



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

# **SARS-CoV-2 Infecção e Replicação: Estudo da Flexibilização Conformacional por Métodos de Simulação Molecular e Química Quântica**

MSc. José Gutembergue de Mendonça

1º Orientador: Prof. Dr. Gerd Bruno da Rocha

2º Orientador: Prof. Dr. Leonardo Henrique Franca de Lima

*João Pessoa*

*- 07 de Agosto, 2024 -*





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

# **SARS-CoV-2 Infecção e Replicação: Estudo da Flexibilização Conformacional por Métodos de Simulação Molecular e Química Quântica**

MSc. José Gutembergue de Mendonça

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

1º Orientador: Prof. Dr. Gerd Bruno da Rocha

2º Orientador: Prof. Dr. Leonardo Henrique Franca de Lima

*João Pessoa*

*- 07 de Agosto, 2024 -*

# **SARS-CoV-2 Infecção e Replicação: Estudo da Flexibilização Conformacional por Métodos de Simulação Molecular e Química Quântica.**

Trabalho de Tese de Doutorado apresentada pelo discente **JOSÉ GUTEMBERGUE DE MENDONÇA** e aprovada pela Comissão Examinadora composta abaixo, realizada no dia 28 de agosto de 2024.

**Prof. Dr. Gerd Bruno da Rocha**  
**DQ/UFPB**  
**Orientador/Presidente**

**Prof. Dr. José Fernando Ruggiero Bachega**  
**UFCS/Porto Alegre-RS**  
**Examinador externo**

**Prof. Dr. Alex Gutterres Taranto**  
**UFSJDel Rei/MG**  
**Examinador externo**

**Prof<sup>a</sup>. Dr<sup>a</sup> Karen Cacilda Weber**  
**DQ/UFPB**  
**Examinadora interna**

**Prof. Dr. Marcus Tullius Scotti**  
**DQ/UFPB**  
**Examinador externo**

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

M539s Mendonça, José Gutemberg de.

SARS-CoV-2 infecção e replicação : estudo da flexibilização conformacional por métodos de simulação molecular e química quântica / José Gutemberg de Mendonça. - João Pessoa, 2024.

130 f. : il.

Orientação: Gerd Bruno da Rocha, Leonardo Henrique Franca de Lima.

Tese (Doutorado) - UFPB/CCEN.

1. Coronavírus - SARS-CoV-1. 2. SARS-CoV-2 - Coronavírus. 3. Papain-Like protease. 4. Proteína Spike. 5. Dinâmica molecular. I. Rocha, Gerd Bruno da. II. Lima, Leonardo Henrique Franca de. III. Título.

UFPB/BC

CDU 578.834(043)

## Agradecimentos

Agradeço, primeiramente, a Deus pelo dom da vida e por sempre me dar força espiritual necessária para realizar meus objetivos. A minha família, em especial a meu pai (Deoclécio) e a minha mãe (Quitéria) que entenderam o motivo de ter saído aos 15 anos de casa para estudar e trabalhar, suportando a distância e a saudade. Jamais esquecerei as últimas palavras da minha mãe quando resolvi sair de casa, pois elas marcaram para sempre a minha vida.

Aos meus orientadores Prof. Dr. Gerd Bruno da Rocha e Prof. Dr. Leonardo Henrique Franca de Lima, pela paciência, orientação, correções e suporte durante a execução deste trabalho.

A todos os professores que contribuíram para a minha formação ao longo da minha vida. Um médico faz uma cirurgia para salvar uma vida em algum órgão vital, já o professor é cirúrgico em salvar vidas, transformando a vida dos seus alunos para sempre. Quantos professores que passaram por minha vida e foram cirúrgicos, sem eles não seria a pessoa que sou hoje. Gratidão eterna a todos, não citarei nomes para não cometer a injustiça de esquecer o nome de alguém.

A minha esposa Lithiely, pela paciência, compreensão e companheirismo, sempre torcendo para que eu realize todos os meus objetivos pessoais, além de todas as palavras de incentivo e conforto ao longo do caminho. Aos amigos de laboratório que mesmo no período da pandemia onde as novas amizades se concretizavam por meio de uma tela de notebook estavam presentes quase que todos os dias no laboratório: Eduardo, Cicero e Élton pelos momentos de descontração, dúvidas, medos, em especial do COVID-19 e apoio na concretização deste trabalho.

Aos amigos que a UFPB me apresentou: Kelvin, Antônio, Ingrid e Daniel Gabriel enfim

a todos que de alguma forma contribuíram direto ou indiretamente na realização deste trabalho.

Por fim, agradeço a concessão da bolsa pelo órgão de fomento a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), à Universidade Federal da Paraíba (UFPB), ao Departamento de Química da UFPB, e ao Programa de Pós-Graduação em Química da UFPB pela estrutura necessária para o desenvolvimento deste trabalho. Assim como, ao Laboratório Nacional de Computação Científica (LNCC) pelo fornecimento de infraestrutura computacional.

**Muito Obrigado!**

*“Atravesso o presente de olhos vendados, mal podendo  
pressentir aquilo que estou vivendo... Só mais tarde,  
quando a venda é retirada, percebo o que foi vivido e  
compreendo o sentido do que se passou...”*

**– Milan Kundera**

## Abreviações

**SARS-CoV-2** Síndrome Respiratória Aguda Grave CoronaVírus 2

**COVID-19** CoronaVirus Disease 2019

**SARS** Síndrome Respiratória Aguda Grave

**MERS** Síndrome Respiratória do Oriente Médio

**NC** Proteína do nucleocapsídeo

**PLpro** *Papain-Like protease*

**SCoV1-PLpro** *Papain-Like protease* do SARS-CoV

**SCoV2-PLpro** *Papain-Like protease* do SARS-CoV-2

**Ubl** *Ubiquitin-Like domain*

**ACE2** *Angiotensin-Converting Enzyme 2*

**RBD** *Receptor-Binding Domain*

**RBM** *Receptor-Binding Motif*

**DM** Dinâmica Molecular

**QM** Mecânica Quântica

**ISG15** Gene Estimulado por Interferon 15

**DFT** *Density Functional Theory*

**CDFT** *Conceptual Density Functional Theory*

**RMSF** *Root Mean Square Fluctuation*

**RMSD** *Root Mean Square Deviation*

**PCA** *Principal Component Analysis*

**Mpro** *Main protease*

**FEL** *Free Energy Landscape*



## Resumo

O surgimento do novo coronavírus SARS-CoV-2, causador da pandemia da COVID-19, teve um impacto sem precedentes na vida de todos os seres humanos. Nesse sentido, as proteases responsáveis pela replicação e infecção viral, estudadas nesse trabalho, assim como as linhagens mutantes, têm sido motivo de preocupação para os cientistas, que buscam entender o comportamento conformacional dessas proteínas. Foi demonstrado que, na Papain-Like Protease (PLpro), o domínio do dedo de zinco é fundamental para a estrutura da proteína e influencia o comportamento conformacional do domínio do tipo ubiquitina. Além disso, mutações no domínio de ligação ao receptor (RBD) podem aumentar a resistência a anticorpos neutralizantes e a afinidade com o receptor. Neste trabalho, foram realizadas simulações de dinâmica molecular e cálculos de descritores quânticos para entender as possíveis diferenças conformacionais inter-domínio e reatividade da PLpro. Adicionalmente, também foram utilizadas para avaliar as diferenças conformacionais do RBD e suas mutantes, classificadas de acordo com a afinidade ao receptor e resistência aos anticorpos. Os resultados sugerem que o cátion  $\text{Zn}^{2+}$  tetracoordenado com as cisteínas na estrutura da PLpro estabiliza a região do dedo de zinco, aumentando a mobilidade e a reatividade no domínio ubiquitina, sobretudo no SARS-CoV-2, sugerindo que essa região exerce influência nas conformações de ligação aos substratos pós-traducionais de ubiquitina e interferon que fazem parte da resposta imune antiviral. A clusterização das trajetórias de dinâmica molecular permitiu extrair as estruturas mais populosas para o espaço conformacional de cada sistema. Os resultados para o RBD responsável por infectar o receptor demonstraram que ocorre um aumento na flutuação dos resíduos 360-374 e no motivo de ligação ao receptor (RBM). Observou-se que na região do RBM ocorreu um desvio padrão maior para as mutações que apresentam menor interação com o receptor e maior resistência aos anticorpos. Esse fenômeno pode estar relacionado a uma maior abrangência conformacional e uma menor reatividade dessas mutações, possivelmente como estratégia para escapar da resposta imune do hospedeiro. Os resultados obtidos nesse trabalho podem contribuir para elucidar comportamentos e padrões conformacionais de duas proteínas envolvidas em diferentes estágios do ciclo de vida viral: A PLpro responsável pela replicação e a proteína *Spike*, que é responsável pela invasão da célula hospedeira.

**Palavras-Chaves:** SARS-CoV-1; SARS-CoV-2; Papain-Like protease; Proteína Spike; Conformações; Dinâmica Molecular; Descritores quânticos.

## Abstract

The emergence of the novel SARS-CoV-2 coronavirus, which has caused the ongoing global pandemic of coronavirus disease 2019 (Covid-19), has had an unparalleled impact on the lives of all humans. In this regard, proteases that are responsible for viral replication and infection, as well as mutant strains, have been a subject of concern for scientists, who are engaged in efforts to comprehend the conformational behavior of these proteases. It has been demonstrated that the zinc finger domain plays a pivotal role in the structural integrity of the Papain-Like Protease (PLpro) protein, exerting an influence on the conformational behavior of the ubiquitin-like domain. Moreover, mutations in the receptor-binding domain (RBD) have been observed to enhance the resistance of the virus to neutralizing antibodies and the affinity for the receptor. In this study, molecular dynamics simulations and quantum descriptor calculations were employed to elucidate potential inter-domain conformational variations and the reactivity of PLpro. Moreover, these methods were employed to assess the conformational variations of the RBD and its mutants, which were categorized based on their affinity for the receptor and resistance to antibodies. The results indicate that the  $\text{Zn}^{2+}$  cation, which is tetraordinated with the cysteines in the PLpro structure, serves to stabilize the zinc finger region, thereby increasing the mobility and reactivity of the ubiquitin domain, particularly in the context of SARS-CoV-2. This indicates that this region exerts an influence on the binding conformations to the post-translational substrates of ubiquitin and ISG15, which are components of the antiviral immune response. The clustering of molecular dynamics trajectories enabled the extraction of the most populated structures within the conformational space of each system. The results for the RBD responsible for receptor binding demonstrated an increase in the fluctuation of residues 360-374 and in the receptor binding motif (RBM). It was observed that in the RBM region, there was a greater standard deviation for the mutations that exhibited reduced interaction with the receptor and increased resistance to antibodies. This phenomenon may be related to a greater conformational range and lower reactivity of these mutations, which may represent a strategy to evade the host's immune response. The results obtained in this study may contribute to a greater understanding of the behavior and conformational patterns of two proteins involved in different stages of the viral life cycle: The PLpro responsible for infection and the spike protein, which is responsible for invading the host cell.

**Key-words:** SARS-CoV-1; SARS-CoV-2; Papain-Like protease; Spike protein; Conformations; Molecular Dynamics; Quantum descriptors.

## Sumário

<b>Agradecimentos</b>	<b>1</b>
<b>Abreviações</b>	<b>1</b>
<b>1 Introdução</b>	<b>13</b>
<b>2 Objetivos</b>	<b>16</b>
2.1 Objetivo Geral . . . . .	16
2.2 Objetivos Específicos . . . . .	16
<b>3 Referencial Teórico</b>	<b>18</b>
3.1 Papain-Like protease (PLpro) . . . . .	19
3.2 Domínio de Ligação ao Receptor da Proteína <i>Spike</i> (S) . . . . .	23
3.3 Métodos de Simulação Aplicados no Estudo do SARS-CoV-2. . . . .	27
3.3.1 Métodos de Simulação Aplicados à Papain-Like Protease (PLpro) . .	28
3.3.1.1 Métodos de Simulação Aplicado ao Domínio de Ligação ao Receptor da Proteína <i>Spike</i> . . . . .	31
3.3.1.2 Métodos QM Aplicados no Estudo do SARS-COV-2 . . . . .	36
3.4 Modelagem Molecular . . . . .	38
3.4.1 Dinâmica Molecular . . . . .	38
3.4.1.1 Campos de Força . . . . .	41
3.4.1.2 Campo de Força: AMBER ff14SB . . . . .	43
3.4.1.3 Campo de Força: ZAFF . . . . .	44
3.4.2 Descritores de Reatividade Químico Quânticos . . . . .	44

<b>4</b>	<b>Métodos</b>	<b>47</b>
4.1	Amostragem das Estruturas Cristalográficas . . . . .	47
4.1.1	Papain-Like Protease (PLpro) . . . . .	47
4.1.2	Domínio de Ligação ao Receptor (RBD) . . . . .	48
4.2	Estado de Protonação e Parametrização . . . . .	49
4.3	Protocolo de Dinâmica Molecular . . . . .	50
4.4	Análises Pós-Produção das Trajetórias de DM . . . . .	51
4.4.1	Papain-Like Protease (PLpro) . . . . .	51
4.4.1.1	Reconstrução dos perfis de energia livre e PCA no estudo da Papain-Like protease (PLpro) . . . . .	51
4.4.2	Domínio de Ligação ao Receptor (RBD) . . . . .	52
4.5	Cálculo de descritores de reatividade . . . . .	52
4.5.1	Papain-Like Protease (PLpro) . . . . .	52
4.5.2	Domínio de Ligação ao Receptor (RBD) . . . . .	53
<b>5</b>	<b>Resultados e Discussão</b>	<b>54</b>
5.1	Estudo Conformacional das proteínas do SARS-CoV-2 . . . . .	54
5.1.1	<i>Papain-Like Protease</i> (PLpro) . . . . .	54
5.1.1.1	Análises do Comportamento Conformacional da <i>Papain-Like Protease</i> por PCA . . . . .	62
5.1.1.2	Descritores de Reatividade na <i>Papain-Like Protease</i> . . . . .	68
5.1.2	Domínio de Ligação ao Receptor (RBD) . . . . .	77
5.1.2.1	Análises do Comportamento Conformacional do Domínio de Ligação ao Receptor . . . . .	80
5.1.2.2	Descritores de Reatividade no Domínio de Ligação ao Receptor	84
<b>6</b>	<b>Conclusões</b>	<b>91</b>
<b>7</b>	<b>Apêndices</b>	<b>115</b>
7.1	Diferentes versões do script escrito em Python para plotar os gráficos de RMSF	115
7.2	Perfis de energia livre PCA . . . . .	120
7.3	Descritores de Reatividade . . . . .	123

## Lista de Figuras

- 1.1 A organização genômica do SARS-CoV-2 é composta por várias regiões funcionais estruturais e não estruturais. O genoma codifica dois grandes genes, ORF1a em verde e ORF1b em azul, processados em 16 proteínas não estruturais (nsp1–nsp16). Os genes estruturais codificam as proteínas estruturais S, E, M e N. Fonte: Adaptado de<sup>[1]</sup> . . . . . 14
- 3.1 A) Representação esquemática de ORF1a do SARS-CoV-2, indicando as proteínas não estruturais processadas (nsps). Os locais de clivagem de PLpro são indicados por tesoura. B) Sequência de alinhamento de aminoácidos P4-P1' dos sítios de clivagem PLpro indicados. Fonte: Adaptada de<sup>[2]</sup> . . . . . 19
- 3.2 Estrutura cristalina do SCoV2-PLpro(PDB ID: 6WRH). Os domínios foram divididos por cores, na cor azul tem-se o domínio do tipo ubiquitina, na cor roxa o *Thumb*, na cor laranja o *Palm* e por último o domínio *Zinc-Finger* (dedo de zinco) na cor verde. Fonte: Imagem do autor. . . . . 21
- 3.3 Alinhamento da estrutura cristalina da PLpro SARS-CoV-1 (PDB ID: 2FE8) em azul com a PLpro SARS-CoV-2 (PDB ID: 6WRH) em verde. A alça flexível BL2 em vermelho e o Sítio catalítico formado por CYS-111, HIS-272 e ASP-286 em destaque. Fonte: Imagem do autor. . . . . 22

3.4	(A) Representação em bloco da proteína <i>Spike</i> SCoV-2 com seus diferentes domínios, a saber, NTD, domínio N-terminal; RBD, domínio de ligação ao receptor; FP, peptídeo de fusão; HR1, repetição heptada 1, HR2, repetição heptada 2; TM, domínio transmembranar; CT, domínio citoplasmático em. (B) Estrutura trímica da Proteína <i>Spike</i> (PDB ID: 7KRR) com um RBD-up e 2 RBD-down. (C) Conformação RBD-up do sítio de ligação entre o RBM (em cinza), localizado no RBD (em azul-marinho) da proteína <i>Spike</i> , com a enzima conservadora de angiotensina humana 2 (ACE2) em roxo. Fonte: Imagem do autor. . . . .	24
4.1	Análise estrutural comparando a proteína SCoV1-PLpro e a SCoV2-PLpro usadas nas simulações de Dinâmica molecular. Os respectivos valores de desvio médio quadrático (RMSD) baixo mostram nenhuma distinção substancial considerando tanto o <i>backbone</i> como o sítio ativo das duas proteínas. Fonte: Imagem própria. . . . .	48
4.2	Conformações do Domínio de Ligação ao Receptor: +inf/+ev em branco, +inf/-ev em azul, -inf/+ev em amarelo e -inf/-ev em cinza. As cores verde, amarelo, azul e vermelho correspondem às mutações. <b>Fonte:</b> Imagem própria. . . . .	53
5.1	RMSD para a SCoV1-PLpro e SCoV2-PLpro com e sem zinco <i>vs.</i> tempo. Fonte: Própria. . . . .	55
5.2	RMSF da SCoV1-PLpro e SCoV2-PLpro com e sem zinco. Resíduos altamente flexíveis 1-62 (Ubl), 182-200 e 223-235 (Dedo de Zn) e 265-272 (Sítio de ligação). Fonte: Própria. . . . .	56
5.3	PLpro e em destaque com cores diferentes as regiões que apresentaram uma maior flutuação, Ubl (Verde), dedo de zinco (vermelho) e sítio de ligação (amarelo). Fonte: Própria. . . . .	57
5.4	(a)RMSF para a SCoV1-PLpro e SCoV2-PLpro sem zinco. (b). RMSF para a SCoV1-PLpro e SCoV2-PLpro com zinco . . . . .	58
5.5	Gráfico RMSD bidimensional para as doze trajetórias combinadas (100 ns cada uma) para cada um dos sistemas da PLpro (a) Backbone da proteína, (b) BL2 loop, (c) Motivo Zn- $\beta 4\beta 5$ , (d) Zn- $\beta 6\beta 7$ , (e) domínio Ubl e (f) Número de Cluster <i>vs.</i> frame plot do Domínio Ubl, Motivo Zn- $\beta 6\beta 7$ e Ubl/Zn- $\beta 6\beta 7$ . . . . .	60

5.6	RMSD-2D para as duas regiões que apresentaram maiores mudanças conformacionais na PLpro.(a) região do domínio do dedo de zinco formado pelo motivo Zn- $\beta 4\beta 5$ e Zn- $\beta 6\beta 7$ . (b) Região formada pelo Ubl com o motivo Zn- $\beta 6\beta 7$ . . . . .	62
5.7	Número de Cluster vs. frame plot. (a) Domínio Ubl, (b) Motivo Zn- $\beta 6\beta 7$ e (c) Ubl/Zn- $\beta 6\beta 7$ . . . . .	63
5.8	Perfis de energia livre para todos os sistemas. (a) Ubl; (b) Motivo Zn-( $\beta 6\beta 7$ ); (c) Ubl mais o motivo Zn-( $\beta 6\beta 7$ ) juntos. Os pequenos símbolos de triângulos, ou quadrados e círculos em B, ao longo do gráfico representam os aglomerados conformacionais mais populosos considerando cada região em específico e todo o conjunto de trajetórias dos sistemas. . . . .	64
5.9	Perfis de energia livre para Zn-( $\beta 6\beta 7$ ) em cada sistema da <i>Papain-Like protease</i> . Os círculos dos gráficos representam as projeções médias. (A) SCoV1-PLpro sem zinco, (B) SCoV1-PLpro com zinco, (C) SCoV2-PLpro sem zinco e (D) SCoV2-PLpro com zinco ao longo das PCs. . . . .	65
5.10	Os 5 clusters mais populosos de cada trajetória com suas respectivas representatividade em porcentagem para cada sistema com ou sem zinco. Esses clusters serão utilizados em análises posteriores. . . . .	66
5.11	Perfis de energia livre para cada sistema da PLpro considerando apenas os resíduos que forma o Ubl. Os círculos dos gráficos representam as projeções médias de cada sistema ao longo das PCs. . . . .	67
5.12	Descritores de reatividade calculados entre as estruturas amostradas para SCoV1-PLpro e SCoV2-PLpro com e sem zinco (A) Eletrofilicidade, (B) Nucleofilicidade, (C) Radicalidade, (D) Multifilicidade, (E) Potencial Eletrostático Molecular, (F) Dureza Local (aproximação potencial químico local), (G) Dureza local (potencial Fukui) e (H) Potencial Fukui radical. . . . .	69
5.13	Resultados da análise de componentes principais (PCA) para os descritores de reatividade. (a) Objetos pelo tipo de vírus SARS-CoV-1 e SARS-CoV-2 (b) Objetos pelo tipo de Proteína com e sem zinco. Os círculos e triângulos representa os valores dos resíduos por descritor. . . . .	71
5.14	Mapas de calor mostrando as médias dos descritores para cada resíduo analisado para a Proteína com e sem zinco. (a) SARS-CoV-1 (b) SARS-CoV-2. . .	72

5.15	Mapas de calor mostrando as médias dos descritores para cada resíduo analisado para o domínio ubiquitina. (A) Nucleofilicidade (B) Eletrofilicidade (C) Potencial Molecular Eletrostático e (D) Potencial Fukui Radical. . . . .	74
5.16	Dureza Local (potencial elétron-elétron) (a) SCoV1-PLpro os resíduos com maiores variação são destacados pelas cores verde, cinza, amarelo, vermelho e azul, (b) Dureza Local (potencial elétron-elétron) para o domínio ubiquitina, (c) SCoV2-PLpro os resíduos com maiores variação são destacados pelas cores verde, cinza, amarelo, vermelho e azul, (d) Resíduos com maiores variação para a dureza local no domínio ubiquitina. . . . .	75
5.17	Dureza Local (potencial químico) (a) SCoV1-PLpro os resíduos com maiores variação são destacados pelas cores verde, cinza, amarelo, vermelho e azul, (b) Dureza Local (potencial químico) para o domínio ubiquitina, (c) SCoV2-PLpro os resíduos com maiores variação são destacados pelas cores verde, cinza, amarelo, vermelho e azul, (d) Resíduos com maiores variação para a dureza local no domínio ubiquitina. . . . .	76
5.18	Nomenclatura adotada para RBD e RBM. Fonte: adaptada de Wang et al. (2020) <sup>[3]</sup> . . . . .	78
5.19	(a) RMSF de toda a região RBD para das spikes classificadas como (+inf/+ev, "++"),(+inf/-ev, "+-"), (-inf/+ev", "-+") e (-inf/-ev,"--"). (B) RMSF obtido alinhando a região do motivo de ligação ao receptor (RBM) e considerando apenas está mesma região (resíduos 438-508). Os RMSFs são apresentados como média (linhas) e desvios (sombra) considerando as três réplicas individualmente e a soma das três de cada mutação e variante. . . . .	79
5.20	Os 5 clusters (c0-azul, c1-verde, c2-vermelho, c3-cinza e c4-oliva) mais populosos de cada trajetória com suas respectivas representatividade em porcentagem para cada sistema +inf/+ev, +inf/-ev, -inf/+ev e -inf/-ev. Esses clusters serão utilizados em análises posteriores. . . . .	81
5.21	Análise de componentes principais para cada fenotípico na região 360-374 para os PCs 1, 2, e 3. Os círculos sobrepostos à superfície de energia livre correspondem à projeção dos frames das trajetórias de DM para cada grupo fenotípico (+inf/+ev (A), +inf/-ev (B), -inf/+ev (C) e -inf/-ev (D). . . . .	82



5.22	Os 5 clusters (c0-azul, c1-verde, c2-vermelho, c3-cinza e c4-oliva) mais populosos de cada trajetória com suas respectivas representatividade em porcentagem para cada sistema +inf/+ev, +inf/-ev, -inf/+ev e -inf/-ev. Esses <i>clusters</i> serão utilizados em análises posteriores. . . . .	83
5.23	Análise de componentes principais para cada fenotípico na região 360-374 para os PCs 1, 2, e 3. Os círculos sobrepostos à superfície de energia livre correspondem à projeção dos frames das trajetórias de DM para cada grupo fenotípico (+inf/+ev (A), +inf/-ev (B), -inf/+ev (C) e -inf/-ev (D). . . . .	84
5.24	Media e desvio padrão da eletrofilicidade (a) região 360-374, (b) região do RBM. . . . .	85
5.25	Eletrofilicidade por resíduos de aminoácidos (a) 360-374, (b) RBM, (c) Eletrofilicidade. . . . .	86
5.26	Media e desvio padrão para a Dureza local (potencial Fukui) (a) região 360-374, (b) região do RBM. . . . .	87
5.27	Dureza local (potencial Fukui) por resíduos de aminoácidos (a) 360-374, (b) RBM. . . . .	88
5.28	Media e desvio padrão para a moleza local (a) região 360-374, (b) região do RBM. . . . .	89
5.29	Resíduos do RBM com maior moleza local (a) região 360-374, (b) região do RBM. . . . .	90
7.1	Perfis de energia livre para a PLpro SCoV-1 sem zinco. Os círculos ao longo dos gráficos representam as projeções da trajetória ao longo de (a) PC1 <i>vs.</i> PC3 e (b) PC2 <i>vs.</i> PC3. . . . .	121
7.2	Perfis de energia livre para a PLpro SCoV-1 com zinco. Os círculos ao longo dos gráficos representam as projeções da trajetória ao longo de (a) PC1 <i>vs.</i> PC3 e (b) PC2 <i>vs.</i> PC3. . . . .	121
7.3	Perfis de energia livre para a PLpro SCoV-2 sem zinco. Os círculos ao longo dos gráficos representam as projeções da trajetória ao longo de (a) PC1 <i>vs.</i> PC3 e (b) PC2 <i>vs.</i> PC3. . . . .	122
7.4	Perfis de energia livre para a PLpro SCoV-2 sem zinco. Os círculos ao longo dos gráficos representam as projeções da trajetória ao longo de (b) PC1 <i>vs.</i> PC3 e (a) PC2 <i>vs.</i> PC3. . . . .	122

7.5	Descritores de reatividade (a) e (b) Carga parcial, (c) e (d) Dureza local (potencial elétron-elétron), (e) e (f) Potencial eletrostático molecular e (g) e (h)Radicalidade. . . . .	124
7.6	Descritores para moleza local (a) Moleza Dual Local (360-374), (b) Moleza Dual Local (RBM), (c) Hiper Moleza Local(360-374) e (d) Hiper Moleza Local(RBM). . . . .	125

O recente surgimento de um novo coronavírus causou um impacto sem precedentes na vida de todos os seres humanos e na economia global, estimulando a mobilização das autoridades públicas de saúde em todo o mundo para conter a rápida disseminação do vírus, classificado como Síndrome Respiratória Aguda Grave CoronaVírus 2 (SARS-CoV-2).<sup>[4]</sup> A infecção respiratória aguda grave (SARI) tem causado a morte de milhões de pessoas em todo o mundo. Já em 2017, com a morte de 4,5 milhões de pessoas, foi classificada como a quarta principal causa de morte para todas as idades em todo o mundo.<sup>[5]</sup> Nas duas primeiras décadas do século XXI, o surgimento de novas infecções respiratórias como o vírus da síndrome respiratória aguda grave (SARS), a gripe aviária, a síndrome respiratória do Oriente Médio (MERS) e a gripe pandêmica H1N1 representaram ameaças significativas para os seres humanos, particularmente em países do continente asiático.<sup>[6]</sup> Contudo, nenhuma delas atingiu a proporção causada pela síndrome respiratória aguda grave do novo coronavírus.

