

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO  
INSTITUTO DE QUÍMICA  
CURSO DE QUÍMICA COM ATRIBUIÇÕES TECNOLÓGICAS

ANA BEATRIZ BARROSO LIMA

DINÂMICA MOLECULAR DA DESCARBOXILASE DE ÁCIDOS GRAXOS OLET:  
AS INTERAÇÕES COM ALCENOS E ÁCIDOS GRAXOS HIDROXILADOS

RIO DE JANEIRO  
2023

ANA BEATRIZ BARROSO LIMA

DINÂMICA MOLECULAR DA DESCARBOXILASE DE ÁCIDOS GRAXOS OLET:  
AS INTERAÇÕES COM ALCENOS E ÁCIDOS GRAXOS HIDROXILADOS

Trabalho de conclusão de curso de graduação apresentado ao Departamento de Físico-Química da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Químico com Atribuições Tecnológicas

Orientador: Prof. Rodrigo Leandro Silveira

RIO DE JANEIRO

2023

## CIP - Catalogação na Publicação

L732d Lima, Ana Beatriz Barroso  
Dinâmica molecular da descarboxilase de ácidos graxos OleT: as interações com alcenos e ácidos graxos hidroxilados / Ana Beatriz Barroso Lima. -- Rio de Janeiro, 2023.  
109 f.

Orientador: Rodrigo Leandro Silveira.  
Trabalho de conclusão de curso (graduação) - Universidade Federal do Rio de Janeiro, Instituto de Química, Bacharel em Química, 2023.

1. Dinâmica Molecular. 2. Simulação Computacional. 3. Modelagem Molecular. 4. Química Computacional. 5. Química Teórica. I. Silveira, Rodrigo Leandro, orient. II. Título.

ANA BEATRIZ BARROSO LIMA

DINÂMICA MOLECULAR DA DESCARBOXILASE DE ÁCIDOS GRAXOS OLET:  
AS INTERAÇÕES COM ALCENOS E ÁCIDOS GRAXOS HIDROXILADOS

Trabalho de conclusão de curso de graduação apresentado ao Departamento de Físico-Química da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Químico com Atribuições Tecnológicas

Aprovado em 11 de J u l h o de 2 0 2 3

BANCA EXAMINADORA:



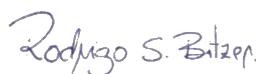
---

Prof. Rodrigo Leandro Silveira  
Universidade Federal do Rio de Janeiro



---

Prof. Ricardo Rodrigues de Oliveira Junior  
Universidade Federal do Rio de Janeiro



---

Prof. Rodrigo da Silva Bitzer  
Universidade Federal do Rio de Janeiro

Dedico este trabalho ao meu pai Emmanoel (*In Memoriam*). *In due time, the distance  
between us will soon quench.*

## AGRADECIMENTOS

Primeiramente, agradeço calorosamente minha família, amigos e colegas que sempre me apoiaram ao longo de minha jornada acadêmica e que foram um alicerce fundamental para que eu completasse esse trabalho. Ao meu querido Rafael, minha mais sincera gratidão pela paciência em perscrutar e corrigir meu texto, reconheço a tarefa torturante que lhe foi incumbida. Pretensiosismo, ou não, reconheço a mim mesma. Pois, de fato houve dúvidas, medos e muitos empecilhos ao longo de minha trajetória, pequenos ou não, mas que se apresentavam colossais diante de mim. Mas, de forma alguma estive sozinha ou desamparada, logo mais gratulações são devidas aos já citados. E agradeço ao Prof. Rodrigo Silveira pela orientação e instrução desse trabalho e pela parceria frutífera que tivemos trabalhando juntos.

Por fim, agradeço ao Laboratório Nacional de Computação Científica (LNCC/MCTI, Brasil) por fornecer recursos HPC do supercomputador SDumont, que contribuíram para os resultados da pesquisa relatados neste trabalho URL: <http://sdumont.lncc.br>.

*“O mais importante da vida não é a situação em que estamos, mas a direção para a qual  
nos movemos”*

**Oliver Wendell Holmes**

## RESUMO

A produção de hidrocarbonetos a partir de fontes vegetais abundantes obtida com a descarboxilação de ácidos graxos por vias biológicas, tem o potencial de remediar a produção de combustíveis de origem fóssil por uma fonte de energia mais sustentável e renovável. A enzima OleT<sub>JE</sub> descoberta por RUDE *et al.* (2011) é capaz de converter ácidos graxos com diferentes tamanhos de cadeia em alcenos terminais, que podem ser utilizados para produção de biocombustíveis ou nas indústrias petroquímicas produzindo polímeros e lubrificantes. Entretanto, a conversão de ácidos graxos também vem acompanhada com a produção indesejada de ácidos  $\alpha$  e  $\beta$ -hidroxilados. Estudos anteriores, mostram que ácidos graxos com cadeias menores, variando entre 10 e 14 carbonos apresentavam maiores atividades enzimáticas, mas acompanhadas por uma perda na seletividade, enquanto, os ácidos de cadeias maiores, entre 16 e 20 carbonos, mostravam menores taxas de conversão, mas com maior seletividade para alcenos (MATTHEWS *et al.*, 2017). Entretanto, os fatores moleculares que regulam esse comportamento catalítico na presença de substratos de diferentes cadeias ainda não eram conhecidos. Uma região da enzima localizada longe do sítio de ligação chamado “loop” F/G foi sugerido por Amaya *et al.* (2018) como responsável por mediar a saída do produto da cavidade de ligação e que a sua dinâmica poderia influenciar a seletividade e a regioseletividade da enzima, mas sem apresentar o mecanismo ou dados experimentais que corroborassem tais especulações. Nesse trabalho foram realizadas simulações de dinâmica molecular da enzima OleT<sub>JE</sub> na presença dos dois produtos majoritariamente formados: alcenos e ácidos  $\beta$ -hidroxilados, variando o tamanho de suas cadeias alifáticas. As dinâmicas comprovaram a relação da flexibilidade da alça F/G com as atividades enzimáticas exibidas pela enzima e sugerem que ela possa influenciar a regioseletividade, mas não são determinantes para reger a seletividade da enzima. Obteve-se que o resultado catalítico é derivado da mobilidade do substrato dentro da cavidade de ligação, e que a saída do produto é mediada pela solvatação do canal de reação. Esse trabalho, portanto, elucidou o comportamento molecular da enzima e amplia o entendimento necessário para o desenvolvimento de novas mutantes mais eficientes.

**Palavras-chave:** dinâmica molecular. simulação computacional. modelagem molecular. química computacional. química teórica.

## ABSTRACT

The production of hydrocarbons from abundant plant sources obtained with the decarboxylation of fatty acids by biological routes, has the potential to remedy the production of fossil fuels by a more sustainable and renewable source of energy. The OleT<sub>JE</sub> enzyme discovered by RUDE *et al.* (2011) is capable of converting fatty acids with different size chains into terminal alkenes, which can be used for the production of biofuels or in the petrochemical industries producing polymers and lubricants. However, the conversion of fatty acids also comes with the undesired output of  $\alpha$  and  $\beta$ -hydroxy acids. Prior studies have shown that fatty acids with shorter chains, varying between 10 and 14 carbons, had higher enzymatic activities, but accompanied by a loss in selectivity, while acids with longer chains, between 16 and 20 carbons, showed lower conversion rates, but with greater selectivity for alkenes (MATTHEWS *et al.*, 2017). However, the molecular factors that regulate this catalytic behavior in the presence of substrates with different chains were still unknown. A region in the enzyme located far from the binding site called the F/G loop was suggested by Amaya *et al.* (2018) as responsible for mediating the output of the product from the binding cavity and that its dynamics could influence the selectivity and regioselectivity of the enzyme, but without presenting the mechanism or experimental data that would corroborate such speculations. In this paper, molecular dynamics simulations of the OleT<sub>JE</sub> enzyme were performed in the presence of the two main products formed: alkenes and  $\beta$ -hydroxylated acids, varying the size of their aliphatic chains. The dynamics proved the relationship between the flexibility of the F/G loop and the enzymatic activities exhibited by the enzyme and suggest that it may influence the regioselectivity, although they are not determinant to govern the selectivity of the enzyme. It was found that the catalytic result is derived from the mobility of the substrate within the binding cavity, and that the output of the product is mediated by the solvation of the reaction channel. This work, therefore, clarified the molecular behavior of the enzyme and broadens the understanding needed for the development of new and more efficient mutants.

**Keywords:** molecular dynamics. computational simulation. molecular modeling. computational chemistry. theoretical chemistry.

## LISTA DE TABELAS

Tabela 1 – Rendimentos na conversão de ácidos graxos, %: . . . . .	72
Tabela 2 – Rendimentos das enzimas OleTP <sub>RN</sub> e OleT <sub>JE</sub> na conversão de ácidos graxos, %: . . . . .	75
Tabela 3 – Seletividade enzimática para alcenos, %: . . . . .	77

## LISTA DE ABREVIATURAS E SIGLAS

ABO	Aproximação de Born-Oppenheimer
AMBER	<i>Assisted Model Building and Energy Refinement</i>
Å	Medida de comprimento Ångström
$c$	Velocidade da luz no vácuo
DCC	Diciclohexilcarbodiimida
ESP	<i>Electrostatic Potential</i> - Potencial Eletrostático
fs	Femtosegundos
$h$	Constante de Planck
K	Escala Kelvin de temperatura
$k_b$	Constante de Boltzmann
$\lambda$	Comprimento de onda térmico de de Broglie
MANIAC	<i>Mathematical Analyzer, Numerical Integrator And Computer</i>
MD	Dinâmica Molecular ( <i>Molecular Dynamics</i> )
ns	Nanosegundos
PDB	<i>Protein data base</i> - Base de dados de proteínas
PME	<i>Particle mesh Ewald</i>
ps	Picossegundos
RESP	<i>Restrained Electrostatic Potential</i> - Potencial Eletrostático Restrito
RMSD	<i>Root mean square deviation</i> - Raiz quadrada do deslocamento quadrático médio
RMSF	<i>Root mean square fluctuation</i> - Raiz quadrada da flutuação média quadrática
SEP	Superfície de energia potencial

## LISTA DE SÍMBOLOS

$\alpha$	Letra grega Alfa
$\beta$	Letra grega Beta
$\varepsilon$	Letra grega Épsilon
$\Gamma$	Letra grega Gama maiúscula
$\Lambda$	Letra grega Lambda maiúscula
$\tau$	Letra grega Tau
$\Omega$	Letra grega Ômega maiúscula
$\omega$	Letra grega Ômega
$\phi$	Letra grega Phi
$\psi$	Letra grega Psi

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO . . . . .</b>	<b>13</b>
1.1	ORGANIZAÇÃO DA MONOGRAFIA . . . . .	17
<b>2</b>	<b>PROTEÍNAS: ESTRUTURAS E DINÂMICA . . . . .</b>	<b>18</b>
2.1	PROTEÍNAS E SUAS ESTRUTURAS . . . . .	18
2.2	DINÂMICA DE PROTEÍNAS . . . . .	22
2.2.1	Flexibilidade Conformacional . . . . .	22
2.2.2	Escala de tempo da dinâmica de proteínas . . . . .	24
<b>3</b>	<b>A DESCARBOXILASE OLET<sub>JE</sub> . . . . .</b>	<b>26</b>
<b>4</b>	<b>DINÂMICA MOLECULAR CLÁSSICA . . . . .</b>	<b>35</b>
4.1	MECÂNICA ESTATÍSTICA . . . . .	35
4.2	APROXIMAÇÃO DE BORN-OPPENHEIMER (ABO) . . . . .	38
4.3	ALGORITMO DE INTEGRAÇÃO: <i>VELOCITY-VERLET</i> . . . . .	40
4.4	CONTROLE DAS SIMULAÇÕES . . . . .	44
4.4.1	Termostato de Langevin . . . . .	44
4.4.2	Barostato de Berendsen . . . . .	45
<b>5</b>	<b>CAMPO DE FORÇA . . . . .</b>	<b>47</b>
5.1	POTENCIAL DE INTERAÇÃO MOLECULAR . . . . .	47
5.2	PARAMETRIZAÇÃO E TRANSFERIBILIDADE . . . . .	53
<b>6</b>	<b>SIMULAÇÃO COMPUTACIONAL DE DINÂMICA MOLE- CULAR . . . . .</b>	<b>55</b>
6.1	CONDIÇÕES DE CONTORNO PERIÓDICAS . . . . .	55
6.1.1	Interações de curto alcance: convenção de imagem mínima . . . . .	56
6.1.2	Interações de longo alcance: somas de Ewald . . . . .	57
6.2	MÉTODO DE INCLUSÃO DE SOLVENTE EXPLÍCITO . . . . .	59
6.3	ALGORITMO “ <i>RATTLE</i> ” . . . . .	60
6.4	MINIMIZAÇÃO E EQUILIBRAÇÃO . . . . .	61
<b>7</b>	<b>OBJETIVOS E METODOLOGIA . . . . .</b>	<b>63</b>
7.1	OBJETIVOS . . . . .	63
7.2	METODOLOGIA . . . . .	64
7.2.1	Configuração inicial do sistema . . . . .	64

7.2.2	Configuração das simulações . . . . .	65
8	RESULTADOS E DISCUSSÕES . . . . .	67
8.1	ANÁLISE DE CONTROLE DAS SIMULAÇÕES . . . . .	67
8.2	ANÁLISES DE FLUTUAÇÕES ESTRUTURAIS . . . . .	69
8.2.1	RMSD (“ <i>Root Mean Square Deviation</i> ”) - Raiz quadrada do deslocamento quadrático médio . . . . .	69
8.2.1.1	RMSD da alça F/G . . . . .	71
8.2.1.2	RMSD dos produtos dentro da cavidade da enzima selvagem OleT <sub>JE</sub> . . . . .	76
8.2.2	RMSF (“ <i>Root mean squared fluctuation</i> ”) - Raiz quadrada da flutuação média quadrática . . . . .	78
8.2.2.1	RMSF por resíduo . . . . .	80
8.2.2.2	RMSF por átomo dos produtos . . . . .	84
8.3	ANÁLISE DE SOLVATAÇÃO DOS LIGANTES . . . . .	85
9	CONCLUSÃO . . . . .	94
	REFERÊNCIAS . . . . .	96
	GLOSSÁRIO . . . . .	102
	ANEXO A – TABELA DOS AMINOÁCIDOS QUE COM- PÕEM AS PROTEÍNAS E SUAS SIGLAS . . . . .	105
	ANEXO B – TABELA PERIÓDICA DE ELEMENTOS QUÍ- MICOS . . . . .	107

## 1 INTRODUÇÃO

Devido à grande demanda global por energia, e a uma maior preocupação da sociedade com políticas de proteção ambiental, em resposta ao aquecimento global e à possibilidade que as mudanças climáticas atinjam um estado de irreversibilidade, desencadeou-se uma busca mundial por fontes de energias alternativas (às de origem fóssil), que sejam renováveis e sustentáveis (**ARFIN; MOHAMMAD; YUSOF, 2014**). Como opção para suprir essa demanda energética, tem-se a Bioenergia, que é toda energia derivada da biomassa, que, por sua vez, é definida como qualquer material biológico recentemente vivo, podendo ser tanto de origem vegetal ou animal (esterco) (**COLEMAN; STANTURF, 2006**). Por ter origem primariamente em processos de fotossíntese, a biomassa é considerada uma fonte renovável de energia, podendo ser renovada de forma potencialmente ilimitada no curto prazo. Em contraste, carvão e combustíveis fósseis, que são derivados de matéria orgânica e mineral, por decorrerem de processos geológicos que levam milhares de anos para se formar, não são considerados oriundos de matéria viva recente e, portanto, não constituem uma fonte renovável de energia (**WILLIAMS; DAHIYA; PORTER, 2020**). Já os biocombustíveis são as formas sólidas, líquidas ou gasosas obtidas do processamento da biomassa, podendo ser obtidos de culturas alimentícias ou não alimentícias e até de algas (**LEE; LAVOIE, 2013**). Quando originadas a partir de alimentos como milho, cana-de-açúcar e óleo de soja, as matérias-primas bioenergéticas derivadas dessas biomassas são majoritariamente compostas, respectivamente, por amido, açúcares e lipídeos (**GOH; LEE, 2010**).

Os lipídeos são grupos de biomoléculas de estruturas químicas diversas que apresentam como característica central a insolubilidade em água, além de serem uma das formas mais comuns de armazenamento de energia utilizada pelos organismos vivos, estando nas células eucarióticas, dispersos no citosol aquoso na forma de triglicerídeos (gorduras neutras). Sendo, os ácidos graxos — compostos formados por uma cadeia alifática, com comprimento variando entre 4 e 36 carbonos que possuem em uma de suas extremidades uma carboxila — um dos principais componentes precursores dessas gorduras, quando estes se encontram unidos a uma mesma molécula de glicerol por meio de uma ligação éster (**NELSON; COX, 2014**). Os ácidos graxos são derivados de hidrocarbonetos saturados com comprimentos de cadeia semelhantes aos hidrocarbonetos de combustíveis líquidos derivados do petróleo utilizados em veículos, que variam entre C5 a C20 (**ZHOU; KERKHOVEN; NIELSEN, 2018**).

Na última década foram desenvolvidos diferentes métodos para a obtenção de hidrocarbonetos que pudessem suprir a demanda energética vigente, a partir de fontes renováveis e sustentáveis por meio de vias biológicas — síntese enzimática e/ou microbiológica (**SCHIRMER *et al.*, 2010**). **WANG e ZHU (2018)** apontam uma vantagem econô-

mica na utilização de hidrocarbonetos renováveis (bio-hidrocarbonetos), que por serem quimicamente equivalentes àqueles derivados do petróleo — biocombustíveis “*drop-in*” —, poderiam ser processados nas mesmas plantas industriais produtoras de combustíveis e petroquímicos já existentes, não sendo necessários novos investimentos dessa natureza. E, **CASPETA e NIELSEN (2013)** apontam que a produção industrial de bio-hidrocarbonetos por fontes microbianas apresenta um balanço energético líquido melhor e com reduções substanciais na emissão de gases de efeito estufa, quando comparada com outras fontes de biocombustíveis.

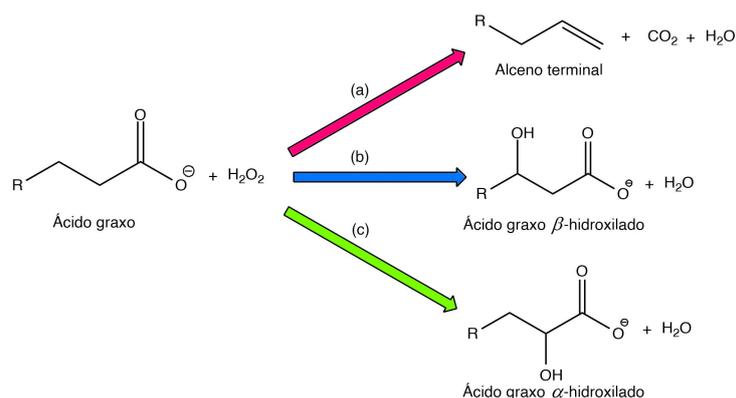
Os biocombustíveis de hidrocarbonetos produzidos por micro-organismos, não possuem desvantagens, quando comparados a outras fontes renováveis, como o bioetanol, que além de ter uma densidade energética menor que os hidrocarbonetos é altamente higroscópico e volátil; ou o biodiesel, obtido pela reação de esterificação de gorduras e óleos vegetais, que resulta em moléculas oxigenadas. Onde, a desvantagem decorre do fato que o grau de oxigenação de uma molécula está diretamente relacionada a sua densidade energética. Além disso, o uso direto do biodiesel como biocombustível veicular é desaconselhável, pois a presença de oxigênio nessas moléculas pode causar danos ao motor e a peças veiculares, sendo necessário previamente a sua remoção. Esse procedimento de desoxigenação ocorre por vias termoquímicas, envolvendo o uso de altas concentrações de hidrogênio ( $H_2$ ) — um composto altamente inflamável —, altas temperaturas e o uso de metais tóxicos, que a depender do composto utilizado e método empregado pode tornar o processo nocivo ao meio ambiente (**SCRUTTON, 2017**).

Estudos conduzidos por **RUDE et al. (2011)** identificaram e caracterizaram em um grupo de bactérias *Jeotgalicoccus* sp., uma enzima capaz de sintetizar olefinas terminais a partir de ácidos graxos, batizada como *Jeotgalicoccus* sp. OleT (OleT<sub>JE</sub>). Tornando possível a obtenção de bio-hidrocarbonetos a partir de óleos e gorduras vegetais. Além disso, esses alcenos terminais podem ser utilizados na indústria petroquímica na produção de lubrificantes, detergentes e polímeros, já que a ligação dupla terminal desses hidrocarbonetos é facilmente funcionalizável (**COOMBS; MORKEN, 2016**) (**WANG; ZHU, 2018**). Dessa forma, a utilização de ácidos graxos por vias biológicas como fonte de energia revela uma rota tangível para a substituição da matriz fóssil em favor de uma fonte renovável e sustentável. Onde essa via metabólica se apresenta, atualmente, como uma alternativa bioenergética potencialmente mais vantajosa, pois permitiria inclusive a produção de combustível de aviação.

Entretanto, a síntese de biocombustível pela OleT<sub>JE</sub> é acompanhada da formação de produtos indesejáveis, como exemplificado na **Figura 1.1**, pois a enzima também é capaz de gerar ácidos graxos  $\beta$ - e até  $\alpha$ -hidroxilados, este último em menor escala (**GRANT; HSIEH; MAKRIS, 2015**), havendo correlações entre o tamanho da cadeia do substrato com a seletividade e rendimento dos produtos. Substratos de cadeias longas, contendo 18 a 20 carbonos, apresentam maior seletividade para alcenos, mas com rendimentos inferiores

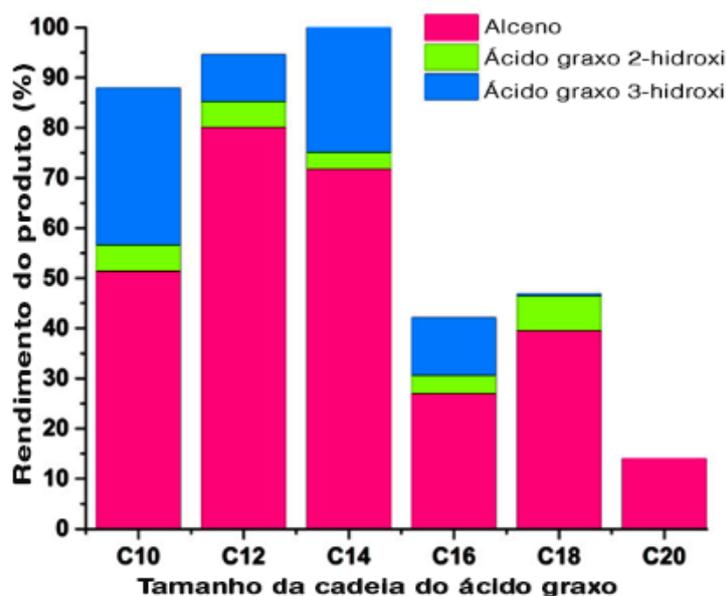
a 50%, ao passo que, para cadeias curtas, contendo de 10 a 14 carbonos, a porcentagem de conversão fica em torno de 90 a 100%, porém, associada a uma perda de seletividade para alcenos (MATTHEWS *et al.*, 2017), sendo esse dados mostrados na **Figura 1.2**. Entretanto, os fatores que regulam esse comportamento enzimático, em relação a maiores atividades estarem associadas a substratos de cadeias menores, e de maior seletividade quando o substrato apresenta cadeias longas, ainda não foram esclarecidos.

Figura 1.1 – Produtos gerados pela ação enzimática OleT<sub>JE</sub> em ácidos graxos



As reações (não balanceadas) catalisadas pela OleT<sub>JE</sub> gera alcenos terminais, mas a enzima também é capaz de hidroxilar o ácido graxo nas posições  $\alpha$  e  $\beta$ . Fonte: Figura elaborada pela autora

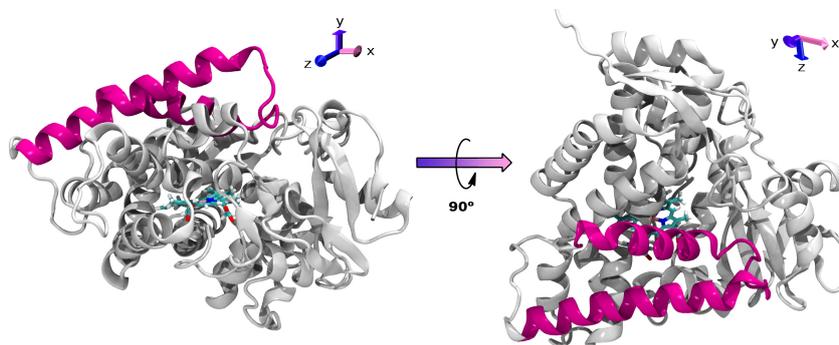
Figura 1.2 – Atividade e seletividade da OleT<sub>JE</sub>



Atividade da OleT<sub>JE</sub>, em termos da porcentagem convertida de substrato e proporção de alcenos e ácidos graxos  $\alpha$ - (ou 2-hidroxi) e  $\beta$ -hidroxilados (ou 3-hidroxi) formados. Fonte: Figura adaptada de Matthews *et al.* (2017).

Estudos realizados por **Amaya *et al.* (2018)**, baseados em fluorescência e FRET, obtiveram que a etapa de dissociação dos produtos pela enzima é a etapa lenta de todo o ciclo catalítico da OleT<sub>JE</sub>, e sugeriram que a regulação dos produtos gerados é influenciada pela mobilidade do substrato dentro do sítio ativo; que a seletividade exibida pela enzima seria mediada por um resíduo localizando longe do sítio ativo, na região do “loop” que conecta as hélices F e G, sendo mostrado na Figura 1.3 todo o enovelamento — alça F/G. Mas, até o presente momento, não é compreendido como essa região poderia atuar regulando o comportamento enzimático por se localizar longe do sítio catalítico ou como o tamanho da cadeia do substrato seria influenciado por essa região.

Figura 1.3 – Representação terciária da enzima OleT<sub>JE</sub>



A OleT<sub>JE</sub> em duas perspectivas distintas entre si por uma rotação de 90° do eixo *y*, com a alça F/G destacada em magenta e o sítio ativo da enzima, com os carbonos, oxigênios, ferro, nitrogênio e hidrogênios respectivamente em ciano, vermelho, rosê, anil e branco, evidenciando a distância entre esses domínios. Fonte: Figura elaborada pela autora.

Logo, o presente trabalho propõe um estudo computacional preliminar da etapa de saída dos produtos pela enzima OleT<sub>JE</sub>, utilizando simulações de dinâmica molecular. Dessa forma, procura-se compreender como a enzima OleT<sub>JE</sub> interage com ambos os produtos majoritariamente formados — alcenos terminais e ácidos graxos β-hidroxilados—, considerando o efeito de diferentes tamanhos de cadeias alifáticas, em paralelo aos dados apresentados por **Matthews *et al.* (2017)**, ao analisar como a mobilidade da alça F/G seria afetada na presença desses diferentes ligantes.

A dinâmica molecular clássica é uma técnica computacional, em que se utiliza programas de computador para simular o movimento de todos os átomos presentes em um sistema molecular em função do tempo. Conhecendo a localização de cada átomo e as interações entre eles, pode-se obter novas posições e velocidades de todos os átomos ao se integrar as equações de movimento de Newton ao longo do tempo e, a partir desses dados, calcular propriedades de interesse (**LEACH, 2001**). Esta técnica é muito utilizada para entender as implicações que as flutuações do movimento das proteínas têm no desempenho de suas funções biológicas. Isso porque as proteínas apresentam grande flexibilidade em sua estrutura terciária, que se refere à sua forma tridimensional única, que surge do

dobramento e flexão do esqueleto de suas cadeias peptídicas, e que são responsáveis pela funcionalidade biológica da molécula. Portanto, a compreensão de características do funcionamento enzimático desses sistemas vai além de sua estrutura tridimensional estática e só podem ser esclarecidas quando propriedades dinâmicas são levadas em consideração (**HOSPITAL *et al.*, 2015**). Dessa forma, a dinâmica molecular encontra grande aplicabilidade no estudo de biomoléculas, sendo uma ferramenta imprescindível para desvendar o seu funcionamento biológico a nível molecular.

Por fim, elucidar os mecanismos enzimáticos da OleT<sub>JE</sub> é de grande interesse, pois possibilita o desenvolvimento de novas enzimas mais eficientes para o uso em bioprocessos devido à produção de biocombustíveis compatíveis com os de origem fóssil (“*drop-in*”) e na utilização industrial em química fina desenvolvendo produtos biorrenováveis.

## 1.1 ORGANIZAÇÃO DA MONOGRAFIA

Nesse trabalho os capítulos 2 e 3 tratam, respectivamente, de uma revisão sobre conceitos relacionados as proteínas e seu funcionamento biológico e sobre a enzima OleT<sub>JE</sub>. Os capítulos 4, 5 e 6 são uma fundamentação teórica e computacional da técnica de dinâmica molecular. Os objetivos do trabalho, assim como a metodologia são abordados no capítulo 7. Os resultados são apresentados e discutidos no capítulo 8, enquanto, as conclusões e sugestões para trabalhos futuros estão no capítulo 9.

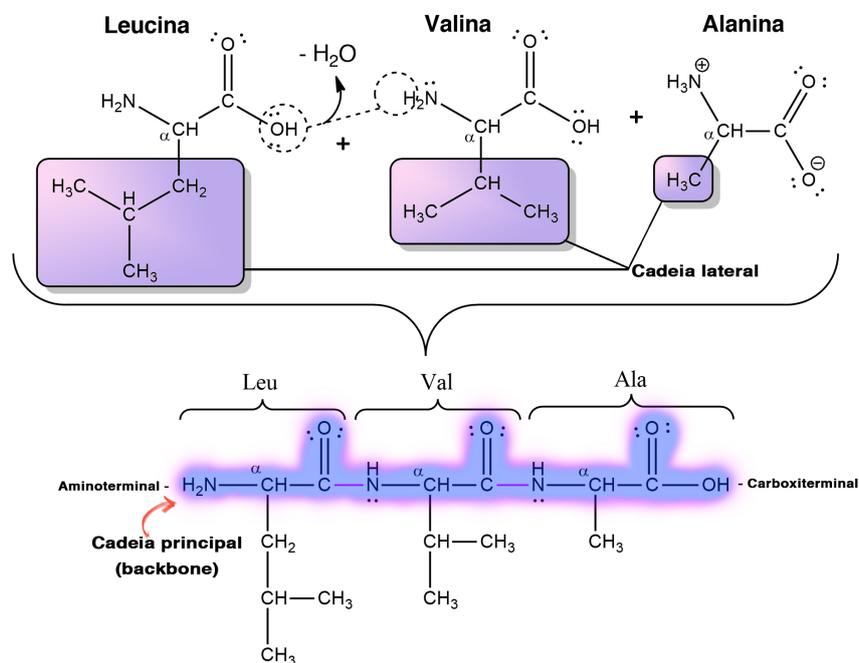
No **Anexo A** é apresentado a lista dos aminoácidos constituintes das proteínas, suas estruturas químicas e as respectivas siglas adotadas nesse trabalho. E, no **Anexo B** é apresentado a tabela periódica de elementos químicos atualizada.

## 2 PROTEÍNAS: ESTRUTURAS E DINÂMICA

### 2.1 PROTEÍNAS E SUAS ESTRUTURAS

As proteínas são macromoléculas formadas a partir da repetição múltipla de monômeros de aminoácidos que se unem a partir de um tipo específico de ligação covalente, denominada ligação peptídica. Esse aglomerado molecular de alta massa relativa<sup>1</sup> é formado a partir da combinação entre 20  $\alpha$ -aminoácidos específicos (ver **Anexo A** — Tabela dos aminoácidos que compõem as proteínas e suas siglas), que são compostos em que um grupo amino e uma carbonila estão ligadas ao mesmo átomo de carbono, definido como carbono  $\alpha$ . Cada aminoácido apresenta propriedades químicas próprias devido às diferenças em suas cadeias laterais, já que tal variação acarreta mudanças de estrutura, solubilidade, tamanho e carga elétrica (**BERG *et al.*, 2015**). A polimerização se dá quando o grupo  $\alpha$ -amino de um aminoácido se une ao grupo  $\alpha$ -carboxila ativado quimicamente de um segundo monômero, a partir de uma reação de condensação, mostrada na **Figura 2.1**. Nessa junção há a perda de uma molécula de água, o que faz com que os aminoácidos constituintes de uma cadeia peptídica passem a ser chamados de resíduos.

Figura 2.1 – Formação da ligação peptídica



Reação de condensação entre os aminoácidos leucina, valina e alanina. Fonte: Figura elaborada pela autora.

A ativação mencionada é necessária para que a formação peptídica seja termodinami-

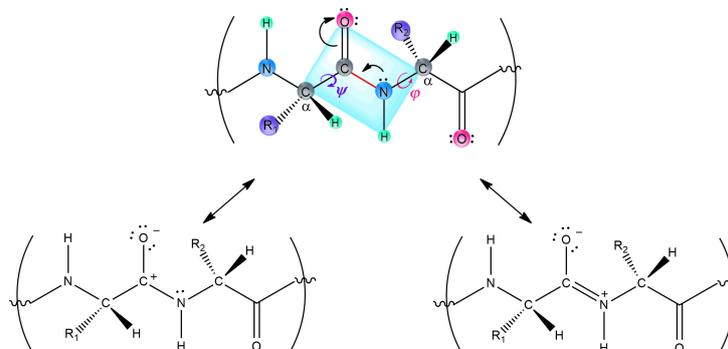
<sup>1</sup> Razão entre a massa da molécula e um duodécimo da massa de carbono.

camente favorável, visto que o grupo hidroxila não é um bom grupo de saída, sendo essa reação, no meio intracelular catalisada por enzimas presentes nos ribossomos, e em laboratório, mediada com o uso do reagente diciclohexilcarbodiimida (DCC) (MACHADO *et al.*, 2004).

De acordo com NELSON e COX (2014), as moléculas classificadas como polipeptídeos são definidas como um polímero de aminoácidos de massa molecular relativa bem menor ( $\leq 10000$ ) quando comparado às proteínas, e que apresentam apenas uma estrutura primária, que se refere à sequência de resíduos de aminoácidos que formam a cadeia polipeptídica. Convencionalmente, a sequência polipeptídica é expressa com a extremidade aminoterminal à esquerda e encerrada com um grupo carboxiterminal à direita, como demonstrado na **Figura 2.1**.

A ligação peptídica, que une os resíduos, exhibe uma geometria planar rígida, decorrente de efeitos de ressonância (**Figura 2.2**). Essa rigidez limita o movimento rotacional desse grupo, por isso, a conformação desses seis átomos é definida por dois ângulos de torção que refletem a rotação sobre as ligações da cadeia principal (esqueleto peptídico):  $\phi$ , referente à ligação N-C $_{\alpha}$  do diedro C-N-C $_{\alpha}$ -C, e  $\psi$ , relativo à ligação C $_{\alpha}$ -C do diedro N-C $_{\alpha}$ -C-N, podendo esses ângulos adotarem valores específicos entre  $\pm 180^{\circ}$  (NELSON; COX, 2014).

Figura 2.2 – Ressonância e planaridade do grupo peptídico



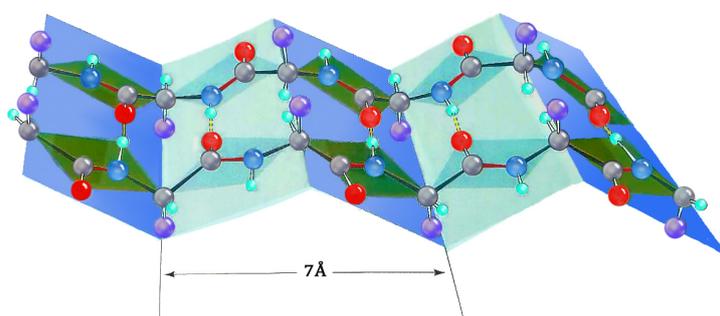
As estruturas de ressonância do grupo peptídico, com o plano evidenciado em ciano. Fonte: Figura elaborada pela autora.

Por outro lado, as proteínas, além de apresentarem uma massa relativa bem mais elevada que a dos polipeptídeos, podem ser compostas por duas ou mais cadeias peptídicas, que manifestam interações que resultam em dobramentos e enovelamentos de cadeias. Devido a essa estrutura complexa, as proteínas são descritas em 4 níveis de arranjos estruturais: a estrutura primária, secundária, terciária e quaternária. A estrutura primária considera apenas o sequenciamento linear do grupo peptídico na cadeia de resíduos. Enquanto a secundária, se refere ao arranjo espacial (conformação) adotado exclusivamente pelos átomos da cadeia principal guiados por ligações de hidrogênio formadas entre aminoácidos próximos, sem levar em consideração as interações com as cadeias laterais ou

interações com outros segmentos. Os padrões estruturais mais comumente observados são as folhas  $\beta$ -pregueadas e as  $\alpha$ -hélices, que ocorrem quando os ângulos de torção  $\phi$  e  $\psi$  permanecem os mesmos (ou quase os mesmos) ao longo de uma sequência de resíduos (NELSON; COX, 2014).

As folhas  $\beta$  surgem devido à coplanaridade do grupo peptídico, que faz com que os grupos ligados ao nitrogênio da amida e à carbonila se arranjam conformacionalmente de maneira que haja alternância entre as cadeias laterais, e em disposições geométricas que reduzam as repulsões interatômicas entre elas, produzindo um filamento  $\beta$ . As folhas  $\beta$ -pregueadas são formadas quando dois filamentos, ou mais, se unem por ligações de hidrogênio, como mostrado na **Figura 2.3**.

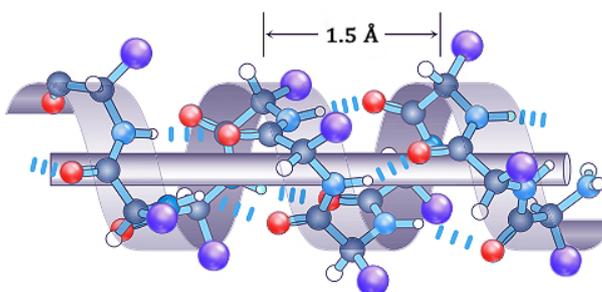
Figura 2.3 – Folha  $\beta$ -pregueada



Formação de uma folha  $\beta$ -pregueada por ligação de hidrogênio, em amarelo, entre dois filamentos beta. Fonte: Figura adaptada de Voet e Voet (1995).

Já as  $\alpha$ -hélices surgem quando ligações de hidrogênio são formadas entre o hidrogênio ligado ao átomo de nitrogênio de um aminoácido e o oxigênio da carbonila de um quarto resíduo de aminoácido da sequência peptídica. Essa conformação produz uma estrutura helicoidal (**Figura 2.4**), onde o esqueleto peptídico se enrola no entorno de um eixo longitudinal imaginário e as cadeias laterais são projetadas para fora do esqueleto da espiral (SOLOMONS *et al.*, 1996).

Figura 2.4 – Estrutura de uma  $\alpha$ -hélice



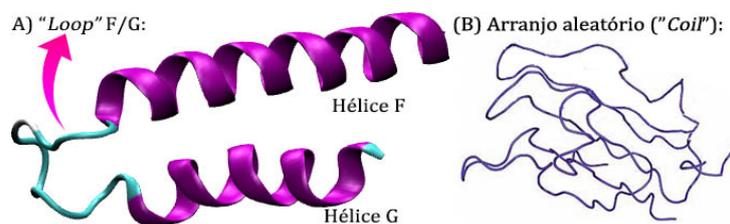
Fonte: Figura adaptada de NELSON e COX (2014).

