



Serviço Público Federal  
Ministério da Educação

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



Instituto de Física - INFI

Programa de Pós-Graduação em Ciência dos Materiais - PPGCM

Marcio Fernandes da Silva

---

**Investigação da Interferência de Mutações Sobre a  
Interação da *IsPETase* com Ligantes via Simulação  
por Dinâmica Molecular**

---

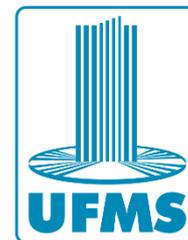
Campo Grande-MS

2024



Serviço Público Federal  
Ministério da Educação

Fundação Universidade Federal de Mato Grosso do Sul



Programa de Pós-Graduação em Ciência dos Materiais - PPGCM

---

**Investigação da Interferência de Mutações Sobre a  
Interação da *IsPETase* com Ligantes via Simulação  
por Dinâmica Molecular**

---

Marcio Fernandes da Silva

Orientador: Prof. Dr. Marcos Serrou do Amaral

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

Campo Grande-MS

2024

*Este trabalho é dedicado à minha mãe, Lucécia Maria Fernandes*

## **Agradecimentos**

A Deus pela oportunidade e por resguardar minha sanidade

À minha mãe, Lucécia Maria Fernandes, minha referência máxima de conquista

À minha família pelo apoio incondicional

Ao meu amigo Bruno Camargo Braga por acompanhar de perto o desenvolvimento deste trabalho e contribuir com sua melhora

Aos amigos Jéssica de Moraes Carretone e Stéphano Andrade de Vasconcelos por me ajudarem sempre que eu precisei

Ao técnico do Laboratório de Modelagem Molecular do Instituto de Física da UFMS, Rodrigo Fioravante, pelo suporte com os (nem sempre amigáveis) computadores e *softwares* fundamentais para a conclusão deste trabalho

Aos professores Danilo da Silva Olivier, Pedro Geraldo Pascutti, Fábio Mallmann Zimmer e Diogo Duarte dos Reis, cujos valiosos comentários e correções foram essenciais para o meu aprendizado acadêmico e para a conclusão deste trabalho

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo financiamento deste trabalho.

*“The hardest battles are fought in the mind.”*  
*Hellblade: Senua’s Sacrifice (Ninja Theory, 2017)*

## Resumo

Uma das principais causas da poluição ambiental é o acúmulo de polímeros sintéticos na natureza. Para lidar com esse problema, o uso de bactérias que produzem enzimas que degradam polímeros (tratamento biológico) pode ser uma alternativa. Para este trabalho a PETase produzida pela *Ideonella sakaiensis* foi escolhida (*Is*PETase) pelo seu alto grau de afinidade com o Polietileno Tereftalato (PET) – de onde a bactéria extrai energia das cadeias de carbono – e por apresentar atividade enzimática à temperatura ambiente. Com o auxílio da simulação via Dinâmica Molecular (DM), foi investigado como mutações no sítio ativo da proteína interferem na sua interação com ligantes – especificamente no processo de ancoramento. Mutar a Isoleucina 180 por uma Valina (variante I) e por uma Alanina (variante II) reduziu o número de resíduos interagindo com o ligante, assim como aumentou a distância proteína-ligante – consequência da redução da formação de ligações de hidrogênio entre a proteína e o ligante. A energia livre de Gibbs também diminuiu pela implementação das mutações, indicando perda de afinidade entre a proteína e os ligantes analisados, comprometendo a manutenção do complexo.

**Palavras-chave:** Tratamento biológico • *Is*PETase • Dinâmica Molecular • Mutação

## Abstract

One of the leading causes of environmental pollution is the stacking of synthetic polymers in nature. To deal with this problem, the use of bacteria that produces polymer-degrading enzymes (biological treatment) might be an alternative. The PETase produced by *I. sakaiensis* (*IsPETase*) we chose, possesses great affinity towards polyethylene terephthalate (PET) as its carbon source and it is active at ambient temperature. Aided by Molecular Dynamics (MD) simulation, we investigated how single mutations on its active site interfere in its interactions with ligands – specifically the docking process. Mutate Isoleucine 180 for a Valine (variant I) and for an Alanine (variant II) reduced how many residues interact with the ligand, as well as increased the protein-ligand distance – caused by the decrease in the hydrogen bond formation between the protein and the ligand. Gibbs free energy also decreased due to the mutations, indicating some affinity loss between protein and ligand analyzed, compromising complex maintenance.

**Key-words:** Biological treatment • *IsPETase* • Molecular Dynamics • Mutation

## Lista de Figuras

FIGURA 1: ACÚMULO DE RESÍDUOS PLÁSTICOS NO MEIO AMBIENTE. ADAPTADO DE <i>NATURE COMMUNICATIONS</i> "THE FUTURE OF PLASTIC" [9].	17
FIGURA 2: ESTRUTURAS QUÍMICAS DO POLIETILENO TEREFTALATO (PET) E DO BIS(2-HIDROXIETIL TEREFTALATO) (BHET) [15].	18
FIGURA 3: MICROSCOPIA ELETRÔNICA DA <i>I. SAKAIENSIS</i> . ADAPTADO DE YOSHIDA ET AL., 2016 [28].	19
FIGURA 4: EM (A) A ESTRUTURA TERCIÁRIA (ESTILO CARTOON) DA <i>IsPETASE</i> (CÓDIGO PDB: 6EQE) COM SUAS PRINCIPAIS ESTRUTURAS SECUNDÁRIAS: 9 FITAS-B (AMARELO) – FORMANDO A FOLHA-CENTRAL DA PROTEÍNA – E 7 HÉLICES-A (LILÁS). AS DUAS PRINCIPAIS ESTRUTURAS DO SÍTIO ATIVO DA <i>IsPETASE</i> ESTÃO AMPLIADAS AO LADO, COM OS RESÍDUOS REPRESENTADOS NA FORMA ESFERAS E BASTÕES: (B) A TRÍADE CATALÍTICA, COM OS RESÍDUOS SER132 (VERMELHO), ASP178 (LARANJA) E HIS209 (VERDE); (C) A FENDA DE OXIÂNION, COMPOSTA PELOS RESÍDUOS TYR59 (ROSA) E MET133 (VERDE).	22
FIGURA 5: ESQUEMA SIMPLIFICADO DO PROCESSO DE DEGRADAÇÃO REALIZADO PELA <i>IsPETASE</i> . BASEADO NO FLUXOGRAMA APRESENTADO POR GE, HU E CHEN [23].	23
FIGURA 6: MODELOS DE <i>DOCKING</i> MAIS CONHECIDOS: (A) <i>CHAVE-FECHADURA</i> E (B) <i>ENCAIXE INDUZIDO</i> . ADAPTADO DE [47].	29
FIGURA 7: REPRESENTAÇÃO 2D DA ESTRUTURA QUÍMICA DA ISOLEUCINA (RESÍDUO NATIVO DA POSIÇÃO 180), VALINA E ALANINA (RESÍDUOS PROPOSTOS PARA AS MUTAÇÕES). IMAGEM ADAPTADA DE LEHNINGER, ALBERT L., DAVID L. NELSON, E MICHAEL M. COX. <i>LEHNINGER PRINCIPLES OF BIOCHEMISTRY</i> . MACMILLAN, 2005. [55].	34
FIGURA 8: PARÂMETROS APLICADOS À MINIMIZAÇÃO SEGUIDOS DE SUAS DESCRIÇÕES.	39
FIGURA 9: PARÂMETROS APLICADOS AO AQUECIMENTO SEGUIDOS DE SUAS DESCRIÇÕES.	39
FIGURA 10: PARÂMETROS APLICADOS À EQUILIBRAÇÃO SEGUIDOS DE SUAS DESCRIÇÕES.	40
FIGURA 11: PARÂMETROS APLICADOS À PRODUÇÃO SEGUIDOS DE SUAS DESCRIÇÕES	41
FIGURA 12: COMPARATIVO DOS RESULTADOS DE RMSD ENTRE A PROTEÍNA NATIVA E SEUS RESPECTIVOS COMPLEXOS: (A) E (B) WT; (C) E (D) I180V; E (E) E (F) I180A. A COLUNA ESQUERDA REPRESENTA AS TRIPLICATAS DOS COMPLEXOS COM O BHET E A COLUNA DIREITA REPRESENTA AS TRIPLICATAS DOS COMPLEXOS COM O PET.	46
FIGURA 13: COMPARATIVO DOS RESULTADOS DE RG ENTRE A PROTEÍNA NATIVA E SEUS RESPECTIVOS COMPLEXOS: (A) E (B) WT; (C) E (D) I180V; E (E) E (F) I180A. A COLUNA ESQUERDA REPRESENTA AS TRIPLICATAS DOS COMPLEXOS COM O BHET E A COLUNA DIREITA REPRESENTA AS TRIPLICATAS DOS COMPLEXOS COM O PET.	48
FIGURA 14: COMPARATIVO DOS RESULTADOS DE RMSF ENTRE A PROTEÍNA NATIVA E SEUS RESPECTIVOS COMPLEXOS: (A) E (B) WT; (C) E (D) I180V; E (E) E (F) I180A. A COLUNA ESQUERDA REPRESENTA AS TRIPLICATAS DOS COMPLEXOS COM O BHET E A COLUNA DIREITA REPRESENTA AS TRIPLICATAS DOS COMPLEXOS COM O PET.	49
FIGURA 15: ESTRUTURA DE LAÇO DA PROTEÍNA CORRESPONDENTE AO INTERVALO DE RESÍDUOS 25-60 – REGIÃO (I).	50

FIGURA 16: AS FLUTUAÇÕES RESIDUAIS MAIS EVIDENTES SE MOSTRAM CENTRADAS NO RESÍDUO 180 (À ESQUERDA, NA COR VERMELHA, A ÍLE DA PROTEÍNA NATIVA) E NAS PROXIMIDADES DA HIS209 (À DIREITA, NA COR ROXA).....	51
FIGURA 17: CONTABILIZAÇÃO DE LIGAÇÕES DE HIDROGÊNIO (HB) AO LONGO DA SIMULAÇÃO (POR FRAMES) PARA OS COMPLEXOS DA PROTEÍNA NATIVA COM O MONÔMERO DE BHET. À ESQUERDA OS PARES RESÍDUO-LIGANTE QUE FORMARAM HB E ENTRE PARÊNTESE OS ÁTOMOS QUE PARTICIPAM DA LIGAÇÃO. DE CIMA PARA BAIXO: WT-BHET-DM1, WT-BHET-DM2 E WT-BHET-DM3. ....	52
FIGURA 18: CONTABILIZAÇÃO DE LIGAÇÕES DE HIDROGÊNIO (HB) AO LONGO DA SIMULAÇÃO (POR FRAMES) PARA OS COMPLEXOS DA PROTEÍNA NATIVA COM O MONÔMERO DE PET. À ESQUERDA OS PARES RESÍDUO-LIGANTE QUE FORMARAM HB E ENTRE PARÊNTESE OS ÁTOMOS QUE PARTICIPAM DA LIGAÇÃO. DE CIMA PARA BAIXO: WT-PET-DM1, WT-PET-DM2 E WT-PET-DM3. ....	53
FIGURA 19: CONTABILIZAÇÃO DE LIGAÇÕES DE HIDROGÊNIO (HB) AO LONGO DA SIMULAÇÃO (POR FRAMES) PARA OS COMPLEXOS DA VARIANTE I COM O MONÔMERO DE BHET. À ESQUERDA OS PARES RESÍDUO-LIGANTE QUE FORMARAM HB E ENTRE PARÊNTESE OS ÁTOMOS QUE PARTICIPAM DA LIGAÇÃO. DE CIMA PARA BAIXO: I180V-BHET-DM1, I180V-BHET-DM2 E I180V-BHET-DM3. ....	54
FIGURA 20: CONTABILIZAÇÃO DE LIGAÇÕES DE HIDROGÊNIO (HB) AO LONGO DA SIMULAÇÃO (POR FRAMES) PARA OS COMPLEXOS DA VARIANTE II COM O MONÔMERO DE BHET. À ESQUERDA OS PARES RESÍDUO-LIGANTE QUE FORMARAM HB E ENTRE PARÊNTESE OS ÁTOMOS QUE PARTICIPAM DA LIGAÇÃO. DE CIMA PARA BAIXO: I180A-BHET-DM1, I180A-BHET-DM2 E I180A-BHET-DM3. ....	55
FIGURA 21: ANÁLISE DO MAPA DE ENERGIA LIVRE BASEADO EM RMSD-R <sub>G</sub> DOS SISTEMAS COM PROTEÍNA NATIVA (A) ISOLADA; (B), (C) E (D) COMPLEXADA COM MONÔMERO DE BHET. ....	57
FIGURA 22: ANÁLISE DO MAPA DE ENERGIA LIVRE BASEADO EM RMSD-R <sub>G</sub> DOS SISTEMAS COM PROTEÍNA NATIVA (A) ISOLADA; (B), (C) E (D) COMPLEXADA COM MONÔMERO DE PET. ....	58
FIGURA 23: ANÁLISE DO MAPA DE ENERGIA LIVRE BASEADO EM RMSD-R <sub>G</sub> DOS SISTEMAS COM A VARIANTE I (A) ISOLADA; (B), (C) E (D) COMPLEXADA COM MONÔMERO DE BHET. ....	59
FIGURA 24: ANÁLISE DO MAPA DE ENERGIA LIVRE BASEADO EM RMSD-R <sub>G</sub> DOS SISTEMAS COM A VARIANTE I (A) ISOLADA; (B), (C) E (D) COMPLEXADA COM MONÔMERO DE PET. ....	60
FIGURA 25: ANÁLISE DO MAPA DE ENERGIA LIVRE BASEADO EM RMSD-R <sub>G</sub> DOS SISTEMAS COM A VARIANTE II (A) ISOLADA; (B), (C) E (D) COMPLEXADA COM MONÔMERO DE BHET.....	61
FIGURA 26: ANÁLISE DO MAPA DE ENERGIA LIVRE BASEADO EM RMSD-R <sub>G</sub> DOS SISTEMAS COM A VARIANTE II (A) ISOLADA; (B), (C) E (D) COMPLEXADA COM MONÔMERO DE PET. ....	62
FIGURA 27: VALOR MÉDIO TEMPORAL DA DECOMPOSIÇÃO DE ENERGIA POR RESÍDUO AO LONGO DA SIMULAÇÃO. OS GRÁFICOS APRESENTADOS REFEREM-SE AOS COMPLEXOS DA PROTEÍNA NATIVA COM O MONÔMERO DE BHET APRESENTADOS DA SEGUINTE MANEIRA: DM1 (A), DM2 (B) E DM3 (C). .	65
FIGURA 28: RESULTADOS DA DISTÂNCIA PROTEÍNA-LIGANTE ( $D_{HC}$ ) DURANTE A SIMULAÇÃO. OS GRÁFICOS REPRESENTAM OS COMPLEXOS DA PROTEÍNA NATIVA COM O BHET (TOPO) E PET (INFERIOR). ....	66
FIGURA 29: VALOR MÉDIO TEMPORAL DA DECOMPOSIÇÃO DE ENERGIA POR RESÍDUO AO LONGO DA SIMULAÇÃO. OS GRÁFICOS APRESENTADOS REFEREM-SE AOS COMPLEXOS DA PROTEÍNA NATIVA COM O MONÔMERO DE BHET: DM2 (A) E DM3 (B). ....	67
FIGURA 30: VALOR MÉDIO TEMPORAL DA DECOMPOSIÇÃO DE ENERGIA POR RESÍDUO AO LONGO DA SIMULAÇÃO. OS GRÁFICOS APRESENTADOS REFEREM-SE AOS COMPLEXOS DA VARIANTE I COM O MONÔMERO DE BHET: DM1 (A), DM2 (B) E DM3 (C). ....	68

FIGURA 31: RESULTADOS DA DISTÂNCIA PROTEÍNA-LIGANTE ( $D_{HC}$ ) DURANTE A SIMULAÇÃO. OS GRÁFICOS REPRESENTAM OS COMPLEXOS DA VARIANTE I COM O BHET (TOPO) E PET (INFERIOR). .....	69
FIGURA 32: VALOR MÉDIO TEMPORAL DA DECOMPOSIÇÃO DE ENERGIA POR RESÍDUO AO LONGO DA SIMULAÇÃO. OS GRÁFICOS APRESENTADOS REFEREM-SE AOS COMPLEXOS DA VARIANTE II COM O MONÔMERO DE BHET: DM1 (A), DM2 (B) E DM3 (C). .....	70
FIGURA 33: RESULTADOS DA DISTÂNCIA PROTEÍNA-LIGANTE ( $D_{HC}$ ) DURANTE A SIMULAÇÃO. OS GRÁFICOS REPRESENTAM OS COMPLEXOS DA VARIANTE II COM O PET (TOPO) E PET (INFERIOR). .....	71
FIGURA 34: COMPLEXO DA PROTÉINA VARIANTE I COM O MONÔMERO DE BHET MOMENTOS APÓS O LIGANTE SE DESPRENDER DA PROTEÍNA, APÓS 250 NS DE SIMULAÇÃO. ....	74
FIGURA 35: COMPLEXO DA DM2 DA VARIANTE II COM O MONÔMERO DE BHET, EM QUE O LIGANTE SAIU DO SÍTIO ATIVO DA PROTEÍNA, SEM SE DESPRENDER E SE MOVEU AO LONGO DE SUA SUPERFÍCIE DURANTE A SIMULAÇÃO. AS CORES REPRESENTAM A SUPERFÍCIE DA SER132 (VERDE), ASP178 E HIS209 (ROXO), ALA180 (LARANJA), TRP157 (AMARELO) E MET133 (VERMELHO). ....	75
FIGURA 36: REPRESENTAÇÃO DOS ESTADOS DE PROTONAÇÃO DO ANEL IMIDAZOL DA HISTIDINA: DUPLAMENTE PROTONADO (A), MONO PROTONADO EM $N_E$ (B) E MONO PROTONADO EM $N_\Delta$ (C) – ADAPTADO DE [78]. ....	77
FIGURA 37: COMPLEXO DA PROTEÍNA NATIVA COM O BHET (DM1). A AMPLIAÇÃO DA REGIÃO DO SÍTIO ATIVO MOSTRA O LIGANTE PRÓXIMO AO NUCLEÓFILO DA PROTÉINA (DISTÂNCIA MÉDIA DE 4 Å DA SER132) E TAMBÉM A UMA DISTÂNCIA FAVORÁVEL A FORMAÇÃO DE LIGAÇÕES DE HIDROGÊNIO DO ASP178. ....	78
FIGURA 38: O LIGANTE SE DESPRENDEU DURANTE A SIMULAÇÃO DM1 DO COMPLEXO DA PROTEÍNA NATIVA COM O PET. ....	78
FIGURA 39: A AMPLIAÇÃO DO SÍTIO ATIVO NO COMPLEXO DA PROTEÍNA NATIVA COM O PET DEMONSTRA O LIGANTE DESLOCADO NO SÍTIO ATIVO, MAIS PRÓXIMO DA ALA155, COMO SUGERIDO PELAS LIGAÇÕES DE HIDROGÊNIO FORMADAS. ....	79
FIGURA 40: SEQUÊNCIA DE IMAGENS DEMONSTRANDO VARIAÇÃO DA POSIÇÃO DO LIGANTE NO SÍTIO ATIVO NA DM1 DO COMPLEXO DA VARIANTE I COM O BHET. DURANTE A SIMULAÇÃO O LIGANTE REALIZOU UM GIRO E ACABOU SE POSICIONANDO DE MANEIRA DESLOCADA NO SÍTIO ATIVO. AS CORES REPRESENTAM A SUPERFÍCIE DA SER132 (VERDE), ASP178 E HIS209 (ROXO), VAL180 (LARANJA), TRP157 (AMARELO) E MET133 (VERMELHO). ....	80
FIGURA 41: SEQUÊNCIA DE IMAGENS QUE DEMONSTRA O DESLOCAMENTO DO LIGANTE NA DM2 DA VARIANTE I COMPLEXADO COM O BHET. O LIGANTE SE MOVE E SE POSICIONA NO SÍTIO ATIVO DESLOCADO NA DIREÇÃO DA MET133 (VERMELHO). AS DEMAIS CORES REPRESENTAM A SUPERFÍCIE DA SER132 (VERDE), ASP178 E HIS209 (ROXO), VAL180 (LARANJA), TRP157 (AMARELO), SER160 E ALA155 (CIANO) E LEU89 (CINZA). ....	80
FIGURA 42: LIGANTE SAIU DO SÍTIO ATIVO E CAMINHOU NA SUPERFÍCIE DA PROTEÍNA NA DM1 DA VARIANTE II COMPLEXADA COM BHET. O FOI EM DIREÇÃO AO RESÍDUO ASP122 (DESTACADO EM ROSA À ESQUERDA) AS DEMAIS CORES REPRESENTAM A SUPERFÍCIE DA SER132 (VERDE), ASP178 E HIS209 (ROXO), ALA180 (LARANJA), TRP157 (AMARELO) E MET133 (VERMELHO). ....	81
FIGURA 43: LIGANTE SE MOVIMENTOU NA SUPERFÍCIE DA PROTEÍNA E SE POSICIONOU FORA DO SEU SÍTIO ATIVO NA DM2 DA VARIANTE II COMPLEXADA COM BHET. AS CORES REPRESENTAM A SUPERFÍCIE DA SER132 (VERDE), ASP178 E HIS209 (ROXO), ALA180 (LARANJA), TRP157 (AMARELO) E MET133 (VERMELHO). ....	82

FIGURA 44: SEQUÊNCIA DE IMAGENS QUE DEMONSTRA O DESLOCAMENTO DO LIGANTE NA DM3 DA VARIANTE II COMPLEXADO COM O BHET. O LIGANTE PERMANECE NO SÍTIO ATIVO, MAS POSICIONADO NO SÍTIO ATIVO DESLOCADO EM RELAÇÃO À SER132 (VERDE), COM $d_{HC}$ MÉDIO DE 6 Å, NA DIREÇÃO DA MET133 (VERMELHO). AS DEMAIS CORES REPRESENTAM A SUPERFÍCIE DA SER132 (VERDE), ASP178 E HIS209 (ROXO), ALA180 (LARANJA) E TRP157 (AMARELO). .....	82
FIGURA 45: MODELO 2D DA ESTRUTURA BÁSICA DE UM A-AMINOÁCIDO [55]. .....	93
FIGURA 46: HIERARQUIA ESTRUTURAL DAS PROTEÍNAS – ESTRUTURA PRIMÁRIA (A), ESTRUTURA SECUNDÁRIA (B), ESTRUTURA TERCIÁRIA (C) E ESTRUTURA QUATERNÁRIA (D). ADAPTADO DE [83]. .....	94
FIGURA 47: REPRESENTAÇÃO DO <i>SCRIPT</i> PARA A GERAÇÃO DOS ARQUIVOS BASE PARA A OBTENÇÃO DOS DADOS SOBRE A DECOMPOSIÇÃO DE ENERGIA.....	96
FIGURA 48: MODELO DE <i>SCRIPT</i> PARA CONVERSÃO DE ARQUIVOS DE TOPOLOGIA E COORDENADAS. ....	97

## Lista de Tabelas

TABELA 1: APRESENTAÇÃO DA CLASSIFICAÇÃO MODELOS DE MONÔMEROS DE BHET OBTIDOS COM O ATD. OS MODELOS ESTÃO DISPOSTOS CONFORME O VALOR OBTIDO PARA O PARÂMETRO $S_{VD}$ . .....	44
TABELA 2: APRESENTAÇÃO DA CLASSIFICAÇÃO MODELOS DE MONÔMEROS DE PET OBTIDOS COM O ATD. OS MODELOS ESTÃO DISPOSTOS CONFORME O VALOR OBTIDO PARA O PARÂMETRO $S_{VD}$ . .....	45
TABELA 3: DISTRIBUIÇÃO DOS ESTADOS DE PROTONAÇÃO DA HIS209 DURANTE A SIMULAÇÃO PARA OS COMPLEXOS COM A PROTEÍNA NATIVA. E1 REPRESENTA O ESTADO 1, MONOPROTONADO EM $N_{\Delta}$ ; E2 REPRESENTA O ESTADO 2, MONOPROTONADO EM $N_E$ . .....	63
TABELA 4: DISTRIBUIÇÃO DOS ESTADOS DE PROTONAÇÃO DA HIS209 DURANTE A SIMULAÇÃO PARA OS COMPLEXOS COM A VARIANTE I. E1 REPRESENTA O ESTADO 1, MONOPROTONADO EM $N_{\Delta}$ ; E2 REPRESENTA O ESTADO 2, MONOPROTONADO EM $N_E$ . .....	64
TABELA 5: DISTRIBUIÇÃO DOS ESTADOS DE PROTONAÇÃO DA HIS209 DURANTE A SIMULAÇÃO PARA OS COMPLEXOS COM A VARIANTE II. E1 REPRESENTA O ESTADO 1, MONOPROTONADO EM $N_{\Delta}$ ; E2 REPRESENTA O ESTADO 2, MONOPROTONADO EM $N_E$ . .....	64
TABELA 6: VALORES DA ENERGIA LIVRE DE GIBBS DOS SISTEMAS ANALISADOS.....	72

## Lista de Abreviaturas e Siglas

A<sup>LIG</sup> – Forma intermediária acil-enzima

AMBER – Assisted Model Building with Energy Refinement

ASM – AMBER Simulation Manager

ATD – AutoDock Vina

BHET – Bis(2-hidroxietil tereftalato)

CpHMD – Constant pH Molecular Dynamics

$d_{HC}$  – Valor da distância entre o oxigênio da Serina 132 da proteína e o carbono de uma das carbonilas do ligante (BHET ou PET)

DM – Dinâmica Molecular

EG – Etileno Glicol

FEL – Free Energy Landscape

GAFF – Generalized AMBER Force Field

GB – Generalized-Born

HB – Hydrogen Bonds

HMR – Hydrogen Mass Repartitioning

LIG – Ligante

MC – Monte Carlo

MHET – Mono(2-hidroxietil) Tereftalato

MM/PB(GB)/SA – Molecular Mechanics / Poisson-Boltzmann (or Generalized Born) Surface Area

P1/P2 – Produto 1 / Produto 2 (no esquema de degradação realizado pela proteína)

PB – Poisson-Boltzman

PET – Poliestireno

pmemd – Particle Mesh Ewald Molecular Dynamics

PP – Polipropileno

PROT – Proteína

PU – Poliuretano

PVC – Cloreto de Polivinila

QM/MM – Quantum Mechanics/Molecular Mechanics

$R_g$  – Radius of gyration

RMSD – Root Mean Square Deviation

RMSF – Root Mean Square Fluctuation

sander – Simulated Annealing with NMR-Derived Energy Restraints

$T_g$  – Temperatura de transição

TPA – Ácido Tereftálico

WT – wild type

# Sumário

Sumário.....	15
1 Introdução.....	17
2 Revisão da Literatura.....	21
2.1 Breves conceitos sobre proteínas.....	21
2.2 A proteína da <i>Ideonella sakaiensis</i> .....	21
2.3 O processo de degradação realizado pela IsPETase.....	22
2.4 Ferramental computacional.....	23
2.4.1 Alguns programas do AmberTools.....	25
2.4.2 Simulação pelo AMBER.....	27
2.4.3 Conceitos de Docking molecular.....	28
2.4.4 Conceitos sobre simulação com pH constante.....	29
2.4.5 Método de redistribuição de massa dos átomos.....	30
2.4.6 Mutações residuais em proteínas.....	31
3 Objetivo.....	32
3.1 Objetivos gerais.....	32
3.2 Objetivos específicos.....	32
4 Metodologia.....	33
4.1 Gerenciamento dos processos.....	33
4.2 Mutações no sítio ativo da proteína.....	33
4.3 Docking e preparação dos ligantes.....	34
4.4 Preparação dos sistemas.....	36
4.5 Pré-produção e produção (simulação).....	36
4.6 Análises de dados.....	41
5 Resultados.....	44
5.1 Modelos de ligante gerados.....	44
5.2 Análises dos sistemas complexados.....	45
5.2.1 Análise de RMSD.....	46

5.2.2	Análise do raio de giro.....	47
5.2.3	Análise da flutuação residual.....	49
5.2.4	Análise da formação das ligações de hidrogênio.....	51
5.2.5	Análise dos mapas de energia livre .....	56
5.2.6	Análise dos estados de protonação .....	63
5.3	Decomposição de Energia e Distância proteína-ligante .....	64
5.3.1	Complexos com a proteína nativa (WT).....	64
5.3.2	Complexos com a variante I (I180V) .....	67
5.3.3	Complexos com a variante II (I180A).....	69
5.3.4	Energia livre Gibbs nos complexos .....	71
6	Discussão dos resultados .....	73
6.1	Sobre os resultados estruturais e mapas de energia .....	73
6.2	Estados de protonação .....	77
6.3	Energia livre de ligação por resíduo e distância nucleofílica .....	77
6.4	Energia livre de Gibbs .....	83
7	Conclusão .....	84
8	Referências .....	86
9	Apêndices .....	93
	Apêndice A – Sobre a estrutura geral das proteínas.....	93
	Apêndice B – Descrição dos <i>scripts</i> em <i>Python</i> .....	94
	Apêndice C – Decomposição de energia com MMPBSA.....	96
	Apêndice D – Conversão de arquivos para produção dos mapas de energia livre no PyMOL.....	97

# 1 Introdução

A poluição é um problema bastante conhecido, de modo que a poluição do ar, a presença de resíduos industriais no ciclo das águas e o enorme acúmulo de polímeros sintéticos (plásticos) na natureza, figuram entre suas principais formas [1–4]. A partir do surgimento dos polímeros sintéticos, o paradigma dos produtos, embalagens e seus sistemas de produção foi profundamente alterado [5]. A indústria se desenvolveu consideravelmente com a criação de várias rotas de produção de plásticos baseadas em fontes petroquímicas. Dessa maneira, a associação do baixo custo de produção com a versatilidade dos polímeros produzidos, acabaram tornando-os essenciais para a sociedade moderna [6,7].

Contudo, da vasta quantidade de polímeros produzidos, aproximadamente 50% desse montante é empregado em atividades de uso único, como embalagens e produtos descartáveis. Assim, as características que os tornam industrialmente interessantes (resistência e impermeabilidade) são as mesmas que os fazem uma parcela significativa das fibras sintéticas que poluem o meio ambiente, como apresentado na Fig. 1 [6–8].



Figura 1: Acúmulo de resíduos plásticos no meio ambiente. Adaptado de *Nature Communications* “The future of plastic” [9].

Comercialmente, os principais polímeros são: polipropileno (PP), poliestireno (PE), cloreto de polivinila (PVC), poliuretano (PU) e polietileno tereftalato (PET) [10,11]. O PET – estrutura monomérica representada na Fig. 2 – é um dos poliésteres sintéticos mais usados no mundo devido a sua excelente combinação de propriedades mecânicas, químicas e térmicas [12]. Sua produção se dá pela polimerização de etileno glicol (EG) e ácido tereftálico (TPA), onde suas unidades básicas são associadas por ligações ésteres [13,14]. O produto dessa reação é um composto com muitas ligações carbono-carbono (C–C) e grupos funcionais pouco reativos na cadeia principal, o que lhe

confere baixa mobilidade e cristalinidade, além de uma superfície hidrofóbica. Essas características acabam dificultando a biodegradação do polímero pela mera exposição ao meio ambiente [5].

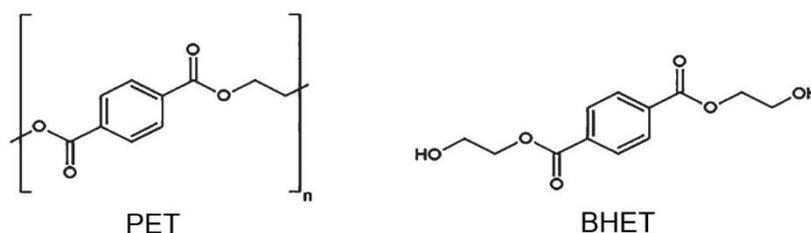


Figura 2: Estruturas químicas do Polietileno Tereftalato (PET) e do bis(2-hidroxiethyl tereftalato) (BHET) [15].

Como a maioria dos polímeros sintéticos, o PET é derivado do petróleo e estima-se que até 2050 sua produção seja responsável por 20% do uso do petróleo mundial [14,16]. Dessa maneira, o gerenciamento do acúmulo de plástico é um assunto de suma importância. Uma das possibilidades seria a substituição dos polímeros sintéticos por seus equivalentes biodegradáveis. No entanto, como suas propriedades físico-químicas não se equiparam às das dos plásticos convencionais, eles ainda correspondem a uma baixa porcentagem do total produzido – além da reciclagem manter os polímeros convencionais como o PET em circulação [4,12].

Dentre as estratégias existentes para a reciclagem/recuperação do PET, as principais são: os tratamentos físico, químico e biológico. O tratamento físico converte o polímero em novos produtos através da remodelagem (trituração) e aquecimento. Entretanto, esse tipo de processo compromete suas propriedades e, apesar da quantidade de PET reciclado através desse método ter aumentado, o baixo custo do PET puro/novo ainda dificulta uma maior implementação desse método de reciclagem [5,12]. No tratamento químico, por outro lado, o polímero é convertido em seus monômeros básicos através de reações químicas sem que ocorra uma degradação significativa. Contudo, entre as desvantagens apresentadas por esse método, as principais são o alto consumo de energia e os resíduos gerados, que acabam contribuindo para a poluição ambiental [12,17]. Diante das limitações nas técnicas físico-químicas tradicionais de reciclagem, o tratamento biológico (baseado no emprego de enzimas produzidas por microorganismos) se destaca ao se propor como uma rota alternativa que não traz os problemas apresentados pelos outros métodos [12,18]. Além disso, o gerenciamento via tratamento biológico se mostra promissor especialmente para polímeros com estruturas químicas passíveis de hidrólise, como é o caso do PET [10].

No que diz respeito ao PET, são conhecidos diversos microorganismos que produzem enzimas com a capacidade de realizar sua degradação. Essas enzimas são conhecidas como PET-hidrolases, podendo ser do tipo esterases, lipases ou cutinases [4,12]. São alguns desses microorganismos: *Thermobifida fusca*, *Thermomyces insolens*, *Humicola insolens*, *Candida antarctica*, *Saccharomonospora viridis*, *Bacillus subtilis*

e *Fusarium Solani*, para citar alguns [4,5,19]. Com isso, mono(2-hidroxiethyl) tereftalato (MHET), ácido tereftálico (TPA) e bis(2-hidroxiethyl) tereftalato (BHET) são compostos bem conhecidos como resultantes da reação de degradação do PET [16,20–22] – o último destes, o BHET, por se tratar de um composto intermediário e com estrutura bastante semelhante ao PET (Fig. 2), é também muitas vezes utilizado como substrato para estudar o funcionamento das enzimas produzidas por esses microorganismos [23–25].

Entretanto, o PET apresenta temperatura ideal para sua degradação próxima dos 70 °C – conhecida como temperatura de transição ( $T_g$ ) [5]. Nessa temperatura a mobilidade das cadeias do polímero aumenta e ele se torna mais suscetível ao processo de hidrólise, condição que obrigaria as enzimas a serem resistentes a altas temperaturas (termoestáveis). Dessa forma, o número de bactérias capazes de hidrolisar o PET próximo de  $T_g$  se torna bastante reduzido, além da exposição prolongada a altas temperaturas comprometer o desempenho hidrolítico das enzimas, sendo essas algumas das razões para o interesse em enzimas capazes de operar em temperaturas mais baixas do que a temperatura de transição do PET [5,21,26,27].

Em 2016, na universidade de Keio, Japão, Yoshida *et al.* identificaram uma nova bactéria: a *Ideonella sakaiensis* (*I. sakaiensis*) – Fig. 3. Ela produz um par de enzimas capaz de degradar o PET através da clivagem de suas cadeias de carbono, tornando-o sua principal fonte de energia. Essas enzimas, nomeadas PETase e MHETase, trabalham em conjunto no processo de degradação do polímero [10–12]. A PETase da *I. sakaiensis* (*IsPETase*), por apresentar atividade enzimática em temperaturas moderadas (por volta de 30 °C) e dispor de alta afinidade com o PET, dispensa as altas temperaturas ( $T_g$ ) que favorecem a degradação do polímero, o que as demais enzimas precisam devido às suas reduzidas afinidades com o esse polímero em comparação com a *IsPETase*. Essas propriedades únicas da *IsPETase* tornam a *I. sakaiensis* uma forte candidata para aplicação na biodegradação do PET [12,24]. Vale notar que a forma de implementação do que se discute neste trabalho, se a enzima será aplicada diretamente ou se suas enzimas serão utilizadas em, por exemplo, biorreatores para a reciclagem, não faz parte do escopo deste trabalho – voltado exclusivamente para o estudo e conhecimento da enzima.