O surto foi relatado pela primeira vez em Wuhan, província de Hubei, na China,<sup>[7]</sup> ocasionado pelo vírus associado à infecção designada como (COVID-19), do inglês "coronavirus disease 2019". O período médio de incubação do coronavírus é de aproximadamente 05 dias, com intervalos que podem chegar a 14 dias, período em que os primeiros sintomas levam para aparecer desde a infecção em 95% dos infectados.<sup>[8]</sup> Declarada como estado de pandemia em 11 de março de 2020, espalhou-se rapidamente por 221 países, sendo responsável até agora por mais de 500 milhões de casos e mais de 6 milhões de fatalidades no mundo. Só no Brasil, são mais de 30 milhões de casos confirmados e mais de 600 mil mortes, ficando atrás apenas dos Estados Unidos da América com mais de 900 mil mortes, conforme a Organização Mundial da Saúde (OMS).<sup>[9]</sup>

O SARS-CoV-2 é um vírus de RNA, com fita simples de sentido positivo e envelope,



As proteínas categorizadas como não estruturais referem-se aos genes ORF1a e ORF1b, os quais codificam proteínas não estruturais essenciais para a replicação do RNA, compreendendo dois terços do genoma e dando origem às proteínas pp1a e pp1b. Estas poliproteínas são subsequentemente processadas por duas enzimas proteases cisteínicas virais ("*papain-like*" e *poliovirus 3C-like*). Tais proteases realizam clivagens nos polipeptídeos pp1a e pp1b, resultando em 16 proteínas não estruturais (designadas como nsp1-nsp16), que constituem os elementos do complexo de replicação.<sup>[1]</sup>

As principais proteases não estruturais mais estudadas pela comunidade científica são: protease semelhante à papaína (nsp3), semelhante a quimotripsina 3CL ou protease principal (nsp5), proteína de ligação de RNA (nsp9), RNA polimerase RNA viral-dependente (nsp12) e helicase (nsp13).<sup>[14]</sup> Essas proteínas são consideradas enzimas virais essenciais por desempenharem papéis importantes na replicação e no empacotamento do vírus nas células hospedeiras. Todavia, as funções de algumas proteínas não estruturais (nsps) ainda são desconhecidas ou precisam ser melhor compreendidas no desenvolvimento de fármacos.<sup>[15]</sup>

Assim, considerando a importância de aprofundar nossa compreensão acerca do papel desempenhado pelas proteínas *Spike* e protease tipo-papaína (do inglês, *Papain-Like protease* (PLpro)), este estudo propõe uma abordagem teórica que emprega técnicas de simulação molecular e química quântica. O objetivo é analisar de forma mais aprofundada características essenciais dessas duas enzimas cruciais para os processos de infecção e replicação do SARS-CoV-2. A proteína *Spike* está relacionada com a etapa de infecção, enquanto a PLpro desempenha um papel crucial na replicação do vírus.

Bem como propomos, um estudo utilizando as diferentes variantes e mutações da proteína *Spike*, sobretudo, as mutações que ocorrem no domínio de ligação ao receptor para compreender o quanto as mutações do tipo único aminoácido alteram estas características para otimizar o planejamento de fármacos, compreensão da infectividade e imunogenicidade, tal como o planejamento de novas vacinas eficientes para as múltiplas variantes do vírus são fundamentais para a comunidade científica. Através deste estudo buscamos ampliar os conhecimentos sobre a infecção e a replicação do SARS-CoV-2.

## 2.1 Objetivo Geral

Aplicar metodologias de simulações computacionais clássicas e quânticas para investigar o comportamento dinâmico de duas proteínas do SARS-CoV-2: a proteína *Spike*, mais especificamente seu domínio de ligação ao receptor, responsável pela infecção, e a protease PLpro, envolvida na replicação, com o objetivo de acessar uma ampla gama de informações sobre essas proteínas, calcular suas características eletrônicas e avaliar o efeito de mutações no domínio de ligação ao receptor.

## 2.2 Objetivos Específicos

- Produzir conformações representativas do RBD da proteína *Spike* e da PLpro através de simulações de DM;
- Identificar diferenças estruturais entre as diferentes variantes do SARS-CoV-2 relacionadas com as diferentes mutações no RBD através de simulações de DM;
- Compreender como as mutações podem diferenciar drasticamente nas características físico-químicas de uma proteína viral, alterando assim sua funcionalidade, acessibilidade ao sistema imune e drogabilidade (responsividade a fármacos);
- Compreender as características chaves da PLpro fundamentais para a replicação viral;
- Avaliar o quanto o dedo de zinco da PLpro influencia nas conformações das proteínas do SARS-CoV-1 e SARS-CoV-2;

- Utilizar técnicas de análise de componentes principais e clusterização estrutural para selecionar as conformações mais representativas das duas proteínas;
- Utilizar o PRIMoRDIA para fazer os cálculos dos descritores quânticos;
- Avaliar as assinaturas relacionadas a densidade eletrônica, não apenas uma conformação mais representativa, mas todo o equilíbrio conformacional representativo da proteína para calcular nucleofilicidade, eletrofilicidade, moleza e dureza locais, localização do orbital molecular, multifilicidade e densidade eletrônica;

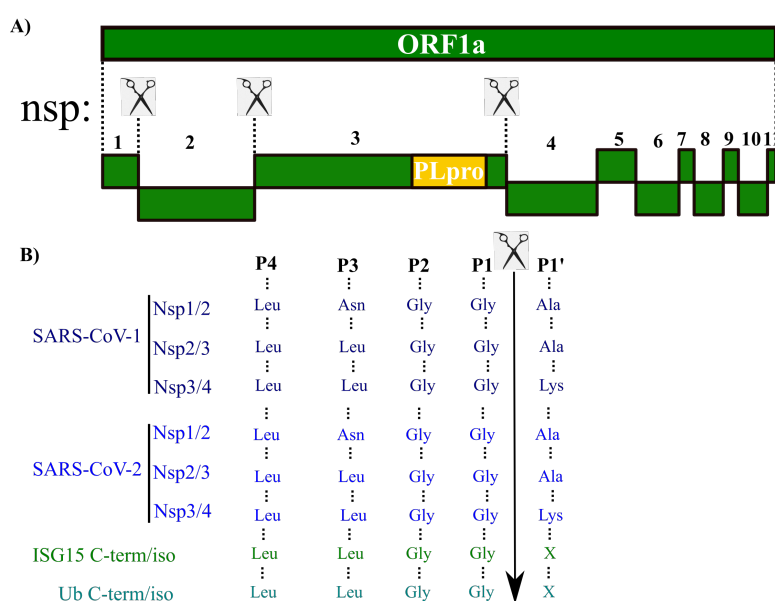
## Referencial Teórico

Neste capítulo serão abordadas as concepções da comunidade sobre a PLpro e a Proteína *Spike*. Serão abordados os métodos de simulação utilizados para compreender as conformações estruturais dessas duas proteínas, bem como, na descoberta de fármacos. Nessa seção também serão abordados brevemente o formalismo matemático, avanços e limitações dos métodos de dinâmica molecular no estudo de macromoléculas e os descritores de reatividade.



### 3.1 Papain-Like protease (PLpro)

O coronavírus SARS-CoV-2 codifica um domínio essencial da PLpro como parte de sua proteína não estrutural nsp3. Essa protease desempenha papéis essenciais na replicação do vírus e na evasão imune, apresentando-se como um excelente alvo farmacológico. Reconhece o local de clivagem de consenso LXGG (L = Leucina, X = qualquer aminoácido e G = glicina) nas posições P4-P1 e cliva o N-terminal da poliproteína replicase em três locais diferentes para liberar as proteínas não estruturais nsp1, nsp2 e nsp3, essenciais para a replicação do vírus.<sup>[16,17]</sup>



**Figura 3.1:** A) Representação esquemática de ORF1a do SARS-CoV-2, indicando as proteínas não estruturais processadas (nsps). Os locais de clivagem de PLpro são indicados por tesoura. B) Sequência de alinhamento de aminoácidos P4-P1' dos sítios de clivagem PLpro indicados. Fonte: Adaptada de<sup>[2]</sup>

Estudos anteriores de caracterização *in vitro* indicaram que a protease PLpro do SARS-CoV-2 (SCoV2-PLpro), além de ter um papel fundamental na replicação viral, ainda desempenham outras duas atividades proteolíticas tais como: remove a proteína ISG15 (proteína do gene 15 estimulada por interferon) e a ubiquitina (Ub) das proteínas celulares dos hospedeiros. Essas duas atividades são importantes para facilitar a fuga do vírus do sistema imunológico.<sup>[2,18]</sup>

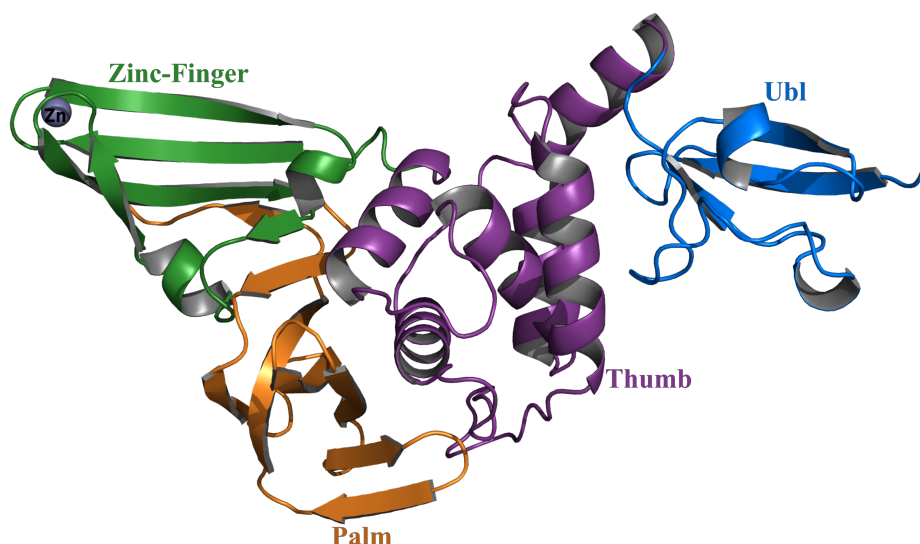
A ubiquitina é uma pequena proteína conservada de 76 resíduos de aminoácidos que atua como marcador na modificação pós-traducional de proteínas,<sup>[19]</sup> forma um sistema com o proteassoma, chamado ubiquitina-proteassoma (UPS). A via ubiquitina-proteassoma

atua como o principal processo de degradação de proteínas em células eucarióticas, ou seja, liga a ubiquitina à proteína alvo; essa adição serve como um marcador de destruição altamente específico. Esse processo é denominado ubiquitinação, o qual é catalisado por três enzimas principais conhecidas como *E1* enzima ativadora de ubiquitina, a *E2* enzima conjugante de ubiquitina e a *E3* enzima ligante de ubiquitina.<sup>[20,21]</sup>

Já o ISG15 é uma proteína semelhante à ubiquitina, produto de um gene estimulado por interferons (IFN) do tipo I, que pode ser ligado covalentemente por meio de uma ligação isopeptídica entre o C-terminal e o grupo  $\epsilon$ -amino da cadeia lateral dos resíduos de lisina da proteína alvo e regular a sua função. Um processo de modificação pós-traducionais denominado ISGilação.<sup>[22]</sup> Tanto o processo de ubiquitinação como o processo de ISGilação são responsáveis pela degradação proteica da proteína alvo. Dessa forma, muitos vírus desenvolvem estratégias para interromper essas duas vias de degradação.<sup>[23]</sup> Devido às atividades desubiquitinantes (remoção de ubiquitina) e deISGilantes (remoção de ISG15) a PLpro desempenha um papel fundamental na resposta imune inata durante a infecção causada pelo vírus.<sup>[2]</sup>

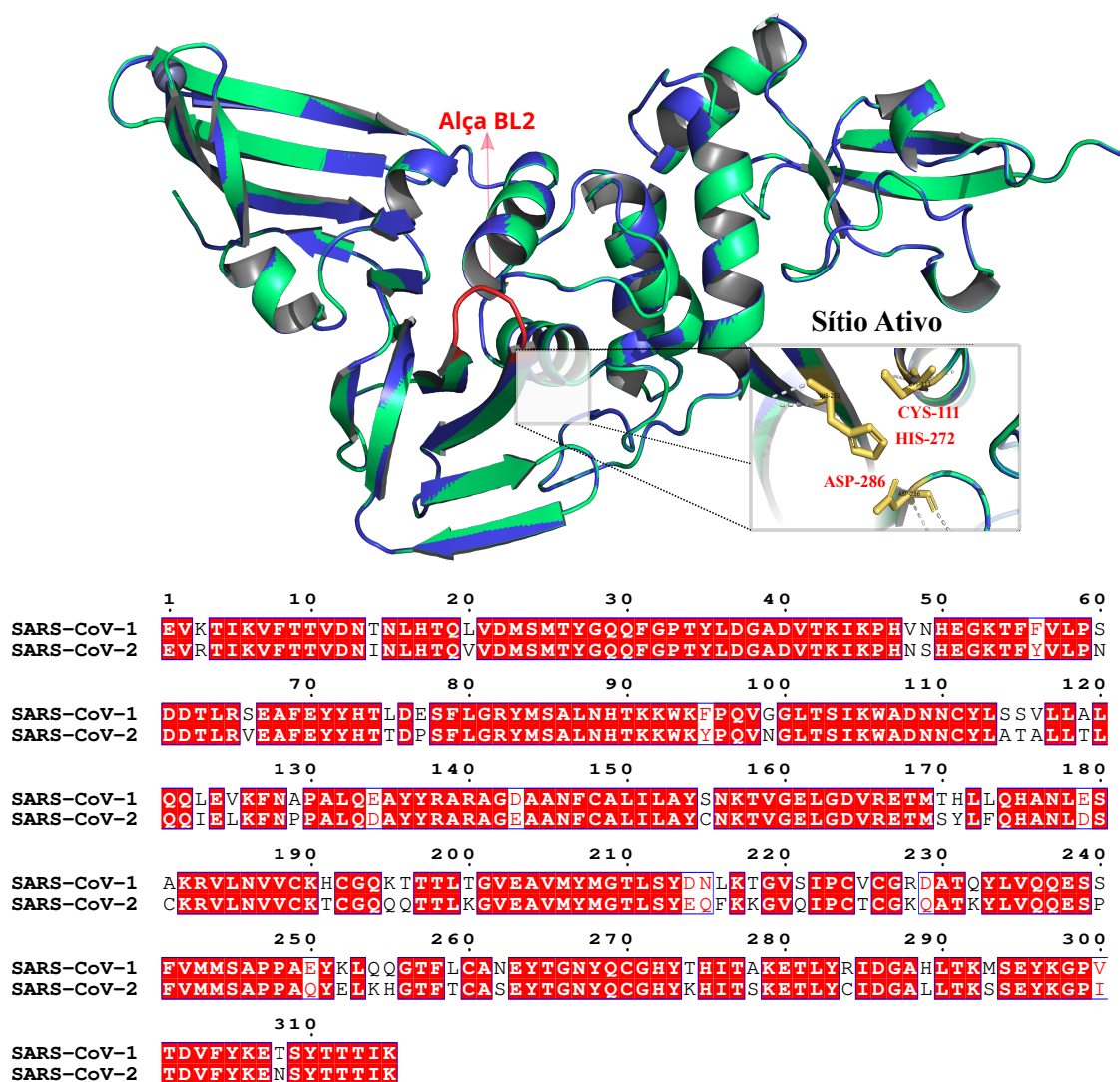
Estudos recentes em in vitro em células humanas mostraram que existe uma preferência do SCoV2-PLpro por ISG15 sobre cadeias de poli ubiquitina ligadas a Lys-48.<sup>[1,13]</sup> Ela cliva o ISG15 do fator responsivo ao interferon 3 (IRF-3), bloqueando sua translocação nuclear e reduzindo as respostas do interferon do tipo I, enquanto, a protease do SARS-CoV-1 (SCoV1-PLpro) cliva preferencialmente a ubiquitina, processo de desubiquitinação, de componentes críticos nas vias de sinalização que ativam o fator de transcrição NF- $\kappa$ B, embora, o ISG15 seja capaz de ligar-se a essa protease. O estudo desses fatores são importantes, visto que, são responsáveis pela ativação das respostas inflamatórias.<sup>[24-26]</sup>

A estrutura cristalográfica do SCoV1-PLpro e do SCoV2-PLpro é um monômero que apresenta uma identidade de 82,9% da sequência de aminoácidos. Onde, a maioria das características estruturais dos ortólogos entre essas duas proteases são conservadas.<sup>[27]</sup> Consiste em quatro subdomínios distintos, três dos quais formam uma arquitetura de “palma da mão ou mão aberta” divididos nos domínios *Palm*, *Thumb* e *Zinc-Finger*. O quarto domínio corresponde aos primeiros 60 resíduos de aminoácidos, formando um subdomínio de N-terminal independente denominado domínio tipo ubiquitina (Ubl, do inglês *Ubiquitin-like domain*) que é bem separado dos outros três domínios acima descrito e adota uma dobra de folhas betas semelhante ao domínio Ubl de várias proteínas, incluindo o ISG15 e a elongina, por exemplo.<sup>[28,29]</sup> Figura 3.2 mostra esses quatro domínios.



**Figura 3.2:** Estrutura cristalina do SCoV2-PLpro(PDB ID: 6WRH). Os domínios foram divididos por cores, na cor azul tem-se o domínio do tipo ubiquitina, na cor roxa o *Thumb*, na cor laranja o *Palm* e por último o domínio *Zinc-Finger* (dedo de zinco) na cor verde. Fonte: Imagem do autor.

A SCoV2-PLpro apresenta 54 mutações em relação ao SCoV1-PLpro, e seu sítio ativo, que fica entre os subdomínios palm e thumb, é 100% conservado. Ele é composto por uma tríade catalítica conservada, formada pelos resíduos de aminoácidos Cys-111, His-272 e Asp-286. O resíduo Cys-111 está localizado a uma distância de 3,6 Å da histidina catalítica His-272 no SARS-CoV-2, enquanto, no SARS-CoV-1 a distância correspondente é de 3,7 Å. O resíduo His-272 forma uma ligação de hidrogênio com o Asp-286 com um comprimento de 3,0 Å, enquanto no SARS-CoV-1 essa mesma ligação tem um comprimento de 2,7 Å. Por fim, a ligação de hidrogênio entre Asp-108 e Trp-93 apresenta um comprimento de 2,8 Å e fortalece a conformação do ânion de oxigênio.<sup>[27]</sup>



**Figura 3.3:** Alinhamento da estrutura cristalina da PLpro SARS-CoV-1 (PDB ID: 2FE8) em azul com a PLpro SARS-CoV-2 (PDB ID: 6WRH) em verde. A alça flexível BL2 em vermelho e o Sítio catalítico formado por CYS-111, HIS-272 e ASP-286 em destaque. Fonte: Imagem do autor.

Outro domínio importante é o *Zinc-Finger*, onde o íon  $Zn^{2+}$  é coordenado por quatro cisteínas (Cys192/189 e Cys224/226 na região de loop entre 2  $\beta$ -folhas) com geometria tetraédrica essencial para a integridade estrutural e atividade da protease.<sup>[29]</sup> O loop (alça) BL2, localizado no domínio Palm, desempenha um papel fundamental no sítio catalítico, apresentando uma conformação aberta na estrutura da PLpro não ligada. Não obstante, na presença de um inibidor, ele se fecha após a formação da ligação do inibidor com a proteína em direção à fenda catalítica, permitindo a formação de interações intermoleculares com os inibidores e facilitando o acesso do substrato ao sítio ativo.<sup>[17]</sup> A alça pode assumir múltiplas

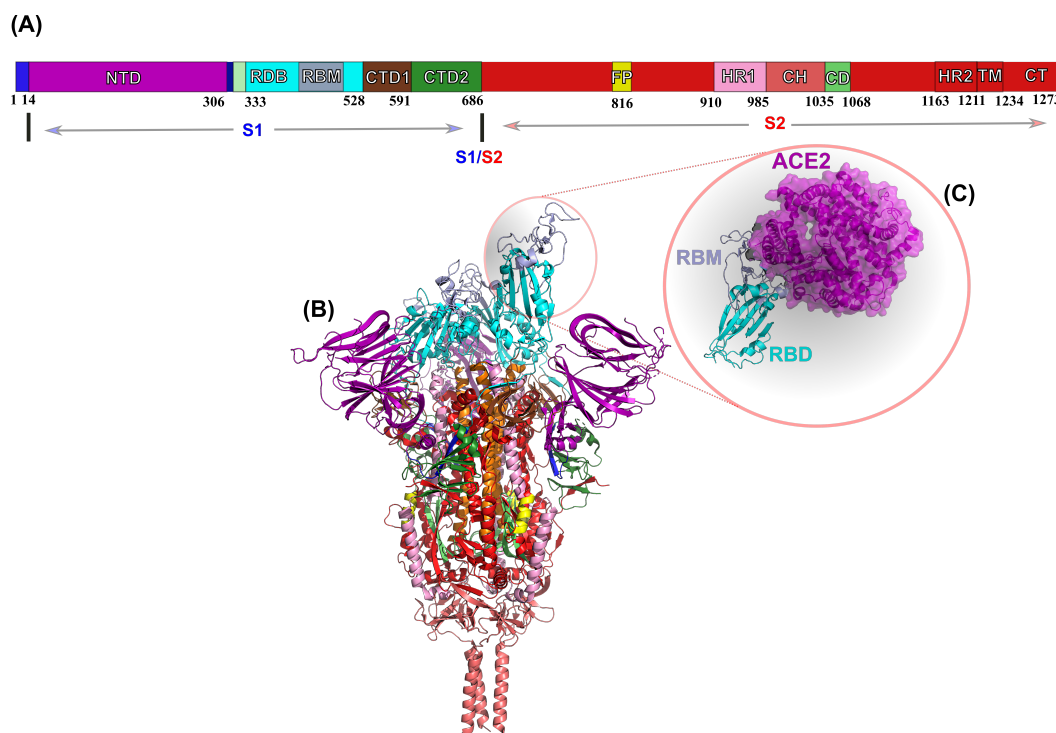
conformações dependendo do tamanho e tipo da molécula inibidora, sendo crucial para os esforços futuros no desenho racional de medicamento para SARS-CoV-2.<sup>[18,29]</sup>

Estudos realizados com a protease SCoV1-PLpro mostraram que enquanto a alça faz um movimento de fechamento acionado pelo mecanismo de ajuste induzido pela ligação com o inibidor, outros movimentos independentes são observados na ponta do domínio do dedo de zinco, essa movimentação pode explicar a capacidade da PLpro de interagir com outras cadeias poli-UB topologicamente diferentes, bem como com o modificador Ub e ISG15.<sup>[17]</sup> Esses estudos são importantes para compreender o comportamento estrutural da SCoV2-PLpro e a partir deles desenvolver mecanismos de inibição dessa protease.

## 3.2 Domínio de Ligação ao Receptor da Proteína *Spike* (S)

Os mecanismos de infecção do SARS-CoV-2 foram amplamente estudados pela comunidade científica, sendo a proteína *Spike* do pico viral um dos principais alvos terapêuticos em virtude do seu papel-chave no ciclo de vida do vírus, densamente glicosilada para entrar nas células hospedeiras.

A *Spike* é uma das principais proteínas do SARS-CoV-2, sendo responsável pelo primeiro contato e invasão do vírus à célula humana através da enzima conversora de angiotensina 2 (ACE2, da sigla em inglês: *Angiotensin-Converting Enzyme 2*). Portanto, desenvolver anticorpos capazes de inibir essa proteína é fundamental para impedir a entrada do vírus nas células humanas, principalmente o domínio de ligação ao receptor na proteína (RBD, do inglês: *Receptor-Binding Domain*), um alvo chave, pois inicia o processo de infecção na célula hospedeira.<sup>[30]</sup>



**Figura 3.4:** (A) Representação em bloco da proteína *Spike* SCoV-2 com seus diferentes domínios, a saber, NTD, domínio N-terminal; RBD, domínio de ligação ao receptor; FP, peptídeo de fusão; HR1, repetição heptada 1, HR2, repetição heptada 2; TM, domínio transmembranar; CT, domínio citoplasmático em. (B) Estrutura trimerica da Proteína *Spike* (PDB ID: 7KRR) com um RBD-up e 2 RBD-down. (C) Conformação RBD-up do sítio de ligação entre o RBM (em cinza), localizado no RBD (em azul-marinho) da proteína *Spike*, com a enzima conservadora de angiotensina humana 2 (ACE2) em roxo. Fonte: Imagem do autor.

A proteína *Spike* é uma glicoproteína transmembranar do tipo I. Ela consiste em uma estrutura trimérica composta por 1273 resíduos de aminoácidos, distribuídos nas subunidades S1 (resíduos de aminoácidos 14-685) e S2 (resíduos de aminoácidos 686-1273), e é precedida por uma cadeia polipeptídica de sinal longa de 1-13 aminoácidos em direção à extremidade N-terminal. A subunidade S1 se dobra em quatro subdomínios, que são o domínio N-terminal (NTD), o domínio de ligação ao receptor (RBD), o domínio C-terminal 1 (CTD1) e o domínio C-terminal 2 (CTD2). Esses subdomínios trabalham juntos para proteger o núcleo S2 pré-fusão trimerizado. A subunidade S2, por sua vez, inclui o peptídeo de fusão (FP), as regiões Heptadas de Repetição 1 (HR1), a hélice central (CH), o domínio conector (CD), as regiões Heptadas de Repetição 2 (HR2), o domínio transmembranar (TM) e o domínio C-terminal (CT).<sup>[31,32]</sup>

A interação das partículas virais com os receptores celulares na superfície das células hospedeiras marca o início do processo de infecção viral. Portanto, o reconhecimento des-

ses receptores desempenha um papel fundamental na entrada viral e, conseqüentemente, torna-se um alvo crucial no desenvolvimento de peptídeos inibidores. O domínio de ligação ao receptor (RBD) é responsável pelo reconhecimento e ligação ao receptor ACE2. Localizado na subunidade S1 da proteína *Spike*, o RBD passa por movimentos semelhantes aos de dobradiças para alternar entre as conformações "down"(para baixo), que representa um estado inacessível ao receptor, e as conformações "up"(para cima), que representam um estado acessível ao receptor ACE2. Cerca de 50% a 70% das conformações abertas (RBDs "up") foram observadas na estrutura da proteína *Spike*. Em contrapartida, aproximadamente 30% a 50% das vezes, o trímero da proteína *Spike* encontra-se em uma conformação fechada (RBDs "down"), como observado em amostras cristalizadas. A conformação "up", também conhecida como forma aberta, é mais frequente, visto que desempenha um papel predominante na mediação da infectividade por meio da interação com o receptor.<sup>[33,34]</sup>

Análise do motivo de ligação ao receptor (RBM) (da sigla em inglês: *Receptor-Binding Motif*), uma parte do RBD compreendida pelos resíduos de aminoácidos 333-528, que faz contato direto com o receptor ACE2, revelou que a maioria dos resíduos essenciais para a ligação ao receptor é conservada nas proteínas *Spike* do SARS-CoV-1 e do SARS-CoV-2. Mutações em resíduos-chave, principalmente no RBM, desempenham um papel importante no aumento ou na diminuição da interação da proteína viral com o receptor ACE2.<sup>[35]</sup>

Na região da subunidade S1, encontramos também o NTD e o CTD. O NTD é um domínio funcional responsável pela ligação ao receptor da célula hospedeira. Embora sua função não seja completamente compreendida, observou-se que anticorpos monoclonais (mAbs) específicos para o NTD contribuem para a neutralização, inibindo a pré-fusão para mudança conformacional da proteína *Spike* após a ligação ao receptor e a subsequente pós-fusão.<sup>[31]</sup> Nesse contexto, esse domínio ganha importância, já que os mAbs direcionados a diferentes regiões da proteína *Spike* se ligam ao NTD formando um complexo NTD/mAb, evitando alterações conformacionais na proteína, bloqueando a fusão da membrana e a entrada viral.<sup>[32]</sup> É notável que a subunidade S1 desempenha um papel crucial na ligação da proteína *Spike* ao receptor ACE2, enquanto a subunidade S2 completa a missão de fusão e entrada viral.

