Chemical Theory and Computation*, v. 10, n. 3, p. 1341–1352, 2014. PMID: 24803862. Disponível em: <<https://doi.org/10.1021/ct401042b>>.
- 26 RAVENZWAAIJ, D. van; CASSEY, P.; BROWN, S. D. A simple introduction to markov chain monte-carlo sampling. *Psychonomic Bulletin & Review*, v. 25, p. 143–154, Feb. 2018. Disponível em: <<https://doi.org/10.3758/s13423-016-1015-8>>.
- 27 ALLWRIGHT, J. C. Conjugate gradient versus steepest descent. *J Optim Theory Appl*, v. 20, n. 1, p. 129–134, 1976. Disponível em: <<https://doi.org/10.1007/BF00933351>>.
- 28 TROTT, O.; OLSON, A. J. Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, v. 31, n. 2, p. 455–461, 2010. ISSN 1096-987X. Disponível em: <<https://doi.org/10.1002/jcc.21334>>.
- 29 ROSSUM, G. V.; DRAKE, F. L. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN 1441412697.
- 30 CHEN, P. P.-S. The entity-relationship model—toward a unified view of data. *ACM Trans. Database Syst.*, Association for Computing Machinery, New York, NY, USA, v. 1, n. 1, p. 9–36, mar. 1976. ISSN 0362-5915. Disponível em: <<https://doi.org/10.1145/320434.320440>>.
- 31 LI, P.; MERZ, K. M. J. Mcpb.py: A python based metal center parameter builder. *Journal of Chemical Information and Modeling*, v. 56, n. 4, p. 599–604, 2016. PMID: 26913476. Disponível em: <<https://doi.org/10.1021/acs.jcim.5b00674>>.
- 32 STIERAND, K.; RAREY, M. From modeling to medicinal chemistry: Automatic generation of two-dimensional complex diagrams. *ChemMedChem*, v. 2, n. 6, p. 853–860, 2007. Disponível em: <<https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cmdc.200700010>>.
- 33 EKICI, z. D.; PAETZEL, M.; DALBEY, R. E. Unconventional serine proteases: Variations on the catalytic ser/his/asp triad configuration. *Protein Science*, v. 17, n. 12, p. 2023–2037, 2008. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.035436.108>>.
- 34 BRAGA, B.; SILVA, M.; AMARAL, M. Variations of the canonical triad of ispetase: In silico insights with molecular dynamics simulation. *Chinese Journal of Physics*, v. 84, p. 282–291, 2023. ISSN 0577-9073. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0577907323000898>>.
- 35 BOX, V. The molecular mechanics of quantized valence bonds. *J Mol Med*, v. 3, p. 124–141, 1997. Disponível em: <<https://doi.org/10.1007/s008940050026>>.
- 36 ADAMS, J. Bonding energy models. In: BUSCHOW, K. J. et al. (Ed.). *Encyclopedia of Materials: Science and Technology*. Oxford: Elsevier, 2001. p. 763 – 767. ISBN 978-0-08-043152-9. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B0080431526001467>>.

- 37 CORNELL, W. D. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, v. 117, n. 19, p. 5179–5197, 1995. Disponível em: <<https://doi.org/10.1021/ja00124a002>>.
- 38 MAPLE, J. R. et al. A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *Journal of Chemical Theory and Computation*, v. 1, n. 4, p. 694–715, 2005. PMID: 26641692. Disponível em: <<https://doi.org/10.1021/ct049855i>>.
- 39 HONIG, B.; NICHOLLS, A. Classical electrostatics in biology and chemistry. *Science*, American Association for the Advancement of Science, v. 268, n. 5214, p. 1144–1149, 1995. ISSN 0036-8075. Disponível em: <<https://science.sciencemag.org/content/268/5214/1144>>.
- 40 TSUI, V.; CASE, D. A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers*, v. 56, n. 4, p. 275–291, 2000. Disponível em: <[https://doi.org/10.1002/1097-0282\(2000\)56:4<275::AID-BIP10024>3.0.CO;2-E](https://doi.org/10.1002/1097-0282(2000)56:4<275::AID-BIP10024>3.0.CO;2-E)>.
- 41 HAWKINS, G. D.; CRAMER, C. J.; TRUHLAR, D. G. Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters*, v. 246, n. 1, p. 122 – 129, 1995. ISSN 0009-2614. Disponível em: <<http://www.sciencedirect.com/science/article/pii/000926149501082K>>.
- 42 MONGAN, J.; CASE, D. A.; MCCAMMON, J. A. Constant ph molecular dynamics in generalized born implicit solvent. *Journal of Computational Chemistry*, v. 25, n. 16, p. 2038–2048, 2004. Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20139>>.
- 43 BERNARDIN, F. E.; RUTLEDGE, G. C. Semi-grand canonical monte carlo (sgmc) simulations to interpret experimental data on processed polymer melts and glasses. *Macromolecules*, v. 40, n. 13, p. 4691–4702, 2007. Disponível em: <<https://doi.org/10.1021/ma062935r>>.
- 44 HUNTER, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007.
- 45 Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 2020.
- 46 MILLER, B. R. I. et al. Mmpbsa.py: An efficient program for end-state free energy calculations. *Journal of Chemical Theory and Computation*, v. 8, n. 9, p. 3314–3321, 2012. PMID: 26605738. Disponível em: <<https://doi.org/10.1021/ct300418h>>.
- 47 BUGG, T. *Introduction to Enzyme and Coenzyme Chemistry*. 2. ed. [S.l.]: Blackwell Publishing, 2004. ISBN 1-4051-1452-5.
- 48 JOHNSON, K. A.; GOODY, R. S. The original michaelis constant: Translation of the 1913 michaelis–menten paper. *Biochemistry*, v. 50, n. 39, p. 8264–8269, 2011. PMID: 21888353. Disponível em: <<https://doi.org/10.1021/bi201284u>>.