Ambas as estruturas secundárias descritas anteriormente são facilmente identificadas

devido à regularidade de recorrências dos grupos peptídicos, sendo típico para as  $\alpha$ -hélice uma distância de repetição de 1.5 Å (**Figura 2.3**), e de 7.0 Å para as folhas  $\beta$ -pregueadas (**Figura 2.4**) (**BERG et al., 2015**).

Existem também estruturas secundárias não regulares, que não apresentam um padrão específico de estruturação para ocorrer, mas que são caracterizadas por uma certa rigidez local possuindo uma estrutura bem definida, chamada de “*loops*”. Nas proteínas globulares<sup>2</sup> essas estruturas auxiliam a inversão de direção das cadeias entre folhas  $\beta$ , ou simplesmente conectam duas estruturas secundárias  $\alpha - \alpha$  ou  $\alpha - \beta$ , visando uma maior compactação da estrutura proteica total, sendo mostradas na figura abaixo.

Figura 2.5 – Regiões de “*loops*” e “*coils*”



(A) Região da alça F/G, com as hélices F e G em magenta conectadas pelo “*loop*” em azul. (B) Um arranjo aleatório ou “*coils*” em violeta. Fonte: Figura elaborada pela autora.

Por fim, as proteínas podem também apresentar regiões onde a cadeia peptídica tem uma disposição aleatória (desordenada) de grande mobilidade, e com filamento extremamente flexível, que não se encaixa em nenhuma das definições anteriores de estruturas secundárias, chamada de espiral aleatória ou “*coils*”, sendo encontradas próximas às regiões N-terminal ou C-terminal da proteína (**KABSCH; SANDER, 1983**).

Já a estrutura terciária decorre de interações entre os aminoácidos de diferentes estruturas secundárias que surgem devido ao dobramento e sobreposição das cadeias peptídicas da proteína, além de envolverem a participação das cadeias laterais. Algumas proteínas contêm mais de uma cadeia polipeptídica, idênticas ou não, que, ao se arranjam entre si, formam complexos tridimensionais a partir da interação entre suas diferentes estruturas terciárias. Como consequência dessa interação, novos dobramentos da cadeia ocorrem e uma nova estrutura tridimensional é formada, intitulada estrutura quaternária.

O arranjo tridimensional final da proteína origina-se da estrutura terciária pelo processo de enovelamento (dobramento) que surge de interações intermoleculares estabilizantes que visam atingir uma conformação energeticamente mais favorável. A principal força estabilizadora, que rege o enovelamento, provém das interações hidrofóbicas que surgem no ambiente aquoso, que impulsionam os resíduos não polares a se aglomerarem no interior da estrutura, de forma a evitar contato com solvente, formando uma região

<sup>2</sup> Proteínas que se doblam em uma forma esférica de forma que os aminoácidos hidrofílicos se projetem para a parte externa, enquanto os resíduos hidrofóbicos se dispõem na parte interna, para garantir a sua solubilidade em água.

hidrofóbica no interior da proteína. Esse processo impede a formação de uma camada de solvatação no entorno dos resíduos hidrofóbicos, que devido a ordenação molecular causa um aumento na entropia do sistema. Como consequência, os resíduos polares são projetados para a superfície, mas de modo a formarem interações intramoleculares entre si por meio de ligações de hidrogênio e com o mínimo de interações com a água (pelo mesmo princípio entrópico anterior).

## 2.2 DINÂMICA DE PROTEÍNAS

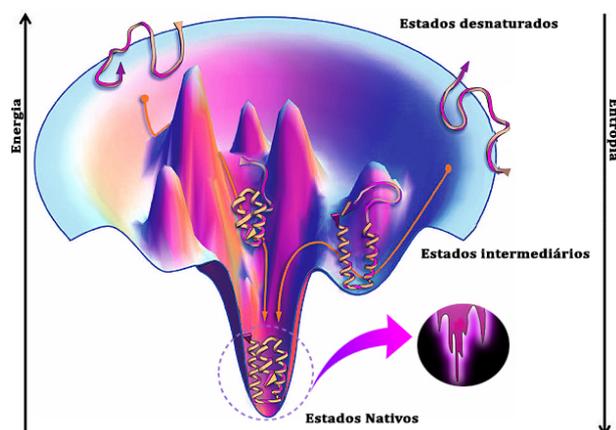
### 2.2.1 Flexibilidade Conformacional

Enzimas são proteínas cuja função é catalisar reações químicas em ambientes biológicos que são caracterizados, em geral, por condições de temperatura e pressão moderadas. A eficiência no uso de enzimas vai além da redução da energia de ativação (comum a todos os catalisadores) e de condições reacionais mais brandas quando comparadas às reações não enzimáticas. Essas estruturas proteicas evoluíram efetivamente para aperfeiçoar sua seletividade – capacidade de atuar em reações de forma a gerar um produto específico –; e sua especificidade – habilidade de atuar em um substrato específico. As enzimas podem acelerar a velocidade de uma reação em até 20 ordens de grandeza em comparação com a correspondente reação não catalisada (**WOLFENDEN; SNIDER, 2001**). E de fato, a descarboxilação de ácidos graxos alifáticos não ocorre facilmente por vias químicas, mas é naturalmente realizada por enzimas específicas, como a OleT<sub>JE</sub>.

Para desempenhar a sua função natural, as enzimas sofrem diferentes mudanças conformacionais, onde tais variações compõem o conjunto (“*ensemble*”) de conformações que caracterizam o seu estado nativo, que é definido como o modo estrutural apto para exercer sua função biológica. Estes estados conformacionais (**Figura 2.6**) envolvem diferentes estruturas que englobam todas as variações necessárias para que as etapas do ciclo catalítico de uma enzima possam ocorrer, e abrangem desde a entrada do substrato; seu deslocamento pela cavidade até o sítio de ligação; a formação do complexo enzima-substrato; sua conversão; dissociação e a saída do produto de dentro do canal de ligação.

A flexibilidade conformacional de uma proteína, se refere a facilidade com a qual ocorre a mudança de estrutura entre seus diferentes confôrmeros, e envolve além do número de conformações possíveis, a estabilidade dessas estruturas, as barreiras de energia que as separam, os parâmetros cinéticos de interconversão entre elas e o movimento térmico dos átomos que compõem o sistema (**HUBER, 1979**). Esses diferentes estados conformacionais possuem diferentes reatividades em relação ao substrato quando este se encontra ligado ao sítio ativo, e a conformação mais cataliticamente ativa pode ser referente ao um estado de mais alta energia, havendo uma maior barreira que precisa ser vencida. Logo, a movimentação da estrutura proteica permite a exploração de diferentes estados conformacionais até que ocorra o cruzamento da barreira. O tempo necessário para que ocorra

Figura 2.6 – Espaço conformacional do enovelamento de proteínas



O espaço conformacional de proteínas possui uma paisagem em formato de funil que decai aos estados nativos, com a presença de morros que representam dobramentos de alta energia. Estados intermediários representam configurações de mínimo de energia local, e se localizam em pequenas depressões no espaço. Fonte: Figura adaptada de **Dill e Maccallum (2012)**

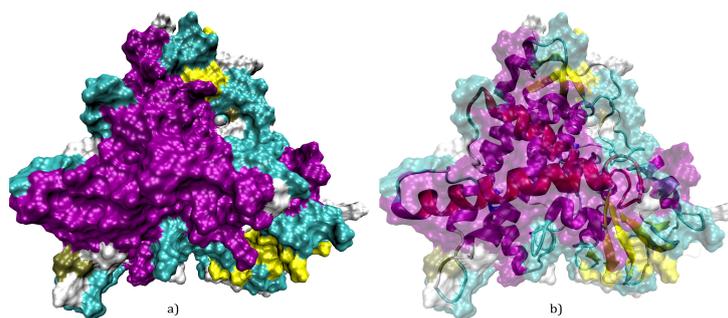
esse tipo de evento é definido pela complexidade do movimento atômico necessário para atingir a mudança estrutural de forma coordenada de todos os átomos do sistema biomolecular, que devido ao grande número de partículas característico desse tipo de sistema, pode ter uma menor probabilidade de ocorrência e levar mais tempo para acontecer.

Além disso, a flexibilidade conformacional possui um papel essencial na acessibilidade do sítio ativo pelo substrato. As proteínas são estruturas sem cavidades acessíveis, como evidenciadas na **Figura 2.7**, e regiões como a alça F/G da OleT<sub>JE</sub> comumente contribuem para a criação de canais transitórios dentro da estrutura, para permitirem a entrada e difusão do substrato da solução até o sítio de ligação. Esse tipo de enovelamento também contribui para isolar o sítio ativo do solvente, de forma a alterar o ambiente químico no entorno do substrato para desencadear o evento catalítico. E a maleabilidade conformacional da proteína atua ainda na orientação do substrato pelo canal de ligação e no seu ajuste em uma posição ótima no sítio ativo, para que a reação química possa ocorrer de forma efetiva.

Ademais, a redução da barreira de energia livre e, conseqüentemente, a aceleração da reação enzimática, depende diretamente da flexibilidade do sítio ativo. E de acordo **Kokinidis, Glykos e Fadoulglou (2012)**, muitos estudos apontam inclusive para uma conexão entre a flexibilidade conformacional do sítio catalítico com a taxa de renovação do substrato (“*turnover*”).

Logo, a compreensão de eventos característicos do funcionamento biológico pode ocorrer apenas quando as propriedades dinâmicas da proteína, por exemplo, a difusão do substrato pelo canal de reação ou o coeficiente da taxa de dissociação do produto, estão sendo levadas em consideração.

Figura 2.7 – Superfície de contorno da OleT<sub>JE</sub>



a) Modelo de contorno da superfície da OleT<sub>JE</sub> evidenciando a falta de cavidades que permita a entrada de substrato. b) Superfície de contorno com a estrutura terciária evidenciando a alça F/G em magenta. Para a entrada e saída do substrato, a alça deve ser flexível o suficiente para criar um canal dentro da proteína. Fonte: Figura elaborada pela autora.

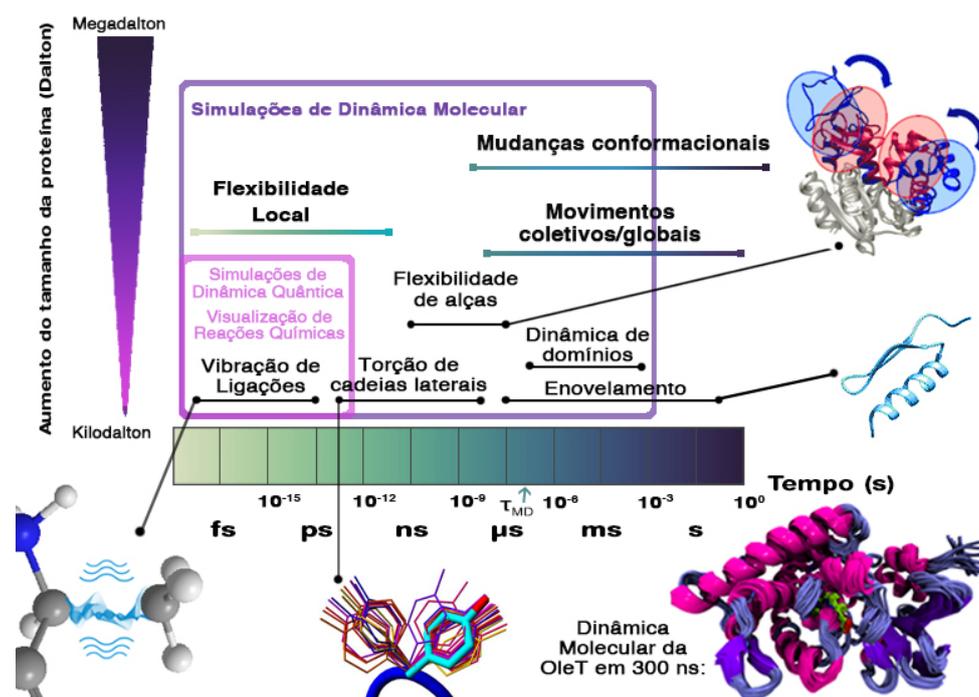
### 2.2.2 Escala de tempo da dinâmica de proteínas

As propriedades dinâmicas de proteínas envolvem qualquer fenômeno que varia ao longo do tempo e que tenha origem nas interações entre as partículas constituintes do sistema molecular, e que regem a posição e a velocidade simultaneamente de todos os seus constituintes. Mais especificamente, que se originam da flutuação das posições atômicas dos átomos, gerando diferentes movimentos característicos para essas biomoléculas em diferentes escalas de tempo. Esses tipos de movimentos originados das flutuações das posições atômicas são chamados de dinâmica de proteínas, e são mostrados na **Figura 2.8**.

A flexibilidade das cadeias peptídicas de uma proteína, é uma propriedade dinâmica, provinda diretamente da flutuação das posições atômicas, e pode se referir tanto a variação de geometria dos átomos (mudança conformacional), quanto a mobilidade translacional dos mesmos, sendo neste caso, chamada de flexibilidade local e envolve o movimento de um grupo pequeno de átomos. Além disso, estruturas secundárias podem compor um conjunto chamado domínio, que englobam uma sequência de aminoácidos independentemente estável em relação à proteína, a ponto da estrutura terciária desses resíduos se manter mesmo após serem separados (clivagem) da cadeia. Essa independência é tal, que o domínio pode se movimentar como uma entidade isolada em relação ao resto da proteína. Quando um processo envolve o movimento coordenado de mais de um domínio ou um conjunto grande de resíduos, a flexibilidade é dita global.

A importância do movimento da estrutura terciária da proteína vai muito além, e está intrinsecamente associado a sua função biológica, sendo essa relação bem estabelecida e reconhecida (**HOSPITAL et al., 2015**). Outra propriedade dinâmica, que resulta das flutuações das posições atômicas é a **plasticidade estrutural**, onde de alguma maneira ainda não totalmente compreendida, a possibilidade da enzima aceitar diferentes tipos de substratos de diferentes funções químicas, abrindo mão da sua especificidade, envolve

Figura 2.8 – Escala de tempo das dinâmicas de uma proteína



Simulações de dinâmica molecular permitem o estudo de sistemas macromoleculares, mas não a visualização de reações químicas. Fonte: Figura elaborada pela autora.

movimentos coletivos/globais de toda a sua estrutura, de forma que ela possa se ajustar a diferentes ligantes mas proporcionando a mesma reação química (KOKKINIDIS; GLYKOS; FADOULOGLOU, 2012).

As proteínas são devido a essa natureza dinâmica, como uma rede de movimentos contínuos, que surgem a partir dos movimentos rápidos e localizados originários das flutuações das posições atômicas, e que influenciam de forma a propagar a dinâmica para movimentações em escalas de tempo maiores (eventos lentos e ou raros). Devido a isso, as diferentes possibilidades de mobilidade biomolecular são consideradas interdependentes e acopladas entre si. Como consequência, as simulações não podem negligenciar os movimentos de pequena escala (rápidos), mesmo que se pretenda estudar dinâmicas mais lentas e de maior importância biológica, como processos de enovelamento e a movimentação de domínios.

As simulações de proteínas utilizando a técnica de dinâmica molecular amostram o espaço conformacional do estado nativo de biomoléculas com uma dependência temporal, o que permite visualizar em escala microscópica os diferentes movimentos dinâmicos da proteína e como essas flutuações se relacionam com o seu papel biológico e sua atividade catalítica.

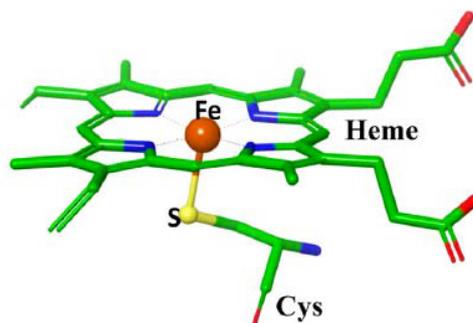
### 3 A DESCARBOXILASE OLE<sub>T<sub>JE</sub></sub>

A produção de hidrocarbonetos a partir de recursos naturais abundantes (óleos e gorduras vegetais) tem atraído grande protagonismo científico devido à possibilidade de substituir, pelo menos em parte, a produção de combustíveis fósseis por alternativas renováveis e mais sustentáveis. Além disso, a geração de biocombustíveis obtida pela rota metabólica de ácidos graxos fornece moléculas compatíveis com a infraestrutura energética pré-existente (DU *et al.*, 2017) (PERALTA-YAHYA *et al.*, 2012). Devido a isso, muitos estudos foram conduzidos sobre a Ole<sub>T<sub>JE</sub></sub> — e que serão abordados nesse capítulo —, uma enzima capaz de sintetizar alcenos terminais a partir de ácidos graxos, os quais podem ser gerados a partir de óleos e gorduras utilizando lipases, ou obtidos diretamente do metabolismo de lipídeos de micro-organismos.

Relatada pela primeira vez em 2011 por RUDE *et al.* (2011), a Ole<sub>T<sub>JE</sub></sub> foi encontrada em um grupo de bactérias do gênero *Jeotgalicoccus* sp. associadas a animais marinhos. Nesse estudo, diferentes cepas de *Jeotgalicoccus in vivo* foram colocadas em um meio de cultivo nutricional contendo uma solução saturada de ácido eicosanóico (C20) produzindo seletivamente 1-nonadeceno (C19). Os ensaios mostraram um aumento de 20% na concentração de C19 em relação ao controle, o que comprovava a presença da reação de descarboxilação do C20 e que a biossíntese de olefinas é um comportamento comum ao gênero *Jeotgalicoccus*. Quando as bactérias foram colocadas em meios que continham exclusivamente ácido palmítico (C16) ou ácido esteárico (C18), observou-se, respectivamente, uma conversão de apenas 6% em C15 e 23% em C17, ocorrendo em ambos os meios a formação majoritária de olefinas C19. Evidenciando predileção biológica por substratos de cadeia longa, visto que os ácidos de cadeias menores foram naturalmente alongados antes da ocorrência da descarboxilação. Estudos conduzidos por Liu *et al.* (2014) mostraram a incapacidade da Ole<sub>T<sub>JE</sub></sub> em promover a descarboxilação de ácido cáprico (C10) ou ácido caprílico (C8) *in vivo*. E Hsieh *et al.* (2017) apontaram uma afinidade enzima-substrato substancialmente menor para ácidos graxos com cadeias curtas (C8-C12) em comparação a cadeias maiores, além da geração de produtos secundários indesejados ( $\alpha$  e  $\beta$ -hidroxi), sustentando que ácidos graxos C<sub>n</sub> com n  $\geq$  16 são de fato os substratos nativos para Ole<sub>T<sub>JE</sub></sub> *in vivo*.

Com base em sua sequência de aminoácidos, a enzima foi incluída na família 152 (CYP152) da superfamília dos citocromos P450 (RUDE *et al.*, 2011). Os citocromos são enzimas que apresentam em seu sítio ativo um átomo de ferro complexado a uma protoporfirina (IX), conhecida como heme. O ferro presente no heme auxilia a transferência de elétrons de acordo com a mudança da sua forma oxidada à reduzida, estando nos citocromos P450, especificamente, ligado à proteína por um resíduo de cisteína, como mostrado na Figura 3.1.

Figura 3.1 – Grupo Heme



Grupo heme com os átomos de carbono, nitrogênio, oxigênio, ferro e enxofre representados em verde, azul, vermelho, laranja e amarelo, respectivamente. Fonte: Figura de Dubey e Shaik (2019).

Os citocromos P450 englobam uma super família de enzimas cuja principal função é a eliminação de xenobióticos<sup>1</sup>, mas podem também atuar na biossíntese de hormônios para vias metabólicas ou na produção de vitaminas (HSIEH *et al.*, 2017). As P450 catalisam diferentes tipos de reações químicas, como epoxidação, desalogenação e hidroxilação, sendo definidas como mono-oxigenases por introduzirem um átomo de oxigênio molecular (advindo de uma molécula de O<sub>2</sub>) em diferentes substratos orgânicos.

O ciclo catalítico dos citocromos P450 envolve a ligação do átomo de ferro do heme a uma molécula de oxigênio (cosubstrato), e a interação do cofator com diferentes espécies redox, que doam cada uma, um único elétron ao complexo. Essas espécies redox são enzimas redutoras que transferem elétrons a partir de cofatores como NADPH ou FADH.

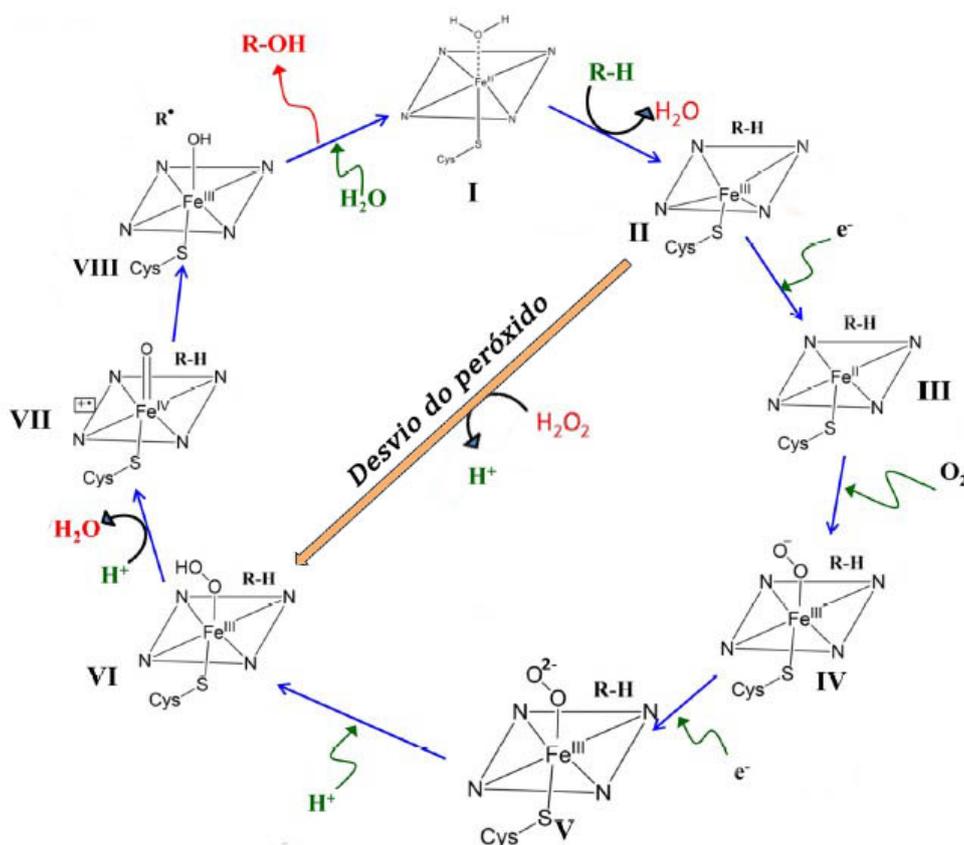
Entretanto, a família de CYP152 são peroxigenases e utilizam o peróxido de hidrogênio (H<sub>2</sub>O<sub>2</sub>) como cosubstrato e acceptor de elétrons, conduzindo a um desvio na rota catalítica das mono-oxigenases conhecido como “*peroxide shunt*”, para formar diretamente o composto 0 (VI), mostrado na **Figura 3.2**.

As CYP152 evoluíram eficientemente para que não houvesse danos oxidativos à proteína ou ao heme, utilizando o peróxido como única fonte de oxigênio e sem a necessidade de usar qualquer tipo de parceiro redox, tornando o processo menos complexo (WANG *et al.*, 2017) (MUNRO *et al.*, 2018). As peroxigenases mais estudadas antes da descoberta da OleT<sub>JE</sub> eram: a P450<sub>SPα</sub> (Matsunaga *et al.*, 1996a; 1997b), que catalisa exclusivamente a α-hidroxilação de ácidos graxos e a P450<sub>BSβ</sub> (MATSUNAGA *et al.*, 1999c) que converte tanto α quanto β-hidroxilação na proporção 40:60 respectivamente.

O ciclo catalítico das peroxigenases se inicia com o heme férrico (Fe<sup>3+</sup>) hexa-coordenado por um tiolato de cisteína (S) e com uma molécula de água (H<sub>2</sub>O) fracamente ligada, definido como estado de repouso (I). Com a entrada do substrato ao sítio ativo, e sua ligação a um resíduo de arginina (Arg242) localizado próximo ao heme, uma mudança conforma-

<sup>1</sup> Compostos químicos não pertencentes a um determinado organismo ou sistema biológico.

Figura 3.2 – Ciclo catalítico da superfamília de P450



O ciclo catalítico das P450 se iniciando pelo complexo no estado de repouso (I). As setas azuis representam o caminho seguido pelas mono-oxigenases, no qual são adicionados 2 elétrons ao sistema: o primeiro (III), para a redução do Fe(III) à Fe(II), que então reage com uma molécula de  $O_2$  (IV); o segundo elétron permite a formação do composto ferro-óxido (V), que ao sofrer uma protonação forma o Composto 0 (VI). O composto 0, então, sofre uma clivagem heterolítica na ligação O-O liberando uma molécula de água e gera o composto I. A seta laranja mostra o desvio do peróxido utilizado pelas CYP152 que se ligam ao  $H_2O_2$  formando diretamente o composto 0 (VI). Fonte: Figura adaptada de Dubey e Shaik (2019).

cional na enzima ocorre, de forma a provocar a expulsão de todas as moléculas de água presentes dentro da cavidade de ligação, incluindo a água coordenada ao ferro, gerando o ferro heme de alto spin penta-coordenado (II), sendo essas etapas comuns a todas as P450.

De alguma forma ainda não conhecida, a ligação do substrato ao sítio ativo permite o surgimento de algum canal por onde possa ocorrer a entrada do peróxido de hidrogênio dentro da cavidade de ligação. O cosubstrato  $H_2O_2$  então se liga ao ferro, onde essa etapa é chamada de ativação do ferro heme por peróxido. Esse processo de entrada e ativação do peróxido nunca ocorre sem que o substrato já esteja ligado a arginina, mas os fatores por trás desse comportamento não são conhecidos. Outro mecanismo ainda não elucidado, envolve o rearranjo que permite a formação da espécie transitente hidroperóxido-férrico

conhecida como composto 0 (**VI**), que ao liberar uma molécula de água, forma a espécie radicalar altamente reativa, óxido-ferryl ( $\text{Fe}^{4+}$ ) porfirina, denominada composto I (**VII**). Sendo este, o estado responsável por promover as reações nos citocromos P450 (**DUBEY; SHAIK, 2019**).

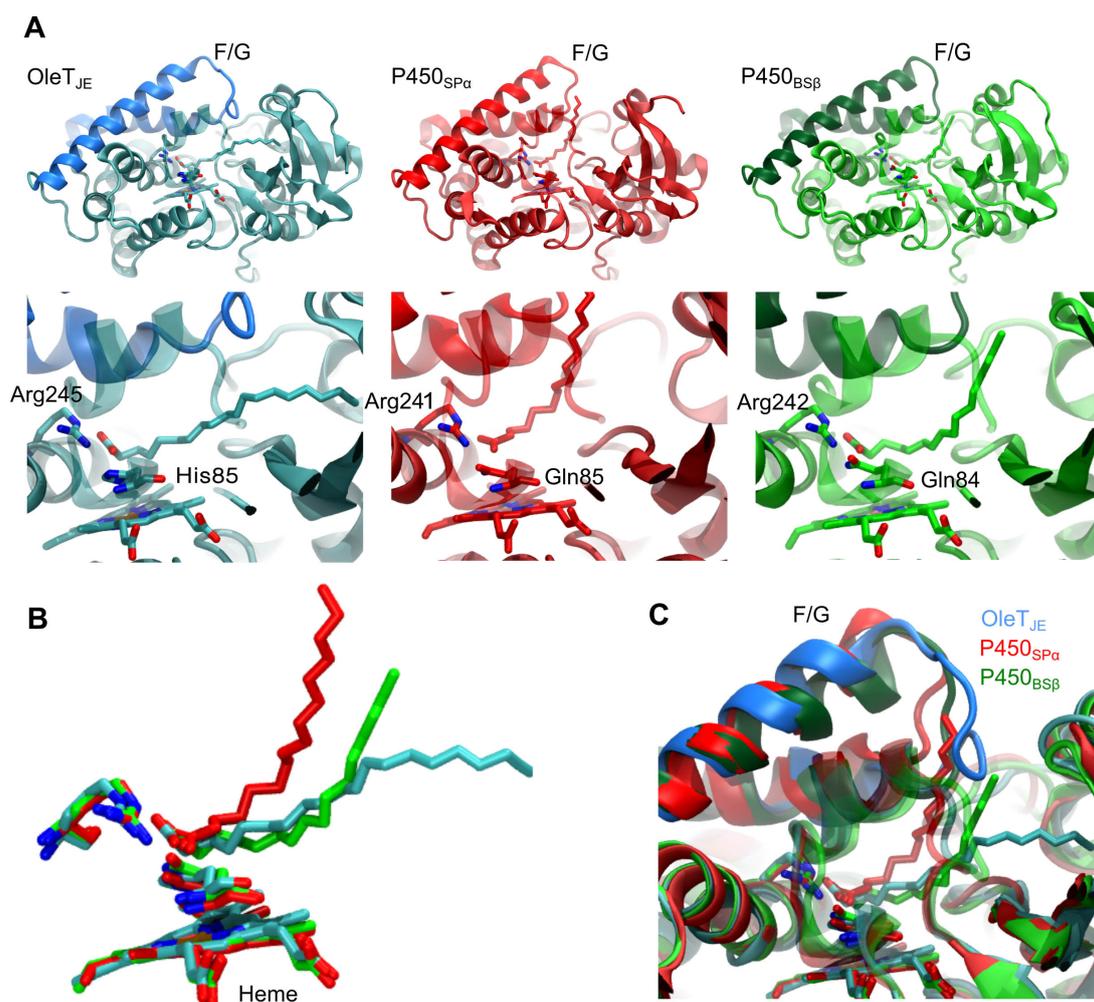
Para promover a hidroxilação nas peroxigenases, o composto I abstrai um hidrogênio do substrato produzindo um ligante radicalar e o hidróxido-ferryl ( $\text{Fe}^{4+}$ ), denominado composto II (**VIII**), onde posteriormente ambos reagem entre si para formar o produto hidroxilado e liberar o ferro heme no estado férrico, que pode ser restaurado ao estado de repouso com a saída do produto e a entrada do ligante aqua, fechando o ciclo catalítico como evidenciado na **Figura 3.2** (**DUBEY; SHAIK, 2019**).

A  $\text{P450}_{SP\alpha}$  e a  $\text{P450}_{BS\beta}$  compartilham 44% de **identidade** da sequência de aminoácidos. Uma enzima é atribuída a uma determinada família quando ela compartilha 25% ou mais dos resíduos que compõe sua sequência peptídica (estrutura primária). Esse valor é obtido ao se alinhar a sequência de aminoácidos entre as duas enzimas, e significa que por compartilharem certas sequências de resíduos, que essas regiões de equivalência geram as mesmas estruturas secundárias que podem refletir em funções biológicas semelhantes (**NELSON; COX, 2014**).

A principal distinção entre a  $\text{P450}_{SP\alpha}$  e a  $\text{P450}_{BS\beta}$  envolve a orientação do canal hidrofóbico — cavidade envolta pelas hélices F e G, que condiciona o substrato no sítio ativo —, que é quase perpendicular ao plano do heme na  $\alpha$ -hidroxilase, mostrado na **Figura 3.3** (B). Além disso, a  $SP\alpha$  apresenta a cavidade do sítio ativo mais estreita, devido a presença de um resíduo fenilalanina de posição 288 na sequência de aminoácidos (Phe288), mediante que sua cadeia lateral se projeta para dentro da cavidade, de forma a interagir diretamente com o ácido graxo. Enquanto, na  $\text{P450}_{BS\beta}$  essa posição é ocupada pela Gly290 que não possui cadeia lateral. Outra diferença é o resíduo Leu78, localizado na borda do canal da  $SP\alpha$ , que por ter menor tamanho torna a entrada da cavidade mais aberta, em comparação ao equivalente na  $BS\beta$ , Phe79. Estudos conduzidos por **Fujishiro et al. (2011)** sobre as  $SP\alpha$  sugerem que tanto a orientação do canal quanto a interação do substrato com os resíduos Phe288 e Arg241 posicionam o ligante perpendicularmente à coenzima de forma que o carbono alfa fique mais próximo ao heme, permitindo apenas a oxidação do grupo  $\alpha\text{-CH}_2$ . Por outro lado, na  $\text{CYP152}_{BS\beta}$  uma maior cavidade e a falta da interação com um resíduo como o Phe288 deixa o substrato mais livre no canal, o que permite o acesso do heme ao grupo  $\beta\text{-CH}_2$  e justificaria a perda da seletividade de acordo com **Dubey e Shaik (2019)**. A disposição do substrato ligado ao sítio ativo em cada peroxigenase é mostrada na figura **Figura 3.3**.

Por sua vez, a  $\text{OleT}_{JE}$  e a  $\text{P450}_{BS\beta}$  compartilham uma similaridade de 41% de **identidade**, com um desvio da raiz quadrada do deslocamento quadrático médio (RMSD) de 0,99 Å para 379 Carbonos- $\alpha$  da sequência. A  $\text{OleT}_{JE}$  é composta por 419 aminoácidos, enquanto a  $BS\beta$ , 416, onde apenas 41% desses resíduos se encontram com uma mesma

Figura 3.3 – As Peroxigenases CYP152



Em A: Comparação estrutural das enzimas OleT<sub>JE</sub>, em azul; P450<sub>SPα</sub>, em vermelho e P450<sub>BSβ</sub>, em verde; com as principais diferenças entre os sítios ativos. B e C mostram o posicionamento do ácido graxo em cada enzima, ilustrando a diferença de orientação dos canais hidrofóbicos nas três proteínas. Fonte: Figura de Rodrigo L. Silveira.

distribuição sequenciada. Para encontrar semelhanças de estrutura terciária entre duas enzimas distintas (até mesmo quando as sequências são pouco similares), pode ser feita com o alinhamento de suas estruturas tridimensionais. Isso porquê, a presença de diferentes aminoácidos ao longo de uma cadeia não reflete necessariamente numa estrutura secundária diferente, logo as proteínas podem se dobrar e enovelar de forma análoga. Quando duas estruturas tridimensionais tem um determinado conjunto de átomos alinhados é possível medir o quanto elas se sobrepõem, onde para duas proteínas idênticas o desvio de RMSD é zero, pois elas ficam perfeitamente sobrepostas. E segundo **Belcher et al. (2014)**, as regiões que apresentam desvios significativos de RMSD na OleT<sub>JE</sub>, envolvem áreas que revestem a cavidade de ligação do substrato sendo referentes ao “loop” F/G e regiões adjacentes ao “coil” C-terminal. E que a principal diferença no sítio ativo entre as CYP152 ocorre na posição 85, que é ocupada por uma histidina na descarboxilase, mas

em ambas as hidroxilases (P450<sub>SP $\alpha$</sub>  e P450<sub>BS $\beta$</sub> ), por uma glutamina.

A importância do resíduo His85 na descarboxilação de ácidos graxos foi levantada, primeiramente, por **RUDE *et al.* (2011)** mediante os resultados obtidos para uma variante mutante da enzima P450<sub>BS $\beta$</sub>  que tem uma histidina na posição 85 (**Gln85His**). Nesse estudo, além da evidência, até então desconhecida, da capacidade do citocromo P450<sub>BS $\beta$</sub>  em descarboxilar ácidos graxos em proporções semelhantes à OleT<sub>JE</sub>, a mutante Gln85His apresentou um aumento de atividade em cerca de 50% na descarboxilação de ácido palmítico (C16), aumentando a razão descarboxilação/hidroxilação para 0,30 em comparação à enzima selvagem (nativa) de 0,19. Entretanto, a P450<sub>SP $\alpha$</sub> , que compartilha 36% de identidade da sequência de aminoácidos com a OleT<sub>JE</sub> não evidenciou mudanças de seletividade ou de atividade com a substituição da glutamina pelo resíduo His85. E apesar da mutante  $\beta$ -hidroxilase Gln85His ter a proporção de  $\alpha$ -hidroxi diminuída com aumento de  $\alpha$ -olefina, o incremento na atividade de descarboxilação veio acompanhado de uma maior produção de ácido  $\beta$ -hidroxilado. Em suma, esses estudos mostram que, embora a His85 possa ter um papel importante na descarboxilação, ela não é essencial. Na **Figura 3.3** são mostradas as principais diferenças estruturais entre as peroxigenases CYP152 mencionadas.

Um mecanismo proposto por **Belcher *et al.* (2014)** para justificar os fatores que levariam uma enzima à descarboxilação e outra à hidroxilação, sugere que a His85 atuaria doando um próton ao composto I concomitantemente à abstração de um elétron do substrato para a redução ao composto II, em oposição à remoção de um átomo de hidrogênio do ácido graxo (mecanismo comumente observado em P450). Nesse mecanismo, ocorreria a cisão homolítica da ligação O<sub>2</sub>C-C $\alpha$  do ácido graxo junto com remoção de um hidrogênio do carbono- $\alpha$ , produzindo o alceno terminal, uma molécula de CO<sub>2</sub> e o composto II.

Entretanto, foi observado em um outro citocromo P450 da família CYP152, encontrada na bactéria *Methylobacterium populi*, que também exibe atividade de descarboxilação, que uma metionina ocupa a posição 85, demonstrando, portanto, que as ligações de hidrogênio envolvendo a histidina também não são essenciais para a descarboxilação, podendo ocorrer por intermédio de outros resíduos próximos ao heme. E de acordo com **Matthews *et al.* (2017)**, a mutante His85Gln da OleT<sub>JE</sub> ainda favorece a produção de alceno em detrimento da hidroxilação, reafirmando que o resíduo His85 não é um fator determinante na descarboxilação.