A subunidade S2 é responsável por intermediar a fusão viral após a clivagem proteolítica pelas proteases da célula hospedeira. Ela é composta por diversas sub-regiões, incluindo o peptídeo de fusão (FP), as regiões Heptadas de Repetição 1 (HR1), as regiões Heptadas de Repetição 2 (HR2), o domínio transmembranar (TM) e o domínio citoplasmático de fusão (CP).<sup>[36]</sup> O peptídeo de fusão (FP), situado no subdomínio S2, consiste em um pequeno seg-



mento de 15 a 20 aminoácidos conservados da família viral. Ele é composto principalmente por resíduos hidrofóbicos, como glicina (G) ou alanina (A), que se ancoram à membrana alvo quando a proteína *Spike* adota a conformação pré-hairpin. Estudos prévios demonstraram que o FP desempenha um papel essencial na mediação da fusão das membranas, conectando as bicamadas lipídicas da membrana da célula hospedeira e interagindo com o hACE2.<sup>[37]</sup>

As duas sub-regiões Heptadas de Repetição, HR1 e HR2, contêm uma sequência de heptapeptídeos repetidos HPPHCPC, em que H denota um resíduo hidrofóbico, P é polar e C é carregado. Essas sub-regiões interagem entre si por meio de interações hidrofóbicas para formar um feixe de seis hélices (6-HB), um componente essencial para a fusão viral e a função de entrada da subunidade S2. A HR1 está localizada no C-terminal de um FP hidrofóbico, enquanto a HR2 se encontra no N-terminal do domínio TM.<sup>[36]</sup> O domínio TM da subunidade S2 é altamente conservado no SARS-CoV-2 e, juntamente com o domínio CP, contribui para a aderência da proteína *Spike* às células hospedeiras.<sup>[37]</sup>

A proteína *Spike* tem sido uma das mais estudadas pela comunidade científica, devido à sua importância na invasão das células hospedeiras. Isso a torna um alvo ideal para o desenvolvimento de um inibidor que possa estimular uma resposta imunológica. Contudo, o SARS-CoV-2 evita essa resposta por meio de heterogeneidade conformacional, glicosilação, mutações e mimetismo peptídico do hospedeiro.<sup>[31]</sup> Isso ressalta a importância do estudo dessa proteína para toda a comunidade científica. A seguir, examinaremos as principais abordagens encontradas na literatura tanto para a proteína *Spike* quanto para a PLpro.



### 3.3 Métodos de Simulação Aplicados no Estudo do SARS-CoV-2.

Ao analisar o estado atual da aplicação dos métodos de modelagem e simulação molecular no estudo do SARS-CoV-2, podemos observar uma grande quantidade de estudos que utilizam essas técnicas para investigar a conformação da proteína viral, a variante do coronavírus, a estrutura-alvo e, principalmente, na busca por novos medicamentos e na identificação de possíveis agentes terapêuticos, bem como nas interações com as proteínas do SARS-CoV-2. Essas interações desempenham um papel fundamental na compreensão da invasão viral e da subsequente replicação nas células humanas. Portanto, é evidente que a utilização desses métodos tem sido amplamente reconhecida como uma abordagem destacada na investigação das propriedades dessas estruturas críticas para a infecção e replicação do vírus.

Dentro do atual contexto, devido à urgência de encontrar medidas preventivas para a Covid-19, a Food and Drug Administration (FDA) instituiu um programa de emergência especial denominado *Coronavirus Treatment Accelerated Program (CTAP)*,<sup>[38]</sup> que realizou a revisão de 270 ensaios clínicos e acompanhou mais de 570 programas de desenvolvimento de medicamentos ainda em fase de planejamento. Vários desses programas foram relatados na literatura e empregaram técnicas de descoberta de medicamentos auxiliada por computador.<sup>[39]</sup> Essas metodologias surgiram como ferramentas potenciadoras e eficazes no processo de descoberta de novos fármacos e têm sido empregadas ao longo da última década para identificar potenciais inibidores de proteínas,<sup>[40]</sup> explorar interações entre proteínas e peptídeos,<sup>[41]</sup> bem como interações entre proteínas.<sup>[42]</sup>

Sob essa perspectiva, os cientistas conseguiram identificar possíveis peptídeos inibidores do SARS-CoV-2 por meio de diferentes abordagens, como a triagem virtual de bibliotecas contendo milhões de compostos, ancoragem molecular, simulações de dinâmica molecular<sup>[43]</sup> e química quântica.<sup>[44]</sup>

Como o desenvolvimento de um medicamento candidato até ser aprovado é um processo longo e caro, em torno de 15 anos e 3 bilhões de dólares, a combinação de metodologias computacionais como triagem virtual, ancoragem molecular, simulação de DM e avaliação de energia livre de ligação, servem como ferramentas alternativas promissoras para auxiliar na identificação de candidatos a fármacos em potenciais.<sup>[45]</sup> A seguir, serão apresentados alguns trabalhos encontrados na literatura que utilizaram métodos de simulação molecular

para estudar e identificar possíveis inibidores, conformação, mutações e características das principais proteases do vírus do SARS-CoV-2.

Os estudos que serão abordados foram pesquisados no portal de periódicos da CAPES usando a pesquisa avançada, selecionando as bases de dados voltadas para as áreas de química e biologia em qualquer campo que contenha “*molecular dynamics*” ou “*molecular simulation*” ou “*molecular modelling*” ou “*biomolecular simulation*” e “SARS-CoV-2”, “*Spike protein*” e “*papain-like protease*” classificados conforme a relevância e revisados por pares.

### 3.3.1 Métodos de Simulação Aplicados à Papain-Like Protease (PLpro)

Enquanto na glicoproteína *Spike* os pesquisadores estudam formas de inibir a ligação viral ao receptor ou a fusão viral na membrana da célula hospedeira, na PLpro os pesquisadores voltam suas atenções para a inibição da replicação viral e reação inflamatória.

Os cientistas já identificaram que esta protease possui a capacidade de clivar e interagir preferencialmente com o gene 15, estimulado pelo interferon humano (hISG15). Essa interação visa suprimir a resposta imunológica inata do hospedeiro. Portanto, é crucial compreender o mecanismo de interação, bem como explorar possíveis inibidores para esta protease do SARS-CoV-2. Neste contexto, Leite e colaboradores<sup>[23]</sup> utilizaram uma abordagem que combinou raios-X de pequeno ângulo e espalhamento de nêutrons com técnicas computacionais. O objetivo era investigar o mecanismo e a dinâmica dessa interação entre o vírus e o hISG15.

Os resultados dos estudos realizados pelos autores revelaram que, após a ligação com a PLpro, o hISG15 passa por uma transição de um estado estendido para um estado compacto. Esta conformação não havia sido observada em estudos anteriores envolvendo complexos de SCoV2-PLpro com ISG15. A análise computacional foi conduzida por meio de um conjunto de técnicas e programas, incluindo o MODELLER para a geração de estruturas e o BioXTAS RAW para a análise da distribuição de pares. Durante a análise teórica, os autores também observaram uma flexibilidade conformacional significativa no domínio N-terminal do ISG15. Isso sugere que o domínio N-terminal pode ter uma interação menos intensa com a PLpro em comparação com o domínio C-terminal, que apresenta uma forte ligação com a PLpro.

No entanto, há estudos na literatura que apontam o contrário. Por exemplo, uma pesquisa conduzida por Roy, R. et al.<sup>[46]</sup> utilizando simulação de dinâmica molecular atomística indica que a interação entre o N-terminal do hISG15 e a PLpro é mais favorável. Além disso,

resultados experimentais publicados por Shin, D. et al.<sup>[25,26]</sup> também sugerem que o domínio de dobras de ubiquitina N-terminal de ISG15 amplifica a interação com a PLpro, essas descobertas corroboram com os resultados teóricos obtidos. Vale destacar que Roy, R. colaboradores<sup>[46]</sup> realizaram simulações de dinâmica molecular em toda a família de coronavírus, incluindo o SARS-CoV-1, o MERS-CoV e o SARS-CoV-2. A análise da energia livre de ligação sugere que o hISG15 se liga mais fortemente à PLpro do SARS-CoV-2 em comparação com os SARS-CoV-1 e o MERS-CoV. Essas pesquisas ressaltam a importância de compreender o complexo de desISGilação do SARS-CoV-2, o que pode ser crucial para orientar o desenvolvimento de terapias direcionadas a esse complexo no contexto do tratamento da COVID-19.

Devido ao papel crucial dessa protease na replicação viral e por ela ainda desempenhar a função de clivar proteína 15 estimulado por interferon semelhante à ubiquitina (ISG15) das proteínas do hospedeiro. Como resultado, vários grupos de pesquisas têm investigado possíveis inibidores da PLpro por meio de triagem in silico de milhões de compostos de grandes bibliotecas.<sup>[47]</sup>

Os pesquisadores Elseginy, S. A. e Anwar, M. M.<sup>[47]</sup> realizaram uma triagem virtual em uma biblioteca contendo 15 milhões de compostos, aplicando técnicas de ancoragem molecular com o auxílio do BIOVIA Discovery Studio. Inicialmente, os autores selecionaram os compostos de maior pontuação, escolhendo-os com base em sua afinidade de ligação com a hidrolase ubiquitina carboxi-terminal humana L1 (UCH-L1), a qual, como uma enzima desubiquitilante, compartilha semelhanças estruturais e funcionais com a PLpro. No decorrer do artigo, eles identificaram 387 compostos como possíveis inibidores da PLpro. Dentre esses candidatos, foram recuperados os 20 compostos promissores, com valores de IC<sub>50</sub> estimados para PLpro através de um modelo de regressão linear múltipla na faixa de 159 a 505 nM. Baseando-se nas análises realizadas, os pesquisadores ranquiaram os possíveis candidatos muito promissores para futuras avaliações in vitro, os quais estão disponíveis para consulta [aqui](#). Esses compostos exibem modos de ligação semelhantes aos inibidores não covalentes das estruturas cristalinas e apresentaram interações adicionais, determinantes para a afinidade de ligação ser potencialmente alta.

Muitos trabalhos estão sendo desenvolvidos com as principais proteínas dos SARS-CoV-2 essenciais para a replicação do vírus e em grande maioria usam metodologia de ancoragem molecular para triagem virtual de milhões de compostos encontrados em grandes bancos de dados. Nessa perspectiva, Rajpoot, S. et.al<sup>[48]</sup> propuseram projetar um inibidor duplo para protease principal (3CLpro; PDB ID: 6M2Q) e para protease semelhante à papaína (PLpro;

PDB ID: 6W9C), responsável por processar a cadeia de poli proteína traduzida do RNA viral, produzindo proteínas virais funcionais. Foram selecionados 9 medicamentos conhecidos que bloqueiam a díade catalítica (His-41 e Cys-145) da 3CLpro, e a tríade catalítica (Cys-111, His-272 e Asp-286) da PLpro. A ideia dos autores de reaproveitarem drogas conhecidas é de acelerar a busca por um fármaco potente anti-coronavírus. Os compostos foram selecionados no ZINC15 Data Base e após fazerem o estudo de docagem molecular, selecionaram 9 hits para realizarem as simulações de dinâmica molecular. Essas simulações foram realizadas no GROMACS 4.5.5 com o campo de força GROMOS 53a6. As análises das equilibrações do complexo formado proteína-ligante foram analisadas através do RMSD dos átomos de backbone em 50 ns, tempo total de simulação. Três ligantes promissores (ZINC-389747, ZINC-1530637 e ZINC-49153) foram selecionados para prosseguimento nas investigações *in vitro*.

Ambas as proteínas 3CLpro e a PLpro complexadas com os ligantes ZINC-389747, ZINC-1530637 e ZINC-49153 mostram um RMSD médio de 0,25-0,5 nm, 0,3-0,4 nm e 0,4-0,45 nm, bem como 0,35-0,4 nm, 0,5-0,55 nm e 0,35-0,4 nm, respectivamente. Todos os complexos proteína-ligante atingiram suas conformações mais estáveis no final das simulações de 50 ns.

Outra metodologia fundamental para compreender a conformação de uma proteína ao longo de uma simulação é a dinâmica molecular. Bosken et al.<sup>[49]</sup> demonstraram, por meio de simulações de dinâmica molecular com duração de 1  $\mu$ s, realizadas com o pacote Amber18 com aceleração para GPU, usando os campos de força ff14SB e Amber GAFF2. Eles investigaram que a dinâmica geral da protease PLpro do SARS-CoV-1, MERS-CoV e SARS-CoV-2 sem ligante e com ligantes, é semelhante, com flexibilidade comparável na alça BL2, domínio do dedo de zinco e domínio Ubl. Isso é importante para elucidar semelhanças e diferenças relevantes para o desenvolvimento de fármacos e a ligação de proteínas semelhantes à ubiquitina. Além disso, os autores conduziram simulações com ligantes a base de naftaleno ligados a PLpro-SCoV1 e PLpro-SCoV2, explorando diversas posições e diferentes locais de ligação conhecidos para analisar os seus modos de ligação.

Nesse contexto, as análises de simulação de DM resultou na criação de um conjunto de conformações da PLpro-SARS-CoV-2, demonstrando possíveis interações entre inibidores e proteínas. Essas informações são essenciais para o desenvolvimento de inibidores com base na estrutura da proteína, bem como para entender a dinâmica relevante da proteína em relação à ligação com proteínas semelhantes à ubiquitina (Ub) ou à Ubl.

Já os pesquisadores Alamri, M. A. e colaboradores<sup>[50]</sup> realizaram uma triagem virtual

de aproximadamente 7.000 compostos candidatos a fármacos. Nesse trabalho, os autores identificaram três compostos candidatos ADM<sub>13083841</sub>, LMG<sub>15521745</sub> e SYN<sub>15517940</sub>. Esses três compostos estabeleceram interações conservadas que foram exploradas por meio de simulações de dinâmica molecular, cálculos de energia livre e contribuição de energia residual estimada pelo método MM-PB(GB)SA para obter uma visão profunda de seu modo de ligação, mecanismo de interação molecular e análise de ADMET. Todos esses compostos apresentaram conformação estável e interagiram bem com os resíduos ativos da SCoV2-PLpro, além de apresentarem um perfil de interação consistente com a SCoV1-PLpro e a PLpro-MERS-CoV.

Todas as simulações DM dos complexos inibidores da SCoV2-PLpro e peptídeo-proteína foram conduzidas ao longo de 50 ns usando o pacote GROMACS 2018, com base em dados previamente relatados na literatura. Em resumo, o UCSF Chimera 1.14 foi utilizado para preparar a estrutura cristalina da PLpro e o complexo com as melhores poses dos compostos ancorados, a fim de realizar a simulação de DM. O estudo de simulação molecular realizado por esses pesquisadores desempenhou um papel importante na compreensão das evidências estruturais, bem como no caminho para o desenvolvimento de possíveis inibidores da PLpro no combate ao vírus causador da COVID-19 e outros coronavírus humanos patogênicos.

Portanto, a utilização cruzada de ferramentas de dinâmica molecular (para ter acesso a uma ampla abrangência informacional das proteínas) em conjunto com ferramentas de busca de assinaturas funcionais de densidade eletrônica para compreender características-chave de duas enzimas fundamentais para a infecção e replicação do SARS-CoV-2, a proteína *Spike* (infecção) e a PLpro (replicação), bem como, mutações do tipo único aminoácido alteram estas características, de forma a otimizar o planejamento de fármacos, compreensão da infectividade e imunogenicidade, tal como o planejamento de novas vacinas eficientes para as múltiplas variantes do vírus são fundamentais para a comunidade científica.

### 3.3.1.1 Métodos de Simulação Aplicado ao Domínio de Ligação ao Receptor da Proteína *Spike*

Como já foi discutido, a proteína *Spike* do SARS-CoV-2, sobretudo o RBD, é responsável pelo primeiro contato e invasão do vírus à célula humana através da ACE2. Para compreender o processo de ligação do domínio de ligação ao receptor (RBD) da proteína *Spike* com a ACE2, Chen e colaboradores<sup>[51]</sup> utilizaram simulação de dinâmica molecular de todos os átomos e métodos de amostragens aprimoradas.

Os resultados desse estudo sugerem a existência de múltiplos estágios no processo de ligação, com as interações polares desempenhando um papel dominante na fase inicial de associação. Nesse contexto, as interações hidrofóbicas entre os resíduos na região do RBM (compreendida entre os resíduos 439-509 do RBD) e a ACE2 são de extrema importância para a formação dos intermediários de ligação. Com base nessas descobertas, a presença de um inibidor nessa região de ligação teria o efeito de estabilizar a estrutura do RBM e facilitar as interações hidrofóbicas, potencialmente reduzindo a barreira de energia livre para a formação de um complexo estável no estágio final do processo.

Consequentemente, o RBD e a ACE2 se destacam como alvos promissores na busca por neutralizar a infecção provocada pelo SARS-CoV-2. É essencial, contudo, compreender a natureza das interações que se estabelecem entre essas duas proteínas. O estudo conduzido por Sartore e colaboradores,<sup>[52]</sup> embora tenha como principal objetivo investigar as razões pelas quais indivíduos com diabetes mellitus desenvolvem formas graves da COVID-19, também revelou descobertas significativas sobre como a glicosilação dos resíduos de aminoácidos de lisina afeta o número de contatos não polares.

Após a ocorrência da glicosilação não enzimática, observou-se uma redução no número de interações entre o ACE2 e a proteína *Spike* do SARS-CoV-2 quando comparado ao modelo não modificado. Para realizar essa análise, os pesquisadores utilizaram a ferramenta de encaixe proteína-proteína MOE, gerando, classificando e examinando visualmente um conjunto de 100 poses de ancoragem molecular. Em seguida, as poses mais relevantes obtidas por meio desse processo de ancoragem molecular foram empregadas para identificar e categorizar os tipos de interações que ocorrem, tanto entre as moléculas como dentro delas, no contexto da ligação entre o RBD e a ACE2.

Desde o início da pandemia da COVID-19, diversas estruturas estáticas têm fornecido informações valiosas sobre a interação entre o SARS-CoV-2 e seu hospedeiro. Entretanto, dado que os processos de reconhecimento molecular, ligação e funções são dinâmicos por natureza, torna-se essencial recorrer a simulações moleculares para uma compreensão mais profunda do processo de interação, particularmente no que diz respeito à ligação entre o RBD da proteína *Spike* do SARS-CoV-2 e o receptor ACE2. Ali, A. et al.,<sup>[53]</sup> por exemplo, utilizaram simulações de dinâmica molecular em triplicata, cada uma com duração de 500 ns, utilizando o software Desmond System Builder (Schrödinger, LLC, New York, NY) e o campo de força OPLS. O objetivo desse estudo com as simulações era aprofundar o entendimento da estabilidade estrutural e das interações intermoleculares entre o RBD da proteína *Spike* do SARS-CoV-2 e o SARS-CoV-1 quando ligados ao receptor ACE2.



Os resultados dessas simulações indicaram que esses dois vírus adotam estratégias distintas para estabelecer uma ligação estável com a ACE2. Essa descoberta é relevante, uma vez que mutações específicas podem ocorrer na proteína *Spike* dos vírus, e essas mutações podem apresentar características e propriedades químicas distintas. Dentre as interações identificadas, os autores destacaram a presença de uma ponte salina estável entre o Lys-417 da proteína do SARS-CoV-2 e o Asp-30 do receptor ACE2, além de três ligações de hidrogênio estáveis entre os resíduos de aminoácidos Tyr-449, Gln-493 e Gln-498 localizados no RBM do SARS-CoV-2 e o Asp-38, Glu-35 e Lys-353 da ACE2. Essas descobertas lançam luz sobre os mecanismos de interação entre esses vírus e seu receptor hospedeiro. Essas interações não foram encontradas na ligação do RBM do SARS-CoV-1 com a ACE2. Estudos de energia livre de ligação baseada na área de superfície de Born (MM-GBSA) foram maior para o SARS-CoV-2 em todas as simulações. Ainda de acordo com Ali, A. e colaboradores,<sup>[53]</sup> esses estudos são importantes porque deve-se dar uma atenção especial a essas interações estáveis ao projetar possíveis fármacos e modalidades de tratamento para atingir ou interromper essas interações.

As interações entre o RBD e ACE2 envolvem ligações específicas, tornando-as alvo de estudos mais aprofundados,<sup>[30]</sup> como demonstrado em um artigo anterior. Isso ocorre porque os resíduos de aminoácidos presentes nessas interações podem possuir propriedades diversas, classificadas de acordo com as características de suas cadeias laterais, incluindo apolares, ácidos, básicos e polares. Por exemplo, a substituição de uma tirosina classificada como polar por uma arginina classificada como básica poderia causar algum impacto na ligação ou conformação da proteína? Essa é uma questão relevante a ser explorada.

Zech, F. et al.,<sup>[54]</sup> em seu artigo, observaram por meio de técnicas experimentais e métodos teóricos que o resíduo 403, localizado no RBD da proteína *Spike*, poderia afetar a ligação da proteína *Spike* do coronavírus ao receptor ACE2 humano por meio da simples mutação T403R. Muitos acreditam que a origem da infecção pelo coronavírus possa ter relação direta com esse vírus de morcego. No entanto, conforme apontado por Zech, F. et. al. eficientemente com o receptor ACE2 humano. No artigo, os autores analisaram o vírus de morcego (RaTG13) do subgênero sarbecovírus da família Coronaviridae, que é um parente próximo do SARS-CoV-2, e introduziram uma única mutação, T403R. Eles demonstraram que essa única mutação aumentou a afinidade do RaTG13 pelo ACE2 humano, possibilitando a infecção de pseudopartículas em células pulmonares humanas e organoides intestinais. No entanto, a mutação R403T no SARS-CoV-2 teve o efeito oposto, reduzindo a infecção por pseudopartículas e a replicação viral.

As simulações de dinâmica molecular, realizadas com ReaxFF (dinâmica molecular reativa) no Amsterdam Modeling Suite 2020, confirmaram a formação de interações de proximidade e carga potencial entre o R403 no SARS-CoV-2 e o E37 no receptor ACE2 humano. Estudos como este destacam como uma única mutação pode ter um impacto significativo na interação entre os resíduos de aminoácidos da proteína *Spike* e o receptor, ilustrando a dinâmica das novas mutações do SARS-CoV-2.

Nessa perspectiva, diversos estudos podem ser encontrados na literatura sobre as principais mutações já ocorrida nas principais proteases do SARS-CoV-2. Muitos desses estudos mostram que a glicosilação da proteína *Spike* do vírus e do receptor da ACE2 na célula hospedeira é fundamental para a infectividade do vírus. Rahnema, S. e colaboradores<sup>[55]</sup> utilizaram simulações de dinâmica molecular realizadas no NAMD2 com campo de força CHARMM36m e modelo de solvatação com água do tipo TIP3P para explorar o papel de um novo sítio ativo de O-glicosilação mutado (D494S) no RBD da proteína *Spike*. Os autores observaram através de seus estudos que o grau de afinidade entre a proteína e o receptor tem relação com o tamanho do O-glicano ligado, ou seja, quanto maior for O-glicano, maiores foram as afinidades de ligações. Portanto, esse fator crucial deve ser levado em consideração para quaisquer outras abordagens inibitórias em relação à interação RBD-ACE2.

O esforço dedicado dos cientistas permitiu a disponibilidade de uma vasta quantidade de informações sobre a COVID-19, o que, por sua vez, pode permitir o desenvolvimento de peptídeos e técnicas de inibição direcionada. Em um trabalho muito interessante, Fernandez-fuentes, N. e colaboradores<sup>[56]</sup> exploraram essa riqueza de dados estruturais disponível na literatura. Eles se concentraram na criação de peptídeos voltados para a interação entre o RBD do SARS-CoV-1 e SARS-CoV-2 com o receptor ACE2, bem como na análise da superfície de dimerização dos monômeros ACE2. Os pesquisadores empregaram o método inovador chamado PiPreD para modelar esses peptídeos. Esse método se baseia nos elementos naturais da interação entre a proteína alvo e seus parceiros cognatos, que são posteriormente incorporados nos peptídeos projetados. Como resultado, esses peptídeos conseguem replicar fragmentos de resíduos encontrados na interface nativa, introduzindo também novas e altamente diversas conformações. Essa abordagem substitui as principais interações na interface, tornando-se uma valiosa orientação para experimentos futuros. Além disso, o modelo estrutural dos peptídeos serve como base para simulações subsequentes.

Em mais de dois anos de pandemia, o aparecimento de novas mutações é um fenômeno natural esperado em um processo evolutivo de qualquer vírus, com isso novas variantes



do SARS-CoV-2 estão surgindo. Dependendo da mutação e de onde ocorra no genoma, a substituição pode ser neutra, benéfica ou prejudicial para o organismo. Na proteína *Spike* essas mutações ocorrem particularmente no RBD por causa do seu papel de reconhecimento e ligação ao ACE2. Portanto, as alterações na sequência de aminoácidos do RBD podem afetar drasticamente a afinidade de ligação da proteína *Spike* com a ACE2, bem como a sua conformação, muitos trabalhos têm sido desenvolvidos nesse sentido.

Em 2020, Chen, J. et al.<sup>[57]</sup> realizaram previsões com base em alinhamento de sequências, estimativas de probabilidade e análise de energia livre de ligação. Eles identificaram que os resíduos 452, 489, 500, 501 e 505 no RBM apresentavam uma probabilidade significativamente maior de sofrer mutações em cepas da COVID-19 mais infecciosas. Essas análises foram fundamentadas na estrutura cristalina de raios-X da proteína *Spike* e do ACE2 (PDB 6M0J).

Em um estudo publicado em 2021 por Gan, H. H. et al.,<sup>[58]</sup> os pesquisadores aplicaram métodos de análise estrutural às variantes altamente transmissíveis. Eles descobriram que as mutações pontuais circulantes, como S477N, E484K e N501Y, apresentavam uma afinidade aproximadamente 40% maior com o ACE2 em comparação com a proteína selvagem. Quando essas mutações apareciam em combinação, como S477N/E484K e E484K/N501Y, a afinidade era cerca de 80% maior do que na proteína não mutada. Esses dados estão correlacionados com a rápida disseminação do vírus em vários países. Por outro lado, mutações como N440K, L452R, E484K/Q/R e K417N/T foram identificadas como capazes de permitir um significativo escape aos anticorpos.

As mutações K417N e E484K, identificadas nas variantes da África do Sul e do Brasil, bem como as variantes que apresentam as mutações L452R e E484Q, encontradas na Índia, demonstraram resistência aos anticorpos e uma maior capacidade de transmissão no contexto das diferentes variantes do SARS-CoV-2. Essas mutações são comumente referidas como "mutações de escape da vacina" e representam uma ameaça para a eficácia das vacinas existentes, uma vez que reduzem sua capacidade de proteção. Uma lista abrangente das principais mutações de escape pode ser encontrada em um estudo conduzido por Wang et al.,<sup>[59]</sup> que inclui mutações como S494P, Q493L, K417N, F490S, F486L, R403K, E484K, L452R, K417T, F490L, E484Q e A475S. Detalhes sobre as principais mutações que evadem o reconhecimento pelos anticorpos podem ser encontrados no estudo publicado por Greaney, A. J. e colaboradores.<sup>[60]</sup> Algumas variantes do vírus apresentam mutações que aumentam tanto sua capacidade de transmissão quanto sua resistência ao sistema imunológico.<sup>[37]</sup> Exemplos disso incluem a linhagem P.1 (anteriormente conhecida como B.1.1.28.1), que

apresenta as mutações E484K, K417T e N501Y, e a variante B.1.351, originada na África do Sul, que apresenta as mutações K417N, E484K e N501Y.

Consequentemente, todas essas mutações descritas até aqui são encontradas nas variantes Alpha, Beta, Gamma, Delta e Omicron. Em março de 2022, essas cinco cepas do SARS-CoV-2 foram rotuladas pela organização mundial da saúde como “variantes preocupantes”, contudo, a comunidade científica está monitorando essas variantes e mutações, pois elas podem comprometer as vacinas e terapias de anticorpos existentes.

### 3.3.1.2 Métodos QM Aplicados no Estudo do SARS-COV-2

Como é possível observar até aqui, os métodos de simulações são importantes para selecionar compostos, estudar conformações e grau de alta afinidade proteínas-alvo, neste caso as proteínas do SARS-CoV-2, contudo, algumas metodologias como, por exemplo: dinâmica molecular não consegue prever os aspectos eletrônicos dos sistemas.<sup>[61]</sup> Consequentemente, são necessários o emprego de métodos que utilizam a mecânica quântica para tratar a estrutura eletrônica dos sistemas.

Muitos trabalhos encontrados na literatura utilizaram métodos de química quântica para compreender as interações eletrônicas na proteína. Nesse sentido, estudos recentes, como o de Watanabe, C. et al.,<sup>[62]</sup> usaram QM para avaliar e caracterizar quantitativamente ligações de hidrogênio  $xH/\pi$  ( $x = N, O$  e  $C$ ) e pontes salinas entre a proteína SARS-CoV-2 com o ACE2 e com o anticorpo B38 Fab envolvido na ligação inibitória do ACE2. Em outro trabalho desenvolvido por Khelifaoui, H. e colaboradores,<sup>[63]</sup> os descritores de reatividade foram utilizados para calcular a dureza e moleza global, apontando a estabilidade do sistema e indicando que o Ramipril é o fármaco mais reativo, apresentando o maior valor de eletrofilicidade seguido pelo ORE-1001, Cloroquina e Lisinopril. Estudos como esse são muito importantes para compreender o mecanismo de reação entre a proteína e o possível inibidor.