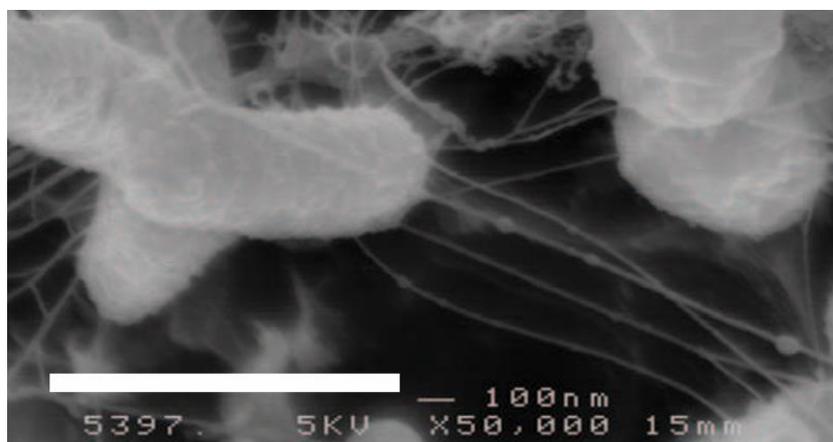


Figura 3: Microscopia eletrônica da *I. sakaiensis*. Adaptado de Yoshida et al., 2016 [28].

Considerando o acúmulo de polímeros sintéticos e a importância de aprimorar as técnicas para reduzir sua presença no meio ambiente, foram investigados neste trabalho aspectos estruturais e funcionais da PETase proveniente da bactéria *I. sakaiensis* (IsPETase). Para avaliar sua viabilidade como microorganismo no tratamento biológico desses poluentes, aspectos relacionados à interação proteína-ligante foram analisados. O estudo foi conduzido via simulação por dinâmica molecular de duas variantes da IsPETase obtidas através de mutação na região de seu sítio ativo (sem alterar os resíduos da tríade catalítica). Foram investigados aspectos estruturais e conformacionais da proteína, quais resíduos contribuem com o processo de ancoramento e a distância média do ligante à proteína. Os resultados obtidos complementam os dados experimentais a respeito das mutações implementadas. Os monômeros utilizados para a montagem dos diferentes sistemas estudados foram BHET e PET.

Adiante, no capítulo 2, são relatados alguns conceitos básicos relacionados às proteínas, a enzima estudada neste trabalho é apresentada, assim como aspectos sobre simulação computacional e algumas técnicas que complementam a simulação. No capítulo 4, a metodologia empregada é descrita – implementação das mutações, preparação dos complexos, os protocolos de dinâmica molecular empregados e as análises de dados realizadas. Os resultados estão no capítulo 5, seguidos da sua discussão no capítulo 6 e da conclusão no capítulo 7. Detalhes extras sobre proteínas e sobre alguns procedimentos apresentados na metodologia encontram-se nos apêndices.

## 2 Revisão da Literatura

### 2.1 Breves conceitos sobre proteínas

As proteínas são estruturas baseadas em aminoácidos (Apêndice A). Das diversas funções que podem apresentar – regulatórias do ciclo celular, estruturais, transportadora, sinalizadora etc –, há grande interesse nas suas funções enzimáticas. Esse é o caso onde elas agem sobre substratos reduzindo a energia de ativação necessária para que reações químicas específicas aconteçam [29]. Elas se ligam aos substratos através de seus sítios ativos que muitas vezes representam de 10 a 20% do volume total da proteína. Os resíduos de aminoácidos que constituem seu sítio ativo podem apresentar diversas sequências, estruturas e propriedades. Isso é uma importante característica que permite à enzima se ligar especificamente a diferentes substratos. Pequenas mudanças nos resíduos encontrados no sítio ativo alteram drasticamente a especificidade do substrato e conseqüentemente a reação que a enzima catalisa [30]. Do ponto de vista de suas funções e atividades, existem seis grandes categorias de enzimas: oxirredutases, transferases, hidrolases, liases, isomerases e ligases, em que cada grupo de enzimas é especializado em um tipo de processo [31].

### 2.2 A proteína da *Ideonella sakaiensis*

A análise completa do genoma da *IsPETase* revelou que, entre as seis grandes categorias de enzimas mencionadas anteriormente, ela pertence ao grupo das hidrolases (especializadas em hidrólise), mais especificamente da superfamília  $\alpha/\beta$  hidrolase [11]. A *IsPETase* apresenta a conformação canônica dessa superfamília que, mesmo com uma sequência de aminoácidos bastante diferente, compartilha um enovelamento semelhante às demais enzimas pertencentes a esse grupo – carboxilesterases e cutinases [32]. O enovelamento clássico de  $\alpha/\beta$ -hidrolase da *IsPETase* foi confirmado através de cristalografia de proteínas e difração de raios X. De acordo com os dados experimentais, a enzima é composta por 265 resíduos de aminoácidos cuja estrutura secundária é predominantemente formada por 9 fitas- $\beta$  e 7 hélices- $\alpha$  – Fig. 4-A [7]. Enquanto outras enzimas análogas apresentam apenas uma ponte dissulfeto em sua estrutura, a *IsPETase* possui duas. A ponte dissulfeto conservada (presente também em outras enzimas) une o terminal-C ao seu último laço, enquanto que a segunda ponte dissulfeto (específica da *IsPETase*) une os dois laços que abrigam a tríade catalítica da proteína [18,19].

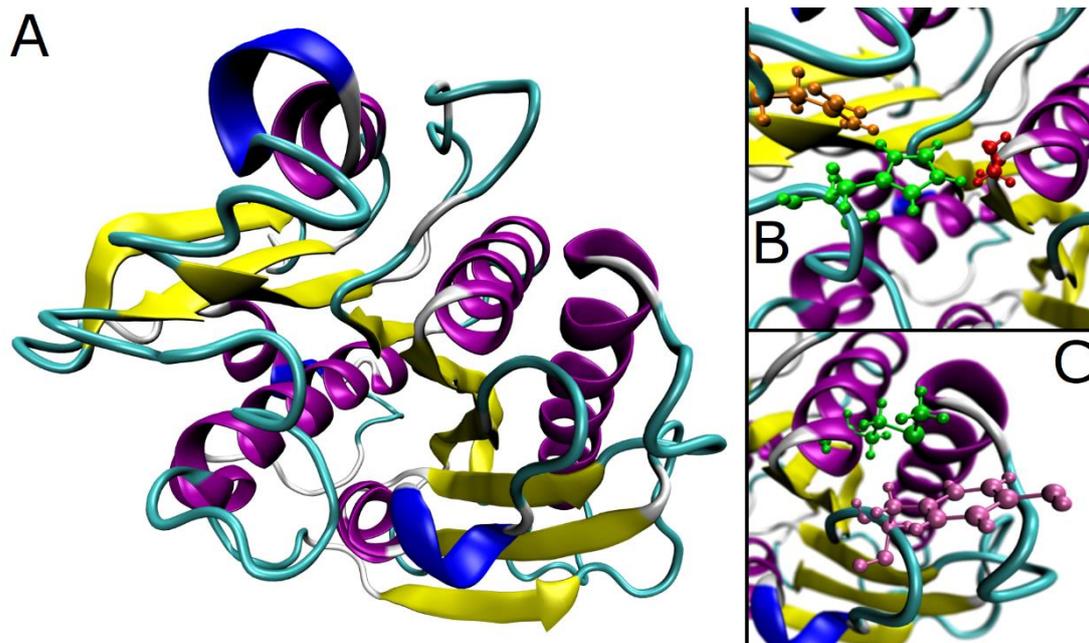


Figura 4: Em (A) a estrutura terciária (estilo cartoon) da *IsPETase* (código PDB: 6EQE) com suas principais estruturas secundárias: 9 fitas- $\beta$  (amarelo) – formando a folha- $\beta$  central da proteína – e 7 hélices- $\alpha$  (lilás). As duas principais estruturas do sítio ativo da *IsPETase* estão ampliadas ao lado, com os resíduos representados na forma esferas e bastões: (B) a tríade catalítica, com os resíduos Ser132 (vermelho), Asp178 (laranja) e His209 (verde); (C) a fenda de oxiânion, composta pelos resíduos Tyr59 (rosa) e Met133 (verde).

O sítio ativo abriga duas estruturas fundamentais para a atividade da proteína: a tríade catalítica e a fenda de oxiânion – Fig. 4-B e 4-C, respectivamente. A tríade catalítica tem a forma geral observada em outras carboxilesterases: Serina (Ser) – Aspartato (Asp) – Histidina (His), em que no caso específico da *IsPETase* compreende as seguintes posições: Ser132–Asp178–His209, onde a Ser132 é o nucleófilo do carbono da carbonila da ligação éster presente no substrato [9,16]. A fenda de oxiânion, por sua vez, é formada pelos resíduos Tirosina (Tyr) e Metionina (Met), respectivamente nas posições 59 e 133 (Tyr59-Met133), estando essa região na vizinhança imediata à tríade catalítica. Juntos, esse par de resíduos é responsável pela estabilização dos intermediários resultantes do processo de degradação. Esse par de estruturas trabalha em conjunto no processo de degradação realizado pela *IsPETase* [19,33].

### 2.3 O processo de degradação realizado pela *IsPETase*

O mecanismo de degradação da *IsPETase* é bastante discutido na literatura. Entre os modelos, há aquele que divide o processo em duas etapas, em que após o ligante se posicionar no sítio ativo da proteína ocorre a geração do *nick* (clivagem na cadeia polimérica) seguida da digestão do terminal polimérico – Joo, Seongjoon, et al., 2018; já o modelo proposto B.K. Ge, G.M. Hu e C.M. Chen., 2021, divide o processo em quatro etapas, em que o mecanismo se baseia em um ataque nucleofílico sobre a cadeia polimérica. Outros modelos ainda procuram explicar o mecanismo com base nos produtos gerados pela degradação (Han, Xu, et al, 2017) ou ainda sob aspectos voltados à bioquímica, como as interações inter-residuais das estruturas da proteína (Burgin, Tucker,

et al, 2024) ou substituições nucleofílicas e os produtos que elas geram (Eugene Shrimpton-Phoenix, John B. O. Mitchell e Michael Bühl, 2022) [23,24,33–35]. Contudo, apesar de haver algumas diferenças nos modelos propostos, é possível fazer uma generalização do mecanismo da *IsPETase*. No presente trabalho isso será feito com base no modelo de Ge, Hu e Chen [23], esquematizada na Fig. 5. Inicialmente, a proteína e o ligante encontram-se separados – Ligante (LIG) + Proteína (Prot). O processo é iniciado a partir do ancoramento do ligante para a formação do complexo ligante-proteína (LIG↔Prot). O ancoramento é tal que a ligação ester do substrato fica posicionada próxima da Ser132 que então realiza o ataque nucleofílico sobre a carbonila do ligante, dando início à reação de degradação. Esse posicionamento é favorável ao ataque nucleofílico desde que  $d_{HC}$  seja de aproximadamente 4 Å ( $d_{HC} \approx 4$  Å), como indicado por Vyas, Bhawna, *et al.*, 2018 [36], em que  $d_{HC}$  é definido como a distância entre oxigênio da hidroxila da Ser132 ( $^{OH}O_{Ser132}$ , o nucleófilo da proteína) e o carbono da carbonila do ligante ( $^{COOH}C_{LIG}$ ) – distância nucleofílica.

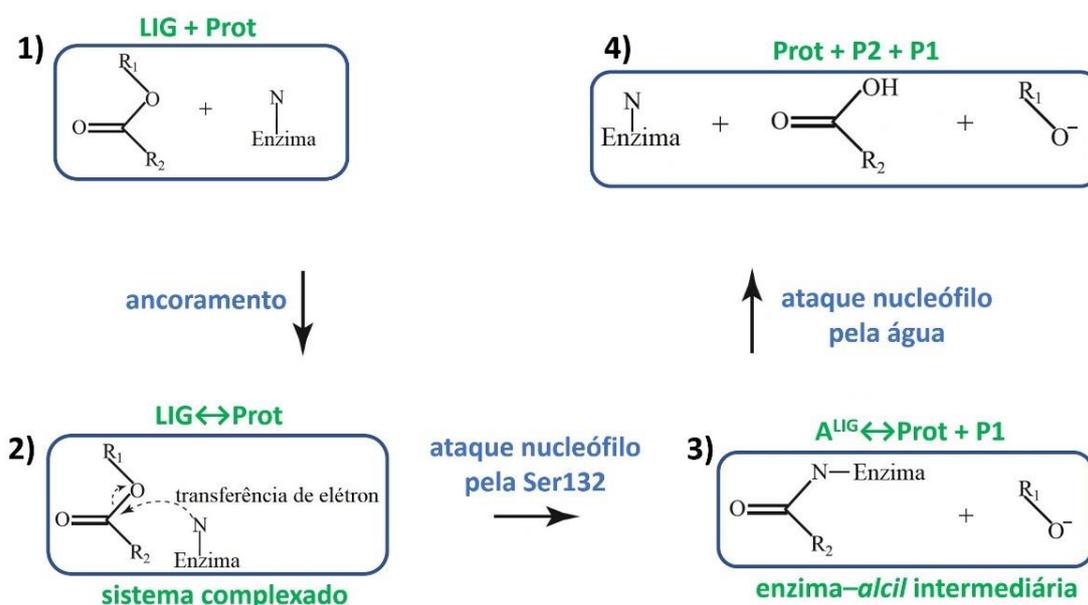


Figura 5: Esquema simplificado do processo de degradação realizado pela *IsPETase*. Baseado no fluxograma apresentado por Ge, Hu e Chen [23].

Então, a partir dessa primeira reação surge o primeiro produto (P1), ao mesmo tempo em que o complexo assume a forma intermediária enzima-ácil ( $A^{LIG↔Prot} + P1$ ). A próxima etapa acontece com um segundo ataque nucleofílico, agora realizado pelo oxigênio da água – novamente sobre a carbonila do ligante. Esse processo gera um segundo produto (P2) e o sistema retorna ao seu estado inicial – com a proteína isolada – com o fim da degradação do ligante (Prot + P1 + P2).

## 2.4 Ferramental computacional

Neste trabalho, foi realizado um estudo teórico computacional sobre o mecanismo da *IsPETase* para degradação do PET e para isso foi utilizado o método de simulações

por Dinâmica Molecular (DM) [37]. Essa é uma das principais técnicas de simulação, baseada na solução numérica das equações clássicas de movimento para uma determinada coleção de átomos [38]. Ela é bastante importante no estudo das estruturas, funções e dinâmicas das biomoléculas [37]. É usualmente o método computacional escolhido para o estudo de grandes moléculas como as proteínas [38]. Uma de suas principais vantagens é a possibilidade de fornecer detalhes em escala atômica sobre a dinâmica de um sistema em um grau muito maior do que aquele fornecido pelo experimento real [37].

A técnica das simulações via DM alcançou uma maturidade que pode ser efetivamente usada para entender as relações entre estrutura e função de macromoléculas [39]. Há duas aproximações básicas inerentes à DM sobre as quais os estudos podem ser conduzidos. A primeira abordagem remete aos casos em que os efeitos quânticos não são desprezados, tornando necessário o uso de integrais de caminho – abordagem capaz de descrever o comportamento dos elétrons sob aspectos quânticos, permitindo o estudo de interações moleculares e propriedades do sistema com maior precisão. Na segunda aproximação, os átomos são entendidos de acordo com a mecânica Newtoniana, ou mecânica clássica (DM clássica) [38].

A DM clássica segue o conceito de que em primeiro lugar deve-se considerar alguma lei que descreva as interações entre os átomos do sistema. Essa lei costuma ser desconhecida, mas pode ser aproximada com diferentes graus de precisão por modelagem via cálculos de estrutura eletrônica ou pelos *force fields* (campos de força). O campo de força é uma expressão matemática que descreve a dependência da energia de um sistema sobre as coordenadas de suas partículas, é uma forma analítica da energia potencial interatômica aliada a alguns parâmetros. Esses parâmetros geralmente são obtidos por cálculos de mecânica quântica *ab initio* ou semiempíricos ou ainda por adequação a dados experimentais conhecidos. O campo de força substitui os potenciais verdadeiros por modelos simplificados para facilitar os cálculos, mas ainda detalhados o suficiente para reproduzir as propriedades de interesse do sistema estudado [40].

Em seguida é necessário um algoritmo para integrar as equações de movimento dos átomos no sistema e por último, para que a integração ocorra, as condições iniciais devem ser estabelecidas. A simulação por DM permite um controle muito maior sobre parâmetros do sistema (temperatura, pressão, configuração atômica, por exemplo) do que nos casos experimentais [37]. Seguindo o caminho da DM clássica e em posse do sistema, as forças sobre cada átomo são obtidas pela derivação das equações Newtonianas (os campos de força) de onde a energia potencial é deduzida a partir da estrutura molecular do sistema. As características simplificadas das macromoléculas biológicas são representadas da seguinte maneira pelo campo de força: ligações covalentes de partículas ligadas são descritas por um termo harmônico simples; a variação angular dessas ligações é descrita por meio de uma aproximação do potencial harmônico; as torsões são representadas normalmente por uma série de cossenos; as interações de van der Waals são descritas a partir de uma forma simplificada do potencial de Lennard-Jones; as

interações eletrostáticas são dadas pela lei de Coulomb. Essa abordagem garante que os cálculos de força e energia sejam realizados de forma mais eficiente até mesmo para sistemas grandes. Entre os códigos de simulação mais populares estão: CHARMM, GROMACS e NAMD e AMBER [39].

Neste trabalho foi utilizado o *software* AMBER, acrônimo para *Assisted Model Building with Energy Refinement*, na sua versão 18 [41]. Trata-se de um conjunto de programas que calcula a trajetória das simulações via DM, oferecendo tanto uma abordagem híbrida (com mecânica quântica e mecânica molecular clássica – QM/MM) quanto a abordagem clássica da DM (através da solução das equações de movimento Newtonianas). Para o estudo desenvolvido neste trabalho foi adotada a abordagem clássica – como descrita em parágrafos anteriores. O AMBER18 é dividido em duas partes: AmberTools (um conjunto de *softwares* gratuitos) e Amber (centrado no programa *pmemd*), este dependente de licença para sua aquisição.

#### 2.4.1 Alguns programas do AmberTools

O AmberTools é constituído por códigos que permitem a criação de campos de força, preparação básica para simulações com AMBER, ferramentas para analisar a estrutura/dinâmica nas trajetórias etc. Entre as ferramentas disponíveis no AmberTools estão: LEaP, *parmed*, Antechamber, *cpinutil.py*, o *script* em Python MMPBSA.py e o módulo *sander*. O LEaP, acrônimo construído a partir dos nomes de antigos softwares do AMBER que ele substituiu – do inglês *link, edit and parm* –, é usado para preparar arquivos de entrada para os programas do AMBER. Ele constrói novos resíduos e moléculas, une resíduos e cria complexos moleculares não ligados, modifica coordenadas internas de moléculas, etc. Ele também é a ferramenta básica para construção de campos de força. O *parmed* edita, predominantemente, arquivos de topologia. Através dele pode ser feita a edição de ângulos interatômicos, mudança de estados de protonação de resíduos, remoção de ligações atômicas etc. Ele também modifica o arquivo de topologia para produzir um novo arquivo para ser trabalhado com o *sander*, *pmemd* e NAB (*Nucleic Acid Builder*). Já o Antechamber é uma reunião de programas escrito para ser utilizado em conjunto com o campo de forças generalizado do AMBER (GAFF), voltado para geração de arquivos de moléculas orgânicas (e alguns metais centrais em proteínas). Entre os programas disponíveis está o *antechamber* (seu principal programa) e o *parmchk2*. O *antechamber* é utilizado para processar o .pdb de praticamente qualquer tipo de molécula para gerar arquivos compatíveis para o LEaP. Apresenta funções como: trocar formatos de arquivos moleculares (.prep, .pdb, .mol2, .gzmat etc), atribuição de cargas atômicas, tipos de átomos e ligações etc [41,42]. Se alguma informação ainda estiver faltando ao final do processo realizado pelo *antechamber*, o *parmchk2* é a ferramenta recomendada para a criação de um arquivo que vai assistir o usuário na geração dos parâmetros necessários. Será escrito um arquivo de modificação de campo de força (*frmod*) contendo quaisquer parâmetros que sejam necessários para a molécula, mas que não foram supridos pelo campo de força [41]. Com relação ao pH, é possível implementar a titulação da cadeia

lateral de alguns resíduos – Asp, Glu, His, Lys, Tyr e Cys. O processo começa pela mudança do nome dos resíduos no arquivo *.pdb* de modo que o *.prmtop* tenha um átomo de hidrogênio definido em todos os possíveis sítios de protonação. Isso é feito com o auxílio da ferramenta *cpinutil.py*. É um *script* em Python é utilizado para a geração do arquivo (*.cpin*) que descreve quais resíduos devem ser titulados, os possíveis estados de protonação e suas energias.

O *script* MMPBSA executa cálculos de *Molecular Mechanics / Poisson-Boltzmann (or Generalized Born) Surface Area*, ou pelo acrônimo MM/PB(GB)/SA. É um método de pós-processamento em que estruturas instantâneas (*snapshots*) de um conjunto de conformações são usadas para calcular as variações na energia livre entre dois estados (geralmente entre um estado ligado e um livre de uma proteína e um ligante). As diferenças de energia são referentes à energia da fase vapor (*gas phase*) e às duas componentes da energia de solvatação (polar e não polar). Na fase vapor a energia está relacionada às interações intra e intermoleculares (proteína, ligante, complexo), calculadas pelo MM. Elas são independentes do modelo de solvente escolhido porque nos cálculos o solvente não é considerado e o sistema é tratado como se estivesse no vácuo. Para a componente polar da energia de solvatação são consideradas as interações eletrostáticas de longo alcance entre a molécula e o solvente, modeladas pelo PB (ou GB, uma forma aproximada da equação de Poisson-Boltzmann). Por outro lado, para a componente não polar, são consideradas as interações hidrofóbicas entre a molécula e o solvente, modelada pelo SA (proporcional à área de superfície acessível ao solvente) [41].

O *Simulated Annealing with NMR-Derived Energy Restraints*, ou *sander*, é o módulo do AMBER que realiza minimização de energia, dinâmica molecular e refinamentos NMR (ressonância magnética nuclear). É usado para diversas simulações que não se limitam a NMR. Entre as funcionalidades do *sander* podem ser citadas como exemplo: procedimento da malha de partículas de Ewald (*Particle Mesh Ewald*), simulações não periódicas, cálculos de energia livre, refinamento estrutural NMR, cálculos QM/MM etc. O *sander* possui o campo de força básico na seguinte forma:

$$\begin{aligned}
 V(\mathbf{r}) = & \sum_{\text{ligações}} \frac{K_b}{2} (b - b_0)^2 + \sum_{\text{ângulos}} \frac{K_\theta}{2} (\theta - \theta_0)^2 \\
 & + \sum_{\text{diedrais}} (V_n/2)(1 + \cos[n\varphi - \delta]) \\
 & + \sum_{\substack{\text{não ligadas} \\ i,j}} (A_{ij}/r_{ij}^{12}) - (B_{ij}/r_{ij}^6) + (q_i q_j / r_{ij})
 \end{aligned} \tag{1}$$

Na equação (1) o primeiro termo é o potencial harmônico que representa as extensões espaciais de átomos ligados; o segundo termo é o potencial harmônico para as variações angulares das ligações; o terceiro termo representa as torsões das ligações

através de uma série de cossenos; o quarto termo representa as interações eletrostáticas (potencial de Lennard-Jones e Coulombiano) [41].

Em seguida foi desenvolvido o *pmemd*, com o objetivo de melhorar a performance dos métodos mais utilizados no *sander*. Não se trata de uma atualização completa do *sander*, na verdade seu objetivo é ser uma implementação da funcionalidade que mais será usada para lidar com um sistema de longa escala temporal solvatado de forma explícita ou implícita.

## 2.4.2 Simulação pelo AMBER

O *pmemd*, Dinâmica Molecular por Malha de Partículas de Ewald (*Particle Mesh Ewald Molecular Dynamics*), é uma técnica de simulação baseada no método de Ewald para calcular forças de natureza Coulombiana em sistemas periódicos. Um exemplo básico de sistema periódico é um sistema de íons pontuais que interagem via potencial de Coulomb. Considerando o sistema em uma célula unitária com condições periódicas de contorno, o potencial desse sistema assume a forma de uma série infinita:

$$E(\mathbf{r}_1, \dots, \mathbf{r}_n) = \frac{1}{2} \sum_{\mathbf{n} \in \mathbb{Z}^3}^* \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j - \mathbf{n}|}, \quad (2)$$

onde  $\mathbf{r}_i$  é a posição de  $q_i$ ,  $\mathbf{r}_j$  é a posição de  $q_j$  e  $\mathbf{n}$  representa a posição das partículas na malha periódica. O método de Ewald faz duas alterações no sistema de modo a superar o problema da série infinita indicada na equação (2). Primeiro, cada íon é neutralizado com uma Gaussiana esférica, de carga oposta e centralizada no próprio íon, de forma que cada par neutralizado íon-gaussiana se torna a parte real da soma de Ewald (agora de curto alcance). Em seguida, uma segunda Gaussiana esférica é adicionada, também centrada no íon. No entanto, a segunda Gaussiana possui a mesma carga do íon, de modo a anular o efeito da primeira Gaussiana introduzida. O potencial para essas Gaussianas é obtido pela equação de Poisson e resolvido no espaço recíproco. De forma geral, o método separa a série infinita (de convergência lenta) em uma soma de duas séries convergentes (de convergência rápida): uma soma direta no espaço Cartesiano (real) e uma soma recíproca no espaço de Fourier – mais o termo de correção a respeito das auto interações das cargas:

$$E_{dir} = \frac{1}{2} \sum_{\mathbf{n} \in \mathbb{Z}^3}^* \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j \text{erfc}(\beta |\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|)}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{n}|}, \quad (3)$$

$$E_{rec} = \frac{1}{2\pi V} \sum_{\mathbf{m} \in \{\mathbb{Z}^3 \setminus \{0\}\}}^* \frac{\exp\left(-\frac{\pi^2 \mathbf{m}^2}{\beta^2}\right)}{\mathbf{m}^2} S(\mathbf{m}) S(-\mathbf{m}), \quad (4)$$

$$E_{corr} = \sum_{(i,j) \in M} \frac{q_i q_j \operatorname{erf}(\beta |\mathbf{r}_i - \mathbf{r}_j|)}{|\mathbf{r}_i - \mathbf{r}_j|} - \frac{\beta}{\sqrt{\pi}} \sum_{i=1}^N q_i^2, \quad (5)$$

Na equação (3)  $E_{dir}$  trata das interações eletrostáticas entre as partículas, com  $\operatorname{erfc}()$  – função erro complementar – reduzindo a contribuição das interações de longo alcance.  $E_{rec}$ , na equação (4), trata das interações de longo alcance, com  $\exp()$  garantindo a convergência da série, com  $S(\mathbf{m})$  e  $S(-\mathbf{m})$  sendo os fatores de estrutura contendo informações sobre as posições das cargas no espaço recíproco. Na equação (5)  $E_{corr}$  representa correções sobre as interações de curto alcance no espaço real (primeiro somatório), com  $\operatorname{erf}()$  – função erro – evitando singularidades e desconsiderando as auto interações advindas da repetição na malha (segundo somatório) [43,44].

### 2.4.3 Conceitos de Docking molecular

A técnica de *docking molecular* (ancoramento molecular) é geralmente utilizada para prever os modos de ligação entre receptores e ligantes. Ela possui diversas aplicações, como desenvolvimento de fármacos, interações de DNA/RNA, modelagem de enzimas e catálise etc. Essa técnica basicamente busca o modo de ligação mais favorável de um ligante no receptor de interesse (geralmente uma proteína) [45,46].

Todas as formas de *docking* precisam de uma *score function* para classificar os vários modos de ligação e um método de busca para explorar as variáveis de estado. A *score function* pode ser empírica ou baseada em campo de força. Já os métodos de busca podem ser classificados conforme a amplitude da busca no espaço de exploração, que pode ser: (a) local, com tendência para encontrar o mínimo mais próximo ou mínimo local de energia da conformação atual; (b) global, que procuram o melhor mínimo ou o mínimo de energia global dentro do espaço de busca definido [46].

Entre os modelos de formação do complexo proteína-ligante, destacam-se dois. O primeiro deles, chave-fechadura (*lock-and-key*) exibido na Fig. 6-A, trata os ligantes e proteínas de forma rígida. Esse modelo enfatiza a complementaridade geométrica. O segundo modelo, indicado na Fig. 6-B é o chamado encaixe-induzido (*induced fit*), onde os receptores e ligantes adotam a estrutura mais estável que minimiza sua energia livre. Os *softwares* de simulação de *docking* procuram a melhor conformação e orientação de acordo com o método de busca implementado, seguido da aplicação de uma *score function* para prever a afinidade da ligação e analisar a interação [45].

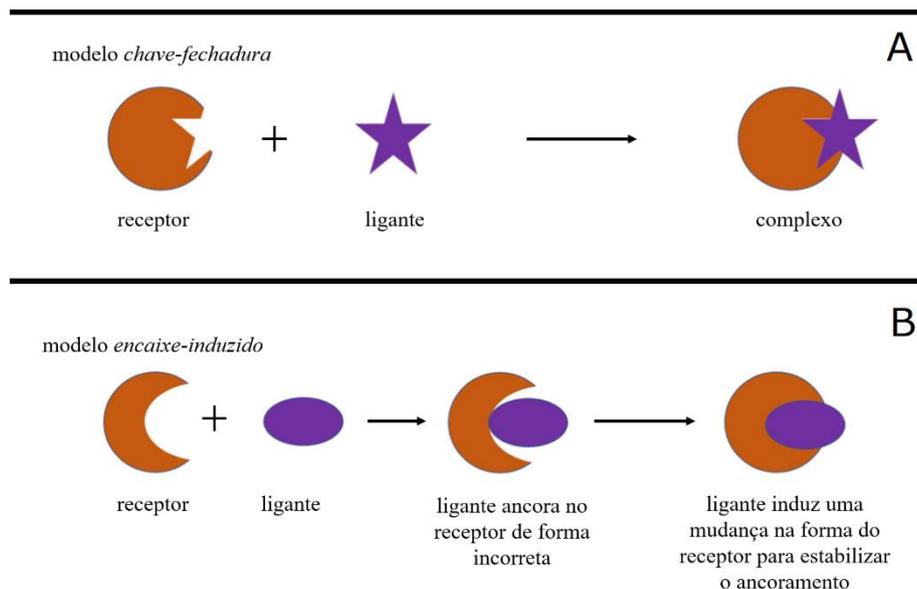


Figura 6: Modelos de *docking* mais conhecidos: (A) *chave-fechadura* e (B) *encaixe induzido*. Adaptado de [47].

#### 2.4.4 Conceitos sobre simulação com pH constante

A técnica de simulação com pH constante (*Constant pH Molecular Dynamics* – CpHMD) disponível no AMBER, faz a amostragem das estruturas no *ensemble* semi-grande canônico com função de probabilidade de distribuição dada pela equação (6):

$$\rho(\mathbf{q}, \mathbf{p}, \mathbf{n}) = \frac{\exp(\beta\mu^*n - \beta H(\mathbf{q}, \mathbf{p}, \mathbf{n}))}{\sum_{n'} \int d\mathbf{p}' d\mathbf{q}' \exp(\beta\mu^*n' - \beta H(\mathbf{p}', \mathbf{q}', \mathbf{n}'))} \quad (6)$$

onde  $\beta = 1/k_bT$ ,  $\mu^*$  é o potencial químico do hidrônio,  $\mathbf{q}$  representa as coordenadas generalizadas do sistema,  $\mathbf{p}$  representa o momento conjugado e  $n$  é o número de prótons tituláveis naquele estado;  $\mathbf{n}$  (em negrito) se refere ao vetor do estado de protonação – o denominador da equação é a função de partição do *ensemble* semi-grande canônico. Para amostrar a função de probabilidade mencionada, métodos de estados de protonação discretos combinam dinâmica molecular (DM) e amostragem de estados de protonação baseada em Monte Carlo (MC) ao longo da trajetória. A DM mantém fixa os estados de protonação para explorar as conformações e momentos do sistema, enquanto o MC mantém fixas as conformações para explorar os estados de protonação [48].

No método proposto por Swails, Jason M., Darrin M. York, e Adrian E. Roitberg, 2014, a DM padrão é realizada com uma quantidade constante de estados de protonação fornecidos no início da simulação. Em algum momento durante a simulação ela é parada, o solvente é removido e um conjunto de N mudanças de estados de protonação são tentadas, onde N é o número de resíduos tituláveis – nesse método o intervalo de tempo para as tentativas é predefinido. Com a DM parada, são propostas mudanças nos estados de protonação para cada resíduo titulável (de forma aleatória). Se a mudança no estado de protonação é aceita, o estado atual do sistema é atualizado e o resíduo seguinte,

escolhido aleatoriamente (sem repetição) é titulado a partir do novo estado. Se qualquer mudança no estado de protonação for aceita, o soluto (molécula ou complexo) é “congelado” enquanto se faz uma DM no solvente para relaxar a distribuição do solvente com base nos novos estados de protonação. Concluída essa relaxação, as velocidades dos átomos do soluto são restabelecidas e a simulação padrão continua normalmente até a próxima parada, reiniciando o processo [48].

#### 2.4.5 Método de redistribuição de massa dos átomos

O tamanho do passo de integração da simulação é limitado pela escala temporal do movimento que apresenta a maior frequência no sistema, que geralmente são ligações que envolvem átomos de hidrogênio – que impõem um “limite de 1 fs”. Simplesmente aumentar o passo de integração pode tornar a simulação instável, aumentando as chances de instabilidades quanto mais longa ela for. Para superar esse problema são implementados algoritmos que restringem a movimentação das ligações de alta frequência, como o SHAKE para ligações não aquosas e o SETTLE para ligações aquosas. Com esses algoritmos o passo de integração pode ser aumentado para 2 fs, valor limite devido às limitações no algoritmo do SHAKE) [49].

No trabalho de Chad W. Hopkins et al., 2015, o conceito que embasa a ideia de alterar massas atômicas é que as médias termodinâmicas de equilíbrio de observáveis não dependem da massa exata do sistema. Considerando o *ensemble* canônico essas médias tomam a forma da equação (7):

$$\langle A \rangle = \frac{\int A e^{-\beta H} dp dx}{\int e^{-\beta H} dp dx} \quad (7)$$

Assumindo que é possível separar posição e momento em  $A = A(x)$  e na Hamiltoniana (como é o caso de DM com campos de força clássicos sem componentes magnéticos), as partes dependentes do momento podem ser separadas e simplificadas. A consequência disso é que a expressão final não tem dependência da distribuição de massa do sistema.

Diante disso, Feenstra et al., 1999, determinaram uma forma de repartição da massa dos átomos no sistema de maneira que a massa total permanecesse constante: a massa dos átomos leves de hidrogênio é aumentada enquanto a massa dos átomos pesados (ligados a esses hidrogênios) é reduzida na mesma quantidade – essa é a base para implementar o *Hydrogen Mass Repartitioning* (HMR). Dessa maneira o aumento no passo de integração pela técnica de HMR não introduz erros significativos em simulações longas e pode manter a simulação estável durante longos períodos. Além disso, vale notar que, com base em várias simulações realizadas, o trabalho de Chad W. Hopkins et al., 2015, conclui que o HMR aplicado junto com o SHAKE melhora a estabilidade do sistema simulado. O método apresentado pelo grupo de pesquisa permite aumentar o passo de integração por um fator de  $\sim 2$  [49].

#### 2.4.6 Mutações residuais em proteínas

Realizar mutações nas proteínas para estudar seu comportamento é um método bastante utilizado na literatura, seja em trabalhos experimentais ou teóricos. Substituir resíduos específicos para investigar sua relevância nos processos realizados pela enzima (Maria Eduarda Sevilla et al., 2023 e Chun-Chi Chen et al., 2018) ou ainda para aprimorar sua atividade enzimática por meio de mudanças estruturais locais (Harry P. Austin et al., 2018), são algumas coisas que podem ser estudadas por meio da aplicação de mutações [7,50,51].

O trabalho Cui, Yinglu, et al., 2021, apresentou uma estratégia computacional para implementar várias mutações na *IsPETase*. Essas mutações foram divididas em grupos e o grupo contendo a mutação I180V, segundo os dados experimentais apresentados, apesar de resultar em um aumento de 1,5 °C na estabilidade térmica, a mutação causou um grande comprometimento da atividade enzimática da proteína [52]. A mutação I180A foi investigada experimentalmente nos trabalhos de Han, Xu, et al., 2017 e Joo, Seongjoon, et al., 2018. Em seus trabalhos, investigando o mecanismo de degradação da *IsPETase*, introduziram a mutação I180A e seus resultados experimentais apontaram, respectivamente, uma redução na atividade enzimática da proteína de aproximadamente 80% e 50% [24,53]. Além disso, os resíduos substituídos são alifáticos, hidrofóbicos e pequenos, com poucas chances de comprometer a estrutura da proteína e bastante importantes para as funções de reconhecimento e ancoramento de ligantes [24,54].

## **3 Objetivo**

### **3.1 Objetivos gerais**

O objetivo deste trabalho é realizar mutações residuais no sítio ativo da *IsPETase* e, tomando como referência a proteína nativa, avaliar seus efeitos sobre a estrutura da proteína e investigar como essas mutações interferem nas interações da proteína com ligantes, através de simulações por Dinâmica Molecular.