Ainda nesse trabalho, **Matthews *et al.* (2017)** investigou diferentes mutações de resíduos no sítio ativo da OleT<sub>JE</sub>, onde foi observado para as mutantes His85Gln e Phe79Ala pequenas variações de atividade enzimática, mas não havendo mudanças significativas nas estruturas cristalográficas em comparação à enzima nativa, sugerindo que as diferenças no perfil de produtos gerados poderiam estar relacionadas à mobilidade do sítio ativo e do substrato. O estudo também sugere que a seletividade (descarboxilação *versus* hidroxilação) tem origem em variações sutis no posicionamento e nas flutuações atômicas dos

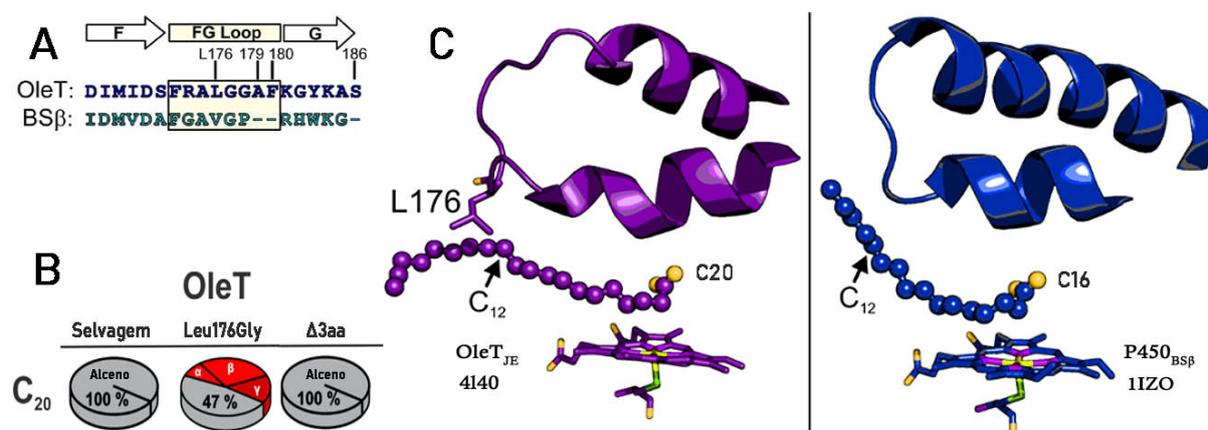
hidrogênios nas posições  $\alpha$  e  $\beta$  do substrato. Além disso, o estudo mostrou a dependência da seletividade e especificidade com o tamanho da cadeia do substrato, mostrada previamente na **Figura 1.2**, onde a enzima apresentou maiores rendimentos na conversão de ácidos graxos C14, sendo de quase 100%, e a menor para C20, com apenas 14%; mas que em contraponto, apresenta seletividade máxima para descarboxilação. Curiosamente, o substrato com C18 apresenta uma inversão de regioseletividade em relação às outras cadeias, apresentando uma conversão maior para ácidos graxos  $\alpha$ -hidroxilados do que para  $\beta$ . Mas os fatores moleculares que justifiquem as atividades e seletividade observadas em diferentes tamanhos de substrato ou a mudança de regioseletividade no sistema C18 ainda não são conhecidos.

Estudos conduzidos por **Amaya et al. (2018)**, compararam as estruturas cristalográficas da descarboxilase na presença de ácido eicosanoico (C20) (PDB:4l40), com a  $\beta$ -hidroxilase na presença de C16 (PDB: 1IZO), onde se sugeriu que devido a OleT<sub>JE</sub> apresentar dois resíduos a mais na região do “loop” entre as hélices F e G, em comparação a BS $\beta$ , que esses aminoácidos seriam responsáveis pela maneira destoante como os ligantes se projetam dentro da cavidade de ligação (**Figura 3.3**). Um desses resíduos é a Leu176, que de acordo com o estudo, tem a sua cadeia lateral interagindo hidrofobicamente com o substrato, de forma a provocar uma torção no ligante a partir do carbono C12 do ácido, sendo mostrada na **Figura 3.4** (C). E que a ausência desse tipo de interação na enzima BS $\beta$  contribuiria para uma atividade de descarboxilação menor. Com base nos estudos conduzidos por **Dodani et al. (2016)** sobre um citocromo P450 TxtE, que catalisa a nitração aromática de L-triptofano, em que pequenas mutações dos resíduos localizados no “loop” F/G provocaram mudanças significativas de regioespecificidade; **Amaya et al. (2018)** sugerem que um comportamento análogo poderia ser esperado para OleT<sub>JE</sub>.

Dessa forma, **Amaya et al. (2018)** estudaram a mutação Leu176Gly na presença de ácido eicosanoico (C20), que supostamente resultaria em uma perda de interação com o substrato, já que a glicina não possui cadeia lateral. Sendo relatado, nessa mutante, um taxa de conversão de 47% para alceno e de 53% para ácidos graxos hidroxilados, com as hidroxilações ocorrendo até nas posições  $\epsilon$ ,  $\gamma$  e  $\sigma$ . Com base na perda de seletividade e regioseletividade demonstrada na mutante, os autores especularam que, na Leu176Gly, o ácido graxo se posicionaria de forma parecida com o observado na enzima P450<sub>BS $\beta$</sub>  na presença de C16. E que portanto, a interação da leucina na posição 176 com o substrato seria responsável por influenciar a seletividade e a regioseletividade da enzima. O artigo também apresentou a mutante da OleT<sub>JE</sub>  $\Delta$ 3aa, em que a região do “loop” F/G foi supostamente modelada para se aproximar da estrutura da BS $\beta$  com a remoção dos dois resíduos extras do “loop” (179 e 180) e um da hélice G adjacente (186), mas mantendo a Leu176. Como essa variante não apresentou qualquer perda de seletividade na conversão de ácidos graxos C20, se reforçou a importância da interação Leu176-substrato para o resultado catalítico.

Mas, pela observação da estrutura cristalográfica da OleT<sub>JE</sub> e da BS $\beta$  as afirmações levantadas por **Amaya et al. (2018)** não são evidentes. Primeiramente, porquê os PDBs indicam que o “loop” F/G da hidroxilase engloba a sequência de resíduos 172 a 177, enquanto na descarboxilase vai da posição 173 ao 181, o que gera uma diferença de 3 aminoácidos nessa região, não sendo justificável a escolha de se remover a Ser186 da hélice G. Além disso, não é perceptível a interação da cadeia lateral da Leu176, que inclusive se projeta de forma diferente ao ilustrado no artigo, não sendo apontado qualquer outro fator que sustente que apenas a interação da leucina com o substrato é determinante para seletividade. Além do mais, dos 9 resíduos que compõem o “loop” da OleT<sub>JE</sub> apenas 3 são conservados na BS $\beta$ , e de acordo com **Dodani et al. (2016)** e com o próprio **Amaya et al. (2018)** a mudança arbitrária de qualquer resíduo no “loop” F/G provoca a perda de seletividade nas P450. Por fim, não existem dados sobre a atividade, seletividade ou estrutura cristalográfica da P450<sub>BS $\beta$</sub>  com ligante C20, de forma que a deleção dos resíduos 179, 180 e 186 apontam somente que a interação desses aminoácidos não influenciam no perfil catalítico da descarboxilase, mas que todos os outros resíduos preservados tanto no “loop” como nas hélices F/G podem ser importantes para ditar a seletividade enzimática da OleT<sub>JE</sub>.

Figura 3.4 – As diferentes alças F/G na (A) OleT<sub>JE</sub> e na (B) P450<sub>BS $\beta$</sub>



Em A: O alinhamento da sequência de aminoácidos dos “loops” F/G e das hélices adjacentes nas enzimas OleT<sub>JE</sub> e P450<sub>BS $\beta$</sub>  sugerida pelo artigo, indicando os resíduos explorados por mutagênese sítio-dirigida. Em B: Comparação da taxa de conversão de substrato C<sub>20</sub> nas diferentes enzimas reportadas: enzima selvagem e a  $\Delta$ 3aa, que apresentam 100% de produção de alcenos, enquanto a mutante Leu176Gly converte apenas 47% em olefinas terminais com a produção de 53% em ácidos hidroxilados. Em C: OleT<sub>JE</sub> apresenta um “loop” F/G mais longo, com o resíduo Leu176 interagindo com a parte apolar do substrato C<sub>20</sub>. Enquanto, na P450<sub>BS $\beta$</sub> , a alça é mais curta, o que abre espaço para que o substrato C<sub>16</sub> se projete em direção oposta, conforme observado em sua estrutura cristalográfica. O resíduo 176 é apontado por **Amaya et al. (2018)** como o principal modulador da descarboxilação. Fonte: Figura adaptada de **Amaya et al. (2018)**.

No mesmo trabalho, **Amaya et al. (2018)** produziram ácidos graxos com sondas

espectroscopicamente ativas anexadas em diferentes pontos da cadeia alifática do substrato, de forma a se preservar o sítio ativo da enzima. Os ácidos graxos ligados às sondas apresentaram constantes de dissociação e afinidades com o sítio ativo com valores semelhantes aos substratos que não tinham as sondas (livres). Logo, se sugeriu que essas sondas se comportavam metabolicamente de maneira análoga aos substratos livres e que esses ligantes poderiam adotar configurações estruturais dentro da cavidade iguais aos substratos nativos da OleT<sub>JE</sub>. A partir das medidas de fluorescência e FRET, o artigo concluiu que a etapa lenta de todo o ciclo catalítico da OleT<sub>JE</sub> é a dissociação do produto, o que conseqüentemente, limita o “*turnover*” da enzima. O baixo número de renovações da enzima selvagem, de acordo com **Amaya et al. (2018)** seria devido ao “*loop*” F/G, pois ambas as mutantes, Leu176Gly e  $\Delta 3aa$ , apresentaram um aumento do “*turnover*” consistente com uma aceleração da taxa de saída de produto. Mas, com ambas as mutações perdendo a quimiosseletividade mista exibida pela selvagem, passando a metabolizar exclusivamente álcoois graxos.

A maior facilidade com que os produtos saem da cavidade quando a enzima sofre alterações em resíduos ao longo da enovelamento F/G (176, 179, 180 e 186), evidenciados por **Amaya et al. (2018)**, sugerem uma redução na interação desse domínio com o substrato, podendo ser decorrente de um alça F/G mais móvel e flexível que auxiliaria na expulsão mais eficiente dos ligantes. Além disso, **Amaya et al. (2018)** sugeriram que o “*loop*” F/G poderia influenciar os modos de ligação dos substratos ou alterar sua mobilidade na cavidade de ligação ao permitir uma maior hidratação do sítio ativo. Mas, não foram apresentados dados que confirmassem essas suposições.

Nas P450, além da região do “*loop*” F/G, o “*loop*” B/C também é comumente apontado como responsável pela criação de canais transitórios na enzima, atuando também como tampa para gerir a entrada e saída do produto. Esse domínio também é capaz de auxiliar o alinhamento do ligante no sítio ativo ou de alterar a solvatação na cavidade de ligação (**DODANI et al., 2016**). Enquanto, estudos realizados por **Du et al. (2017)** sugerem que regiões de encontro de “*loop*” na OleT<sub>JE</sub>, poderiam formar canais para a entrada de água no sítio ativo da enzima, de forma a mediar a saída dos ligantes. A cavidade de ligação da OleT<sub>JE</sub> é predominada por resíduos hidrofóbicos, e a solvatação da cavidade contribuiria para expulsar as olefinas terminais hidrofílicas, pois diminui a solvatação hidrofóbica, resultando numa redução de entropia.

No presente trabalho, simulações de dinâmica molecular são utilizadas para estudar a possível flexibilidade da alça F/G e a sua interação com diferentes produtos (alcenos/ácidos  $\beta$ -hidroxi), considerando o efeito do tamanho de suas cadeias alifáticas. Além de se investigar como essa região pode estar relacionada a uma maior taxa de saída de produto; com a criação de canais de solvatação que auxiliam a expulsão dos ligantes e com a mobilidade dos produtos dentro da cavidade de reação.

## 4 DINÂMICA MOLECULAR CLÁSSICA

### 4.1 MECÂNICA ESTATÍSTICA

As primeiras técnicas de simulação computacional de dinâmica molecular foram desenvolvidas a partir de 1952, quando o Laboratório Nacional de Los Alamos, no Novo México (EUA) disponibilizou para fins não militares, a máquina **MANIAC I**, criada inicialmente para realizar cálculos precisos dos processos termonucleares de uma bomba de hidrogênio (**WOOD, 1985**). Os primeiros cálculos submetidos buscavam solucionar problemas inerente à física de estado líquido, e culminaram no desenvolvimento do método de dinâmica molecular criado por **Alder e Wainwright (1957)**, onde se estudou a colisão entre esferas rígidas submetidas a um potencial atrativo do tipo poço quadrado. Diferentes implementações no método, desenvolvidas ao longo dos anos, contribuíram para que, em 1977, **McCammon, Gelin e Karplus (1977)** publicassem a primeira dinâmica molecular de uma biomolécula, uma inibidora de tripsina, selando a fusão entre a mecânica estatística computacional e a bioquímica.

A falta de uma teoria molecular universal para descrever líquidos e sistemas biológicos complexos, onde as interações entre os constituintes não podem ser negligenciadas, fez com que tais sistemas macroscópicos fossem estudados sob um ponto de vista microscópico, a partir das propriedades individuais das moléculas que o englobam. A mecânica estatística é área da física responsável por fazer essa conexão entre o mundo microscópico (propriedades de átomos e moléculas) com o macroscópico (propriedades termodinâmicas). Dessa maneira, propriedades macroscópicas, como capacidade calorífica e difusão, podem ser inferidas a partir de propriedades moleculares partindo-se de dados estruturais ou espectroscópicos.

O estado que descreve o comportamento físico de um sistema microscópico requer que um grande número de graus de liberdade (todas as posições e velocidades de todas as partículas) sejam definidos. Enquanto, o mesmo sistema, em equilíbrio, para ser descrito macroscopicamente necessita de número muito menor de parâmetros, tais como pressão, temperatura e volume. Num sistema clássico de partículas pontuais, um microestado define a posição e o momento de cada partícula integrante do sistema. Onde, é possível a partir desse microestados obter todos os valores de todas as variáveis microscópicas<sup>1</sup> possíveis. Podendo um dado sistema existir em um conjunto discreto, mas, possivelmente infinito, de microestados (**MCQUARRIE, 1976**).

Para um sistema atômico de  $N$  partículas, define-se o espaço fásico ( $\Gamma$ ), que é um espaço  $6N$ -dimensional cuja base é formada pelo conjunto de todas as  $3N$  posições ( $r^N$ )

<sup>1</sup> Variável relativa aos átomos e moléculas individuais que compõem o sistema. Por exemplo, a componente  $x$  da velocidade de uma partícula; o número quântico para um modo vibracional específico de uma molécula ou o vetor de momento magnético de um íon paramagnético.

num sistema cartesiano e seus respectivos  $3N$  momentos ( $p^N$ ) de cada uma das partículas. Cada “ponto”<sup>2</sup> desse espaço representa, um microestado possível do sistema que evolui ao longo do tempo segundo as leis de movimento, gerando uma trajetória no espaço de fases. Nessa região, defini-se a densidades de probabilidade para as coordenadas ou dos microestados individuais “acessíveis”, obtidos com a restrição do espaço fásico ao se fixar determinadas variáveis ( $E, V, N$ , etc.). Sendo, a densidade de probabilidade uma função que se conserva pois não tem uma dependência explícita com o tempo. O tratamento estatístico, a partir da medida de probabilidade tem o propósito de estimar o tamanho relativo do subconjunto de microestados “acessíveis” do sistema.

A passagem da descrição micro para a macroscópica de um sistema é feita pelo princípio de Boltzmann-Gibbs, que postula que toda função de estado pertinente a um volume restrito do espaço de fase, possui um valor macroscópico em equilíbrio correspondente, que pode ser obtido pelo valor esperado (ou valor médio) da função com respeito a uma medida de probabilidade “adequada”. A partir dessa formulação, surge a definição de “ensemble” estatístico, que é o conjunto de medidas de probabilidade invariantes indexadas pelos parâmetros termodinâmicos ( $T, V, E$ , etc.) do sistema estudado que satisfazem o Princípio de Boltzmann-Gibbs (**VOLCHAN, 2006**).

Um sistema macroscópico em equilíbrio naturalmente emerge de microestados se valendo de uma dependência temporal, onde, diferentes pontos no espaço fásico podem descrever um mesmo macroestado definido pelas mesmas constantes termodinâmicas. Onde, o sistema tem igual probabilidade de estar em qualquer microestado que seja consistente com seu macroestado atual - hipótese do “igual a priori” de Gibbs. Os graus de liberdade de um sistema atômico flutua constantemente ao longo do tempo, logo, defini-se a probabilidade de o sistema estar num microestado como a fração de tempo que ele passa nesse estado. As médias calculadas usando esta definição de probabilidade são chamadas de médias temporal. Dessa forma, quando uma propriedade mecânica<sup>3</sup> de um sistema macroscópico é medida experimentalmente, existe o tempo de leitura que o instrumento leva para obter um valor. Durante esse tempo, os microestados das partículas do sistema flutuam, ou seja, diferentes valores são contabilizados, de forma que o valor observado macroscopicamente é a média temporal do equivalente microscópico da grandeza medida. Isso significa, que quando uma propriedade experimental é auferida, o valor observado é na realidade uma média da grandeza ao longo do tempo da medição (**GOODISMAN, 1997**). Quando médias são calculadas considerando a probabilidade do sistema estar em determinado microestado, como a fração de sistemas imaginários de um mesmo macroestado que estão no mesmo microestado, ou seja, a probabilidade desse sistema de se encontra num determinado microestado, são chamadas de médias de “ensemble”. Mas

<sup>2</sup> Devido ao princípio de Heisenberg, existem incertezas acerca das posições e momentos, logo, um microestado corresponde a um hipervolume centrado no ponto  $\Gamma = [r^N, p^N]$ , composto por uma infinidade de pontos.

<sup>3</sup> Grandeza que não necessita do conceito de temperatura para ser definida.

que, pelo postulado de hipótese de ergodicidade ambos os valores, média temporal e média de “ensemble”, são equivalentes.

Logo, para se calcular as propriedades mecânicas de um sistema, como energia e pressão, é possível expressá-las em termos de seu equivalente microscópico, ou seja, a propriedade é calculada como função das coordenadas  $(r^N, p^N)$  no espaço de fases. Para uma propriedade qualquer cujo equivalente microscópico é  $A(r^N, p^N)$ , haverá uma flutuação de  $A$  a medida que ela evolui no tempo, com cada microestado tendo um valor diferente de  $A(r^N, p^N)$ . Conhecendo-se a evolução temporal de  $A(r^N, p^N)$ , podemos calcular sua média temporal, durante um intervalo de tempo  $\tau$ , como:

$$\bar{A} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(p^N(t), r^N(t)) dt \quad (4.1)$$

Quando  $\tau$  tende ao infinito, o valor médio de  $A$  é igual ao seu valor macroscópico (termodinâmico). Em outras palavras, uma propriedade termodinâmica pode ser calculada como uma média de seu equivalente microscópico sobre todos os microestados acessíveis pelo sistema. Como, em simulação de dinâmica molecular, se obtém as posições e momentos de cada átomo em função do tempo, este é um método para obter a evolução temporal de um sistema no espaço de fases. Logo, qualquer propriedade termodinâmica (mecânica) pode ser calculada através da **Equação 4.1**. Entretanto, como as equações de movimento são integradas numericamente ao longo de intervalos de tempo discretos ( $\delta t$ ) e, ainda, como o tempo de simulação não é infinito, consistindo em um número  $n$  grande de intervalos  $\delta t$ , a média da **Equação 4.1** toma a seguinte forma:

$$\bar{A} = \frac{1}{n} \sum_{n=0}^n A(n\Delta t) \quad (4.2)$$

O sistema se move pelo espaço de fase segundo as equações de Hamilton, um formalismo mais geral, utilizado para expressar as leis de movimento de Newton independentemente do sistema de coordenadas, mas com relação direta com a energia total do sistema. As equações de Hamilton para uma partícula  $i$  são:

$$\frac{dr_i}{dt} = \frac{\partial H(r^N, p^N)}{\partial p_i}, \quad (4.3)$$

$$\frac{dp_i}{dt} = -\frac{\partial H(r^N, p^N)}{\partial r_i}, \quad (4.4)$$

onde  $H(r^N, p^N)$  é a função Hamiltoniana que descreve a energia total do sistema, sendo expressa como:

$$H(r^N, p^N) = \sum_{i=1}^N \frac{p_i^2}{2m_i} + U(r^N) \quad (4.5)$$

onde  $U(r^N)$  é a função energia potencial e  $m_i$  a massa da partícula  $i$ .

Ao se resolver numericamente as equações de movimento de Hamilton, as simulações de dinâmica molecular geram trajetórias no espaço fásico para um sistema isolado, onde o número de partículas ( $N$ ), volume ( $V$ ) e energia total do sistema ( $E$ ) são mantidos constantes, produzindo um conjunto de pontos chamado de “*ensemble*” microcanônico, ou NVE. Esse conjunto estatístico amostra o espaço fásico ao longo de um contorno de energia constante onde se define uma função de partição microcanônica  $\Omega(E, V, N)$  para representar todos os microestados degenerados, isto é, de igual energia, que o sistema pode admitir de distribuição “*a priori*” de probabilidade.

Entretanto, nesse tipo de “*ensemble*” existem variações entre as energias cinéticas e potenciais do sistema, que refletem em grandes flutuações de temperatura. E, quando se busca estudar sistemas atômicos cuja dinâmica é diretamente influenciada pela temperatura, como no caso de biomoléculas, é necessário transformar a densidade de probabilidade para um conjunto estatístico que reflete o comportamento de interesse. No limite termodinâmica, em que o número de átomo do sistema tende ao infinito, e fora do espaço regiões de transição de fase, os “*ensembles*” são equivalentes. Entretanto, as simulações computacionais não são realizadas nesse limite, além disso, determinados “*ensembles*” são mais recomendados que outros para se calcular determinadas propriedades físicas.

## 4.2 APROXIMAÇÃO DE BORN-OPPENHEIMER (ABO)

A **Dinâmica Molecular** - “*Molecular Dynamics*” (MD) propõe o estudo de um sistema em escala atômica, sob perspectiva da mecânica clássica. O modelo admite o átomo como a menor unidade de matéria que compõe o sistema. Os átomos são descritos em maior contribuição pelos seus respectivos núcleos, quando esses adotam a representação de uma partícula pontual única. Os elétrons não são considerados explicitamente no modelo, mas influenciam indiretamente, de forma que cada átomo apresente diferentes parâmetros que descrevem suas interações, de acordo com o elemento que se pretende descrever. Como consequência dessa interpretação, qualquer aspecto quântico referente tanto à estrutura atômica quanto ao movimento nuclear e eletrônico é negligenciado. Dessa forma, não são contemplados durante simulações de **MD** efeitos de origem quântica, como o rompimento ou formação de ligações químicas, presença de ressonância (MARTÍNEZ; BORIN; SKAF, 2007).

A abordagem adotada no modelo provém da Aproximação de Born-Oppenheimer (ABO), um método utilizado para solucionar o operador Hamiltoniano de um sistema molecular (não-relativístico), ao desacoplar os movimentos eletrônicos e nucleares. Esse desacoplamento de núcleos e elétrons tem por base o fato de que devido ao tamanho e massa dos elétrons, muito mais leves que o núcleo atômico, que o seu movimento é proporcionalmente mais rápido em comparação ao movimento nuclear. Por sua vez, os núcleos são tão mais pesados e lentos, que pequenas variações em sua posição fazem com que os

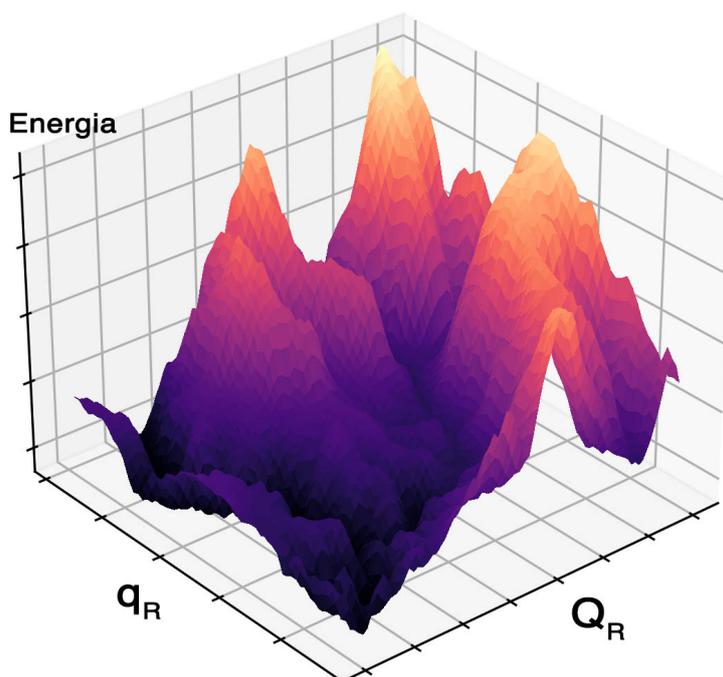
elétrons se adaptem de forma instantânea à nova configuração. Com isso, o sistema pode ser descrito como uma função de onda molecular ( $\psi(r, R)$ ) separável em termos função de onda eletrônica ( $\phi_e(r; R)$ ) e da função de onda nuclear ( $\phi_N(R)$ ):

$$\psi(r, R) = \phi_e(r; R)\phi_N(R) \quad (4.6)$$

Essa diferença de massa entre o núcleo e os elétrons é expressiva, a ponto das funções de onda que compõem o pacote de onda nuclear sofrerem interferências construtivas em pequenas regiões do espaço, e destrutivas nas demais regiões, gerando uma função de onda altamente localizada, com posição bem definida. Dessa forma, na dinâmica molecular convencional, os núcleos podem ser considerados como partículas puntiformes clássicas que podem se propagar em trajetórias regidas pela mecânica newtoniana.

Na ABO, o movimento nuclear não engloba diferentes estados (quânticos) eletrônicos, possibilitando que a energia eletrônica do sistema molecular possa ser escrita em função da posição dos núcleos ( $R$ ). A resolução da equação de Schrödinger eletrônica para diferentes geometrias conformacionais (que fornece seus autovalores  $\varepsilon_i$ ) gera uma superfície de energia potencial (SEP) descrita pela presença dos elétrons em diferentes posições nucleares, mostrada na **Figura 4.1**. Esses autovalores ( $\varepsilon_i$ ) junto com a repulsão

Figura 4.1 – Superfície de Energia Potencial



Superfície de energia potencial obtida para um sistema molecular fictício de coordenadas genéricas  $q_R$  e  $Q_R$ , considerando diferentes conformações geométricas, na qual os núcleos se movem. Fonte: Figura elaborada pela autora.

eletrostática dos núcleos compõem a energia potencial total do sistema molecular,  $U(R)$ . Como o hamiltoniano nuclear obtido na **ABO** é a soma da energia cinética dos núcleos

com a energia potencial total, os elétrons contribuem indiretamente ao modelo, já que o efeito causado pela sua presença é considerado no potencial total. E a energia total do sistema molecular pode ser descrito como uma função apenas das coordenadas nucleares e do momento dos núcleos, equivalente a função Hamiltoniana clássica na **Equação 4.5**.

Na dinâmica molecular clássica, a função potencial total para diferentes configurações nucleares é expressa em termos de uma função analítica simples, cujos parâmetros são obtidos por meio da energia fornecida por cálculos quânticos *ab initio* de configuração eletrônica. No **Capítulo 5** serão abordados os termos que compõem essa função potencial.

A abordagem clássica para o movimento atômico reduz o tempo e o esforço computacional necessário para o estudo de sistemas biológicos, que podem contabilizar de dezenas de milhares a milhões de átomos após a adição do ambiente aquoso de fase condensada na macromolécula de interesse. Devido ao caráter dinâmico desses sistemas e a mobilidade do ambiente químico que o rodeia, números muito grandes de conformações são gerados e que precisam ser submetidos a cálculos de energia, tornando o tratamento quântico proibitivo para tais sistemas (**MACKERELL, 2001**). Dessa forma, o Hamiltoniano para um sistema isolado – energia e massa mantidas constantes –, em que o potencial é uma função apenas das coordenadas (potencial conservativo), obtendo-se a segunda Lei de Newton considerando que a função Hamiltoniana não varia no tempo:

$$\frac{dH(\mathbf{r}, \dot{\mathbf{r}})}{dt} = \frac{m_i}{2} \frac{d\dot{\mathbf{r}}^2}{dt} + \frac{dU(\mathbf{r})}{dt} = 0; \quad (4.7)$$

$$\frac{m_i}{2} \frac{d\dot{\mathbf{r}}^2}{dt} = -\frac{dU(\mathbf{r})}{dt}; \quad (4.8)$$

$$\mathbf{F}_i = m_i \ddot{\mathbf{r}}_i = -\frac{dU(\mathbf{r})}{d\mathbf{r}_i}; \quad (4.9)$$

onde  $\mathbf{F}_i$  é a força resultante atuando sobre um átomo  $i$ , de massa  $m_i$  e posição  $\mathbf{r}_i$  sob a influência do potencial  $U(\mathbf{r})$  num sistema de  $N$  átomos interagentes.

### 4.3 ALGORITMO DE INTEGRAÇÃO: *VELOCITY-VERLET*

Em dinâmica molecular o termo potencial  $U(\mathbf{r}^N)$  é chamado de campo de força, pois é a partir desse potencial que as forças newtonianas instantâneas que atuam nos átomos são geradas, segundo a **Equação 4.10**:

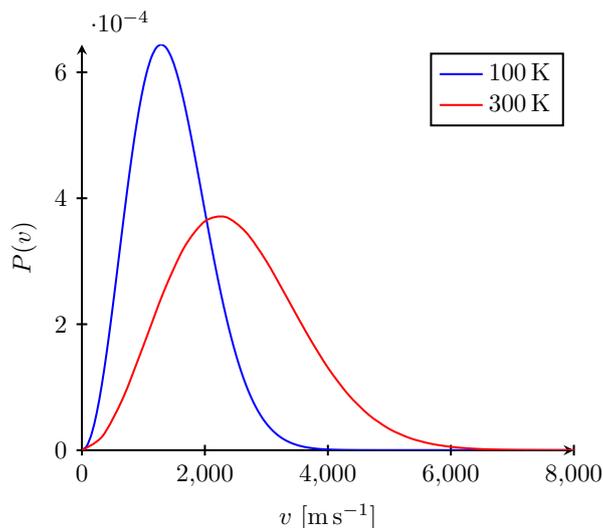
$$\mathbf{F}_i = -\frac{dU(\mathbf{r}_N)}{d\mathbf{r}_i} \quad (4.10)$$

A trajetória de cada partícula  $i$  só pode ser obtida ao se resolver numericamente as equações de movimento. Pois, devido à não-linearidade do potencial de interação e ao grande número de átomos envolvidos é impossível resolver analiticamente as equações

Newtonianas. A impossibilidade sendo necessário o uso de uma sequência de operações aritméticas que sob certas condições culminem numa solução aproximada do problema.

O processo é feito partindo-se de uma configuração inicial do sistema de  $N$  átomos, no instante inicial  $t = 0$  que se encontra nas posições  $r^N(t = 0)$ , onde velocidades aleatórias são atribuídas a cada partícula segundo a distribuição de Maxwell-Boltzmann correspondente à temperatura simulada, de acordo com a **Figura 4.2**.

Figura 4.2 – Distribuição de Maxwell-Boltzmann



Distribuição de Maxwell-Boltzmann para as velocidades prováveis nas temperaturas 100 e 300K de uma partícula em equilíbrio térmico com o seu entorno. Fonte: Figura elaborada pela autora.

A designação é feita de forma que o momento linear total seja zero, para que não haja um deslocamento do sistema como um todo e para que a energia cinética total seja consistente com a temperatura desejada. A partir da configuração inicial, o potencial de interação agindo em uma partícula  $i$  pode ser calculado, e, conseqüentemente, a força resultante atuando nesse átomo; onde, partindo de sua respectiva massa, a aceleração pode ser obtida. Entretanto, como o cálculo da força depende da posição de todos os outros átomos do sistema, isso acarreta o acoplamento das equações de movimento individuais de cada partícula. Logo, para se calcular a evolução do sistema para o tempo  $t + \Delta t$  é necessário armazenar a posição e velocidade das  $N$  partículas num tempo  $t$ , e calcular todas as forças de interação entre os átomos. Devido a isso, diferentes métodos para se integrar a **Equação 4.9** foram desenvolvidos em dinâmica molecular de forma a se obter a evolução temporal do sistema.

O método *Velocity-Verlet* (**SWOPE et al., 1982**), é uma variação do algoritmo baseado no método de integração numérico popularizado por Loup Verlet (**VERLET, 1967**) para calcular trajetórias em simulações de dinâmica molecular. Suas principais vantagens são a conservação da energia total do sistema; o momento linear não acumular

erros de precisão; a preservação do volume no espaço fásico e de permitir a reversibilidade temporal das equações de movimento. Assumindo que as trajetórias obtidas classicamente são contínuas, a posição da partícula  $i$  num instante posterior  $t + \delta t$ , onde,  $\delta t$  representa o intervalo de tempo entre **iterações**. Quando  $\delta t$  é pequeno o suficiente (infinitesimal), a posição pode ser expressa por uma expansão em série de Taylor:

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t)(\delta t) + \frac{1}{2!}\ddot{\mathbf{r}}_i(t)(\delta t)^2 + \dots \quad (4.11)$$

Essa série pode ser truncada no termo de segunda ordem e reescrita de acordo com as definições físicas das propriedades do sistema, de acordo com a **Equação 4.12**:

$$\mathbf{r}_i(t + \delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(\delta t) + \frac{1}{2}\mathbf{a}_i(t)(\delta t)^2 + O(\delta t)^n \quad (4.12)$$

onde  $\mathbf{v}_i$  é a velocidade da partícula  $i$  e  $\mathbf{a}_i$  a sua aceleração. Os termos não considerados da sequência devido ao truncamento geram uma propagação de erro,  $O(\delta t)^n$ , devido à ausência das contribuições  $(\delta t^3, \dots, \delta t^n)$ .

No instante inicial  $t_0$ , a posição inicial  $\mathbf{r}_i(t_0)$  provém da configuração tridimensional da proteína;  $\mathbf{v}_i$  é gerado aleatoriamente pelo algoritmo de **Monte Carlo** segundo a distribuição de probabilidade (Maxwell-Boltzmann) na temperatura do sistema e  $\mathbf{a}_i(t_0)$  pela equação:

$$\mathbf{a}_i(t = t_0) = -\frac{1}{m_i} \frac{dU(r^N(t_0))}{d\mathbf{r}_i} \quad (4.13)$$

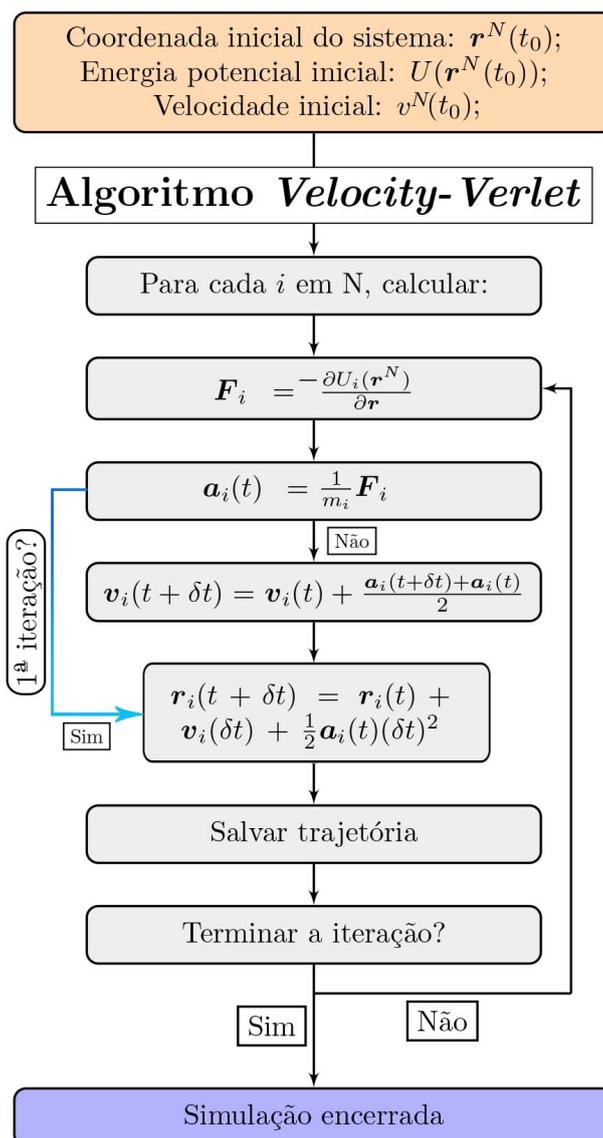
Mas para um instante posterior a  $t_0$ , a velocidade em  $v_0 + dt$  passa a ser calculada a partir de duas expansões em série de Taylor que resultam na expressão:

$$\mathbf{v}_i(t + \delta t) = \mathbf{v}_i(t) + \frac{\mathbf{a}_i(t + \delta t) + \mathbf{a}_i(t)}{2} \delta t \quad (4.14)$$

Entretanto, como essa expressão envolve uma aceleração num tempo futuro  $t + \delta t$ , antes é necessário computá-la, usando a **Equação 4.13**, e considerando a posição obtida na primeira **iteração**. Na **Figura 4.3** é representado um esquema computacional do algoritmo.

O processo de integração das equações ao longo das iterações acumula erros discretizados devido ao truncamento das séries utilizadas, sendo o valor uma função exponencial do intervalo de tempo ( $\delta t$ ) escolhido e acarreta num aumento da energia do sistema ao longo da simulação. Para mitigar o erro total acumulado, escolhe-se um intervalo de tempo ínfimo. Numa simulação de MD, ( $\delta t$ ) deve ser pequeno o suficiente para englobar os movimentos de vibração atômica e, como mencionado anteriormente, permitir a expansão das variáveis físicas numa expansão de série. Utilizar um passo de tempo extremamente pequeno leva a resultados mais precisos, mas acarreta em um custo de tempo de computação mais longo, que por sua vez, reduziria o tempo de simulação total devido ao esforço computacional exigido. Além disso, um tempo de simulação menor reduz a

Figura 4.3 – Diagrama do Algoritmo Velocity-Verlet



A partir das posições, velocidades e potenciais iniciais, o algoritmo calcula as forças e acelerações atuando em cada átomo. Na primeira iteração, a velocidade é definida pela distribuição de Maxwell-Boltzmann correspondente à temperatura em que sistema está sendo simulado. Mas, após o tempo inicial, a velocidade é calculada seguindo a **Equação 4.14**. As trajetórias são salvas e o processo reiniciado até a alcançar o tempo de simulação desejado. Fonte: Figura elaborada pela autora.

amostragem conformacional do espaço fásico, diminuindo a confiabilidade estatística do processo, havendo um “*tradeoff*” entre a acurácia física e a eficiência computacional. O aumento do intervalo ( $\delta t$ ) permite simular uma escala maior de tempo, o que é necessário para o estudo de biomoléculas de acordo com a **Figura 2.8**, mas acarretando em uma perda de acurácia (LEACH, 2001).

## 4.4 CONTROLE DAS SIMULAÇÕES

O algoritmo “*velocity-Verlet*” gera trajetórias no “*ensemble*” microcanônico NVE, como mencionado no final da **seção 4.1**, onde o sistema apresenta uma conservação da energia, o número de partículas se mantém o mesmo durante toda a simulação, assim como o volume do sistema. Entretanto, em simulações de biomoléculas não se tem o interesse em visitar diferentes configurações de mesma energia, e sim, percorrer o espaço fásico acessível na temperatura em que se dá os processos biológicos, sendo referente a um “*ensemble*” canônico, NVT. Ou ainda, submeter o sistema a variações de volume ao longo do tempo, mas mantendo a pressão e a temperatura constantes, gerando conformações NPT chamadas de “*ensemble*” isotérmico-isobárico. Dessa forma, algoritmos específicos são utilizados para controle da pressão e da temperatura de acordo com as condições desejadas (MACKERELL, 2001).

### 4.4.1 Termostato de Langevin

Para simular utilizando o “*ensemble*” canônico (NVT), onde T é mantido constante durante a simulação, um “termostato” é usado para manter as velocidades das partículas em acordo com a distribuição de Maxwell-Boltzmann na temperatura de interesse. Nesse tipo de “*ensemble*”, o sistema não se encontra mais isolado, e as partículas não interagem mais apenas entre si, mas também com um “banho térmico”, onde o sistema é acoplado a um reservatório para ajudar a controlar a temperatura. A interação com o “banho” pode ocasionar uma aceleração das partículas ao fornecer uma força adicional, ou absorver energia para desacelerá-las com um amortecimento ou atrito. Dessa forma, a equação de movimento é modificada para incluir essa nova interação (ALLEN; TILDESLEY, 1996).