O estudo da estrutura eletrônica de biomoléculas é desafiador devido à quantidade de elétrons a serem tratados. Por isso, inúmeros trabalhos encontrados na literatura utilizam a química quântica para fazer o tratamento eletrônico de resíduos-chaves para entender o processo de inibição, conformação ou transferência de elétrons. Nessa perspectiva, Fernandes, Henrique S. e colaboradores<sup>[64]</sup> mostraram o mecanismo catalítico da Main protease (Mpro), caracterizando o estado de transição de cada etapa descrito com os reagentes e produtos. É importante ressaltar que nos métodos de simulação baseados na mecânica

molecular não temos o tratamento eletrônico.

Muitos trabalhos<sup>[65-67]</sup> com a Mpro tentaram caracterizar o mecanismo inibitório, sobretudo no sítio ativo dessa protease na busca por um péptido inibidor. Esses trabalhos são de fundamental importância no desenvolvimento de uma nova vacina para o SARS-CoV-2. Trabalhos similares são desenvolvidos com outras proteases vitais para a replicação do vírus como a PLpro.<sup>[44,68]</sup>

O estudo desenvolvido por Rafael et al.<sup>[69]</sup> utilizou o PRIMoRDiA para calcular descritores quânticos para analisar a trajetórias de dinâmica molecular e estabelecer correlações entre esses descritores e a abertura do sítio ativo da Mpro do SARS-CoV-1 e SARS-CoV-2. Os resultados dos descritores de netfilicidade e dureza local indicaram que a Mpro do SARS-CoV-1 modela a reatividade da cisteína para a transferência de prótons no sítio ativo, conforme descrito pela dureza local, e que os valores observados não foram os mesmos para a reatividade dessa transferência de carga, descrita pelas interações mole-mole pelo descritor de netfilicidade.

## 3.4 Modelagem Molecular

Inicialmente, modelagem é a ação ou o efeito de modelar, ou seja, a modelagem molecular é a ciência de representar estruturas moleculares numericamente através da simulação de seu comportamento, abrangendo uma ampla variedade de métodos teóricos e computacionais usados para representar a estrutura das moléculas, íons e/ou partículas por meio de equações da física quântica e clássica.<sup>[70]</sup>

Os métodos teóricos de modelagem molecular geralmente se dividem em duas abordagens principais: aqueles que aplicam aproximações quânticas, com base na Mecânica Quântica (QM), e aqueles que empregam aproximações clássicas, fundamentadas na Mecânica Molecular (MM). Além disso, existem métodos híbridos que combinam elementos tanto da Mecânica Quântica quanto da Mecânica Molecular, conhecidos como métodos QM/MM. Uma das diferenças fundamentais entre esses métodos reside no custo computacional envolvido. Em sistemas com muitos átomos,<sup>[71]</sup> como macromoléculas, particularmente proteínas, os métodos quânticos podem se tornar impraticáveis devido ao seu elevado custo computacional. Nesses casos, as simulações de dinâmica molecular frequentemente representam a escolha mais viável.

### 3.4.1 Dinâmica Molecular

Dinâmica Molecular (DM) é um método baseado na mecânica clássica que utiliza as equações de movimento Newtonianas para calcular e gerar informações em nível microscópico, tais como posições e velocidades de cada átomo que constitui um sistema em diferentes e consecutivos instantes de tempo. Essas propriedades podem ser calculadas em diferentes condições de temperatura e pressão. Esse método é amplamente empregado para simular os movimentos de proteínas e outros sistemas moleculares, permitindo prever como cada átomo em uma proteína ou em sistemas semelhantes se deslocará ao longo do tempo, em um modelo geral da física que rege as interações interatômicas de acordo com a dinâmica clássica.<sup>[72]</sup>

A dinâmica molecular clássica consiste na integração das equações de movimento de Newton para um sistema molecular. Logo, o movimento da  $i$ -ésima partícula de um sistema é dada por:

$$\frac{d^2 r_i(t)}{dt^2} = \frac{F_i(t)}{m_i} \quad (3.1)$$

$$a_i(t) = \frac{F_i(t)}{m_i} = -\frac{dU(r(t))}{m_i dr_i(t)} \quad (3.2)$$

Aqui,  $F_i(t)$  representa a força que atua sobre cada partícula do sistema em um dado momento  $t$ ,  $r_i(t)$  denota o vetor posição no instante  $t$ , e  $m_i$  é a massa da partícula. A aceleração do sistema, por sua vez, é calculada com base nas forças que agem sobre cada átomo  $i$ . Essas forças são determinadas por um campo de força, no qual  $dU(r(t))$  representa a função de energia potencial associada a esse campo.<sup>[73,74]</sup>

A princípio, a ideia básica por trás da DM é bem simples. Dadas as posições de todos os átomos em um sistema molecular, seus núcleos são tratados como partículas clássicas. A partir dessa consideração, pode-se calcular a força exercida em cada átomo por todos os outros átomos, baseando-se nas leis do movimento de Newton para prever a posição espacial de cada átomo em função do tempo. Consequentemente, a cada passo no tempo, as forças são calculadas sobre cada átomo e, posteriormente, utilizadas para atualizar a posição e a velocidade de cada átomo.<sup>[75]</sup>

Nesse contexto, poder-se-ia pensar que a aplicação desse método de simulação é extremamente nova, algo descoberto no século XXI; contudo, não é bem assim. As primeiras simulações de DM foram realizadas em gases simples por Alder, B. J. e Wainwright, T. E.<sup>[76]</sup> no século passado, por volta de 1950. Já as primeiras simulações de DM em proteínas foram realizadas cerca de 27 anos depois, no final da década de 1970, por McCammon et al.<sup>[77]</sup> em um trabalho publicado com o seguinte nome “*Dynamics of folded proteins*”. Esse foi o primeiro passo para a utilização de simulações de DM em sistemas biológicos.

Os trabalhos de Lifson, S. et al. e Levitt, M. et al.,<sup>[78,79]</sup> que serviram de base para essas simulações em macromoléculas, receberam reconhecimento da comunidade científica, agraciados com o Prêmio Nobel de Química em 2013. Contudo, as simulações tornaram-se populares nos últimos anos, sobretudo devido ao poder computacional existente atualmente. Desde 1950 até os dias atuais, já se passaram cerca de seis décadas, e durante esse tempo, as simulações de DM se tornaram ferramentas consagradas em estudos de sistemas moleculares, complementando as técnicas experimentais, principalmente em biomoléculas.<sup>[80]</sup>

As simulações de DM são importantes em processos biomoleculares para capturar uma grande variedade de informações, tais como mudança conformacional, interação do ligante com a proteína alvo e dobramento das proteínas, demonstrando as posições atômicas com resolução temporal de femtossegundos. Outra característica fundamental é que essas

simulações podem prever como as proteínas responderão em nível atômico a perturbações, como mutações, fosforilação, protonação, adição ou remoção de um peptídeo em sistemas reais, como, por exemplo, em uma célula humana.<sup>[75]</sup>

Contudo, apesar de todo o progresso e sucesso desse método de simulação, dependendo do estudo que se deseja realizar, o método de DM pode apresentar vantagens e limitações, assim como qualquer outro método de simulação. Apesar de ser uma ferramenta poderosa de pesquisa em níveis atômico e molecular, complementar aos estudos experimentais, essas limitações inerentes devem ser consideradas para o uso adequado desta técnica.

A principal vantagem desse método de simulação é conectar fenômenos em escala atômica às propriedades medidas experimentalmente.<sup>[81]</sup> Oferece uma grande variedade de parametrização com o potencial de evitar a necessidade de realizar experimentos longos, de alto custo e com taxas de sucesso pequenas,<sup>[82]</sup> além de conseguir descrever propriedades de macromoléculas com centenas de átomos. Principalmente, permite explorar a dinâmica das estruturas-alvo nos mecanismos de inibição, fornecendo informações detalhadas sobre os processos de mudança estrutural e de energia, e, desta forma, revelar o mecanismo molecular de resistência a drogas de uma estrutura-alvo, algo ainda impossível de ser realizado pelos métodos baseados na mecânica quântica. Esse mecanismo é importante para orientar o desenvolvimento de uma nova geração de inibidores para um alvo específico, como, por exemplo, o SARS-CoV-2.<sup>[83]</sup>

Se em 1959, as limitações do método estavam relacionadas ao pequeno número de partículas que podiam ser utilizadas, pois o tamanho do sistema de moléculas era limitado pela capacidade de armazenamento e velocidade das máquinas de computação.<sup>[84]</sup> Atualmente, essas limitações não são tão pertinentes, mas ainda existem desafios para abranger todas as mudanças conformacionais relevantes para uma determinada função biológica. É possível realizar simulações abrangentes através da Dinâmica Molecular, desde que haja recursos computacionais e, principalmente, de armazenamento suficiente. No entanto, permanecem algumas limitações, por exemplo, no caso de uma proteína-inibidor, onde as simulações são limitadas a uma escala de tempo de micro ou milissegundos.<sup>[83]</sup> Outra limitação significativa das simulações de dinâmica molecular é a sua geralmente menor precisão em comparação com os métodos baseados na mecânica quântica.<sup>[83]</sup>

Como mencionado anteriormente, a dinâmica molecular emprega uma função de energia potencial fundamentada na física, conhecida como campo de força, para realizar simulações em função do tempo, seguindo os princípios da mecânica newtoniana clássica. A seguir, abordaremos os campos de força e destacaremos os parâmetros cruciais para conduzir uma

simulação de dinâmica molecular.

### 3.4.1.1 Campos de Força

De forma geral, entendemos campo de força como um conjunto de parâmetros capaz de proporcionar descrições precisas das proteínas<sup>[85]</sup> por meio do cálculo da energia potencial total do sistema. Essa energia é resultante da soma das forças intra e intermoleculares. As funções de forças específicas (intra e intermoleculares) correspondem a uma única molécula ou a uma classe de moléculas relacionadas. Esses conjuntos de funções e parâmetros, ou seja, o campo de força, são derivados de dados experimentais e cálculos de mecânica quântica.<sup>[86]</sup> Muitos pesquisadores têm se dedicado à parametrização desses campos de força, o que resultou em melhorias significativas nos conjuntos de funções e parâmetros ao longo da última década.<sup>[75]</sup>

Há vários campos de força disponíveis para a modelagem molecular, sendo os mais utilizados nas simulações: AMBER,<sup>[87]</sup> OLPS,<sup>[88]</sup> CHARMM<sup>[89]</sup> e GROMOS.<sup>[90]</sup> Esses campos de força dividem a função de energia potencial em termos das contribuições dos átomos ligados (ligações químicas) e não ligados (interações intermoleculares). Essa separação em termos de átomos ligados e não ligados é de extrema importância para auxiliar no desenvolvimento dos algoritmos usados em simulações computacionais.<sup>[85]</sup>

Um campo de força típico, usado nas simulações de macromolécula, pode ser descrito pela Equação 3.3 a seguir:

$$\begin{aligned}
 U(r_1, \dots, r_N) = & \sum_{\text{Ligação}} \frac{a_i}{2} (l_i - l_{i0})^2 \\
 & + \sum_{\text{Ângulo}} \frac{b_i}{2} (\theta_i - \theta_{i0})^2 \\
 & + \sum_{\text{Diedros}} \frac{c_i}{2} [1 + \cos(n\varphi_i - \gamma_{i0})] \\
 & + \sum_{\text{par átomos}} 4 \varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
 & + \sum_{\text{par átomos}} k \frac{q_i q_j}{r_{ij}}
 \end{aligned} \tag{3.3}$$

os três primeiros termos (na cor vermelha) correspondem a todas as ligações, ângulos e ângulos diedrais, respectivamente, definidos pela estrutura covalente do sistema, enquanto nos dois últimos termos (na cor azul) os índices de soma representam todos os pares de átomos, separados por distâncias  $r$  e não ligados quimicamente.<sup>[91]</sup>



Fisicamente, o vetor  $(r_1, \dots, r_N)$  corresponde às coordenadas dos átomos, enquanto  $U(r_1, \dots, r_N)$  representa a energia potencial do sistema. Os dois primeiros termos da equação Equação 3.3,  $l_i$  e  $\theta_i$ , representam os comprimentos e ângulos de ligação, respectivamente, enquanto  $l_{i0}$  e  $\theta_{i0}$  denotam seus valores de equilíbrio. Os coeficientes  $a_i$  e  $b_i$  correspondem às constantes de força que restauram esses valores de equilíbrio, garantindo a estrutura química correta, ou seja, são as constantes de força para o comprimento e o ângulo de ligação, respectivamente.<sup>[73]</sup> Por outro lado, o terceiro termo descreve os ângulos diédricos em torno das ligações químicas, representando o potencial energético caracterizado pelos termos de periodicidade ( $n$ ) e pelas barreiras rotacionais definidas por  $c_i$ .<sup>[91]</sup>

Os termos de ligações atômicas geralmente são modulados utilizando potencial harmônico que apresenta uma boa aproximação da energia de um sistema apenas para desvios pequenos do comprimento da ligação de referência. Contudo, para desvios maiores, o potencial de Morse fornece uma representação mais precisa da energia potencial para o alongamento da ligação.<sup>[92]</sup>

Os dois últimos termos da equação correspondem aos efeitos de *Coulomb* e *Lennard-Jones* 6-12. Aqui,  $r_{ij}$  denota a distância entre os átomos,  $q_i$  e  $q_j$  representam as cargas parciais dos átomos,  $\epsilon_{ij}$  é a constante dielétrica do meio, e  $\sigma_{ij}$  são os parâmetros específicos de *Lennard-Jones*, cujos valores variam conforme os tipos de átomos envolvidos. Esses dois últimos termos da equação exigem um poder computacional substancial em comparação com os três primeiros termos, pois os átomos estão apenas ligados a alguns vizinhos, mas interagem com uma grande quantidade de outros átomos.

O termo de *Lennard-Jones* consiste em duas partes distintas: um termo atrativo ( $r^{-6}$ ), que descreve as forças de *van der Waals*, relacionadas a interações de curto alcance, como as forças de *Keesom*, *Debye* e *London*. Há também um termo repulsivo ( $r^{-12}$ ) que modela a repulsão de Pauli, associada às forças de curto alcance entre átomos que estão suficientemente próximos para que ocorra uma sobreposição parcial de suas nuvens eletrônicas.<sup>[93]</sup> Essas interações diminuem rapidamente à medida que a distância aumenta, e é possível aplicar um limite para reduzir o custo computacional. As interações de *Coulomb*, por outro lado, são consideradas de longo alcance, mas, para otimizar o desempenho computacional, é importante reconhecer que qualquer função que seja de longo alcance no espaço real se torna de curto alcance no espaço recíproco, ou espaço de Fourier.<sup>[94]</sup>

A parametrização de um campo de força depende de um grande número de parâmetros a serem definidos, isso requer uma otimização cuidadosa para fornecer resultados precisos. Nesse sentido, os parâmetros são ajustados empiricamente a dados experimentais e/ou



dados de cálculos de nível superiores baseadas na mecânica quântica. Outro aspecto muito importante na parametrização de um campo de força é a percepção química, ou seja, é fundamental que o pesquisador entenda o processo pelo qual o software de simulação molecular reconhece a química de uma molécula com o objetivo final de atribuir parâmetros de campo de força apropriados para descrever as ligações, ângulos de ligação, ângulos diedrais e as interações inter e intramolecular dessa molécula.<sup>[95,96]</sup>

No próximo tópico será discutido um pouco sobre o campo de força amber ff14SB e ZAFF utilizados nas simulações de DM realizadas neste trabalho.

### 3.4.1.2 Campo de Força: AMBER ff14SB

dinâmica molecular de todos os átomos é certamente o método de amostragem de simulação de biomoléculas como proteínas mais utilizado. Esse método emprega funções de energia denominados de campos de força que geralmente têm muitos parâmetros ajustáveis, normalmente obtidos por ajustes de dados experimentais ou QM, apresentam formas funcionais semelhantes, contudo diferem significativamente na escolha dos sistemas-modelo e na fonte dos dados de treinamento.<sup>[97]</sup>

Portanto, grupos de pesquisa concentram seus esforços em aprimorar ou desenvolver novos campos de força que representem de maneira mais precisa os sistemas de estudo por meio de modelos. Nesse contexto, Maier, J. A. et al.<sup>[98]</sup> apresentaram o campo de força AMBER ff14SB, comumente utilizado na parametrização de proteínas e amplamente empregado para parametrizar as principais proteínas do SARS-CoV-2 em simulações de DM. Este campo de força foi desenvolvido para melhorar a precisão dos parâmetros relacionados às cadeias laterais e à espinha dorsal das proteínas, superando uma das limitações do campo de força ff99SB. Além disso, foi demonstrado que o ff14SB mantém a estrutura cristalina e as conformações proteicas da lisozima triclinica de maneira mais eficaz do que os campos de força ff99SB, ff14ipq ou CHARMM36, este último também amplamente utilizado em simulações de dinâmica molecular.

Com base nos resultados obtidos no *benchmark* mencionado no artigo, os autores recomendam o uso do campo de força ff14SB para a simulação de proteínas e peptídeos. Além disso, Abriata, L. A.<sup>[99]</sup> demonstrou, por meio de simulações de dinâmica molecular de 10  $\mu$ s com a ubiquitina, que a complexação proteína-proteína ocorre mais rapidamente com o campo de força ff14SB-TIP3P do que com o CHARMM36m-TIP3P. Ambos os campos de força apresentaram uma ligação bastante forte, a ponto de preverem que a ubiquitina

fosse insolúvel. Esses efeitos também se refletem nos gráficos das áreas de superfície total acessíveis ao solvente ao longo do tempo, onde o ff14SB-TIP3P parece favorecer agregados mais compactos, seguido pelo CHARMM36m-TIP3P, depois pelo ff19SB-OPC e, por fim, pelo a99SBdisp-TIP4PD, onde as proteínas permanecem totalmente solúveis e monoméricas. Com base nos resultados apresentados nesses artigos, o campo de força ff14SB, em conjunto com o modelo de água TIP3P, destaca-se como um dos principais campos de força para a parametrização de proteínas.

### 3.4.1.3 Campo de Força: ZAFF

Na PLpro do SARS-CoV-2, encontra-se o dedo de zinco formado pelo íon  $Zn^{2+}$ , ligado a quatro cisteínas. Não é surpresa que íons metálicos possam desempenhar um papel fundamental na função, estrutura e estabilidade das proteínas. O campo de força AMBER ZAFF<sup>[100]</sup> foi desenvolvido para a parametrização de proteínas que, assim como a PLpro, possuem zinco em sua estrutura. Cerca de 47% dos átomos de zinco em estruturas de proteínas são encontrados na forma tetraédrica. O ambiente tetraédrico mais comum para o Zn é o CCCC, ou seja, com quatro cisteínas coordenadas, exatamente o que é observado no domínio do dedo de zinco da PLpro do SARS-CoV-2, seguido por CCCH, CCHH, DHHH, HHHO, HHHX, CCHX e CCHO. Peters, M. B., et al.<sup>[100]</sup> desenvolveram o campo de força para facilitar as simulações de DM da maioria das proteínas que contêm Zn, denominado campo de força de zinco AMBER, ou simplesmente ZAFF, que foi parametrizado utilizando o método DFT-B3LYP com o conjunto de base 6-31G(d).

Muitas simulações de proteínas são realizadas utilizando os campos de força AMBER<sup>[101]</sup> que incluem pacotes de programas desenvolvidos para trabalhar em conjunto, desde a preparação das estruturas, produção das simulações até a análise dos resultados. No nosso caso, tanto o preparo como a análise foi feita utilizando os pacotes disponíveis no AMBER.

## 3.4.2 Descritores de Reatividade Químico Quânticos

Os descritores de reatividade (DR) mais usados são definidos pela Teoria Funcional da Densidade Conceitual (CDFT) (do inglês, Conceptual Density Function Theory) e são descritos como quantidades teóricas utilizadas para descrever e entender a reatividade global e a seletividade com base na densidade eletrônica.<sup>[102,103]</sup> Os DR são obtidos por meio de manipulações matemáticas que fornece uma abordagem matematicamente rigorosa e formalmente exata para entender as mudanças químicas provenientes da estrutura eletrônica.

A partir da CDFT através da equação diferencial fundamental da DFT (Equação 3.4) e suas derivadas, os descritores de reatividade são definidos fornecendo informações quantitativas das estruturas eletrônicas. Essa equação pode ser expressa como:

$$dE = \mu dN + \int \rho(r) dv(r) dr \quad (3.4)$$

em que mostra a relação entre a energia eletrônica ( $E$ ) com o potencial químico eletrônico ( $\mu$ ) do sistema e ainda a densidade eletrônica ( $\rho(r)$ ) e o potencial externo ( $v(r)$ ).<sup>[104]</sup> Os descritores de reatividade podem ser utilizados para calcular as interações eletrônicas e a reatividade química em sistemas químicos em geral, incluindo as macromoléculas.

No estudo com a PLpro e com a *Spike*, utilizaremos o software PRIMoRDiA.<sup>[105]</sup> Na versão 1.25v ([repositório no GitHub](#)), esse software calcula 14 descritores de reatividade globais e 19 locais para sistemas com muitos elétrons, como as proteínas estudadas neste trabalho. Recentemente, alguns estudos obtiveram sucesso com o PRIMoRDiA, incluindo a análise de interação proteína-ligante para o sistema HIV1-protease,<sup>[106]</sup> a interação da ricina A (RTA) com diversos inibidores,<sup>[107]</sup> e uma investigação similar à nossa que examinou o impacto da variação da estrutura eletrônica na reatividade da Mpro dos SARS-CoV-1 e SARS-CoV-2.<sup>[69]</sup>

As funcionalidades do PRIMoRDiA podem ser dividida em três etapas diferentes:

- I. Analisar e armazenar informações moleculares;
- II. Cálculos de Descritores Quânticos Moleculares (DQM);
- III. Geração de arquivos (outputs) que podem ser visualizados no Pymol ou R.

A partir de um protocolo computacional relativamente simples, algumas informações são geradas automaticamente quando o PRIMoRDiA é executado, por exemplo, a densidade eletrônica (Equação 3.5), os orbitais moleculares (Equação 3.6) e potencial eletrostático molecular. Vale ressaltar que um dos pontos-chave desse software é a utilização dos descritores baseados na química quântica em sistemas biológicos tratadas com simulações anteriores.

$$\rho(r) = \sum_i^{occ} |\phi(r)_i|^2 \quad (3.5)$$

$$\phi(r)_i = \sum_i c_i \chi(r) \quad (3.6)$$

Logo, matematicamente os descritores de reatividade locais são determinados por uma função no espaço tridimensional, melhor dizendo, para um dado ponto no espaço,  $r$ , o descritor assume um determinado valor.<sup>[105]</sup> A densidade eletrônica  $\rho$  é uma quantidade local e seu valor é definido através do produto dos orbitais moleculares ocupados  $\phi(r)$ , que por sua vez nada mais é do que combinações lineares de orbitais atômico  $\chi(r)$ , onde os coeficientes  $C_i$  são calculados por métodos quânticos.

De forma geral, o PRIMoRDiA extrai as propriedades eletrônicas dos arquivos de saída de pacotes de métodos baseados na química quântica, calculadas em programas como o MOPAC ou GAUSSIAN, por exemplo. Com essas informações, realiza o tratamento matemático necessário para obter os descritores de reatividade. Esses descritores podem ser classificados em duas categorias: globais, calculados para todo o sistema, e locais, calculados para cada átomo no espaço tridimensional.

Nessa perspectiva, usaremos os descritores de reatividade para explorar características da estrutura eletrônica da PLpro, protease responsável pela replicação viral e das mutações no RBD da proteína *spike*, responsável pela infecção na célula hospedeira. Entre os descritores de reatividade disponível no PRIMoRDiA pretendemos calcular descritores de reatividade global (dureza, potencial de ionização e afinidade eletrônica), além de calcular descritores locais (localização do orbital molecular, eletrofilicidade, nucleofilicidade, multifilicidade e densidade eletrônica).<sup>[105,106]</sup> Os descritores de eletrofilicidade e nucleofilicidade indicam a capacidade de um sistema aceitar e doar carga, respectivamente, fornecendo informações sobre a distribuição da densidade eletrônica. A dureza e moleza locais são grandezas químicas conhecidas que analisam as interações favoráveis entre átomos ou moléculas (interações duro-duro e mole-mole). Esses descritores são importantes para a compreensão de propriedades eletrônicas fundamentais dos sistemas estudados.<sup>[105,106]</sup>

Neste capítulo serão abordadas as principais estratégias empregadas, as etapas desenvolvidas e os programas computacionais utilizados para avaliar as propriedades estruturais das duas proteínas do SARS-CoV-2.

## 4.1 Amostragem das Estruturas Cristalográficas

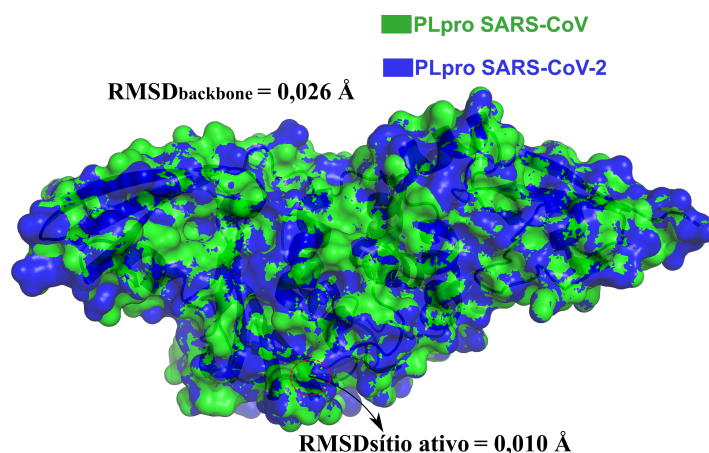
### 4.1.1 Papain-Like Protease (PLpro)

Foram escolhidas duas conformações cristalográficas da PLpro no Protein Data Bank (PDB): uma para o SARS-CoV e outra para o SARS-CoV-2. O cristal escolhido para a SCoV1-PLpro foi identificado com o código PDB: 2FE8, considerado a forma apo da proteína com uma resolução de 1,85 Å.<sup>[108]</sup> Já o cristal selecionado para a SCoV2-PLpro foi identificado com o código PDB: 6WRH e também é considerado a forma apo da proteína, porém com uma mutação no sítio ativo C111S e uma resolução de 1,60 Å.<sup>[109]</sup>

Como a SCoV2-PLpro possui uma mutação no sítio ativo, optou-se por utilizar a sequência FASTA da estrutura 6WRH (utilizada como *template*) para gerar um modelo comparativo da PLpro sem a mutação, utilizando o SWISS-MODEL.<sup>[110]</sup> A mesma metodologia foi aplicada para garantir que ambos os cristais tivessem a mesma quantidade de resíduos de aminoácidos. Ambas as proteínas analisadas apresentaram um RMSD (*Root-Mean-Square Deviation*) entre seus *backbones* de apenas 0,026 Å e um RMSD entre seus respectivos sítios ativos de 0,010 Å.

A escolha dessas duas estruturas teve como objetivo inicial aumentar a diversidade conformacional disponível para as análises de dinâmica molecular. Dado os baixos valores de RMSD entre essas duas proteínas, foi inferido que ambas as estruturas poderiam ser

consideradas como subconjuntos de um mesmo espaço de fase, permitindo, posteriormente, a análise das duas simulações como se pertencessem a um único espaço amostral para fins estatísticos.



**Figura 4.1:** Análise estrutural comparando a proteína SCoV1-PLpro e a SCoV2-PLpro usadas nas simulações de Dinâmica molecular. Os respectivos valores de desvio médio quadrático (RMSD) baixo mostram nenhuma distinção substancial considerando tanto o *backbone* como o sítio ativo das duas proteínas. Fonte: Imagem própria.

#### 4.1.2 Domínio de Ligação ao Receptor (RBD)

A estrutura cristalográfica da glicoproteína *Spike* do SARS-CoV-2 do tipo selvagem foi obtida a partir do PDB: 6M0J. Utilizando o software PyMOL,<sup>[111]</sup> selecionou-se a região que compreende os resíduos 333 a 527, correspondentes ao domínio de ligação ao receptor (RBD). As mutações no RBD foram feitas através de modelagem por homologia realizada no SWISS-MODEL.<sup>[110]</sup> No total, foram realizadas 37 mutações no RBD.