### **3.2 Objetivos específicos**

- I. Gerar duas variantes da proteína pela substituição da Isoleucina 180 por uma Valina (variante I) e por uma Alanina (variante II)
- II. Fazer o *docking* molecular com dois modelos de ligante (BHET e PET) na forma de monômeros
- III. Gerar os complexos proteína-ligante com todos os três modelos de proteína (nativa, variante I e variante II)
- IV. Realizar simulações por Dinâmica Molecular (DM) com os complexos para gerar dados sobre as interações proteína-ligante
- V. Estudar os efeitos estruturais sobre a proteína devido às mutações implementadas
- VI. Estudar como a formação do complexo foi alterada pelas mutações implementadas

## 4 Metodologia

### 4.1 Gerenciamento dos processos

O trabalho foi realizado predominantemente com o auxílio do pacote AMBER18/AmberTools18. Contudo, para simplificar a execução, foram empregados diversos *scripts* em *Python* para gerenciar as ferramentas fornecidas pelo pacote AMBER: *ASM.py* (*Amber Simulation Manager*) para administrar a simulação e seus complementos (*Analysis\_plot.py*, *Prot\_transition.py*, *Multidock\_Vna.py*, *AvR\_Docking.py* e *decode\_mmpbsa.py*) para auxiliar o processo de análise de dados – mais informações em <https://github.com/Blodhor/Molecular-Dynamics> (o Apêndice B traz os detalhes sobre cada um dos *scripts* utilizados).

### 4.2 Mutações no sítio ativo da proteína

As mutações foram implementadas por meio do *script* em *Python* ASM apresentado na seção anterior. Contudo, cabe a descrição do processo automatizado pelo *script*. Considere o interesse em substituir o resíduo Serina 30 (Ser30) por uma Histidina (His). Em posse do arquivo *.pdb* da proteína, observa-se os átomos que compõem a Ser30 e mantém-se apenas aqueles que também fazem parte da His, apagando as linhas referente aos átomos exclusivos da Ser. Em seguida, substitui-se o código de três letras referente à serina (SER) pelo da histidina (HIS) nas linhas remanescentes salvando o *.pdb* com essas alterações – mesmo que falte alguns átomos ao resíduo substituído. O processo é finalizado quando o *.pdb* é carregado no *tleap* e processado conforme as funcionalidades apresentadas na seção 2.4.1, completando os átomos ausentes, corrigindo suas posições e cargas parciais etc.

Neste trabalho, as mutações foram implementadas no resíduo 180 da proteína, parte do seu sítio ativo. Nativamente esse resíduo é uma Isoleucina (Ile, I), que foi posteriormente substituído por uma Valina (Val, V) e Alanina (Ala, A) – estruturas químicas exibidas na Fig. 7 – para gerar as duas variantes estudadas neste trabalho: a variante I (I180V) e a variante II (I180A). Dessa forma, as proteínas para os complexos são: proteína nativa (WT), variante I e variante II. Os resíduos substituídos foram escolhidos com base em resultados experimentais discutidos na literatura e apresentados na sec. 2.4.6 [24,52,53].

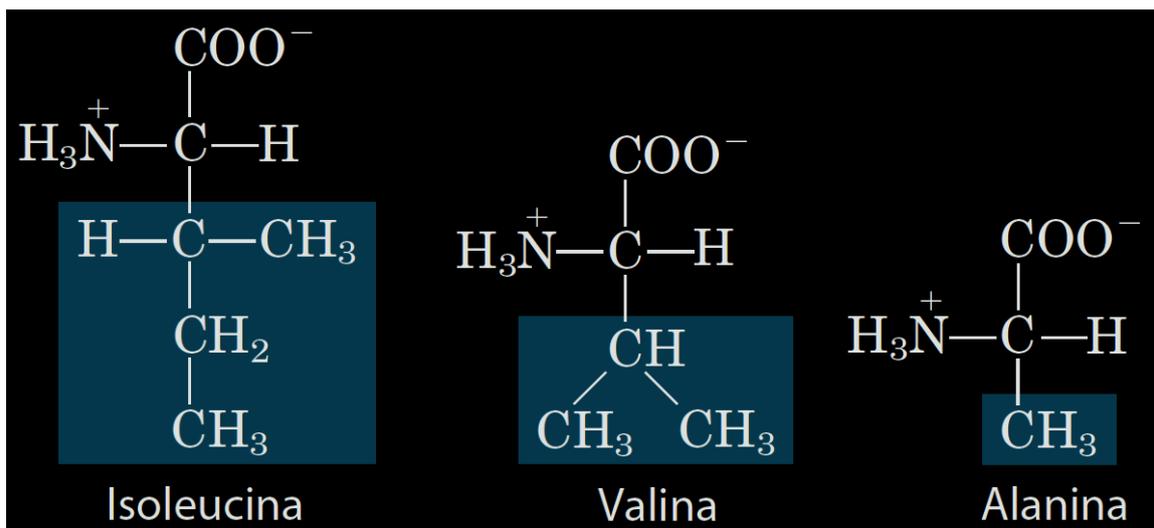


Figura 7: Representação 2D da estrutura química da Isoleucina (resíduo nativo da posição 180), Valina e Alanina (resíduos propostos para as mutações). Imagem adaptada de Lehninger, Albert L., David L. Nelson, e Michael M. Cox. *Lehninger principles of biochemistry*. Macmillan, 2005. [55].

### 4.3 Docking e preparação dos ligantes

Para o processo de *docking* (ancoramento) foi utilizada uma técnica que, dentro de uma área predefinida, procura a melhor forma de posicionar o ligante na proteína. No caso da *IsPETase* o local onde o ligante se posiciona já é conhecido: o sítio ativo da proteína, mais especificamente referenciado na Ser132. Então, com essa premissa, o programa realiza as buscas e atribui um *score* para cada uma das possíveis maneiras de posicionar o ligante na área de busca.

Para o entendimento da *score function* do AutoDock Vina (ATD), faz-se necessário o esclarecimento de alguns conceitos. O primeiro deles é o *consensus scoring*. Ele faz a classificação a partir de diferentes métodos e considera a interseção dos melhores *N%* de cada uma das listas – os resultados são ainda melhores a partir de uma combinação otimizada de funções de classificação [56]. Isso serviu de base para o desenvolvimento da *score function* X-CSCORE de Wang, R., Lai, L., & Wang, S. 2002. Nesse caso, uma série de algoritmos (cada um deles considerando efeitos importantes que envolvem o processo de ancoramento) foi utilizado para derivar funções de *score*. Vale notar que foram empregados três algoritmos diferentes para o efeito hidrofóbico, gerando assim três funções distintas. Assim, aplicando a estratégia *consensus scoring*, a média aritmética das funções foi obtida e nomeada X-CSCORE.

A implementação da função de classificação utilizada no Vina foi feita predominantemente inspirada pela X-CSCORE. A forma funcional geral da parte dependente da conformação da função de classificação com a qual o Vina foi desenvolvido para trabalhar é:

$$c = \sum_{i>j} f_{t_i t_j}(r_{ij}) \quad (8)$$

onde o somatório na equação (8) é sobre todos os pares atômicos que podem se mover uns em relação aos outros, desconsiderando átomos separados por menos de quatro ligações covalentes consecutivas. Para cada átomo  $i$  foi atribuído um tipo  $t_i$ , com o requerimento da definição de uma função simétrica de interação ( $f_{t_i t_j}$ ) com distância interatômica  $r_{ij}$ . O valor pode ser entendido como a soma das contribuições intermoleculares e intramoleculares, como na equação (9):

$$c = c_{inter} + c_{intra} \quad (9)$$

Houve uma atualização na função de classificação do ATD por Eberhardt, Jerome, et al., 2021. A função de classificação AD4 foi disponibilizada e se mostrou bastante diferente das outras funções do Vina. Ela usa um modelo baseado em van der Waals, com potenciais eletrostáticos e de ligação de hidrogênio direcionais (derivados de versões anteriores do campo de força AMBER), um termo de dessolvatação baseado em cargas parciais e uma penalidade conformacional entrópica simples. Contudo, usar as funções de classificação AD4 pode levar até 3x mais tempo do que as demais. É importante observar que o ATD disponibiliza um parâmetro que influencia na quantidade de tempo dedicada a cada busca por um posicionamento do ligante: *exhaustiveness* (cujo valor padrão é 8). Então, esse parâmetro permite alterar a robustez da busca realizada [57].

Dessa forma, para posterior geração dos sistemas complexados, os ligantes foram preparados como descrito a seguir. Foram utilizados dois modelos: 1) um monômero de PET que foi desenhado no Avogadro [58]; e 2) um monômero de BHET baseado no arquivo C8X.cif obtido a partir do PDB RCSB, com seu .pdb gerado via PyMOL [59]. Então, com o auxílio do ATD, as buscas para o melhor ancoramento foram realizadas em uma região cúbica com aresta de 20 Å, centralizada no oxigênio da hidroxila da Ser132 ( $^{\text{OH}}\text{O}_{\text{Ser132}}$ , o nucleófilo da proteína). Essa região foi amostrada 300 vezes para cada complexo e o valor de exaustividade (*exhaustiveness*, que controla o tempo gasto na busca pelo melhor ancoramento) foi definido em 10 e todas as conformações geradas (em torno de 2700) foram salvas.

Para a organização dos modelos de ligantes obtidos pelo *docking* no ATD, dois parâmetros foram considerados: o *vina.score* e a distância  $d_{\text{HC}}$ . O *vina.score* é a forma como o ATD qualifica o ajuste do ligante (energia de ligação) no interior da caixa cúbica definida para a busca, enquanto que  $d_{\text{HC}}$  representa a distância entre a Ser132 ( $^{\text{OH}}\text{O}_{\text{Ser132}}$ ) e o ligante ( $^{\text{COOH}}\text{C}_{\text{LIG}}$ ) – conforme definido na seção 2.3, com valor de interesse em torno de 4 Å. Dessa maneira, ficou definido o fator de minimização adimensional  $s_{\text{vd}}$  conforme a equação (10) apresentada a seguir:

$$s_{vd} = d_{HC} + vina.score, \quad (10)$$

que representa o melhor cenário quanto menor for o valor obtido. Feito isso, os modelos foram organizados em tabelas com base nos valores de  $s_{vd}$  – tabelas apresentadas na seção de resultados. Em seguida os ligantes foram parametrizados pelo *antechamber*, que dispõem dos seguintes métodos para atribuição de carga: RESP, AM1-BCC, CM2, Gasteiger e AM1 Mulliken [60]. Para este trabalho foi utilizado o método AM1-BCC, que visa reproduzir os potenciais eletrostáticos de moléculas dados pelo método de Hartree-Fock [61].

## 4.4 Preparação dos sistemas

A estrutura com PDB ID 6EQE foi a base para o modelo computacional da *IsPETase* [7,62]. Todos os sistemas com a proteína, seja com ela isolada ou complexada, foram preparados sob as mesmas condições. O campo de força *ff14SB* foi utilizado para estabelecer as propriedades da proteína – no caso das proteínas isoladas. Em se tratando dos complexos, o campo de força *GAFF* foi utilizado para estabelecer as propriedades do ligante. A solvatação foi feita em água modelo TIP3P (solvente explícito) em uma caixa octaédrica com arestas de 12 Å, com a adição de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ) para neutralizar a carga do sistema [63,64]. Além disso, a distância de corte para as interações entre átomos não ligados foi de 12 Å e as distâncias de ligações que possuem hidrogênio foram limitadas com o SHAKE [65]. Ao final do processo os arquivos de coordenadas dos átomos (.rst7) e topologia do sistema (.prmtop) foram gerados. Nesse estágio, os arquivos para simulação também foram configurados para simulação com pH constante – CpHMD, como apontado na seção 2.4.4 –, em que Tirosina 59 (Tyr59), Ácido Aspártico 178 (Asp178) e Histidina 209 (His209) foram escolhidos como resíduos tituláveis por pertencerem ao sítio ativo da proteína e à tríade catalítica, respectivamente. O valor de pH escolhido teve como base o trabalho de Son, Hyeoncheol Francis, et al., 2019. Nos experimentos realizados pelo grupo, conforme indicado em seu material suplementar, o estudo da atividade enzimática da *IsPETase* foi conduzido em uma solução de glicina e hidróxido de sódio em pH 9. Dessa forma, para a investigação do processo de ancoramento (parte do processo de degradação), foi adotado o valor de pH 9 em simulação com CpHMD. Também foi implementado o método HMR (apresentado na seção 2.4.5), o que possibilitou a realização das simulações com passo de integração maior – aumentado de 2 fs para 4 fs.

## 4.5 Pré-produção e produção (simulação)

Todas as simulações via DM foram realizadas com base no mesmo procedimento de quatro estágios – minimização, aquecimento, equilíbrio e produção – detalhados nos parágrafos a seguir.

Para a minimização do sistema são empregados em conjunto dois algoritmos. O primeiro deles, o *steepest descent* (gradiente descendente), é o método clássico mais antigo para a minimização de uma função não linear. A partir de uma função  $f(x)$  contínua, onde  $x$  é um vetor  $n$ -dimensional, o algoritmo considera que essa função deve diminuir, pelo menos inicialmente, se o sentido de seu gradiente negativo for tomado [66]. O método busca, por iteração, a derivada de  $f(x)$  que fornece o mínimo mais adequado com base na equação (11):

$$x_{i+1} = x_i - \alpha_t \nabla f(x_i) \quad (11)$$

onde  $\alpha_t$  é a taxa de incremento em cada iteração e  $\nabla f(x_i)$  é o gradiente em  $x_i$ . Contudo, uma das limitações do algoritmo está na lenta convergência, causada pela interferência de cada nova iteração na anterior [67]. O segundo algoritmo, o *conjugate gradient*, vem com o intuito de resolver esse problema. Ele trabalha de modo a desviar a direção do *steepest descent* adicionando a ele um múltiplo positivo da direção usada no passo anterior [68], conforme a equação (12):

$$x_{i+1} = x_i + \alpha_i d_i \quad (12)$$

onde  $\alpha_i$  é uma atualização no passo e  $d_i$  a direção de busca. Podem ser feitas  $n$  atualizações ( $n$  buscas em  $d_1, \dots, d_n$ ) se essas direções forem ortogonais entre si. Dessa maneira os erros advindos do *steepest descent* são reduzidos [68].

Para as demais etapas – aquecimento, equilíbrio e produção – a metodologia do AMBER é semelhante. Em todos os casos a variável  $ntb$  é definida de acordo com o *ensemble* escolhido: no aquecimento e na produção  $ntb=1$  (*ensemble* NVT) e na equilíbrio  $ntb=2$  (*ensemble* NPT). Com isso definido, as bases teóricas que sustentam cada uma das etapas são as mesmas. A variável  $ntt$ , por exemplo, define a escala de temperatura. Neste trabalho foi utilizada a configuração  $ntt=3$  – valor que define a dinâmica de Langevin – que traz consigo a condição de definição da variável  $gamma\_ln$  (a frequência de colisão,  $\gamma$ ). O termostato de Langevin (base de  $ntt=3$ ) mantém a temperatura escolhida através da aplicação de uma força de atrito  $\zeta$  e uma força aleatória  $R_i$  a todas as partículas  $i$  (de massa  $m_i$ ) para simular colisões aleatórias entre as partículas na simulação e as partículas imaginárias em um banho externo mantido em uma temperatura  $T$  escolhida. Nesse sistema, colisões com uma partícula  $i$  acontecem com frequência  $\gamma_i$ , na forma da equação (13) apresentada adiante:

$$\gamma_i = \frac{\zeta}{m_i} \quad (13)$$

de modo que a equação (14) fornece a força central da seguinte maneira:

$$\dot{\mathbf{p}}_i = \mathbf{f}_i - \gamma_i \mathbf{p}_i + \mathbf{R}_i \quad (14)$$

Vale destacar que a frequência de colisão  $\gamma$  não precisa necessariamente ser definida com alta precisão. Por exemplo:  $\gamma$  real para água tem o valor em torno de  $50 \text{ ps}^{-1}$ , mas os algoritmos entregam bons resultados mesmo com valores muito menores – os valores mais baixos são mais vantajosos em termos de amostragem ou estabilidade de integração, como 10 ou valores ainda mais baixos (entre 2 e  $5 \text{ ps}^{-1}$ ) [41,69].

No que diz respeito ao controle da pressão, a variável *barostat* permite a escolha de dois modelos: Berendsen (*barostat* = 1) ou Monte Carlo (*barostat* = 2). Neste trabalho foi empregado o barostato de Monte Carlo (MCB) – estocástico – por se tratar de um modelo mais eficiente e rigoroso, em que uma de suas vantagens é a possibilidade de ser usado para simular qualquer *ensemble* de interesse – como mencionado anteriormente, foram implementados dois *ensembles* diferentes. Uma das formas de aplicação do método de Monte Carlo tem como premissa considerar a probabilidade de se chegar em um determinado ponto  $\mathbf{r}^N$  proporcional ao fator de Boltzman  $\exp[-\beta U(\mathbf{r}^N)]$  [70]. A partir de então faz-se o seguinte:

- I. seleciona-se uma partícula aleatória e sua energia  $U(\mathbf{r}^N)$  é calculada
- II. para a partícula é dado um deslocamento aleatório,  $\mathbf{r}' = \mathbf{r} + \Delta\mathbf{r}$ , e sua nova energia  $U(\mathbf{r}'^N)$  é obtida
- III. a probabilidade de aceitação do deslocamento de  $\mathbf{r}^N$  para  $\mathbf{r}'^N$  é fornecida pela equação (15):

$$prob = \min (1, \exp \{-\beta[U(\mathbf{r}'^N) - U(\mathbf{r}^N)]\}) \quad (15)$$

Com isso, para a execução das simulações via DM, partiu-se do processo de minimização de energia com o objetivo de relaxar o sistema. Como essa etapa é realizada com a combinação de dois algoritmos, (gradiente descendente e gradiente conjugado, nessa ordem), a transição de um algoritmo para o outro é dada pelo valor de *ncyc* e o número máximo de passos do processo é definido por *maxcyc*. A minimização foi realizada estritamente sobre os átomos da cadeia principal da proteína (restrição dada por *restraintmask*) usando uma constante estabelecida por *restraint\_wt*. A Fig. 8 apresenta a padronização do protocolo para a minimização.

### Etapa de Minimização de Energia (Energy Minimization Stage)

```
&cntrl
imin=1,                !escolhe uma minimização
igb=0,                !não usa nenhum termo generalizado de Born
ntb=1,                !condições de contorno periódicas com volume constante
cut=12.0,             !distância de corte de interações não ligadas
ntmin=1,              !método de minimização
ncyc=2000,            !quantidade de ciclos antes da troca de algoritmo STEEPEDES DESCENT pelo CONJUGATE GRADIENT
maxcyc=10000,         !número máximo de ciclos de minimização
restraintmask='@CA,C,O,N', !restringe o movimento dos átomos da cadeia principal da proteína
restraint_wt=10.0,    !constante de força de restrição definida em 10Kcal/mol/Å²
```

Figura 8: Parâmetros aplicados à minimização seguidos de suas descrições.

Concluída a minimização, a temperatura do sistema foi elevada até 300 K a uma taxa de 0,48 Kps<sup>-1</sup>. Como em baixas temperaturas o cálculo da pressão não é preciso e o barostato pode levar a inconsistências, a etapa de elevação da temperatura do sistema foi feita no *ensemble* NVT (volume constante), com duração de 8 ns. Na etapa seguinte, o sistema passou por um processo para ajustar as energias (cinética e potencial) e a densidade. O surgimento de vácuo e espaços vazios entre o solvente e o sistema acaba reduzindo o valor da densidade, tornando necessária sua correção. Sendo assim, essa etapa foi feita no *ensemble* NPT (pressão constante), com duração de 15 ns. Os parâmetros detalhados para cada uma dessas etapas estão representados nas Fig. 9 e Fig. 10, respectivamente.

### Etapa de Aquecimento do Sistema (System Annealing Stage)

```
&cntrl
ntt=3,                !controle de temperatura com termostato de Langevin
ntx=1,                !leitura das coordenadas, mas não as velocidades do arquivo .rts7
gamma_ln=5.0,         !frequência de colisão de Langevin
ntc=2,                !restringe as oscilações de ligações de hidrogênio
ntf=2,                !impede o cálculo de forças das ligações restritas
ntp=0,                !desliga o controle de pressão
barostat=2,           !habilita o barostato de monte Carlo
pres0=1.0,            !valor alvo da pressão
temp0=10.0,           !temperatura inicial
temp0=300.00,         !temperatura de referência onde o sistema deve permanecer
taup=2.0,             !tempo de relaxação da pressão em ps
ntb=1,                !condições de contorno periódicas com volume constante
cut=12.0,             !distância de corte de interações não ligadas
dt=0.004,             !duração do passo de integração em ps
nstlim=2000000,      !quantidade de passos na simulação
ig=-1,                !fornece valores aleatórios para as velocidades iniciais
igb=0,                !não se aplica modelo de Born generalizado
ntwr=10000,           !a cada NTWR passos as informações são salvas no arquivo de reinicialização (.restrf)
ntwx=10000,           !salva informações sobre a simulação no arquivo de trajetória (.nc) a cada NTWX passos
nptr=10000,           !salva informações sobre a simulação no arquivo de saída (.out) a cada nptr passos
ioutfm=1,             !estabelece o formato dos arquivos gerados
nmropt=1,             !considera mudanças nas restrições e pesos de parâmetros
```

Figura 9: Parâmetros aplicados ao aquecimento seguidos de suas descrições.

## Etapa de Equilibração do Sistema (System Equilibration Stage)

```
&cntrl
ntt=3,                !controle de temperatura com termostato de Langevin
gamma_ln=1.0,        !frequência de colisão de Langevin
imin=0,              !não realizar minimização
ntc=2,               !restringe as oscilações de ligações de hidrogênio
ntf=2,               !impede o cálculo de forças das ligações restritas
ntb=2,               !habilita pressão constante
tempi=300.0,         !temperatura inicial
temp0=300.00,        !temperatura de referência onde o sistema deve permanecer
ntp=1,               !pressão constante (escala isotrópica de pressão)
barostat=2,          !habilita o barostato de Monte Carlo
pres0=1.0,           !pressão de referência
taup=1.0,            !tempo de relaxação da pressão em ps
cut=12.0,            !distância de corte de interações não ligadas
dt=0.004,            !duração do passo de integração em ps
nstlim=3750000,     !quantidade de passos na simulação
ig=-1,               !fornece valores aleatórios para as velocidades iniciais
igb=0,               !não se aplica modelo de Born generalizado
ntwr=18750,          !a cada NTWR passos as informações são salvas no arquivo de reinicialização (.restrt)
ntwx=18750,          !salva informações sobre a simulação no arquivo de trajetória (.nc) a cada NTWX passos
ntpr=18750,          !salva informações sobre a simulação no arquivo de saída (.out) a cada NPTR passos
ioutfm=1,            !estabelece o formato dos arquivos gerados
irest=1              !reinicia a simulação (para dar continuidade à etapa anterior)
nmropt=1,            !considera mudanças nas restrições e pesos de parâmetros
ntx=5                !habilita a leitura das coordenadas iniciais e das velocidades a partir de arquivos .rst
```

Figura 10: Parâmetros aplicados à equilibração seguidos de suas descrições.

Por fim, a etapa de produção (a simulação, onde o sistema é observado sob as condições preestabelecidas), foi realizada a volume e temperatura constantes, com o intuito de analisar aspectos conformacionais e de interações com ligante. Dessa forma foi utilizado o *ensemble* NVT. As produções foram realizadas em triplicata e preparadas para simular 300 ns. Seus parâmetros podem ser observados em detalhes na Fig. 11.

## Produção (Dinâmica Molecular) – Production (Molecular Dynamics)

```
&cntrl
ntt=3,           !controle de temperatura com termostato de Langevin
imin=0,         !não realizar minimização
gamma_ln=5.0,   !frequência de colisão de Langevin
ntc=2,          !restringe as oscilações de ligações de hidrogênio
ntf=2,          !impede o cálculo de forças das ligações restritas
ntb=1,          !habilita pressão constante
tempi=300.0,    !temperatura inicial
temp0=300.00,   !temperatura de referência onde o sistema deve permanecer
ntp=0,          !pressão constante (escala isotrópica de pressão)
barostat=2,     !habilita o barostato de Monte Carlo
pres0=1.0,      !pressão de referência
taup=1.0,       !tempo de relaxação da pressão em ps
cut=12.0,       !distância de corte de interações não ligadas
dt=0.004,       !duração do passo de integração em ps
nstlim=7500000, !quantidade de passos na simulação
ig=-1,          !fornece valores aleatórios para as velocidades iniciais
igb=0,          !não se aplica modelo de Born generalizado
ntwr=75000,     !a cada NTWR passos as informações são salvas no arquivo de reinicialização (.restrt)
ntwx=75000,     !salva informações sobre a simulação no arquivo de trajetória (.nc) a cada NTWX passos
nptr=75000,     !salva informações sobre a simulação no arquivo de saída (.out) a cada NPTR passos
ioutfm=1,       !estabelece o formato dos arquivos gerados
irest=1         !reinicia a simulação (para dar continuidade à etapa anterior)
ntx=5           !habilita a leitura das coordenadas iniciais e das velocidades a partir de arquivos .rst
saltcon=0.1     !concentração de sal para a CpHMD
ntxo=2,         !formata as coordenadas finais, velocidades e tamanho da caixa em NetCDF
icnstph=2,      !indica pH constante com solvente explícito
ntcnstph=100   !a cada NTCNSTPH passos há uma tentativa de troca nos estados de protonação
ntrelax=100,    !NTRELAX passos de relaxação do solvente a cada mudança bem sucedida de estado de protonação
solvph=9.0,     !estabelece o valor de pH que será utilizado
nmropt=1,       !considera mudanças nas restrições e pesos de parâmetros
```

Figura 11: Parâmetros aplicados à produção seguidos de suas descrições

## 4.6 Análises de dados

As análises pós-simulação foram feitas predominantemente com *cpptraj*, principal programa do AMBER para processar trajetórias e dados. Para avaliar os aspectos estruturais gerais da proteína, o quanto a estrutura muda ao longo do tempo em relação a uma estrutura de referência – no caso o primeiro *frame* da trajetória –, foram gerados dados sobre o valor médio para o desvio médio quadrático (*Root Mean Square Deviation* – RMSD), baseado na equação (16) a seguir:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (|r_i^{(sim)} - r_i^{ref}|)^2} \quad (16)$$

O Raio de Giro (*Radius of gyration* –  $R_g$ ) calcula o grau de dispersão dos átomos de uma estrutura molecular com respeito ou ao seu centro de massa ou a um determinado eixo de rotação. No caso do AMBER, adota-se o primeiro método a partir da equação (17):

$$R_g = \sqrt{\frac{\sum_i m_i (\vec{r}_i - \vec{r}_{cm})^2}{\sum_i m_i}} \quad (17)$$

para observar o grau de enovelamento da proteína, assim como possíveis passagens por diferentes estados conformacionais. Para o cálculo da flexibilidade individual dos resíduos, ou o quanto um resíduo em particular se movimenta (flutua) ao longo da simulação foi utilizado o RMSF (*Root Mean Square Fluctuation*), baseado na equação (18):

$$RMSF = \sqrt{\frac{1}{T} \sum_{i=1}^T |r_i^{(t)} - \langle r_i \rangle|^2} \quad (18)$$

Para contabilizar as ligações de hidrogênio formadas ao longo da dinâmica, utilizou-se o *hbonds* (complementado pelo *lifetime*), que determina as ligações com base em um critério geométrico: uma ligação de hidrogênio é formada quando a distância entre o átomo receptor e o átomo doador é menor do que 3 Å – distância de corte padrão. Os estados de protonação dos resíduos tituláveis foram obtidos pelo *cphstats*, que indica os estados de protonação em simulações com pH constante. Ele foi configurado para fornecer a estatística da população de cada resíduo (titulável) em cada estado de protonação sobre toda a trajetória. Os dados foram reunidos nas tabelas apresentadas na seção 5.2.6.

Para a obtenção da energia livre de ligação por resíduo na interação proteína-ligante dos sistemas analisados foi utilizada a técnica *Molecular Mechanics/Poisson-Boltzmann (Generalized Born) Surface Area* – MM/PB(GB)/SA. Esses dados foram obtidos com dois *scripts* em Python: *ante-MMPBSA* e *MMPBSA.py* (ferramentas do AmberTools). O primeiro *script* cria arquivos de topologia do complexo, receptor (proteína) e do ligante a partir do arquivo de topologia do sistema solvatado – trata-se de quatro arquivos para a etapa seguinte. Esses são os arquivos de entrada para o *MMPBSA.py*, que neste trabalho foi ajustado para realizar os cálculos com base no modelo Generalizado de Born (GB) – Apêndice C. A variação da distância entre o ligante e a Serina do sítio ativo (Ser132) da proteína (o parâmetro  $d_{HC}$ ) foi obtida com o comando *distance (cpptraj)*, que calcula a distância entre átomos (ou conjunto de átomos) – no caso o oxigênio da serina na posição 132 e o carbono do ligante, como definido da seção 2.3.

Para mapear todas as possíveis conformações do sistema e seus níveis de energia correspondentes, foram gerados mapas de energia livre (*Free Energy Landscape*, FEL) baseados em RMSD vs Raio de Giro. Esses dados foram obtidos através da conversão dos arquivos de topologia, coordenada e trajetória do formato de leitura do AMBER para o formato de leitura do GROMACS [71] com auxílio do *mdconverter* – ferramenta do MDTraj [72]. Em seguida, os arquivos convertidos foram então carregados no PyMOL e

o *plugin* Geo-Measures [73] calculou o RMSD e o Raio de Giro (no formato adequado) para posteriormente gerar os gráficos de FEL – os *scripts* para a conversão dos dados estão no Apêndice D.

## 5 Resultados

### 5.1 Modelos de ligante gerados

É importante observar que pela perspectiva biológica, a formação do complexo acontece pelo ancoramento da enzima (ligante) sobre o polímero (substrato). Por outro lado, nos trabalhos de simulação por DM, as funções de substrato e ligante são diferentes. Considerando o objetivo de estudar como a enzima age sobre o polímero, trata-se a enzima como o substrato onde o ligante (uma representação do polímero usualmente com até quatro monômeros) realiza o ancoramento. A partir disso, como mencionado na seção 4.3, diferentes modelos de ligante (BHET e PET) foram gerados com o auxílio do ATD. Esse processo foi realizado tendo como receptor cada um dos três modelos de proteína (nativa, variante I e variante II), totalizando 6 complexos proteína-ligante.

Tabela 1: Apresentação da classificação modelos de monômeros de BHET obtidos com o ATD. Os modelos estão dispostos conforme o valor obtido para o parâmetro  $S_{vd}$ .

Sistema	Modelo do Ligante	Vina score (Kcal/mol)	$d_{HC}$ (Å)	$S_{vd}$
WT	<sup>WT</sup> BHET <sub>0</sub>	-5,7	2,87	-2,83
	<sup>WT</sup> BHET <sub>1</sub>	-5,7	2,91	-2,79
	<sup>WT</sup> BHET <sub>2</sub>	-5,8	3,04	-2,76
	<sup>WT</sup> BHET <sub>3</sub>	-5,7	2,98	-2,72
I180V	<sup>I180V</sup> BHET <sub>0</sub>	-5,6	2,91	-2,69
	<sup>I180V</sup> BHET <sub>1</sub>	-5,6	2,95	-2,65
	<sup>I180V</sup> BHET <sub>2</sub>	-5,6	2,95	-2,65
	<sup>I180V</sup> BHET <sub>3</sub>	-5,7	3,06	-2,64
I180A	<sup>I180A</sup> BHET <sub>0</sub>	-5,5	2,85	-2,65
	<sup>I180A</sup> BHET <sub>1</sub>	-5,5	2,95	-2,55
	<sup>I180A</sup> BHET <sub>2</sub>	-5,5	2,97	-2,53
	<sup>I180A</sup> BHET <sub>3</sub>	-5,4	2,88	-2,52

Tabela 2: Apresentação da classificação modelos de monômeros de PET obtidos com o ATD. Os modelos estão dispostos conforme o valor obtido para o parâmetro  $s_{vd}$ .

Sistema	Modelo do Ligante	Vina score (Kcal/mol)	$d_{HC}$ (Å)	$s_{vd}$
<b>WT</b>	<sup>WT</sup> PET <sub>0</sub>	-5,0	2,91	-2,10
	<sup>WT</sup> PET <sub>1</sub>	-4,9	2,97	-1,93
	<sup>WT</sup> PET <sub>2</sub>	-4,9	2,98	-1,92
	<sup>WT</sup> PET <sub>3</sub>	-4,9	2,98	-1,92
<b>I180V</b>	<sup>I180V</sup> PET <sub>0</sub>	-5,0	2,89	-2,11
	<sup>I180V</sup> PET <sub>1</sub>	-5,0	2,90	-2,10
	<sup>I180V</sup> PET <sub>2</sub>	-5,0	2,90	-2,10
	<sup>I180V</sup> PET <sub>3</sub>	-5,0	2,90	-2,10
<b>I180A</b>	<sup>I180A</sup> PET <sub>0</sub>	-4,9	2,84	-2,06
	<sup>I180A</sup> PET <sub>1</sub>	-4,9	2,84	-2,06
	<sup>I180A</sup> PET <sub>2</sub>	-4,9	2,84	-2,06
	<sup>I180A</sup> PET <sub>3</sub>	-4,9	2,85	-2,05

A organização dos modelos conforme o parâmetro  $s_{vd}$  gerou as tabelas apresentadas adiante: Tab. 1 para os monômeros de BHET e Tab. 2 para os monômeros de PET. A organização apresentada nas tabelas – apresentando os quatro melhores modelos para cada complexo – permitiu a escolha dos modelos de ligantes com melhor valor de  $s_{vd}$  (valores mais negativos).

## 5.2 Análises dos sistemas complexados

Adiante são apresentados os resultados sobre a estrutura geral da proteína (RMSD, Rg e RMSF), seguidos dos resultados sobre a formação de ligações de hidrogênio e dos mapas de energia livre (FEL). Logo depois, os dados sobre a decomposição de energia por resíduo (MMGBSA) e distância ligante-proteína ( $d_{HC}$ ) são apresentados juntos – proteína nativa (WT), variante I (I180V) e variante II (I180A), nessa ordem.

## 5.2.1 Análise de RMSD

A Fig. 12 apresenta o comparativo dos resultados de RMSD das proteínas isoladas com seus respectivos complexos. Os resultados a respeito da proteína nativa e seus complexos com BHET (Fig. 12-A) indicam que mesmo após a formação do complexo não houve grandes variações nos valores de RMSD até próximo dos 200 ns de simulação, período em que permaneceram bastante próximos de 1 Å. Em seguida houve um aumento no RMSD – especialmente na WT-BHET-DM2 – que ao final apresentou uma tendência de convergência para um valor um pouco acima da média. De forma semelhante, os complexos com PET (Fig. 12-B) se mostraram regulares em torno de 1 Å (valor médio) até cerca de 150 ns – com exceção da WT-PET-DM1 que permaneceu acima da média (próximo de 1,25 Å) no intervalo 75-180 ns e posteriormente convergiu para próximo de 1 Å novamente. Após os 150 ns as WT-PET DM2 e DM3 apresentam oscilações próximas de 1,25 Å.

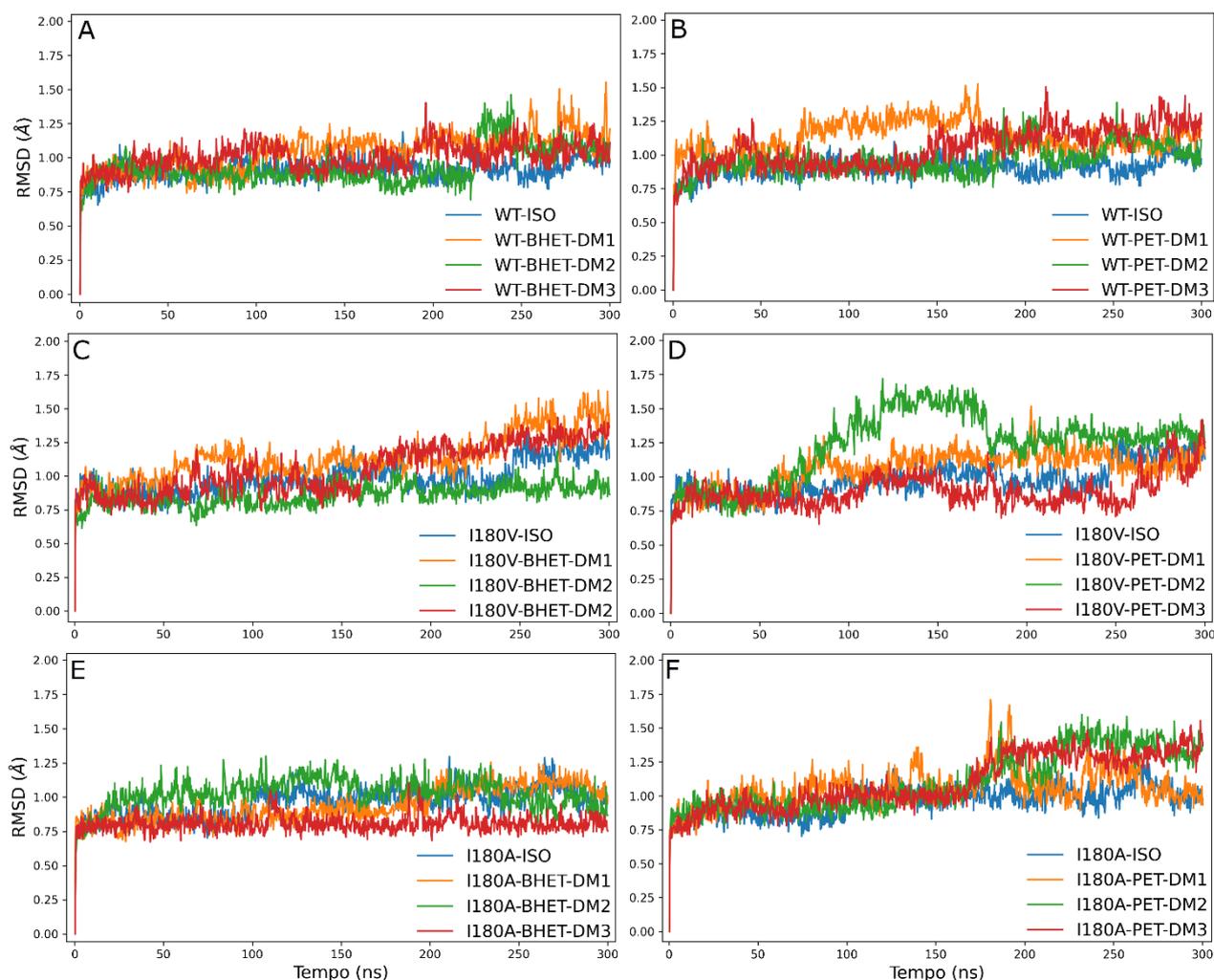


Figura 12: Comparativo dos resultados de RMSD entre a proteína nativa e seus respectivos complexos: (A) e (B) WT; (C) e (D) I180V; e (E) e (F) I180A. A coluna esquerda representa as triplicatas dos complexos com o BHET e a coluna direita representa as triplicatas dos complexos com o PET.