Em dinâmica molecular existe diferente forma de se implementar esse “banho térmico”, sendo o Termostato de Langevin o algoritmo usado nas simulações executadas nesses trabalho. Esse termostato, baseado na equação de movimento homônima, é um método estocástico de simulação, onde cada partícula fica sujeita a forças aleatórias que fazem o papel de colisões com partículas do banho térmico, e de uma força dissipativa que depende de um coeficiente de atrito e da velocidade dos átomos, de acordo com a seguinte equação:

$$m_i \ddot{\mathbf{r}}_i = -\frac{\partial U(\mathbf{r}^N)}{\partial \mathbf{r}_i} - \xi m_i \dot{\mathbf{r}}_i + R_i \quad (4.15)$$

onde  $\xi$  é o coeficiente de atrito de Langevin e  $R_i$  é uma força aleatória definida a partir de uma distribuição gaussiana centralizada em zero com variância,  $\sigma^2 = 2m_i \xi k_B T / \Delta t_{MD}$ , sendo  $k_B$  a constante de Boltzmann e  $\Delta t_{MD}$  o “*timestep*” da MD.

A imposição do atrito tem o efeito de retirar energia cinética das partículas, enquanto a força tem o papel de fornecer energia acelerando os átomos. No AMBER define-se a temperatura que se deseja manter o sistema e a frequência média de colisão com o banho,

sendo nesse trabalho de  $1,0 \text{ ps}^{-1}$ , um valor baixo para que as trajetórias das partículas não sejam influenciadas diretamente pela flutuação de temperatura imposta ao sistema. Quando esse algoritmo é utilizado numa simulação a NVT, o sistema percorre o espaço fásico segundo a superfície de energia potencial (SEP) de acordo com as equações de movimento de Newton, até que as velocidades de todos os átomos sejam recalculadas considerando as forças introduzidas pelas colisões, segundo a **Equação 4.15**, deslocando-o para uma SEP diferente, que o sistema passa a explorar até que ocorra uma nova iteração, gerando uma amostragem final canônica (ALLEN; TILDESLEY, 1996).

#### 4.4.2 Barostato de Berendsen

Para se recriar de forma fidedigna processos que ocorrem à temperatura e pressão constantes, é necessário amostrar um “ensemble” NPT. O método implementado nesse trabalho para controle da pressão, é feito com o barostato de Berendsen (**BERENDSEN et al., 1984**) que acopla a simulação a um “banho de pressão”, semelhantemente ao termostato de Langevin, ao adicionar um termo de interação à equação de movimento que permite um controle de pressão de acordo com a seguinte equação:

$$\frac{d\mathfrak{P}}{dt} = (P - \mathfrak{P})/t_p, \quad (4.16)$$

onde  $\mathfrak{P}$  é pressão instantânea do sistema,  $P$  é a pressão desejada (do “banho”) e  $t_p$  é o tempo de relaxamento. A pressão é obtida em uma MD pelo teorema do Virial:

$$P = \frac{Nk_bT}{V} + \frac{1}{3V} \langle \sum \mathbf{r}_i \cdot \mathbf{F}_i \rangle \quad (4.17)$$

onde o primeiro termo é uma componente cinética da pressão que independe das interações do sistema, logo, a mesma pressão caso o sistema fosse um gás ideal; e o segundo, provém da contribuição das forças internas que atuam em cada átomo do sistema. Como as forças são calculadas a cada etapa de iteração, obter a pressão por meio desse teorema é conveniente do ponto de vista computacional (**ALLEN; TILDESLEY, 1996**).

A cada passo de tempo, o volume de simulação é redimensionado por um fator  $\chi$ , e as coordenadas do centro de massa das moléculas por um fator  $\chi^{1/3}$ , em que  $\chi$  é dado por:

$$\chi = 1 - \beta_\tau \frac{\delta t}{t_p} (P - \mathfrak{P}), \quad (4.18)$$

onde  $\beta_\tau$  representa a compressibilidade isotérmica do sistema, sendo um escalar cujo valor não precisa ser exato, mediante que pode ser incluído na constante de tempo de relaxamento,  $t_p$ . Esta técnica permite ajustar a densidade do solvente (água) em simulações biomoleculares, sendo utilizada para o relaxamento do sistema, tópico abordado no **Capítulo 6**. Entretanto, o algoritmo de Berendsen não amostra o “ensemble” NPT corretamente, devido a isso utiliza-se a notação NpT para o “ensemble” gerado. Mas, a sua aplicabilidade é justificada durante as etapas de relaxação porque a pressão converge

mais rapidamente do que quando outros métodos são utilizados (**BERENDSEN *et al.*, 1984**).

## 5 CAMPO DE FORÇA

### 5.1 POTENCIAL DE INTERAÇÃO MOLECULAR

Um sistema molecular pode ter suas interações moduladas classicamente quando o comprimento de onda térmico de *de Broglie* ( $\Lambda$ ) for menor que a distância média entre as partículas interagentes (MARTÍNEZ; BORIN; SKAF, 2007).

$$\Lambda = \frac{h}{(2\pi mk_B T)^{1/2}} \quad (5.1)$$

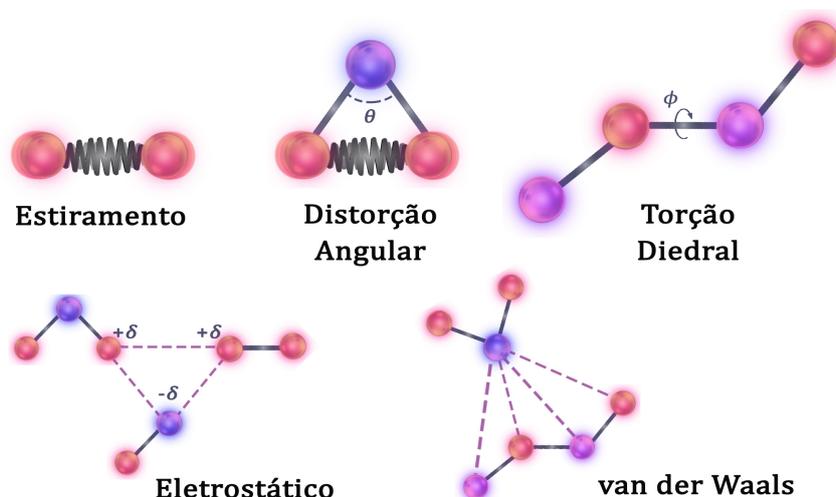
No caso das simulações de **MD**, a distância mínima entre os átomos é bem maior que 1 Å, e para um núcleo de carbono à temperatura ambiente, por exemplo, seu comprimento térmico é aproximadamente 0,3 Å, portanto, a aproximação clássica é válida. Para a grande maioria das biomoléculas, as ligações covalentes são restringidas, nessa temperatura, a distâncias que correspondem a conformações de mínimos de energia (**comprimento de equilíbrio**), e podem ser moduladas por um potencial harmônico. Entretanto, o mesmo não é verdade para o átomo de hidrogênio, que possui um  $\Lambda$  de 1 Å e uma energia vibracional de alta frequência quando ligado a átomos pesados (MARTÍNEZ; BORIN; SKAF, 2007). Devido a isso, um formalismo próprio para as ligações que envolvem átomos de hidrogênio são empregados nas simulações de dinâmica molecular, e serão discutidos no **Capítulo 6**.

Na **MD**, a energia potencial do sistema biomolecular é descrita em função da sua estrutura tridimensional ( $r^N$ ) por uma equação matemática intitulada campo de força, mostrada na **Equação 5.2**, que descreve as interações físicas que ditam as propriedades químicas do sistema e moldam a sua estrutura. Comumente são descritos 5 termos principais, que são divididos em modos que englobam interações em que os átomos envolvidos estão ligados covalentemente, ou seja, restringidos por ligações químicas, chamados de potenciais ligados ( $V_b(l) + V_a(\theta) + V_t(\phi)$ ), e, em interações entre átomos que não estão diretamente ligados entre si, denominados de potenciais não-ligados ( $V_{el}(r_{ij}) + V_{vdW}(r_{ij})$ ).

$$U(r^N) = V_b(l) + V_a(\theta) + V_t(\phi) + V_{el}(r_{ij}) + V_{vdW}(r_{ij}) \quad (5.2)$$

O primeiro termo do campo, mostrado na Equação 5.3, propõe uma representação da vibração da ligação química entre dois átomos, modelada como uma mola por um potencial harmônico, conhecido como lei de Hooke, onde:  $k_i$  é a constante da mola, que adota um caráter de força de ligação;  $l_{i,0}$  é o comprimento de referência da ligação, que é a distância internuclear de equilíbrio do modo normal de vibração que simula o estiramento da ligação entre dois átomos; e  $l_i$ , o quanto a ligação se distancia do valor de referência, provocando um aumento de energia no sistema.

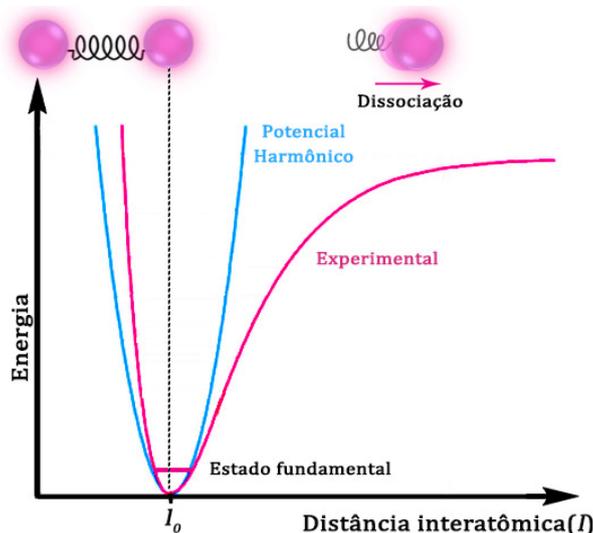
Figura 5.1 – Termos do campo de força



Os principais termos de interação de um campo de força: Estiramento ( $V_b$ ), distorção angular ( $V_a$ ), torção diedral ( $V_t$ ), interação eletrostático ( $V_{el}$ ) e van der Waals ( $V_{vdW}$ ); propostos para descrever o potencial total de energia de um sistema molecular. Fonte: Figura elaborada pela autora.

$$V_b(l) = \sum_{\text{Ligação}} \frac{k_i}{2} (l_i - l_{i,0})^2 \quad (5.3)$$

Figura 5.2 – Energia de estiramento de um sistema diatômico



Curva experimental de energia potencial de um sistema diatômico em magenta. À temperatura de eventos biológicos, a maioria das moléculas vão estar no estado fundamental e apresentam um estiramento próximo ao descrito pela lei de Hooke em azul. Fonte: Figura elaborada pela autora.

O potencial harmônico é uma representação aproximada da parte inferior do poço da curva de energia potencial obtida experimentalmente para sistemas diatômicos, exem-

plificada na **Figura 5.2**, e corresponde a distâncias de ligação em moléculas no estado fundamental, o que é previsto à temperatura (ambiente) em que as simulações são executadas. Essa aproximação é considerada razoável, pois, devido as restrições impostas às ligações, uma simulação não contempla o rompimento das mesmas, logo não é necessário que a curva se ajuste à dissociação dos átomos, além de ser mais fácil e barato de se computar. Além disso, a falta de termos anarmônicos que melhor moldem o potencial só é crítica quando se quer ajustar com maior precisão as energias para moléculas altamente deformadas ou quando se tem o interesse de reproduzir as frequências vibracionais das ligações.

O segundo termo, mostrado na **Equação 5.4**, descreve a distorção angular entre três átomos ligados, modelados também pela lei de Hooke, a partir de um ângulo de referência,  $\theta_0$  (quando todos os outros termos do campo de força são nulos),  $k_\theta$  a constante da mola e  $\theta$  o quanto o ângulo se distancia do valor referencial.

$$V_a(\theta) = \sum_{\text{Ângulo}} \frac{k_\theta}{2} (\theta - \theta_0)^2 \quad (5.4)$$

O último potencial intramolecular representa a periodicidade das torções diedrais entre uma sequência de quatro átomos ligados A-B-C-D, em relação a rotação da ligação B-C, equacionada através de uma expansão em série de Fourier em cosseno:

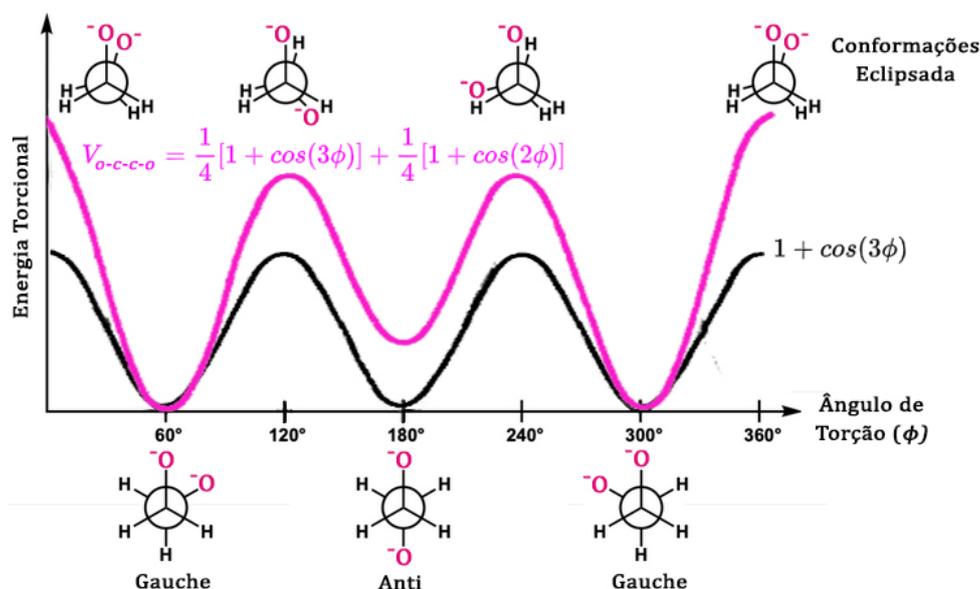
$$V_t = \sum_{n=0}^N \frac{k_{\phi_n}}{2} [1 + \cos(n\phi - \gamma)], \quad (5.5)$$

onde  $k_{\phi_n}$  é a altura das barreiras de energia relativa à rotação da ligação B-C;  $n$  é a multiplicidade do potencial e indica o número de pontos mínimos da função quando a ligação é rotacionada em  $360^\circ$ .  $\gamma$  é o ângulo de fase, que determina quando o ângulo de torção, ( $\phi$ ), passa pelo seu valor mínimo.

Nos campos de força do *software* AMBER (**CASE et al., 2005**) utilizado nesse trabalho, ff14SB (**MAIER et al., 2015**) para a proteína e GAFF (**WANG et al., 2006**) — *general amber force field* — para os ligantes, se utiliza parâmetros diedrais genéricos, de acordo com os átomos da ligação central da rotação  $\phi$ , e não abrangem os átomos terminais. Além disso, a expansão em cosseno envolve apenas o primeiro termo da série de Fourier para cada estiramento, exceto no caso especial da sequência O-C-C-O (**LEACH, 2001**). Para o fragmento  $OCH_2 - CHO_2$  presente em muitos peptídeos, o diedro tende a adotar a conformação gauche, e para essa sequência ser modelada corretamente dois termos são utilizados, exclusivamente, como demonstrado na **Figura 5.3**.

As distâncias e ângulos de ligação, geralmente, são graus de liberdade muito rígidos, sendo muita vezes necessário que uma quantidade expressiva de energia seja fornecida ao sistema para se observar deformações significativas em relação aos valores de referência, que são bem conhecidos na literatura. Variações na energia potencial do sistema que provocam mudanças na estrutura tridimensional são devidas, majoritariamente, à complexa

Figura 5.3 – Torção diedral para a sequência O-C-C-O



Em magenta, a função diedral para a sequência O-C-C-O é descrita por dois termos de uma função em cosseno, com os mínimos em  $\theta = 60^\circ$  e  $300^\circ$ . Em preto é descrito uma função genérica de torção para um caso onde  $k_{\phi_n} = 2$ ,  $n = 3$  e  $\gamma = 0$ . Fonte: Figura elaborada pela autora.

inter-relação entre o potencial de torção e as interações intermoleculares, já que essas influenciam a rotação dos átomos 1,4 diedrais, e, conseqüentemente, no valor da constante  $k_{\phi_n}$ .

As interações intermoleculares ocorrem entre elementos que não estão covalentemente ligados entre si, mas a interação se manifesta através do espaço sob influência da distância entre os envolvidos, divididas em dois termos não covalentes. O primeiro representa as interações eletrostáticas, e descreve a interação interatômica que surge devido a uma redistribuição interna dos elétrons numa molécula gerando partes positivas e negativas. Essa distribuição heterogênea da densidade eletrônica em uma molécula é representada por um valor de carga parcial atômica. Admitindo que essas cargas estão restritas ao centro do núcleo atômico, o modelo eletrostático de carga pontual pode ser utilizado para se obter a energia potencial, modelado pela lei de Coulomb:

$$V_e(r_{ij}) = \sum_{i=1}^N \sum_{j=i+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}, \quad (5.6)$$

onde:  $q_i$  e  $q_j$  são as cargas parciais de dois átomos  $i$  e  $j$ , respectivamente;  $r_{ij}$  é a distância entre eles e  $\epsilon_0$  é a permissividade elétrica no vácuo.  $N$  é o número de cargas pontuais presentes entre moléculas diferentes, mas o potencial também é calculado para diferentes cargas parciais localizadas em diferentes partes de uma mesma molécula.

A carga parcial atômica não é uma grandeza quanticamente observável, pois não é um autovalor de um operador quântico, não sendo possível medi-la experimentalmente pois

não é uma grandeza física bem definida. Entretanto, ela pode ser obtida indiretamente a partir de medidas experimentais de observáveis que resultam diretamente do efeito da carga parcial ou que possam com ela ser correlacionadas, como o **momento de dipolo** ou pelo gradiente de um campo elétrico. Devido a isso, existem diferentes métodos para se determinar cargas parciais atômicas, sendo o modelo denominado Potencial Eletrostático Restrito - “*Restrained Electrostatic Potential*” (BAYLY *et al.*, 1993) o método empregado pelo AMBER.

Nesse método, as cargas parciais são derivadas a partir do potencial eletrostático molecular quântico (ESP), calculado para um determinado número de pontos em uma malha regular distribuída no entorno das moléculas. Onde, esse potencial é definido como a força eletrostática que atua sobre uma carga positiva de valor unitário em uma dada posição da molécula. A escolha de reproduzir o campo ESP no entorno de uma molécula descreve mais acuradamente as interações coulombianas entre diferentes cargas parciais considerando dois corpos. Além disso, o valor do potencial ESP em um ponto específico no espaço ao redor de uma molécula é um observável quântico, pois é o valor esperado para o operador de um elétron.

Em seguida, como o potencial ESP criar as cargas fora da superfície de van der Waals das moléculas, e o campo de força busca reproduzir parametricamente cargas centralizadas em átomos, o mesmo é ajustado a um potencial *ab initio* que considera que toda a densidade eletrônica associada a um átomo está localizada onde a função de base atômica usada está centralizada. Sendo no RESP, esse potencial obtido pelo método de análise populacional de Mulliken<sup>1</sup>. As cargas ajustadas são então restringidas por um termo hiperbólico, de forma a aumentar a transferibilidade (seção 5.2) do parâmetro e reduzir erros associados às cargas de átomos internos que se localizam longe da superfície de van der Waals da molécula, onde as cargas são projetadas pelo ESP (JANEČEK *et al.*, 2021).

O último termo do campo de força descreve as interações de van der Waals modeladas pelo potencial de Lennard-Jones. Essas interações se originam do equilíbrio entre forças de carácter atrativo e de forças repulsivas que atuam em curtas distâncias. As forças atrativas resultam de dipolos instantâneos que surgem devido às flutuações das densidades eletrônicas, onde tais oscilações induzem um dipolo em átomos vizinhos, dando origem a momentos dipolares e multipolares que criam um efeito de atração entre as partículas. Já as forças repulsivas surgem pela aproximação significativa dos átomos (superposição), a ponto de haver repulsão eletrostática entre os envolvidos, tendo origem no princípio de

---

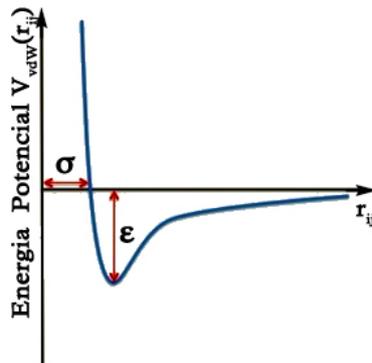
<sup>1</sup> O ajuste não utiliza as cargas geradas pelo método de Mulliken, e sim as funções de densidade.

exclusão de Pauli<sup>2</sup>. O potencial de Lennard-Jones é expresso pela Equação 5.7:

$$V_{vdW}(r_{ij}) = \sum_{i=1}^N \sum_{j=i+1}^N 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right], \quad (5.7)$$

onde  $\varepsilon_{ij}$  é a energia associada à interação entre os átomos  $i$  e  $j$  e representa a profundidade do poço da curva potencial. Já  $\sigma_{ij}$  é o comprimento referente à distância em que o potencial é nulo.

Figura 5.4 – Potencial de Lennard-Jones



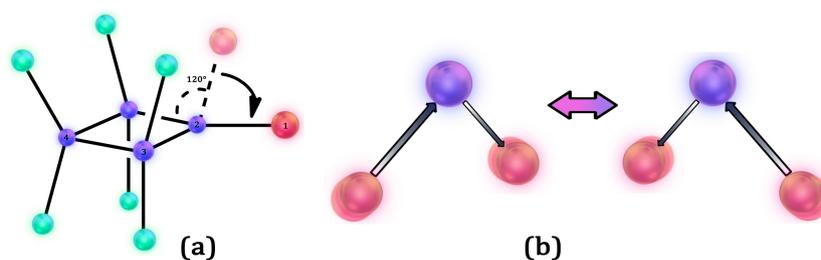
Quanto mais acentuada a profundidade do “poço”,  $\varepsilon_{ij}$ , mais forte é a interação entre as duas partículas. Fonte: Figura elaborada pela autora.

Os termos apresentados na **Equação 4.10** são comuns a todos os campos de força, e se mostram suficientes para reproduzir as propriedades estruturais necessárias em biomoléculas. Entretanto, outros termos de interações podem ser encontrados de acordo com o software utilizado, como potenciais de termos cruzados e de torção diedral imprópria. Os potenciais de termos cruzados modelam o acoplamento do movimento entre os estiramentos, ângulos e torção. Um dos principais exemplos envolve as interações do tipo estiramento-estiramento entre duas ligações sequenciais modulados pelo potencial de Urey-Bradley, que é descrito, a partir das distâncias entre os átomos 1 e 3, por um potencial harmônico. Já a torção diedral imprópria decorre da necessidade de ajustar a planaridade e quiralidade de grupos de 4 átomos, onde três estão confinados em algum tipo de ciclo, e o último átomo acaba adotando uma conformação fora do plano do grupo. Ambos os termos são exemplificados na **Figura 5.5**.

As interações que modulam os potenciais de um campo apresentam grandes correlações entre si, e a falta de tais termos não implicam em erros significativos nas simulações geradas pelo AMBER, ocorrendo uma compensação paramétrica devido a interdependência que as dinâmicas de biomoléculas apresentam em seus movimentos. Entretanto, essa variação nos componentes inclusos por diferentes campos faz com que os parâmetros de

<sup>2</sup> O princípio afirma que dois férmions idênticos não podem possuir o mesmo estado quântico em um mesmo sistema.

Figura 5.5 – Torção diedral imprópria e termo de correlação cruzada



(a) Torção imprópria, que corrige o átomo 1 que se projeta fora do plano da molécula. (b) Termos de correlação cruzada tipo estiramento-estiramento para uma sequência de 3 átomos. Fonte: Figura elaborada pela autora.

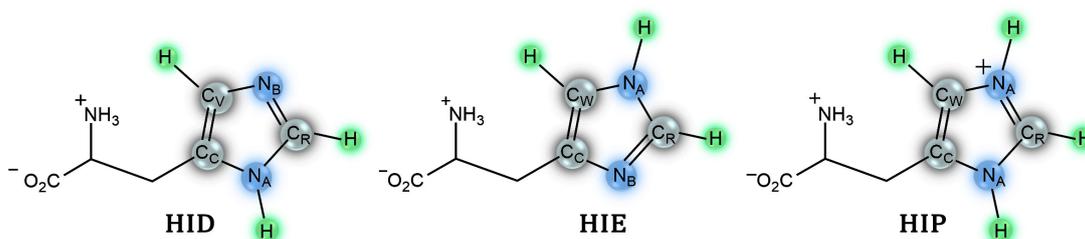
campos de força não sejam considerados transferíveis entre si. Outro motivo dessa intransferibilidade, envolve um tipo de classificação atribuída aos átomos de um sistema, nos quais os parâmetros são expressos, que envolvem além do número atômico, a hibridização e informações a respeito do ambiente químico no entorno molecular, chamado de tipo atômico (“*atom type*”). Essa categorização varia entre diferentes campos de força, e se relaciona com o tipo de sistema que o software se propõe a estudar. No caso, o AMBER foi desenvolvido, inicialmente, para o estudo específico de proteínas e DNA, logo, os tipos atômicos são específicos para aminoácidos e bases nitrogenadas. Conseqüentemente, a uma determinada molécula são atribuídos tipos atômicos com determinados parâmetros em um campo de força, mas em outro, receber uma identificação diferente, acarretando outros valores para as constantes do potencial.

Por exemplo, o campo de força ff14SB (MAIER *et al.*, 2015), utilizado nesse trabalho para modular o potencial da proteína, define 50 tipos atômicos diferentes para cada átomo dos 20 aminoácidos naturais. Aminoácidos que apresentam diferentes possibilidades de protonação, como as histidinas, recebem uma atribuição de tipos atômicos diferentes. A **Figura 5.6** mostra como os átomos de carbono do anel da histidina mudam de tipo atômico de acordo com o ambiente químico, onde para um campo de força para moléculas orgânicas simples seriam todos atribuídos o mesmo tipo: carbono  $sp^2$ . Implementar uma tipificação específica para cada átomo permite a atribuição de parâmetros que melhor descrevem as distâncias, cargas parciais etc. e conseqüentemente as interações do sistema, produzindo melhores correlações experimentais.

## 5.2 PARAMETRIZAÇÃO E TRANSFERIBILIDADE

A forma como os parâmetros do campo de força são obtidos também varia entre diferentes programas. Comumente, esses valores são parametrizados de forma a replicar propriedades termodinâmicas ou estruturais, como entalpia de vaporização, densidade e espectros vibracionais. Mas devido à complexidade de sistemas biomoleculares e muitas

Figura 5.6 – Tipos atômicos para o anel de histidinas

**Tipos atômicos do AMBER:**

**C<sub>c</sub>:** Carbono  $sp^2$  aromático em anel de 5 membros com 1 substituinte e próximo a um nitrogênio;

**C<sub>v</sub>:** Carbono  $sp^2$  aromático em anel de 5 membros ligado a 1 N (nitrogênio) e a um hidrogênio;

**C<sub>w</sub>:** Carbono  $sp^2$  aromático em anel de 5 membros ligado a 1 N-H e a um hidrogênio;

**C<sub>r</sub>:** Carbono  $sp^2$  aromático em anel de 5 membros entre 2 nitrogênios e ligado a um hidrogênio (em histidina);

**N<sub>A</sub>:** Nitrogênio  $sp^2$  em anel de 5 membros com um hidrogênio ligado;

**N<sub>B</sub>:** Nitrogênio  $sp^2$  em anel de 5 membros com par de elétrons livres;

Diferentes estados de protonação para a histidina implementados pelo AMBER: protonada na posição  $\delta$  (HID), protonada na posição  $\epsilon$  (HIE) e duplamente protonada (HIP). Para cada caso, os átomos recebem uma tipificação que traduz o ambiente químico que estão inseridos. Fonte: Figura elaborada pela autora.

vezes, à indisponibilidade de tais dados experimentais, a parametrização no AMBER é feita a partir de cálculos quânticos *ab initio* que geram uma curva de energia para os quais os potenciais do campo são ajustados, sendo um processo de elevada complexidade, que foge do escopo desse trabalho.

Uma propriedade importante nos campos de força, e que possibilita sua grande aplicabilidade em diferentes sistemas moleculares é a transferibilidade. Em que, a partir de parâmetros otimizados para um pequeno grupo molecular alvo, as mesmas constantes podem ser transferidas para sistemas de moléculas com esse mesmo grupo comum de átomos. Essa relação emerge, do fato das frequências de vibrações moleculares definidas para um sistema de coordenadas internas não variarem significativamente de uma molécula para outra, onde essa relação se estende devido à correlação entre os termos do potencial.

## 6 SIMULAÇÃO COMPUTACIONAL DE DINÂMICA MOLECULAR

### 6.1 CONDIÇÕES DE CONTORNO PERIÓDICAS

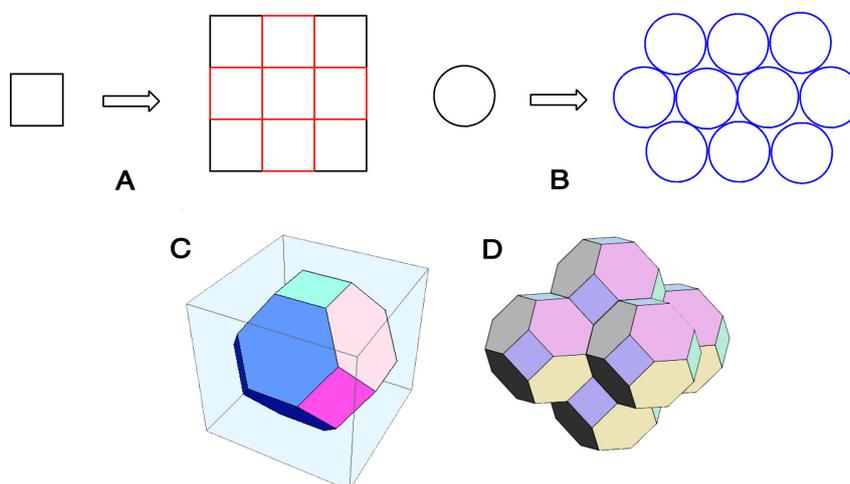
Antes de uma simulação de dinâmica molecular ser produzida, o sistema precisa ser configurado para representar de forma realista o seu ambiente químico natural, sendo para sistemas biológicos uma solução. Comumente, o sistema parte de estruturas cristalográficas que representam a célula unitária da biomolécula, apresentando apenas algumas águas de hidratação na estrutura sólida. Para se recriar um ambiente solvatado, de forma a exibir propriedades inerentes à matéria condensada, um número expressivo de moléculas do soluto e de solvente precisariam ser inseridas à caixa de simulação e que devem ser computadas.

Entretanto, adicionar indiscriminadamente moléculas de soluto e solvente à caixa produz artefatos indesejados à simulação, conhecidos como efeito de borda, que decorrem da interação das moléculas com o vácuo circundante. A interação com a extremidade reproduz o comportamento de uma gotícula, pois há um número maior de partículas que interagem com a borda do que interações no seio do fluido. Dessa forma, as flutuações de interação com vácuo sobrepõem as do seio da solução, a ponto de moléculas do solvente evaporarem da caixa. Porém, como a simulação de um sistema macroscópico, da ordem de  $10^{24}$  átomos, não é plausível computacionalmente, condições de contorno periódicas - “*Periodic Boundary Conditions*” (PBC) são utilizadas para se manter um número menor de partículas, de  $10^4$  a  $10^6$ , e reduzir o efeito de borda. No método PBC, as partículas que estão sendo simuladas são colocadas em uma caixa, que tem então suas paredes cercadas por replicadas virtuais em todas as três dimensões e indefinidamente, formando uma matriz periódica artificialmente infinita de imagens que permite a obtenção de propriedades volumétricas “*bulk*” de um fluido.

A escolha do formato da caixa para reproduzir um contorno periódico PBC exige que, ao se adicionar as réplicas, que seja preenchido todo o espaço no entorno da caixa central. Este requisito restringe as formas possíveis de serem usadas, mas o fato de uma forma preencher o espaço, significa que é possível replicá-la em diferentes direções. Por esse motivo, não existem caixas de simulação esféricas, mesmo que proteínas globulares fossem melhor representadas nesse tipo de simetria. Apesar da forma cúbica ser mais fácil de programar as periodicidades e o efeito “torus” das fronteiras, simulações de sistema biomoleculares em solução utilizam como caixa um dodecaedro rômbo ou o octaedro truncado. Isso porque, a célula unitária de uma caixa cúbica é muito maior do que as outras duas estruturas mencionadas. Logo, usar uma caixa octaédrica truncada ou um dodecaedro rômbo reduz significativamente a quantidade de solvente que precisa ser inserido, gerando uma economia no tempo computacional necessário. Uma segunda

razão, é que o dodecaedro rômbo e o octaedro truncado têm uma forma mais próxima de uma esfera do que o cubo, tornando o espaço mais simétrico, entretanto, o segundo é preferencial por ser mais fácil de implementar computacionalmente. Os critérios que ilustram a escolha de uma caixa de simulação são mostrados na **Figura 6.1**.

Figura 6.1 – Escolha da caixa de simulação



A: Contorno periódico de um quadrado. B: Contorno periódico de um círculo, mostrando que o espaço no entorno não é completamente preenchido, e por isso as PBC não podem ser empregadas. C: Um octaedro truncado dentro de um cubo, evidenciando a redução de solvente necessário para preencher a caixa. D: uma caixa periódica de octaedros truncados. Fonte: Figura elaborada pela autora.

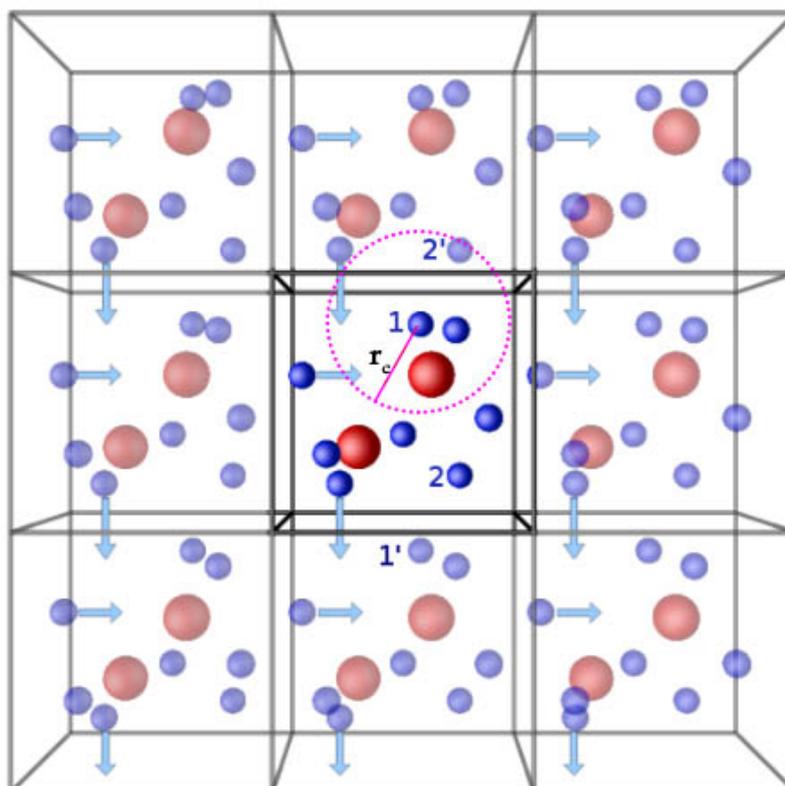
As moléculas nas réplicas se movem identicamente às da caixa original, e de modo que, quando um átomo cruza a fronteira de uma caixa, um outro átomo idêntico entra pela face oposta com a mesma velocidade, de forma que a densidade de partículas ( $N/V$ ) em todas as imagens permaneçam constantes. Entretanto, simular réplicas sincronizadas e idênticas que interagem entre si impõe uma artificialidade incondizente com a natureza de sistemas líquidos, mediante que nas interações de longo alcance o sistema irá interagir com versões de si mesmo, impondo uma simetria de estrutura de célula unitária a um fluido. Para reduzir esse efeito duas abordagens são utilizadas para lidar com as interações intermoleculares de curto e de longo alcance entre as caixas de simulação.

### 6.1.1 Interações de curto alcance: convenção de imagem mínima

A primeira abordagem, chamada de convenção da imagem mínima, propõe limitar as interações entre a caixa original e suas réplicas, ao estabelecer que cada átomo só pode interagir com uma única réplica de um segundo átomo do sistema — a que estiver mais próxima. Para isso, é definido o raio de corte,  $r_c$ , uma distância máxima para as quais as interações entre pares atômicos seriam computadas. O potencial entre átomos afastados numa distância maior que o raio de corte é considerado de magnitude desprezível. Como

consequência dessa imposição, temos que na **Figura 6.2**, as partículas 1 e 2, localizadas na mesma caixa, não interagem entre si já que sua distância seria maior que o raio de corte. Entretanto, elas interagem com a imagem virtual da partícula na caixa vizinha, ou seja, o átomo 1 interage com 2' e o átomo 2 com 1'.

Figura 6.2 – Condições de contorno periódicas.



Condições de contorno periódicas. Quando um átomo cruza as fronteiras da caixa de simulação central, um outro átomo entra pela face oposta com a mesma velocidade. Em rosa, é mostrado o raio de corte para o átomo 1. Fonte: Figura elaborada pela autora.

### 6.1.2 Interações de longo alcance: somas de Ewald

Na convenção de imagem mínima as interações de van der Waals são truncadas no raio de corte para impedir interações de longo alcance, de forma que a caixa central consiga interagir com as réplicas de formas distintas. Entretanto, as interações eletrostáticas, por serem de longo alcance, não podem ser truncadas. Como consequência, um átomo da caixa central interage com todos os outros átomos do sistema e com todos os átomos virtuais das infinitas réplicas, incluindo sua própria imagem em outras caixas. Para o sistema periódico, a energia potencial eletrostática é da forma:

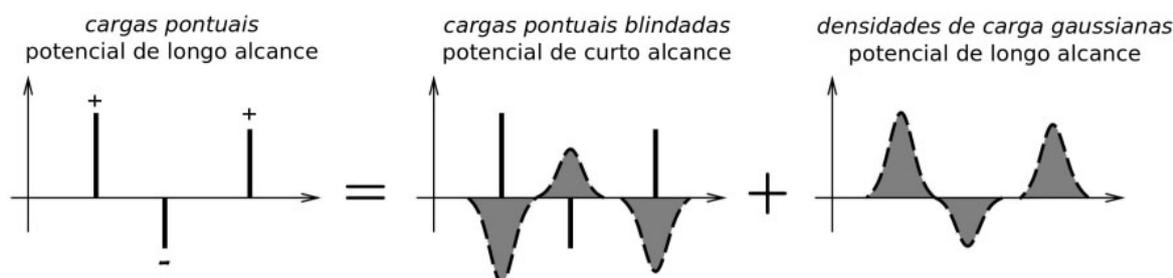
$$V_{el} = \frac{1}{2} \sum_{\mathbf{n}}' \sum_{j=1}^N \sum_{i=1}^N \frac{q_i q_j}{\epsilon |\mathbf{r}_{ij} + \mathbf{n}|}, \quad (6.1)$$

onde,  $\mathbf{n}$  é um vetor que determina a posição de cada caixa de simulação, e o apóstrofo no somatório de  $\mathbf{n}$  indica que  $i \neq j$  para  $n = 0$ , ou seja, a partícula não interage consigo mesma, mas pode interagir com sua imagem virtual. A utilização dessa equação nas simulações é impraticável, pois a série converge muito lentamente e é influenciada pela ordem em que os termos positivos e negativos são somados (convergência condicional). Para contornar esse problema, é implementado no AMBER um algoritmo baseado nas somas de Ewald (**EWALD, 1921**), que converte a **Equação 6.1** em duas séries distintas que convergem mais rapidamente, se baseando na relação matemática (**LEACH, 2001**):

$$\frac{1}{r} = \frac{f(r)}{r} + \frac{1 - f(r)}{r}, \quad (6.2)$$

onde,  $f(r)$  é escolhida de forma a lidar com a variação rápida de  $1/r$  que surge em interações coulombianas a curtas distâncias, e com o decaimento lento de  $r$  em longas distâncias. Dessa maneira, a primeira função é construída para ser uma função de variação rápida de curto alcance que pode ser tratada pela soma direta de termos, sendo possível empregar um raio de corte; enquanto a segunda, é uma função periódica suave de longo alcance, adequada para solução por métodos de transformada de Fourier (**ALLEN; TILDESLEY, 1996**).