Essas mutações foram rotulados conforme suas características fenotípicas nas variantes e em mutações individuais, classificadas conforme a literatura em quatro categorias, no que diz respeito à infectividade e evasividade ao sistema imune: mutantes que aumentam infectividade e evasividade ao sistema imune foram rotulados (+inf/+ev), mutantes que aumentam somente infectividade foram rotulados (+inf/-ev), mutantes que aumentam apenas evasividade foram rotulados (-inf/+ev), e as mutantes neutras que foram rotuladas (-inf/-ev) como mostra **Tabela 4.1**. Essa classificação será mais detalhada na seção de resultados. Ressalta-se que essa categorização segue a classificação presente na literatura citada.

**Tabela 4.1:** Lista das variantes e rótulos dos mutantes do RBD da *Spike* do SARS-CoV-2 utilizados no presente estudo

-inf/-ev			+inf/+ev			+inf/-ev			-inf/+ev		
Variante	Linhagem	Mutação	Variante	Linhagem	Mutação	Variante	Linhagem	Mutação	Variante	Linhagem	Mutação
-	-	S:Wild type <sup>[112]</sup>	Epsilon	linhagem B.1.427/9	S:L452R <sup>[1,113]</sup>	-	-	S:T478K <sup>[114]</sup>	-	-	S:E484K <sup>[22,115,116]</sup>
-	-	S:V445L <sup>[112]</sup>	-	-	S:E484Q <sup>[22,115,116]</sup>	-	-	S:N501Y <sup>[22,115,116]</sup>	-	-	S:K417T <sup>[22,115,116]</sup>
-	-	S:Q498L <sup>[112]</sup>	Kappa	linhagem B.1.617.1	S:L452R+E484Q <sup>[117,118]</sup>	-	-	S:S477N <sup>[22,115,116]</sup>	-	-	S:K417N <sup>[22,115,116]</sup>
-	-	S:S494Q <sup>[112]</sup>	Delta	linhagem B.1.617.2	S:L452R+T478K <sup>[119,120]</sup>	-	-	S:N501S <sup>[22,115,116]</sup>	Lambda	linhagem C.37	S:L452Q+F490S <sup>[22,121]</sup>
-	-	S:Q498V <sup>[112]</sup>	Gamma	linhagem P.1	S:K417T+E484K+N501Y <sup>[122-124]</sup>	-	-	S:V367F <sup>[22,115,116]</sup>	-	-	S:R346K <sup>[22,115,116]</sup>
-	-	S:N439R <sup>[112]</sup>	Theta	linhagem P.3	S:E484K+N501Y <sup>[125,126]</sup>	-	-	S:N354D <sup>[22,115,116]</sup>	-	-	S:F490S <sup>[22,115,116]</sup>
-	-	S:Y489F <sup>[112]</sup>	Mu	linhagem B.1.621	S:R346K+E484K+N501Y <sup>[127,128]</sup>				-	-	S:G446V <sup>[22,115,116]</sup>
-	-	S:S494K <sup>[112]</sup>	Iota	linhagem B.1.526	S:S477N+E484K <sup>[129,130]</sup>						S:N439K <sup>[22,115,116]</sup>
-	-	S:Q498I <sup>[112]</sup>	Beta	linhagem B.1.351	S:K417N+E484K+N501Y <sup>[99,131]</sup>				-	-	S:F490L <sup>[22,115,116]</sup>
-	-	S:K417Y <sup>[112]</sup>	-	-	S:L452Q <sup>[22,115,116]</sup>				-	-	S:Y508H <sup>[22,115,116]</sup>
-	-	S:Y505F <sup>[112]</sup>									

## 4.2 Estado de Protonação e Parametrização

O procedimento de protonação e parametrização foi aplicado a ambas as proteínas estudadas neste trabalho. Foi realizado o cálculo dos estados de protonação para todos os resíduos em ambos os sistemas, considerando o pH fisiológico padrão do sangue (7,4), uma salinidade de 0,15 M e permissividades dielétricas relativas internas e externas de 10 e 80, respectivamente. Essa estimativa foi realizada por meio do servidor web H++,<sup>[132]</sup> que é um sistema automatizado capaz de calcular os valores de pK de grupos ionizáveis em macromoléculas e adicionar átomos de hidrogênio ausentes de acordo com o pH especificado do ambiente.

Os sistemas foram parametrizados utilizando a ferramenta tleap do AmberTools21.<sup>[133]</sup> Para a solvatação, os sistemas foram colocados em uma caixa cúbica com uma distância mínima da borda de 12 Å. As porções de água, proteína e glicídica dos sistemas foram descritas respectivamente com os campos de força TIP3P,<sup>[134]</sup> ff14SB<sup>[98]</sup> e GLYCAM-06<sup>[135]</sup> AMBER para a S:RBD, PLpro do SARS-CoV e SARS-CoV-2 sem o dedo de zinco.

No caso da PLpro do SARS-CoV e SARS-CoV-2 com o dedo de zinco, os sistemas também foram solvatados em uma caixa cúbica com uma distância mínima da borda de 12 Å. As porções de água, proteína, glicídica e do dedo de zinco dos sistemas foram descritas respectivamente com os campos de força TIP3P,<sup>[134]</sup> ff14SB,<sup>[98]</sup> GLYCAM-06<sup>[135]</sup> e o ZAFF<sup>[100]</sup> AMBER. Para atingir a salinidade fisiológica de 0,15 M, íons Na<sup>+</sup> e Cl<sup>-</sup> foram adicionados aos sistemas. Os arquivos de coordenadas e topologia são escritos com o Tleap.<sup>[133]</sup>

Os sistemas finais, incluindo moléculas de água e contra-íons, resultaram em aproxima-

damente 50.000 átomos para a S:RBD e 70.000 átomos para a PLpro.

### 4.3 Protocolo de Dinâmica Molecular

Após a solvatação, as etapas de minimização, equilíbrio e simulação DM produtiva adequada para ambos os sistemas foram executadas usando verbos-CUDA NAMD 2.13<sup>[136]</sup> utilizando os campos de força AMBER. Todas as simulações foram realizadas com o ensemble NPT, com temperatura e pressão mantidas constantes respectivamente em 310 K e 1 atm por um termostato de Langevin e um pistão de Langevin. Foram utilizadas condições periódicas de contorno e corte de interações eletrostáticas 10 Å para as interações não-ligadas. As interações eletrostáticas de longo alcance foram calculadas por Particle Mesh Ewald (PME).

O protocolo de relaxamento em condição NPT foi aplicado para todos os sistemas, antes da simulação DM produtiva adequada, foi realizado um protocolo de equilíbrio em várias etapas. Este protocolo consistiu de:

- (i) 500 passos de minimização com restrições harmônicas nos átomos de proteína;
  - (ii) 500 passos de minimização sem restrições;
  - (iii) 300 ps de equilíbrio com restrições harmônicas nos átomos de proteína;
  - (iv) 300 ps de equilíbrio com restrições harmônicas nos átomos do backbone;
  - (v) 300 ps de equilíbrio sem restrições;
  - (vi) 1 ns de simulação DM pré-produtiva com velocidades reinicializadas, sem restrições.
- Nas simulações com a PLpro com dedo de zinco foi adicionado uma passo a mais no protocolo;
- (vii) 300 ps de equilíbrio com restrições harmônicas nos átomos de Zn.

Por fim, três simulações independentes de DM produtivas de 100 ns (resultando em 25000 *frames* que foram utilizados para análises posteriores) foram executadas para cada sistema (S:RBD, SCoV1-PLpro e SCoV2-PLpro sem zinco e SCoV1-PLpro e SCoV2-PLpro com zinco) sobre as coordenadas finais do relaxamento que foram realizadas também sobre condições NPT, mantendo os conjuntos gerais do sistema conforme mencionado acima, no supercomputador SDumont, no Laboratório Nacional de Computação Científica (LNCC).



## 4.4 Análises Pós-Produção das Trajetórias de DM

### 4.4.1 Papain-Like Protease (PLpro)

A análise de flutuação quadrática média (RMSF, do inglês *Root Mean Square Fluctuation*) por resíduo em relação ao *frame* médio para os carbonos C $\alpha$  foi realizada utilizando o plugin cpptraj<sup>[137]</sup> do AmberTools 21. Em ambos os sistemas, realizamos o alinhamento pelos dois subdomínios Palm e Thumb, considerando que são as partes menos móveis das duas proteínas, abrangendo os resíduos (62-179, 241-263 e 273-305). O desvio quadrático médio (RMSD, do inglês *Root Mean Square Deviation*) também foi alinhado seguindo a mesma metodologia utilizada para o cálculo de RMSF. Para a visualização gráfica dos resultados de RMSD e RMSF, utilizamos dois scripts em Python de nossa autoria, denominados RMSD.py e RMSF.py (anexados no apêndice).

Posteriormente, empregamos um *script* denominado make-backbone.sh para converter todas as trajetórias em formato *backbone*. Logo após, utilizamos um *script* fornecido pelo Dr. Klaus R. Liedl, da University of Innsbruck, Áustria, para plotar o RMSD-2D e verificar a convergência ao longo das diferentes simulações para cada sistema distinto e gerar os *clusters* utilizando o algoritmo de *hierarchical averagelinkage* com um *cutoff* de 3 Å. Nesta etapa, realizamos o alinhamento pelas duas partes menos móveis e pelo *backbone* da proteína. Essas estruturas foram utilizadas nas análises de componente principal (PCA, do inglês *Principal Component Analysis*).

#### 4.4.1.1 Reconstrução dos perfis de energia livre e PCA no estudo da Papain-Like protease (PLpro)

Em nossos estudos foram realizados a reconstrução dos perfis de energia livres para todos os sistemas em estudo (SCoV1-PLpro e SCoV2-PLpro com e sem o dedo de zinco) através do kit de ferramentas PyReweight2D desenvolvido em Python.<sup>[138,139]</sup> Consideraram-se as projeções nos três primeiros autovetores, PC1, PC2 e PC3. Com o intuito de melhorar a visualização de deslocamento das diferenças estruturais em cada sistemas, projetou-se a projeção média de cada sistema. Todas as simulações usadas na dinâmica molecular foram projetadas nos perfis, para isso fez-se o cluster hierárquico de todos os sistemas juntos para selecionar conformações mais populosas. Uma segunda reconstrução dos perfis de energia livre foi realizada sobre os quatro conjuntos de trajetórias: trajetória da SCoV1-PLpro sem o dedo de zinco, trajetória da SCoV1-PLpro com o dedo de zinco, trajetória da SCoV2-PLpro

sem o dedo de zinco e trajetória da SCoV2-PLpro com o dedo de zinco. Projetou-se sobre essa reconstrução as estruturas com projeções médias de cada sistema em busca de um padrão de mobilidade dentro do espaço PC1 *vs.* PC2, PC1 *vs.* PC3 e PC2 *vs.* PC3.

#### 4.4.2 Domínio de Ligação ao Receptor (RBD)

A análise de RMSF por resíduo em relação ao *frame* médio para os carbonos C $\alpha$  da S:RBD foi realizada usando o plug-in cpptraj.<sup>[137]</sup> Para ambos os sistemas, foram realizados dois conjuntos de análises. Primeiro, alinhamos todo o RBD considerando os resíduos menos móveis (333-437, 454-456, 492-494 e 509-526). Em seguida, analisamos as flutuações para todo o domínio. Em seguida, alinhamos e analisamos apenas os carbonos alfas dos resíduos correspondentes ao motivo de ligação ao receptor (RBM) (438 a 508), a fim de observar seus movimentos internos.

### 4.5 Cálculo de descritores de reatividade

#### 4.5.1 Papain-Like Protease (PLpro)

Com o intuito de investigar características eletrônicas das estruturas, foram realizados cálculos de descritores de reatividade utilizando o software PRIMoRDIA.<sup>[105]</sup> Este software permite o cálculo de propriedades eletrônicas de macromoléculas graças a implementação de diversos descritores de reatividade globais e locais.

Ao todo foram selecionadas 37 estruturas referentes aos clusters mais representativos das simulações de DM para as duas proteínas utilizadas no estudo.

Para a obtenção das estruturas eletrônicas utilizadas no PRIMoRDIA, foram realizados cálculos single-point com as palavras-chave "PM7 PL 1SCF TIMES MOZYME PDB ALLVECS VECTOR AUX LARGE", utilizando o pacote MOPAC.<sup>[140]</sup> O hamiltoniano semi-empírico PM7<sup>[141]</sup> foi empregado em conjunto com o algoritmo de escalonamento linear MOZYME<sup>[142]</sup> e o modelo de solvente implícito COSMO sem a presença dos contra íons. O zinco na PLpro apresenta carga 2+ e multiplicidade singlete.

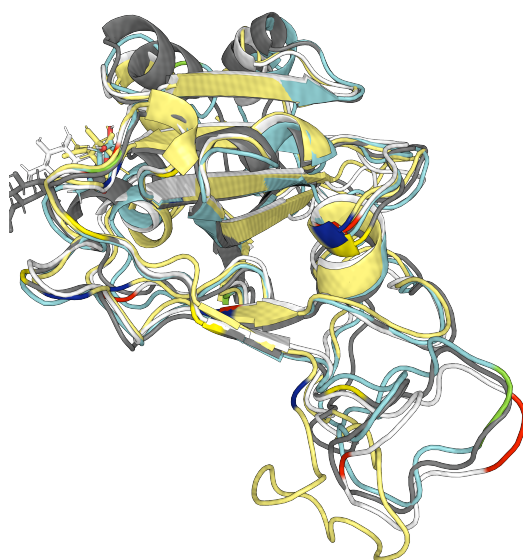
Para investigar características da estrutura eletrônica, a avaliação de descritores de reatividade local foi direcionada para os descritores: nucleofilicidade, eletrofilicidade, netfilicidade, moleza e dureza locais, suscetibilidade radical, localização do orbital molecular, multifilicidade e densidade eletrônica. Todos esses avaliados em relação aos resíduos dos domínios da PLpro com e sem zinco.

### 4.5.2 Domínio de Ligação ao Receptor (RBD)

Com o objetivo de analisar o impacto das mutações nas características eletrônicas do RBD da proteína *Spike*, utilizou-se o software PRIMoRDIA<sup>[105]</sup> para calcular os descritores de reatividade. Inicialmente, selecionaram-se os 2 clusters mais representativos das mutações (+inf/+ev), (+inf/-ev), (-inf/+ev) e (-inf/-ev), totalizando 76 estruturas alinhadas pela região dos resíduos 360-374 e 76 alinhadas pelo RBM.

Para selecionar a mesma conformação para cada fenótipo, primeiro selecionaram-se as conformações mais populosas de cada tipo de mutação da PCA: (+inf/+ev), (+inf/-ev), (-inf/+ev) e (-inf/-ev). Em seguida, as estruturas foram convertidas para a sequência FASTA, e as mutações foram realizadas através de modelos por homologia no SWISS-MODEL,<sup>[110]</sup> a fim de obter a mesma conformação para cada grupo de mutações (+inf/+ev), (+inf/-ev), (-inf/+ev) e (-inf/-ev).

Para a obtenção das estruturas eletrônicas utilizadas no PRIMoRDIA, seguiu-se a mesma metodologia empregada na PLpro, foram realizados cálculos single-point com as palavras-chave "PM7 PL 1SCF TIMES MOZYME PDB ALLVECS VECTOR AUX LARGE", utilizando o pacote MOPAC.<sup>[140]</sup> O hamiltoniano semiempírico PM7<sup>[141]</sup> foi empregado em conjunto com o algoritmo de escalonamento linear MOZYME<sup>[142]</sup> e o modelo de solvente implícito COSMO sem a presença dos contra íons.



**Figura 4.2:** Conformações do Domínio de Ligação ao Receptor: +inf/+ev em branco, +inf/-ev em azul, -inf/+ev em amarelo e -inf/-ev em cinza. As cores verde, amarelo, azul e vermelho correspondem às mutações. **Fonte:** Imagem própria.

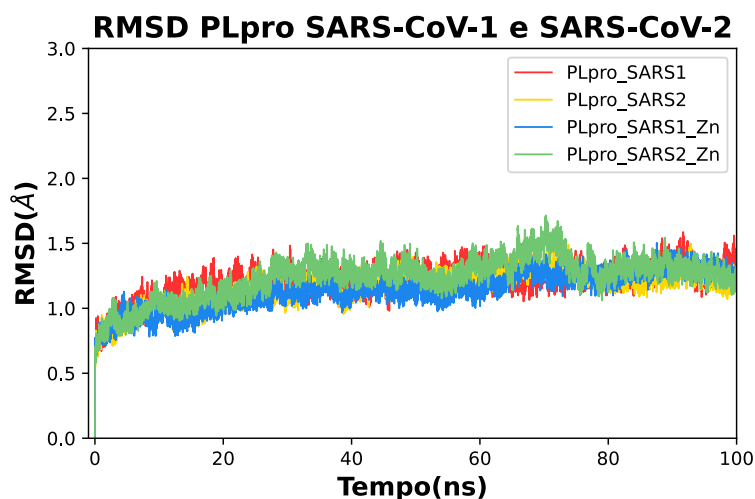
## 5.1 Estudo Conformacional das proteínas do SARS-CoV-2

Nesta seção serão apresentados os resultados referentes ao estudo conformacional da papain-like protease, responsável pela replicação do vírus e do Domínio de Ligação ao receptor da Proteína *Spike* (RBD), responsável pela infecção da ACE2, para as diferentes mutações e variantes.

### 5.1.1 *Papain-Like Protease (PLpro)*

A princípio, pretendemos nesta seção investigar amostras do espaço conformacional da SCoV1-PLpro e SCoV2-PLpro com foco nas alterações conformacionais do domínio do tipo ubiquitina, alça BL2 e domínio do dedo de zinco, uma vez que esse domínio se ligar a um íon zinco essencial para a atividade estrutural da protease, dessa forma sendo um alvo importante para o desenvolvimento de um inibidor. Por conseguinte, existem compostos que modificam as cisteínas coordenadas com o íon zinco, agindo assim como um quelante de zinco.<sup>[143,144]</sup>

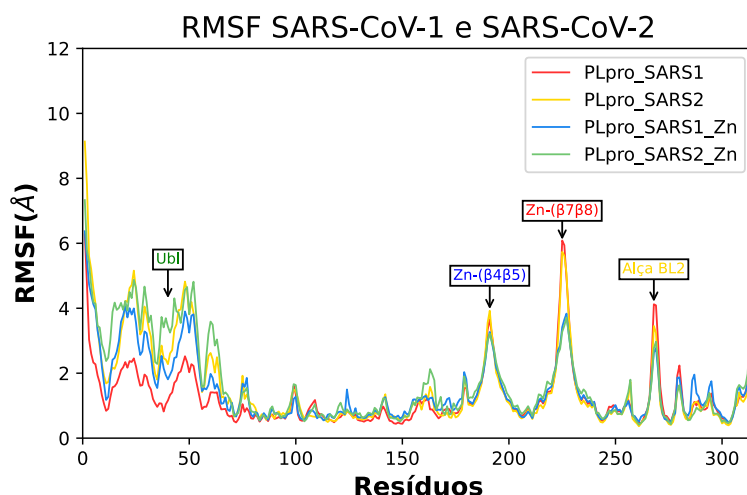
Os cálculos de RMSD foram realizados alinhando essas duas proteínas pelos domínios thumb e palm usando todos os átomos do  $C\alpha$ , usando o primeiro frame como referência. Essa análise geralmente fornece informações importantes sobre a estabilidade dos domínios das duas proteínas, o resultado mais detalhado é mostrado na Figura 5.1, na qual traçamos a evolução do RMSD *vs.* tempo (ns) para às duas proteínas em estudo.



**Figura 5.1:** RMSD para a SCoV1-PLpro e SCoV2-PLpro com e sem zinco *vs.* tempo. Fonte: Própria.

O gráfico RMSD indica que tanto a SCoV1-PLpro e SCoV2-PLpro sem zinco e com zinco atingiram estabilidade conformacional ao longo da simulação, onde é possível notar que as conformações da trajetória atingiram um valor relativamente constante. No entanto, observou-se um pequeno aumento nos desvios, aproximadamente de 0,7 Å para a SCoV2-PLpro, entre 70 e 80 ns. Os valores médios de RMSD para as estruturas da SCoV1-PLpro e SCoV2-PLpro sem zinco variaram de  $0,7 \pm 1,4$  Å e  $0,7 \pm 1,2$  Å, respectivamente. Já os valores médios de RMSD para as estruturas da SCoV1-PLpro e SCoV2-PLpro com zinco variaram de  $0,6 \pm 1,1$  Å e  $0,8 \pm 1,5$  Å. Como é possível observar, a SCoV2-PLpro com zinco apresentou um RMSD levemente superior as demais. Contudo, esses resultados confirmaram que tanto a SCoV1-PLpro e SCoV2-PLpro com e sem zinco foram equilibradas.

Para uma maior estabilidade conformacional termodinâmica, são realizadas as análises de RMSF a flutuação de cada um dos resíduos de aminoácidos da PLpro. Essas análises podem ajudar a identificar resíduos ou regiões da PLpro que apresentam maior flexibilidade ou mobilidade, o que pode influenciar sua atividade catalítica e interação com substratos ou inibidores, impactando diretamente a eficiência da replicação viral. A Figura 5.2 mostra a flutuação quadrática média calculada alinhando as proteínas pelos domínios thumb e palm usando todos os átomos do C $\alpha$ . Os resultados mostram flutuações relativamente altas nas regiões de aminoácidos: 1-63, 182-200, 223-235 e 306-315, regiões distantes do sítio de ligação. Também é possível perceber uma flutuação alta na alça BL2 localizada no sítio de ligação, formada pelos resíduos 265-272.



**Figura 5.2:** RMSF da SCoV1-PLpro e SCoV2-PLpro com e sem zinco. Resíduos altamente flexíveis 1-62 (Ubl), 182-200 e 223-235 (Dedo de Zn) e 265-272 (Sítio de ligação). Fonte: Própria.

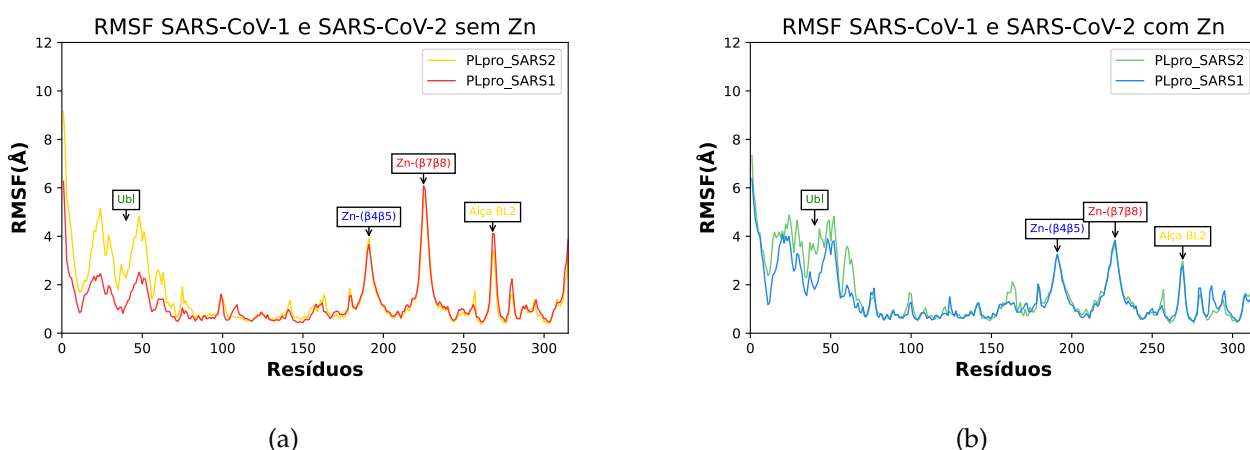
Outra região da proteína que iremos dar ênfase será o Ubl, como mostra a Figura 5.2. O gráfico expõe grandes flutuações no Ubl e no domínio do dedo de zinco, além disso, a análise de RMSF mostra uma flutuação em torno de 4 Å na região dos aminoácidos T265-H272, região da alça BL2.

Aqui, para apresentar e discutir nossos resultados de DM sobre tais aspectos, utilizaremos a nomenclatura de Gao, X. et al (2021)<sup>[27]</sup> para a PLpro, Figura 5.3, onde podemos observar que o domínio do dedo de zinco é formado por 4 fitas  $\beta$ . Logo, para melhorar nossa compreensão, os dedos de zinco compreendidos entre os resíduos 182-200, formados por duas fitas  $\beta 4$  e  $\beta 5$  e 224-238, também formados por duas fitas  $\beta 6$  e  $\beta 7$  foram aqui denominados como o motivo Zn-( $\beta 4\beta 5$ ) e Zn-( $\beta 6\beta 7$ ), nomenclatura utilizada para discutir os resultados para essa região específica da proteína.



6 Å no domínio dedo de zinco da SCoV1-PLpro e SCoV2-PLpro sem zinco, o que já era esperado, uma vez que, o zinco liga-se as 4 cisteínas em um arranjo tetraédrico, diminuindo a flutuação do dedo de zinco tanto para a SCoV1-PLpro como para a SCoV2-PLpro como mostra a Figura 5.4(b).

É interessante notar que existe uma flutuação maior no Ubl na SCoV2-PLpro que apresentou uma flutuação acima de 4 Å, enquanto, na SCoV1-PLpro com zinco a flutuação foi inferior a 4 Å no geral. Esses resultados mostram que o domínio do dedo de zinco tem influência no aumento da flutuação dos resíduos de aminoácidos do Ubl. Esse resultado confirma os resultados publicados no trabalho de Krishna, S. S. et al.<sup>[149]</sup> Os autores afirmam que o domínio do dedo de zinco desempenha um papel fundamental e importante para a estrutura da proteína.



**Figura 5.4:** (a)RMSF para a SCoV1-PLpro e SCoV2-PLpro sem zinco. (b). RMSF para a SCoV1-PLpro e SCoV2-PLpro com zinco

Gráficos de RMSF foram plotados para sondar as regiões nas quais as flutuações de resíduos foram mais altamente alocadas ao longo da simulação. Os resultados, mostram que as regiões mais afetadas são o Ubl, domínio do dedo de zinco e alça BL2. Logo, para sondar o equilíbrio e a convergência da amostragem para as diferentes simulações e sistemas de MD, construímos gráficos bidimensionais de desvio quadrático médio (2D-RMSD) para verificar qual das regiões de alta mobilidade descreve melhor as distinções na amostragem conformacional entre a SCoV1-PLpro e SCoV2-PLpro com e sem zinco, Figura 5.5.

Na Fig. 5.5(a), podemos notar que a amostragem conformacional do backbone da SCoV1-PLpro e SCoV2-PLpro com ou sem zinco não sofreram mudanças significativas em suas estruturas. Contudo, é possível observar que em especial a PLpros com zinco sofreram pequenas flutuações com valores de rmsd entre  $2 \pm 4$  Å. Além disso, a amostragem con-



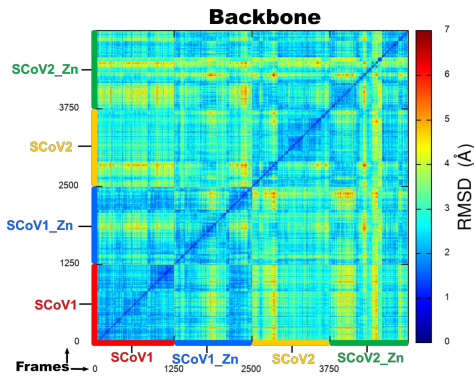
formacional da proteína sugere que nas proteínas formadas com o domínio do dedo de zinco formado pelas duas fitas betas  $\beta 4$  e  $\beta 5$  está passando por pequenas mudanças de conformação entre os valores 4-7 Å o que pode influenciar na estabilidade do sistema, como é possível observar na Fig. 5.5(c).

A alça BL2 foi outra região que apresentou pequenas regiões com um RMSD-2D alto em torno de 7 Å para as PLpros sem zinco, onde podemos notar diferenças substanciais (áreas amareladas e vermelhas) no espaço conformacional na região da alça BL2. Isso indica que zinco pode influenciar na conformação e estabilidade dinâmica do sítio de ligação, ou seja, quando temos o zinco no domínio do dedo a flutuação da alça BL2 é menor e consequente vão ocorrer poucas mudanças conformacionais no sítio de ligação da proteína.