Os resultados a respeito da variante I isolada e de seus complexos com o BHET na Fig. 12-C demonstram que na proteína isolada houve um aumento no valor de RMSD de 1 Å para 1,25 Å nos últimos 50 ns da simulação – uma provável mudança de estado conformacional. As I180V-BHET DM1 e DM3 apresentaram a partir de 150 ns um aumento no valor de RMSD – de 1 Å para próximo de 1,5 Å, enquanto que a DM2 permaneceu bastante estável em torno da média durante toda a simulação. Nos complexos da variante I com o PET, a Fig. 12-D revelou que nos primeiros 80 ns de simulação todos os complexos estavam alinhados com a proteína isolada, momento a partir do qual diante o RMSD dos complexos se alterou. A DM2 passou de 0,9 Å para próximo de 1,75 Å até os 180 ns e logo em seguida estabilizou em torno de 1,3 Å. Enquanto isso, a DM3 apresenta variações predominantemente abaixo da média (em torno de 0,9 Å) que alcançam valores de 1,25 Å a partir de aproximadamente 260 ns e a DM1 permanece um pouco acima do valor médio de RMSD de 80 ns até o final da simulação.

Os gráficos sobre os resultados da variante II isolada e seus complexos com o BHET estão na Fig. 12-E. Observando a proteína isolada, nota-se que a partir dos 100 ns seu RMSD passa de 0,9 Å para cerca de 1,1 Å e permanece dessa forma até o final da simulação. De maneira semelhante, os complexos I180A-BHET DM1 e DM3 começaram a divergir entre si a partir dos 100 ns, com a DM1 aumentando seu RMSD gradativamente até 200 ns (onde acontece um salto de 1 Å para 1,2 Å) e a DM3 permanecendo em torno de 0,9 Å até o final da simulação. A DM2 apresentou RMSD acima da média desde o início da simulação (por volta de 30 ns) e permaneceu em torno de 1,2 Å durante toda a simulação. A Fig. 12-F demonstra os resultados da variante II isolada em relação aos seus complexos com o PET. Os gráficos dos complexos apresentam uma tendência de aumento gradativa até aproximadamente 150 ns. Desse ponto em diante as DM2 e DM3 tendem para o valor médio de 1,3 Å, enquanto que a DM1 – apesar das oscilações no seu valor de RMSD – permanece mais alinhada com a proteína isolada, mais próxima de 1 Å.

### 5.2.2 Análise do raio de giro

A Fig. 13 (A-F) apresenta os resultados de variação do Raio de Giro ( $R_g$ ) de todos os sistemas estudados: proteína nativa, variante I, variante II e seus respectivos complexos com os monômeros de BHET e PET. O valor médio geral de  $R_g$ , entre todas as simulações, foi de 16,4 Å e como um todo os valores permaneceram predominantemente em torno desse valor médio. Contudo, de forma mais detalha, houve diferenças mais evidentes como apontado adiante. Nos complexos da proteína nativa com o monômero de BHET (Fig. 13-A) cabe o destaque para a DM2 que apresentou um salto em  $R_g$  após os 200 ns, concordando com os resultados de RMSD apresentados anteriormente. No caso dos complexos com PET (Fig. 13-B) as DM1 e DM3 apresentaram, respectivamente, uma redução e aumento nos valores de  $R_g$  após 150 ns – como apontado nos resultados de RMSD.

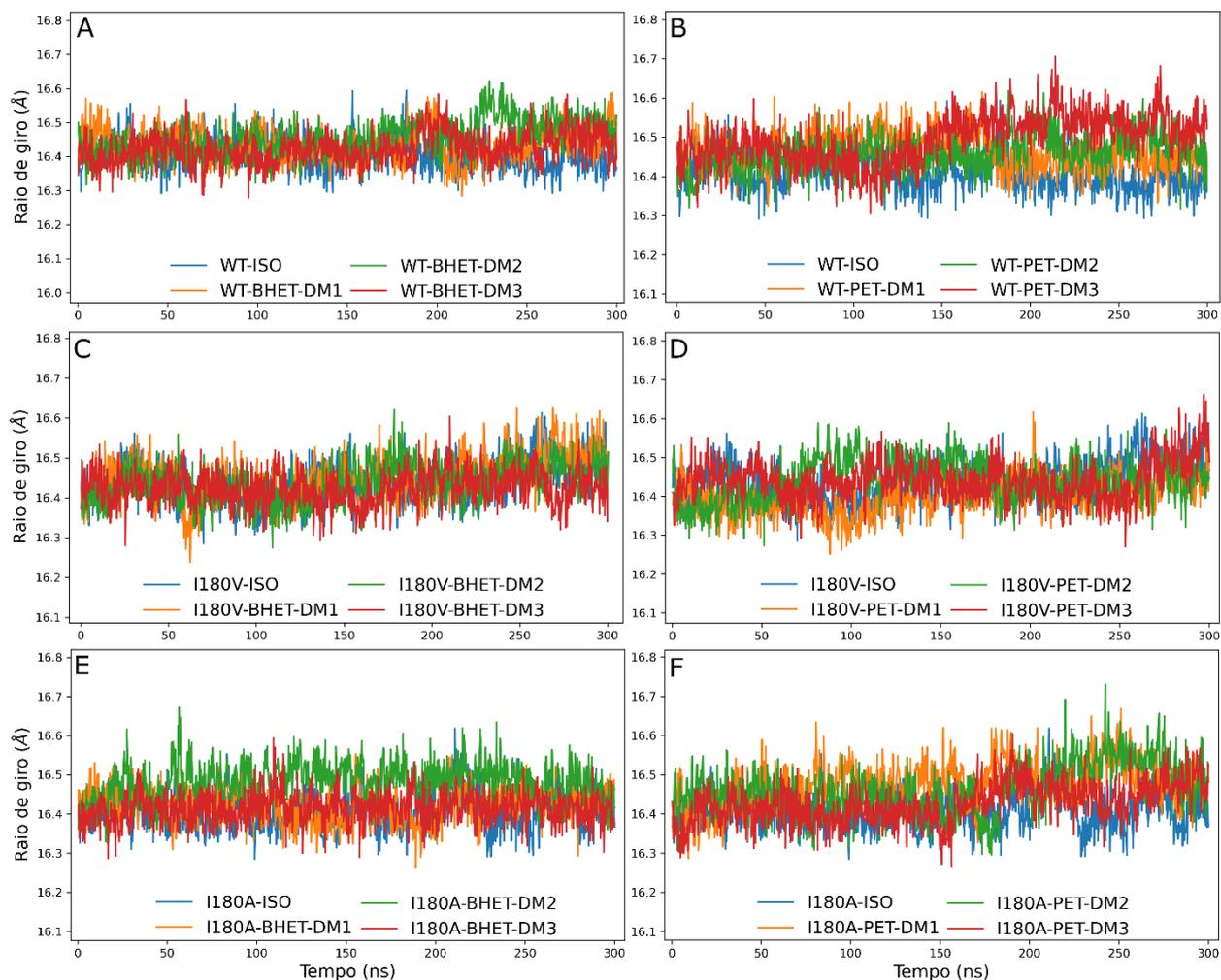


Figura 13: Comparativo dos resultados de  $R_g$  entre a proteína nativa e seus respectivos complexos: (A) e (B) WT; (C) e (D) I180V; e (E) e (F) I180A. A coluna esquerda representa as triplicatas dos complexos com o BHET e a coluna direita representa as triplicatas dos complexos com o PET.

Os resultados trazidos na Fig. 13-C e 13-D relacionados à variante I e seus complexos, sugerem que a variante I apresentou uma tendência de aumento nos seus valores de  $R_g$  em relação à proteína nativa isolada, já que a variante I após 200 ns de simulação teve um aumento no seu raio de giro – redução no grau de compactação. Observando os complexos com BHET (Fig. 13-C), de modo geral eles se mostraram bastante regulares e alinhados com a proteína isolada, exibindo apenas uma oscilação rápida na DM1 entre 50-100 ns. Enquanto isso, os complexos com PET (Fig. 13-D) apresentaram oscilações mais evidentes, diferindo de forma mais clara da proteína isolada. A I180V-PET-DM2 tem um aumento no valor de  $R_g$  entre 50-100 ns e sofre uma redução para próximo do valor médio por volta dos 200 ns e se mantém em torno desse valor até o final da simulação – no caso da I180V-PET-DM3 aconteceu um aumento no raio de giro no final da simulação, após 250 ns.

Observando os resultados em relação à variante II isolada e seus respectivos complexos (Fig. 13-E e 13-F), os dados indicam que a proteína isolada já apresentava

diversas oscilações ao longo da simulação – possível indicativo de conformações transitórias como resultado da mutação implementada. Essa proteína apresentou uma grande oscilação de  $R_g$  entre 200-230 ns. Especificamente nos complexos com BHET (Fig. 13-E) DM1 da variante II esteve bastante alinhada à proteína isolada, até mesmo apresentando uma oscilação no raio de giro entre 200-230 ns, ao passo que a DM2 se apresentou  $R_g$  acima da média durante toda a simulação – sem grandes mudanças. Para os complexos com PET (Fig. 13-F), suas DM1 e DM2 se mostraram consideravelmente instáveis, com diversas oscilações durante toda a simulação – de forma semelhante ao que foi observado nos respectivos resultados de RMSD –, enquanto que sua DM3 apresentou variações mais claras no raio giro apenas após os 150 ns.

### 5.2.3 Análise da flutuação residual

A Fig. 14 representa os gráficos da flutuação residual ao longo das simulações realizadas. De maneira geral as flutuações permaneceram predominantemente abaixo de 1,5 Å, à exceção das extremidades da proteína (que naturalmente apresentam flutuações mais altas) e de três regiões específicas apresentadas e discutidas a seguir.

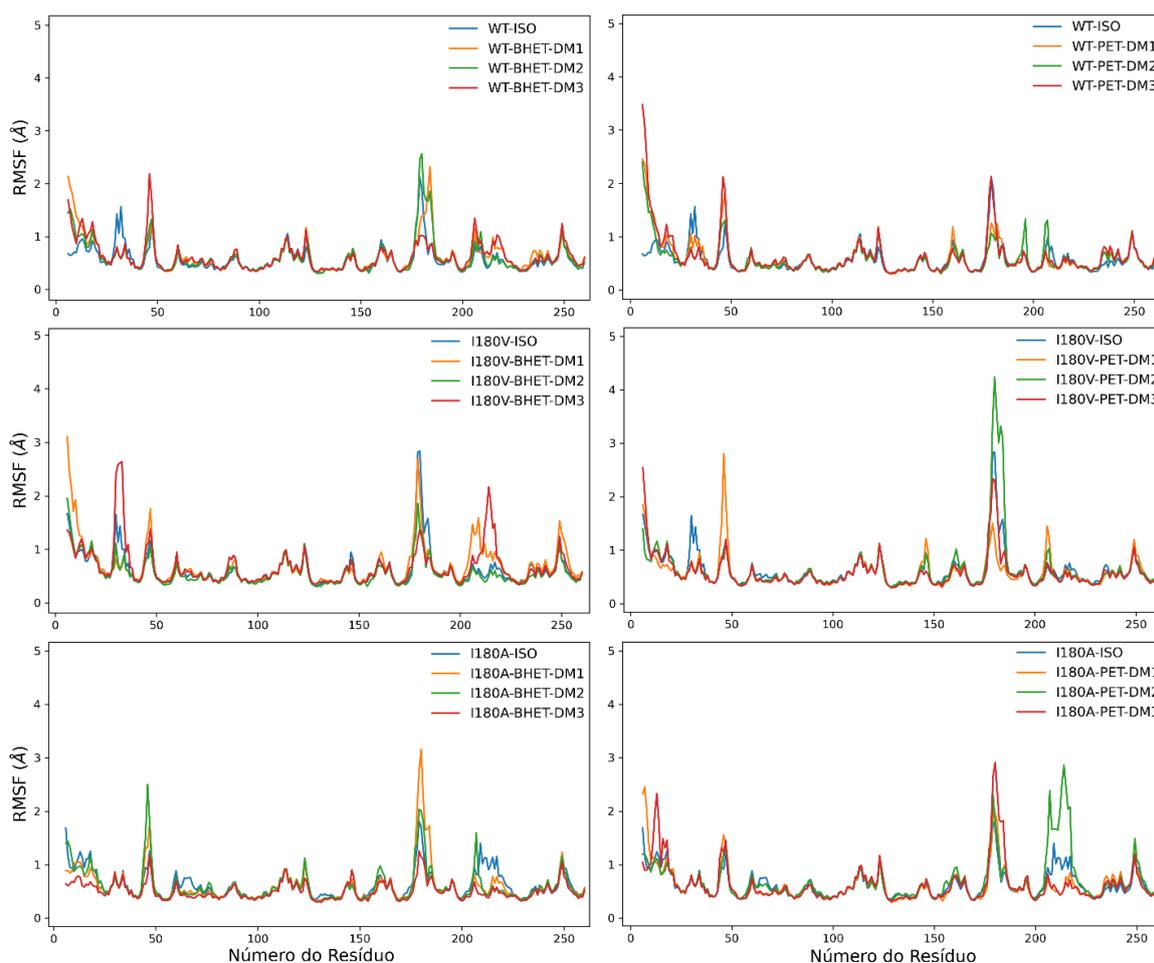


Figura 14: Comparativo dos resultados de RMSF entre a proteína nativa e seus respectivos complexos: (A) e (B) WT; (C) e (D) I180V; e (E) e (F) I180A. A coluna esquerda representa as triplicatas dos complexos com o BHET e a coluna direita representa as triplicatas dos complexos com o PET.

Essas regiões correspondem aos seguintes intervalos de resíduos: (I) 25-60; (II) 170-190, centrado no sítio 180 e (III) 205-220, que abriga a His209 (integrante da tríade catalítica). As flutuações da região (I) se mantiveram entre 1,5-3 Å, sendo esta região correspondente a uma estrutura de laço que une as duas primeiras fitas- $\beta$  na estrutura da proteína (Fig. 15).

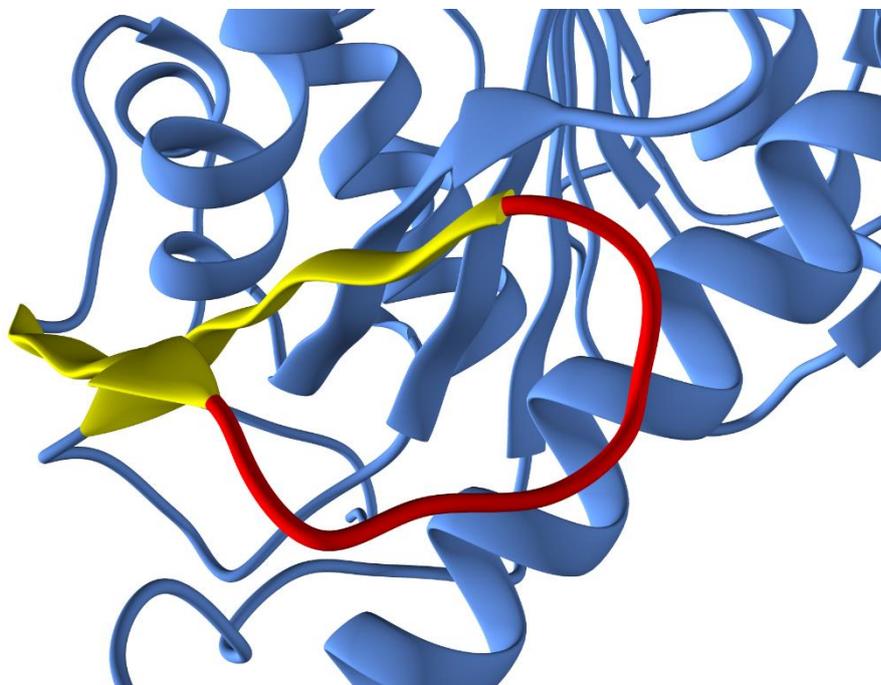


Figura 15: Estrutura de laço da proteína correspondente ao intervalo de resíduos 25-60 – região (I).

A região (II), centrada no sítio 180 (sítio da mutação) apresentou variações maiores, que vão desde 1 Å até valores acima de 4 Å. A começar pelas proteínas isoladas, a variante I (I180V) exibiu maior flutuação nessa região (próxima de 3 Å), ao passo que a proteína nativa e a variante II (I180A) permaneceram ambas próximas de 2 Å. Observando especificamente o comparativo entre a proteína nativa isolada e seus complexos com BHET (Fig. 14-A), notou-se na sua DM3 uma flutuação bastante baixa em relação às outras (alcançando uma máxima de 1 Å), enquanto que as DM1 e DM2 se mantiveram mais alinhadas com a proteína isolada (em torno de 2 Å). Nos complexos da proteína nativa com o com PET (Fig. 14-B) o cenário se inverteu com a DM3, mais alinhada à proteína isolada – aproximadamente 2 Å –, enquanto as DM1 e DM2 apresentaram flutuações mais baixas (abaixo de 1,5 Å). Na Fig. 14-C, os resultados da variante I e de seus complexos com BHET indicaram que apenas a DM1 se manteve próxima à variante I isolada (próximo de 3 Å), ao passo que as DM2 e DM3 exibiram flutuações abaixo de 2 Å (menos flexíveis). Nos complexos da variante I com o PET (Fig. 14-D), foi onde as maiores flutuações foram registradas, com a DM2 ultrapassando a marca de 4 Å, enquanto as DM1 e DM3 (mais próxima da variante I isolada) se mantiveram abaixo da linha de 3 Å. Para a variante II e seus complexos com BHET (Fig. 14-E), sua DM1 exibiu a maior flutuação (acima de 3 Å), ao passo que a DM3 apresentou flutuações menores (pouco acima de 1 Å) e a DM2 se mostrou mais alinhada com a

variante II isolada. Os complexos da variante II com o PET (Fig. 14-F) apresentaram flutuações mais altas do que a variante II isolada, com sua DM3 alcançando 3 Å e as DM1 e DM2 mais próximas de 2,5 Å.

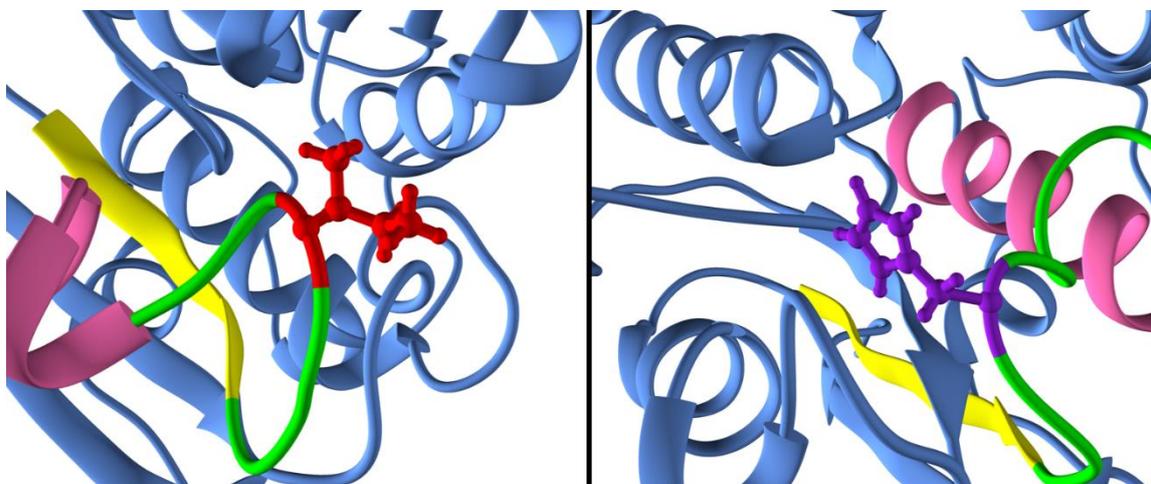


Figura 16: As flutuações residuais mais evidentes se mostram centradas no resíduo 180 (à esquerda, na cor vermelha, a Ile da proteína nativa) e nas proximidades da His209 (à direita, na cor roxa).

Em relação à região (III), que compreende os resíduos 205-220, a variante II apresentou flutuação mais alta em relação à proteína nativa (e também à variante I). De forma geral, agora levando em conta também seus complexos, as flutuações apresentaram variações que vão, predominantemente, de 0,5 até ligeiramente acima dos 2 Å, com exceção dos complexos da variante II com o PET (Fig. 14-F), onde sua DM2 alcançou os 3 Å, ao passo que as DM1 e DM3 exibiram flutuações bem mais baixas (próximas de 1 Å); assim como foi o caso da variante II isolada. Nos demais sistemas, a começar pela proteína nativa e seus complexos com o BHET (Fig. 14-A), houve um pequeno aumento na flutuação dessa região para os complexos – apesar de não excederem 1,5 Å. Para os complexos com PET (Fig. 14-B) as DM1 e DM3 apresentaram uma redução na flutuação, ficando abaixo de 1 Å – menos flexíveis que a DM1 e a proteína isolada. Passando para o comparativo da variante I isolada e seus complexos com BHET (Fig. 14-C), os resultados apontaram um aumento na flexibilidade das DM1 e DM3 (acima de 1,5 Å e 2 Å, respectivamente) em relação a DM2 e à variante II isolada (que se mantiveram abaixo de 1 Å). Na Fig. 14-D, os complexos da variante I com PET apresentaram flutuações mais baixas em relação aos seus respectivos complexos com BHET, com valores abaixo de 1,5 Å. Os gráficos relacionados às simulações da variante II e complexada com BHET (Fig. 14-E) apontaram que as três simulações do complexo tiveram flutuações mais baixas nas proximidades da His209 em relação à variante II isolada, com valores abaixo de 1 Å.

#### 5.2.4 Análise da formação das ligações de hidrogênio

A série temporal de ligações de hidrogênio (HB) ao longo da simulação foi calculada para observar esse tipo de interação entre os resíduos da proteína e o ligante. Alguns pares resíduo-ligantes se repetem, pois podem atuar alternadamente como doadores ou receptores das ligações de hidrogênio. No que diz respeito aos gráficos, suas

duas cores representam: presença de HB (amarelo) e ausência de HB (preto). Além disso, a organização da apresentação dos pares no eixo y deixa o receptor sempre à esquerda.

Começando com os resultados relacionados aos complexos da proteína nativa com o monômero de BHET, a Fig. 17 (topo) apresenta as ligações de hidrogênio da WT-BHET-DM1. Nesse complexo houve HB quase que exclusivamente pelo Asp178. Esse resíduo teve papel de receptor por mais de 90% da simulação (pelo seu átomo de O n° 2) e de doador por 45% da simulação (pelo H de seu O n° 1). Os dados de *lifetime* apontaram que essas ligações se formaram 24 e 452 vezes ao longo da simulação. Por outro lado, as demais HB contabilizadas têm permanência muito baixa e não contribuíram com a estabilidade da interação proteína-ligante (permanecem menos de 1% da simulação).

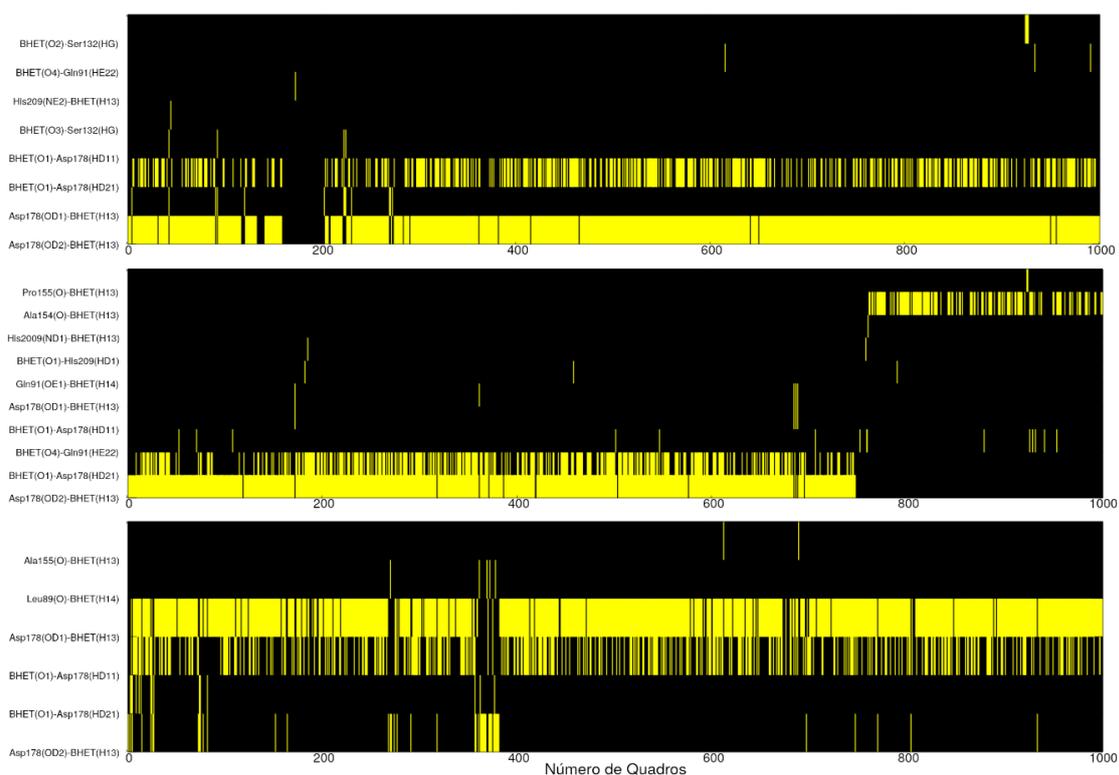


Figura 17: Contabilização de ligações de hidrogênio (HB) ao longo da simulação (por frames) para os complexos da proteína nativa com o monômero de BHET. À esquerda os pares resíduo-ligante que formaram HB e entre parêntese os átomos que participam da ligação. De cima para baixo: WT-BHET-DM1, WT-BHET-DM2 e WT-BHET-DM3.

A Fig. 17 (centro) apresenta a série temporal das ligações de hidrogênio contabilizadas durante a simulação do complexo WT-BHET-DM2. Assim como na DM1, predominaram HB entre o ligante e o Asp178. Esse resíduo interagiu com o ligante por meio de HB mais de 70% como receptor (átomo de O n° 2) e aproximadamente 37% do tempo como doador (H de seu O n° 1), sendo as duas ligações com maior permanência na simulação. Além disso, vale notar que elas se formaram, respectivamente, 14 vezes e 160 vezes. Houve ainda certa interação do ligante com a Ala155, mas assim como outras ligações de hidrogênio, apresentaram permanência abaixo dos 15% da simulação e não contribuíram para a estabilidade do complexo.

Na Fig. 17 (inferior) é indicada a série temporal das ligações de hidrogênio a respeito da WT-BHET-DM3. Novamente as HB com o Asp178 apresentaram permanência alta durante a simulação, sendo esse resíduo receptor (átomo de O n° 2) por quase 90% da simulação em uma ligação que se formou aproximadamente 64 vezes; o Asp178 também interagiu como doador (H de seu O n° 1) durante quase 40% da simulação em uma ligação que se formou aproximadamente 230 vezes. As poucas ligações de hidrogênio apontadas no gráfico tiveram permanência bastante baixa (abaixo de 5%) e não contribuíram para a estabilidade do complexo.

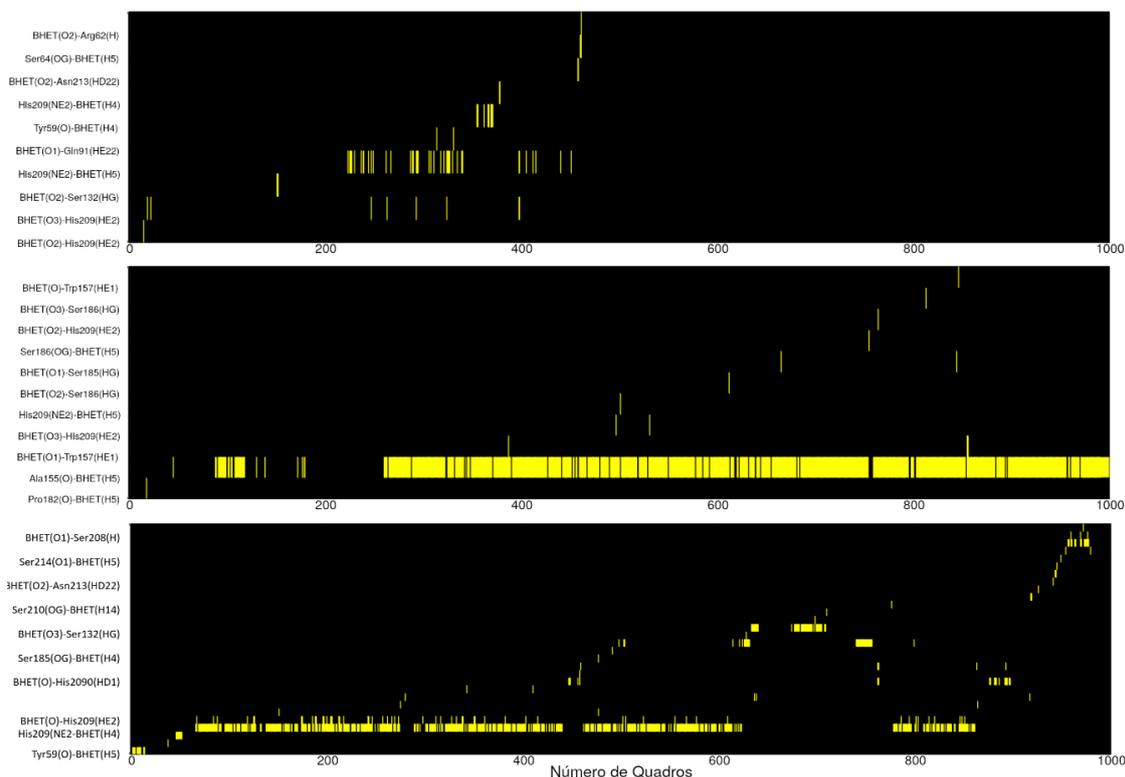


Figura 18: Contabilização de ligações de hidrogênio (HB) ao longo da simulação (por frames) para os complexos da proteína nativa com o monômero de PET. À esquerda os pares resíduo-ligante que formaram HB e entre parêntese os átomos que participam da ligação. De cima para baixo: WT-PET-DM1, WT-PET-DM2 e WT-PET-DM3.

A ausência quase completa de ligações de hidrogênio na simulação do complexo WT-PET-DM1, apresentado na Fig. 18 (topo) é indicativo do desprendimento do ligante. A Fig. 18 (centro) demonstra a predominância de HB com Ala155 (como receptora com seu O) de maneira quase que exclusiva na WT-PET-DM2. Essa interação permaneceu por mais de 70% da simulação e se formou aproximadamente 54 vezes. Como já sugerido anteriormente, as demais ligações de hidrogênio não contribuem para a estabilidade do complexo. Na simulação do complexo WT-PET-DM3 (Fig. 18, inferior), foram contabilizadas ligações de hidrogênio com o ligante por diversos resíduos. No entanto, a maioria dessas interações teve permanência inferior a 5% da simulação, não contribuindo para a estabilidade do complexo. Por outro lado, a His209 foi o resíduo que formou ligações de hidrogênio por mais tempo – por mais de 45% da simulação, uma ligação que se formou aproximadamente 109 vezes.

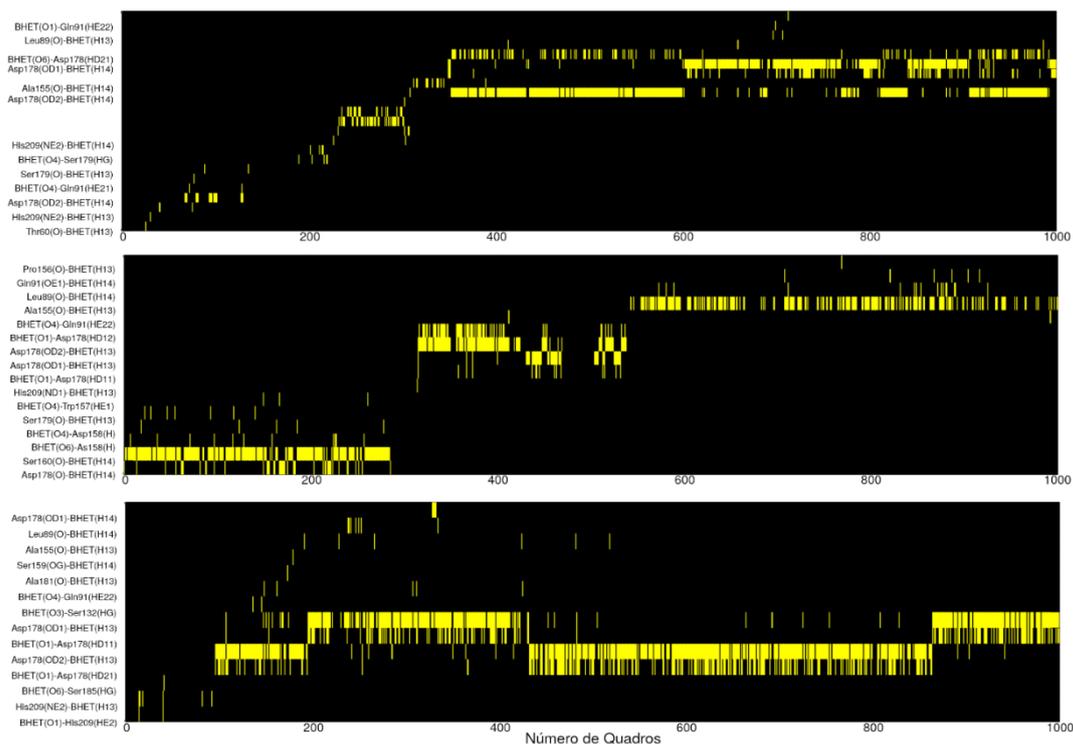


Figura 19: Contabilização de ligações de hidrogênio (HB) ao longo da simulação (por frames) para os complexos da variante I com o monômero de BHET. À esquerda os pares resíduo-ligante que formaram HB e entre parêntese os átomos que participam da ligação. De cima para baixo: I180V-BHET-DM1, I180V-BHET-DM2 e I180V-BHET-DM3.

Para melhor apresentar os resultados relacionados à I180V-BHET-DM1 (Fig. 19, topo), seu eixo Y foi alterado de forma a apresentar as ligações de hidrogênio mais relevantes formadas juntas com alguns exemplos das muitas ligações de baixa permanência contabilizadas. Sendo assim, de forma semelhante àquilo que foi observado nas simulações com a proteína nativa complexada com o BHET, a Fig. 19 (topo) indica a predominância de ligações de hidrogênio com o Asp178 por volta de 120 ns (400 frames) até o final da simulação. A interação com o Asp passa de seu oxigênio n° 2 para o seu oxigênio n° 1 no intervalo mencionado – como receptor. Essas interações têm, respectivamente, permanência por volta de 37% e 26% da simulação no referido intervalo de tempo, se formando aproximadamente 45 e 40 vezes. O Asp178 interage com o ligante em outros momentos como doador, mas assim com as demais ligações de hidrogênio registradas não contribuem o suficiente para a estabilidade do complexo.

A formação de HB representada na Fig. 19 (centro), relativas à I180V-BHET-DM2 não ficou concentrada apenas em um resíduo por simulação como demonstrado anteriormente nas simulações com a proteína nativa. No complexo em questão as interações foram distribuídas principalmente entre os resíduos Ser160 e Ala155 – apesar de a permanência dessas interações ser consideravelmente menor do que aquela observada nos casos com a proteína nativa. As interações com a Ser160 existiram durante 22% da simulação, se formando aproximadamente 33 vezes e com a Ala155 as interações perduraram por 21% da simulação, se formando cerca de 86 vezes. O Asp178, bastante

evidente nas simulações anteriores, junto com as demais ligações de hidrogênio contabilizadas, não apresentaram contribuição para a estabilidade da interação proteína-ligante, possuindo permanência por períodos muito curtos (abaixo de 11%).