Figura 6.3 – Método de soma de Ewald

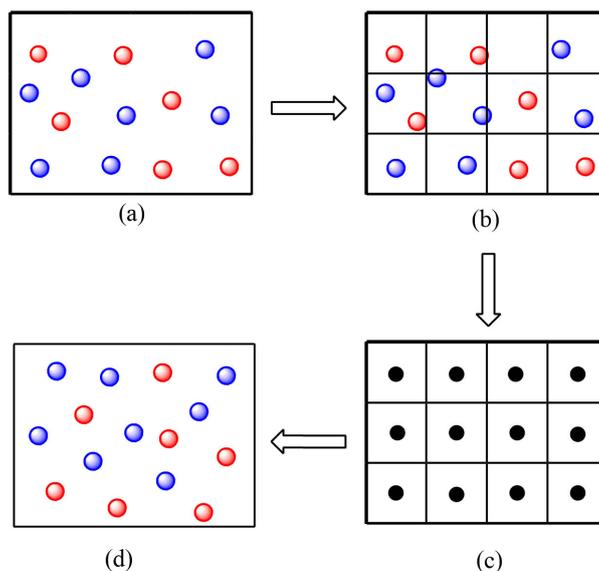


Decomposição de cargas pontuais em densidades de carga. Fonte: Figura adaptada de **Griebel, Knapek e Zumbusch (2007)**.

No método original de Ewald, a cada carga, representada na **Figura 6.3** por uma barra positiva ou negativa, são adicionadas uma densidade de carga gaussiana de magnitude equivalente mas de sinal contrário - de forma a blindar o potencial gerado pela carga pontual. Quando essa distribuição é incorporada à equação (primeira função), o potencial associado às cargas blindadas obtido converge rapidamente e pode ser truncado a partir de uma distância de corte e computado utilizando a convenção da imagem mínima. Para compensar a densidade de distribuição adicionada anteriormente, uma distribuição gaussiana de mesma magnitude que as cargas parciais e de mesmo sinal é adicionada para que o efeito líquido total seja nulo. O potencial final gerado, de longo alcance, converge rapidamente quando somado no espaço recíproco (**LEACH, 2001**).

No algoritmo “*Particle mesh Ewald*” (PME) (ESSMANN *et al.*, 1995) implementado no AMBER e esquematizado na **Figura 6.4**, as densidades de carga são discretizadas numa malha tridimensional, onde os potenciais e forças são calculados em cada ponto da grade usando transformadas de Fourier discretas. Em seguida, as forças são interpoladas de volta às partículas, para que as coordenadas sejam atualizadas. É importante frisar, que apesar das infinitas imagens virtuais do sistema, apenas as posições e momento das partículas que compõem a caixa central são contabilizadas.

Figura 6.4 – Esquema do algoritmo “*Particle mesh Ewald*”



Um sistema bidimensional de cargas parciais (a) é interpolado em uma malha (b), onde os potenciais e forças são calculados para cada ponto na grade (c). Por fim, as forças são interpolados de volta aos átomos para gerarem novas coordenadas. Fonte: Figura elaborada pela autora.

## 6.2 MÉTODO DE INCLUSÃO DE SOLVENTE EXPLÍCITO

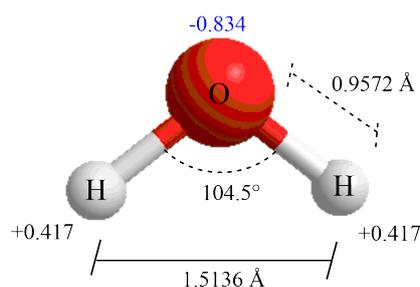
Em simulações de MD, o solvente pode ser adicionado implicitamente, ou seja, é considerado uma imitação da sua presença dentro da caixa de simulação ao supô-lo como um meio contínuo que apresenta a mesma constante dielétrica do solvente que se pretende reproduzir, resultando numa influência média, mas que acelera o tempo computacional por reduzir o número de moléculas do sistema. Entretanto, esse modelo não consegue reproduzir os detalhes microscópicos da interface proteína-água; a ponto de os ensembles produzidos diferirem significativamente dos gerados por solvente explícito, não sendo capazes de identificar o estado nativo da proteína.

Quando se deseja reproduzir adequadamente uma fase condensada e as interações intermoleculares solvente-solvente e solvente-soluto, o método de inclusão de solvente explícito é o mais recomendado. Nesse caso, moléculas de água são incluídas à caixa

composta por átomos de hidrogênio e oxigênio. Entretanto, com essa adição, as moléculas do solvente representarão mais de 80% dos átomos da simulação, e as interações água-água irão dominar o custo computacional. Dessa forma, é necessário a utilização de um modelo para descrever o solvente de forma a acelerar os cálculos computacionais, mas mantendo a acurácia.

Na literatura, existem muitos modelos de água possíveis de serem utilizados em dinâmica molecular, que foram obtidos empiricamente para reproduzir um certo número de propriedades do “*bulk*”. Alguns modelos reproduzem bem as energias de hidratação das proteínas, enquanto outros predizem melhor a estrutura geométrica da molécula, mas não exibem uma energia livre de hidratação adequada. E de fato, nenhum dos modelos de água atuais conseguem reproduzir com acurácia todas as principais propriedades da água em “*bulk*” simultaneamente, mediante a possibilidade de se assumir estruturas rígidas ou flexíveis, permitindo todos os graus de liberdade para as ligações OH e ângulo de ligação HOH, ou não. Entretanto, é recomendado ao aplicar um campo de força, usar o modelo de água para o qual esse campo foi especificamente desenvolvido e testado. Dessa forma, como os campos de força do AMBER (ff14SB e GAFF) foram parametrizados utilizando o modelo TIP3P (“*transferable intermolecular potential 3P*”), o mesmo foi utilizado nesse trabalho. O modelo de água TIP3P (**JORGENSEN *et al.*, 1983**) descreve a molécula com uma geometria fixa e possuindo três locais de interação correspondentes aos átomos da molécula, com as interações eletrostáticas sendo moduladas pela lei de Coulomb, mas apenas o oxigênio possui os parâmetros de Lennard-Jones.

Figura 6.5 – Modelo de água TIP3P



Modelo de geometria fixa para a molécula de água TIP3P, com as respectivas distâncias interatômicas e densidade eletrônica dos átomos. Fonte: Figura elaborada pela autora.

### 6.3 ALGORITMO “*RATTLE*”

Como discutido anteriormente, o passo de tempo máximo permitido por uma simulação MD é determinado de acordo com a frequência máxima de vibração das ligações, sendo as ligações de hidrogênio com átomos mais pesados (H-X), da ordem de  $3000 \text{ cm}^{-1}$ ;

determinantes numa dinâmica, e limitam o “*time step*” para 0.5 fs. O Modelo TIP3P considera uma geometria fixa da água ao restringir as distâncias de ligação, mas é necessário fixar também todas as ligações de tipo H-X. Para garantir que essas distâncias permaneçam as mesmas durante a simulação é utilizado o algoritmo “*RATTLE*” (ANDERSEN, 1983), baseado no método originário “*SHAKE*” (RYCKAERT; CICCOTTI; BERENDSEN, 1977), criado para confinar as distâncias por meio de vínculos ao longo da simulação. Na mecânica Lagrangiana, um vínculo é definido como qualquer restrição de movimento que reduz o grau de liberdade do sistema. O algoritmo percorre cada par de átomos, ajustando-os para re-estabelecer o vínculo. O procedimento é iterado de forma cíclica até atingir uma convergência. A principal diferença com o “*SHAKE*” é que o “*RATTLE*” implementa adicionalmente o vínculo nas velocidades, sendo mais adequado para o uso com a versão de integrador de velocidade “*velocity-Verlet*”. A implementação de vínculos aos comprimentos de ligação em simulações de dinâmica molecular de proteínas permite aumentar o passo de integração em até quatro vezes, de acordo com Ryckaert, Ciccotti e Berendsen (1977), mantendo uma boa conservação de energia do sistema e sem sobrecarregar computacionalmente com o procedimento iterativo que exige uma convergência dos vínculos (FORESTER; SMITH, 1998).

#### 6.4 MINIMIZAÇÃO E EQUILIBRAÇÃO

A introdução da caixa de solvatação em formato de octaedro truncado ao redor da biomolécula de interesse ocorre em posições pré-determinadas pelo AMBER, com a inserção de águas do tipo TIP3P pré-equilibradas. Além disso, o sistema é neutralizado com a adição de íons de sódio e de cloreto próximos a regiões com alto potencial eletrostático. Essa etapa é chamada de configuração inicial do sistema, e engloba também a designação dos campos de força para todos os átomos presentes no arquivo **PDB**. Como o arquivo PDB usado nesse trabalho foi gerado por difração de raios X, há a ausência dos átomos de hidrogênio em toda a estrutura, já que devido a sua baixa densidade eletrônica, não consegue espalhar a radiação. Dessa forma, também são inseridos átomos de hidrogênio por toda a estrutura tridimensional. Devido ao fato dessas introduções ocorrerem sem levar em consideração a posição ou qualquer interação prévia, é necessário a realização de etapas de relaxamento antes da produção da dinâmica, de forma a garantir uma maior estabilização do sistema. Isso porque, por exemplo, a presença de lacunas entre o solvente e o soluto durante a dinâmica formam bolhas de “vácuo” que causam instabilidades à simulação. Além disso, átomos muito próximos uns dos outros, podem causar forças de repulsão que irão provocar deslocamentos atômicos drásticos, que podem ser difíceis de serem corrigidos durante a dinâmica, pois a simulação prosseguiria em uma direção irreal.

O relaxamento engloba dois procedimentos intercalados e distintos. O primeiro é a etapa de otimização de geometria, um procedimento que desloca os átomos com base nas

forças atuantes para localizar uma conformação de mínimo de energia na SEP, também conhecido como de etapa de minimização. O segundo é a etapa de equilíbrio, que aquece progressivamente o sistema até uma temperatura alvo e variando a retenção das ligações, sob condições de pressão constante e/ou volume constante. É recomendado que a relaxação seja executada em várias etapas diferentes com as restrições sendo liberadas gradualmente.

No AMBER, utiliza-se a combinação de dois algoritmos distintos - “*Steepest Descent*” e gradiente descendente - para a minimização da energia mecânica do sistema. A partir da conformação inicial do sistema, com o “*Steepest Descent*” são dados repetidos passos na direção contrária ao gradiente da função potencial, pois essa é direção da descida mais íngreme que leva ao mínimo local da função de energia em relação as coordenadas. Dessa forma, a equilibração pode partir de uma conformação mais estável. Entretanto, métodos mais robustos podem ser necessários, quando a dinâmica parte de uma estrutura que está longe de um ponto de mínimo. Mas, quando se parte de uma estrutura cristalográfica essa combinação de algoritmos é suficiente, pois o sistema se encontra próximo a um mínimo.

Nesse trabalho, inicialmente a caixa de simulação foi relaxada considerando apenas as moléculas do solvente e os íons livres, com todos os átomos do soluto (proteína, heme e ligantes) restringidos por uma constante de força num potencial harmônico, de forma a mantê-los fixos na mesma posição. Utilizar uma constante  $k$  alta indica que mover o átomo da posição de equilíbrio vai resultar numa força que atua para restaurá-lo à posição inicial, como uma mola rígida. A diferença entre restringir uma ligação e fixá-la por meio de um vínculo, é que o vínculo impõe uma contenção absoluta ao cálculo, enquanto a restrição introduz um viés energético que tende a forçar o cálculo em direção a restrição.

A segunda etapa de relaxação, considera apenas os carbonos alfa do “*backbone*” fixo, permitindo dessa forma, que as cadeias laterais, a coenzima e ligante relaxam com o solvente sem que haja uma translação da proteína dentro da caixa. Em seguida, a relaxação ocorre com todos os átomos livres, mas são feitos dois tipos de equilibração. Até então, a etapa de equilibração era feita em condições de NVT, onde o sistema era aquecido de 0K a 300K gradualmente, utilizando o termostato de Langevin para controle da temperatura. Dessa maneira, as ligações químicas eram relaxadas gradualmente para serem condizentes com o estado condensado. Mas com a suspensão das restrições, uma etapa em condições NpT é feita utilizando além do termostato para controle da temperatura, o barostato de Berendsen para controle da pressão, a fim de que a densidade e o volume da caixa possam relaxar, removendo as “bolhas” na caixa de simulação. A segunda equilibração parte das posições, velocidade e temperaturas salvas da etapa anterior e simula uma dinâmica curta, onde o objetivo central é o monitoramento do sistema, para garantir que as propriedades termodinâmicas durante a MD, flutuaram em torno de valores médios constantes. Pois assim, o sistema pode ser considerado apto para prosseguir a etapa de produção da dinâmica, onde as trajetórias serão geradas e posteriormente analisadas.

## 7 OBJETIVOS E METODOLOGIA

### 7.1 OBJETIVOS

O objetivo deste trabalho consiste na execução de simulações computacionais de dinâmica molecular da enzima OleT<sub>JE</sub> em solução com diferentes produtos dentro de sua cavidade, visando obter detalhes da interação da proteína com os compostos majoritariamente formados. Além da investigação da influência da mobilidade da alça F/G na presença desses ligantes, durante a etapa de saída do produto. Os sistemas simulados foram gerados a partir da estrutura cristalográfica da OleT<sub>JE</sub> de **PDB: 4140 (BELCHER *et al.*, 2014)**. Também foi investigado se a hidratação da cavidade poderia desempenhar um papel auxiliando a dissociação do produto para a solução, e quais regiões da enzima seriam responsáveis pelo processo. Dessa forma, foram realizados 300 ns de simulação em triplicata dos seguintes sistemas:

- A dinâmica molecular da OleT<sub>JE</sub> na presença de alcenos de cadeias compostas de 9 a 19 carbonos, ou seja, n-1, pois são resultados da descarboxilação de ácidos graxos;
- A dinâmica molecular da OleT<sub>JE</sub> na presença de ácidos graxos  $\beta$ -hidroxilados de cadeias compostas de 10 a 20 carbonos;
- A dinâmica molecular da mutante OleT<sub>JE</sub> Leu176Gly na presença de um alceno de 19 carbonos;
- A dinâmica molecular da mutante OleT<sub>JE</sub> Leu176Gly na presença do ácido graxo  $\beta$ -hidroxilado de 20 carbonos;

A partir, das dinâmicas moleculares, foram feitas as análises:

- Das flutuações estruturais (**RMSF**) das hélices F e G na presença de cada ligante;
- Das flutuações estruturais (**RMSF**) de cada resíduo que compõe a enzima;
- Das flutuações estruturais (**RMSF**) de cada ligante que compõe a enzima;
- De deslocamento quadrático médio (**RMSD**) de cada ligante durante as simulações;
- O RMSD das hélices F e G na presença de cada ligante;
- De solvatação;

O tempo de simulação escolhido nesse trabalho, de 300 ns, visa abranger a torção de cadeias laterais, a flexibilidade das alças e mudanças conformacionais próximas à estrutura cristalográfica da enzima na presença de diferentes produtos, mostrados na **Figura 2.8**.

Um citocromo CYP152 que compartilha 32,3% de identidade da sequência de aminoácidos com a OleT<sub>JE</sub> e que também exibe atividades de descaboxilação havia sido anteriormente simulado pelo grupo utilizando o mesmo tempo computacional. Como as simulações dessa enzima produziram boas correlações com dados experimentais, assumiu-se que 300 ns poderiam ser suficientes para amostrar significativamente o espaço de fases da OleT<sub>JE</sub> quando os produtos ainda se encontram dentro da cavidade de ligação. A etapa do ciclo catalítico que se pretende reproduzir é referente ao momento posterior à reação química, em que o produto ainda permanece dentro da cavidade reacional e a coenzima, heme, se mantém no estado pentacoordenado de alto spin. Dessa forma, pretende-se investigar como se dá a saída do produto de dentro da cavidade de reação, e se tal processo é de fato mediado pela alça F/G como sugerido por **Amaya *et al.* (2018)**.

## 7.2 METODOLOGIA

### 7.2.1 Configuração inicial do sistema

As simulações de dinâmica molecular foram realizadas a partir da estrutura cristalográfica da enzima P450 OleT<sub>JE</sub> com um substrato ácido graxo contendo uma cadeia de 20 carbonos ligado ao sítio ativo por meio de uma ponte salina com o resíduo Arg245, e o grupo heme (C<sub>34</sub>H<sub>32</sub>FeN<sub>4</sub>O<sub>4</sub>) no estado pentacoordenado de alto spin, obtida por difração de raio x com resolução 2.50 Å (**BELCHER *et al.*, 2014**) (PDB: 4l40). A partir da estrutura do ligante de ácido graxo C20, foram geradas estruturas de cadeias alifáticas contendo C18, C16, C14, C12 e C10 pela remoção progressiva de átomos de carbono do substrato. Para a formação de sistemas contendo como ligante os produtos de ácidos graxos  $\beta$ -hidroxilado, houve a adição de uma hidroxila na posição  $\beta$ , utilizando o programa tLeap (**CASE *et al.*, 2005**). E para os alcenos, o grupo carboxilato (-OC=O) foi removido, resultando em produtos C(n-1). Além disso, a estrutura mutante da OleT<sub>JE</sub> Leu176Gly foi criada substituindo o aminoácido leucina de posição 176 por uma glicina, sendo gerados dois sistemas, um com o ligante  $\beta$ -C20 e outro com o alceno-C19.

Os estados de protonação dos resíduos tituláveis foram estimados utilizando o programa H++ (**GORDON *et al.*, 2005**) em pH 7,4. Para a OleT<sub>JE</sub> nesse pH, todos os aspartatos e glutamatos se encontram desprotonados, as His92, His222, His325 e His363 estão duplamente protonadas; as His120, His210 e His259 se encontram protonadas na posição N <sub>$\epsilon$</sub> ; e apenas a His85 está protonada na posição N <sub>$\delta$</sub> .

Com o programa tLeap (**CASE *et al.*, 2005**) foram adicionados todos os hidrogênios, e a proteína foi solvatada com uma caixa de água pré-equilibrada do tipo TIP3P (**JORGENSEN *et al.*, 1983**) de formato octaedro truncado, de forma que a distância mínima entre a enzima e a borda da caixa seja de 15 Å. Além disso, íons de sódio e cloreto foram adicionados à caixa, em diferentes regiões próximas à da proteína, de forma a tornar o sistema eletricamente neutro. Com o programa também foram inseridos em todos

os sistemas os parâmetros do campo de força referentes ao grupo heme no estado de alto spin pentacoordenado e do resíduo de cisteína que está coordenado com o átomo de Ferro (heme), ambos fornecidos por **Shahrokh *et al.* (2012)**. Onde, os parâmetros de otimização de geometria foram obtidos por DFT (Teoria do Funcional da Densidade) usando o funcional densidade B3LYP/LACVP e as cargas parciais por **RESP**. Já o campo de força ff14SB (**MAIER *et al.*, 2015**) foi usado para a proteína, e o Gaff (**WANG *et al.*, 2004**) para moléculas orgânicas simples, de acordo com o ligante presente no sistema. Os parâmetros específicos para ácido graxo e alcenos foram obtidos do “*Generalized Amber Force Field*” (**WANG *et al.*, 2004**)(**WANG *et al.*, 2006**). Os sistemas gerados pelo programa tLeap abrangem em torno de 73.600 átomos.

### 7.2.2 Configuração das simulações

Condições de contorno periódicas foram empregadas em todas as simulações, com as interações de van der Waals de curto alcance truncadas em uma distância de corte de 10,0 Å, e as interações eletrostáticas de longo alcance tratadas com o algoritmo “*Particle mesh Ewald*” (PME) (**DARDEN; YORK; PEDERSEN, 1993**). Todas as equações de movimento foram integradas numericamente com um passo de integração de 2,0 fs, e as ligações envolvendo átomos de hidrogênio foram fixadas aos seus comprimentos de ligação na posição de equilíbrio usando o algoritmo “*RATTLE*” (**ANDERSEN, 1983**). O termostato de Langevin foi usado com frequência de colisão de 1,0 ps<sup>-1</sup> para manter a temperatura em 300 K durante as etapas de equilíbrio e de produção. E o barostato de Berendsen foi usado para manter em 1 bar a pressão durante as etapas de equilíbrio, com tempo de relaxamento de 2,0 ps.

O protocolo utilizado teve como primeira etapa uma minimização da energia do sistema de 2000 passos, mantendo fixos os átomos da proteína, do grupo heme e do produto sob um potencial harmônico com constante de força,  $k = 10 \text{kcal/mol. \AA}^2$ , ou seja, mantendo apenas os átomos de água e íons livres. Seguido por uma etapa de equilíbrio de 200 ps de dinâmica molecular a volume constante, mantendo a mesma contenção da etapa anterior para os átomos, e com a temperatura variando entre 0 K e 300 K. O sistema foi então submetido a uma segunda etapa de minimização de energia de 2000 passos, agora mantendo apenas os carbonos  $\alpha$  da proteína sob o mesmo potencial harmônico usado anteriormente. Depois, o sistema passou por uma equilíbrio de 500 ps a volume constante e preservando a mesma contenção de átomos da etapa anterior de minimização, com a temperatura variando também entre 0 K e 300 K. Esse ciclo de relaxamento mantendo os carbonos  $\alpha$  fixos, permite apenas que as cadeias laterais da proteína se ajustem a presença do solvente, pois ainda é possível a existência de contatos ruins que podem provocar um deslocamento do esqueleto peptídico. Em seguida, o sistema foi submetido a uma terceira etapa de minimização da energia de 2000 passos, onde não há qualquer tipo de contenção, com todos os átomos livres, e seguido por uma equilíbrio de 400 ns de dinâmica

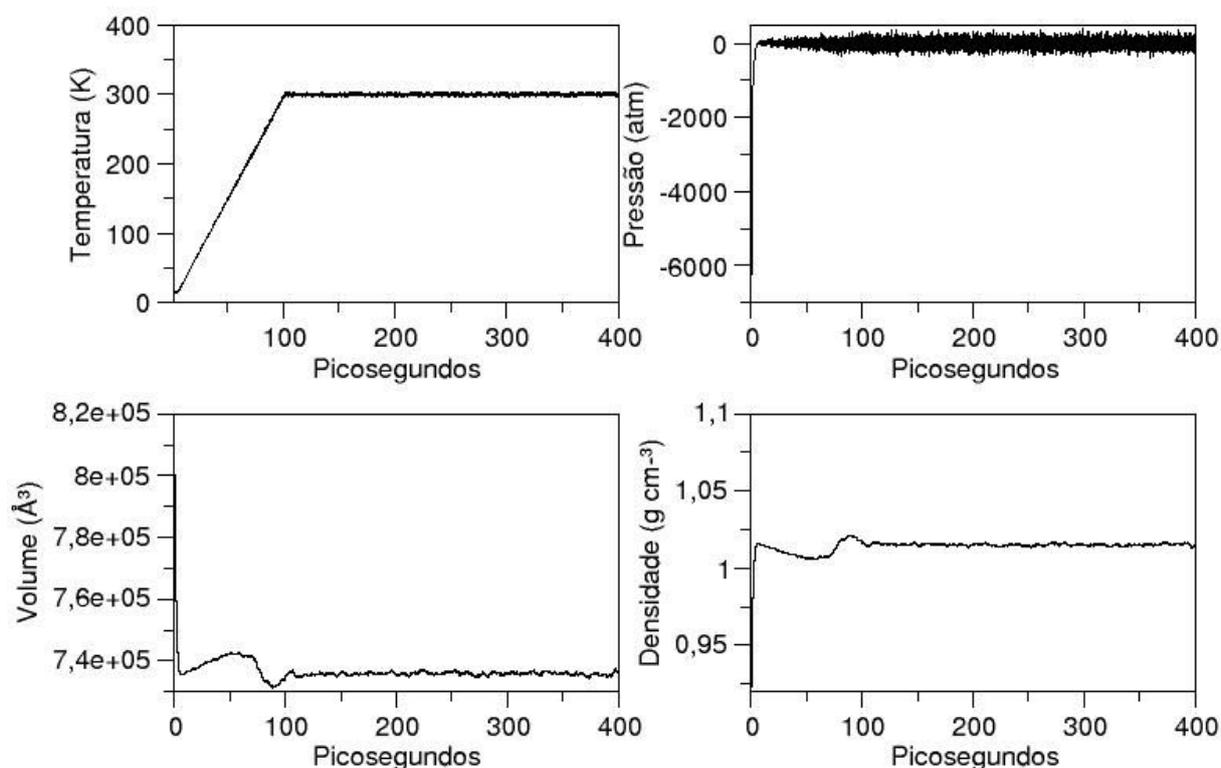
molecular à pressão constante e, também sem restrições do movimento dos átomos, com a temperatura variando de 0 K a 300 K. Uma última etapa de equilíbrio foi feita, onde uma dinâmica molecular de 100 **ps** a volume constante com todos os átomos livres e temperatura constante de 300 K foi realizada. Por fim, a etapa de produção foi conduzida a volume e temperatura constantes, com o sistema mantido a 300 K por 300 **ns**. Para cada sistema foram realizadas 3 simulações independentes, onde todas as velocidades iniciais foram atribuídas aleatoriamente seguindo uma distribuição de Maxwell-Boltzmann. As simulações foram então analisadas utilizando o programa cpptraj (ROE; CHEATHAM, 2013).

## 8 RESULTADOS E DISCUSSÕES

### 8.1 ANÁLISE DE CONTROLE DAS SIMULAÇÕES

Em simulações de dinâmica molecular é necessário fazer o monitoramento das propriedades do sistema (temperatura, densidade, energia cinética etc.) durante as etapas de equilíbrio, de forma a garantir a qualidade do sistema gerado e se o mesmo se encontra apto para prosseguir para a etapa de produção. Todos os 14 sistemas foram devidamente acompanhados, mas para exemplificação são retratados no **Gráfico 8.1** apenas as etapas de equilíbrio 3 e 4 do sistema com ligante  $\beta$ -hidroxilado C18:

Gráfico 8.1 – Monitoramento da equilibração 3

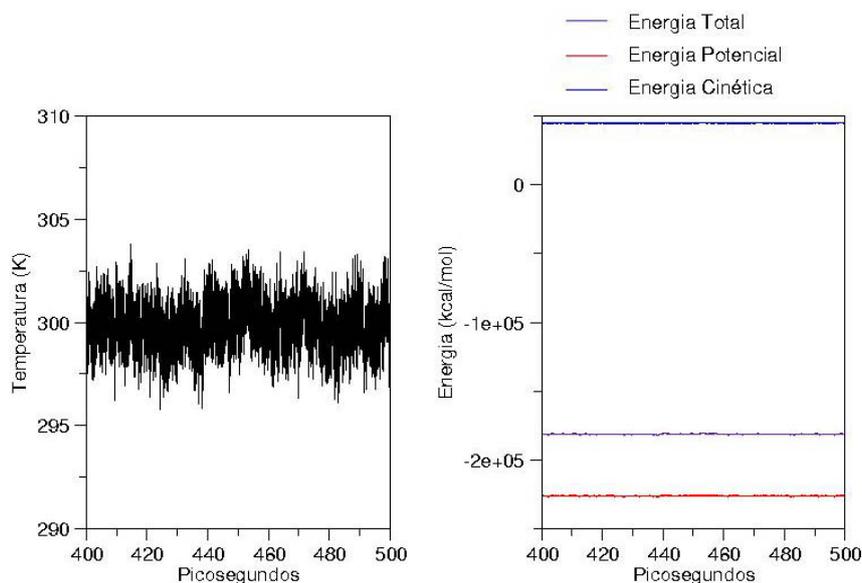


Os gráficos da variação de temperatura, pressão, volume e densidade ao longo de 400 ps numa equilibração NpT.

Na equilibração 3, a primeira em que não há qualquer tipo de contenção nos átomos, o sistema foi aquecido gradualmente de 0 à 300 K em condições NpT usando o termostato de Langevin, atingindo a temperatura alvo após 100 ps. No resto do processo, a temperatura passou a flutuar em torno do valor desejado, sendo esse o comportamento esperado. Já a pressão foi regulada com o barostato de Berendsen, partindo de um valor negativo condizente com uma “força” que age para diminuir o tamanho da caixa; e as pressões positivas, a uma “força” agindo para aumentar o tamanho dela. Pelo gráfico,

observa-se que a pressão ao longo do processo oscila mais em comparação a temperatura. Isso ocorrer porque, no caso da temperatura, o coeficiente de atrito de Langevin,  $\xi$ , impõe o quão rápido será a transferência de calor ao longo da caixa de simulação, e a força estocástica,  $R_i$ , atuante nos átomos, aumenta diretamente a energia do sistema, gerando uma maior flutuação tanto na energia cinética quanto potencial. Essa maior flutuação faz com que a temperatura convirja mais rapidamente. Enquanto, no barostato de Berendsen, se controla o tempo de relaxamento,  $t_p$ , que introduz uma escala de tempo para o sistema responder quando submetido a uma pressão externa (banho). Esse algoritmo cria um amortecimento exponencial entre a diferença de pressão interna instantânea,  $\mathfrak{P}$ , e a pressão externa ao longo do tempo, e não representa o comportamento físico real de um sistema termodinâmico, afetando diretamente as equações de movimento e trajetórias. Os artefatos introduzido nas simulações por esse tipo de barostato faz que o algoritmo não consiga recuperar as flutuações corretas da pressão instantânea, reduzindo artificialmente a sua oscilação, o que conseqüente desfavorece as flutuações na energia cinética, proporcionando uma convergência mais lenta da propriedade termodinâmica. Mas, ao longo do processo a pressão passa a oscilar em torno de 1 atm, indicando que ocorreu a equilíbrio. Sincronizado com o aumento da pressão, ocorre a diminuição do volume da caixa, onde após o sistema alcançar a temperatura desejada, o mesmo passa a oscilar em torno de um valor médio; assim como, paralelamente, a densidade se equilibra. Dessa forma, com todas as propriedades convergindo, a equilíbrio 3 pode ser considerada bem-sucedida.

Gráfico 8.2 – Monitoramento da equilíbrio 4



À esquerda, a variação de temperatura ao longo dos 100 ps de equilíbrio NVT. À direita, o monitoramento das energias cinética, potencial e total, evidenciando a conservação do sistema.

No **Gráfico 8.2**, mostram-se propriedades do sistema simulado por 100 ps na última

etapa de equilibração a NVT. Ao longo do processo, a temperatura se manteve flutuando em torno do valor alvo de 300 K, e as energias cinética e potencial se mantiveram constantes em todo o tempo. Conclui-se, portanto, que o relaxamento final sem contenções nos átomos foi bem-sucedido e que o sistema está apto para ser submetido à etapa de produção da dinâmica.

## 8.2 ANÁLISES DE FLUTUAÇÕES ESTRUTURAIS

### 8.2.1 RMSD (“*Root Mean Square Deviation*”) - Raiz quadrada do deslocamento quadrático médio

O **RMSD** é uma medida comumente utilizada em simulações de **MD** para quantificar o quanto a estrutura tridimensional da proteína varia durante a simulação, ou seja, sua flexibilidade. Por meio da **Equação 8.1**, a evolução temporal das flutuações das posições atômicas pode ser representada graficamente. As análises de RMSD podem ser feitas considerando todos os átomos dos resíduos da proteína, apenas os carbonos-alfa ou para qualquer conjunto de átomos presentes na estrutura. O valor é calculado considerando a diferença de coordenadas atômicas entre o átomo alvo ao longo da **MD** e a sua posição em um sistema de referência fixo. Onde essa referência pode ser definida a partir das coordenadas da estrutura cristalográfica do sistema (aderida nesse trabalho), da estrutura média originada com a simulação ou do instante final gerado após a última etapa de equilibração. Antes de serem submetidas ao cálculo, o “*ensemble*” de estruturas produzidas ao longo da trajetória é alinhado e sobreposto com a estrutura de referência em um mesmo sistema de coordenadas. O alinhamento é feito para remoção dos movimentos translacionais e rotacionais da molécula de forma a manter apenas o movimento interno da estrutura. O cálculo do RMSD é feito conforme a equação abaixo:

$$RMSD_{\Omega}(t) = \left( \frac{1}{N_{\Omega}} \sum_{i \in \Omega} [r_i(t) - r_i^{ref}]^2 \right)^{\frac{1}{2}} \quad (8.1)$$

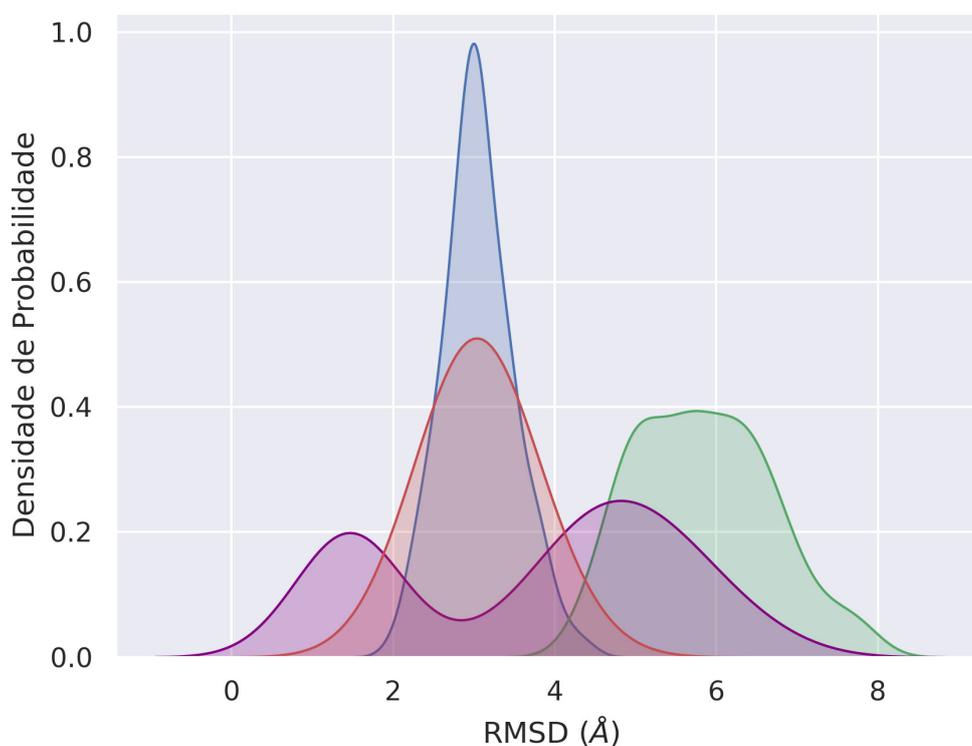
onde  $N_{\Omega}$  é o número de átomos do grupo  $\Omega$  e  $r^{ref}$  são as coordenadas de referência em torno da qual as flutuações do grupo  $\Omega$  são calculadas. A diferença entre os pares de coordenadas é elevada ao quadrado para obter apenas valores positivos; o resultado é dividido por  $N_{\Omega}$  para se obter a média e a raiz quadrada é usada para anular a potência de 2.

Quando o **RMSD** é projetado em relação aos instantes (“*frames*”) da simulação, obtém-se a relação de quanto a proteína ou o grupo de átomos selecionados se desvia de sua conformação original ao longo do tempo de simulação, o que acontece constantemente ao longo de uma **MD**. Mas, para se obter informações acerca da mobilidade atômica plota-se um histograma normalizado do RMSD na forma de densidade de probabilidade,

onde a área abaixo da curva é sempre 1, sendo possível, dessa forma, obter uma relação semiquantitativa da flexibilidade do grupo analisado.

Estruturas mais rígidas e de baixa mobilidade, como mostradas no **Gráfico 8.3**, apresentam um gráfico mais alongado e estreito (a), ou seja, alta probabilidade na ocupação de estados conformacionais com baixa variação em valores de RMSD entre si, indicado por uma base mais curta. Estruturas mais flexíveis apresentam um gráfico de base larga (b), entretanto, nesses casos pode haver a presença de mais de um pico (c), o que sinaliza a presença de mais de uma conformação principal, se os picos apresentarem probabilidades semelhantes, ou, caso contrário, que estados metaestáveis foram visitados durante a simulação. Quando o gráfico apresenta uma base muito larga com igual probabilidade de ocorrência, o sistema é dito desordenado (d), mediante que ele apresenta uma distribuição ampla de conformações com probabilidades iguais de ocorrência.