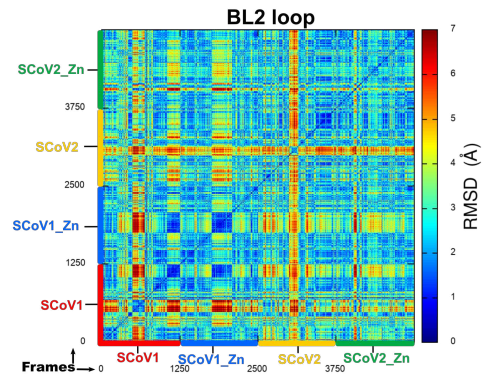
A região correspondente ao Ubl (Figura 5.5(e)) demonstrou uma notável divergência conformacional da proteína, com maior mobilidade para SCoV2-PLpro com ou sem zinco, com valores de RMSD-2D de aproximadamente 7 Å. Na Figura 5.5(f), que representa o número de clusters versus frames, observamos uma quantidade maior de clusters para a SCoV2-PLpro. Esse resultado mostra que a PLpro do SARS-CoV-2 ganhou mobilidade no Ubl, sobretudo, quando adicionamos zinco no dedo de zinco dessa proteína, mostrando um alosterismo entre o domínio do dedo de zinco e Ubl o que pode alterar a atividade enzimática dessa protease.

Este resultado está em concordância com as descobertas experimentais de Tan et al.<sup>[150]</sup> Nesse estudo, os autores demonstram como os inibidores podem desempenhar um papel crucial na conformação do domínio do dedo de zinco, que, por sua vez, desempenha um papel essencial na integridade estrutural e, consequentemente, na atividade enzimática da PLpro, bem como nos possíveis efeitos dos inibidores da PLpro em DUBs humanas. De fato, tanto a SCoV1-PLpro quanto a SCoV2-PLpro demonstraram que a remoção do zinco reduz o seu papel estrutural, resultando em uma menor distinção conformacional ou dinâmica na região do Ubl. Isso sugere que o equilíbrio conformacional e dinâmico nessa região, no contexto da PLpro, está intimamente ligado ao papel estrutural do domínio do dedo de zinco dentro desta proteína.

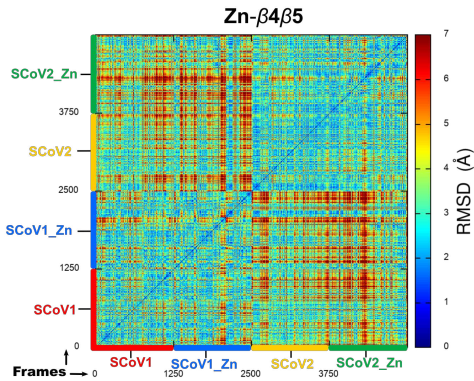
Assim, independentemente do vírus, seja o SARS-CoV-1 ou o SARS-CoV-2, onde a PLpro desempenha um papel crucial na replicação viral, é de se esperar que isso resulte em alterações significativas no comportamento do Ubl e do domínio do dedo de zinco. Essa expectativa está em consonância com a distinção constante no comportamento observado entre a SCoV1-PLpro e a SCoV2-PLpro sem zinco, quando comparadas à SCoV1-PLpro e a SCoV2-PLpro com zinco em nossas pesquisas.



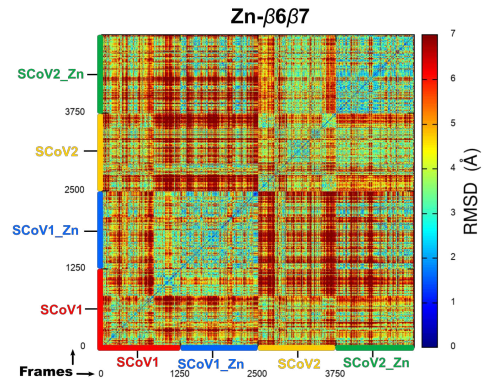
(a)



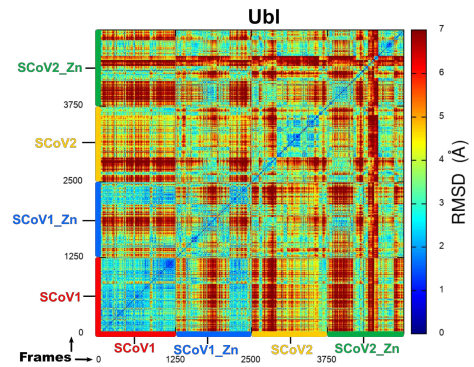
(b)



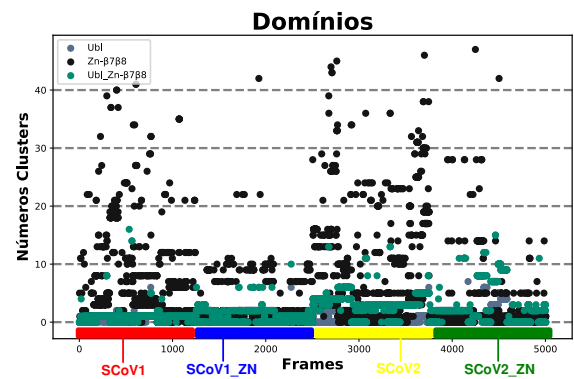
(c)



(d)



(e)



(f)

**Figura 5.5:** Gráfico RMSD bidimensional para as doze trajetórias combinadas (100 ns cada uma) para cada um dos sistemas da PLpro (a) Backbone da proteína, (b) BL2 loop, (c) Motivo Zn- $\beta 4\beta 5$ , (d) Zn- $\beta 6\beta 7$ , (e) domínio Ubl e (f) Número de Cluster vs. frame plot do Domínio Ubl, Motivo Zn- $\beta 6\beta 7$  e Ubl/Zn- $\beta 6\beta 7$ .

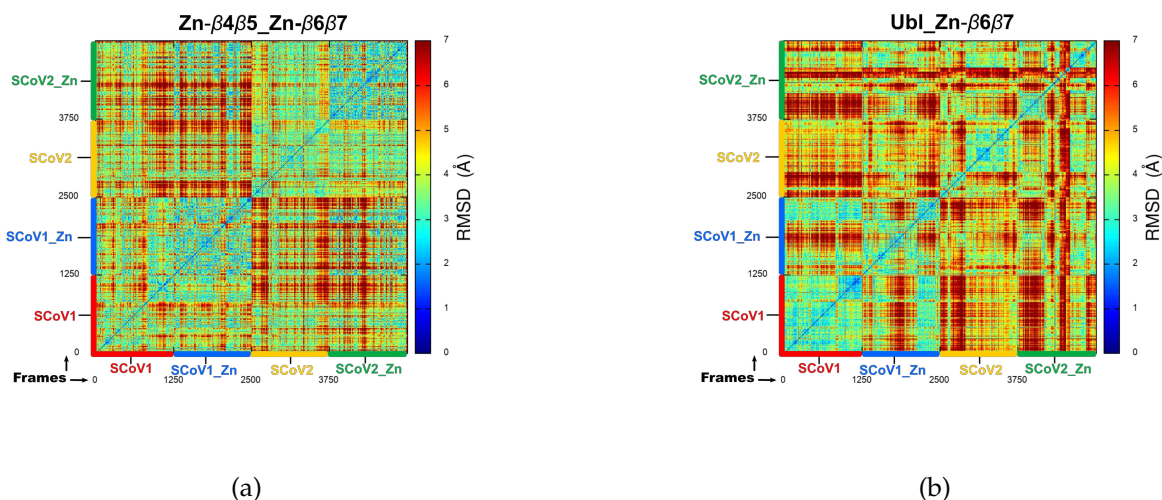
Na região do motivo Zn-( $\beta 6\beta 7$ ), representada na Figura 5.5(d) e 5.5(f), podemos observar mudanças conformacionais variando entre 2 e 7 Å para a PLpro dos SCoV<sub>1</sub> sem o zinco. Assim, a protease PLpro parece reverter o espaço da PLpro dos SCoV<sub>1</sub> sem o zinco na região do Zn-( $\beta 6\beta 7$ ), gerando um comportamento estrutural dinâmico. Além disso, é evidente na Figura 5.5(d) que os valores do RMSD para a SCoV<sub>1</sub>-PLpro e SCoV<sub>2</sub>-PLpro com zinco são menores, o que indica que o zinco contribui para estabilizar as fitas:  $\beta 4$ ,  $\beta 5$ ,  $\beta 6$  e  $\beta 7$ . O alosterismo inter-domínios pode explicar essa maior flutuação do Ubl, influenciada pela estabilização do domínio do dedo de zinco, conforme observado nos gráficos de RMSF. Essa comunicação entre os domínios pode estar relacionada à transmissão de sinais em vias de sinalização celular das células de defesa do organismo.

Além das análises de RMSD-2D individual, realizamos uma investigação focada nos resíduos que compõem o domínio do dedo de zinco e o Ubl para identificar a região que mais significativamente contribuiu para o aumento da flexibilidade na PLpro. Também buscamos identificar possíveis distinções comparáveis entre a região do domínio do dedo de zinco e do Ubl, conforme ilustrado na Figura 5.6.

Dessa forma, nossa análise dinâmica e conformacional indica que tanto o domínio do dedo de zinco quanto o Ubl desempenham papéis fundamentais na compreensão da influência do íon bivalente  $Zn^{2+}$  na estabilidade do complexo e, por conseguinte, na atividade da Papain-Like protease. Essa conclusão se alinha com a literatura, citando trabalhos de Ma, C. et al.,<sup>[151]</sup> que destacam esses dois domínios como áreas altamente promissoras para o desenvolvimento de inibidores destinados ao combate do SARS-CoV-2.

O resultado do 2D-RMSD para essas duas regiões que apresentaram maiores mudanças conformacionais ao longo da simulação de DM mostra que a região do domínio do dedo de zinco sofre mudança conformacional e dinâmica significativa semelhante ao observado para o Zn-( $\beta 6\beta 7$ ), o que pode sugerir que essa região pode influenciar mais na mudança conformacional do domínio. Por outro lado, o Zn-( $\beta 4\beta 5$ ) tem uma contribuição maior no grau de liberdade do domínio, “estabilizando” mais o domínio do dedo de zinco, Figura 5.6(a).

Já, quando fizemos as análises considerando os resíduos do Ubl e do Zn-( $\beta 6\beta 7$ ) juntos, Figura 5.6(b), observamos que essa duas regiões sofre mudanças maiores para a PLpro do SARS-CoV<sub>1</sub> com Zn. Uma análise mais aprofundada pode agora ser feita nessa região específica, como, por exemplo, uma análise de componentes principais, pois isso é importante para compreendemos as mudanças conformacionais e estruturais dessa protease fundamental para a replicação do vírus SARS-CoV-2.



**Figura 5.6:** RMSD-2D para as duas regiões que apresentaram maiores mudanças conformacionais na PLpro. (a) região do domínio do dedo de zinco formado pelo motivo Zn- $\beta 4\beta 5$  e Zn- $\beta 6\beta 7$ . (b) Região formada pelo Ubl com o motivo Zn- $\beta 6\beta 7$ .

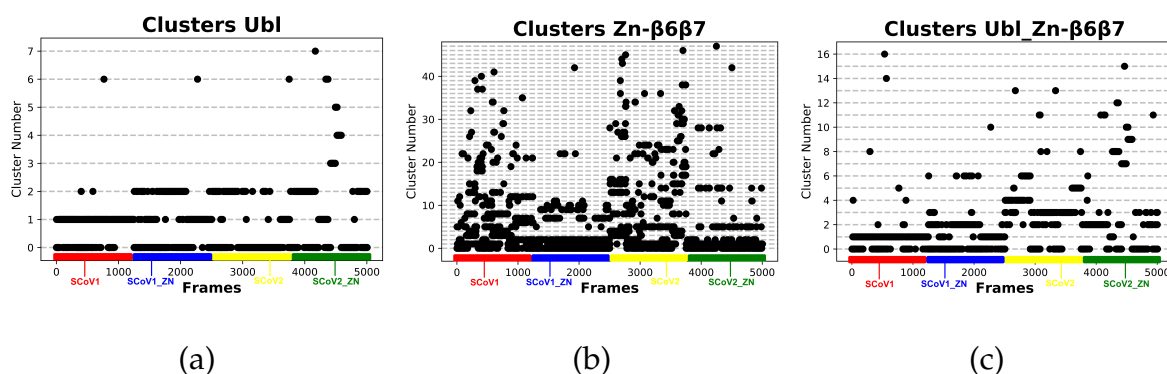
#### 5.1.1.1 Análises do Comportamento Conformacional da *Papain-Like Protease* por PCA

Além das análises convencionais realizadas (RMSD, RMSF, 2D-RMSD), apresentamos também dados da mobilidade dos sistemas proteicos por meio da PCA, essas análises foram feitas através da reconstrução dos perfis de energia livre para todos os sistemas (SCoV1-PLpro e SCoV2-PLpro com e sem zinco) com o kit de ferramentas PyReweighting2D.

A clusterização hierárquica de todas as conformações do Ubl acessado por todas as conformações das proteínas em nossos estudos de DM, retornaram oito agrupamentos estruturais com representatividade maior ou igual a 5%: c0-7, figura 5.7(a). Notamos através da figura 5.7(a) um número de *clusters* maior para SCoV2-PLpro com zinco do que para as demais. Nosso foco no domínio do dedo de zinco será no motivo Zn-( $\beta 6\beta 7$ ), pois os resultados encontrados até aqui indicam que esse motivo pode influenciar na estabilidade conformacional da PLpro.

A clusterização hierárquica para o Zn-( $\beta 6\beta 7$ ) resultou em quarenta e oito agrupamentos estruturais, cada um com uma representatividade de pelo menos 5%: observa-se um número maior de *clusters* para PLpro sem zinco, especialmente para a SCoV2-PLpro sem zinco, conforme ilustrado na Figura 5.7(b). Por último, a clusterização hierárquica de todas as conformações do Ubl em conjunto com o Zn-( $\beta 6\beta 7$ ) resultou em dezessete agrupamentos estruturais, com representatividade igual ao demais. Nesse contexto, podemos observar um maior espaço conformacional para SCoV2-PLpro, como ilustrado na Figura 5.7(c).





**Figura 5.7:** Número de Cluster vs. frame plot. (a) Domínio Ubl, (b) Motivo Zn- $\beta 6\beta 7$  e (c) Ubl/Zn- $\beta 6\beta 7$ .

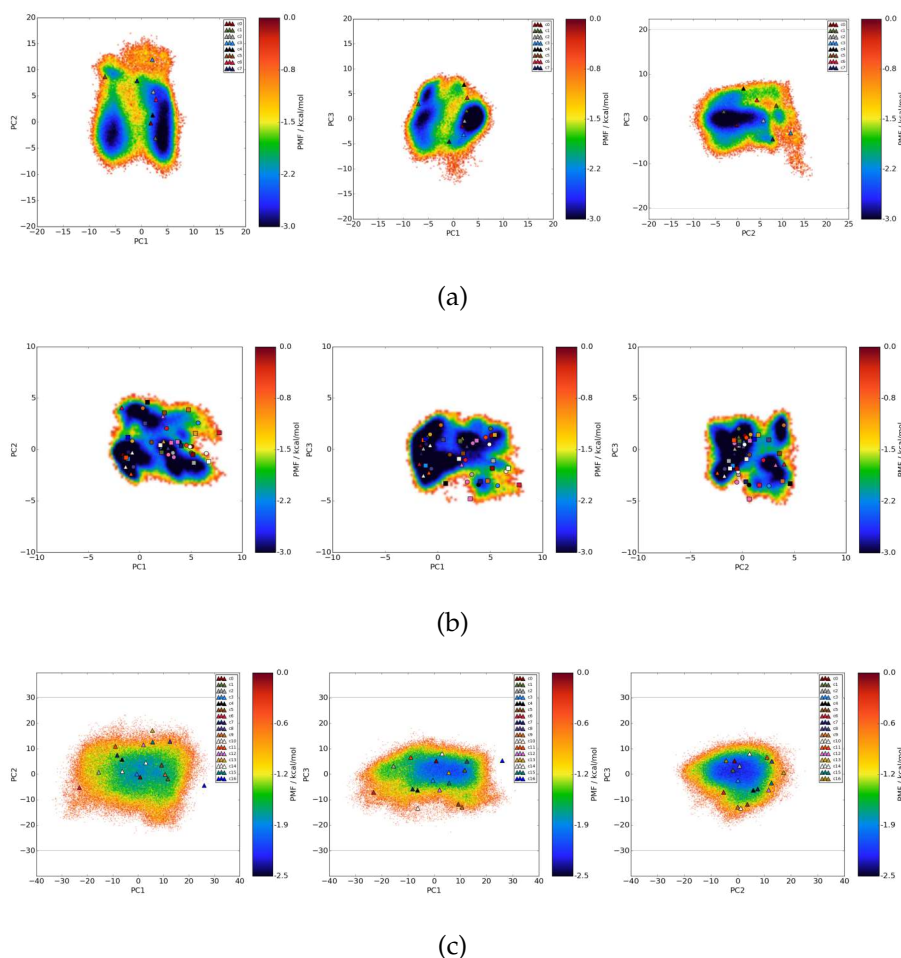
A superposição dos confórmeros representativos dos clusters nos perfis paisagem de energia livre (do inglês, *Free Energy Landscape*, FEL) ao redor do espaço representado pelos autovetores de PCA (PC1 *vs.* PC2, PC1 *vs.* PC3 e PC2 *vs.* PC3) permite uma melhor compreensão e fornecer diretrizes estruturais para entender a natureza das semelhanças ou discrepâncias conformacionais e dinâmicas do domínio dedo de zinco e do Ubl na estrutura da PLpro. Essas sobreposições são mostradas na Figura 5.8.

Enquanto o Ubl mais o domínio do dedo de zinco possui um espaço conformacional restrito (Figura 5.8(c)). Contudo, o Ubl e o domínio do dedo de zinco separado adotam um conjunto de outros mínimos e conformações e, os perfis de FEL mostram que todos eles têm seu espaço conformacional expandido em relação ao essas duas regiões da proteína juntas, em menor ou maior extensão (Figura 5.8(a)-5.8(b)). Tal expansão resulta em uma superposição de seu FEL com todos os agrupamentos representativos (*clusters*) observados para cada região da PLpro em estudos.

É fundamental destacar que a expansão do espaço FEL pela SCoV1-PLpro e SCoV2-PLpro está em consonância com o que os dados experimentais disponíveis na literatura. No nosso estudo é possível notar Figura 5.8(b), que o FEL do Zn-( $\beta 6\beta 7$ ) apresenta três mínimos bem característicos em projeções de PC1 *vs.* PC2 onde temos movimentos de maior amplitude e intrínsecos da proteína sofrendo expansões substancialmente.

Os resultados indicam que há um conjunto de diferentes conformações em todas as componentes principais analisadas. Isso sugere que os movimentos em direção ao Zn-( $\beta 6\beta 7$ ) são menos abrangentes. Ao analisar todas as simulações de Dinâmica Molecular em uma única trajetória, esse comportamento pode ser observado. Isso ocorre quando o cátion  $Zn^{2+}$  está tetracoordenado às cisteínas por meio da ligação Zn-S, o que é essencial para a formação do domínio do dedo de zinco.<sup>[152]</sup> O zinco atua como eletrófilo, enquanto os resíduos de

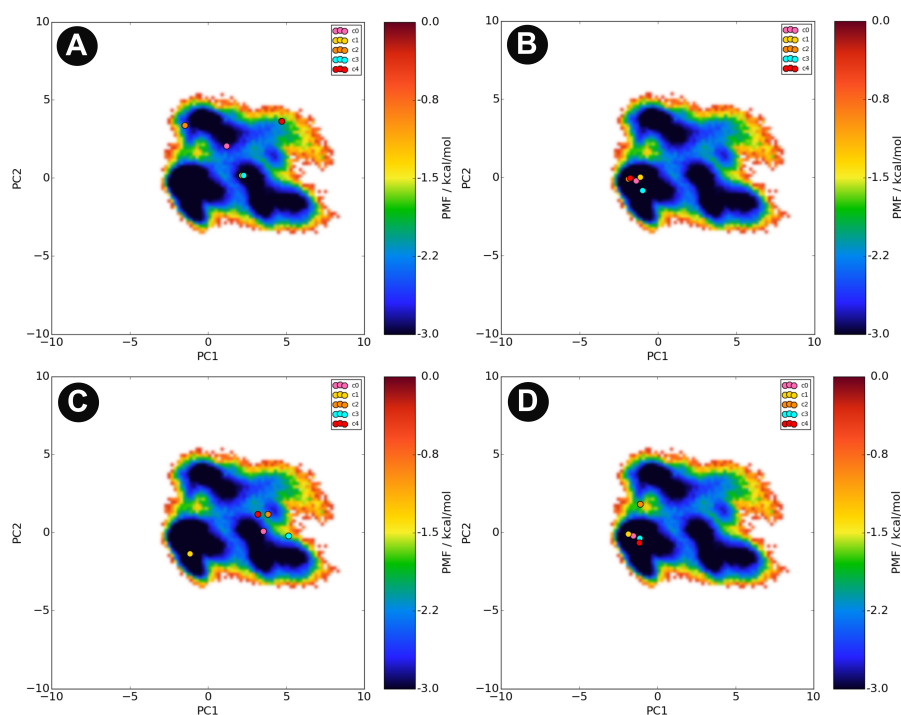
aminoácidos das cisteínas, devido à presença de enxofre em suas cadeias laterais, conferem caráter nucleofílico e redox ativo. Para uma compreensão mais detalhada dessa interação, uma abordagem quântica será realizada, utilizando descritores de reatividade quântica.



**Figura 5.8:** Perfis de energia livre para todos os sistemas. (a) Ubl; (b) Motivo Zn-( $\beta 6\beta 7$ ); (c) Ubl mais o motivo Zn-( $\beta 6\beta 7$ ) juntos. Os pequenos símbolos de triângulos, ou quadrados e círculos em B, ao longo do gráfico representam os aglomerados conformacionais mais populosos considerando cada região em específico e todo o conjunto de trajetórias dos sistemas.

A partir dessa perspectiva, conduzimos uma segunda análise visando a reconstrução dos perfis de energia livre, focando exclusivamente nos resíduos associados ao Zn-( $\beta 6\beta 7$ ) nas projeções de maior amplitude PC1 *vs.* PC2. No entanto, também observamos variações significativas nas demais projeções internas (consulte os apêndices para mais informações).

Conforme esperado, os resultados na Figura 5.9 demonstram que tanto a SCoV1-PLpro quanto a SCoV2-PLpro com íons de zinco exibem movimentos menos extensos, ao passo que a SCoV1-PLpro e a SCoV2-PLpro sem zinco revelam movimentos mais amplos nesse domínio da proteína.



**Figura 5.9:** Perfis de energia livre para Zn-( $\beta 6\beta 7$ ) em cada sistema da *Papain-Like protease*. Os círculos dos gráficos representam as projeções médias. (A) SCoV1-PLpro sem zinco, (B) SCoV1-PLpro com zinco, (C) SCoV2-PLpro sem zinco e (D) SCoV2-PLpro com zinco ao longo das PCs.

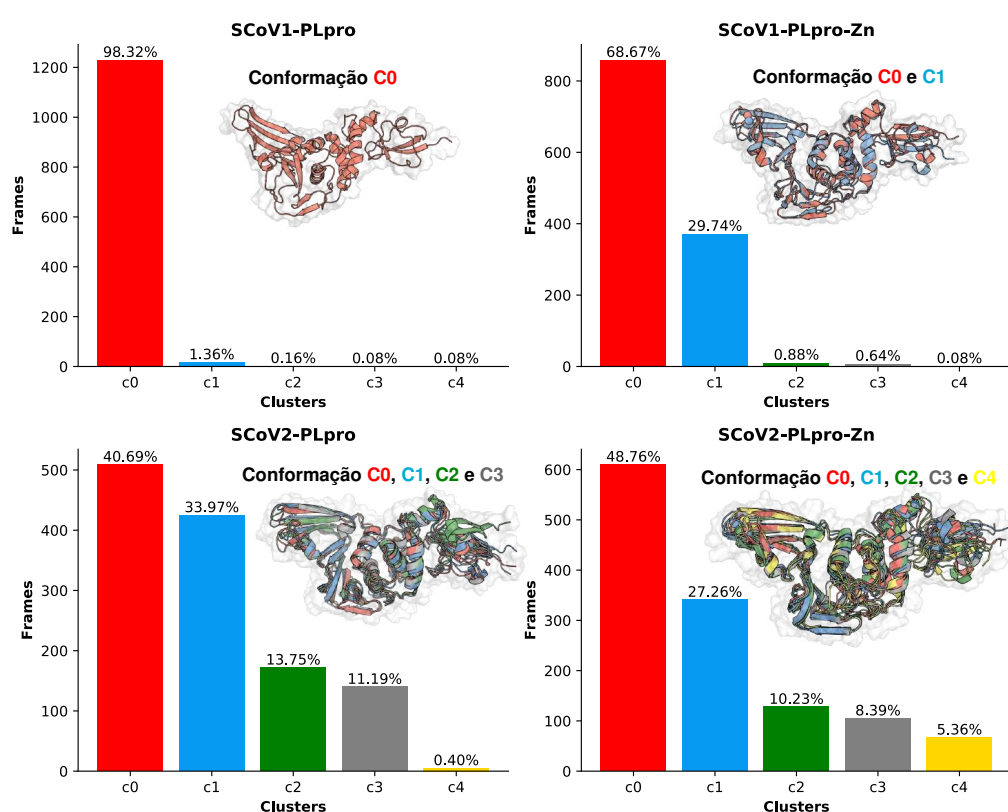
Os movimentos observados durante a simulação de Dinâmica Molecular (DM), conforme evidenciado pelos gráficos de RMSF, foram recuperados por meio da PCA. Esta análise considerou exclusivamente os resíduos envolvidos na formação do domínio Zn-( $\beta 6\beta 7$ ).

Notavelmente, as proteínas contendo zinco tendem a ser projetadas na centralidade da PCAs, formando um aglomerado coeso. Por outro lado, as proteínas desprovidas de zinco têm uma tendência a se afastar dessa centralização, espalhando-se praticamente por todos os perfis de energia.

Realizamos a análise de cluster hierárquico, a partir da qual obtivemos cinco clusters representativos de cada trajetória dos sistemas, conforme ilustrado na Figura 5.10. Esses clusters foram empregados na reconstrução dos perfis de energia livre, concentrando-se unicamente nos resíduos relacionados ao motivo Zn-( $\beta 6\beta 7$ ) (os gráficos correspondentes encontram-se no apêndice).

Nossos resultados evidenciam que a presença de íons de zinco no domínio do dedo de zinco desempenha um papel fundamental na estabilização dessa região, influenciando de modo significativo os movimentos dinâmicos do domínio Ubl. Além disso, observamos diferenças conformacionais notáveis entre as proteases PLpro do SCoV-1 e PLpro do

SCoV-2 na região do *loop* BL2, Ubl e dedo de zinco, o que sugere que essas variações conformacionais podem estar relacionadas às 53 mutações que o vírus de SCoV-2 fez na PLpro. Essas descobertas destacam o papel essencial da protease PLpro na identificação e captura de substratos cruciais para a resposta imunológica, como o interferon ISG15 ou o K48-Ub2.<sup>[153,154]</sup> A compreensão dessas diferenças conformacionais oferece *insights* valiosos que podem ser explorados na busca por inibidores eficazes, apresentando alternativas promissoras para a inibição da função da PLpro.



**Figura 5.10:** Os 5 clusters mais populosos de cada trajetória com suas respectivas representatividade em porcentagem para cada sistema com ou sem zinco. Esses clusters serão utilizados em análises posteriores.