A Fig. 19 (inferior) retrata as ligações de hidrogênio referentes à DM3 da variante I com o BHET. Novamente se destacam interações do ligante com o Asp178. Durante aproximadamente 90% da simulação o ligante apresenta ligação de hidrogênio com o Asp, onde predominam interações com os oxigênios do Asp – ambos como receptores. O oxigênio n° 2 interage durante 46% da simulação com uma ligação que se formou cerca de 76 vezes, enquanto que o oxigênio n° 1 interage durante 32% da simulação com uma ligação que se formou por volta de 63 vezes. Além do Asp178 interagir como doador durante 21% da simulação, as demais ligações de hidrogênio não se mantiveram por tempo suficiente para contribuir com a estabilidade do complexo.

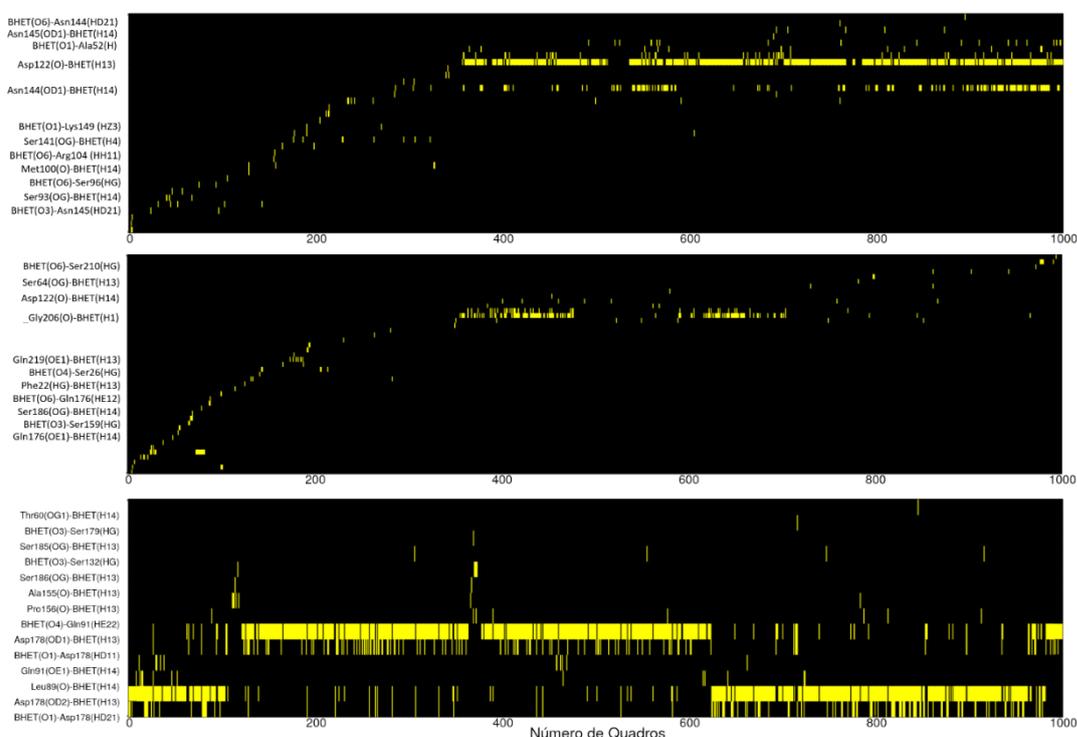


Figura 20: Contabilização de ligações de hidrogênio (HB) ao longo da simulação (por frames) para os complexos da variante II com o monômero de BHET. À esquerda os pares resíduo-ligante que formaram HB e entre parêntese os átomos que participam da ligação. De cima para baixo: I180A-BHET-DM1, I180A-BHET-DM2 e I180A-BHET-DM3.

Assim como na Fig. 19 (topo), o eixo Y das Fig. 20 (topo e centro) foi alterado de forma a exibir as ligações de hidrogênio mais relevantes junto com alguns exemplos das demais. Assim, nas simulações com a variante II complexada com o BHET, a Fig. 20 (topo) junto com os dados gerados pelo *lifetime* indicam que na DM1 as ligações de hidrogênio foram predominantemente devido ao Asp122. Esse resíduo interagiu com o ligante por mais de 50% da simulação, formando ligações de hidrogênio por cerca de 72 vezes. O Asn144, formando ligações por menos de 15% das simulações, junto com as

demais interações contabilizadas (por períodos ainda mais curtos) não contribuem com a estabilização do complexo.

Em se tratando da DM2 do complexo da variante II com o BHET, indicado pela Fig. 20 (centro) os dados de *lifetime* indicam que nenhuma das ligações de hidrogênio perduraram por mais de 12% da simulação, sendo a Gly206 o resíduo que permaneceu mais tempo formando ligações com o ligante. A Fig. 20 (inferior) traz a série temporal das ligações de hidrogênio para a DM3, onde voltam a predominar ligações com o Asp178 atuando como receptor com seus dois átomos de oxigênio. Seus oxigênios interagem com o BHET cerca de 50% e 43% da simulação com ligações que se formaram, respectivamente, cerca de 70 e 81 vezes. As interações em que seus oxigênios atuam como doadores não excedem os 12% da simulação e, assim como as demais ligações contabilizadas não têm grandes contribuições para a estabilização do complexo.

As ligações de hidrogênio contabilizadas pelo *lifetime* para as simulações da variante I com o PET foram extremamente baixas, dado que as interações mais duradouras não se mantiveram por mais de 5% da simulação – indicativo do desprendimento do ligante do sítio ativo da proteína. Para os complexos da variante II com o PET o *lifetime* não apontou ligações de hidrogênio ao longo da simulação devido ao desprendimento do ligante desde o início da simulação – como indicado nos gráficos de  $d_{HC}$  e na decomposição de energia residual ao longo do tempo. Por esse motivo não foram gerados gráficos da série temporal de HB para esses sistemas.

### 5.2.5 Análise dos mapas de energia livre

Para medir a variação conformacional entre as formas isoladas e complexadas das proteínas [74] e determinar regiões de mínimos globais de energia [75], foram gerados os gráficos de Mapa de Energia Livre (*Free Energy Landscape*, FEL) baseados em RMSD e  $R_g$ .

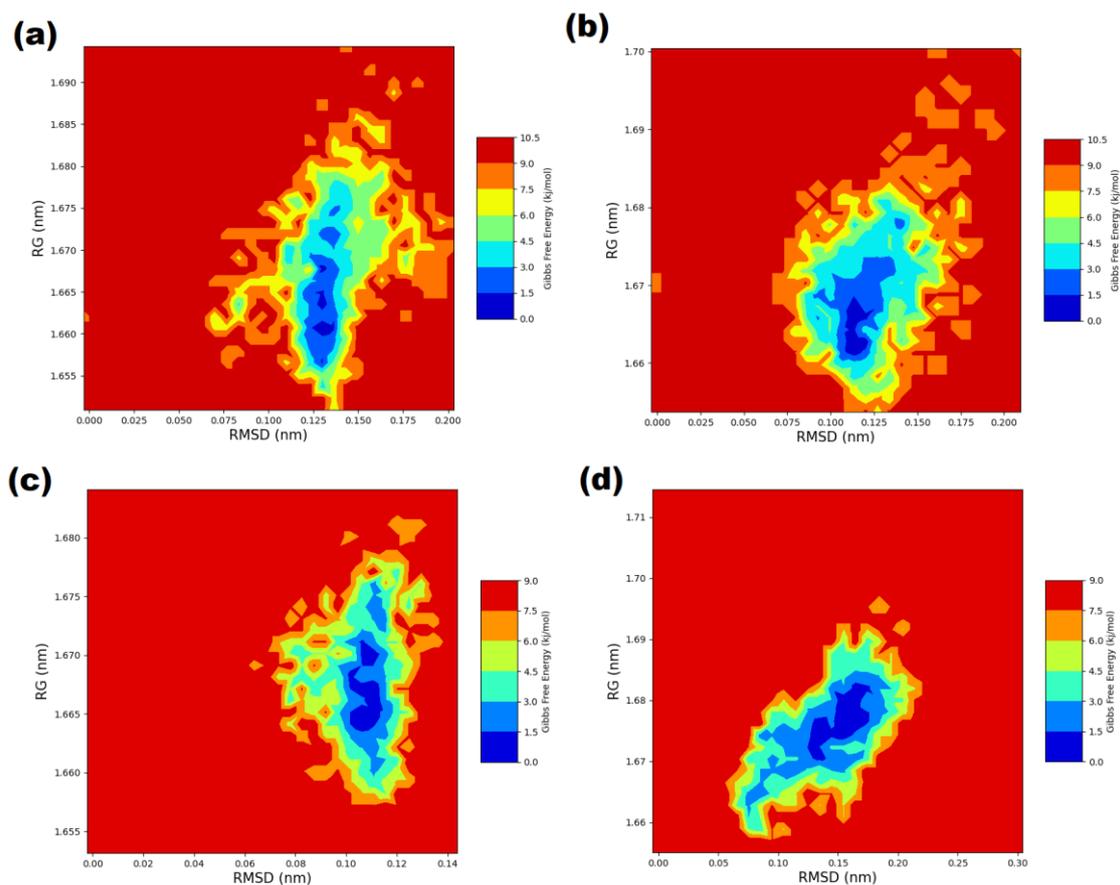


Figura 21: Análise do Mapa de Energia Livre baseado em RMSD– $R_g$  dos sistemas com proteína nativa (a) isolada; (b), (c) e (d) complexada com monômero de BHET.

A Fig. 21 traz o comparativo entre a proteína nativa isolada e seus complexos com o monômero de BHET (WT-BHET). A proteína isolada apresentou quatro poços de energia mínima (na cor azul escuro no gráfico), todas bastante centralizadas e compactos na região de aproximadamente 0,125 nm no RMSD e entre 1,65-1,67 nm no  $R_g$ , como indica a Fig. 21-a. Os gráficos de FEL dos complexos revelaram uma redução nos estados conformacionais acessíveis à proteína complexada: na DM2, há dois poços de energia mínima próximos (Fig. 21-c), enquanto nas DM1 e DM3 (Fig. 21-b e 21-d) foi identificado apenas um poço de energia mínima global. Os resultados também apontaram que o valor máximo da Energia Livre de Gibbs nos quatro sistemas em questão permaneceram entre 9 e 10,5 kJ/mol. De forma semelhante, todas as proteínas demonstraram conformações estáveis em níveis abaixo de 1,5 kJ/mol (indicado pela cor azul escuro no gráfico).

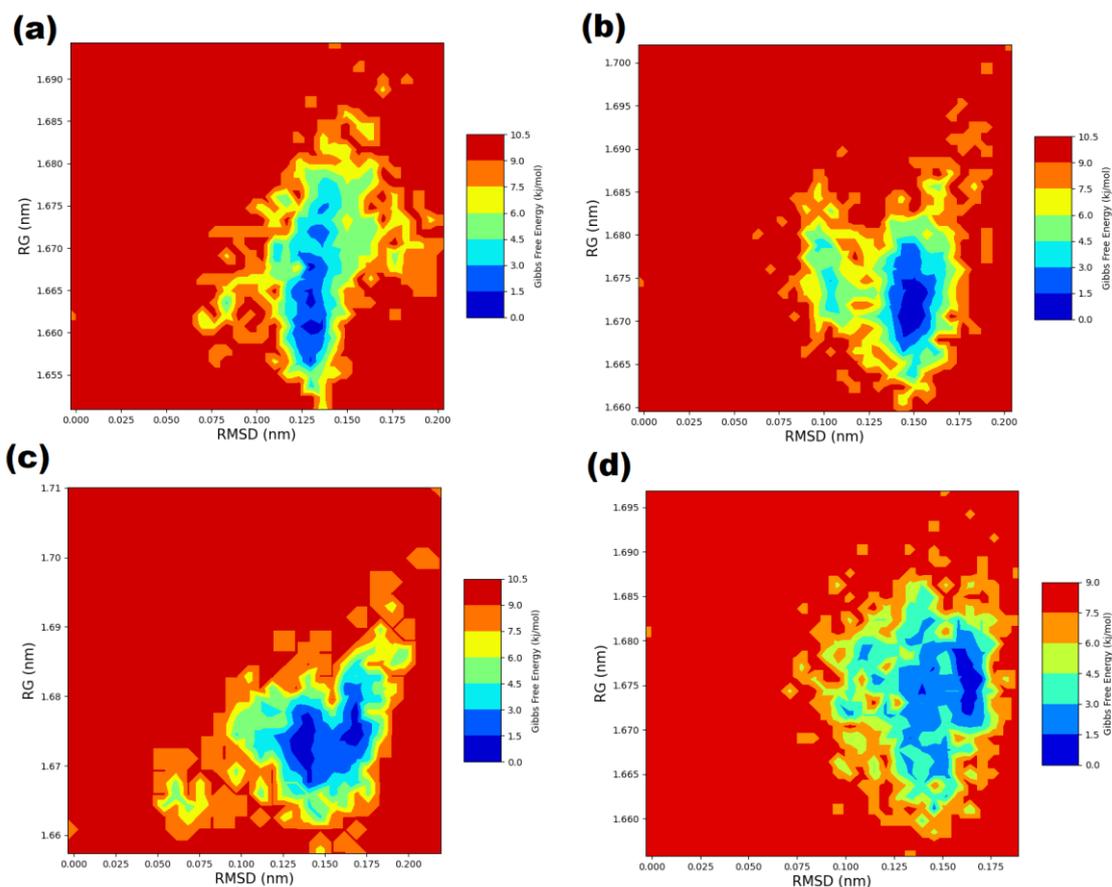


Figura 22: Análise do Mapa de Energia Livre baseado em RMSD– $R_g$  dos sistemas com proteína nativa (a) isolada; (b), (c) e (d) complexada com monômero de PET.

A Fig. 22 traz o comparativo da proteína nativa isolada com seus complexos com o monômero de PET. A situação indicada pelos gráficos de FEL dos complexos foi bastante semelhante àquela observada nos complexos com BHET, em que houve a diminuição dos estados conformacionais acessíveis para a proteína: dos quatro poços de energia mínima que a nativa isolada dispunha (Fig. 22-a), aos complexos permaneceram um poço para DM1 (Fig. 22-b) e 2 poços de energia mínima na DM2 (mais evidentemente separadas como indica a Fig. 22-c) e DM3, apontado na Fig. 22-d, e. Vale destacar que o segundo poço de energia mínima na DM3 muito provavelmente corresponde a uma conformação transitória, mais difícil de ocorrer devido à sua área reduzida – centrada por volta de 0,14 nm (RMSD) e 1,67 ( $R_g$ ). Novamente o valor máximo da Energia Livre de Gibbs permaneceu entre 9 e 10,5 kJ/mol, assim como as conformações mais estáveis apresentaram níveis de energia abaixo de 1,5 kJ/mol

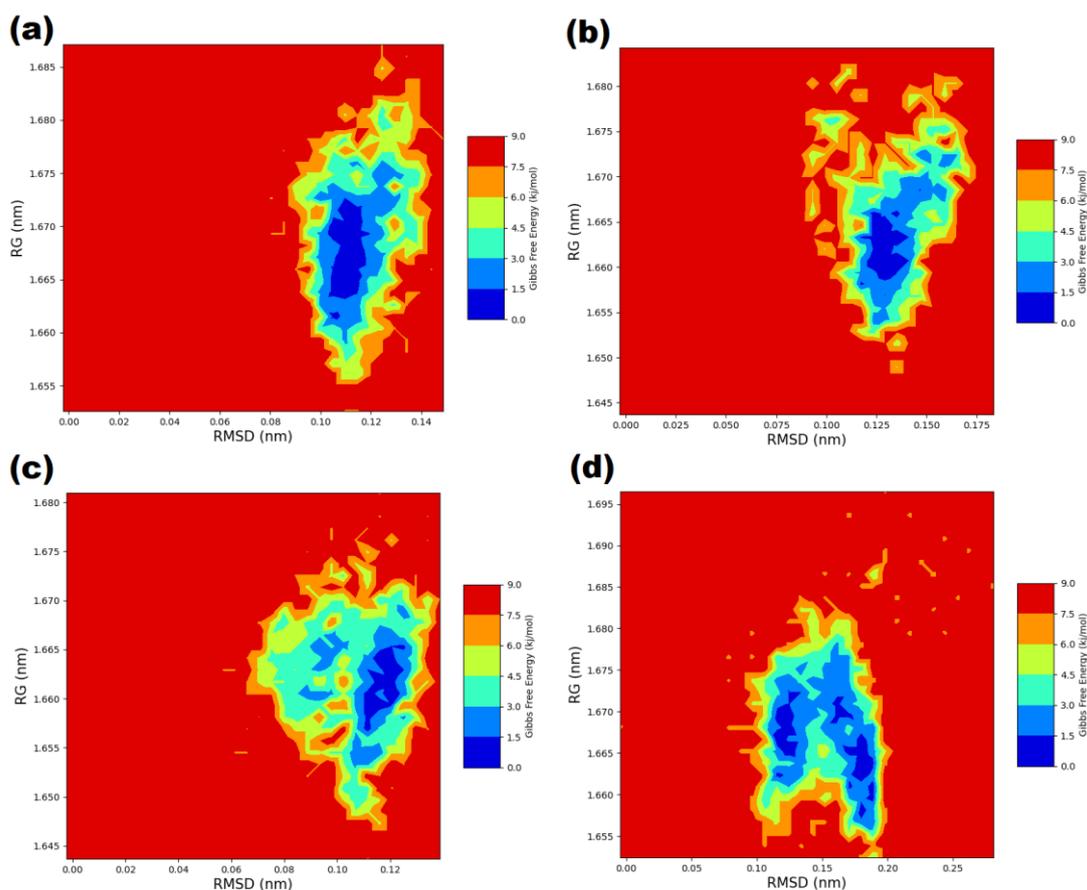


Figura 23: Análise do Mapa de Energia Livre baseado em RMSD–R<sub>g</sub> dos sistemas com a variante I (a) isolada; (b), (c) e (d) complexada com monômero de BHET.

Na Fig. 23 estão representados os gráficos de FEL da variante I isolada e complexada com o monômero de BHET. A variante I isolada, na Fig. 23-a, apresentou uma grande região de energia mínima (na cor azul escuro) localizada entre 0,10-0,12 nm (RMSD) e 1,66-1,67 nm (R<sub>g</sub>) e uma pequena região localizada aproximadamente em rmsd 0,13 nm (RMSD) e 1,66 nm (R<sub>g</sub>), que muito provavelmente representa uma conformação transitória (estado metaestável). Nos complexos também foram identificadas possíveis regiões representando conformações transitórias acompanhando a região de energia mínima. Na DM1 (Fig. 23-b), a única região de energia mínima é acompanhada por uma pequena área em torno de 0,13 nm (RMSD) e 1,66 nm (R<sub>g</sub>). Na DM2 (Fig. 23-c), essa região de conformação transitória, com energia 1,5-3,0 kJ/mol, está no interior da única região de energia mínima – aproximadamente em 0,11 nm (RMSD) e 1,66 nm (R<sub>g</sub>). Já para a DM3 (Fig. 23-d), a região que fica em 0,10-0,20 nm (RMSD) e 1,65-1,67 nm (R<sub>g</sub>) abriga os três poços de energia mínima do complexo (na cor azul escuro) e as 4 possíveis conformações metaestáveis contabilizadas pelo FEL. Ainda assim, o valor máximo da Energia Livre de Gibbs foi de 9 kJ/mol em todos os sistemas, assim como as conformações mais estáveis apresentaram níveis de energia abaixo de 1,5 kJ/mol.

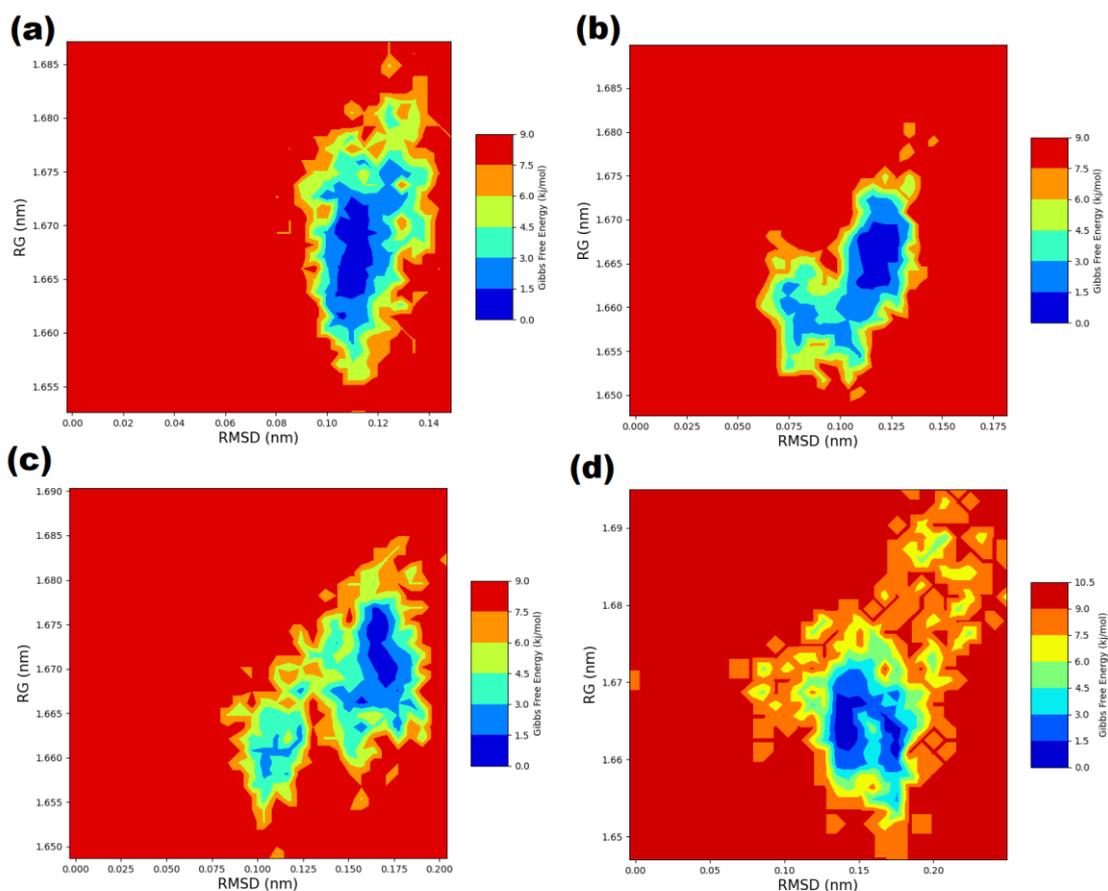


Figura 24: Análise do Mapa de Energia Livre baseado em RMSD–R<sub>g</sub> dos sistemas com a variante I (a) isolada; (b), (c) e (d) complexada com monômero de PET.

Os resultados de FEL das simulações da variante I complexada com PET estão demonstrados na Fig. 24. As DM1 e DM2 (Fig. 24-b e 24-c) são semelhantes no que diz respeito a apresentarem apenas um poço de mínima energia, respectivamente nas regiões 0,10-0,12 nm (RMSD) e 1,66-1,67 nm (R<sub>g</sub>) e 0,15-0,17 nm (RMSD) e 1,66-1,67 nm (R<sub>g</sub>), com o detalhe de que na DM2 existe uma região de provável conformação transitória aproximadamente em 0,16 nm (RMSD) e 1,66 nm (R<sub>g</sub>). A região 0,17 nm (RMSD) e 1,67 nm (R<sub>g</sub>) abriga os dois poços de energia mínima do complexo. Para a variante I isolada e seus complexos o valor máximo da Energia Livre de Gibbs permaneceu entre 9 e 10,5 kJ/mol e as conformações mais estáveis apresentaram níveis de energia abaixo de 1,5 kJ/mol.

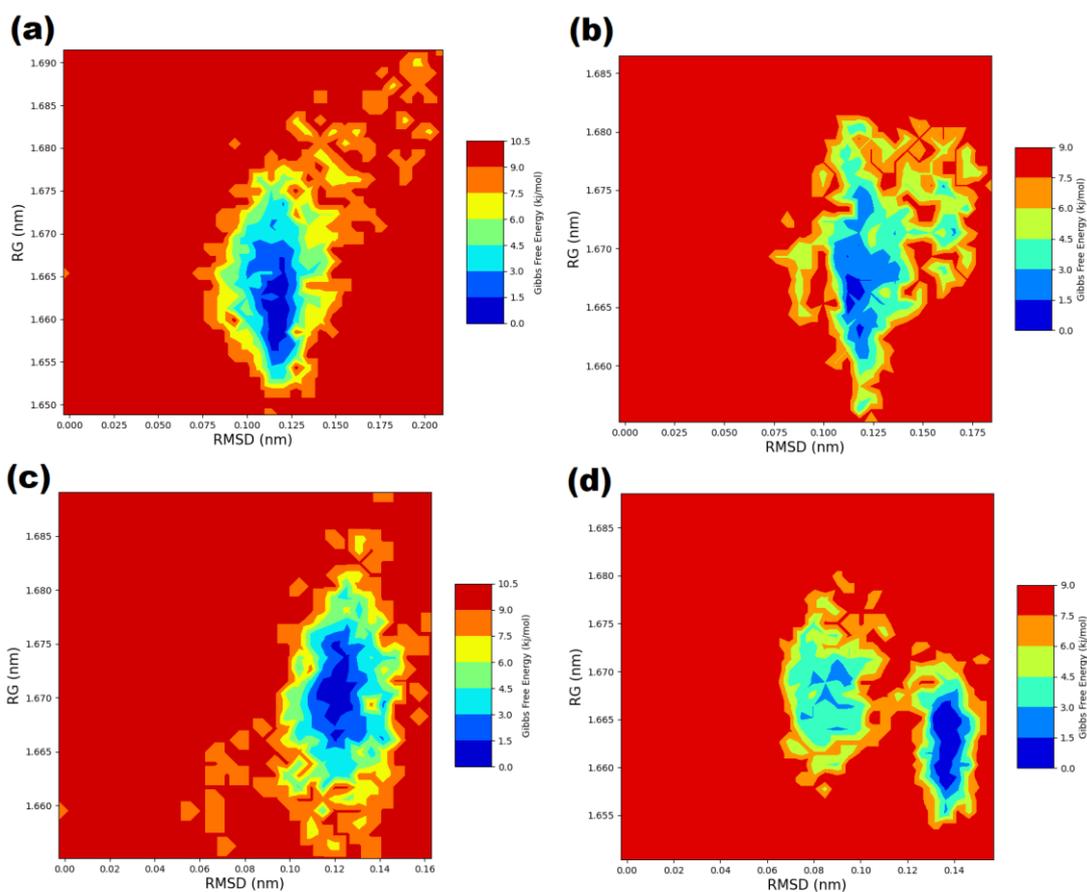


Figura 25: Análise do Mapa de Energia Livre baseado em RMSD–R<sub>g</sub> dos sistemas com a variante II (a) isolada; (b), (c) e (d) complexada com monômero de BHET.

Os gráficos de FEL da variante II isolada e complexada com o monômero de BHET são apresentados na Fig. 25. Foi observado um único poço de energia mínima global para a variante II isolada na região 0,11-0,12 nm (RMSD) e 1,65-1,66 nm (R<sub>g</sub>). A DM3 da variante I complexada apresentou resultado semelhante, com uma única região de mínimo global em 0,13-0,14 nm (RMSD) e 1,65-1,65 nm (R<sub>g</sub>). Para as outras simulações (DM1 e DM2 nas Fig. 25-b e 25-c, respectivamente) a única região de energia mínima está acompanhada de uma provável região de conformação transitória: (I) na DM1 a região de mínima energia fica em 0,11-0,12 nm (RMSD) e 1,66-1,67 nm (R<sub>g</sub>), enquanto que sua pequena região de conformação transitória está aproximadamente em 0,12 nm (RMSD) e 1,66 nm (R<sub>g</sub>); e (II) na DM2 a região de mínima energia fica em 0,11-0,13 nm (RMSD) e 0,16-0,17 nm (R<sub>g</sub>), ao passo que o espaço representando sua conformação transitória está aproximadamente em 0,12 nm (RMSD) e 1,66 nm (R<sub>g</sub>). Os resultados também demonstram a Energia Livre de Gibbs máxima, nos quatro sistemas em questão, entre 9 e 10,5 kJ/mol. De forma semelhante aos casos anteriores, todas as proteínas demonstraram conformações estáveis em níveis abaixo de 1,5 kJ/mol (indicado pela cor azul escuro no gráfico).

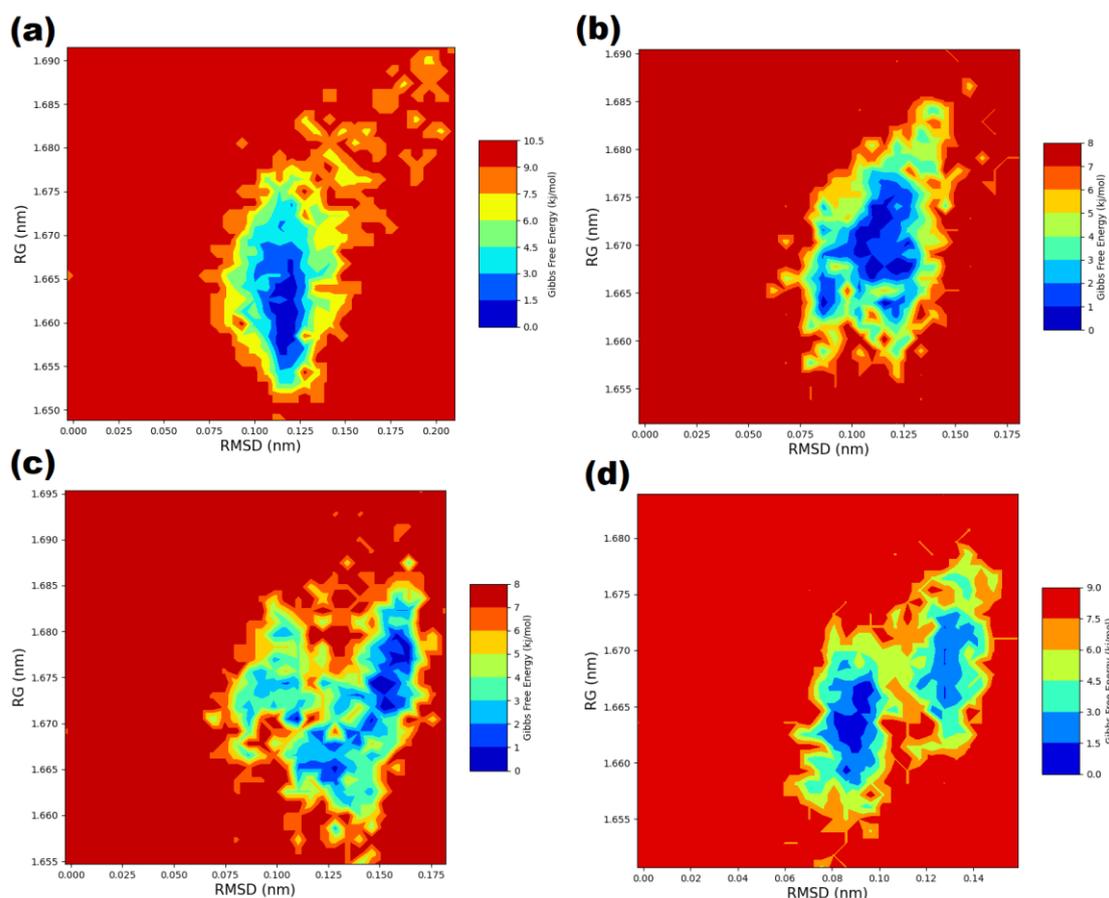


Figura 26: Análise do Mapa de Energia Livre baseado em RMSD–R<sub>g</sub> dos sistemas com a variante II (a) isolada; (b), (c) e (d) complexada com monômero de PET.

Os complexos da variante II com o monômero de PET comparadas à variante II isolada estão representadas na Fig. 26. A região 0,10-0,12 nm (RMSD) e 1,66-1,67 nm (R<sub>g</sub>) abriga o que parecem ser duas regiões acessíveis ao complexo na DM1 (Fig. 26-b). Além disso, há duas regiões que possivelmente representam conformações transitórias do complexo: (1) 0,08 nm (RMSD) e 0,16 nm (R<sub>g</sub>); (2) 0,12 nm (RMSD) e 0,16 nm (R<sub>g</sub>). Os gráficos relacionados à DM2, Fig. 26-c, indicam três regiões de mínimo energético bastante compactas representando conformações acessíveis à proteína. Essas regiões são separadas por barreiras de energia não muito altas (1,5-3,0 kJ/mol). No cenário da DM2 também ficaram evidenciadas algumas regiões de conformações transitórias, como por exemplo em 0,12 nm (RMSD) e 1,66 nm (R<sub>g</sub>). Na Fig. 26-d, os resultados da DM3 apontam um único poço de energia mínima em 0,08-0,1 nm (RMSD) e 1,66-1,67 nm (R<sub>g</sub>) acompanhada de uma região com estado metaestável aproximadamente em 0,09 nm (RMSD) e 1,66 nm (R<sub>g</sub>). Esse conjunto de simulações apresentou a maior variação da Energia Livre de Gibbs máxima que foi de 8-10,5 kJ/mol, mas manteve os níveis de conformações estáveis abaixo de 1,5 kJ/mol como nos demais casos apresentados.

## 5.2.6 Análise dos estados de protonação

Os estados de protonação apresentaram um comportamento bem definido para os resíduos passíveis de titulação estabelecidos na seção 4.4. Os resíduos Asp178 e Tyr59 permaneceram, respectivamente, protonados e desprotonados em todos os sistemas analisados. O algoritmo fez poucas ou nenhuma tentativa de alterar o estado de protonação desses resíduos durante a simulação, de modo que pode ser dito que se mantiveram estáveis.

Com base no caráter alcalino do meio (pH 9), os grupos carbonila e amino da His estiveram, respectivamente, protonados e desprotonados durante toda a simulação. Entretanto, o anel imidazol (que possui  $pK_a$  de aproximadamente 6) foi quem determinou a carga resultante da His. Isso porque em meio alcalino o anel apresenta 1 de seus 2 átomos de Nitrogênio desprotonados – podendo ser  $N_\delta$  ou  $N_\epsilon$  –, caso em que a His é neutra. Para os casos analisados, os resultados reunidos nas Tab. 3, 4 e 5 demonstram a preferência da His209 pelo estado 1 (E1) – com  $N_\delta$  protonado.

Tabela 3: Distribuição dos estados de protonação da His209 durante a simulação para os complexos com a proteína nativa. E1 representa o estado 1, monoprotonado em  $N_\delta$ ; E2 representa o estado 2, monoprotonado em  $N_\epsilon$ .

Sistema	Simulação	His209 – Estado de protonação
WT-BHET	DM1	100% E1
	DM2	85% E1 – 15% E2
	DM3	100% E1
WT-PET	DM1	96% E1 – 4% E2
	DM2	83% E1 – 17% E2
	DM3	78% E1 – 22% E2

Tabela 4: Distribuição dos estados de protonação da His209 durante a simulação para os complexos com a variante I. E1 representa o estado 1, monoprotonado em N<sub>δ</sub>; E2 representa o estado 2, monoprotonado em N<sub>ε</sub>.

I180V-BHET	DM1	60% E1 – 40% E2
	DM2	73% E1 – 27% E2
	DM3	87% E1 – 13% E2
I180V-PET	DM1	98% E1 – 2% E2
	DM2	89% E1 – 11% E2
	DM3	82% E1 – 18% E2

Tabela 5: Distribuição dos estados de protonação da His209 durante a simulação para os complexos com a variante II. E1 representa o estado 1, monoprotonado em N<sub>δ</sub>; E2 representa o estado 2, monoprotonado em N<sub>ε</sub>.

I180A-BHET	DM1	83% E1 – 17% E2
	DM2	59% E1 – 41% E2
	DM3	100% E1
I180A-PET	DM1	84% E1 – 16% E2
	DM2	35% E1 – 65% E2
	DM3	68% E1 – 32% E2

## 5.3 Decomposição de Energia e Distância proteína-ligante

### 5.3.1 Complexos com a proteína nativa (WT)

A Fig. 27 apresenta a decomposição de energia média por resíduo dos complexos da proteína nativa com o monômero de BHET. Os dados indicaram de forma clara o mesmo conjunto de resíduos (chamados aqui de conjunto  $\lambda$ ) nas três simulações conduzidas: Tyr59, Ser132, Trp157, Asp178, Ile180 e His209, com energias que se mantiveram semelhantes em extensão em todos os cenários: variando entre -1 e -4 kcal/mol e a His209 em torno de 2 kcal/mol.