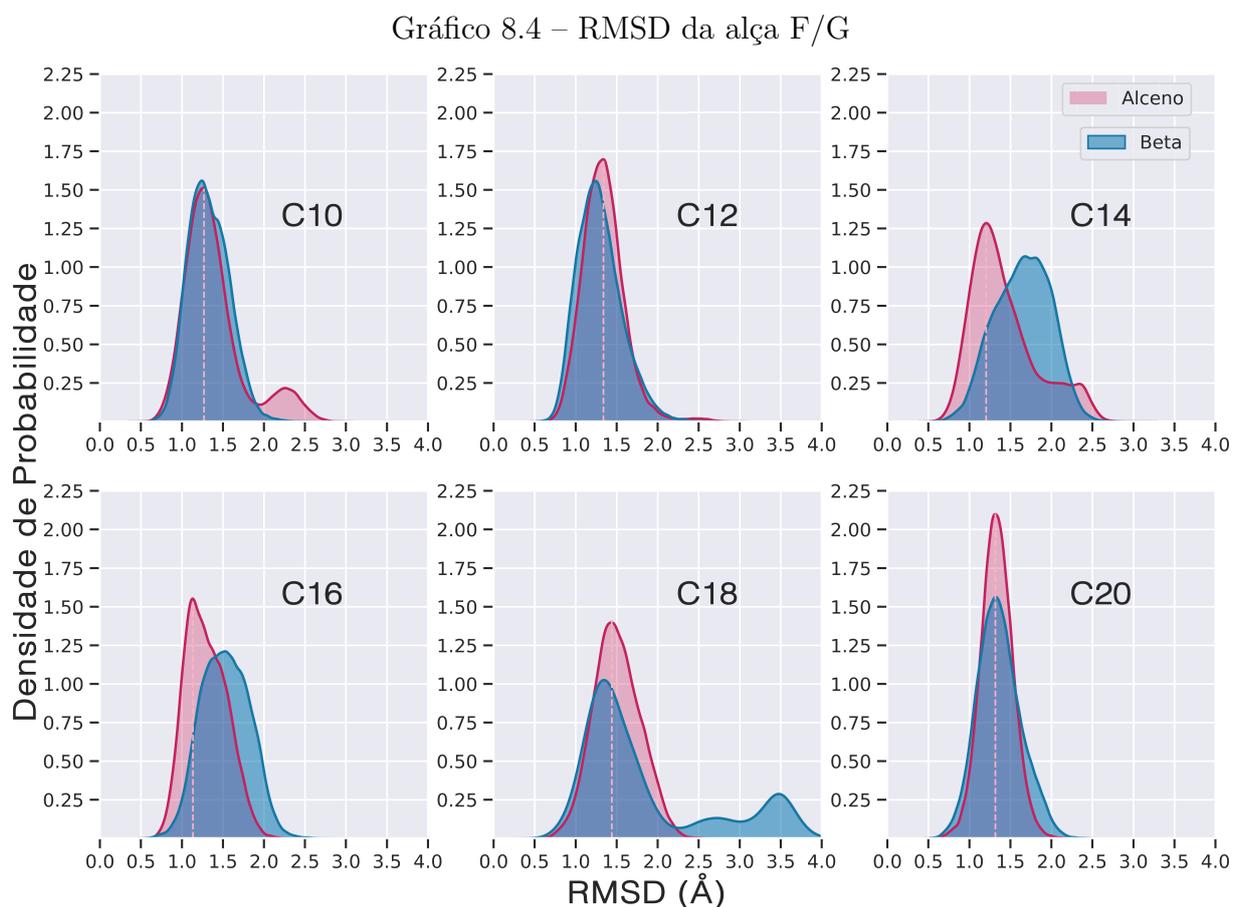
Gráfico 8.3 – Tipos de gráficos de densidade de probabilidade



Tipos de gráficos de distribuição de probabilidades de densidade gerados a partir de histogramas. (a) Em azul, gráfico centralizado com baixa dispersão de valores indica uma estrutura rígida. (b) Em vermelho, um gráfico com base mais larga aponta para uma maior dispersão de valores em relação à média e sugerem maior flexibilidade ou mobilidade. (c) Em roxo, dois picos são formados de baixa frequência, com probabilidades semelhantes de ocorrência. (d) Em verde, um gráfico tipo “platô”, onde os valores centrais apresentam a mesma frequência de ocorrência.

### 8.2.1.1 RMSD da alça F/G

O Gráfico 8.4 mostra o RMSD para os resíduos que compõem a alça F/G para cada sistema simulado, e permite analisar a flexibilidade da alça na presença de diferentes produtos com diferentes tamanhos de cadeia, onde os produtos  $\beta$ -hidroxilados de comprimento  $C(n)$  estão sendo comparados aos alcenos,  $C(n - 1)$ , abstraídos do átomo de carbono C1.



Gráficos de RMSD da alça F/G na presença de diferentes produtos de cadeias variadas, com os ligantes alcenos  $C(n - 1)$  sendo representados em rosa e os  $\beta$ -hidroxilados  $C(n)$  em azul.

Pela análise dos gráficos é possível inferir, tanto por apresentarem uma base mais larga, quanto pela tendência central de desvio de RMSD, que os sistemas com produtos  $\beta$ -hidroxilados dentro da cavidade de ligação apresentaram uma alça F/G levemente mais flexível do que na presença dos alcenos, exceto nos casos C10 e C12. Em relação aos sistemas com produtos alcenos, nos originados de ácidos graxos de cadeias menores, entre 10 a 14 carbonos, a alça F/G apresentou maior flexibilidade do que nos sistemas de cadeias maiores. Havendo nos sistemas C10 e C14 o acesso a estados metaestáveis durante a trajetória, com desvios chegando a 2.5 Å em relação à estrutura cristalográfica, porém com menor probabilidade de ocorrência. Tal observação não fica facilmente perceptível

no sistema C12, mas ele acessou uma conformação metaestável - há uma pequena linha vermelha acima de 2.5 Å; entretanto, o tempo de simulação usado não foi o suficiente para amostrar significativamente esse processo. Dentre os sistemas  $\beta$ -hidroxilados, o C14 e o C18 apresentaram comportamentos mais distintos, sendo o primeiro caracterizado por uma desordem maior na alça; e o segundo, por visitar estados metaestáveis com grandes desvios de RMSD, em torno de 3.5 Å. E, em ambos os produtos com ligantes gerados a partir do ácido graxo C20, a alça F/G apresentou a maior rigidez entre todas as cadeias analisadas.

As análises de RMSD da alça F/G na presença dos produtos alcenos indicam uma ordem crescente de mobilidade entre os sistemas como: C20<C16<C12<C18<C10<C14, onde o critério se baseia na comparação da dispersão de valores em relação à tendência central (linha pontilhada). Com base em enzimas semelhantes à OleT<sub>JE</sub> que também apresentam uma alça F/G, é sugerido que após a catálise, esse **motivo** se movimentaria de forma a “guiar” o produto para fora da cavidade de reação, o que exigiria grande flexibilidade estrutural. O papel da alça F/G como possível responsável por guiar a saída do produto da cavidade de ligação está diretamente relacionada a sua mobilidade e pode refletir a atividade enzimática de uma proteína, onde se expressa o quanto de substrato é convertido em produto. Ou seja, a alta taxa de conversão de substrato reflete a facilidade com o qual a enzima expulsa o produto, e vai estar diretamente relacionada a movimentação do enovelamento responsável pelo processo. A comparação da flexibilidade da F/G na presença de alcenos com diferentes cadeias apresentou um comportamento similar a atividade enzimática exibida pela OleT<sub>JE</sub>, pelos dados de conversão (%) obtidos por **Matthews *et al.* (2017)** listados na **Tabela 1**.

Tabela 1 – Rendimentos na conversão de ácidos graxos, %:

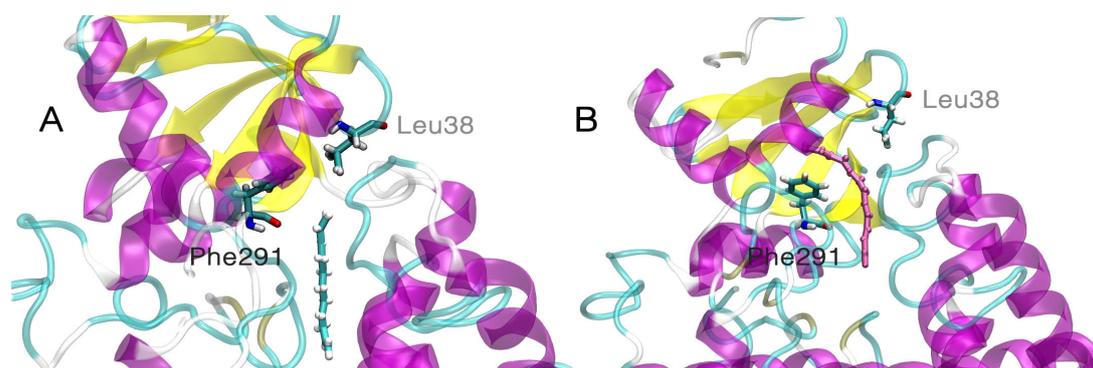
	Sistemas:					
	C20	C16	C18	C10	C12	C14
Conversão, %:	14	43	47	87	94	100

Conversão de substratos de ácidos graxos com diferentes tamanhos de cadeia promovida pela OleT<sub>JE</sub>. Dados obtidos por **Matthews *et al.* (2017)** (não foram fornecidas incertezas acerca das medidas).

A análise das simulações aponta que a carboxila dos produtos  $\beta$ -hidroxi permaneceram ligadas a Arg245, mantendo esses ligantes próximos ao sítio de ligação, exceto para o sistema C18, onde o produto se dirigiu a “saída” da cavidade, próximo a região do “loop” F/G. Como nessas simulações o produto não se dissociou do sítio ativo, não se amostrou o espaço físico referente ao processo de saída do ligante de dentro da enzima, não sendo possível associar o comportamento da alça F/G desses sistemas com a atividade enzimática.

Enquanto, os produtos alcenos conseguiram circular pela cavidade de ligação durante a simulação, com exceção para o sistema C12 que, como mencionado anteriormente, não simulou tempo suficiente para visitar as conformações metaestáveis possíveis de serem acessadas na temperatura simulada. A análise das trajetórias apontara um confinamento do alceno C11 no sítio de ligação devido a interações com a cadeia lateral dos resíduos Leu38 e Phe291, ambos localizados no sítio de ligação, que fecham o canal ao redor do produto, mostrado na **Figura 8.1**. O ancoramento do produto no sítio não permitiu que a proteína visitasse mais conformações estruturais possíveis à 300 K.

Figura 8.1 – Ancoramento do ligante alceno C11

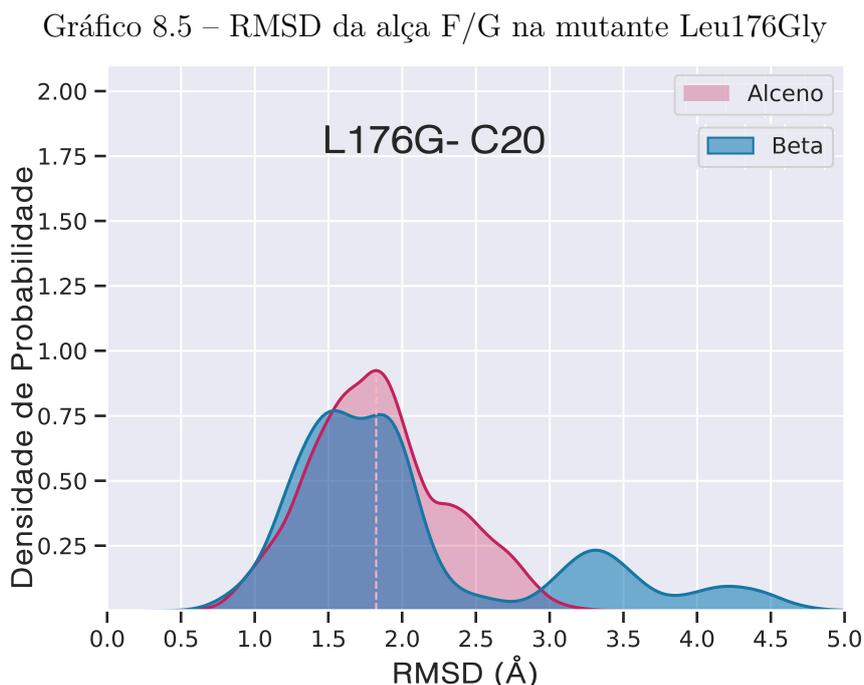


(A) Ancoramento do ligante C11 na cavidade pelas cadeias laterais dos resíduos Leu38 e Phe291, próximo às hélices A e A'. (B) O ligante C13 no mesmo local evidencia um comportamento distinto em torno dos resíduos 38 e 291, que se projetam para longe da cavidade, permitindo que o ligante se desloque livremente pela cavidade de reação.

O que permite dizer se tempo de uma simulação é suficiente para extrair informações pertinentes do funcionamento de uma proteína, envolve quanto tempo é necessário para capturar os movimentos biologicamente relevantes do processo. No caso, o movimento relativo de domínios, como a alça F/G, ocorre em escalas de nanossegundos até milissegundos, mas os recursos computacionais típicos atualmente, limitam as simulações com solvente explícito a escala de nanossegundos. Entretanto, o tempo de simulação pode ser considerado suficiente quando as propriedades dinâmicas obtidas numa simulação convergem com os dados experimentais. Como o comportamento do alceno C11 foi o único destoante em relação as atividades, se assume que todos os outros alcenos simularam por tempo suficiente. Apesar das simulações não mostrarem a saída direta dos produtos, mediante que esses movimentos globais podem demorar de  $10^7$  a  $10^8$  segundos, as dinâmicas dos sistemas na presença dos produtos alcenos reproduziram o comportamento natural da alça F/G especulada a partir de dados experimentais de **Matthews *et al.* (2017)**. Logo, com a análise de RMSD da F/G na presença dos alcenos e do produto  $\beta$ hidroxilado C18 é possível apontar a influencia do enovelamento F/G com a saída desses produtos da cavidade.

**Amaya et al. (2018)** apresentaram dados da mutante Leu176Gly, relatando uma maior taxa de liberação de produto em comparação a enzima selvagem. Esses dados podem sugerir que a perda da interação entre o “loop” e substrato, torne a alça F/G da mutante mais flexível e desordenada, a ponto de provocar a perda de seletividade e regioseletividade, causando a hidroxilação até nas posições  $\epsilon$ ,  $\gamma$  e  $\sigma$  do ácido graxo C20.

A análise de RMSD da alça F/G da mutante Leu176Gly na presença de produtos de cadeia C20  $\beta$ -hidroxi e alceno C19, plotados no **Gráfico 8.5**, mostra um comportamento distinto ao observado na enzima selvagem na presença de ligantes correspondentes.



Gráficos de RMSD da alça F/G da enzima mutante Leu176Gly na presença do ligante alceno C(19), representado em rosa, e o produto beta-hidroxilados (C20), em azul.

Assim como sugerido pelo estudo experimental de **Amaya et al. (2018)**, o **motivo** F/G da variante apresentou grande flexibilidade na presença de ambos os produtos, sendo mais expressiva a mobilidade ao interagir com o produto  $\beta$ -hidroxilado. O caráter mais flexível demonstrado pela F/G da Leu176Gly na presença do alceno C19, associado a uma taxa de conversão de substratos de 98.3%, em comparação à enzima selvagem, que converte apenas 14% dos ácidos graxos e possui uma alça mais rígida, corrobora a relação direta entre a flexibilidade desse domínio com a taxa de atividade enzimática e seu papel como mediadora da saída dos produtos de dentro da cavidade.

O comportamento da alça F/G da mutante na presença do produto  $\beta$ -hidroxi C20 é similar ao observado pela cepa selvagem no sistema C18  $\beta$ -hidroxilado, que foi apontado por **Matthews et al. (2017)** como o único a apresentar uma inversão de regioseletividade, promovendo a hidroxilação quase que exclusivamente na posição  $\alpha$ . Isso sugere uma relação entre a alta mobilidade da alça F/G com a regioseletividade. Visto que os

produtos beta-hidroxilados são semelhantes em nível molecular aos substratos de ácido graxo que os geram, é possível inferir um comportamento análogo da alça antes da catálise. O que sugere que na presença do ácido graxo C18 na selvagem e C20 na mutante Leu176Gly, a alça se manteria tão móvel, a ponto de permitir ao substrato se movimentar pela cavidade de forma a não ancorá-lo devidamente, tornado de alguma forma, outras posições mais acessíveis para a abstração de hidrogênio.

Outro sistema que reforça a relação da flexibilidade da alça F/G com a atividade enzimática envolve a mais recente descarboxilase descoberta por colaboradores do grupo, OleTP<sub>RN</sub> (RADE *et al.*, 2023), isolada da bactéria *Rothia nassimurium*, que também atua em ácidos graxos poli-insaturados, e compartilha 32,3% de identidade da sequência de aminoácidos com a OleT<sub>JE</sub>. A OleTP<sub>RN</sub> apresenta a alça F/G mais curta, sendo mais assemelhada à da hidroxilase P450<sub>BSβ</sub>. A atividade enzimática da OleTP<sub>RN</sub> é maior para ácidos graxos de cadeia curta assim como na OleT<sub>JE</sub>. Entretanto, os rendimentos para os sistemas de cadeias longas, equivalentes aos analisados por Matthews *et al.* (2017), é expressivamente maior em comparação ao demonstrado pela OleT<sub>JE</sub>, como evidenciado na Tabela 2:

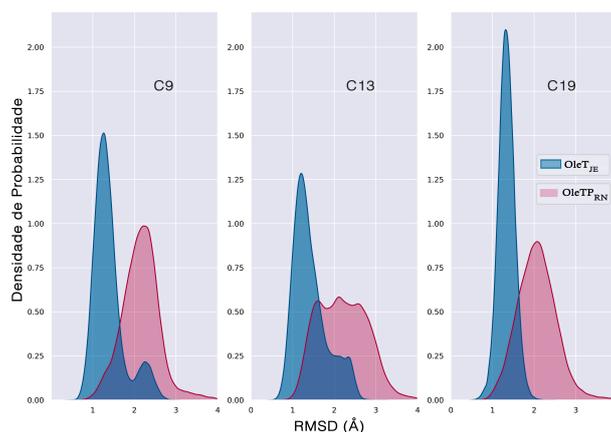
Tabela 2 – Rendimentos das enzimas OleTP<sub>RN</sub> e OleT<sub>JE</sub> na conversão de ácidos graxos, %:

	Sistemas:					
	C9	C11	C13	C15	C17	C19
OleTP <sub>RN</sub>	70.8±2.5	73.9±3.3	68.6±1.8	66.6±0.3	69.2±1.8	65.9±2.0
OleT <sub>JE</sub>	87	94	100	43	47	14

Atividade enzimática das enzimas OleTP<sub>RN</sub>, fornecida por Rade *et al.* (2023), e da enzima OleT<sub>JE</sub>, obtida por Matthews *et al.* (2017), com os rendimentos na conversão de ácidos graxos com diferentes tamanhos de cadeias.

Para entender a base molecular por trás das maiores atividades observadas para a OleTP<sub>RN</sub>, comparou-se o RMSD da F/G das duas descarboxilases, cuja simulação já havia sido feita anteriormente pelo grupo. Quando comparado o RMSD da alça F/G na presença dos sistemas alcenos de cadeia C9, C13 e C19, fica evidente que a OleTP<sub>RN</sub> apresenta esse **motivo** mais flexível que a OleT<sub>JE</sub>, reforçando a relação da mobilidade desse grupo atômico com a liberação de produto pela enzima, como demonstrado no Gráfico 8.6. O fato da OleTP<sub>RN</sub> apresentar o “loop” entre as hélices F e G menor, com apenas 5 resíduos, enquanto a OleT<sub>JE</sub> apresenta quase o dobro, sugere que essa região desempenha um papel central na flexibilidade da alça F/G. Ou seja, mesmo se encontrando longe do sítio ativo, as interações resultante dos resíduos envolvidos no “loop” provocariam uma maior mobilidade da alça, resultando em maiores atividades na enzima OleTP<sub>RN</sub>.

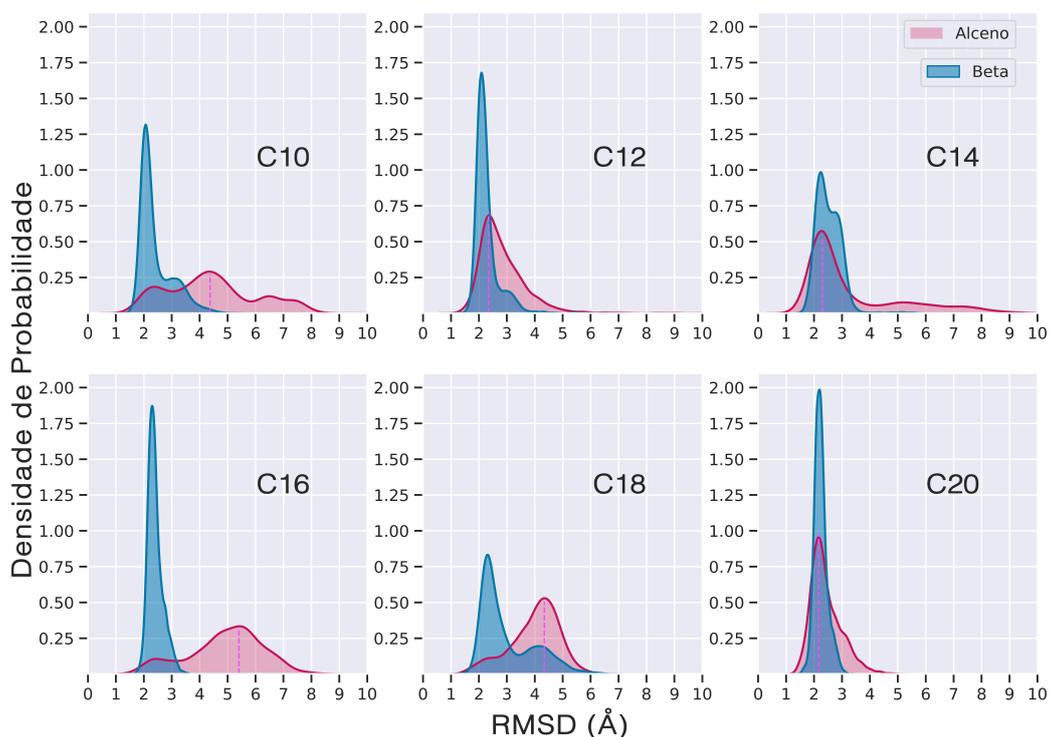
Gráfico 8.6 – Comparação do RMSD da alça F/G nas descarboxilases



Comparação do RMSD da alça F/G das enzimas OleTP<sub>RN</sub>, representada em rosa, e da OleT<sub>JE</sub>, em azul, na presença de diferentes produtos alcenos. Esses resultados foram incluídos no artigo RADE *et al.* Proc. Natl. Acad. Sci. USA, 2023, 120, e2221483120.

### 8.2.1.2 RMSD dos produtos dentro da cavidade da enzima selvagem OleT<sub>JE</sub>

Gráfico 8.7 – Comparação do RMSD dos produtos



Gráficos de RMSD de cada produto na enzima selvagem OleT<sub>JE</sub>, com as olefinas terminais C( $n - 1$ ) representadas em rosa, e os ácido beta-hidroxilado C $n$ , em azul.

O Gráfico 8.7 mostra o RMSD dos produtos em relação à proteína, evidenciando, em geral, um confinamento maior dos produtos beta-hidroxilados em comparação aos alcenos C( $n - 1$ ), refletindo o confinamento dos produtos que não conseguem se dissociar do

sítio ativo. Entre as olefinas terminais, o produto alceno C19 foi o que apresentou a menor e do C9 a maior mobilidade. Entretanto, não se deve esperar uma correlação entre o tamanho da cadeia e a mobilidade, com substratos maiores sendo menos móveis que os ligantes menores. Isso porque, pela teoria do encaixe induzido (**KOSHLAND, 1959**), o substrato ao se ligar no sítio ativo induz uma mudança conformacional na enzima, onde dependendo dos resíduos que envolvem o ligante, a cavidade se adapta mais ou menos ao seu redor. E como mencionado anteriormente, o sistema C12 ficou “ancorado” na cavidade, apresentando uma mobilidade bem menor que sistemas de cadeias maiores como o C18 e C16, com este último sendo um dos mais móveis. Dessa forma, a partir dos gráficos, se infere a ordem crescente de mobilidade para os produtos alcenos dentro da cavidade como:  $C20 < C12 < C18 < C14 < C16 < C10$ . Esta ordem sugere uma relação entre a baixa flexibilidade desses produtos na cavidade de ligação com a alta seletividade, de acordo com os valores apresentados por **Matthews *et al.* (2017)** e exibidos na **Tabela 3**.

Tabela 3 – Seletividade enzimática para alcenos, %:

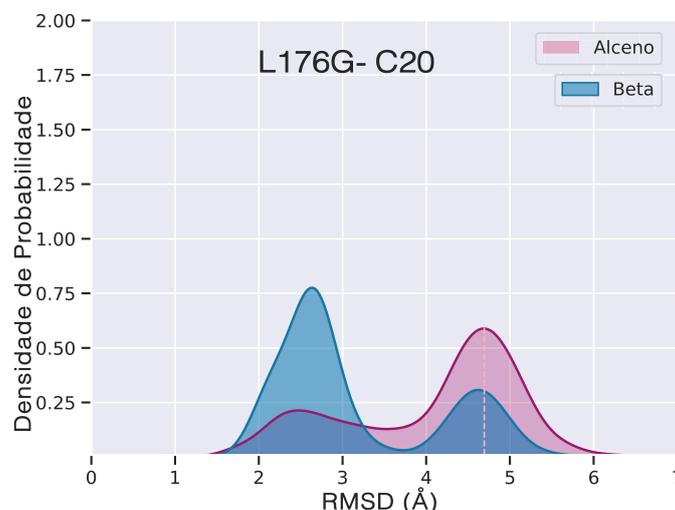
	Sistemas:					
	C20	C12	C18	C14	C16	C10
Seletividade	100.00	85.11	85.11	72.00	62.79	58.62

A partir de dados fornecidos por **Matthews *et al.* (2017)**, a seletividade para cada sistema foi obtida como a razão entre a porcentagem de alcenos produzidos pelo total de substrato convertido. Não foram fornecidas incertezas acerca das medidas.

Os resultados sugerem que as interações que guiam os ácidos graxos pela cavidade atuam de forma análoga para expulsar os produtos após a reação, gerando uma mobilidade equivalente nas cadeias mesmo após a abstração do grupo  $CO_2$ . Ou seja, a difusão exibida pelos substratos seria semelhante a de seus respectivos produtos, pois o comportamento seria regido em maior contribuição pela interação com a cadeia alifática. Tanto **Amaya *et al.* (2018)** quanto **Matthews *et al.* (2017)** especulam que a mobilidade do substrato no sítio ativo, assim como a flexibilidade da cavidade de reação, influenciaria o perfil de produtos gerados. Como houve uma correlação entre a mobilidade apresentada pelos produtos com a seletividade observada em cada substrato, infere-se que para a descarboxilação ocorrer é necessário que o ligante apresente menor mobilidade dentro da cavidade, de forma a ser guiado mais lentamente até o sítio de reação e melhor ancorado. Enquanto, a alta mobilidade do substrato no sítio ativo permitiria a hidroxilação. E o **Gráfico 8.8** do RMSD para os ligantes C20 da mutante Leu176Gly corroboram essa suposição, mediante que a olefina C19 apresentou uma mobilidade maior dentro da cavidade em comparação à enzima selvagem, sendo esse sistema associado a uma perda de seletividade. Visto que, dados de conversão apresentados por **Amaya *et al.* (2018)** mostram

uma conversão de 52.5% em alcenos pela enzima mutante, em oposição aos 100% obtidos com a selvagem por **Matthews *et al.* (2017)**.

Gráfico 8.8 – Comparação do RMSD dos produtos C20 na mutante Leu176Gly



Gráficos de RMSD de cada tipo de produto (olefina ou ácido  $\beta$ -hidroxi) na enzima mutante Leu176Gly OleT<sub>JE</sub>.

Entretanto, a flexibilidade da F/G não apresentou relação com a seletividade expressa pela enzima, como foi proposto por **Amaya *et al.* (2018)**. Logo, a alça F/G não atuaria influenciando o encaixe do substrato no sítio ativo. Ou seja, a ancoragem sugerida, se de fato existente, ocorreria por intermédio de outros resíduos dentro da cavidade, e não apenas por intermédio da F/G. Além disso, apesar da sua relação com a atividade enzimática sugerir uma mediação na saída do produto, não há uma influência direta entre a flexibilidade da F/G com a mobilidade dos alcenos dentro da cavidade.

Destaca-se que os produtos  $\beta$ -hidroxi C18 da selvagem e C20 da mutante Leu176Gly apresentaram maior flexibilidade em relação as outras cadeias, o que reforça a suposição que a alta mobilidade do ácido graxo dentro da cavidade nesses sistemas estaria associada de alguma maneira à perda de regioseletividade das enzimas durante a hidroxilação, como mencionado anteriormente.

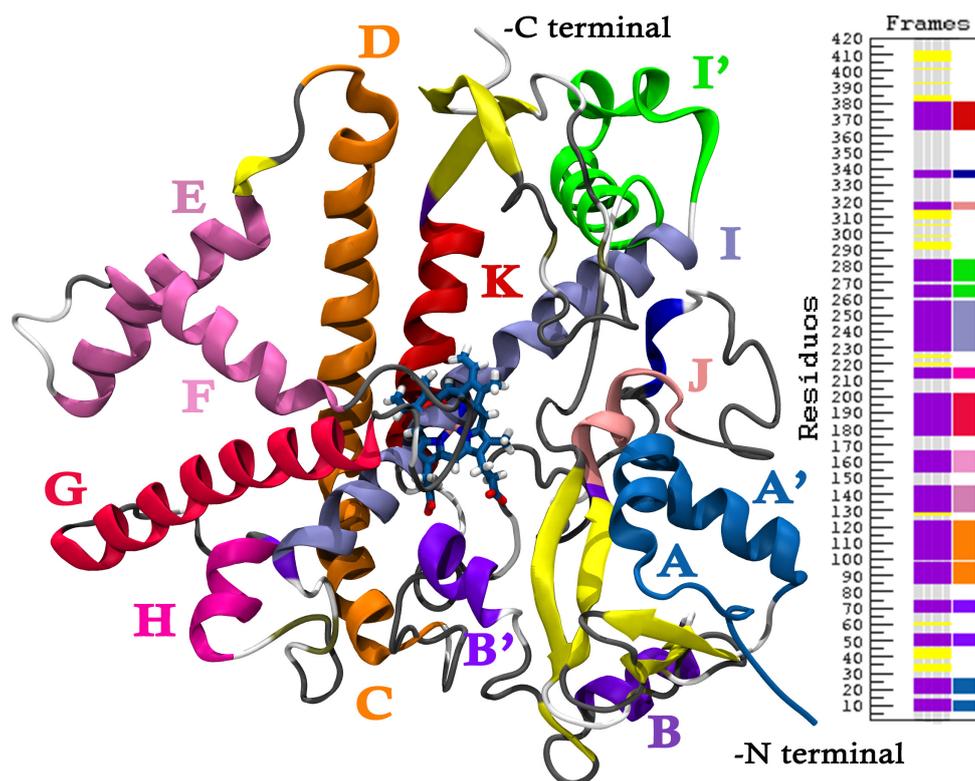
### 8.2.2 RMSF (“*Root mean squared fluctuation*”) - Raiz quadrada da flutuação média quadrática

Os desvio das posições atômicas em relação a estrutura de referência em um dado grupo de resíduos  $\Omega$  podem ser obtidas a partir da média temporal do RMSD do conjunto, onde o valor é conhecido como raiz quadrada da flutuação média quadrática, **RMSF** - (“*Root mean squared fluctuation*”), obtido pela **Equação 8.2**:

$$RMSF_{res} = \langle RMSD_{res} \rangle = \frac{1}{N_p} \sum_{n=0}^{N_p} RMSD_{res}(t_n), \quad (8.2)$$

onde o  $N_p$  é o número total de configurações do “ensemble” gerado durante a simulação, e  $t_n$  é o tempo correspondente à  $n$ -ésima conformação do conjunto estatístico. Em simulações de biomoléculas, o valor do RMSF dos carbonos  $\alpha$  da cadeia peptídica fornece o perfil de mobilidade de cada resíduo ao longo da estrutura proteica primária, evidenciando áreas de maior flexibilidade e variações localizadas. Diferentemente dos gráficos de RMSD apresentados anteriormente, que analisam mudanças conformacionais na estrutura, o RMSF mede o desvio de porções da estrutura que estão fluando mais, ou menos, em relação à sua estrutura média. Além disso, os valores de RMSF dos carbonos alfa do esqueleto peptídico podem ser usados para calcular e obter as estruturas secundárias de uma proteína, segundo os critérios estipulados por **Kabsch e Sander (1983)**, e com o auxílio do algoritmo STRIDE (**FRISHMAN; ARGOS, 1995**) implementado pelo programa de análise de trajetórias cpptraj (**ROE; CHEATHAM, 2013**). Onde, aos resíduos que constituem uma hélice são atribuídos a cor violeta; às folhas beta são designadas a cor amarela; regiões de “loops” a cor cinza e aos “coils” a cor branca.

Figura 8.2 – Estruturas secundárias da OleT<sub>JE</sub>



Identificação por cores das estruturas secundárias da enzima OleT<sub>JE</sub> obtidas pela análise das simulações com o uso do programa STRIDE.

Na **Figura 8.2**, todas as estruturas secundárias da OleT<sub>JE</sub> foram identificadas de acordo para o sistema na presença do ligante C19-alceno. As hélices foram ordenadas alfabeticamente e assinaladas com diferentes cores para melhor identificação dentro da estrutura.

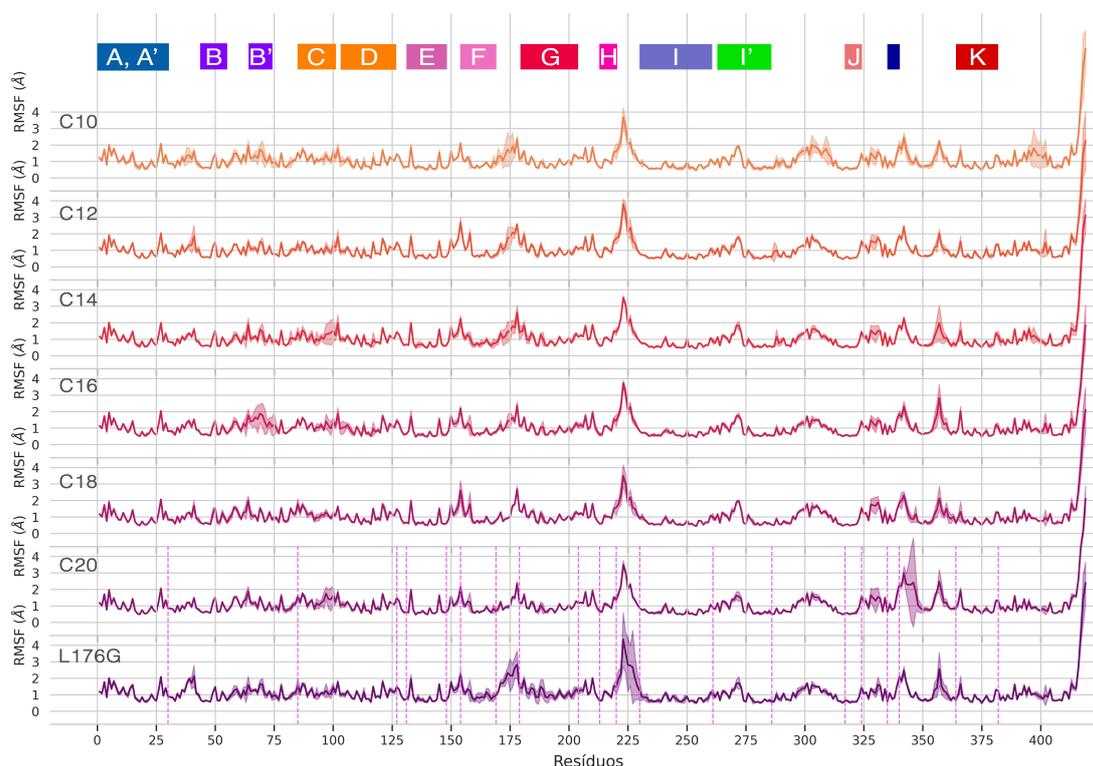
### 8.2.2.1 RMSF por resíduo

No **Gráfico 8.9** e **Gráfico 8.10**, o “*joyplot*”<sup>1</sup> com os **RMSF** em relação aos resíduos da cadeia peptídica foi plotado de acordo com o tipo de produto e as diferentes cadeias dos ligantes. No cpptraj (**ROE; CHEATHAM, 2013**), a flutuação média de um resíduo é obtido pela soma das flutuações de todos os átomos ponderada pela massa de cada átomo que compõe o resíduo, segundo a **Equação 8.3**. As regiões em que se espera maiores desvios de RMSF são nas regiões de “*loops*” entre as hélices e as folhas beta, mediante que as estruturas secundárias envolvem certa coesão estrutural.

$$\langle RMSF \rangle = \frac{\sum Flut_i \times Massa_i}{\sum Massa_i} \quad (8.3)$$

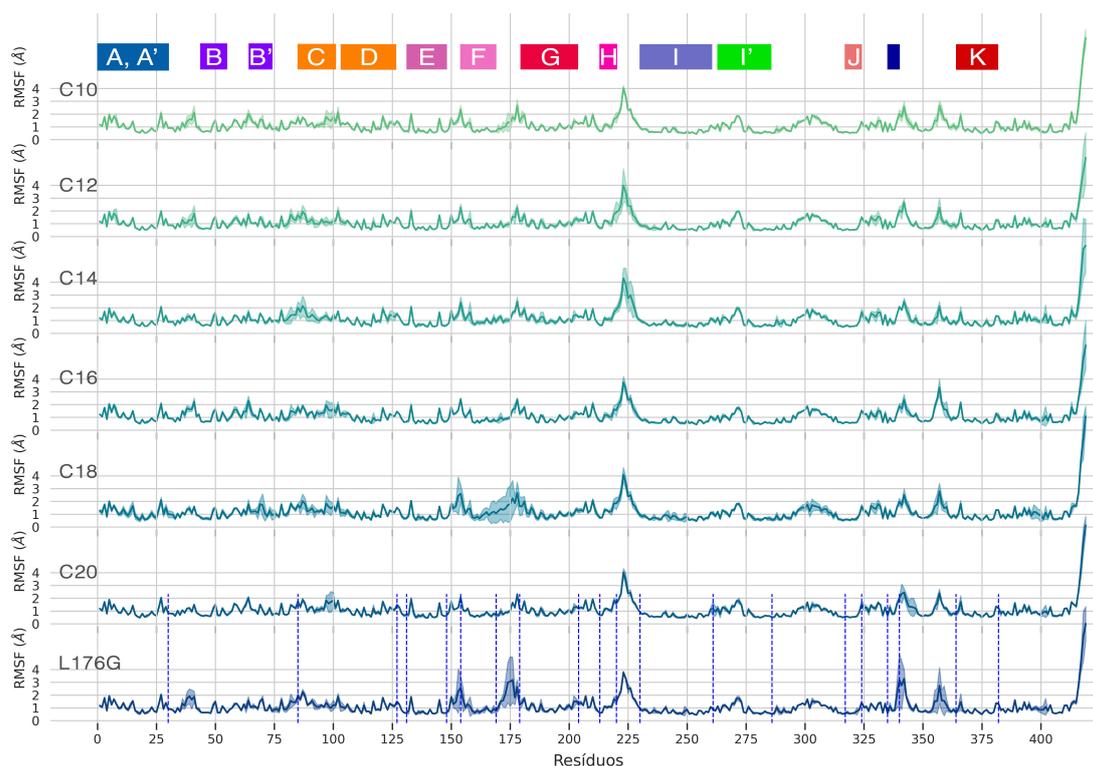
Com a análise dos gráficos, observa-se que os alcenos de cadeias menores apresentaram maiores flutuações na região do “*loop*” F/G (resíduos 173 a 181), que os de cadeias maiores. Já a mutante Leu176Gly apresenta flutuações por toda alça F/G (154 a 204), assim como o produto 3-hidroxi do sistema C18, evidenciando para tais sistemas o desordenamento dessa região.

Gráfico 8.9 – “*Joyplot*” com os RMSFs por resíduos na presença de alcenos



Gráficos de RMSF de cada resíduo da proteína, na presença de diferentes produtos alcenos.

<sup>1</sup> Conjunto de gráficos lineares combinados em um empilhamento vertical e levemente sobrepostos, compartilhando de um mesmo eixo x. Permite a fácil visualização e comparação de mudanças de distribuições ou flutuações (eixo y) ao longo do espaço ou tempo (eixo x).

Gráfico 8.10 – “Joyplot” com os RMSFs por resíduos na presença de  $\beta$ -hidroxilados

Gráficos de RMSF de cada resíduo da proteína, na presença de diferentes produtos beta-hidroxilados.