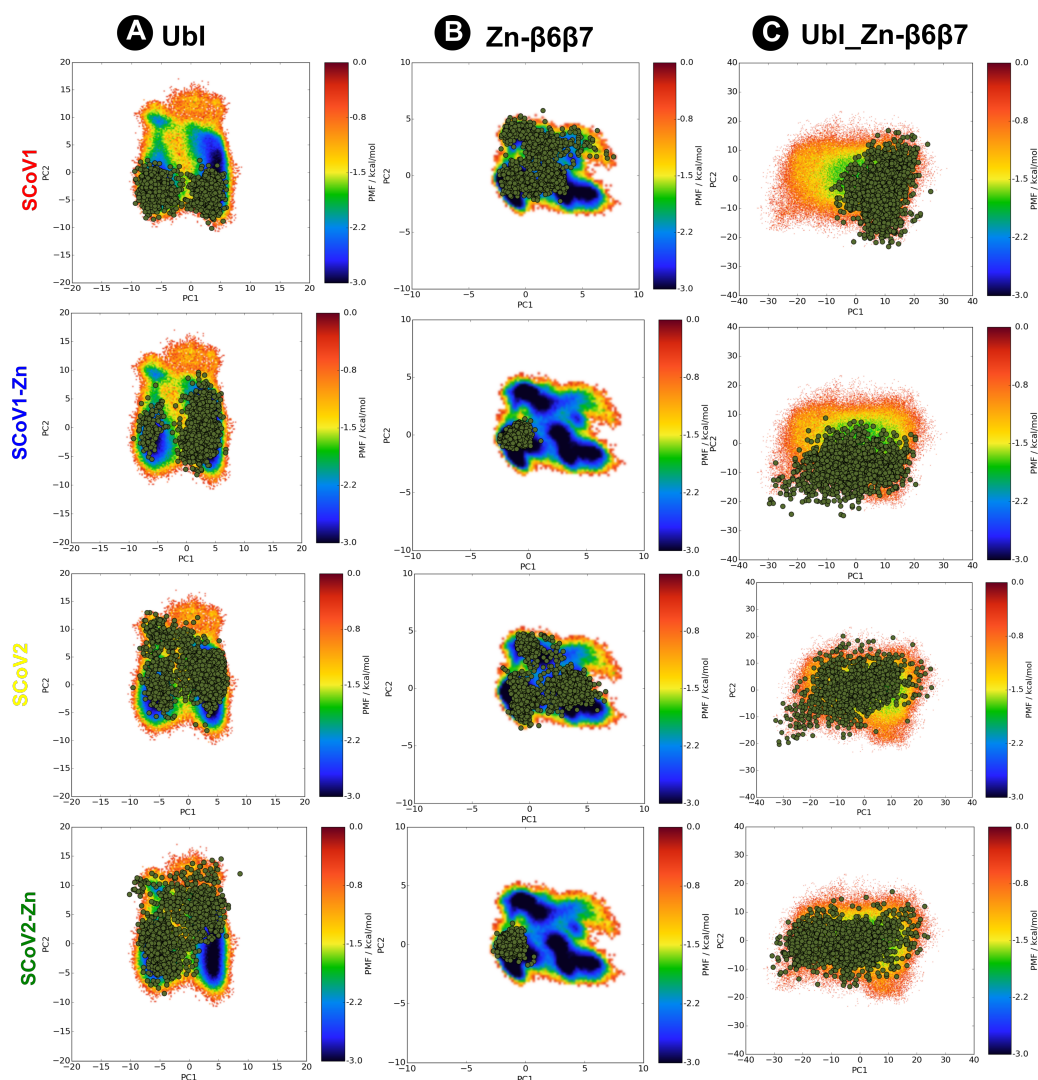
Como já foi discutido, a PLpro também tem como alvo modificações pós-traducionais de ubiquitina e ISG15, que fazem parte da resposta imune antiviral. Portanto, o Ubl tem uma função importante nesse contexto, acredita-se que esse domínio é responsável por capturar as células do sistema imune do organismo do hospedeiro e levar para ser digerido no sítio ativo dessa protease.<sup>[155]</sup>

É importante ressaltar que a presença de zinco na PLpro influencia a conformação da alça BL2. Conforme ilustrado na Figura 5.10, observamos que, quando o zinco está presente na estrutura da proteína, a alça BL2 adota conformações mais fechadas. No entanto, na



ausência de zinco, a alça BL2 assume conformações mais abertas. Esse resultado é de relevância significativa para o desenvolvimento de novos antivirais.

Como foi possível observar, os gráficos de RMSD-2D mostraram que esse domínio sofre alterações conformacionais significativas ao longo da trajetória. Explorando esses resultados fizemos a reconstrução dos perfis de energia livre levando em consideração apenas os resíduos do Ubl nas projeções de PCs, Figura 5.11. Os resultados encontrados para as projeções foram similares, contudo, podemos notar pequenas diferenças nas projeções.



**Figura 5.11:** Perfis de energia livre para cada sistema da PLpro considerando apenas os resíduos que forma o Ubl. Os círculos dos gráficos representam as projeções médias de cada sistema ao longo das PCs.

Os resultados, quando consideramos apenas os resíduos do Ubl, revelam maiores discrepâncias entre a SCoV2-PLpro e a SCoV1-PLpro. Ao observarmos o gráfico de RMSF, notamos que a região do Ubl demonstra maior flexibilidade na SCoV2-PLpro. Entretanto, na análise de PCA (Figura 5.11), observamos que a SCoV2-PLpro, tanto com zinco quanto

sem zinco, tende a se dispersar por toda a FEL, o que corrobora os resultados de RMSF, indicando que essa protease é capaz de realizar movimentos mais amplos..

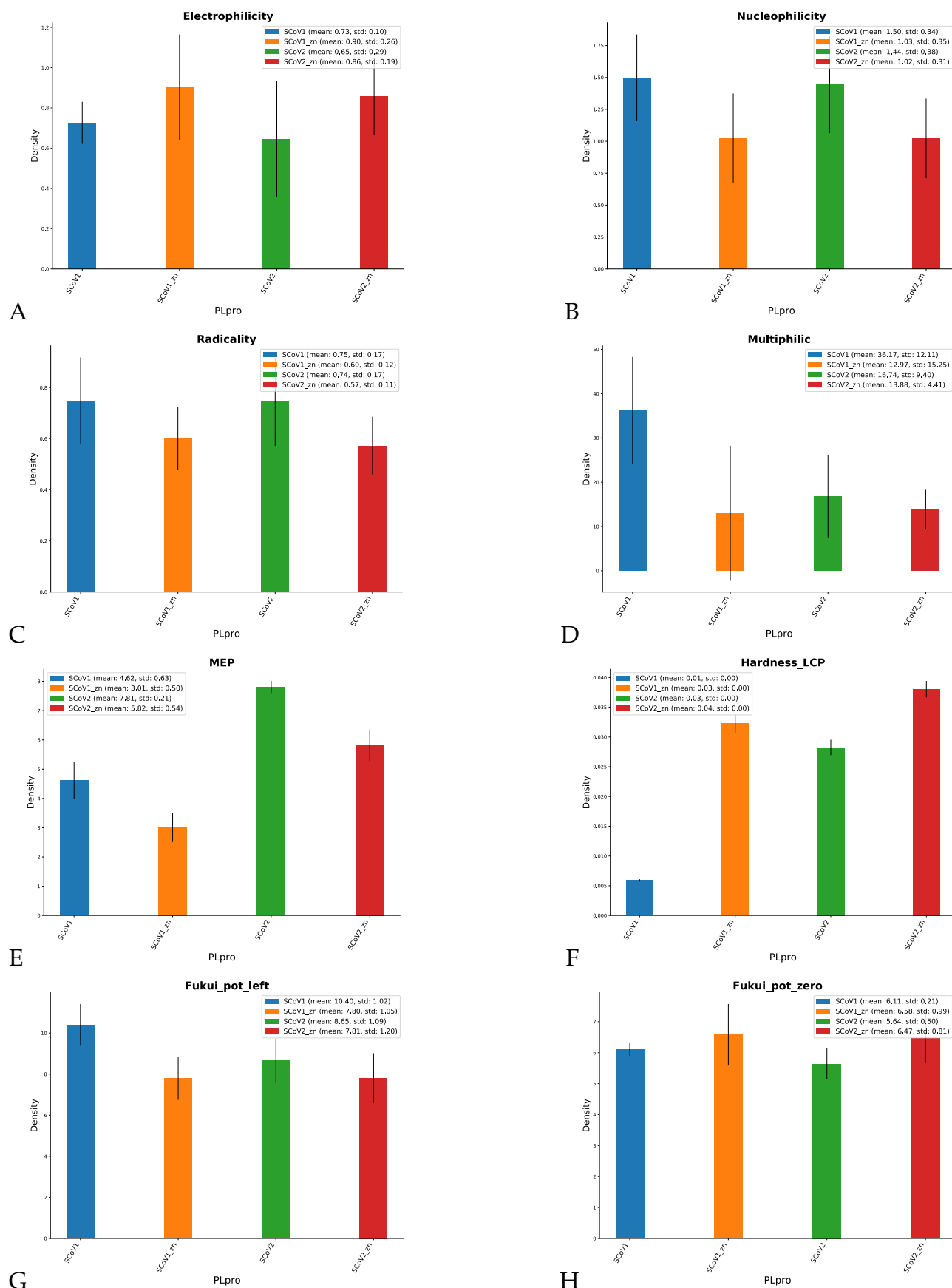
Esses movimentos mais amplos são particularmente evidentes na SCoV2-PLpro que contém zinco, sugerindo que o zinco, coordenado tetraedralmente com as cisteínas que possuem enxofre em suas cadeias laterais, reduz o grau de liberdade nessa região da protease. Isso resulta em um aumento no movimento de fechar a "palma da mão" da PLpro e uma maior mobilidade do Ubl. Essas observações sugerem que essas duas sub-regiões desempenham um papel crítico nas conformações de ligação ao substrato, o que as torna uma via importante para o desenvolvimento de inibidores potenciais.

#### 5.1.1.2 Descritores de Reatividade na *Papain-Like Protease*

Os descritores de reatividade para os resíduos de aminoácidos para a PLpro do SCoV-1 e SCoV-2 para as conformações alinhadas pela região do Ubl e do dedo de zinco, foram calculados e utilizados nas análises gráficas com o objetivo de caracterizar a distribuição de valores e encontrar diferenças estruturais e eletrônicas nessas duas proteínas. A Figura 5.12 apresenta um resumo dos resultados dos descritores, mostrando a média e o desvio padrão para cada descritor em cada sistema.

A eletrofilicidade mostra uma pequena diferença entre as duas proteínas, com uma média maior para SCoV1-PLpro com zinco, no valor de 0,90 eV, e um desvio padrão de 0,26 eV, indicando que os resíduos apresentam uma variabilidade nesse descritor. Já SCoV2-PLpro apresentou uma média de 0,86 eV e um desvio padrão de 0,19 eV. Isso demonstra que a adição de zinco na estrutura aumenta a reatividade, ou seja, a capacidade de aceitar elétrons e formar uma ligação com uma espécie nucleofílica. Podemos observar que as proteínas sem zinco em sua estrutura apresentaram os menores valores de eletrofilicidade para os dois vírus.

A nucleofilicidade e a radicalidade são descritores que, de forma geral, apresentaram uma distribuição de valores semelhante tanto para as proteínas SCoV1-PLpro e SCoV2-PLpro sem zinco quanto para as proteínas com zinco, cujos resultados também são semelhantes entre si. Essa consistência sugere que a Susceptibilidade ao ataque eletrofílico (nucleofilicidade) e a suscetibilidade radical (radicalidade) são significativamente alteradas pela presença do zinco na estrutura das proteínas.



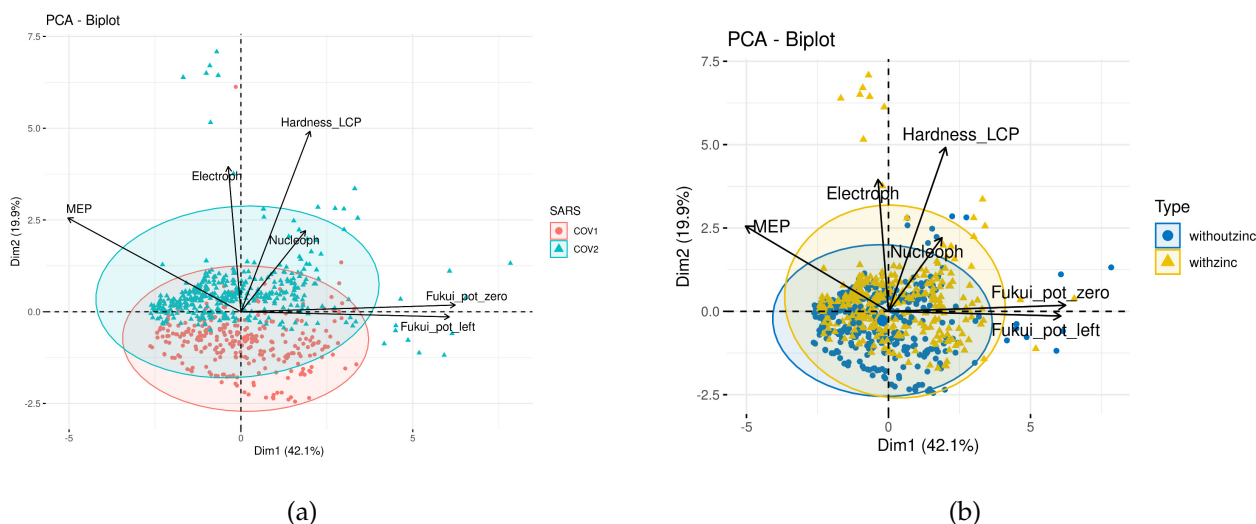
**Figura 5.12:** Descritores de reatividade calculados entre as estruturas amostradas para SCoV1-PLpro e SCoV2-PLpro com e sem zinco (A) Eletrofilicidade, (B) Nucleofilicidade, (C) Radicalidade, (D) Multifilicidade, (E) Potencial Eletrostático Molecular, (F) Dureza Local (aproximação potencial químico local), (G) Dureza local (potencial Fukui) e (H) Potencial Fukui radical.

Outro descritor que descreve a reatividade é a multifilicidade, que se refere à capacidade de uma molécula de interagir com outras espécies químicas. A Figura 5.12-D mostra que a SCoV2-PLpro com zinco é um pouco mais reativa com outras espécies químicas do que a SCoV1-PLpro com zinco. No entanto, a SCoV1-PLpro sem zinco apresentou os maiores valores de multifilicidade, demonstrando que a presença de zinco influencia na reatividade dessa protease com outras espécies químicas. Quando observamos o Potencial Molecular Eletrostático (Figura 5.12-E), vemos que a SCoV2-PLpro é mais reativa do que a SCoV1-PLpro.

Para os descritores de dureza local, que abordam as interações controladas por carga, especificamente aquele baseado no potencial químico local (Figura 5.12-F), a dureza local apresentou uma diferença muito pequena entre as duas proteases. Já a dureza local baseada no potencial de Fukui (Figura 5.12-G e H) mostrou resultados semelhantes para a SCoV2-PLpro, tanto com zinco quanto sem zinco. Para a protease do SARS-CoV1 foi possível evidenciar uma diferença maior para a sem zinco em comparação com a mesma protease com zinco. É possível demonstrar uma pequena diferença na dureza local entre a SCoV1-PLpro e SCoV2-PLpro. Considerando que ainda havia diversas fontes de possíveis efeitos nas duas proteases, como diferentes resíduos, coordenadas tridimensionais e descritores quânticos, foi conduzida uma análise exploratória utilizando PCA.

A PCA foi aplicada primeiro para as duas proteases separando apenas pelo tipo de vírus SARS-CoV-1 e SARS-CoV-2, segundo pelo tipo de proteína com e sem zinco e, a variância explicada por componente principal. Observamos uma diferenciação maior quando comparamos os dois tipos de SARS.

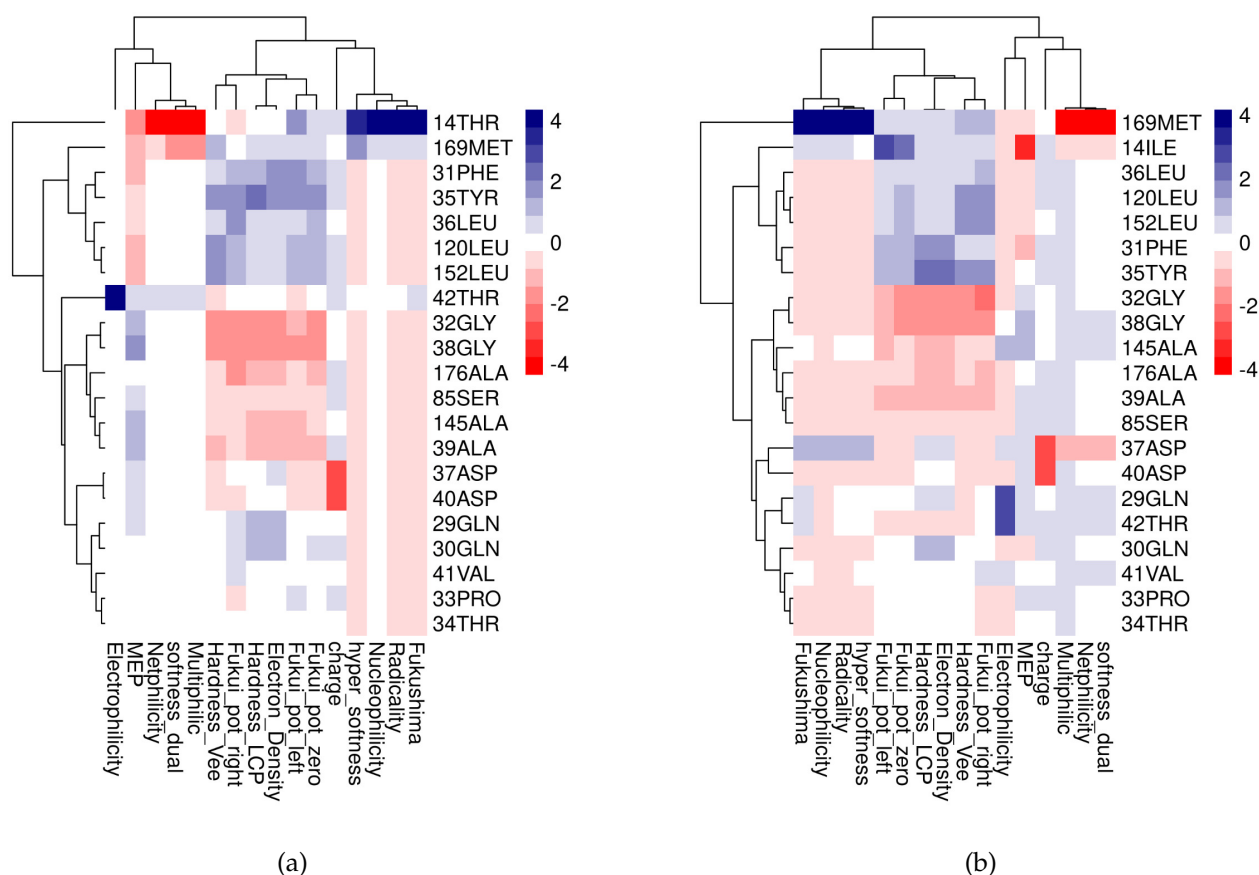
Os descritores mostrados na figura (5.12) estão altamente associados ao PC que separa os dados por tipo de vírus, mas com contribuição significativa também na segunda PC que compara o tipo da protease. Por outro lado, a Nucleofilicidade, a radicalidade, Eletrofilicidade, Potencial Eletrostático Molecular e a dureza Local (aproximação potencial químico local) demonstram uma maior contribuição para PC2, com loadings mais altos, o que significa que podem representar a diferenciação dos dois tipos de proteínas.



**Figura 5.13:** Resultados da análise de componentes principais (PCA) para os descritores de reatividade. (a) Objetos pelo tipo de vírus SARS-CoV-1 e SARS-CoV-2 (b) Objetos pelo tipo de Proteína com e sem zinco. Os círculos e triângulos representa os valores dos resíduos por descritor.

Nas duas rodadas, foram explicados pelo menos 62% da variância nos dois primeiros componentes principais (PCs), com o PC1 explicando cerca de 42% e o PC2 explicando 20%. Uma das vantagens da PCA é a capacidade de explorar uma vasta matriz de dados, segmentando a variância por blocos de variáveis e classes de dados. Com base nessas análises, é possível afirmar que essas duas proteínas apresentam diferentes propriedades de estrutura eletrônica quando comparadas de acordo com o tipo de vírus.

Complementando as análises realizadas até aqui, foram construídos mapas de calor com os valores médios de cada resíduo para todos os descritores (Figura 5.14), sendo a figura A para SARS-CoV-1 e a figura B para SARS-CoV-2. Os dados de entrada para os gráficos foram dimensionados pelo seu desvio padrão e centralizados pela sua média, de modo que os descritores fossem comparáveis. Nestas figuras, é possível visualizar o agrupamento entre os resíduos e descritores, bem como as mudanças nos valores médios dos descritores para alguns dos resíduos.



**Figura 5.14:** Mapas de calor mostrando as médias dos descritores para cada resíduo analisado para a Proteína com e sem zinco. (a) SARS-CoV-1 (b) SARS-CoV-2.

Através dos mapas de calor com os descritores agrupados por tipo, como interações duro-duro e mole-mole, é possível verificar a variação dos resíduos de aminoácidos para as duas proteínas. O ASP-37 e o ASP-40 apresentam similaridade nas duas proteínas para os descritores que descrevem a carga parcial e a eletrofilicidade, com uma diminuição nas médias desses descritores. Isso pode indicar uma menor reatividade ou uma mudança na interação desses resíduos com outras moléculas.

Para a MET-169 e a mutação de THR-14 para ILE-14, observa-se uma diminuição nas médias dos descritores Fukushima, radicalidade, nucleofilicidade e moleza local para a SCoV2-PLpro. Indicando que essas alterações específicas podem afetar a reatividade e a interação da proteína com outras moléculas, além de, impactar na estrutura e função da proteína, possivelmente diminuindo sua capacidade de participar em reações que envolvem suscetibilidade radical ou sua reatividade em geral (nucleofilicidade e moleza local).

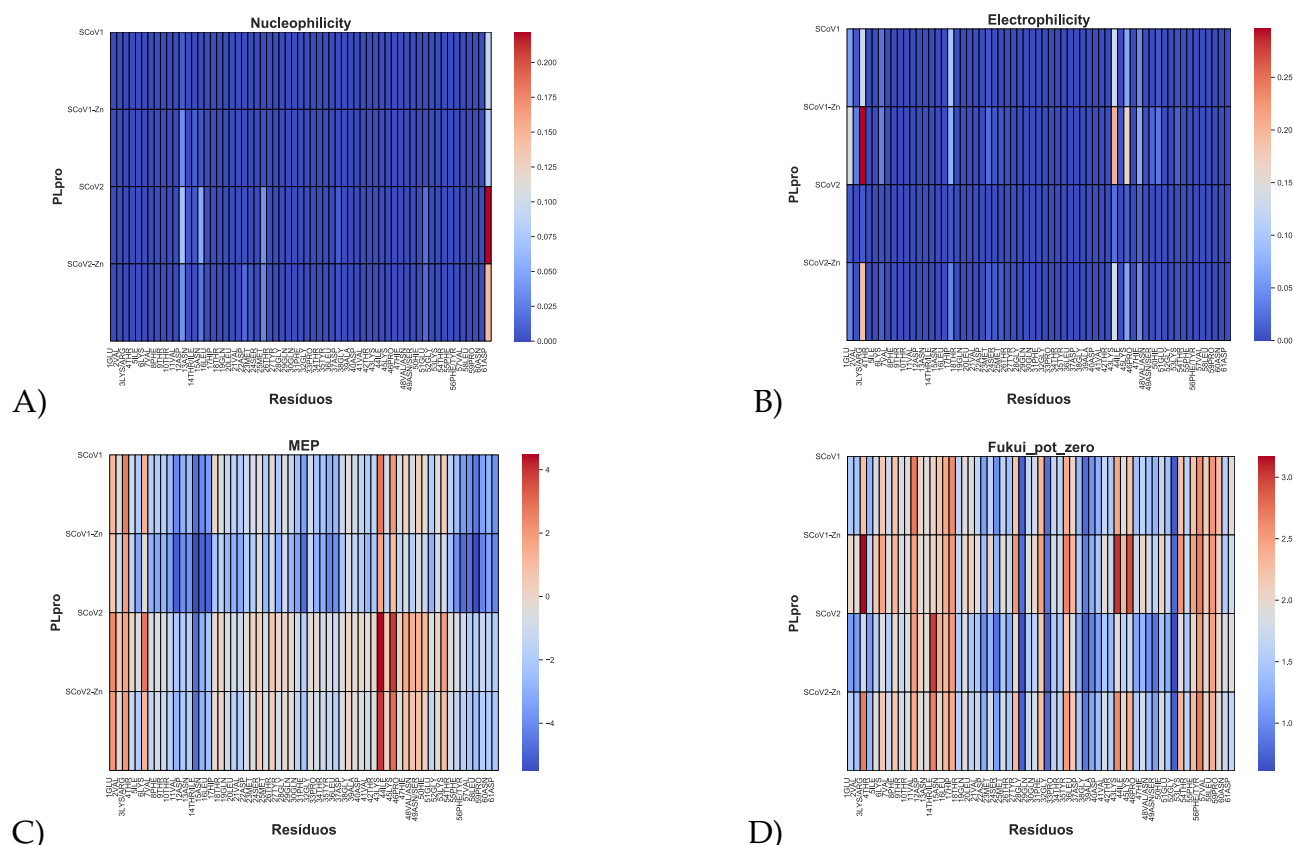
Com base na Figura 5.14, é possível observar que a maioria dos resíduos de aminoácidos

com valores de dureza ou moleza local estão concentrados no domínio ubiquitina dessas duas proteases, sobretudo, para SCoV2-PLpro. Com essa informação, é interessante estudar a reatividade desse domínio específico para identificar possíveis diferenças entre os dois tipos de SARS e a influência da presença de zinco.

Através das análises de PCA e dos mapas de calor, a Figura 5.15 mostra os descritores de nucleofilicidade, eletrofilicidade, potencial molecular eletrostático e potencial Fukui radical para os resíduos do domínio da ubiquitina que estão apresentando maior variação.

Na nucleofilicidade, observamos uma pequena variação nos resíduos ASP-12, ASN-15, MET-25 e ASP-61. Notamos que os maiores valores estão associados à SCoV2-PLpro, especialmente na protease sem zinco. Esses resíduos podem aumentar a reatividade dessas proteases com, por exemplo, o ISG15. Já na eletrofilicidade, não há diferenças significativas quando comparamos os dois SARS; apenas os resíduos histidina-17, histidina-47, lisina-43 e lisina-45 são sensíveis a esse descritor tanto para SARS-CoV-1 quanto para SARS-CoV-2 com zinco, que também descreve a reatividade.

O descritor do potencial molecular eletrostático revela que o SCoV2-PLpro apresenta as maiores diferenças em sua distribuição ao comparar com a outra proteína, mostrando uma reatividade maior para o domínio da ubiquitina. Essa reatividade diminui um pouco quando o zinco é adicionado à estrutura. Por outro lado, o potencial Fukui radical exibe uma distribuição semelhante para SCoV1-PLpro e SCoV2-PLpro com zinco, identificando resíduos mais reativos que podem se ligar a substratos.



**Figura 5.15:** Mapas de calor mostrando as médias dos descritores para cada resíduo analisado para o domínio ubiquitina. (A) Nucleofilicidade (B) Eletrofilicidade (C) Potencial Molecular Eletrostático e (D) Potencial Fukui Radical.