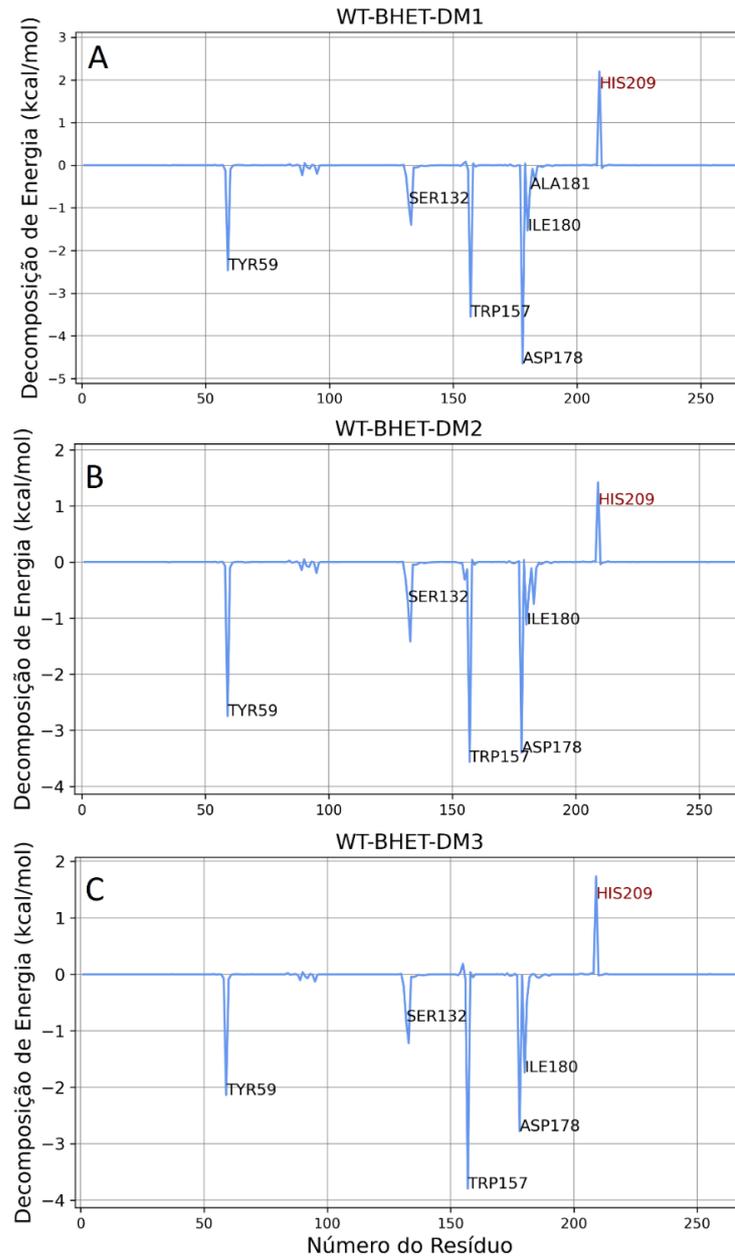


Figura 27: Valor médio temporal da decomposição de energia por resíduo ao longo da simulação. Os gráficos apresentados referem-se aos complexos da proteína nativa com o monômero de BHET apresentados da seguinte maneira: DM1 (A), DM2 (B) e DM3 (C).

Desse conjunto, quatro desses resíduos (His209, Trp157, Ile180 e Tyr59) também são indicados por Han, Xu *et al.*, 2017 [33] como interagentes com o ligante para a formação do complexo com a *IsPETase*. A interação desses resíduos está de acordo com o que apontaram os resultados de  $d_{HC}$  para os complexos da proteína nativa com o BHET, que indicaram a permanência do ligante no sítio ativo da proteína (Fig. 28-A) a uma distância favorável ao ataque nucleofílico –  $d_{HC}$  médio de 4,6 Å.

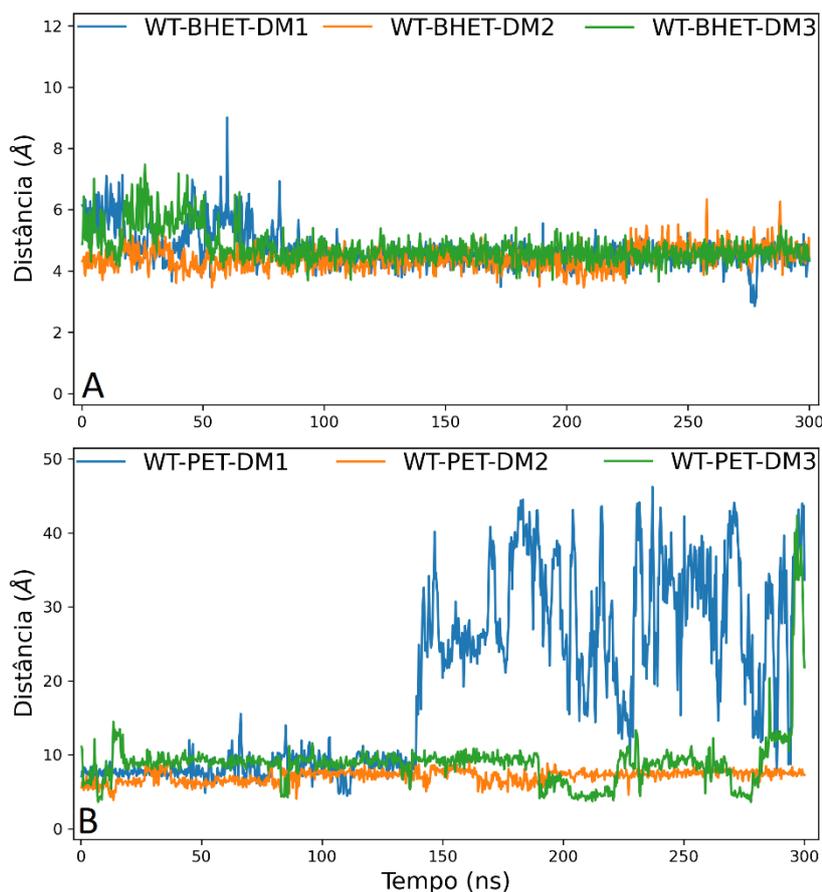


Figura 28: Resultados da distância proteína-ligante ( $d_{HC}$ ) durante a simulação. Os gráficos representam os complexos da proteína nativa com o BHET (topo) e PET (inferior).

Para os complexos da proteína nativa com o monômero de PET (Fig. 29) a decomposição de energia da DM1 foi desconsiderada, dado que, pela Fig. 28-B, o ligante se desprende da proteína próximo dos 150 ns. Por outro lado, nota-se na decomposição de energia da DM2 (Fig. 29-A) a presença de resíduos diferentes e em menor número em comparação ao conjunto  $\lambda$  evidenciado nos complexos com BHET. Suas energias são menos favoráveis ao ancoramento do ligante e o resíduo mais evidente foi a Ala155 (que não contribuiu anteriormente). Na DM3 do complexo com o PET (Fig. 29-B), evidenciou-se a His209 que apresentou uma energia de interação bastante alta (6 kcal/mol) em relação aos demais resíduos, o que pode ter levado ao desprendimento do ligante ao final da simulação como demonstrado na Fig. 28-B.

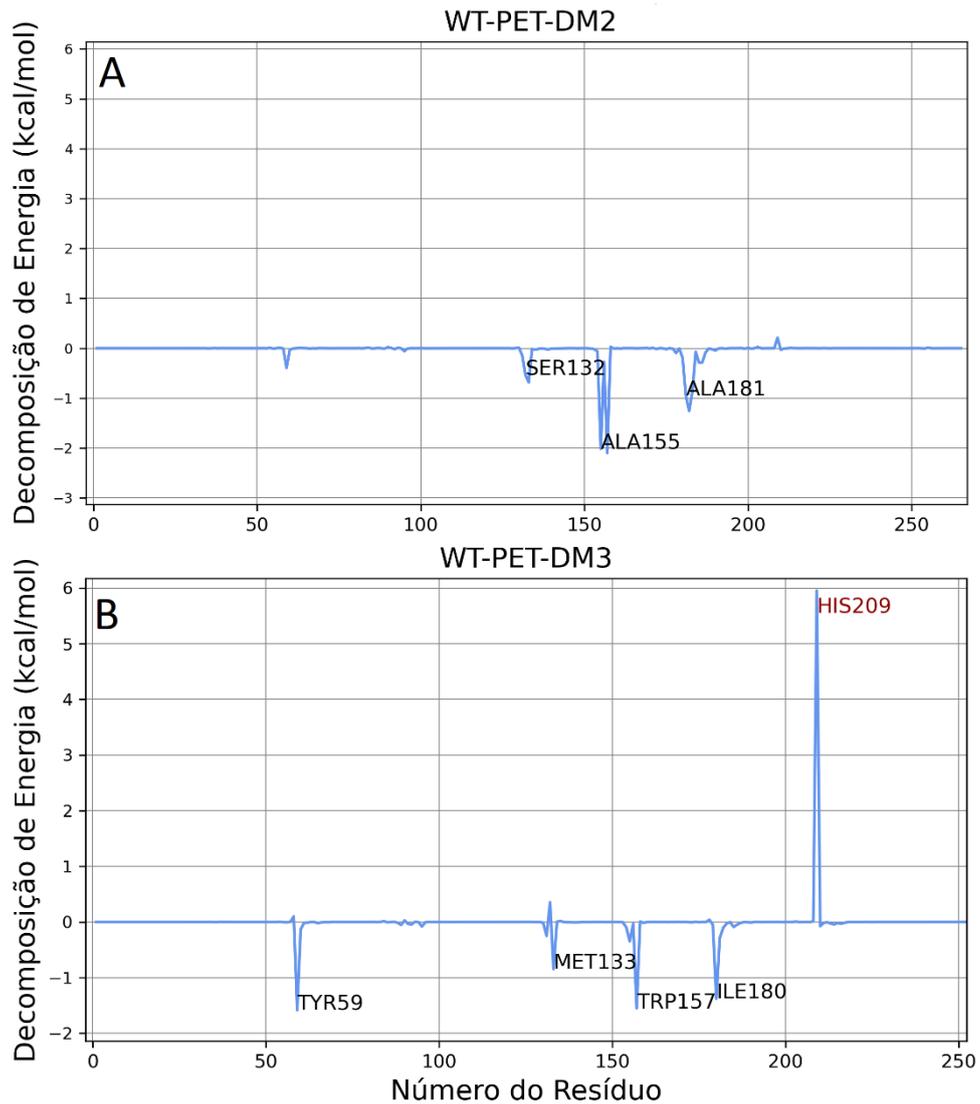


Figura 29: Valor médio temporal da decomposição de energia por resíduo ao longo da simulação. Os gráficos apresentados referem-se aos complexos da proteína nativa com o monômero de BHET: DM2 (A) e DM3 (B).

### 5.3.2 Complexos com a variante I (I180V)

Em relação à decomposição de energia por resíduo dos complexos da variante I com o monômero de BHET (Fig. 30), observou-se a presença da Tyr59 e Trp157 (do conjunto  $\lambda$  definido anteriormente) em todas as três simulações, com energias que se mantiveram relativamente estáveis – cerca de -2 kcal/mol e -3 kcal/mol, respectivamente.

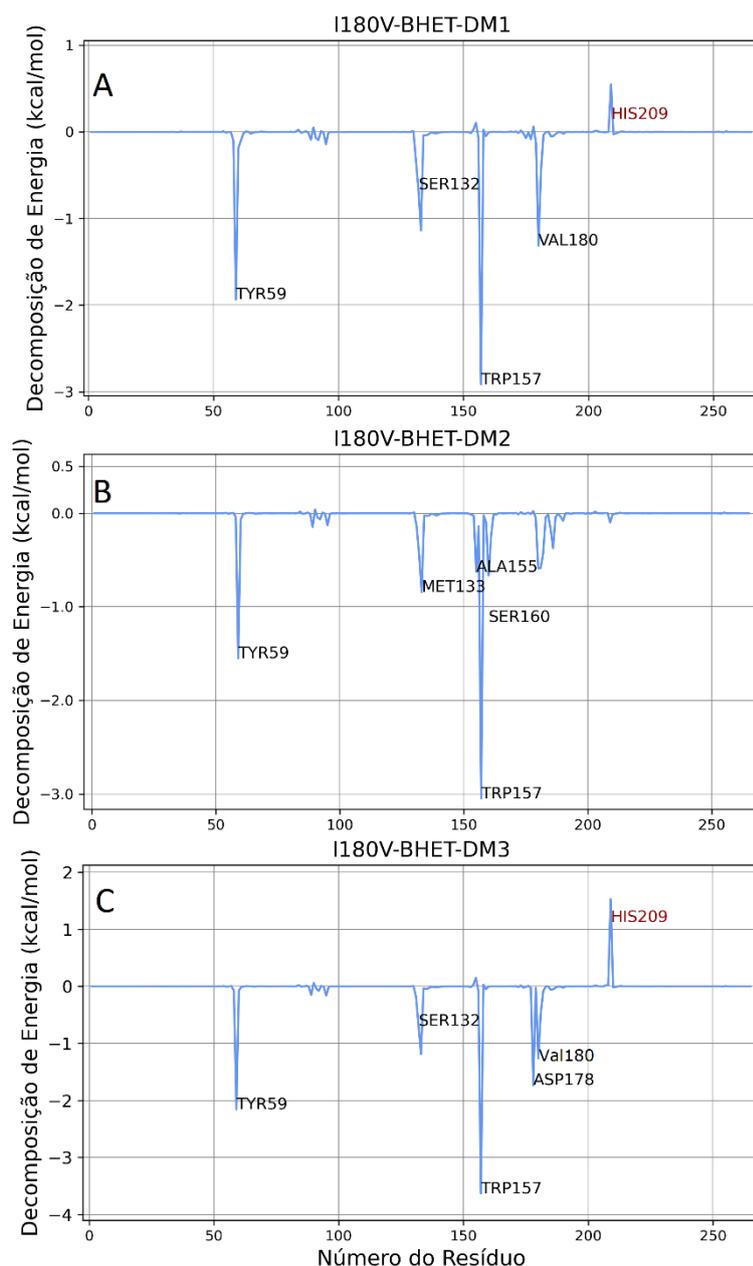


Figura 30: Valor médio temporal da decomposição de energia por resíduo ao longo da simulação. Os gráficos apresentados referem-se aos complexos da variante I com o monômero de BHET: DM1 (A), DM2 (B) e DM3 (C).

Os demais resíduos de  $\lambda$  (Ser132, Asp178, res180 e His209) também contribuíram, mas não em todas as simulações. Na Fig. 30-A, entre os resíduos da DM1 da variante I complexada com BHET, apenas o Asp178 não apresentou interação com o ligante para completar o conjunto  $\lambda$ , possivelmente dificultando o posicionamento do ligante no sítio ativo, como indicado no início da simulação pela Fig. 31-A; o que ainda gerou um  $d_{HC}$  médio de 4,7 Å (favorável ao ataque nucleofílico). Na DM2 (Fig. 30-B), do conjunto  $\lambda$ , apenas a Tyr59 e Trp157 interagiram com o ligante, acompanhados por interações pouco favoráveis por parte dos resíduos Ala155, Met133 e Ser160, o que pode ter levado ao resultado indicado na Fig. 31-A ( $d_{HC}$  médio de 11,8 Å). Já no caso da DM3 (Fig. 30-C), apenas a Val180 (substituída da Ile180) não contribuiu para completar o

conjunto  $\lambda$ , com os demais resíduos destacados apresentando energias bastante semelhante às observadas nos complexos da proteína nativa com o BHET, de forma a contribuírem com o ancoramento do ligante na proteína a uma distância favorável ao ataque nucleofílico ( $d_{HC}$  médio de 4,7 Å) como indicado na Fig. 31-A. Nos complexos da variante I com o monômero de PET, e decomposição de energia média por resíduo foi desconsiderada por, como apontam os resultados de  $d_{HC}$  na Fig. 31-B, os complexos não se mantiveram, com o ligante se despreendendo da proteína e/ou caminhando na sua superfície de forma errante, sem estabelecer ligações estáveis com resíduo algum.

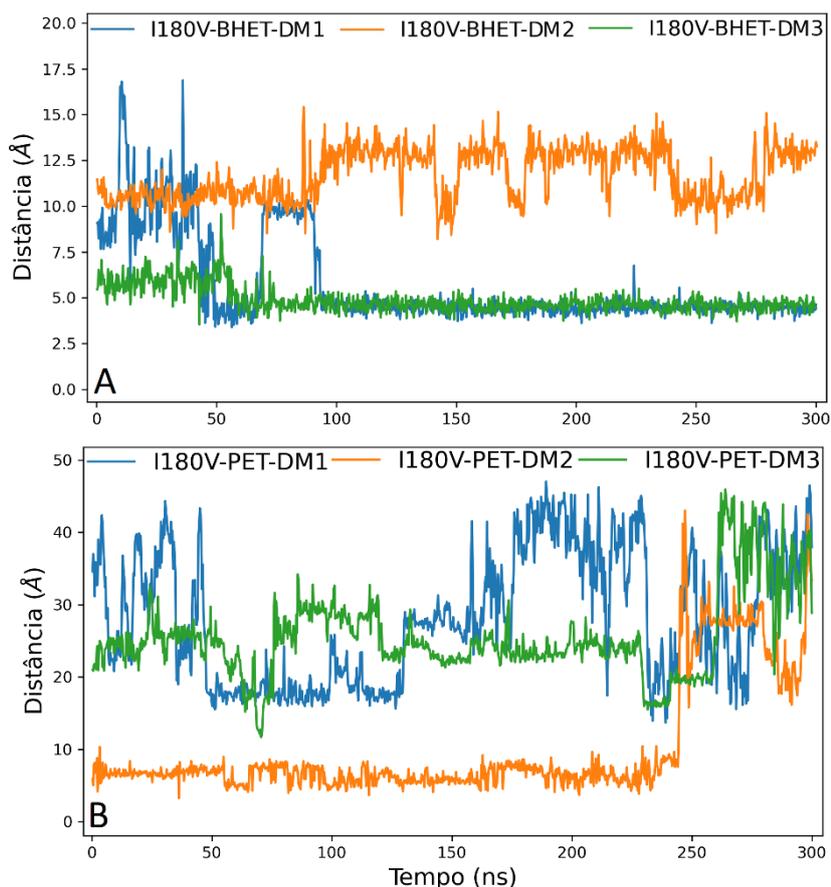


Figura 31: Resultados da distância proteína-ligante ( $d_{HC}$ ) durante a simulação. Os gráficos representam os complexos da variante I com o BHET (topo) e PET (inferior).

### 5.3.3 Complexos com a variante II (I180A)

A decomposição de energia por resíduo dos complexos da variante II com o monômero de BHET (Fig. 32) exibe três cenários distintos para cada uma das simulações. Na DM1 da variante II (Fig. 32-A), foi observado um conjunto de resíduos completamente distinto do conjunto  $\lambda$  (definido anteriormente na apresentação dos complexos da proteína nativa).

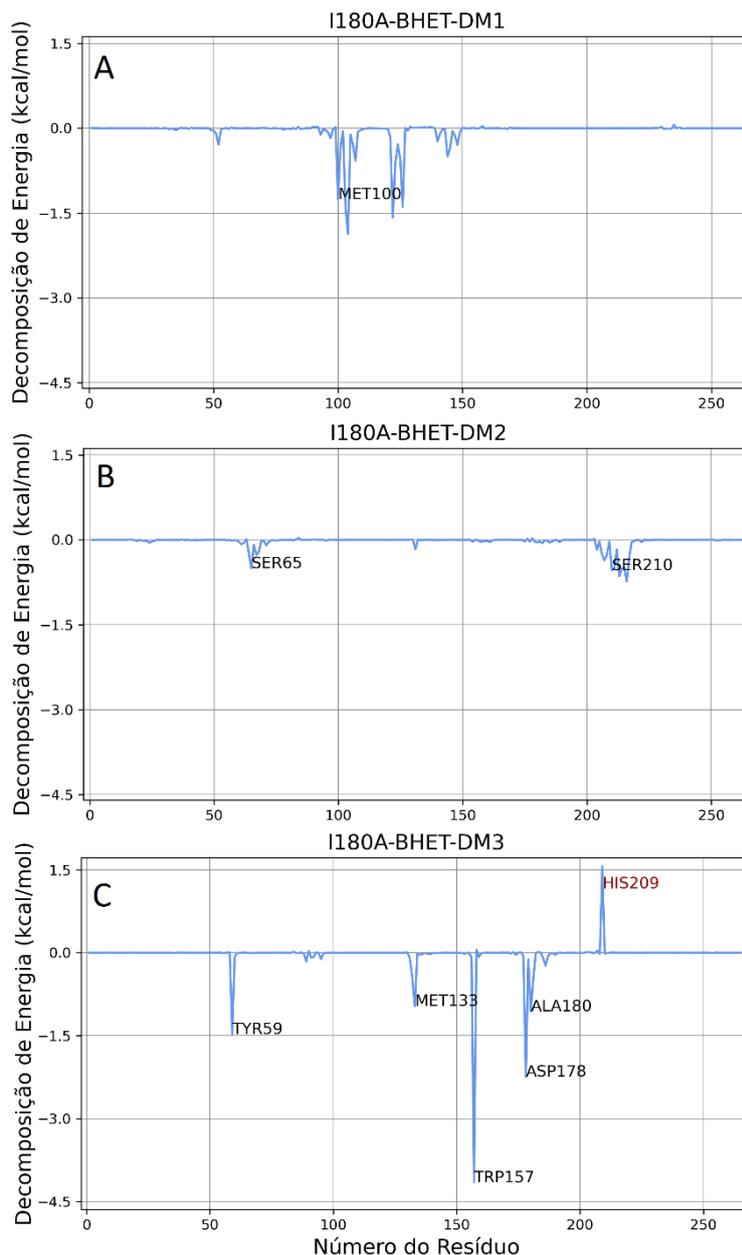


Figura 32: Valor médio temporal da decomposição de energia por resíduo ao longo da simulação. Os gráficos apresentados referem-se aos complexos da variante II com o monômero de BHET: DM1 (A), DM2 (B) e DM3 (C).

Na DM1 contribuíram os resíduos Arg104 e Asp122 (com energias entre -2 e -1,5 kcal/mol, respectivamente), seguidos pelos resíduos Met100, Met126 (com energias mais próximas de -1,5 kcal/mol). A DM2 da variante II complexada com BHET (Fig. 32-B) indicou resíduos vizinhos ao sítio ativo da proteína (Ser65, Ser210 e Asn216) com energia pouco favoráveis ao ancoramento. No caso da DM3 (Fig. 33-C), contribuíram quatro dos seis resíduos do conjunto  $\lambda$  (Tyr59, Trp157, Asp178, His209), todos eles com valores de energia equiparável àqueles vistos nos complexos da proteína nativa exibidos na Fig. 27. Além disso, houve ainda uma fraca participação energética por parte dos resíduos Ala180 e Met133 (com aproximadamente -1,3 Kcal/mol). Os resultados da decomposição de energia concordam com as distâncias ligante-proteína apontadas na Fig. 7-A para cada

uma das simulações, com valores que vão de 6 até 24 Å – muito acima da distância favorável ao ataque nucleofílico estabelecida na seção 2.3. Para os complexos da variante II com o monômero de PET, a decomposição de energia média por resíduo foi desconsiderada por, como apontam os resultados de  $d_{HC}$  na Fig. 7, os complexos não se mantiveram, com o ligante se desprendendo da proteína e/ou caminhando na sua superfície de forma errante, sem estabelecer ligações estáveis com resíduo algum.

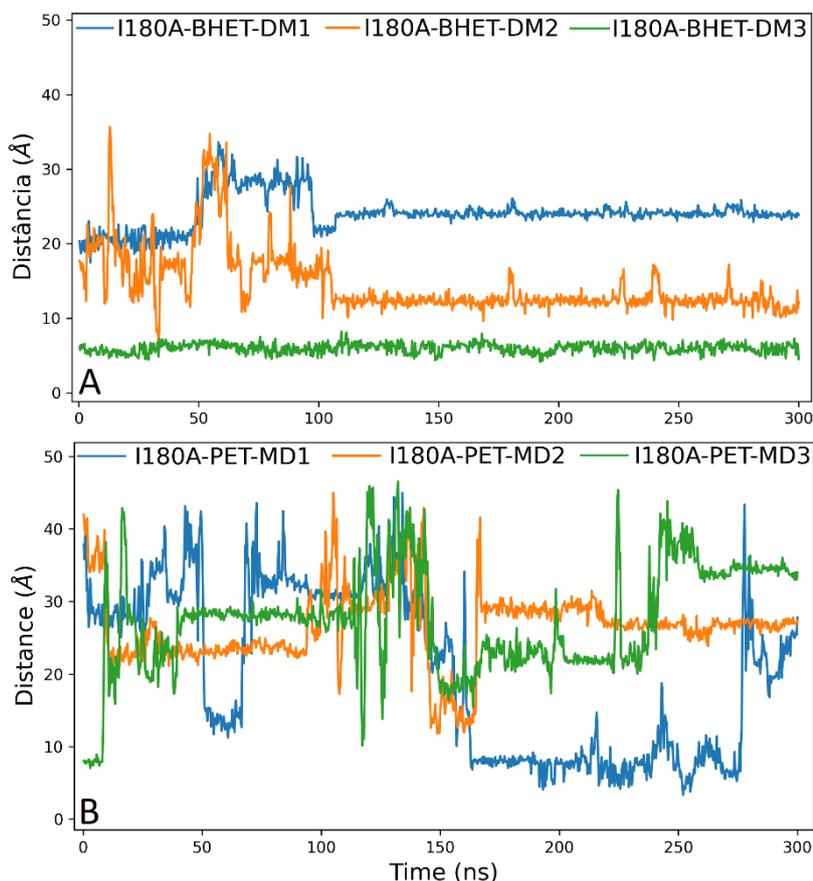


Figura 33: Resultados da distância proteína-ligante ( $d_{HC}$ ) durante a simulação. Os gráficos representam os complexos da variante II com o PET (topo) e PET (inferior).

### 5.3.4 Energia livre Gibbs nos complexos

O MMPBSA.py também forneceu informações sobre a energia de cada uma das componentes do sistema: complexo, proteína, ligante e a diferença da energia livre de ligação dada pela equação (19).

$$\Delta\Delta G = \Delta G_{complexo} - \Delta G_{receptor} - \Delta G_{ligante} \quad (19)$$

A segunda coluna da Tab. 6 apresenta as energias de cada complexo onde o ligante permaneceu ancorado na proteína durante todo o período da simulação – seja no sítio ativo ou em algum outro local na superfície. Uma análise mais detalhada demonstra que a variação da energia dos complexos foi desprezível. O valor médio para cada triplicata

foi de: -4934,65 kcal/mol (complexo da proteína nativa), -4944,63 kcal/mol (complexo da variante I) e -4931,72 kcal/mol (complexo da variante II). Com base no complexo da proteína nativa, a variação entre os valores médios de  $\Delta G_{\text{complexo}}$  ficou abaixo de 1%.

Tabela 6: Valores da energia livre de Gibbs dos sistemas analisados.

Sistema	Simulação	$\Delta G_{\text{complexo}}$ (kcal/mol)	$\Delta\Delta G$ (kcal/mol)
<b>Nativa-BHET</b>	DM1	-4937,25	-27,66
	DM2	-4936,84	-27,50
	DM3	-4929,87	-24,66
<b>I180V-BHET</b>	DM1	-4932,55	-19,81
	DM2	-4948,39	-20,56
	DM3	-4952,96	-21,85
<b>I180A-BHET</b>	DM1	-4932,24	-23,80
	DM2	-4913,97	-15,95
	DM3	-4948,97	-22,34

Por outro lado, as variações na energia livre de ligação dos complexos foram mais evidentes. A média desse valor para cada uma das triplicatas tem-se: -26,60 kcal/mol (complexo da proteína nativa), -20,74 kcal/mol (complexo da variante I) e -20,69 kcal/mol (complexo da variante II). A comparação entre esses valores, com base no complexo da proteína nativa, indica que  $\Delta\Delta G$  sofre uma redução de aproximadamente 22%.

Nas simulações da proteína nativa com o PET, em sua DM1 o ligante se despreendeu, assim como na DM3 em que  $\Delta\Delta G$  foi aproximadamente 0. Na sua DM2, onde o ligante permaneceu ancorado,  $\Delta\Delta G = -16,75$  kcal/mol. Já nas simulações das variantes com o PET, o ligante não apresentou interação estável com a proteína.

## 6 Discussão dos resultados

Nas seções seguintes são discutidos os resultados apresentados no capítulo 5. Na primeira seção se discute aspectos predominantemente estruturais, tratando dos resultados de RMSD, raio de giro ( $R_g$ ) e RMSF, mas também os relaciona com os resultados de mapas de energia (FEL). Adiante, na segunda seção, são discutidos os resultados sobre os estados de protonação e logo depois, na terceira seção, os resultados à respeito da energia livre de ligação por resíduo (dada pela decomposição de energia do MMPBSA.py) e distância proteína-ligante ( $d_{HC}$ ). Na última seção deste capítulo, discute-se as implicações da energia livre de Gibbs, outro resultado fornecido pela decomposição de energia do MMPBSA.py.

### 6.1 Sobre os resultados estruturais e mapas de energia

Os resultados de RMSD da proteína nativa isolada, que se mostrou mais regular em torno da média, podem ser entendidos como a estabilização de energia da proteína em um dos três poços de energia indicados pelos resultados de FEL, bastante concentrados e centralizados – sinal de estabilidade energética da proteína. Por outro lado, os complexos com o monômero de BHET exibiram poços de mínimo global de energia mais amplos, com destaque para a DM2 com dois poços de energia bastante próximos separados por uma barreira de energia não muito alta – o que pode explicar a variação repentina no RMSD e  $R_g$  no intervalo 200-250 ns, possivelmente passando de um poço para outro.

Os complexos com o monômero de PET apresentaram variações no  $R_g$  que podem ser atribuídas a mudanças conformacionais dentro de um mesmo poço de energia. A diminuição de RMSD e  $R_g$  após 150 ns de simulação vistas na DM1 pode ser a proteína mudando sua conformação devido ao desprendimento do ligante pouco antes dos 150 ns, como indicado pelos resultados de  $d_{HC}$ . Na DM2 dos complexos com PET (WT-PET-DM2), a variação repentina observada tanto em seu RMSD quanto em seu  $R_g$  em torno dos 200 ns de simulação pode ser atribuída a uma mudança de estado de energia, dado que pelos resultados de FEL esse complexo possui dois poços de energia mínima. A DM3 dos complexos com PET apresenta um aumento no raio de giro e no RMSD a partir de aproximadamente 150 ns, o que muito provavelmente desfavoreceu o ancoramento do ligante, que se despreendeu após os 250 ns de simulação.

A variante I isolada apresentou uma tendência de aumento de  $R_g$  em relação à proteína nativa isolada. Além do raio de giro, seu RMSD também aumentou no intervalo entre 250-300 ns muito provavelmente devido ao tamanho do poço de energia mínima exibido pelos seus resultados de FEL, sugerindo uma mudança conformacional dentro do mesmo estado de energia. Nos complexos com o BHET, o aumento no RMSD e a diminuição de  $R_g$  vistos na DM1 entre 50 e 100 ns, sugerem uma variação conformacional dentro do poço de energia apontado pelo FEL do complexo. Essa variação pode ser o que permitiu o movimento do ligante nesse mesmo período – indicado no gráfico de  $d_{HC}$  da

I180V-BHET-DM1 – e retorna para a posição que mantém até o final da simulação, com  $d_{HC}$  médio de 4,7 Å, favorável ao ataque nucleofílico.

A estabilidade exibida pela DM2 no RMSD pode ser advinda do único poço de energia mínima apresentado pelos resultados de FEL. O poço de energia indicado também comporta as oscilações vistas no raio de giro – variações conformacionais dentro do poço de energia. Na DM3 as variações observadas, tanto no RMSD quanto no  $R_g$ , podem ser reflexo da proteína passando pelos três poços de energia ou ainda pelos vários estados conformacionais transitórios – todos separados por barreiras de energia não tão altas – indicados nos gráficos de FEL.

A DM1 dos complexos da variante I com o PET apresentou variações no RMSD (e  $R_g$ ) provavelmente por mudanças conformacionais dentro do mesmo poço de energia – assim como foi com a proteína isolada. Na DM2, a grande oscilação de RMSD no intervalo 80-180 ns (também presente no seu gráfico de  $R_g$ ) pode ser atribuída a uma variação conformacional que levou ao desprendimento do ligante próximo aos 250 ns de simulação, como indicado pelos resultados de  $d_{HC}$  e ligação de hidrogênio, e demonstrado na Fig. 34. O aumento nos valores de RMSD e  $R_g$  observados a partir dos 200 ns na DM3 da variante I complexada com o PET pode ser um indicativo da proteína transicionando entre os dois poços de energia indicados pelos seus resultados de FEL.

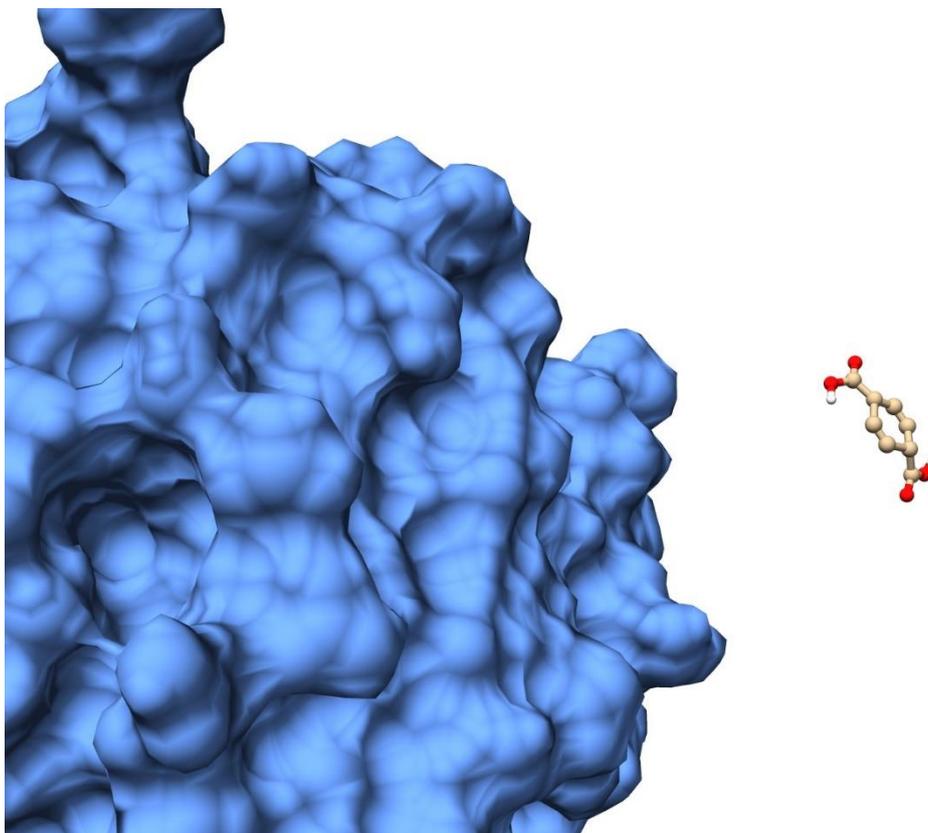


Figura 34: Complexo da proteína variante I com o monômero de BHET momentos após o ligante se desprender da proteína, após 250 ns de simulação.

A observação dos dados sobre os sistemas com a variante II (I180A), aponta para uma variação conformacional da proteína isolada pela mudança no seu RMSD a partir dos 100 ns de simulação. Além disso, houve também uma variação de raio de giro acima da média no intervalo 200-230 ns, possivelmente uma conformação transitória dentro do poço, que apresenta uma região bastante estreita que quase o divide em dois. As DM1 e DM3 da variante II complexada com o BHET se mostraram bastante alinhadas com a proteína isolada. A DM1, no mesmo intervalo mencionado (200-230 ns) também apresentou variação em seu raio de giro, possivelmente por seu poço de energia apresentar estrutura semelhante àquela da variante II isolada – indicando uma variação conformacional. Ao passo que a DM2, ao apresentar  $R_g$  maior do que a proteína isolada, sugere que a mutação pode ter causado interferência sobre as conformações acessíveis à proteína que são favoráveis ao ancoramento do ligante, dado que desde o início da simulação o ligante se encontrava fora do sítio ativo, como indicado por seus resultados de  $d_{HC}$  e exibido na Fig. 35.