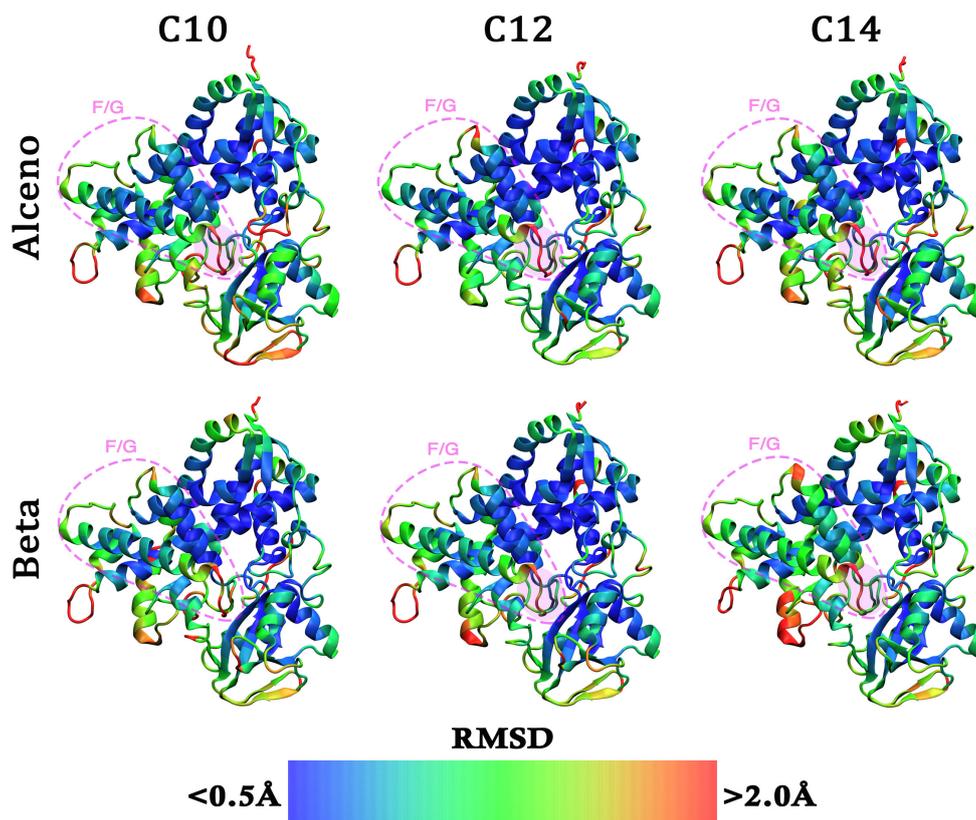
Com a análise dos gráficos, observa-se que os alcenos de cadeias menores apresentaram maiores flutuações na região do “loop” F/G (resíduos 173 a 181), que os de cadeias maiores. Já a mutante Leu176Gly apresenta flutuações por toda alça F/G (154 a 204), assim como o produto 3-hidroxi do sistema C18, evidenciando para tais sistemas o desordenamento dessa região.

Para ambos os tipos de produtos, a região da enzima que apresentou maior flutuação envolve o “loop” entre as hélice H e I (resíduos 221 a 229). Estudos realizados por **Du et al. (2017)** sugerem que regiões de encontro de vários “loops” na OleT<sub>JE</sub>, como as regiões entre as hélices H/I e entre os “loops” B/B'/C (resíduos 64 a 85), poderiam formar canais para entrada de água no sítio ativo da enzima, de forma a mediar a dissociação dos ligantes. Além disso, o canal de reação da OleT<sub>JE</sub> é predominado por resíduos hidrofóbicos, e a solvatação da cavidade contribuiria para expulsar as olefinas terminais hidrofílicas, pois reduziria a penalidade entrópica da solvatação hidrofóbica vigente. Na presença dos ligantes alcenos C19, a mutante Leu176Gly obteve flutuações bem mais intensas na alça H/I (acima de 4 Å), quando comparada à enzima selvagem. Tais resultados favorecem a hipótese de **Amaya et al. (2018)** de que a diferença de atividade observada na mutante seria consequência de uma maior solvatação dentro da cavidade, sendo sustentada pela especulação de **Du et al. (2017)** sobre a influência das flutuações dos encontros de “loops”

com a entrada de águas na cavidade. Enquanto, a selvagem com o ligante C19, apresentou variações expressivas no “loop” J/K, não sendo observado comportamento análogo em nenhum outro sistema.

As análises de RMSF por resíduo podem ser visualizadas diretamente na estrutura tridimensional da proteína com o auxílio do programa “*Visual Molecular Dynamics*” (VMD), sendo mostradas na **Figura 8.3**, **Figura 8.4** e **Figura 8.5**.

Figura 8.3 – RMSF na estrutura tridimensional da OleT<sub>JE</sub>, C10 a C14

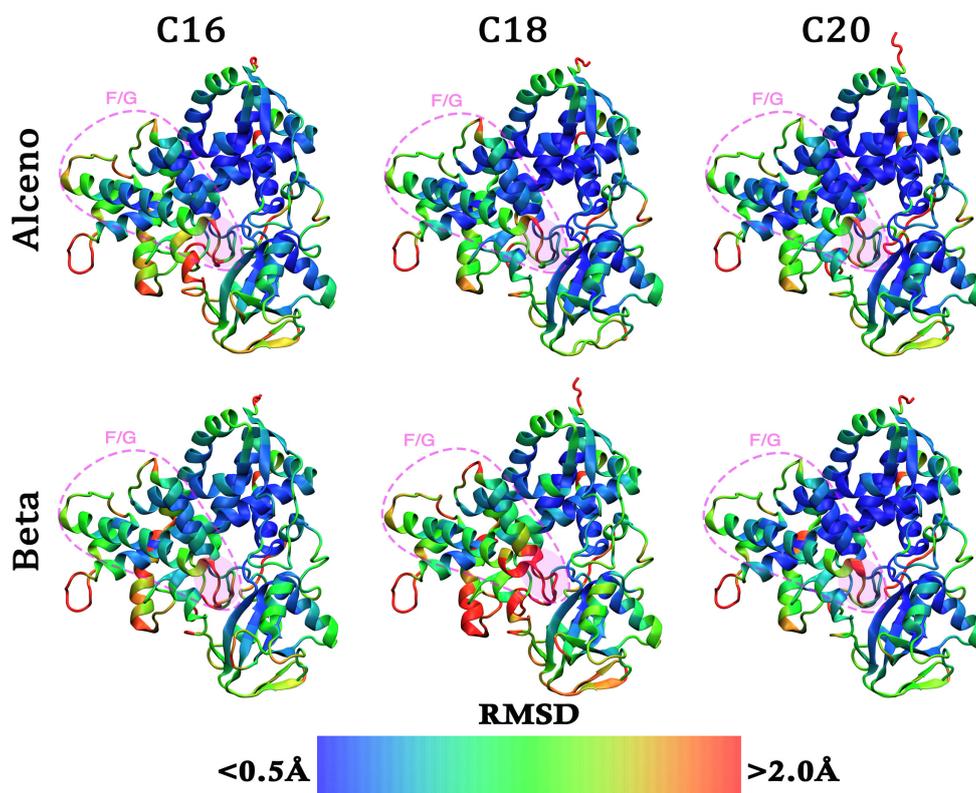


O RMSF da estrutura média da enzima selvagem OleT<sub>JE</sub> na presença dos produtos originários de ácidos de cadeia entre C10 a C14.

Onde, regiões com variações acima de 2.0 Å adquirem uma coloração em vermelho, e abaixo de 0.5 Å são representadas em azul, regiões intermediárias entre esse números são definidas de acordo com a escala degradê das figuras.

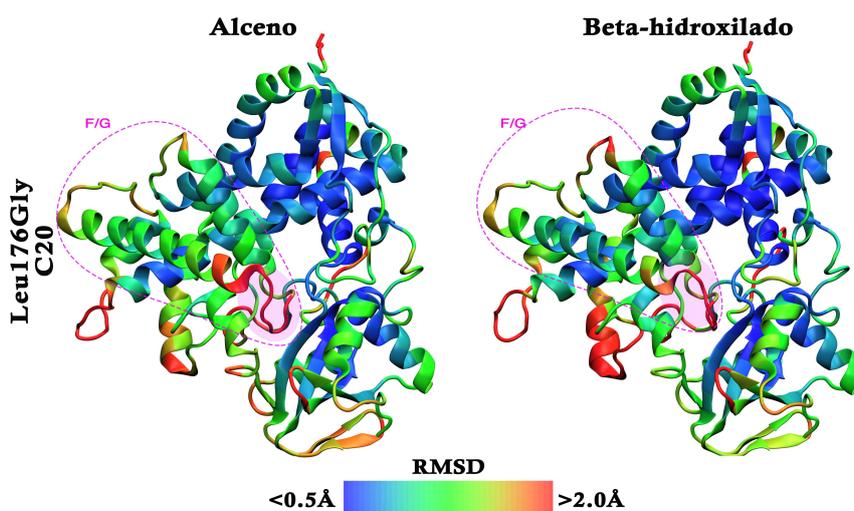
Agora é possível observar, considerando apenas a região do “loop” entre as hélices F e G, que a flexibilidade observada corresponde exatamente ao perfil de atividade enzimática obtido por **Matthews *et al.* (2017)** na **Tabela 1** (C20<C16<C18<C10<C12<C14). Dessa forma, observa-se também que os sistemas alcenos provindos dos ácidos graxos C18 e C20 apresentam flutuações no início da hélice G ao invés da região do “loop” F/G. E que a mutante Leu176Gly também apresenta flutuações bem mais intensas na alça, que também envolvem parte da hélice G. Assim como os produtos 3-hidroxi do sistema C18 da

Figura 8.4 – RMSF na estrutura tridimensional da OleT<sub>JE</sub>, C16 a C20



O RMSF da estrutura média da enzima selvagem OleT<sub>JE</sub> na presença dos produtos originários de ácidos de cadeia entre C16 a C20.

Figura 8.5 – RMSF na estrutura tridimensional da mutante Leu176Gly



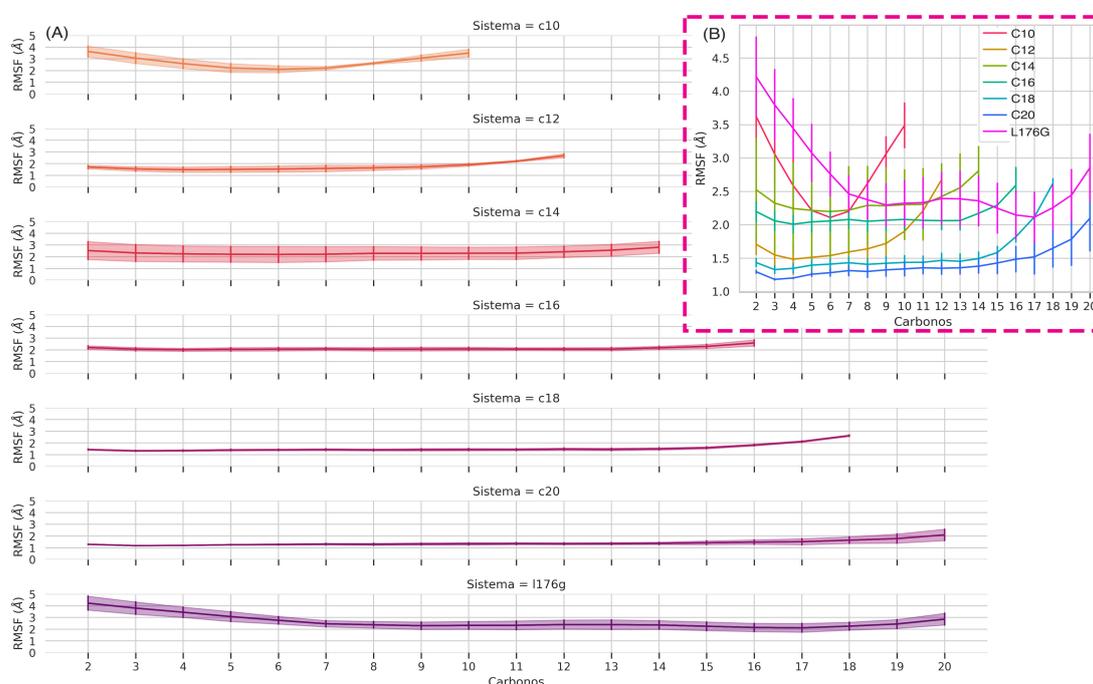
O RMSF da estrutura média da enzima mutante Leu176Gly OleT<sub>JE</sub> na presença dos produtos originários de ácido graxo C20.

selvagem, com flutuação se estendendo a hélice F, evidenciando um maior desordenamento do envelhecimento.

### 8.2.2.2 RMSF por átomo dos produtos

As análises de RMSF por átomo para os carbonos presentes nos ligantes são representadas nos **Gráfico 8.11** e **Gráfico 8.12**. Na análise de RMSF por átomo, a proteína foi alinhada à estrutura inicial para desconsiderar o seu movimento translacional na caixa durante a simulação, de forma que a mobilidade dos ligantes seja expressa em relação à enzima fixa, ou seja, o quanto o produto se move dentro da cavidade de reação. E o desvio de RMSF do ligante é calculado em relação a sua posição inicial na estrutura cristalográfica.

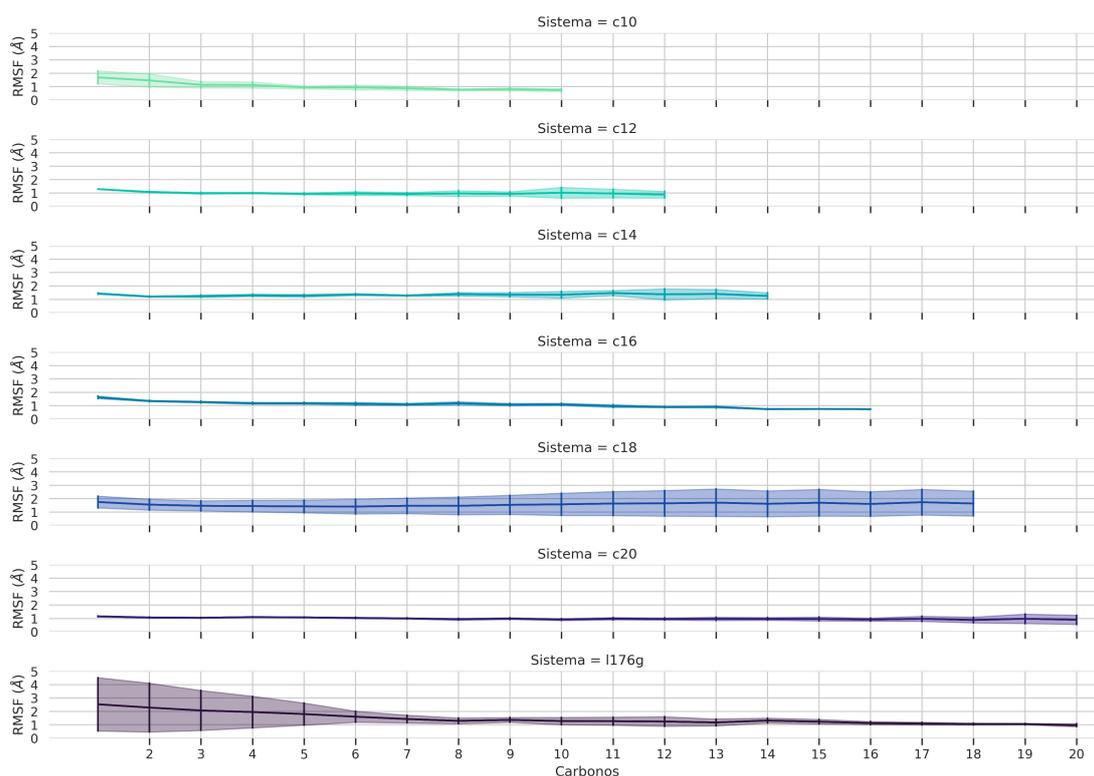
Gráfico 8.11 – RMSF dos produtos alcenos por átomo de carbono



Gráficos de RMSF para cada átomo de carbono que compõe os produtos alcenos na enzima selvagem e na mutante Leu176Gly. Onde, a ligação dupla terminal é entre os carbonos 2 e 3

Ao se alinhar cada curva (B) dos produtos alcenos, é possível inferir que o sistema proveniente do ácido graxo C12 é mais móvel que o C18. Porém, como a seletividade de ambos os sistemas é a mesma, essa mudança de ordem não altera a interpretação sobre a relação da mobilidade dos substratos com a seletividade discutido anteriormente. Além disso, é perceptível que as extremidades dos ligantes próximos à saída da cavidade apresentam maior mobilidade, mas o mesmo também ocorre para a ligação dupla (carbonos 1 e 2) dos produtos de cadeias menores, exceto para o C12. Fica também evidente a disparidade de comportamento dos sistemas provenientes do ácido C20, com a mutante Leu176Gly

Gráfico 8.12 – RMSF dos produtos beta-hidroxilados por átomo de carbono



Gráficos de RMSF para cada átomo de carbono que compõe os produtos beta-hidroxilados na enzima selvagem e na mutante Leu176Gly. Onde, o carbono 1 é referente a carboxila.

apresentando um ligante mais móvel que a enzima selvagem. Ademais, observa-se maiores desvios em relação à média das dinâmicas nos sistemas C14, seguido pelo C10 e C12. Enquanto, os ligantes de cadeias maiores ficam mais próximos à estrutura de referência e apresentando menores variações entre as simulações.

Nos produtos  $\beta$ -hidroxilados, exibido no **Gráfico 8.12**, o sistema C18 apresentou os maiores desvios de flutuação entre as dinâmicas, assim como a mutante Leu176Gly, apesar desta ter os desvios no início da cadeia alifática na região próxima à Arg245. Pelos gráficos fica evidente a disparidade de comportamento entre as enzimas (mutante/selvagem) na presença do produto C20  $\beta$ -hidroxilado.

### 8.3 ANÁLISE DE SOLVATAÇÃO DOS LIGANTES

A análise de solvatação, por falta de um melhor termo, realizada pelo programa cpptraj (**ROE; CHEATHAM, 2013**), conta o número de moléculas de água a cada “frame” de simulação dentro de um raio de distância de 3.4 Å do ligante. Esse valor não representa a primeira camada de hidratação de moléculas de água no entorno do ligante, e foi utilizado nas análises para um estudo preliminar, não refletindo na solvatação real dos produtos dentro da cavidade, pois não é possível acompanhar como as águas se distribuem ao longo

dos ligantes. Os “*frames*” são as “fotografias” do sistema, e foram armazenadas a cada 10.000 intervalos de tempo, ou seja, a frequência  $N$  de passos que a simulação foi salva, onde o intervalo de tempo,  $dt$ , usado foi de 2 femtossegundos. O tempo de simulação,  $t_n$ , pode ser obtido multiplicando:  $t_n = \text{“frames”} \times \text{frequência} \times dt$ .

Os dados dessa análise mostram uma relação entre o tempo de simulação com uma variável inteira, que representa o número de águas presentes dentro da distância de corte. Apesar da cronologia que as observações são contabilizadas, cada evento é independente entre si devido às lacunas que os dados são armazenados. Dessa forma, não se observa uma razão contínua entre os resultados, havendo uma amplitude aleatória de valores entre as medidas. Por exemplo, o primeiro “*frame*” pode conter 5 moléculas, o segundo apenas uma água e o terceiro 12, ou seja, a solvatação não está ocorrendo progressivamente e de maneira uniforme. Para se visualizar a tendência de distribuições aleatórias, os dados de uma das variáveis são divididos em categorias, no caso, o número de moléculas de águas; que são projetados como pontos de ocorrência em um gráfico de dispersão do tipo “*enxame*” (“*swarmplot*”) ao longo do tempo. Devido ao grande número de observações produzidas em apenas uma MD (25.000 pontos), com cada sistema sendo simulado em triplicata, ou seja, para cada valor no eixo x existem 3 observações em y; indicando dados multivariáveis independentes, há um número muito grande de pontos (75.000 pontos). Os gráficos de “*enxame*” são caracterizados por ajustarem as observações categóricas ao longo do seu respectivo eixo para evitar a sobreposição dos pontos, mas devido a escala do gráfico e o grande números de marcações, para categorias com muitas incidências, uma faixa contínua acaba sendo formada.

No **Gráfico 8.13** e **Gráfico 8.14** são mostrados todos os gráficos de “*enxame*”, com as observações — quantidade de água a um raio de 3.4 Å do produto — das três simulações de cada sistema. Além do RMSD de cada ligante ao longo do tempo, para dessa forma, se investigar a relação entre a solvatação ao redor da molécula com a sua mobilidade dentro da cavidade, o que pode sugerir o auxílio na expulsão do produto. Os gráficos de RMSD ao longo do tempo demonstraram o confinamento do sistema alceno C11 na cavidade de reação, como discutido anteriormente.

Os “*swarmplots*” evidenciam que os sistemas  $\beta$ -hidroxilados apresentam maiores quantidades de moléculas de água ao redor dos produtos ao longo das simulações. Entretanto, tanto pela análise das trajetórias quanto pelos desvios de RMSD, verifica-se que nesses casos, apesar de muitos sistemas apresentarem um grande número de moléculas solvatando-os, isso não reflete em um aumento de mobilidade dos produtos  $\beta$ -hidroxilados. Pois, como mencionado anteriormente, os produtos  $\beta$ -hidroxi ficaram durante as simulações restritos ao sítio ativo, ligados a Arg245, com exceções para duas MDs do sistema C18, onde os ligantes após 150 ns, conseguiram romper a ponte salina e se dirigiram à “saída” da cavidade; e a mutante Leu176Gly, que em todas as simulações o produto chegou bem próximo de ser “expelido”.

Gráfico 8.13 – Análise de solvatação C10-C16

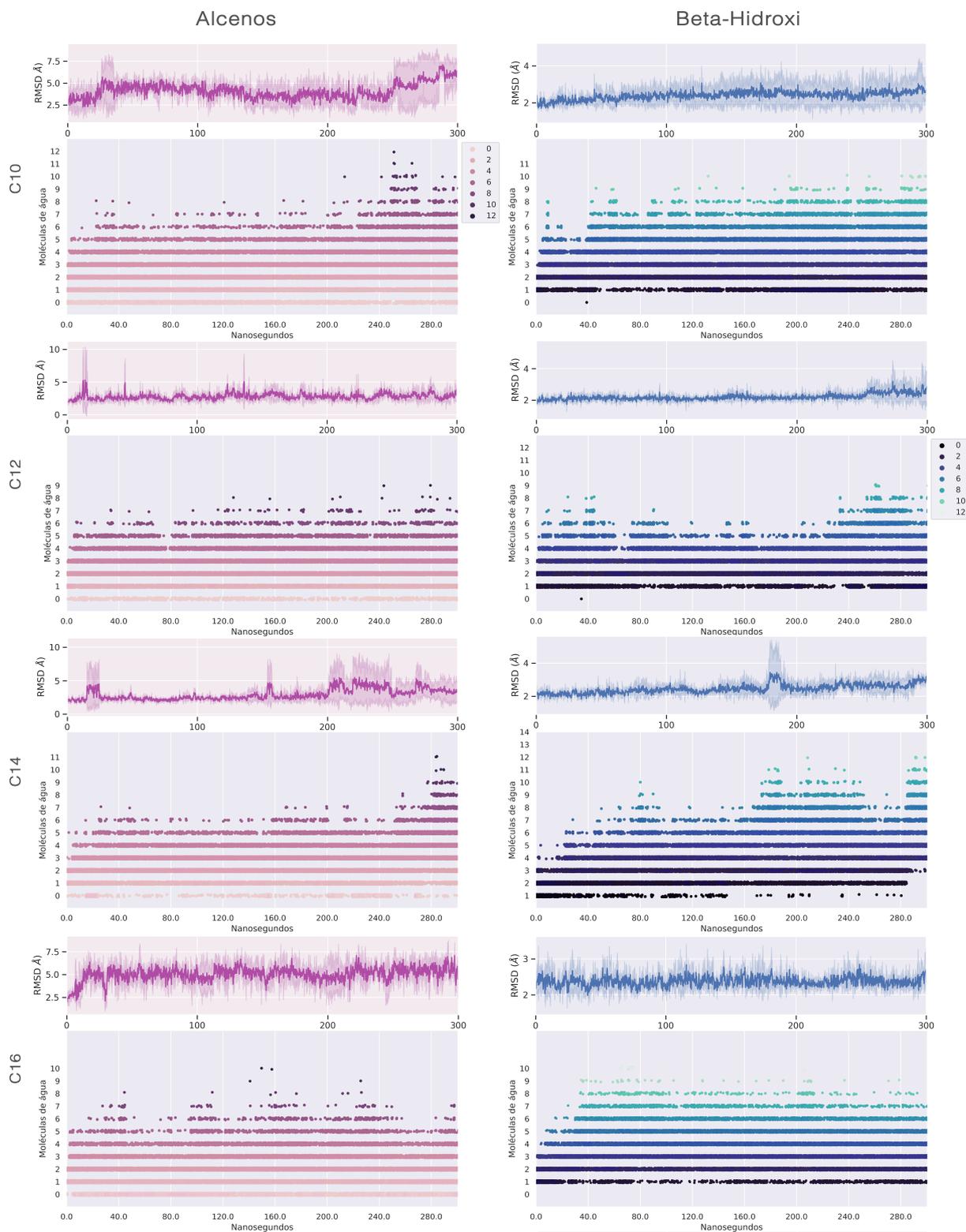


Gráfico “enxame” com a solvatação dos sistemas C10 a C16 ao longo do tempo de simulação. Acima, um gráfico do RMSD dos respectivos produtos ao longo do tempo.



A relação entre a “solvatação” e a mobilidade do produto fica evidente nas olefinas terminais, em especial para os sistemas originados dos ácidos graxos C10 e C20, onde instantes exibindo grandes flutuações de RMSD estão sincronizados com ocorrências de muitas moléculas de água circundando o ligante, nos “*swarmplots*”. Sistemas em que as flutuações se mantiveram em torno de um valor central, ou seja, estruturas mais rígidas — no caso, C12, C18 e C20 —, mostraram uma distribuição com menos visitas a estados mais “solvatados” (Acima de 6 águas). Enquanto, os sistemas mais móveis, C10, C14 e C16, apresentaram um número maior de incidências com mais de 6 águas. Ao passo que, a mutante Leu176Gly apresentou mais observações de pontos em estados mais “solvatados” ao longo de toda a simulação, com muitos picos referentes a grandes desvios de RMSD correlacionadas a pontos de alta “solvatação”. O final da dinâmica para esses sistemas, mostram o produto próximo ao final da cavidade, na iminência de sua expulsão do sítio, sendo mostrada na **Figura 8.6**. Dessa forma, a diminuição na solvatação ao final pode ser condizente com produto em uma região mais confinada, evidenciada pela leve queda na média do desvio de RMSD. Além disso, o perfil entre a C19 da mutante e a enzima selvagem refletem comportamentos distintos, e demonstram que a expulsão do produto de dentro da cavidade é mediado pela presença de moléculas de água dentro da cavidade de ligação, de acordo com as especulações de **Amaya *et al.* (2018)**.

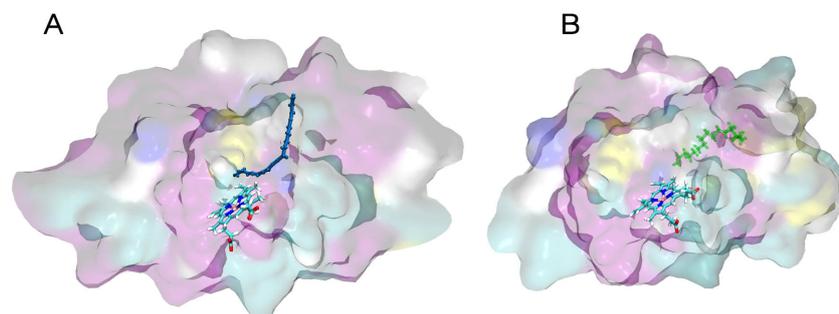


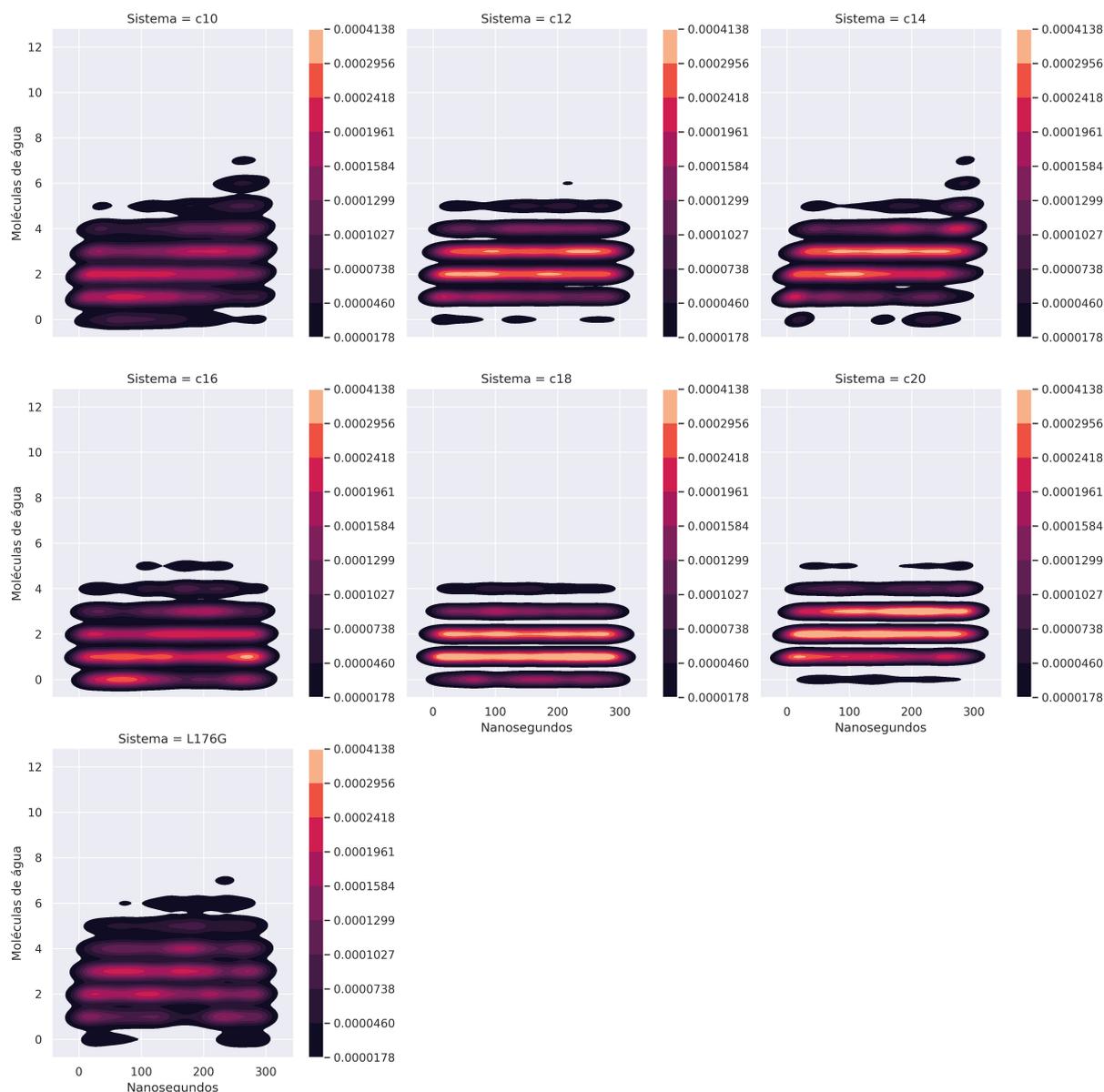
Figura 8.6 – Comparação dos produtos C19 dentro da cavidade

(A) O “*frame*” final da simulação da mutante Leu176Gly na presença do produto C19. A imagem mostra o quão próximo o ligante está de sair da cavidade. (B) Comparativamente, o último “*frame*” do ligante C19 na enzima selvagem, demonstra o produto encaixado na cavidade.

Para melhor visualizar como a distribuição das observações nas categorias mais populadas (de mais baixa solvatação) se dá durante as MDs, para investigar se o número de moléculas de água se mantém constante ou aumenta ao longo da simulação, foram plotados mapas 2-D de estimativa de densidade “*Kernel*” bivariacional, com as probabilidades sendo representadas em uma escala de cores por curvas de níveis (**Gráfico 8.15** e **Gráfico 8.16**). Regiões com uma concentração maior de observações apresentam uma coloração clara, indicando maior densidade de probabilidade e sinalizam um “pico” topográfico; enquanto, regiões mais escuras indicam regiões de baixa densidade. Essa representação

é necessária porque as análises dos gráficos “enxame” se basearam nos pontos que são humanamente perceptíveis, e que justamente representam eventos que não são estatisticamente relevantes. Não sendo possível inferir um comportamento global da distribuição das observações nas categorias mais populadas.

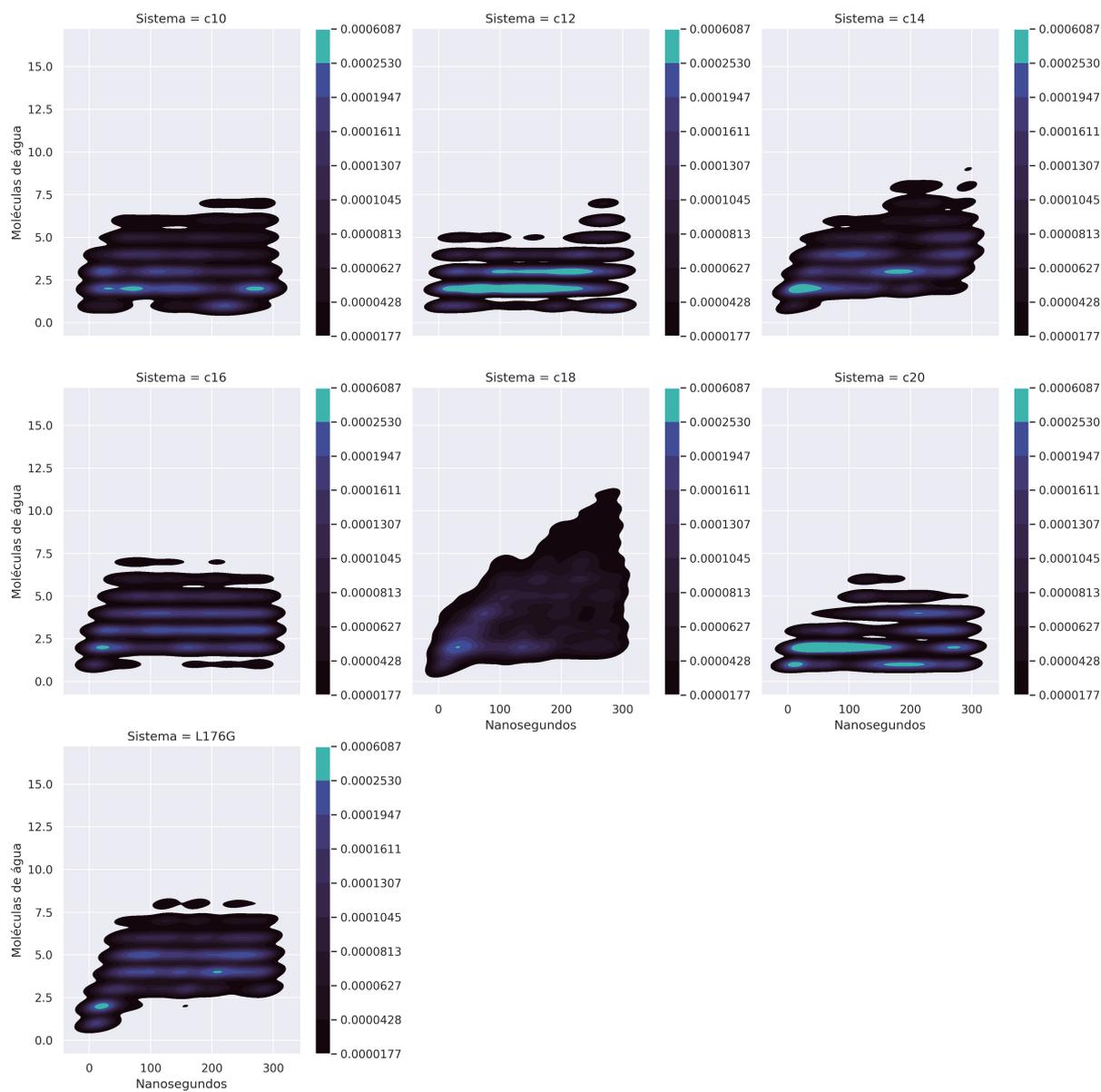
Gráfico 8.15 – Curvas de níveis da solvatação para os alcenos



Curvas de níveis de densidade de probabilidade “Kernel” para as análises de solvatação dos produtos alcenos terminais ao longo da simulação. Todos os sistemas foram nivelados numa escala em comum, para melhor comparação. Observações de baixa probabilidade não são contempladas no ajuste estatístico.

Pela análise dos gráficos **Gráfico 8.15**, observa-se que a mutante Leu176Gly na presença do alceno C20 exemplifica bem um sistema com grande dispersão de valores igualmente prováveis de solvatação ao longo de toda dinâmica. Enquanto nos ligantes alcenos

Gráfico 8.16 – Curvas de níveis da solvatação para os beta-hidroxiados



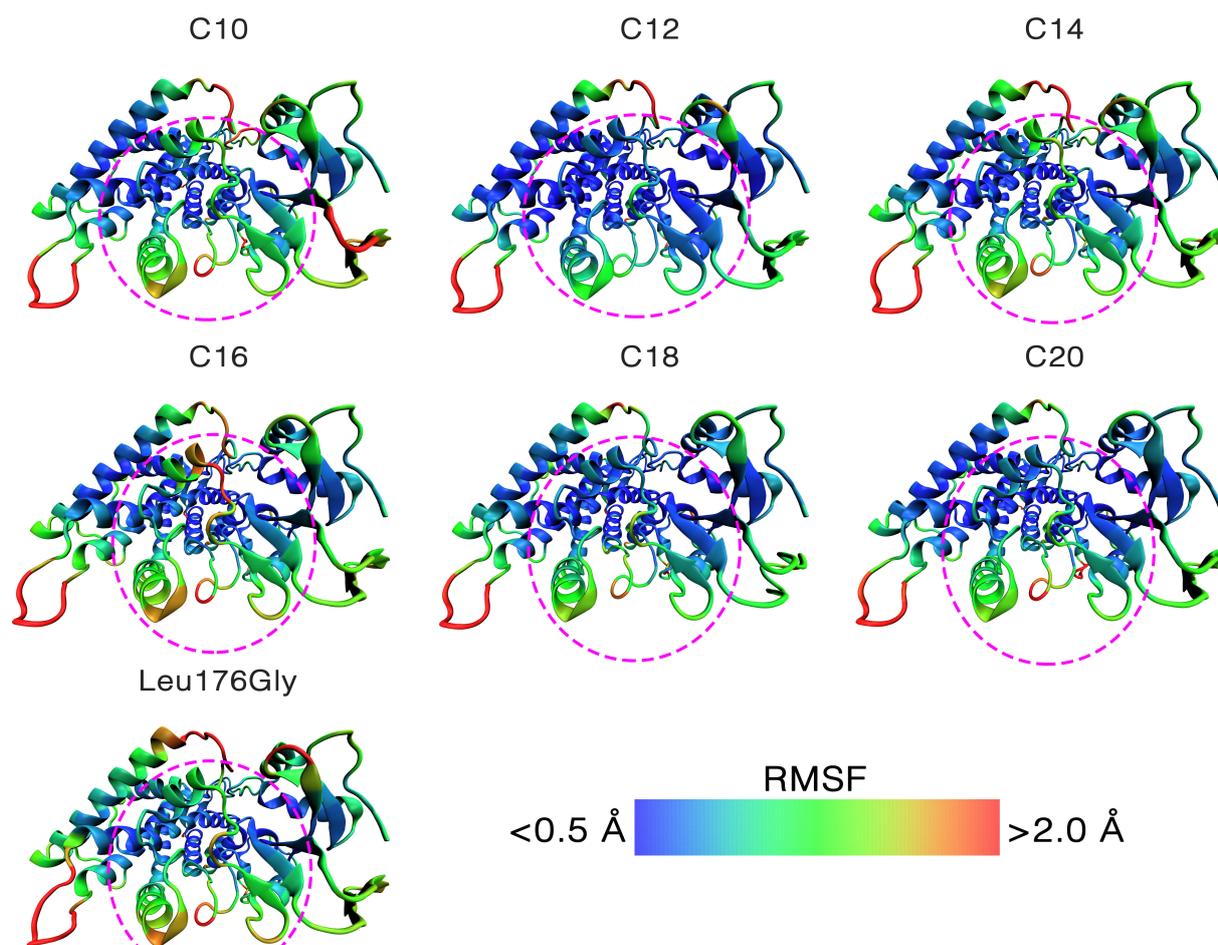
Curvas de níveis de densidade de probabilidade “Kernel” para as análises de solvatação dos produtos beta-hidroxiados ao longo da simulação de MD.

da selvagem, o sistema C10 apresentou a maior solvatação, indicada pela distribuição de probabilidade semelhante em todas as categorias, e com um aumento gradual do número de moléculas com o tempo de simulação, descrito por manchas rosadas se deslocando a categorias mais altas e com o aumento do preenchimento desses estados, conforme o tempo avança na simulação. Seguido pelo C16, que apesar de apresentar certa solvatação com maiores probabilidades em estados equivalente a apenas uma água ou nenhuma, existe uma dispersão maior entre as categorias, além de um aumento gradual no preenchimento de estados mais solvatados. Sendo sua posição reforçada com o gráfico de “enxame”, que mostrou pontos espalhados em categorias de alta solvatação. Dessa forma, combinando os dois gráficos é possível obter uma ordem crescente de “solvatação” para as olefinas terminais como  $C18 < C20 < C12 < C14 < C16 < C10$ , sustentando uma relação entre a mobilidade do produto dentro da cavidade (Tabela 3); e sendo essa observação reforçada com os sistemas  $\beta$ -hidroxilados C20 da mutante e C18 da selvagem que também apresentaram um comportamento distinto de alta solvatação e mobilidade.