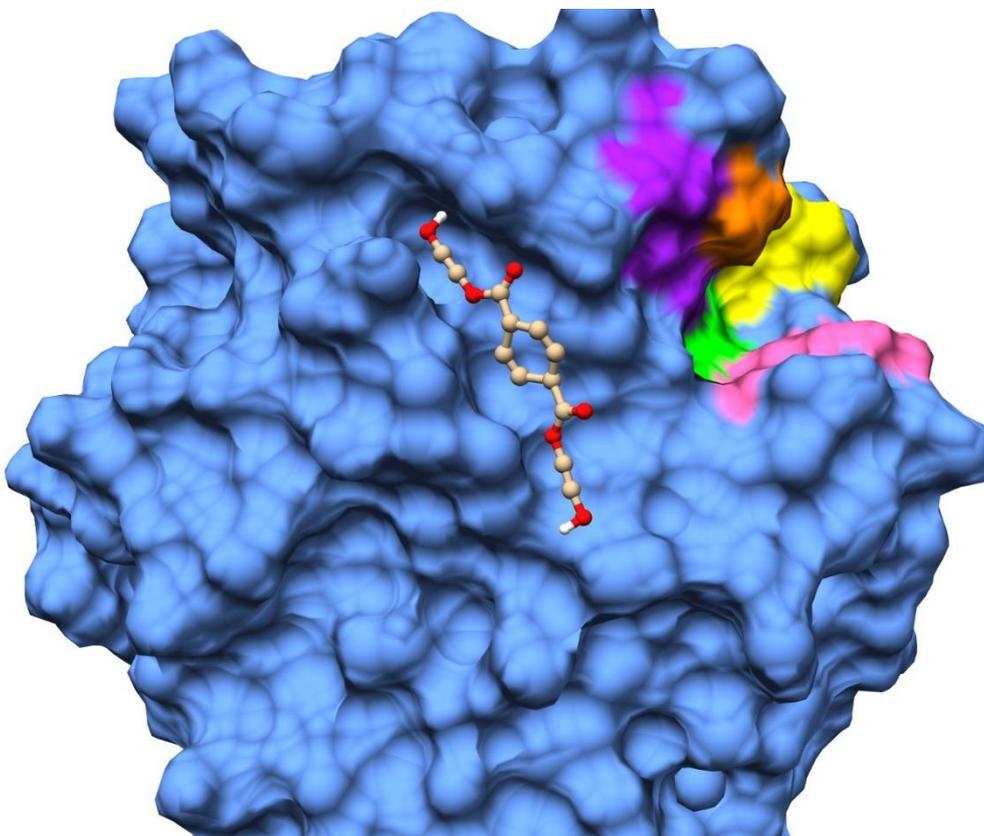


Figura 35: Complexo da DM2 da variante II com o monômero de BHET, em que o ligante saiu do sítio ativo da proteína, sem se desprender e se moveu ao longo de sua superfície durante a simulação. As cores representam a superfície da Ser132 (verde), Asp178 e His209 (roxo), Ala180 (laranja), Trp157 (amarelo) e Met133 (vermelho).

Além disso, a forma do poço de energia apresentado pelos resultados de FEL também pode ser o que permitiu a DM2 permanecer acima da média (RMSD e  $R_g$ ) de forma estável durante toda a simulação. Na DM3 dos complexos da variante II com BHET, os valores de RMSD e raio de giro permanecem em torno de 0,9 Å durante toda a simulação, possivelmente uma consequência do único poço de energia mínima obtido

por FEL, bastante restrito no RMSD, mas mais extenso no  $R_g$ . Contudo, apesar de o ligante permanecer no sítio ativo, como indicado pelo seu gráfico de  $d_{HC}$ , a distância nucleofílica obtida foi 50% maior do que aquela favorável ao ataque nucleófilo, consequência da reduzida presença de ligações de hidrogênio.

Sobre os complexos da variante II com o PET, as oscilações no RMSD depois de 150 ns observadas na DM1 se dão pela presença de mais de um poço de energia mínima que se encontram em um intervalo mais amplo no eixo de RMSD do gráfico de FEL. Como esses poços estão bastante próximos e são separados por uma barreira de energia não tão alta, esse cenário torna a proteína mais suscetível a mudanças conformacionais e energéticas. É provável que, entre os estados acessíveis a I180A-PET-DM1, o complexo adotou uma conformação desfavorável ao ligante, dado que este não permaneceu no sítio ativo da proteína, como indicado nos resultados de  $d_{HC}$ . Cenário bastante semelhante se dá com a DM2 das I180A-PET: oscilações no RMSD e  $R_g$  pela facilidade com que a proteína pode mudar de estado conformacional e energético (como indicado pelos resultados de FEL), levando às mesmas consequências discutidas na DM1 – desprendimento do ligante. As variações no RMSD na DM3 não só correspondem às observadas no  $R_g$  como também à forma do poço de energia, que permite variações conformacionais sem alterações de estado de energia. Contudo, assim como nas DM1 e DM2, o ligante se desprendeu da proteína muito provavelmente pela variante II ter adotado uma conformação desfavorável ao ancoramento.

Com base no que foi apresentado sobre a flutuação dos resíduos nos resultados de RMSF, houve uma tendência geral de diminuição na flexibilidade da proteína nas regiões discutidas devido às mutações implementadas na proteína. Os complexos apresentaram suas regiões II e III – região II abrangendo os resíduos 170-190 e região III os resíduos 205-220 – com uma tendência de redução na flexibilidade, acompanhando a tendência geral mencionada anteriormente. Observando mais atentamente a flutuação dessas regiões, foi possível notar que na variante I houve uma tendência de diminuição na flexibilidade no entorno do sítio 180 (região II), ao mesmo tempo que houve uma tendência de aumento na flexibilidade nas proximidades da His209 (região III). Por outro lado, na variante II notou-se uma tendência de aumento na flexibilidade no entorno do sítio 180 e uma tendência de redução na flexibilidade nas proximidades da His209. Essa alteração “inversa” sobre a flexibilidade dessas regiões vizinhas no sítio ativo, pode desequilibrar a flutuação média das estruturas na região, fazendo com que o processo de ancoramento se torne mais difícil e acarretando ao que foi demonstrado nas figuras de  $d_{HC}$  nas diferentes simulações apresentadas: ligante posicionado distante da ser132 (nucleófilo), ligante posicionado fora do sítio, caminha pela superfície da proteína ou se desprendendo ao longo da simulação

## 6.2 Estados de protonação

Existe uma tendência de a His permanecer mono protonada no Estado 2 (E2) – com H em  $N_\epsilon$  –, uma das três possibilidades em que possui prótons (Fig. 36). Contudo, em situações em que a His pertence à tríade catalítica de enzimas, como nas serina proteases (sendo a *IsPETase* um exemplo desse tipo de enzima), ela tem preferência pelo Estado (1) – com H em  $N_\delta$  –, que pode ser entendido como favorável ao ancoramento de ligantes na proteína [76,77].

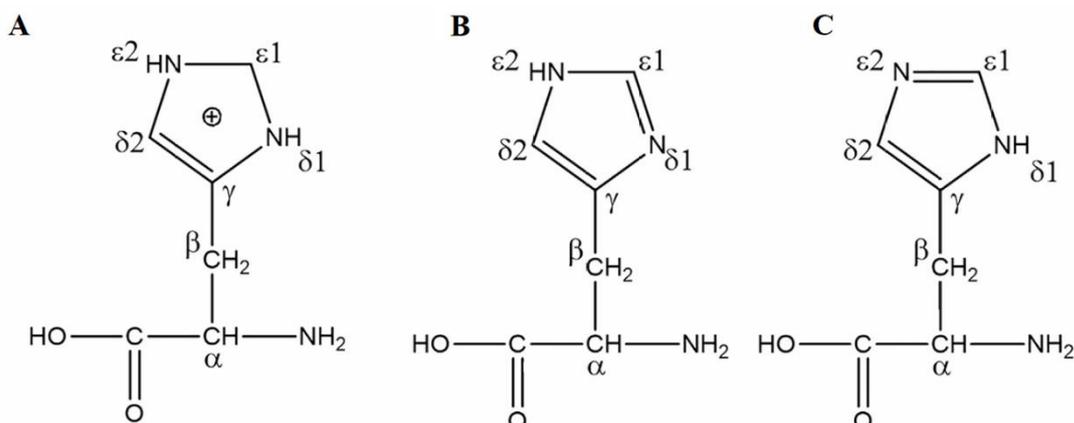


Figura 36: Representação dos estados de protonação do anel imidazol da Histidina: duplamente protonado (A), mono protonado em  $N_\epsilon$  (B) e mono protonado em  $N_\delta$  (C) – Adaptado de [78].

Considerando o comportamento dos estados de protonação da His209 apresentados na seção 5.2.6 (exibidos na Tab. 3, 4 e 5), foi possível notar sua tendência de permanecer mono protonada no Estado 1 (E1) em todos os complexos analisados – como era esperado. Contudo, o valor médio de permanência em E1 é diferente para cada um dos complexos, sendo aproximadamente: 91% (proteína nativa), 81% (variante I) e 71% (variante II). Como a carga da His em  $N_\delta$  ou  $N_\epsilon$  é a mesma, alternar entre esses estados acarreta em alterações geométricas no seu anel imidazol [79]. A essas alterações podem ser atribuídas as mudanças na formação de ligações de hidrogênio entre a proteína e o ligante, assim como o comprometimento do ancoramento, como indicado pelos resultados de  $d_{HC}$ .

## 6.3 Energia livre de ligação por resíduo e distância nucleofílica

Os resultados dos complexos da proteína nativa com o monômero de BHET, como apresentado na seção de resultados se mostraram os mais bem sucedidos entre as simulações conduzidas, o que se evidencia pela presença do que se convencionou chamar conjunto  $\lambda$  – resíduos Tyr59, Ser132, Trp157, Asp178, Ile180 e His209. Esse conjunto, presente nas três simulações realizadas, é bastante importante para o ancoramento do ligante, como evidenciado pelos seus resultados de  $d_{HC}$ , tanto pela regularidade do gráfico quanto pelo valor médio favorável ao ataque nucleofílico ( $d_{HC} = 4,6 \text{ \AA}$ ). Além disso, os dados sobre as ligações de hidrogênio destacam a predominância da interação do Asp178

(evidenciado nas três simulações pela decomposição de energia) tanto como receptor quanto como doador, indicando um papel estabilizador do Asp178 no ancoramento do ligante – Fig. 37.

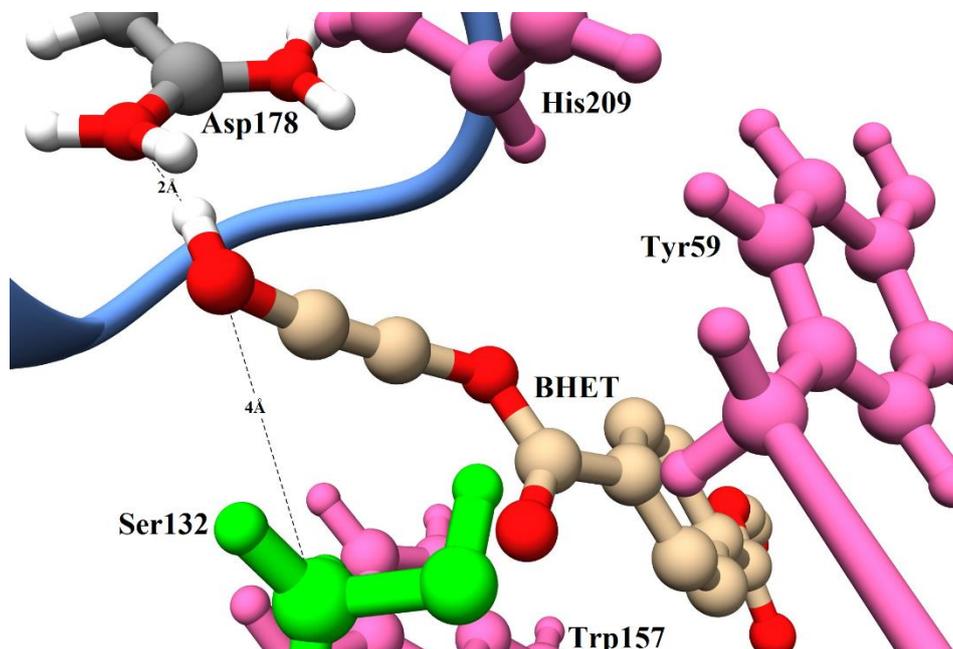


Figura 37: Complexo da proteína nativa com o BHET (DM1). A ampliação da região do sítio ativo mostra o ligante próximo ao nucleófilo da proteína (distância média de 4 Å da Ser132) e também a uma distância favorável a formação de ligações de hidrogênio do Asp178.

Para os complexos com o PET (com exceção da DM1, onde o ligante se desprende da proteína ao longo da simulação, como indicado na Fig. 38), a predominância da interação com a Ala155 está de acordo com a distância proteína-ligante exibida na Fig. 39 para a DM2, pois dada sua posição espacial ela pode afastar o ligante da Ser132, aumentando a distância proteína-ligante adequada para o início do processo de degradação – o que parece ter sido o caso, dado que seu  $d_{HC}$  médio foi de aproximadamente 7 Å. No caso da DM3, o alto valor de energia devido à His209 se coloca como causa provável do desprendimento do ligante após os 250 ns de simulação.

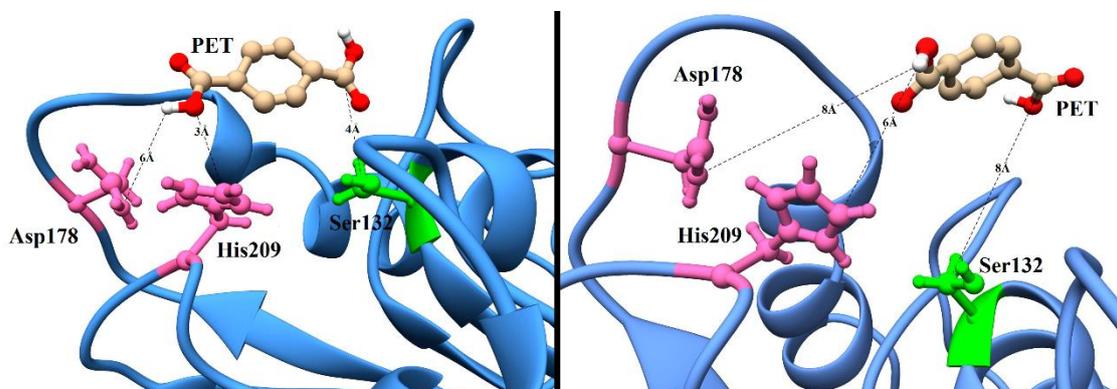


Figura 38: O ligante se desprende durante a simulação DM1 do complexo da proteína nativa com o PET.

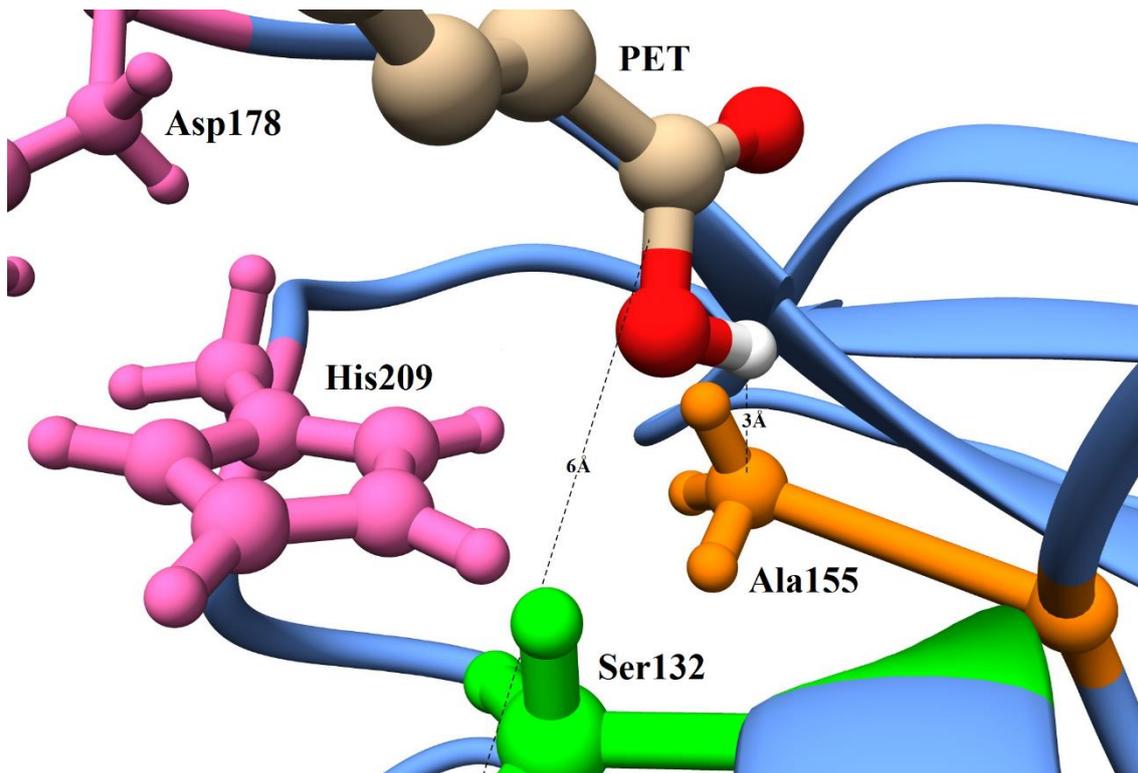


Figura 39: A ampliação do sítio ativo no complexo da proteína nativa com o PET demonstra o ligante deslocado no sítio ativo, mais próximo da Ala155, como sugerido pelas ligações de hidrogênio formadas.

Como foi mencionado na apresentação dos resultados, alguns resíduos do conjunto  $\lambda$  não contribuíram pela decomposição de energia. Na DM1 dos complexos da variante I com o BHET a ausência do Asp178 pela decomposição de energia, aliada às ligações de hidrogênio com permanência abaixo de 40% da simulação, podem ser considerados como fatores que permitiram a movimentação do ligante no sítio ativo da proteína de modo a realizar um giro (como indicado na Fig. 40), mudando completamente a sua posição de ancoramento, mas mantendo  $d_{HC}$  baixo (conforme apontado nos gráficos na seção de resultados).

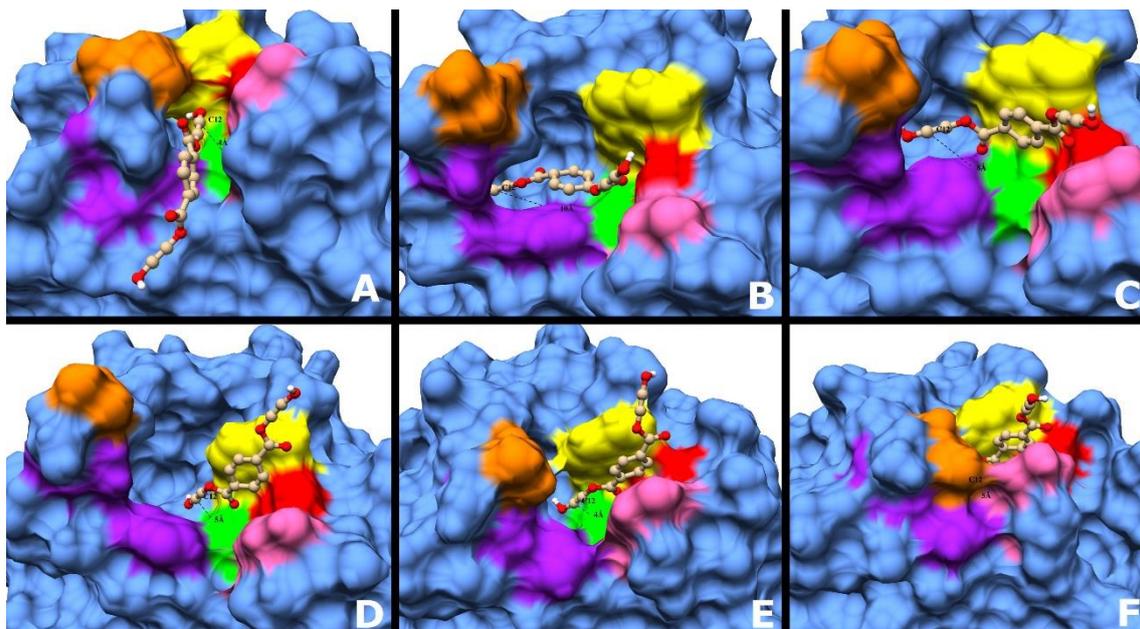


Figura 40: Sequência de imagens demonstrando variação da posição do ligante no sítio ativo na DM1 do complexo da variante I com o BHET. Durante a simulação o ligante realizou um giro e acabou se posicionando de maneira deslocada no sítio ativo. As cores representam a superfície da Ser132 (verde), Asp178 e His209 (roxo), Val180 (laranja), Trp157 (amarelo) e Met133 (vermelho).

Esse cenário dificultou a permanência do ligante no sítio ativo da proteína. No caso da DM2, apesar da interação dos resíduos Tyr59, Trp157 com o ligante, houve ligações de hidrogênio principalmente por parte dos resíduos Ser160 e Ala155 por cerca de 20% da simulação – Asp178 formou ligações de hidrogênio por apenas aproximadamente 11% da simulação. Como Ser160 e Ala155 estão deslocados do sítio ativo, essas interações corroboram o que foi apontado nos resultados de  $d_{HC}$ , assim como o deslocamento do ligante indicado na sequência de imagens na Figura 41, onde o ligante se fixa na proteína fora do sítio ativo.

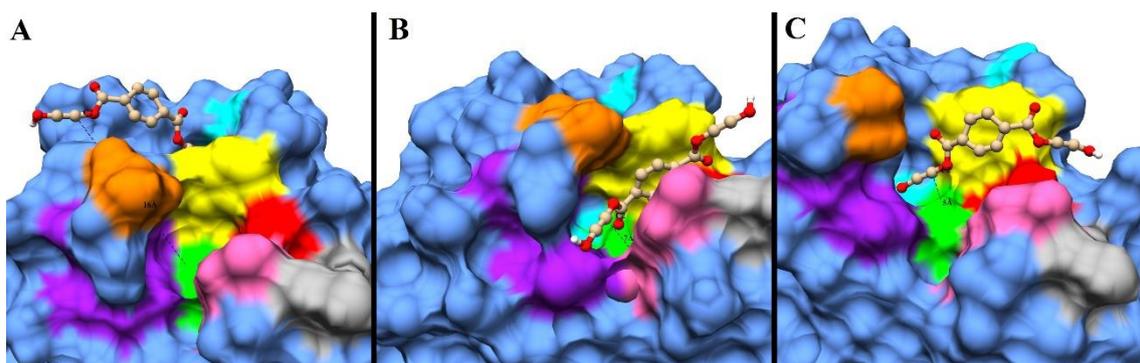


Figura 41: Sequência de imagens que demonstra o deslocamento do ligante na DM2 da variante I complexado com o BHET. O ligante se move e se posiciona no sítio ativo deslocado na direção da Met133 (vermelho). As demais cores representam a superfície da Ser132 (verde), Asp178 e His209 (roxo), Val180 (laranja), Trp157 (amarelo), Ser160 e Ala155 (ciano) e Leu89 (cinza).

O cenário da DM3 favoreceu o ancoramento do ligante devido à presença de todos os resíduos do conjunto  $\lambda$ . Contudo, como o ancoramento apresentou instabilidade (nos primeiros 75 ns). Isso pode ser atribuído à energia de ligação pelo Asp178 ter sido

desfavorável ao ancoramento (bem diferente daquilo que foi exibido nos complexos da proteína nativa) e à baixa quantidade de ligações de hidrogênio – houve ligações predominantemente pelo Asp178, mas que se formaram por menos de 50% da simulação.

Sobre os complexos da variante I com o monômero de PET, as energias dos resíduos nas DM1 e DM2 foram muito pequenas, sendo a causa do desprendimento do ligante junto com as ligações de hidrogênio que se formaram por menos de 2% da simulação. Ao passo que sua DM3, apesar da interação dos resíduos do conjunto  $\lambda$ , suas energias foram 50% mais baixas e as ligações de hidrogênio se formaram 40% menos do que nos casos em que o ligante permaneceu na proteína – como nos complexos da proteína nativa com o BHET.

Nos complexos da variante II com o BHET, além da energia de ligação não ser o suficiente (como demonstrado nos resultados), a falta de ligações de hidrogênio com resíduos importantes (Asp178 ou His209) sugere que o ligante não permaneceu fixo na proteína na DM1. Na verdade, a constatação de ligações de hidrogênio que se formaram menos de 10 vezes durante a simulação (bem abaixo de 1% do tempo) com diferentes resíduos ao longo da proteína indica que o ligante se movimentou ao longo de sua superfície – como exibe a Fig. 42, confirmando o ligante fora do sítio ativo da proteína.

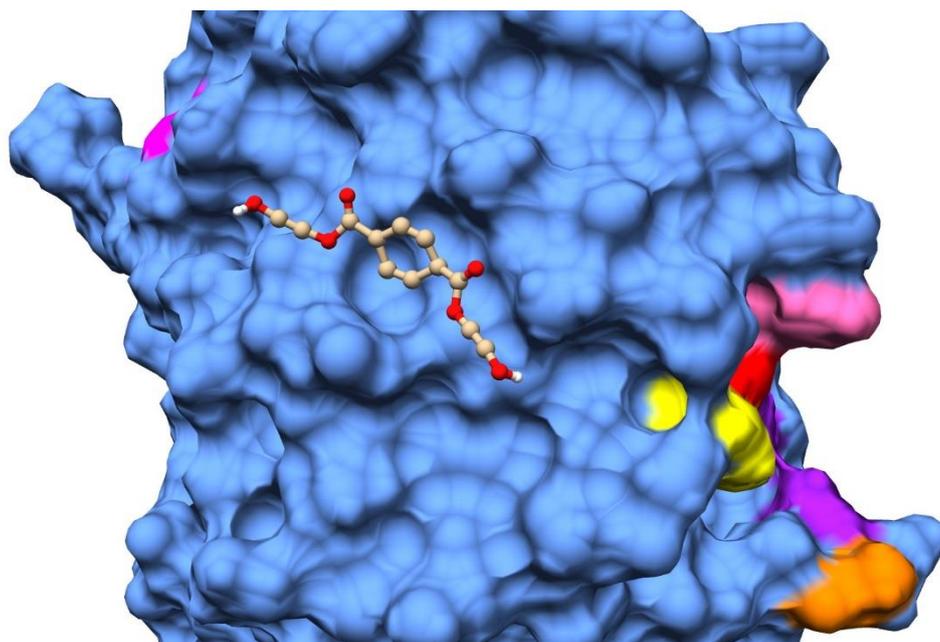


Figura 42: Ligante saiu do sítio ativo e caminhou na superfície da proteína na DM1 da variante II complexada com BHET. O foi em direção ao resíduo Asp122 (destacado em rosa à esquerda) As demais cores representam a superfície da Ser132 (verde), Asp178 e His209 (roxo), Ala180 (laranja), Trp157 (amarelo) e Met133 (vermelho).

Apesar de  $d_{HC}$  médio ter se estabilizado após 100 ns da simulação com o ligante afastado do sítio ativo na DM2, a baixa energia de ligação dos poucos resíduos interagindo com o ligante no ancoramento (próximo de -1 kcal/mol), aliada com a disposição espacial desses resíduos (Ser65, Ser210 e Asn216), podem ser um indicativo de um ancoramento instável que em algum momento adiante se desprenderia da proteína

ou continuaria a se mover pela sua superfície (Fig. 43) como indicam os dados sobre a formação de ligação de hidrogênio, apontando pouquíssimas ligações se formando ao longo da simulação, com permanência abaixo de 15% do tempo da simulação.

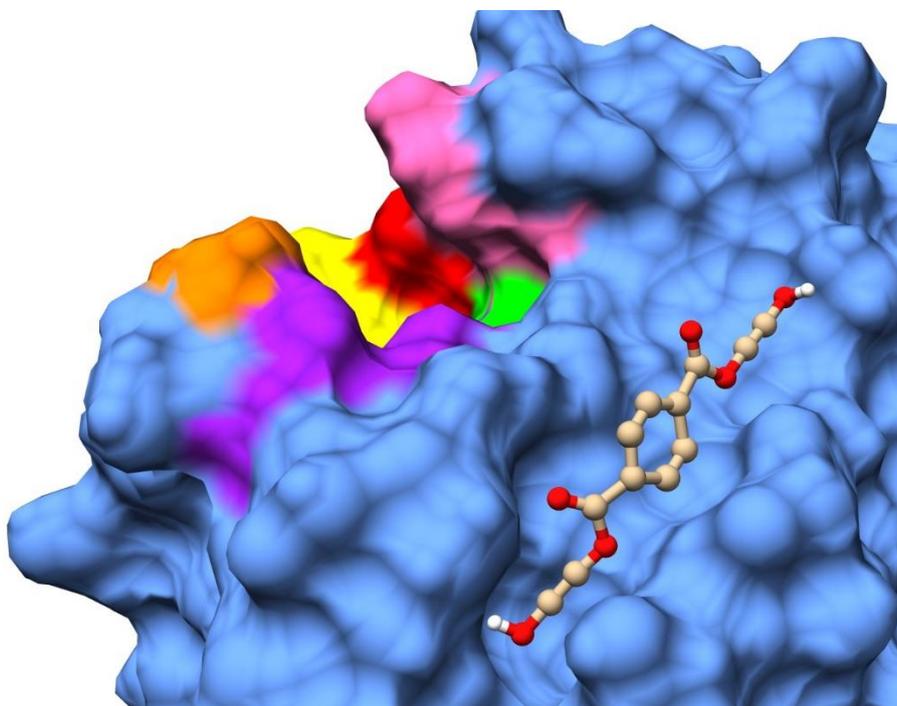


Figura 43: Ligante se movimentou na superfície da proteína e se posicionou fora do seu sítio ativo na DM2 da variante II complexada com BHET. As cores representam a superfície da Ser132 (verde), Asp178 e His209 (roxo), Ala180 (laranja), Trp157 (amarelo) e Met133 (vermelho).

Na DM3, o posicionamento do ligante com  $d_{HC} = 6 \text{ \AA}$  (Fig. 44), apesar da presença do conjunto  $\lambda$  quase que na sua totalidade (com exceção da Ser132), pode ser explicada pela formação insuficiente de ligações de hidrogênio do Asp178, que não ultrapassou 50% da simulação, valor abaixo daquele observado nos complexos da proteína nativa – casos em que o ligante se posicionou de forma correta no sítio ativo da proteína.

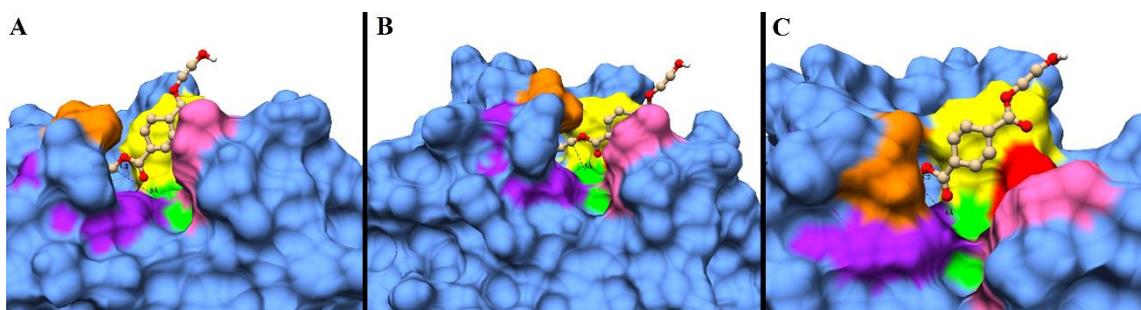


Figura 44: Sequência de imagens que demonstra o deslocamento do ligante na DM3 da variante II complexado com o BHET. O ligante permanece no sítio ativo, mas posicionado no sítio ativo deslocado em relação à Ser132 (verde), com  $d_{HC}$  médio de  $6 \text{ \AA}$ , na direção da Met133 (vermelho). As demais cores representam a superfície da Ser132 (verde), Asp178 e His209 (roxo), Ala180 (laranja) e Trp157 (amarelo).

Nos complexos com da variante II o PET, semelhante ao que foi visto nos complexos da variante I com o PET, foram as energias pouco favoráveis ao ancoramento do ligante (que não ultrapassaram -1 kcal/mol), aliadas às quase inexistentes ligações de hidrogênio ao longo da simulação (que se formaram por menos 1% do tempo), que levaram ao desprendimento do ligante em todas as três simulações realizadas – como apontado pelos resultados de  $d_{HC}$ .

## 6.4 Energia livre de Gibbs

Com o que foi apresentado na seção 5.3.4, observar os valores médios da energia livre de Gibbs fornece informações adicionais sobre as simulações. O valor médio de  $\Delta G_{\text{complexo}}$  para cada conjunto de triplicata foi de: -4934,65 kcal/mol (complexo da proteína nativa), -4944,63 kcal/mol (complexo da variante I) e -4931,72 kcal/mol (complexo da variante II). Com isso, tomando como referência o complexo da proteína nativa, a variação entre os valores médios de  $\Delta G_{\text{complexo}}$  ficou abaixo de 1%. Isso indica que do ponto de vista de energia os complexos (proteína nativa, variante I e variante II) se mantiveram estáveis.

Ao se fazer a média de  $\Delta\Delta G$  (energia livre de Gibbs devido à ligação) para cada uma das triplicatas tem-se: -26,60 kcal/mol (complexo da proteína nativa), -20,74 kcal/mol (complexo da variante I) e -20,69 kcal/mol (complexo da variante II). A comparação entre esses valores, tendo como referência o complexo da proteína nativa, indica que  $\Delta\Delta G$  sofreu uma redução de aproximadamente 22% nos sistemas mutados. Nesse caso, diferente do que foi observado em  $\Delta G_{\text{complexo}}$  médio, o comportamento de  $\Delta\Delta G$  médio indica que as mutações desfavoreceram a formação do complexo.

## 7 Conclusão

Com o objetivo de estudar os efeitos de mutações residuais no sítio ativo da IsPETase e como essas mutações interferem nas interações da proteína com ligantes, foram geradas duas variantes da IsPETase: a variante I, com a mutação que substituiu sua Ile180 por uma Val, e a variante II, com a mutação que substituiu a sua Ile180 por uma Alanina. O método empregado foi bem sucedido em gerar as proteínas mutadas que, junto com a proteína nativa, foram os três modelos de proteína usados para a condução dos estudos. A técnica de *docking* foi efetiva ao fornecer aproximadamente 2700 modelos de conformação de ligante (BHET e PET) para cada modelo de proteína, dos quais o modelo que apresentou o parâmetro  $s_{vd}$  com menor valor foi utilizado para a geração do complexo – seis complexos ao todo, com cada modelo de proteína complexada com BHET e PET. Os sistemas então passaram pela simulação por Dinâmica Molecular e os dados gerados foram analisados.

No estudo dos efeitos estruturais das mutações, a comparação dos resultados da proteína nativa com as variantes indicou uma tendência de aumento nas variações estruturais e conformacionais. Além disso, houve mudanças consideráveis no que diz respeito aos estados de energia acessíveis às proteínas, estados que não são necessariamente favoráveis ao ancoramento do ligante – como em alguns casos em que o mapa de energia livre apresentou algumas regiões metaestáveis inexistentes na proteína nativa ou mesmo na proteína isolada. Ficou demonstrado a diminuição da formação de ligações de hidrogênio (assim como sua permanência e frequência) ao longo da simulação quando se observou as variantes – sistemas mais propensos ao ligante se deslocar pela superfície da proteína. Além disso, os resultados sobre as ligações de hidrogênio estão de acordo com os resultados de decomposição de energia livre de ligação por resíduo. Essa decomposição de energia também indica um comprometimento da interação proteína-ligante devido às mutações, com diminuição na intensidade da interação dos resíduos com o ligante, redução da quantidade de resíduos interagindo com o ligante, vários resíduos fora do sítio ativo interagindo com o ligante, confirmando a movimentação do ligante na superfície da proteína. A redução da permanência da His209 no estado mono protonada em  $N_{\delta}$  foi outro fator que acompanhou a redução na formação de ligações de hidrogênio nos complexos, dado que as mudanças geométricas devido a alternância entre  $N_{\delta}$  e  $N_{\epsilon}$  interferem no posicionamento do ligante no sítio ativo, o que também influencia nos resíduos que interagem com o ligante.

Os valores médios da distância proteína ligante ( $d_{HC}$  definido na seção 2.3) também confirmam a mudança na interação proteína ligante, conforme demonstram mudanças de posição do ligante dentro do sítio ativo, sua movimentação ao longo da superfície (como apontado por outros resultados) e eventualmente seu desprendimento da proteína. Apesar de ser comum o desprendimento do ligante como parte do ciclo de degradação, para o caso em questão, esse ciclo não foi representado por completo, apenas

as etapas de formação do complexo e de sua manutenção – parte do objeto de estudo deste trabalho.

A redução de aproximadamente 22% na energia livre de ligação dos sistemas estudados, também evidencia o desfavorecimento causado pelas mutações na formação dos complexos. Dessa forma, conforme os resultados apresentados e considerando os resultados experimentais conhecidos – Cui, Yinglu, et al., 2021, de Han, Xu, et al., 2017 e Joo, Seongjoon, et al., 2018, demonstrando a diminuição do desempenho da atividade enzimática da *IsPETase* pelas mesmas mutações estudadas aqui –, este trabalho evidencia o comprometimento do processo de degradação da proteína a partir de um ponto de vista teórico sobre processo de formação do complexo proteína-ligante.

A partir do que foi apresentado aqui para contribuir com o estudo da PETase da *Ideonella sakaiensis* e sua eventual aplicação na biodegradação do PET, pode-se propor estudos mais aprofundados sobre o processo de formação do complexo, implementando a técnica de *Quantum mechanics/Molecular Mechanics* (QM/MM) para considerar efeitos quânticos no processo, a implementação das mutações em outros resíduos do sítio ativo para avaliar seus efeitos sobre o processo de ancoramento, ou propor mutações sobre a Ile180 que possam favorecer a formação do complexo.