Para investigar se a região de encontro de “loops” entre as hélices H/I ou B'/C poderia contribuir com a solvatação do canal reacional ao permitir a criação de canais até a cavidade de reação, as análises de RMSF da enzima mostradas nos “joyplot” foram visualizadas diretamente na estrutura tridimensional da proteína, como mostrada na **Figura 8.7**. Onde, regiões que apresentam variações acima de 2.0 Å adquirem uma coloração em vermelho, abaixo de 0.5 Å são representadas em azul e regiões intermediárias entre esse números são definidas em verde, de acordo com a escala degradê.

Pela **Figura 8.7** é possível observar que a mutante apresenta a hélice B' e no início da C uma flutuação mais intensa que a enzima selvagem. Os sistemas que apresentaram maior solvatação (C10, C14, C16) também mostraram maiores flutuações de RMSF na região pontilhada. Dentre esses, o C16 foi o que apresentou maiores desvios, o que sugere que por ser um ligante maior, haveria uma necessidade de mais mobilidade nessa região para produzir um canal de entrada de água. Enquanto, os sistemas que apresentaram menores solvatações ao longo da simulação (C20, C18 e C12), manifestaram baixas flutuações na região pontilhada. E como todos os sistemas apresentaram uma flexibilidade semelhante no “loop” H/I, é pouco provável que a mobilidade dessa região seja responsável pelos diferentes perfis de “solvatação” obtidos. Dessa forma, os dados sugerem que a região de encontros de “loops” B'/C é a mais provável na criação de um canal de “solvatação” na OleT<sub>JE</sub>. Pois, a sua flexibilidade comparativa em diferentes cadeias se relacionaram melhor com a mobilidade dos produtos do que o perfil da mobilidade do “loop” F/G.

Figura 8.7 – Desvios de RMSF das regiões de B'/C na estrutura tridimensional



Desvios de RMSF indicados por uma escala de cores na estrutura secundária da proteína OleT<sub>JE</sub> para os sistemas com ligantes alceno. Em pontilhado, os enovelamentos B'/C.

## 9 CONCLUSÃO

A partir da técnica computacional de simulação de dinâmica molecular, foi estudada a enzima OleT<sub>JE</sub>, uma descarboxilase capaz de converter ácidos graxos em alcenos terminais com a geração indesejável de ácidos  $\beta$  e  $\alpha$ -hidroxilados. A compreensão dos fatores estruturais e moleculares por trás da regulação da OleT<sub>JE</sub> na presença de substratos de diferentes cadeias é necessária para o desenvolvimento de novas enzimas mais eficientes para produção de bio-hidrocarbonetos, de forma a se ampliar a matriz energética vigente, por uma produção renovável e mais sustentável. Esse trabalho preliminar, fornece informações relevantes sobre os fatores que regem a seletividade e a atividade exibida pela enzima, ao simular o comportamento dessa proteína na presença de diferentes tipos de produtos (olefinas terminais e beta-hidroxilados) com diferentes tamanhos de cadeia alifática.

As análises de RMSD e RMSF da alça F/G na presença de produtos alcenos evidenciaram um padrão de comportamento análogo à atividade enzimática exibida pela enzima selvagem, indicando uma relação direta entre a flexibilidade do “loop” F/G com a liberação do produto dentro da cavidade. Sendo esses dados reforçados com os resultados obtidos na mutante Leu176Gly e na enzima OleTP<sub>RN</sub>. É possível supor, que para se obter uma atividade enzimática alta, sem perdas de seletividade na descarboxilação, é necessário que a flexibilidade da alça F/G se restrinja apenas a região do “loop” entre essas hélices, de forma a garantir a saída facilitada do produto, que ocorre próxima a esse motivo. Ou, alternativamente, com uma redução no número de resíduos que compõem o “loop” F/G, de forma a criar uma entrada mais ampla para a cavidade reacional. Uma maior mobilidade nas regiões das hélices, localizadas próximas ao sítio catalítico, podem influenciar indiretamente no encaixe do substrato no sítio de ligação, de forma a promover a hidroxilação, mas com reflexos diretos na perda da regiosseletividade enzimática.

Adicionalmente, as análises de RMSD dos produtos dentro da cavidade, evidenciaram uma relação entre mobilidade dos ligantes alcenos com a seletividade da enzima selvagem e da mutante. O que pode sugerir que a descarboxilação necessita de um encaixe minucioso do substrato no sítio catalítico, que é intermediado por interações com a cavidade que limitam a mobilidade do ligante. Entretanto, como não foi observado uma relação entre a flexibilidade da F/G com a seletividade exibida pela enzima, esse encaixe e ancoragem não é direcionado pelo respectivo enovelamento. Além disso, não foi observada uma relação entre a flexibilidade da F/G com a mobilidade dos produtos, o que sugere que a mediação da saída dos produtos de dentro da cavidade é possivelmente realizada pela entrada de moléculas de água dentro do canal reacionário. Sendo esse processo evidenciado pelas análises de solvatação. Por fim, as análises de RMSF da enzima para a região “loop” B/B'/C sugerem que a solvatação ocorreria por intermédio desses resíduos, e não pela alça F/G como sugerido por **Amaya et al. (2018)**.

Os objetivos propostos nesse trabalho foram alcançados, e as hipóteses levantadas foram elucidadas. Entretanto, mais estudos são necessários para compreender os fatores que regem a seletividade da enzima e como ampliá-la para substratos de cadeias menores. Portanto, como perspectivas futuras, o presente trabalho propõe que sejam realizadas simulações da enzima OleT<sub>JE</sub> na presença de ácidos graxos de cadeias variadas, análogas a metodologia utilizada e considerando o grupo heme de alto spin pentacoordenado. Dessa forma, seria embasada a suposição que a seletividade é influenciada pela mobilidade dos ligantes no sítio ativo, ou se essa relação obtida no trabalho deriva na realidade da solvatação dos produtos. Além disso, essas simulações também elucidariam se a solvatação prévia do canal reacional poderia influenciar o resultado catalítico da enzima, favorecendo a descarboxilação do ácido graxo, sendo, portanto, de grande interesse no desenvolvimento de futuras mutações da OleT<sub>JE</sub>.

## REFERÊNCIAS

- ALDER, B. J.; WAINWRIGHT, T. E. Phase transition for a hard sphere system. **Journal of Chemical Physics**, v. 27, p. 1208–1209, 1957.
- ALLEN, M.; TILDESLEY, D. **Computer Simulation of Liquids**. New York : Oxford University Press: Clarendon Press, 1996. ISBN 9780198556459.
- AMAYA, J. A.; RUTLAND, C. D.; LESCHINSKY, N.; MAKRIS, T. M. A distal loop controls product release and chemo- and regioselectivity in cytochrome p450 decarboxylases. **Biochemistry**, v. 57, n. 3, p. 344–353, dezembro 2018. doi:10.1021/acs.biochem.7b01065.
- ANDERSEN, H. C. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. **Journal of Computational Physics**, v. 52, n. 1, p. 24–34, outubro 1983. doi:10.1016/0021-9991(83)90014-1.
- ARFIN, T.; MOHAMMAD, F.; YUSOF, N. Biomass resources in environmental and socio-economic analysis of fuel-wood consumption. In: HAKEEM, K.; JAWAID, M.; RASHID, U. **Biomass and Bioenergy: Processing and Properties**. [S.l.]: Springer International Publishing, 2014. p. 191–206. doi:10.1007/978-3-319-07641-6\_12.
- BAYLY, C. I.; CIEPLAK, P.; CORNELL, W.; KOLLMAN, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the resp model. **The Journal of Physical Chemistry**, American Chemical Society, v. 97, p. 10269–10280, outubro 1993. doi:10.1103/PhysRev.56.340.
- BELCHER, J.; MCLEAN, K. J.; MATTHEWS, S.; WOODWARD, L. S.; FISHER, K.; RIGBY, S. E. J.; NELSON, D. R.; POTTS, D.; BAYNHAM, M. T.; PARKER, D. A.; LEYS, D.; MUNRO, A. W. Structure and biochemical properties of the alkene producing cytochrome p450 oletje (cyp15211) from the *Jeotgalicoccus* sp. 8456 bacterium. **Journal of Biological Chemistry**, v. 289, n. 10, p. 6535–6550, março 2014. doi:10.1074/jbc.M113.527325.
- BERENDSEN, H. J. C.; POSTMA, J. P. M.; GUNSTEREN, W. F. van; DINOLA, A.; HAAK, J. R. Molecular dynamics with coupling to an external bath. **The Journal of Chemical Physics**, v. 81, n. 8, p. 3684–3690, 10 1984. doi:10.1063/1.448118.
- BERG, J.; STRYER, L.; TYMOCZKO, J.; GATTO, G. **Biochemistry**. 8. ed. New York: Macmillan Learning, 2015.
- CASE, D. A.; III, T. E. C.; DARDEN, T.; GOHLKE, H.; LUO, R.; JR., K. M. M.; ONUFRIEV, A.; SIMMERLING, C.; WANG, B.; WOODS, R. J. The amber biomolecular simulation programs. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1668–1688, 2005. doi:10.1002/jcc.20290.
- CASPETA, L.; NIELSEN, J. Economic and environmental impacts of microbial biodiesel. **Nature Biotechnology**, v. 31, n. 9, p. 1546–1696, setembro 2013. doi:10.1038/nbt.2683.

- COLEMAN, M.; STANTURF, J. A. Biomass feedstock production systems: Economic and environmental benefits. **Biomass and Bioenergy**, v. 30, n. 8, p. 693–695, agosto 2006. doi:10.1016/j.biombioe.2006.04.003.
- COOMBS, J. R.; MORKEN, J. P. Catalytic enantioselective functionalization of unactivated terminal alkenes. **Angew Chem Int Ed Engl**, v. 55, n. 8, p. 2636–2649, janeiro 2016. doi:10.1002/anie.201507151.
- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh ewald: An  $n \cdot \log(n)$  method for ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089–10092, 1993. doi:10.1063/1.464397.
- DILL, K.; MACCALLUM, J. The protein-folding problem, 50 years on. **Science (New York, N.Y.)**, v. 338, p. 1042–6, novembro 2012. doi:10.1126/science.1219021.
- DODANI, S. C.; KISS, G.; CAHN, J. K. B.; SU, Y.; PANDE, V. S.; ARNOLD, F. H. Discovery of a regioselectivity switch in nitrating p450s guided by molecular dynamics simulations and markov models. **Nature Chemistry**, v. 8, n. 5, p. 419–425, março 2016. doi:10.1038/nchem.2474.
- DU, J.; LIU, L.; GUO, L. Z.; YAO, X. J.; YANG, J. M. Molecular basis of p450 oletje: an investigation of substrate binding mechanism and major pathways. **Journal of Computer-Aided Molecular Design**, v. 31, n. 5, p. 483–495, maio 2017. doi:10.1007/s10822-017-0013-x.
- DUBEY, K. D.; SHAIK, S. Cytochrome p450—the wonderful nanomachine revealed through dynamic simulations of the catalytic cycle. **Accounts of Chemical Research**, v. 52, n. 2, p. 389–399, fevereiro 2019. doi:10.1021/acs.accounts.8b00467.
- ESSMANN, U.; PERERA, L.; BERKOWITZ, M. L.; DARDEN, T.; LEE, H.; PEDERSEN, L. G. A smooth particle mesh Ewald method. **The Journal of Chemical Physics**, v. 103, n. 19, p. 8577–8593, novembro 1995. doi:10.1063/1.470117.
- EWALD, P. P. Die berechnung optischer und elektrostatischer gitterpotentiale. **Annalen der Physik**, v. 369, n. 3, p. 253–287, 1921. doi:10.1002/andp.19213690304.
- FORESTER, T. R.; SMITH, W. Shake, rattle, and roll: Efficient constraint algorithms for linked rigid bodies. **Journal of Computational Chemistry**, v. 19, n. 1, p. 102–111, janeiro 1998. doi:10.1002/(SICI)1096-987X(19980115)19:1<102::AID-JCC9>3.0.CO;2-T.
- FRISHMAN, D.; ARGOS, P. Knowledge-based protein secondary structure assignment. **Proteins: Structure, Function, and Bioinformatics**, v. 23, n. 4, p. 566–579, 1995. doi:10.1002/prot.340230412.
- FUJISHIRO, T.; SHOJI, O.; NAGANO, S.; SUGIMOTO, H.; SHIRO, Y.; WATANABE, Y. Crystal structure of  $h_2O_2$ -dependent cytochrome p450<sub>SP $\alpha$</sub>  with its bound fatty acid substrate. insight into the regioselective hydroxylation of fatty acids at the  $\alpha$  position. **Journal of Biological Chemistry**, v. 286, n. 34, p. 29941–29950, agosto 2011. doi:10.1074/jbc.M111.245225.
- GOH, C. S.; LEE, K. T. A visionary and conceptual macroalgae-based third-generation bioethanol (tgb) biorefinery in sabah, malaysia as an underlay for renewable and

sustainable development. **Renewable and Sustainable Energy Reviews**, v. 14, n. 2, p. 842–848, fevereiro 2010. doi:10.1016/j.rser.2009.10.001).

GOODISMAN, J. **Statistical Mechanics for Chemists**. New York: Wiley, 1997.

GORDON, J. C.; MYERS, J. B.; FOLTA, T.; SHOJA, V.; HEATH, L. S.; ONUFRIEV, A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. **Nucleic Acids Research**, v. 33, n. suppl\_2, p. W368–W371, julho 2005. doi:10.1093/nar/gki464.

GRANT, J. L.; HSIEH, C. H.; MAKRIS, T. M. Decarboxylation of fatty acids to terminal alkenes by cytochrome p450 compound i. **Journal American Chemical Society**, v. 137, n. 15, p. 4940–4943, abril 2015. doi:10.1021/jacs.5b01965.

GRIEBEL, M.; KNAPEK, S.; ZUMBUSCH, G. W. Numerical simulation in molecular dynamics: Numerics, algorithms, parallelization, applications. In: . [S.l.: s.n.], 2007.

HOSPITAL, A.; GOÑI, J. R.; OROZCO, M.; GELPI, J. L. Molecular dynamics simulations: advances and applications. **Advances and Applications in Bioinformatics and Chemistry**, v. 8, p. 37–47, novembro 2015. doi:10.1042/BST20170218.

HSIEH, C. H.; HUANG, X.; AMAYA, J. A.; RUTLAND, C. D.; KEYS, C. L.; GROVES, J. T.; AUSTIN, R. N.; MAKRIS, T. M. The enigmatic p450 decarboxylase olet is capable of, but evolved to frustrate, oxygen rebound chemistry. **Biochemistry**, v. 56, n. 26, p. 3347–3357, setembro 2017. doi:10.1021/acs.biochem.7b00338.

HUBER, R. Conformational flexibility and its functional significance in some protein molecules. In: HOLZER HELMUTAND TSCHESCHE, H. (Ed.). **Biological Functions of Proteinases**. Berlin, Heidelberg: Springer Berlin Heidelberg, 1979. p. 1–16. ISBN 978-3-642-81395-5. doi:10.1007/978 - 3 - 642 - 81395 - 5<sub>1</sub>.

JANEČEK, M.; KÜHROVÁ, P.; MLÝNSKÝ, V.; OTYEPKA, M.; ŠPONER, J.; BANÁŠ, P. W-resp: Well-restrained electrostatic potential derived charges. revisiting the charge derivation model. **Journal of Chemical Theory and Computation**, v. 17, n. 6, p. 3495–3509, maio 2021. doi:acs.jctc.0c00976.

JORGENSEN, W. L.; CHANDRASEKHAR, J.; MADURA, J. D.; IMPEY, R. W.; KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. **The Journal of Chemical Physics**, v. 79, n. 2, p. 926–935, 1983. doi:10.1063/1.445869.

KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, v. 22, n. 12, p. 2577–2637, dezembro 1983. doi:10.1002/bip.360221211.

KOKKINIDIS, M.; GLYKOS, N. M.; FADOULOGLOU, V. E. Chapter 7 - protein flexibility and enzymatic catalysis. In: CHRISTOV, C.; KARABENCHEVA-CHRISTOVA, T. (Ed.). **Structural and Mechanistic Enzymology**. [S.l.]: Academic Press, 2012, (Advances in Protein Chemistry and Structural Biology, v. 87). p. 181–218. doi:10.1016/B978-0-12-398312-1.00007-X.

KOSHLAND, D. E. J. Enzyme flexibility and enzyme action. **Journal of Cellular and Comparative Physiology**, v. 54, n. S1, p. 245–258, 1959. doi:10.1002/jcp.1030540420.

LEACH, A. R. **Molecular Modelling: Principles and Applications**. 2. ed. England: Prentice Hall, 2001.

LEE, R. A.; LAVOIE, J.-M. From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity. **Animal Frontiers**, v. 3, n. 2, p. 6–11, abril 2013. doi:10.2527/af.2013-0010.

LIU, Y.; WANG, C.; YAN, J.; ZHANG, W.; GUAN, W.; LU, X.; LI, S. Hydrogen peroxide-independent production of  $\alpha$ -alkenes by *olet<sub>JE</sub> p450* fatty acid decarboxylase. **Biotechnology for Biofuels**, v. 7, n. 28, p. 1754–6834, fevereiro 2014. doi:10.1186/1754-6834-7-28.

MACHADO, A. P.; LIRIA, C. W.; PROTI, P. B.; REMUZGO, C.; MIRANDA, M. T. M. de. Sínteses química e enzimática de peptídeos: princípios básicos e aplicações. **Química Nova**, v. 27, p. 781–789, 2004.

MACKERELL, A. D. J. Atomistic models and force fields. In: BECKER, O.M. AND MACKERELL, A.D. AND ROUX, B. AND WATANABE, M. **Computational Biochemistry and Biophysics**. New York: Marcel Dekker, 2001.

MAIER, J. A.; MARTINEZ, C.; KASAVAJHALA, K.; WICKSTROM, L.; HAUSER, K. E.; SIMMERLING, C. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. **Journal of Chemical Theory and Computation**, v. 11, n. 8, p. 3696–3713, 2015. doi:10.1021/acs.jctc.5b00255.

MARTÍNEZ, L.; BORIN, I.; SKAF, M. Fundamentos de simulação por dinâmica molecular. In: MORGON, NELSON H. AND COUTINHO, KALINE (EDS). **Métodos de Química Teórica e Modelagem Molecular**. São Paulo: Editora Livraria da Física, 2007.

MATSUNAGA, I.; UEDA, A.; FUJIWARA, N.; SUMIMOTO, T.; ICHIHARA, K. Characterization of the *ybdt* gene product of *Bacillus subtilis*: Novel fatty acid  $\beta$ -hydroxylating cytochrome p450. **Lipids**, v. 34, p. 841–846, 1999c. doi:10.1007/s11745-999-0431-3.

MATSUNAGA, I.; YAMADA, M.; KUSUNOSE, E.; NISHIUCHI, Y.; YANO, I.; ICHIHARA, K. Direct involvement of hydrogen peroxide in bacterial  $\alpha$ -hydroxylation of fatty acid. **FEBS Letters**, v. 386, n. 2-3, p. 252–254, 1996a. doi:10.1016/0014-5793(96)00451-6.

MATSUNAGA, I.; YOKOTANI, N.; GOTOH, O.; KUSUNOSE, E.; YAMADA, M.; ICHIHARA, K. Molecular cloning and expression of fatty acid  $\alpha$ -hydroxylase from *Sphingomonas paucimobilis*\*. **Journal of Biological Chemistry**, v. 272, n. 38, p. 23592–23596, setembro 1997b. doi:10.1074/jbc.272.38.235927.

MATTHEWS, S.; BELCHER, J. D.; TEE, K. L.; GIRVAN, H. M.; MCLEAN, K. J.; RIGBY S. E. J. AND LEVY, C. W.; LEYS, D.; PARKER, D. A.; BLANKLEY, R. T.; MUNRO, A. W. Catalytic determinants of alkene production by the cytochrome p450 peroxygenase *olet<sub>JE</sub>*. **Journal of Biological Chemistry**, v. 292, n. 12, p. 5128–5143, março 2017. doi:10.1074/jbc.M116.762336.

MCCAMMON, J. A.; GELIN, B. R.; KARPLUS, M. Dynamics of folded proteins. **Nature**, v. 267, p. 585–590, junho 1977. doi:10.1038/267585a01.

- MCQUARRIE, D. **Statistical Mechanics**. New York: Harper & Row, 1976.
- MUNRO, A. W.; MCLEAN, K. J.; GRANT, J. L.; MAKRIS, T. M. Structure and function of the cytochrome p450 peroxygenase enzymes. **Biochemical Society Transactions**, v. 46, n. 1, p. 183–196, fevereiro 2018. doi:10.1042/BST20170218.
- NELSON, D. L.; COX, M. M. **Lehninger Principles of Biochemistry**. 6. ed. New York: WH Freeman and Company, 2014.
- PERALTA-YAHYA, P. P.; ZHANG, F.; CARDAYRE, S. B. del; KEASLING, J. D. Microbial engineering for the production of advanced biofuels. **Nature**, v. 488, n. 7411, p. 320–328, agosto 2012. doi:10.1038/nature11478.
- RADE, L. L.; GENEROSO, W. C.; DAS, S.; SOUZA, A. S.; SILVEIRA, R. L.; AVILA, M. C.; VIEIRA, P. S.; MIYAMOTO, R. Y.; LIMA, A. B. B.; ARICETTI, J. A.; MELO, R. R. de; MILAN, N.; PERSINOTI, G. F.; BONOMI, A. M. F. L. J.; MURAKAMI, M. T.; MAKRIS, T. M.; ZANPHORLIN, L. M. Dimer-assisted mechanism of (un)saturated fatty acid decarboxylation for alkene production. **Proceedings of the National Academy of Sciences**, v. 120, n. 22, p. e2221483120, 2023. doi:10.1073/pnas.2221483120.
- ROE, D. R.; CHEATHAM, T. E. I. Ptraaj and cptraaj: Software for processing and analysis of molecular dynamics trajectory data. **Journal of Chemical Theory and Computation**, v. 9, n. 7, p. 3084–3095, 2013. doi:10.1021/ct400341p.
- RUDE, M. A.; BARON, T. S.; BRUBAKER, S.; ALIBHAI, M.; CARDAYRE, S. B. D.; SCHIRMER, A. Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. **Applied and Environmental Microbiology**, v. 77, n. 5, p. 1718–1727, março 2011. doi:10.1128/AEM.02580-10.
- RYCKAERT, J.-P.; CICCOTTI, G.; BERENDSEN, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. **Journal of Computational Physics**, v. 23, n. 3, p. 327–341, 1977. doi:10.1016/0021-9991(77)90098-5.
- SCHIRMER, A.; RUDE, M. A.; LI, X.; POPOVA, E.; CARDAYRE, S. B. del. Microbial biosynthesis of alkanes. **Science**, v. 329, n. 5991, p. 559–562, 2010. doi:10.1126/science.1187936.
- SCRUTTON, N. S. Enzymes make light work of hydrocarbon production. **Science**, v. 357, n. 6354, p. 872–873, 2017. doi:10.1126/science.aao4399.
- SHAHROKH, K.; ORENDET, A.; YOST, G. S.; III, T. E. C. Quantum mechanically derived amber-compatible heme parameters for various states of the cytochrome p450 catalytic cycle. **Journal of Computational Chemistry**, v. 33, n. 2, p. 119–133, 2012. doi:10.1002/jcc.21922.
- SOLOMONS, T.; FRYHLE, C.; SNYDER, S.; SONS, J. W. . **Solomons' Organic Chemistry**. 6. ed. New York: John Wiley & Sons, Limited, 1996.
- SWOPE, W. C.; ANDERSEN, H. C.; BERENS, P. H.; WILSON, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of

physical clusters of molecules: Application to small water clusters. **The Journal of Chemical Physics**, v. 76, n. 1, p. 637–649, janeiro 1982. ISSN 0021-9606. doi:10.1063/1.442716.

VERLET, L. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. **Phys. Rev.**, v. 159, p. 98–103, julho 1967. doi:10.1103/PhysRev.159.98.

VOET, D.; VOET, J. G. **Biochemistry**. 2. ed. New York: John Wiley and Sons, 1995. v. 23. doi:10.1016/0307-4412(95)90658-4.

VOLCHAN, S. B. A probabilidade na mecânica estatística clássica. **Revista Brasileira de Ensino de Física**, Sociedade Brasileira de Física, v. 28, n. 3, p. 313–318, 2006. ISSN 1806-1117. doi:10.1590/S1806-11172006000300007.

WANG, J.; WANG, W.; KOLLMAN, P. A.; CASE, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. **Journal of Molecular Graphics and Modelling**, v. 25, n. 2, p. 247–260, 2006. doi:10.1016/j.jmgm.2005.12.005.

WANG, J.; WOLF, R. M.; CALDWELL, J. W.; KOLLMAN, P. A.; CASE, D. A. Development and testing of a general amber force field. **Journal of Computational Chemistry**, v. 25, n. 9, p. 1157–1174, 2004. doi:10.1002/jcc.20035.

WANG, J.; ZHU, K. Microbial production of alka(e)ne biofuels. **Current Opinion in Biotechnology**, v. 50, p. 11–18, abril 2018. doi:10.1016/j.copbio.2017.08.009.

WANG, Y.; LAN, D.; DURRANI, R.; HOLLMANN, F. Peroxygenases en route to becoming dream catalysts. what are the opportunities and challenges? **Current Opinion in Chemical Biology**, v. 37, p. 1–9, abril 2017. doi:10.1016/j.cbpa.2016.10.007.

WILLIAMS, C. L.; DAHIYA, A.; PORTER, P. Introduction to bioenergy and waste to energy. In: DAHIYA, A. **Bioenergy: Biomass to Biofuels and Waste to Energy**. 2. ed. [S.l.]: Academic Press, 2020. p. 5–44. doi:10.1016/B978-0-12-815497-7.00001-4).

WOLFENDEN, R.; SNIDER, M. J. The depth of chemical time and the power of enzymes as catalysts. **Accounts of Chemical Research**, v. 34, n. 12, p. 938–945, 2001. doi:10.1021/ar000058i.

WOOD, W. **Early history of computer simulations in statistical mechanics**. New Mexico, USA, 1985. 32 p.

ZHOU, Y. J.; KERKHOVEN, E. J.; NIELSEN, J. Barriers and opportunities in bio-based production of hydrocarbons. **Nature Energy**, v. 3, n. 11, p. 925–935, novembro 2018. doi:10.1038/s41560-018-0197-x.

## GLOSSÁRIO

**ab initio** Conjunto de técnicas computacionais que buscam solucionar diretamente a equação de Schrödinger para o Hamiltoniano não-relativístico de um sistema poliatômico.

**ABO** Aproximação de Born-Oppenheimer.

**AMBER** “*Assisted Model Building and Energy Refinement*”.

**comprimento de equilíbrio** Distância intermolecular de menor energia conformacional adotada perante a contribuição de todas as interações a que o sistema é submetido.

**ESP** Potencial Eletrostático - “*Electrostatic Potential*”.

**família** *Proteínas individuais são associadas a famílias com base no grau de semelhança nas sequências de aminoácidos (estrutura primária).*

**fs** Femtossegundos:  $10^{-15}$  s.

**Gln85His** Em bioquímica, proteínas mutantes são expressas pelo aminoácido conservado na enzima selvagem (WD -*wild type*) com a sua respectiva posição na sequência peptídica, seguido pelo resíduo mutado.

**identidade** Identidade da sequência de aminoácidos refere-se à presença dos mesmos aminoácidos em uma mesma ordem em duas proteínas distintas quando suas respectivas sequências de aminoácidos são previamente alinhadas.

**iteração** Processo computacional de repetir uma ação, onde esta representa o procedimento de obter o resultado de uma equação através de sucessivos cálculos; por um período de tempo até que uma determinada condição seja alcançada.

**MANIAC** Mathematical Analyzer, Numerical Integrator And Computer.

**MD** Dinâmica Molecular - “*Molecular Dynamics*”.

**momento de dipolo** Grandeza vetorial ( $\vec{\mu}$ ) cuja magnitude ( $\mu$ ) é igual ao produto da carga elétrica pela distância que separa as cargas, tem a direção do segmento de reta que une os centros das cargas e o sentido da carga negativa para a positiva.

**Monte Carlo** Algoritmo capaz de gerar valores aleatórios baseados em uma distribuição de probabilidades para qualquer variável que tenha incertezas inerentes.

**motivo** Padrão de enovelamento identificável, envolvendo dois ou mais elementos da estrutura secundária e a conexão (ou conexões) entre eles (“loops”).

**ns** Nanosegundos:  $10^{-9}$  s.

**PBC** “*Periodic Boundary Conditions*”.

**PDB** Base de dados de proteínas “*Protein data base*”.

**plasticidade estrutural** Capacidade de se ligar a diferentes substratos promovendo a mesma reação química. Promiscuidade proteica.

**PME** *“Particle mesh Ewald”*.

**ps** Picossegundos:  $10^{-12}$  s.

**RESP** Potencial Eletrostático Restrito - *“Restrained Electrostatic Potential”*.

**RMSD** Raiz quadrada do deslocamento quadrático médio - *“Root mean square deviation”*.

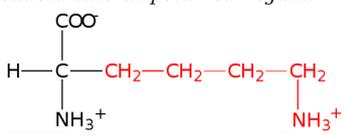
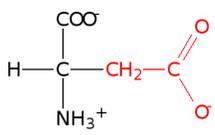
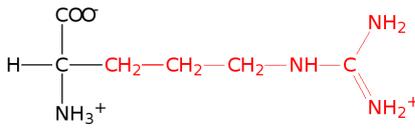
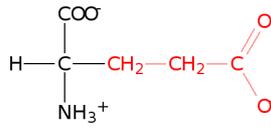
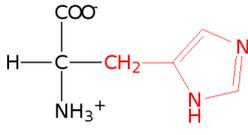
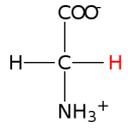
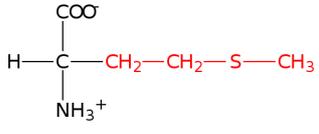
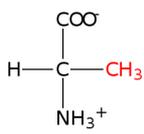
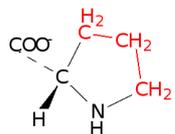
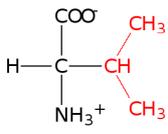
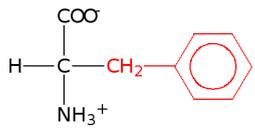
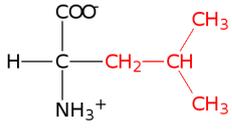
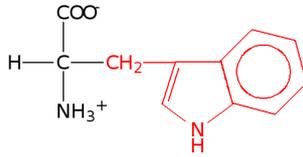
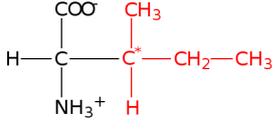
**RMSF** Raiz quadrada da flutuação média quadrática - *“Root mean square fluctuation”*.

**SEP** Superfície de energia potencial.

**turnover** Número de moléculas de substrato transformadas por tempo por uma única molécula de enzima, quando a enzima é o fator limitante da velocidade de reação.

**ANEXOS**

## ANEXO A – TABELA DOS AMINOÁCIDOS QUE COMPÕEM AS PROTEÍNAS E SUAS SIGLAS

Nome, Abreviações	Fórmula estrutural	Nome, Abreviações	Fórmula estrutural
<i>Aminoácidos de cadeia lateral polar carregada</i>			
Lisina Lys K		Ácido aspártico Asp D	
Arginina Arg R		Ácido Glutâmico Glu E	
Histidina His H			
<i>Aminoácidos de cadeia lateral apolar</i>			
Glicina Gly G		Metionina Met M	
Alanina Ala A		Prolina Pro P	
Valina Val V		Fenilalanina Phe F	
Leucina Leu L		Triptofano Trp W	
Isoleucina Ile I			

Nome, Abreviações	Fórmula estrutural	Nome, Abreviações	Fórmula estrutural
<i>Aminoácidos de cadeia lateral polar sem carga</i>			
Serina Ser S	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2-\text{OH} \\   \\ \text{NH}_3^+ \end{array}$	Glutamina Gln Q	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2-\text{CH}_2-\text{C}(=\text{O})\text{NH}_2 \\   \\ \text{NH}_3^+ \end{array}$
Treonina Thr T	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{C}^*(\text{H})-\text{CH}_3 \\   \quad   \\ \text{NH}_3^+ \quad \text{OH} \end{array}$	Tirosina Tyr Y	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2-\text{C}_6\text{H}_4-\text{OH} \\   \\ \text{NH}_3^+ \end{array}$
Asparagina Asn N	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2-\text{C}(=\text{O})\text{NH}_2 \\   \\ \text{NH}_3^+ \end{array}$	Cisteína Cys C	$\begin{array}{c} \text{COO}^- \\   \\ \text{H}-\text{C}-\text{CH}_2-\text{SH} \\   \\ \text{NH}_3^+ \end{array}$

# ANEXO B – TABELA PERIÓDICA DE ELEMENTOS QUÍMICOS

Figura B.1 – Tabela Periódica de Elementos Químicos

1 IA																												18 VIIIA								
1	1.0079																	2		4.0026																
1	H Hidrogênio																	He Hélio																		
2	3 6.941																4 9.0122																			
2	Li Lítio																Be Berílio																			
3	11 22.990												12 24.305																							
3	Na Sódio												Mg Magnésio																							
4	13 IIIA																			14 IVA	15 VA	16 VIA	17 VIIA	18 VIIIA												
4	19 39.098																			20 40.078	21 44.956	22 47.867	23 50.942	24 51.996	25 54.938	26 55.845	27 58.933	28 58.693	29 63.546	30 65.39	31 69.723	32 72.64	33 74.922	34 78.96	35 79.904	36 83.8
4	K Potássio																			Ca Cálcio	Sc Escândio	Ti Titânio	V Vanádio	Cr Crômio	Mn Manganês	Fe Ferro	Co Cobalto	Ni Níquel	Cu Cobre	Zn Zinco	Ga Gálio	Ge Germânio	As Arsênio	Se Selênio	Br Bromo	Kr Criptônio
5	37 85.468																			38 87.62	39 88.906	40 91.224	41 92.906	42 95.94	43 96	44 101.07	45 102.91	46 106.42	47 107.87	48 112.41	49 114.82	50 118.71	51 121.76	52 127.6	53 126.9	54 131.29
5	Rb Rubídio																			Sr Estrôncio	Y Ítrio	Zr Zircônio	Nb Nióbio	Mo Molibdênio	Tc Tecnécio	Ru Rutenio	Rh Ródio	Pd Paládio	Ag Prata	Cd Cádmio	In Índio	Sn Estanho	Sb Antimônio	Te Telúrio	I Iodo	Xe Xenônio
6	55 132.91																			56 137.33	57-71 La-Lu Lantanídeos	72 178.49	73 180.95	74 183.84	75 186.21	76 190.23	77 192.22	78 195.08	79 196.97	80 200.59	81 204.38	82 207.2	83 208.98	84 209	85 210	86 222
6	Cs Césio																			Ba Bário	La-Lu Lantanídeos	Hf Háfnio	Ta Tântalo	W Tungstênio	Re Rênio	Os Osmio	Ir Iridio	Pt Platina	Au Ouro	Hg Mercúrio	Tl Tálio	Pb Chumbo	Bi Bismuto	Po Polônio	At Astatato	Rn Radônio
7	87 223																			88 226	89-103 Ac-Lr Actinídeos	104 261	105 262	106 266	107 264	108 277	109 268	110 281	111 280	112 285	113 284	114 289	115 288	116 293	117 292	118 294
7	Fr Frâncio																			Ra Rádio	Ac-Lr Actinídeos	Rf Rutherfordório	Db Dubnio	Sg Seabúrgio	Bh Bóhrio	Hs Hásso	Mt Meitnério	Ds Darmstádio	Rg Roentgênio	Cn Copernício	Nh Nihônio	Fl Fleróvio	Mc Moscóvio	Lv Livermório	Ts Tenessio	Og Oganesônio
																			57 138.91	58 140.12	59 140.91	60 144.24	61 145	62 150.36	63 151.96	64 157.25	65 158.93	66 162.50	67 164.93	68 167.26	69 168.93	70 173.04	71 174.97			
																			La Lantânio	Ce Cério	Pr Praseodímio	Nd Neodímio	Pm Promécio	Sm Samário	Eu Európio	Gd Gadolínio	Tb Terbio	Dy Disprósio	Ho Hólmio	Er Érbio	Tm Túlio	Yb Ítrio	Lu Lutécio			
																			89 227	90 232.04	91 231.04	92 238.03	93 237	94 244	95 243	96 247	97 247	98 251	99 252	100 257	101 258	102 259	103 262			
																			Ac Actínio	Th Tório	Pa Protactínio	U Urânio	Np Netúnio	Pu Plutônio	Am Americo	Cm Cúrio	Bk Berquílio	Cf Califórnio	Es Einstênio	Fm Férnio	Md Mendelévio	No Nobélio	Lr Laurêncio			

Tabela elaborada pela autora.