## 8 Referências

- [1] Maione, M., Mocca, E., Eisfeld, K., Kazepov, Y. and Fuzzi, S. (2021) Public perception of air pollution sources across Europe. *Ambio*, Springer Science and Business Media B.V. **50**, 1150–8. <https://doi.org/10.1007/S13280-020-01450-5/TABLES/2>
- [2] Purchase, D., Abbasi, G., Bisschop, L., Chatterjee, D., Ekberg, C., Ermolin, M. et al. (2020) Global occurrence, chemical properties, and ecological impacts of e-wastes (IUPAC Technical Report). *Pure and Applied Chemistry*, **92**, 1733–67. <https://doi.org/10.1515/pac-2019-0502>
- [3] Völker, C., Kramm, J., Kerber, H., Schramm, E., Winker, M. and Zimmermann, M. (2017) More than a potential hazard—approaching risks from a social-ecological perspective. *Sustainability (Switzerland)*, **9**, 1–16. <https://doi.org/10.3390/su9071039>
- [4] Kawai, F., Kawabata, T. and Oda, M. (2020) Current State and Perspectives Related to the Polyethylene Terephthalate Hydrolases Available for Biorecycling. *ACS Sustainable Chemistry and Engineering*, **8**, 8894–908. <https://doi.org/10.1021/acssuschemeng.0c01638>
- [5] Taniguchi, I., Yoshida, S., Hiraga, K., Miyamoto, K., Kimura, Y. and Oda, K. (2019) Biodegradation of PET: Current Status and Application Aspects. *ACS Catalysis*, American Chemical Society. **9**, 4089–105. [https://doi.org/10.1021/ACSCATAL.8B05171/ASSET/IMAGES/MEDIUM/CS-2018-05171U\\_0012.GIF](https://doi.org/10.1021/ACSCATAL.8B05171/ASSET/IMAGES/MEDIUM/CS-2018-05171U_0012.GIF)
- [6] Hopewell, J., Dvorak, R. and Kosior, E. (2009) Plastics recycling: Challenges and opportunities. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **364**, 2115–26. <https://doi.org/10.1098/rstb.2008.0311>
- [7] Austin, H.P., Allen, M.D., Donohoe, B.S., Rorrer, N.A., Kearns, F.L., Silveira, R.L. et al. (2018) Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proceedings of the National Academy of Sciences of the United States of America*, National Academy of Sciences. **115**, E4350–7. [https://doi.org/10.1073/PNAS.1718804115/SUPPL\\_FILE/PNAS.1718804115.SAPP.PDF](https://doi.org/10.1073/PNAS.1718804115/SUPPL_FILE/PNAS.1718804115.SAPP.PDF)
- [8] Biundo, A., Ribitsch, D. and Guebitz, G.M. (2018) Surface engineering of polyester-degrading enzymes to improve efficiency and tune specificity. *Applied Microbiology and Biotechnology*, Springer Verlag. **102**, 3551–9. <https://doi.org/10.1007/S00253-018-8850-7/METRICS>
- [9] World, O., Day, E. and Communications, N. (2018) The future of plastic. *Nature Communications*, **9**, 2157. <https://doi.org/10.1038/s41467-018-04565-2>
- [10] Kawai, F. (2021) The current state of research on PET hydrolyzing enzymes available for biorecycling. *Catalysts*, **11**, 1–10. <https://doi.org/10.3390/catal11020206>
- [11] O-Reilly, M. and Stubbe, J.A. (2020) PET polymer recycling. *Biochemistry*, **59**, 2316–8. <https://doi.org/10.1021/acs.biochem.0c00457>

- [12] Chen, Z., Wang, Y., Cheng, Y., Wang, X., Tong, S., Yang, H. et al. (2020) Efficient biodegradation of highly crystallized polyethylene terephthalate through cell surface display of bacterial PETase. *Science of the Total Environment*, Elsevier B.V. **709**, 1–9. <https://doi.org/10.1016/j.scitotenv.2019.136138>
- [13] Son, H.F., Joo, S., Seo, H., Sagong, H.Y., Lee, S.H., Hong, H. et al. (2020) Structural bioinformatics-based protein engineering of thermo-stable PETase from *Ideonella sakaiensis*. *Enzyme and Microbial Technology*, Elsevier Inc. **141**, 109656. <https://doi.org/10.1016/j.enzmictec.2020.109656>
- [14] Sagong, H.Y., Seo, H., Kim, T., Son, H.F., Joo, S., Lee, S.H. et al. (2020) Decomposition of the PET Film by MHETase Using Exo-PETase Function. *ACS Catalysis*, American Chemical Society. **10**, 4805–12. [https://doi.org/10.1021/ACSCATAL.9B05604/SUPPL\\_FILE/CS9B05604\\_SI\\_001.PDF](https://doi.org/10.1021/ACSCATAL.9B05604/SUPPL_FILE/CS9B05604_SI_001.PDF)
- [15] Zanela, T., Muniz, E. and Almeida, C. (2018) Chemical Recycling of Poly(ethylene terephthalate) (PET) by Alkaline Hydrolysis and Catalyzed Glycolysis. *Orbital: The Electronic Journal of Chemistry*, **10**. <https://doi.org/10.17807/orbital.v10i3.1104>
- [16] de Castro, A.M., Carniel, A., Nicomedes Junior, J., da Conceição Gomes, A. and Valoni, É. (2017) Screening of commercial enzymes for poly(ethylene terephthalate) (PET) hydrolysis and synergy studies on different substrate sources. *Journal of Industrial Microbiology and Biotechnology*, Springer Berlin Heidelberg. **44**, 835–44. <https://doi.org/10.1007/s10295-017-1942-z>
- [17] Son, H.F., Cho, I.J., Joo, S., Seo, H., Sagong, H.Y., Choi, S.Y. et al. (2019) Rational Protein Engineering of Thermo-Stable PETase from *Ideonella sakaiensis* for Highly Efficient PET Degradation. *ACS Catalysis*, **9**, 3519–26. <https://doi.org/10.1021/acscatal.9b00568>
- [18] Liu, B., He, L., Wang, L., Li, T., Li, C., Liu, H. et al. (2018) Protein crystallography and site-direct mutagenesis analysis of the poly(Ethylene terephthalate) hydrolase petase from *Ideonella sakaiensis*. *ChemBioChem*, **19**, 1471–5. <https://doi.org/10.1002/cbic.201800097>
- [19] Joo, S., Cho, I.J., Seo, H., Son, H.F., Sagong, H.Y., Shin, T.J. et al. (2018) Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nature Communications 2018 9:1*, Nature Publishing Group. **9**, 1–12. <https://doi.org/10.1038/s41467-018-02881-1>
- [20] Vertommen, M.A.M.E., Nierstrasz, V.A., Veer, M. Van Der and Warmoeskerken, M.M.C.G. (2005) Enzymatic surface modification of poly(ethylene terephthalate). *Journal of Biotechnology*, **120**, 376–86. <https://doi.org/10.1016/j.jbiotec.2005.06.015>
- [21] Kawai, F., Kawabata, T. and Oda, M. (2019) Current knowledge on enzymatic PET degradation and its possible application to waste stream management and other fields. *Applied Microbiology and Biotechnology*, Applied Microbiology and Biotechnology. **103**, 4253–68. <https://doi.org/10.1007/s00253-019-09717-y>
- [22] Ribitsch, D., Heumann, S., Trotscha, E., Herrero Acero, E., Greimel, K., Leber, R. et al. (2011) Hydrolysis of polyethyleneterephthalate by p-nitrobenzylesterase

- from *Bacillus subtilis*. *Biotechnology Progress*, **27**, 951–60. <https://doi.org/10.1002/btpr.610>
- [23] Ge, B.K., Hu, G.M. and Chen, C.M. (2021) Plastic bioconversion: Reaction mechanism of PETases. *Chinese Journal of Physics*, Elsevier. **73**, 331–9. <https://doi.org/10.1016/J.CJPH.2021.07.027>
- [24] Joo, S., Cho, I.J., Seo, H., Son, H.F., Sagong, H.Y., Shin, T.J. et al. (2018) Structural insight into molecular mechanism of poly(ethylene terephthalate) degradation. *Nature Communications*, Springer US. **9**. <https://doi.org/10.1038/s41467-018-02881-1>
- [25] Pfaff, L., Gao, J., Li, Z., Jäckering, A., Weber, G., Mican, J. et al. (2022) Multiple Substrate Binding Mode-Guided Engineering of a Thermophilic PET Hydrolase. *ACS Catalysis*, **12**, 9790–800. <https://doi.org/10.1021/acscatal.2c02275>
- [26] Biundo, A., Ribitsch, D. and Guebitz, G.M. (2018) Surface engineering of polyester-degrading enzymes to improve efficiency and tune specificity. *Applied Microbiology and Biotechnology*, Applied Microbiology and Biotechnology. **102**, 3551–9. <https://doi.org/10.1007/s00253-018-8850-7>
- [27] Kan, Y., He, L., Luo, Y. and Bao, R. (2021) IsPETase Is a Novel Biocatalyst for Poly(ethylene terephthalate) (PET) Hydrolysis. *ChemBioChem*, **22**, 1706–16. <https://doi.org/10.1002/cbic.202000767>
- [28] Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y. et al. (2016) A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*, **353**, 759.
- [29] Lewis, T. and Stone, W.L. (2020) Biochemistry, Proteins Enzymes [Internet]. StatPearls. StatPearls Publishing.
- [30] Murray, J.E., Laurieri, N. and Delgoda, R. (2017) Proteins. *Pharmacognosy: Fundamentals, Applications and Strategy*, Elsevier Inc. p. 477–94. <https://doi.org/10.1016/B978-0-12-802104-0.00024-X>
- [31] Robinson, P.K. (2015) Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, **59**, 1–41. <https://doi.org/10.1042/BSE0590001>
- [32] Fecker, T., Galaz-Davison, P., Engelberger, F., Narui, Y., Sotomayor, M., Parra, L.P. et al. (2018) Active Site Flexibility as a Hallmark for Efficient PET Degradation by *I. sakaiensis* PETase. *Biophysical Journal*, **114**, 1302–12. <https://doi.org/10.1016/j.bpj.2018.02.005>
- [33] Han, X., Liu, W., Huang, J.W., Ma, J., Zheng, Y., Ko, T.P. et al. (2017) Structural insight into catalytic mechanism of PET hydrolase. *Nature Communications* 2017 8:1, Nature Publishing Group. **8**, 1–6. <https://doi.org/10.1038/s41467-017-02255-z>
- [34] Burgin, T., Pollard, B.C., Knott, B.C., Mayes, H.B., Crowley, M.F., McGeehan, J.E. et al. (2024) The reaction mechanism of the *Ideonella sakaiensis* PETase enzyme. *Communications Chemistry*, Springer US. **7**. <https://doi.org/10.1038/s42004-024-01154-x>
- [35] Shrimpton-Phoenix, E., Mitchell, J.B.O. and Bühl, M. (2022) Computational

- Insights into the Catalytic Mechanism of Is-PETase: An Enzyme Capable of Degrading Poly(ethylene) Terephthalate. *Chemistry - A European Journal*, **28**. <https://doi.org/10.1002/chem.202201728>
- [36] Vyas, B., Choudhary, S., Singh, P.K., Singh, A., Singh, M., Verma, H. et al. (2018) Molecular dynamics/quantum mechanics guided designing of natural products based prodrugs of Epalrestat. *Journal of Molecular Structure*, Elsevier. **1171**, 556–63. <https://doi.org/10.1016/j.molstruc.2018.06.030>
- [37] Wang, J. and Hou, T. (2011) Application of molecular dynamics simulations in molecular property prediction II: Diffusion coefficient. *Journal of Computational Chemistry*, **32**, 3505–19. <https://doi.org/10.1002/jcc.21939>
- [38] Hernández, E.R., Zetina, L.M.M., Vega, G.T., Rocha, M.G., Ochoa, L.F.R. and Fernandez, R.L. (2008) Molecular Dynamics: from basic techniques to applications (A Molecular Dynamics Primer). 95–123. <https://doi.org/10.1063/1.3040265>
- [39] Hospital, A., Goñi, J.R., Orozco, M. and Gelpí, J.L. (2015) Molecular dynamics simulations: Advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, **8**, 37–47. <https://doi.org/10.2147/AABC.S70333>
- [40] González, M.A. (2011) Force fields and molecular dynamics simulations. *École Thématique de La Société Française de La Neutronique*, **12**, 169–200. <https://doi.org/10.1051/sfn/201112009>
- [41] Amber, C. (2018) Amber 2018 Reference Manual.
- [42] Wang, J., Wang, W., Kollman, P.A. and Case, D.A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, **25**, 247–60. <https://doi.org/10.1016/j.jmngm.2005.12.005>
- [43] Kholmurodov, K., Smith, W., Yasuoka, K., Darden, T. and Ebisuzaki, T. (2000) A Smooth-Particle Mesh Ewald Method for DL\_POLY Molecular Dynamics Simulation Package on the Fujitsu VPP700. *Journal of Computational Chemistry*, **21**, 1187–91. [https://doi.org/10.1002/1096-987X\(200010\)21:13<1187::AID-JCC7>3.0.CO;2-7](https://doi.org/10.1002/1096-987X(200010)21:13<1187::AID-JCC7>3.0.CO;2-7)
- [44] Iupinov, A. (2016) Implementation of the Particle Mesh Ewald method on a GPU.
- [45] Fan, J., Fu, A. and Zhang, L. (2019) Progress in molecular docking. *Quantitative Biology*, **7**, 83–9. <https://doi.org/10.1007/s40484-019-0172-y>
- [46] Morris, G.M. and Lim-Wilby, M. (2008) Molecular docking. *Methods in Molecular Biology*, **443**, 365–82. [https://doi.org/10.1007/978-1-59745-177-2\\_19](https://doi.org/10.1007/978-1-59745-177-2_19)
- [47] Mohanty, M. and Mohanty, P.S. (2023) Molecular docking in organic, inorganic, and hybrid systems: a tutorial review. *Monatshefte Fur Chemie*, Springer Vienna. **154**, 683–707. <https://doi.org/10.1007/s00706-023-03076-1>
- [48] Swails, J.M., York, D.M. and Roitberg, A.E. (2014) Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: Implementation, testing, and validation. *Journal of Chemical Theory and Computation*, **10**, 1341–52. <https://doi.org/10.1021/ct401042b>

- [49] Hopkins, C.W., Le Grand, S., Walker, R.C. and Roitberg, A.E. (2015) Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*, American Chemical Society. **11**, 1864–74. <https://doi.org/10.1021/ct5010406>
- [50] Sevilla, M.E., Garcia, M.D., Perez-Castillo, Y., Armijos-Jaramillo, V., Casado, S., Vizuete, K. et al. (2023) Degradation of PET Bottles by an Engineered *Ideonella sakaiensis* PETase. *Polymers*, **15**, 1–15. <https://doi.org/10.3390/polym15071779>
- [51] Chen, C.C., Han, X., Ko, T.P., Liu, W. and Guo, R.T. (2018) Structural studies reveal the molecular mechanism of PETase. *FEBS Journal*, **285**, 3717–23. <https://doi.org/10.1111/febs.14612>
- [52] Cui, Y., Chen, Y., Liu, X., Dong, S., Tian, Y., Qiao, Y. et al. (2021) Computational Redesign of a PETase for Plastic Biodegradation under Ambient Condition by the GRAPE Strategy. *ACS Catalysis*, **11**, 1340–50. <https://doi.org/10.1021/acscatal.0c05126>
- [53] Han, X., Liu, W., Huang, J.-W., Ma, J., Zheng, Y., Ko, T.-P. et al. (2017) Structural insight into catalytic mechanism of PET hydrolase. *Nature Communications*, Nature Publishing Group UK. **8**, 2106. <https://doi.org/10.1038/s41467-017-02255-z>
- [54] Betts, M.J. and Russell, R.B. (2003) Amino Acid Properties and Consequences of Substitutions. *Bioinformatics for Geneticists*, **4**, 289–316. <https://doi.org/10.1002/0470867302.ch14>
- [55] Boyle, J. (2005) Lehninger principles of biochemistry (4th ed.): Nelson, D., and Cox, M. *Biochemistry and Molecular Biology Education*, John Wiley & Sons, Ltd. **33**, 74–5. <https://doi.org/10.1002/bmb.2005.494033010419>
- [56] Charifson, P.S., Corkery, J.J., Murcko, M.A. and Walters, W.P. (1999) Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *Journal of Medicinal Chemistry*, American Chemical Society. **42**, 5100–9. <https://doi.org/10.1021/JM990352K/ASSET/IMAGES/MEDIUM/JM990352KN00001.GIF>
- [57] Eberhardt, J., Santos-Martins, D., Tillack, A.F. and Forli, S. (2021) AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal of Chemical Information and Modeling*, American Chemical Society. **61**, 3891–8. [https://doi.org/10.1021/ACS.JCIM.1C00203/SUPPL\\_FILE/CI1C00203\\_SI\\_002.ZIP](https://doi.org/10.1021/ACS.JCIM.1C00203/SUPPL_FILE/CI1C00203_SI_002.ZIP)
- [58] Hanwell, M.D., Curtis, D.E., Lonie, D.C., Vandermeersch, T., Zurek, E. and Hutchison, G.R. (2012) Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, **4**, 17. <https://doi.org/10.1186/1758-2946-4-17>
- [59] Schrodinger LLC. (2015) The PyMOL Molecular Graphics System, Version 1.8.
- [60] Wang, J., Wang, W., Kollman, P. a and Case, D. a. (2001) Antechamber, An Accessory Software Package For Molecular Mechanical Calculations. *J Am Chem*

*Soc*, **222**, U403.

- [61] Schauerl, M., Nerenberg, P.S., Jang, H., Wang, L.-P., Bayly, C.I., Mobley, D.L. et al. (2019) Force Field Partial Charges with Restrained Electrostatic Potential 2 (RESP2). *ChemRxiv*, **2**, 1–25.
- [62] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H. et al. (2000) The Protein Data Bank. *Nucleic Acids Research*, **28**, 235–42. <https://doi.org/10.1093/nar/28.1.235>
- [63] Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. and Simmerling, C. (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, American Chemical Society, **11**, 3696–713. <https://doi.org/10.1021/acs.jctc.5b00255>
- [64] Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, American Institute of Physics. **79**, 926–35. <https://doi.org/10.1063/1.445869>
- [65] Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C. (1977) Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, Academic Press. **23**, 327–41. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)
- [66] Meza, J.C. (2010) Steepest descent. *Wiley Interdisciplinary Reviews: Computational Statistics*, John Wiley & Sons, Ltd. **2**, 719–22. <https://doi.org/10.1002/WICS.117>
- [67] Shewchuk, J. (1994) An Introduction to the Conjugate Gradient Method Without the Agonizing Pain.
- [68] Djordjevic, S.S. and Djordjevic, S.S. (2019) Unconstrained Optimization Methods: Conjugate Gradient Methods and Trust-Region Methods. *Applied Mathematics*, IntechOpen. <https://doi.org/10.5772/INTECHOPEN.84374>
- [69] Cerutti, D.S., Duke, R., Freddolino, P.L., Fan, H. and Lybrand, T.P. (2008) A vulnerability in popular molecular dynamics packages concerning Langevin and Andersen dynamics. *Journal of Chemical Theory and Computation*, American Chemical Society, **4**, 1669–80. [https://doi.org/10.1021/CT8002173/SUPPL\\_FILE/CT8002173\\_SI\\_001.PDF](https://doi.org/10.1021/CT8002173/SUPPL_FILE/CT8002173_SI_001.PDF)
- [70] Frenkel, D. and Smit, B. (2001) Understanding Molecular Simulation. 2nd ed. Academic Press, Inc., USA.
- [71] Abraham, M., Alekseenko, A., Bergh, C., Blau, C., Briand, E., Doijade, M. et al. (2023) GROMACS 2023 Manual. <https://doi.org/10.5281/ZENODO.7588711>
- [72] McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X. et al. (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal*, Biophysical Society. **109**, 1528–32. <https://doi.org/10.1016/j.bpj.2015.08.015>
- [73] Kagami, L.P., das Neves, G.M., Timmers, L.F.S.M., Caceres, R.A. and Eifler-

- Lima, V.L. (2020) Geo-Measures: A PyMOL plugin for protein structure ensembles analysis. *Computational Biology and Chemistry*, **87**, 107322. <https://doi.org/10.1016/j.compbiolchem.2020.107322>
- [74] Baidya, A.T.K., Kumar, A., Kumar, R. and Darreh-Shori, T. (2022) Allosteric Binding Sites of A $\beta$  Peptides on the Acetylcholine Synthesizing Enzyme ChAT as Deduced by In Silico Molecular Modeling. *International Journal of Molecular Sciences*, **23**. <https://doi.org/10.3390/ijms23116073>
- [75] Bordoloi, S., Prasad, R., Lakshmi, V.S., Chandramohanadas, R., Natarajan, K. and Nelson-Sathi, S. (2023) Structure-based virtual screening and Molecular Dynamic Simulations identified FDA-approved molecules as potential inhibitors against the surface proteins of H1N1. *BioRxiv*, 2023.12.02.569695.
- [76] Pan, X., Kirsch, Z.J. and Vachet, R.W. (2022) Distinguishing Histidine Tautomers in Proteins Using Covalent Labeling-Mass Spectrometry. *Analytical Chemistry*, **94**, 1003–10. <https://doi.org/10.1021/acs.analchem.1c03902>
- [77] Sudmeier, J.L., Bradshaw, E.M., Coffman Haddad, K.E., Day, R.M., Thalhauser, C.J., Bullock, P.A. et al. (2003) Identification of histidine tautomers in proteins by 2D  $^1\text{H}/^{13}\text{C}$  one-bond correlated NMR. *Journal of the American Chemical Society*, **125**, 8430–1. <https://doi.org/10.1021/ja034072c>
- [78] Nam, Y., Kalathingal, M., Saito, S. and Lee, J.Y. (2020) Tautomeric Effect of Histidine on  $\beta$ -Sheet Formation of Amyloid Beta 1–40: 2D-IR Simulations. *Biophysical Journal*, Biophysical Society. **119**, 831–42. <https://doi.org/10.1016/j.bpj.2020.07.009>
- [79] Malinska, M., Dauter, M., Kowiel, M., Jaskolski, M. and Dauter, Z. (2015) Protonation and geometry of histidine rings. *Acta Crystallographica Section D: Biological Crystallography*, International Union of Crystallography. **71**, 1444–54. <https://doi.org/10.1107/S1399004715007816>
- [80] Lopez, M.J. and Mohiuddin, S.S. (2020) Biochemistry, Essential Amino Acids [Internet]. StatPearls. StatPearls Publishing.
- [81] Berg, J.M., Tymoczko, J.L. and Stryer, L. (2002) Biochemistry. 5th ed. TA - TT -. W.H. Freeman, New York SE -. <https://doi.org/LK> - <https://worldcat.org/title/48055706>
- [82] LaPelusa, A. and Kaushik, R. (2021) Physiology, Proteins [Internet]. StatPearls. StatPearls Publishing.
- [83] Heim, M., Römer, L. and Scheibel, T. (2010) Hierarchical structures made of proteins. The complex architecture of spider webs and their constituent silk proteins. *Chemical Society Reviews*, **39**, 156–64. <https://doi.org/10.1039/b813273a>
- [84] Murray, R.K., Granner, D.K. and Rodwell, V.W. (2006) Harper's Illustrated Biochemistry [Internet]. Lange Medical Books/McGraw-Hill.
- [85] Mathews, C.K. and van Holde, K.E. (1996) Biochemistry Benjamin/Cummings Pub. Redwood.

## 9 Apêndices

### Apêndice A – Sobre a estrutura geral das proteínas

As proteínas são constituídas por aminoácidos, dos quais os 20 mais comuns são  $\alpha$ -aminoácidos, ou seja, possuem um grupo carbonila ( $-COOH$ ) e um grupo amina ( $-NH_3^+$ ) ligados ao mesmo átomo de carbono – carbono- $\alpha$  ( $C_\alpha$ ). Na Fig. 45 apresenta-se um modelo da estrutura geral de um  $\alpha$ -aminoácido [55].

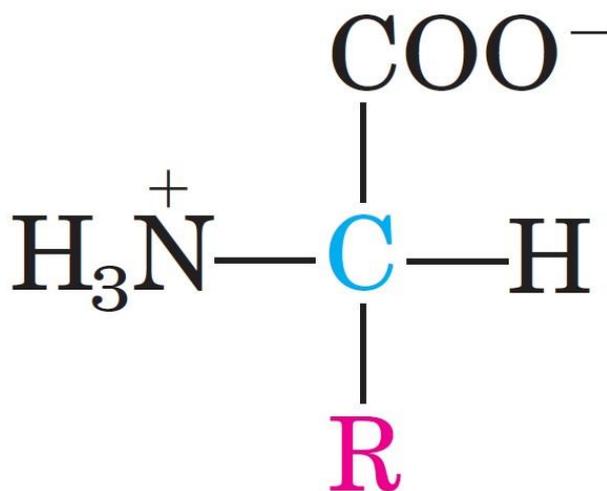


Figura 45: Modelo 2D da estrutura básica de um  $\alpha$ -aminoácido [55].

Quando constituintes de uma proteína, os aminoácidos se unem por ligações peptídicas e passam a ser chamados de resíduos de aminoácidos ou simplesmente resíduos, podendo variar numericamente de pouco menos de 100 até mais de 2000 resíduos. As diferenças nas proteínas são determinadas a partir de seus resíduos e como eles se organizam. Essa organização pode se apresentar em uma ou várias cadeias peptídicas que se enovelam em estruturas tridimensionais específicas determinadas pela sua sequência de resíduos [31,80,81].

As possíveis estruturas podem ser definidas da seguinte maneira: primária, secundária, terciária e quaternária – Fig. 46. A estrutura primária é o nível mais básico de organização, sendo ela a ordem linear em que os resíduos se apresentam na cadeia principal. A estrutura secundária consiste nas possíveis formas adquiridas por essas sequências de resíduos – sustentadas por ligações de hidrogênio –, onde as principais formas são hélices- $\alpha$  e fitas- $\beta$ , das quais um conjunto de fitas- $\beta$  (paralelas e/ou antiparalelas) forma uma estrutura conhecida como folha- $\beta$ . Quando resíduos na estrutura primária podem interagir uns com os outros para formar as estruturas secundárias, possibilitando que resíduos distantes interajam, isso acaba mudando a forma geral da proteína, de modo que o enovelamento a que se chega é a estrutura terciária. Com isso, a estabilidade termodinâmica é alcançada com os resíduos hidrofílicos se posicionando na superfície da proteína e aqueles hidrofóbicos em seu interior. Por último, os complexos

formados por múltiplas cadeias de polipeptídios dão origem à estrutura quaternária, ou seja, é o arranjo tridimensional de dois ou mais polipeptídios em uma proteína [82].

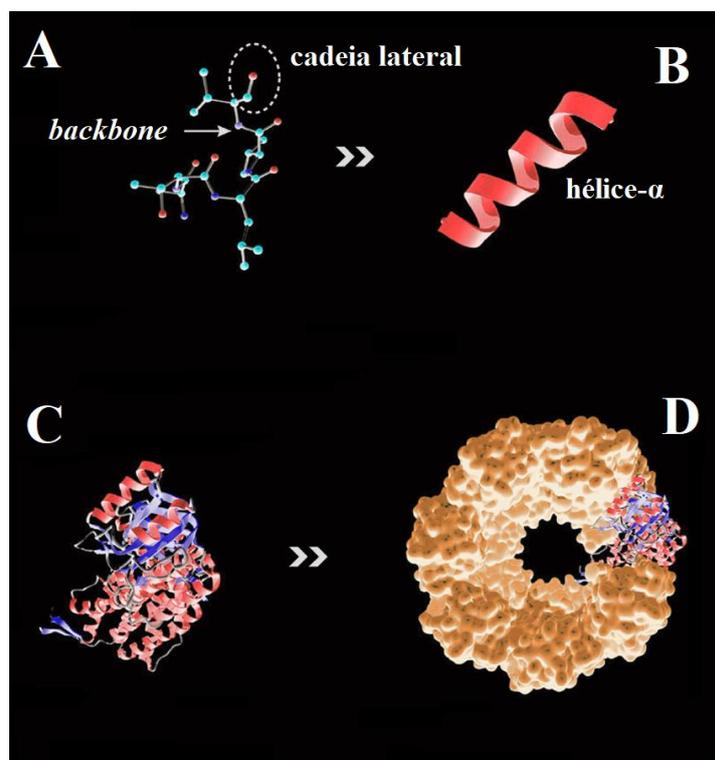


Figura 46: Hierarquia estrutural das proteínas – estrutura primária (A), estrutura secundária (B), estrutura terciária (C) e estrutura quaternária (D). Adaptado de [83].

A forma natural com que uma proteína se enovela (segundo a hierarquia apresentada anteriormente) é conhecida como sua conformação nativa. Essa estrutura é estabilizada por múltiplas interações fracas – como interações hidrofóbicas, ligações de hidrogênio e interações covalentes – e muitas vezes é adotada sem assistência nenhuma, apenas através das propriedades químicas de seus resíduos [55,84]. A estrutura tridimensional da proteína é fundamental na definição de sua função, bastante dependente de sua flexibilidade [30,85].

## Apêndice B – Descrição dos *scripts* em *Python*

Nos processos anteriores à simulação (minimização, aquecimento e equilíbrio) e para a própria simulação (produção) foi utilizado o *ASM.py* para conduzir os processos. O *AMBER Simulation Manager* (*ASM*) automatiza a preparação e a realização da simulação via *DM* com o pacote *Amber/AmberTools*. Com uma única linha de comando é possível definir as condições e parâmetros da simulação (definidos na seção 4.4). Essas informações são armazenadas nos respectivos arquivos de entrada (*min.in*, *annealing.in*, *equilibration.in* e *production.in*) e com a adição da *flag -prepstop* à linha de comando, o sistema (proteína isolada ou complexada) é construído e o processo é interrompido antes do início da minimização. Essa parada permite a edição manual (caso necessário) de cada

um dos arquivos antes da retomada do processo – o que pode ser feito prontamente com um comando *bash*.

Parte dessa linha de comando é dedicada a implementar as mutações que se deseja. A *flag -mut* serve para informar o resíduo padrão a ser substituído seguido do novo resíduo, como por exemplo: *-mut SER\_160\_MET*, caso em que se substitui a Serina na posição 160 por uma Metionina. Na seção 4.2, onde foram apresentadas as mutações implementadas neste trabalho, foram utilizadas as seguintes *flags*: *-mut ILE\_208\_VAL*, substituindo a Isoleucina na posição 180 por uma Valina, para gerar a variante I e *-mut ILE\_208\_ALA*, substituindo a Isoleucina na posição 180 por uma Alanina, para gerar a variante II. Vale lembrar que o *pdb4amber* corrige a contagem dos resíduos da proteína, por isso o comando utiliza a numeração 208 para alterar o resíduo 180. Quanto à análise de dados os seguintes *scripts* foram empregados (todos baseados em comandos presentes no próprio *cpptraj*):

1) **Analysis\_plot.py**: Com os arquivos gerados pelo *cpptraj* esse *script* padroniza a apresentação gráfica dos dados. As análises produzidas são: RMSD, RMSF,  $R_G$ , *prot\_state* (organizando a distribuição dos estados de protonação) e *dist.* (a variação da distância nucleofílica ao longo da simulação). O *script* possibilita a geração de múltiplos gráficos (seja em coluna ou todos juntos) e ainda permite a escolha do tamanho da fonte, idioma, conversão do eixo X de *frames* para *ns*, entre outros.

2) **Prot\_transition.py**: Munido dos dados gerados pelo *cphstats* (ferramenta associada ao *cpptraj*), fornece arquivos com as informações organizadas sobre a distribuição dos estados de protonação dos resíduos – resíduos tituláveis selecionados previamente à DM.

3) **Multidock\_Vna.py**: Semelhante ao *ASM.py*, gerencia o uso do AutoDock Vina na geração de diversas conformações do ligante com base nas configurações definidas previamente ao processo como área de busca, número de execuções da rotina etc.

4) **AvR\_Docking.py**: Baseado nos diversos modelos conformacionais do ligante gerados com o *Multidock\_Vna.py*, esse *script* gera um gráfico da “afinidade de energia vs distância” – no caso a distância nucleofílica definida na seção 2.3 – de todos os modelos gerados. Ao mesmo tempo, no terminal, o *script* imprime os melhores modelos em ordem decrescente de afinidade, além de salvar essas informações em um arquivo de texto.

5) **decode\_mmpbsa.py**: Em posse dos arquivos gerados com o MMPBSA (Apêndice C), esse *script* gera os gráficos da decomposição de energia de ligação (por resíduo). É possível gerar os gráficos separadamente ou em colunas, escolher a quantidade de resíduos destacados no gráfico, definir a numeração dos resíduos (baseado na numeração do PDB ou na correção do *pdb4amber*), entre outros.

## Apêndice C – Decomposição de energia com MMPBSA

Para a obtenção das energias de ligação foi utilizado o *script* em Python *MMPBSA.py*. É uma técnica de pós-processamento para calcular a variação de energia livre entre dois estados. Geralmente emprega uma aproximação onde o espaço de configurações explorado é bastante similar tanto para o estado isolado quanto para o complexado. Nesse método, o solvente explícito e os íons são removidos da trajetória para acelerar a convergência, evitando que termos de interações solvente-solvente prevaleçam.

Em uma simulação realizada com solvente explícito (como definida na seção 4.4), o *MMPBSA* requer um arquivo de topologia do complexo solvatado, do complexo sem solvente e dos arquivos de topologia individuais da proteína e do ligante (4 arquivos no total). A ferramenta *ante-MMPBSA* (outro *script* em *Python*) permite a geração desses arquivos de topologia – Fig. 47. Em posse desses arquivos, o *script* para dar início ao processo é gerado. Esse arquivo de entrada foi desenvolvido para ser o mais familiar possível a outros programas do AMBER. Entre os nomes permitidos nos cabeçalhos dos arquivos de entrada (&general, &gb, &pb, e &decomp, para citar alguns), foi escolhido o &general para elaborar o *script* utilizado neste trabalho e &gb para que os cálculos fossem feitos com base no modelo Generalizado de Born (GB) – consultar manual do AMBER 18 para maiores detalhes [41].

### Input file for running GB

#### &general

use\_sander=1

strip\_mask=:WAT:Na+:Cl-,

startframe=1, endframe=1000, interval=2, verbose=2,

entropy=0, keep\_files=0, ligand\_mask=":266", receptor\_mask=":1-265",

debug\_printlevel=5,

/

#### &gb

igb=8, saltcon=0.1

/

#### &decomp

idecomp=1, dec\_verbose=1,

/

Figura 47: Representação do *script* para a geração dos arquivos base para a obtenção dos dados sobre a decomposição de energia.

O *script* foi feito considerando uma análise sobre parte dos *frames* da trajetória (startframe=701, endframe=1000) e desconsiderando 1 *frame* a cada 2 (interval=2), como o intuito de agilizar o processo sem grandes perdas de informação. Foram definidos os arquivos de topologia do ligante (ligand\_mask=266) e da proteína (receptor\_mask=1-265). Também foi indicada a remoção das moléculas de água e dos íons (strip\_mask=:WAT:Na+:Cl-) na etapa de geração do arquivo de topologia do complexo – tudo isso sem levar em conta qualquer termo relacionado à entropia (entropy=0). Foi utilizado o método generalizado de Born (igb=8) – para mais detalhes consultar o manual do AMBER18 [41] – e concentração molar de sal predefinida (saltcon=0.1). O esquema de decomposição de energia escolhido foi por resíduo, onde os termos 1-4 (referentes a átomos separados por três ligações covalentes consecutivas) foram adicionados aos termos de energia potencial interna (idecomp=1). O parâmetro dec\_verbose=1 permitiu o registro de informações relacionadas à variação de energia total, cadeia lateral e cadeia principal. Com os arquivos gerados, o cálculo da decomposição de energia pode ser iniciado – neste trabalho foi empregado a versão MPI.

## Apêndice D – Conversão de arquivos para produção dos mapas de energia livre no PyMOL

Como descrito na seção 4.6 da metodologia, para a geração dos gráficos de energia livre (FEL) arquivos do formato AMBER foram convertidos para formato GROMACS por questões de compatibilidade com o Geo-Measures, *plugin* do PyMOL. O processo de conversão foi realizado com o auxílio de dois *scripts* em Python: o *convert.py* e *mdconverter* (parte do MDTraj). O primeiro *script* usa o ParmEd para realizar a conversão (por isso nomeado dessa forma) dos arquivos .prmtop e .rst7 para o formato .top e .gro, respectivamente – Fig. 48.

```
import parmed as pmd  
  
amber = pmd.load_file('.prmtop', '.rst7')  
  
amber.save('.top')  
  
amber.save('.gro')
```

Figura 48: Modelo de *script* para conversão de arquivos de topologia e coordenadas.

Para a conversão do arquivo de trajetória foi usado o *mdconverter*, ferramenta que acompanha o pacote MDTraj [72]. Esse *script* realiza a conversão com o seguinte comando: **mdconvert -o .xtc -c nframes .ntc**, em que “-o .xtc” é o arquivo de saída (convertido no formato do GROMACS), “-c nframes” indica a quantidade de frames que se deseja converter (para este trabalho nframes=1000, ou seja, a trajetória inteira) e “.ntc” representa o arquivo de trajetória do AMBER que será convertido.