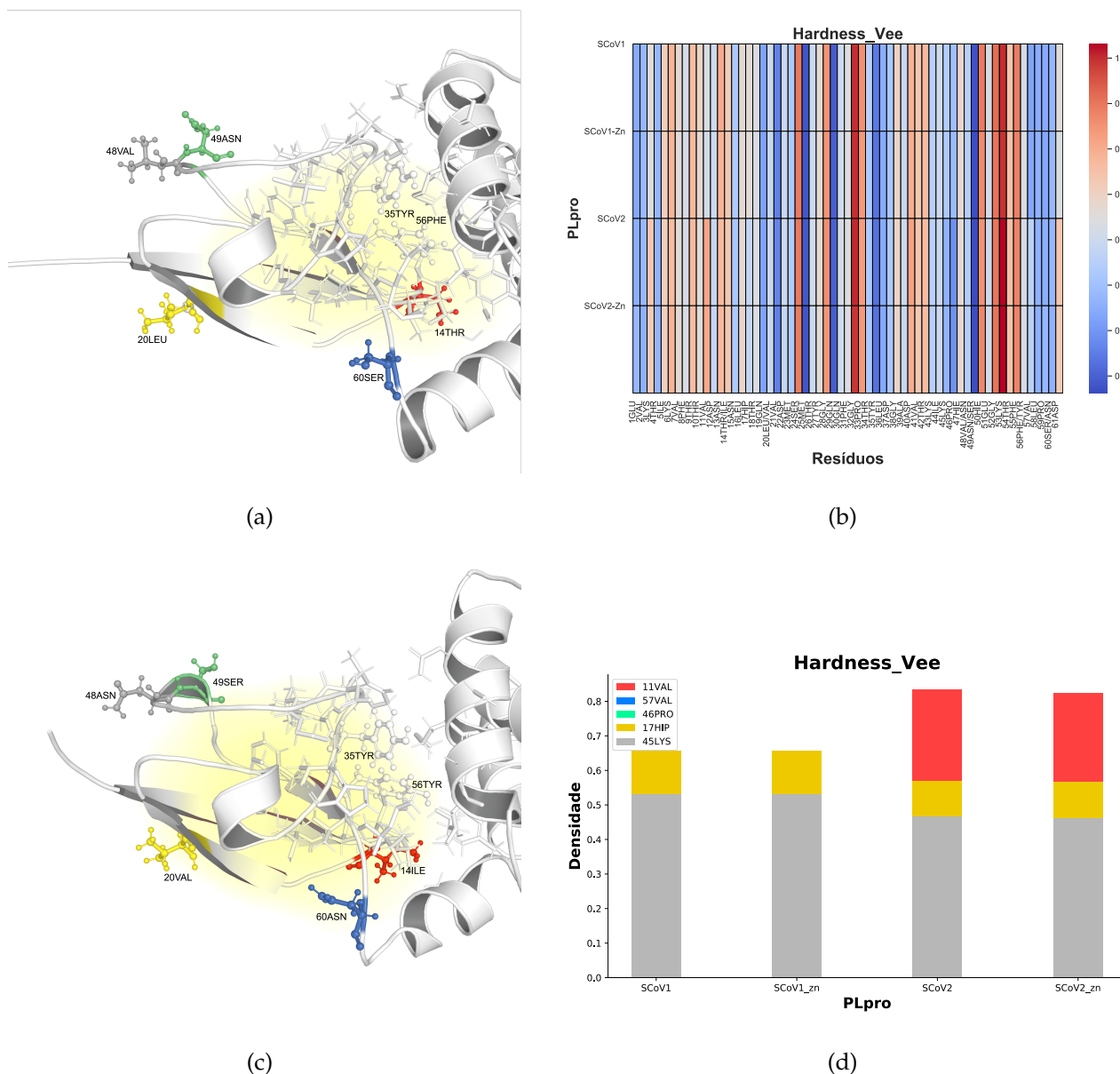
A Figura 5.16 apresenta o resultado da dureza local (potencial elétron-elétron) para o região do Ubl para o cluster c0, mais populoso de cada protease com e sem zinco. Conforme mostrado na Figura 5.16, a dureza local (potencial elétron-elétron) tem valores mais altos para os resíduos TYR-35 e PHE-56 para a SCoV1-PLpro e para os resíduos TYR-35 e TYR-56 para a SCoV2-PLpro.

A tirosina apresentou uma pequena diminuição quando adicionamos zinco na protease, enquanto que, mesmo com a SCoV2-PLpro fazendo uma mutação na fenilalanina-56 para uma tirosina, os resultados são bem semelhantes mostrando que esses resíduos são menos sensíveis à mudança na sua densidade eletrônica causada pela interação elétron-elétron, ou seja, são menos que suscetíveis a ataques nucleofílicos ou eletrofílicos.

Algumas mutações influenciaram na variação da dureza local para os resíduos THR-14, LEU-20, VAL-48, ASN-49 e SER-60 encontrados na protease do SARS-CoV-1, e para as suas respectivas mutações ILE-14, VAL-20, ASN-48, SER-49 e ASN-60 que ocorreram na protease do SARS-CoV-2, Figura 5.16(b). Os resíduos que sofreram uma maior variação



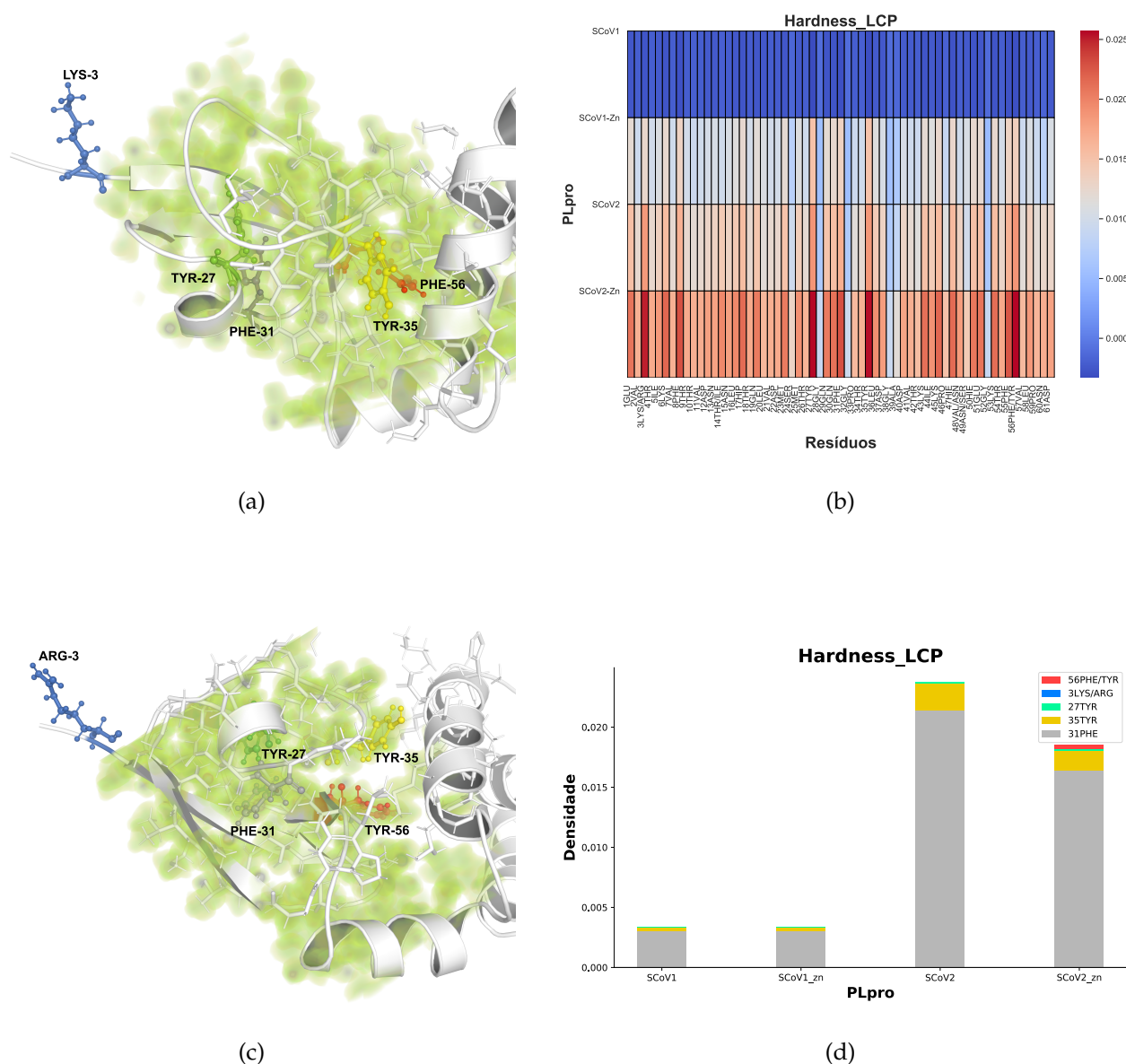
na dureza local para a SCoV1-PLpro foram para leucina-20 e valina-48, Figura 5.16(d). A mutação de treonina-14 para isoleucina aumentou a variação da dureza local na SCoV2-PLpro; observamos também uma variação maior na dureza local dessa protease nos resíduos valina-20 e asparagina-48, Figura 5.16(d).



**Figura 5.16:** Dureza Local (potencial elétron-elétron) (a) SCoV1-PLpro os resíduos com maiores variação são destacados pelas cores verde, cinza, amarelo, vermelho e azul, (b) Dureza Local (potencial elétron-elétron) para o domínio ubiquitina, (c) SCoV2-PLpro os resíduos com maiores variação são destacados pelas cores verde, cinza, amarelo, vermelho e azul, (d) Resíduos com maiores variação para a dureza local no domínio ubiquitina.

Os resíduos de glicina-28, glicina-32 e glicina-52 apresentaram os menores valores de dureza local para a SCoV2-PLpro com e sem zinco, isso significa que a glicina é mais mole

e podem facilitar a interação entre diferentes partes da PLpro ou sofrer ataques nucleofílicos ou eletrofílicos de um potencial inibidor dessa protease.



**Figura 5.17:** Dureza Local (potencial químico) (a) SCoV1-PLpro os resíduos com maiores variação são destacados pelas cores verde, cinza, amarelo, vermelho e azul, (b) Dureza Local (potencial químico) para o domínio ubiquitina, (c) SCoV2-PLpro os resíduos com maiores variação são destacados pelas cores verde, cinza, amarelo, vermelho e azul, (d) Resíduos com maiores variação para a dureza local no domínio ubiquitina.

Ao observarmos a dureza local (potencial químico) Figura 5.17 para o domínio de ubiquitina, notamos que a SCoV1-PLpro é ligeiramente mais duro, menos polarizável do que a SCoV2-PLpro (Figura 5.17(b)). No SARS-CoV-2, a PLpro com zinco apresentou resultados maiores do que a protease sem zinco, sugerindo que o dedo de zinco pode influenciar na

estabilidade eletrônica dos outros domínios dessa protease.

Os valores de dureza local para as duas proteases foram mais elevados nos resíduos ARG-3, TYR-27, PHE-31 e TYR-35 na SCoV2-PLpro com zinco são um pouco menores para esses mesmos resíduos na SCoV2-PLpro sem zinco, como podemos observar, indicando que o domínio do dedo de zinco influenciou a dureza local dos resíduos no domínio de ubiquitina, tornando-os mais duro após a adição do zinco na protease (Figura 5.17(d)). As equações de trabalho de dureza local são descritores que demonstram graficamente as diferenças entre o domínio ubiquitina das duas proteínas, especialmente em relação à aproximação pelo Potencial de Fukui.

Com base nos resultados obtidos dos *clusters* mais representativos da simulação de DM, foi demonstrado que as propriedades eletrônicas e de reatividade calculadas para os resíduos das proteínas SCoV1-PLpro e SCoV2-PLpro, mesmo considerando apenas as estruturas mais representativas, são significativamente diferentes entre as proteínas.

Os resultados indicam que SCoV2-PLpro é mais reativa e, portanto, mais polarizável o que pode explicar a interação maior dessa protease com células de defesas do nosso organismo. O zinco em sua estrutura também pode influenciar na estrutura eletrônica e a reatividade entre domínios. Embora, neste momento, não seja possível estabelecer uma correlação direta entre essas diferenças e os arranjos tridimensionais específicos dos resíduos das proteínas que podem levar a mudanças em suas propriedades de reatividade, esses resultados abrem novas possibilidades de estudo para compreendermos a função catalítica dessa protease.

### 5.1.2 Domínio de Ligação ao Receptor (RBD)

Como descrito na metodologia, fizemos 37 mutações no RBD da proteína *Spike* do SARS-CoV-2 que liga ao receptor celular ACE2. Esse trabalho é feito em parceria com o grupo de pesquisa da UFSJ e os primeiros resultados das simulações de dinâmica molecular desse parceria, realizadas para a selvagem e a P.1 foram publicados na revista PeerJ (anexo).<sup>[156]</sup>

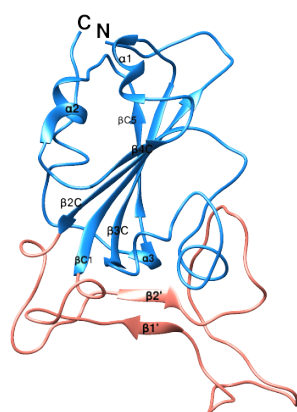
Nosso principal objetivo é avaliar as diferenças conformacionais entre a proteína *Spike* selvagem e suas variantes que apresentam maior afinidade de interação com o receptor celular ACE2 e níveis variados de resistência aos anticorpos. Estas variantes incluem aquelas com maior infectividade ao receptor ACE2 e evasividade ao sistema imune foram rotulados (+inf/+ev), aquelas mutantes que aumentam somente infectividade ao receptor ACE2 (+inf/-ev), aquelas que aumentam apenas evasividade ao sistema imune (-inf/+ev),

bem como as mutações neutras (-inf/-ev). As mutações neutras são caracterizadas por não influenciar na infectividade receptor ACE2 e nem na evasividade ao sistema imune.

Além disso, uma vez que o Domínio de Ligação ao Receptor é o principal alvo dos anticorpos neutralizantes induzidos por vacinas (conhecidos como nAbs, do inglês, *vaccine-induced neutralizing antibodies*),<sup>[157]</sup> nossa intenção é investigar e compreender como essas mutações podem afetar as interações do RBD com o ACE2 e os anticorpos. Isso é particularmente relevante, uma vez que tais mutações podem potencialmente impactar a eficácia das vacinas de mRNA, como as desenvolvidas pela Moderna e Pfizer, sem necessariamente comprometer sua competência imunogênica, conforme ressaltado por Wang et al.<sup>[59]</sup>

De acordo com Lima et al. (2020) e Rocha et al. (2021),<sup>[69,158]</sup> o estudo das mudanças na distribuição conformacional de proteínas é importante, pois desempenha um papel fundamental na alteração da afinidade em complexos macromoleculares. Enquanto o RBD, localizado na proteína *Spike*, é determinante para a interação com o pico-receptor, os resíduos ativos estão no RBM, uma região crucial localizada no RBD, desempenhando um papel fundamental na ligação à superfície externa de seus receptores.<sup>[30]</sup>

Para auxiliar na apresentação e discussões dos nossos resultados de MD sobre tais aspectos, utilizaremos a nomenclatura de Wang et al.<sup>[3]</sup> para os motivos RBD e RBM, Figura 5.18.



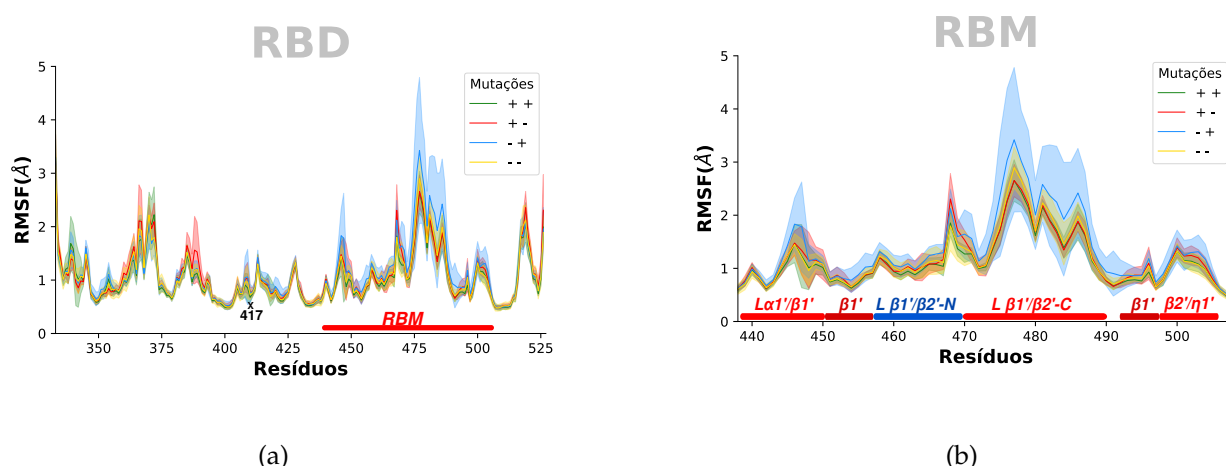
**Figura 5.18:** Nomenclatura adotada para RBD e RBM. Fonte: adaptada de Wang et al. (2020)<sup>[3]</sup>

Nossos resultados de dinâmica molecular (MD) para todas as variantes e mutações realizadas no RBD mostram maiores flutuações e desvio padrão na região dos resíduos 360 a 374 e RBM. Na região dos resíduos 360 a 374 houve um desvio padrão maior para

as mutações que possuem maior infectividade ao receptor ACE2 e evasividade ao sistema imune +inf/+ev e para aquelas que apresenta maior infectividade ao receptor e menor evasividade ao sistema imune +inf/-ev, os resultados são apresentados na Figura 5.19.

Já para a região do RBM os resultados mostram que houve um desvio padrão maior para as mutações que possuem menor infectividade ao receptor celular ACE2 e maior evasividade aos anticorpos -inf/+ev (Figura 5.19(b)).

Conforme evidenciado nos gráficos de RMSF, as flutuações são mais pronunciadas na porção C-terminal da alça  $\beta'_1/\beta'_2$  (a seguir denominada alça  $\beta'_1/\beta'_2 - C$ , que abrange os resíduos 470 a 490), em comparação com o restante do RBD (Figs. 5.19(a)-5.19(b)). Essa alça engloba a posição 484, onde ocorre uma das mutações significativas, a mutação E484K, presente na maioria das variantes classificadas como preocupantes pela OMS. A maior variação na mobilidade da alça  $\beta'_1/\beta'_2 - C$  foi observada em relação ao seu movimento, tanto para o RBD como em sua mobilidade interna no RBM. Além disso, os loops  $\beta'_2/\eta'_1$  (que abrangem os resíduos 497 a 506, no centro da substituição N501Y) e o loop  $L\alpha'_1/\beta'_1$  (resíduos 438 a 450) são *hotspots* de flutuação local no RBM quanto no RBD como um todo.



**Figura 5.19:** (a) RMSF de toda a região RBD para das spikes classificadas como (+inf/+ev, "++"),(+inf/-ev, "+-"), (-inf/+ev, "-+") e (-inf/-ev, "--"). (B) RMSF obtido alinhando a região do motivo de ligação ao receptor (RBM) e considerando apenas está mesma região (resíduos 438-508). Os RMSFs são apresentados como média (linhas) e desvios (sombra) considerando as três réplicas individualmente e a soma das três de cada mutação e variante.

Este sistema de loops compreende uma estrutura semelhante a uma "garra molecular" de três dedos, com o dedo superior representado pelo loop  $\beta'_1/\beta'_2$  em um lado e os dedos menores representados pelos loops  $\beta'_2/\eta'_1$  e  $\alpha'_1/\beta'_1$  no lado oposto. Essa "garra molecular" atua como um mecanismo de "beliscar", prendendo as bordas das duas hélices

N-terminais de ACE2 (as hélices H1 e H2) durante a ligação, posicionando-as de forma a estabelecer um contato mais profundo com a folha antiparalela curta  $\beta'_1/\beta'_2$  no centro do RBM. Isso implica que há uma certa plasticidade nessa região para facilitar esse movimento de "beliscar" e, assim, estabelecer contato com o receptor. Além disso, a posição 417 (onde ocorre a substituição K417T) ocupa uma posição central na hélice 3, anterior ao RBM, e é menos suscetível a movimentos, como ilustrado na Fig. 5.19(a). Devido a esse empacotamento estratégico no centro do RBM, variações sutis no empacotamento e na mobilidade da posição 417 provavelmente exercerão influência sobre as flutuações em todo o domínio.

É possível observar uma abrangência de flutuação mais significativa em várias simulações para as mutações classificadas como tendo menor infectividade ao receptor celular ACE2 e maior evasividade aos anticorpos -inf/+ev em comparação com as outras mutações, principalmente nos loops onde estão localizadas as posições 484 e 501 do RBM (Fig. 5.19(a)A-B). Nota-se que nessa região de alças como  $\beta_{1'}/\beta_{2'} - C$  e  $\beta_{2'}/\eta_{1'}$ , exibem uma mobilidade máxima cerca de 5 Å no RBD das mutações -inf/+ev do que suas respectivas contrapartes no S:RBD +inf/+ev, S:RBD +inf/-ev, e S:RBD -inf/-ev (Figs. 5.19(a)-5.19(b)).

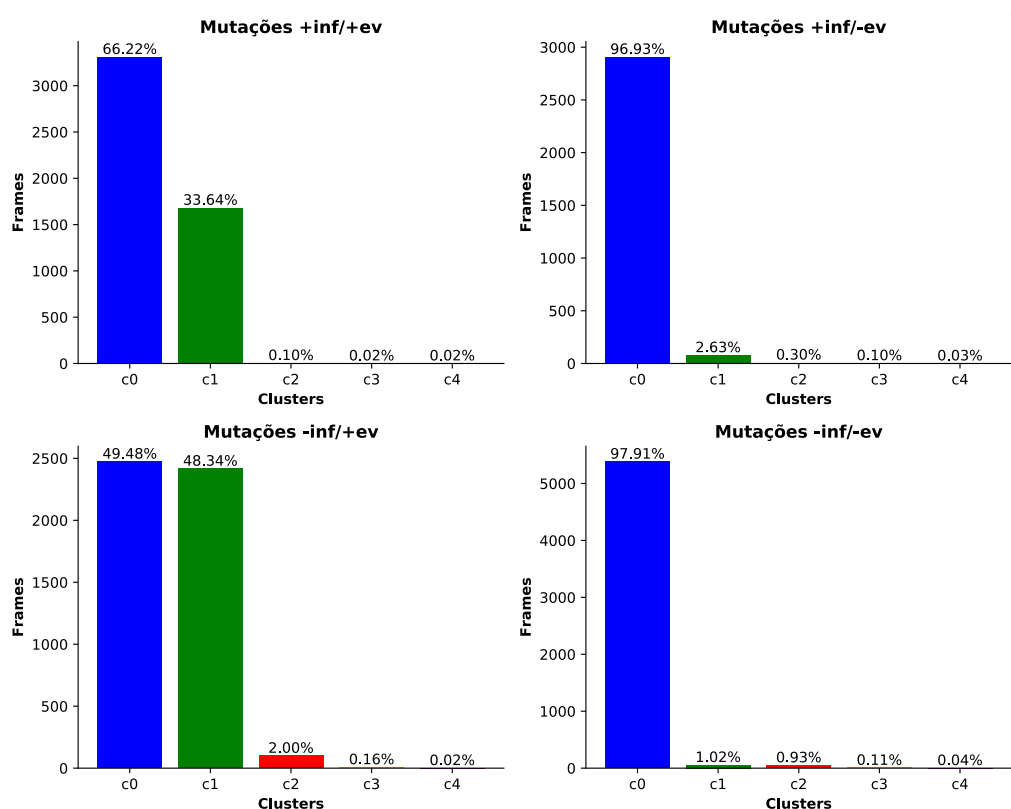
Essa abrangência de flutuação ao longo de diferentes simulações, pode estar relacionada ao fato dessas mutações apresentarem maior resistência aos anticorpos e com isso apresentarem um número maior de conformações do RBM sobretudo do loop  $\beta_{1'}/\beta_{2'} - C$  para justamente "fugir" da resposta imune do hospedeiro. O mesmo acontece para a mutação chave N501Y localizada no loop  $\beta_{2'}/\eta_{1'}$  e G446V localizada  $L\alpha_{1'}/\beta_{1'}$  são *hotspots* de flutuação. No estudo publicado por Harvey, W. T. e Carabelli, A. M. et al. (2021)<sup>[159]</sup> mostram que a mutação G446V conseguiu escapar de três dos 4 soros testados, esse resultado pode estar relacionado com o desvio observado na (Fig. 5.19(b)) em torno de 2,8 Å, nesse mesmo estudo os autores observaram altas pontuações de acessibilidade de anticorpos, centrado nos resíduos 444-447, localizados na região de loop  $\alpha_{1'}/\beta_{1'}$ .

### 5.1.2.1 Análises do Comportamento Conformacional do Domínio de Ligação ao Receptor

As regiões 360-374 e RBM exibiram as maiores flutuações ao longo das simulações de dinâmica molecular. Com base nessa observação, elaboramos as superfícies de energia livre para as regiões 360-374 e RBM para cada fenótipo: +inf/+ev, +inf/-ev, -inf/+ev e -inf/-ev. Utilizamos as trajetórias das simulações tanto para o RBD tipo selvagem quanto para todos os mutantes, obtivemos três componentes principais: PC1, PC2 e PC3 (Figura 5.21). Nessas

superfícies, sobreposamos os *clusters* mais representativos das trajetórias das simulações de DM para cada grupo de mutantes (Figura 5.20).

Como mostrado na Figura 5.20, as conformações c0 apresentaram a maior representatividade: 66,22% para as mutações +inf/+ev, 93,93% para as +inf/-ev, 49,48% para as mutações -inf/+ev, e 97,91% para as mutações -inf/-ev. A conformação c1 apresentou uma representatividade de 33,64% para as mutações +inf/+ev e 48,34% para as mutações -inf/+ev. As demais Conformações possuem representatividade inferior a 5% para todas as mutantes.



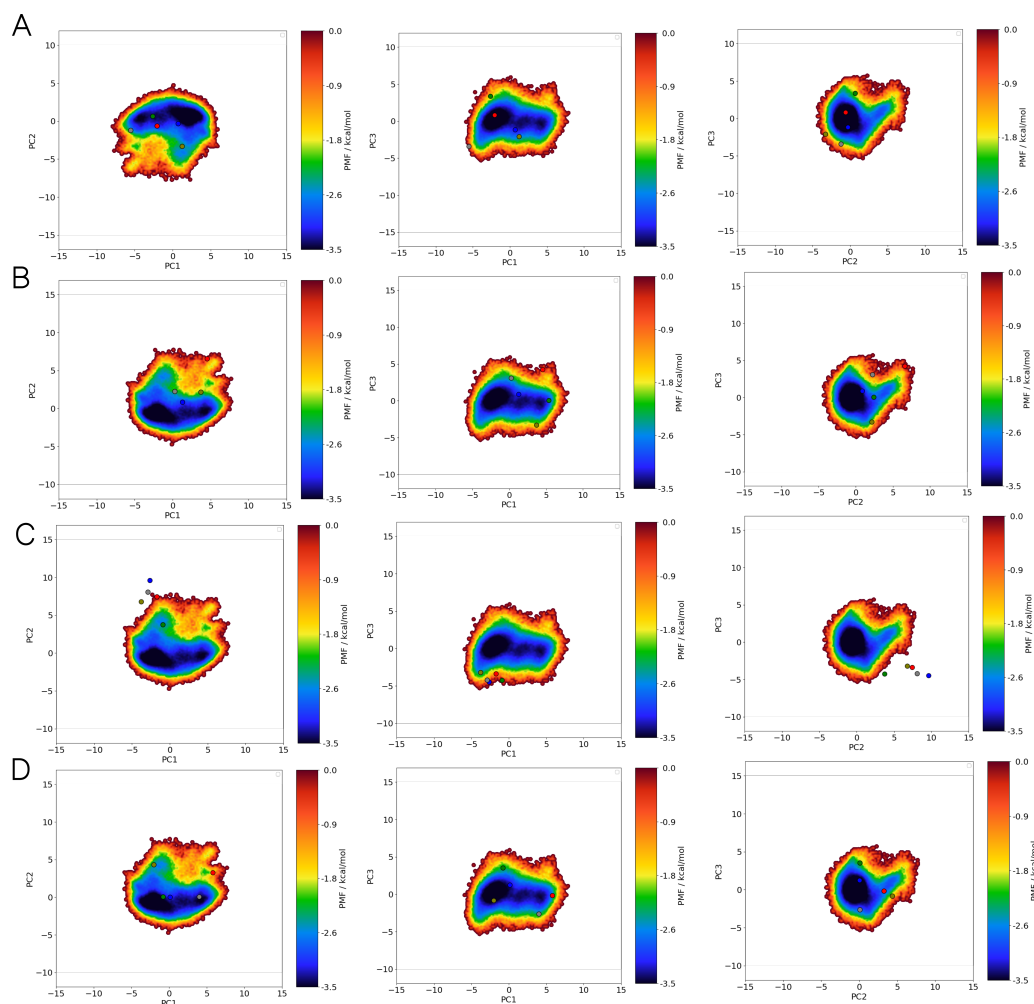
**Figura 5.20:** Os 5 clusters (c0-azul, c1-verde, c2-vermelho, c3-cinza e c4-oliva) mais populosos de cada trajetória com suas respectivas representatividade em porcentagem para cada sistema +inf/+ev, +inf/-ev, -inf/+ev e -inf/-ev. Esses clusters serão utilizados em análises posteriores.

Esses *clusters* foram projetados nas superfícies de energia livre, onde é possível observar suas projeções nas regiões mais frequentes (áreas mais escuras das superfícies de energia livre) e nas regiões menos frequentes (áreas mais claras e periféricas das superfícies de energia livre), com exceção das mutações que apresentam evasividade ao sistema imune (-inf/+ev) que apresentaram projeções apenas nas regiões menos frequentes, além disso os *clusters* estão nas mesmas regiões Figura 5.21-C.

Podemos observar nas Figura 5.21(A-B) que as conformações correspondentes aos mu-



tantes +inf/+ev e +inf/-ev, que apresentam um aumento na infectividade, exibem uma distribuição mais ampla em conformações tanto nas regiões de maior quanto de menor frequência. Nota-se uma pequena diferença para as conformações das mutações +inf/-ev, onde apenas a conformação c0 (em azul) está localizado na região de maior frequência (Figura 5.21-B).

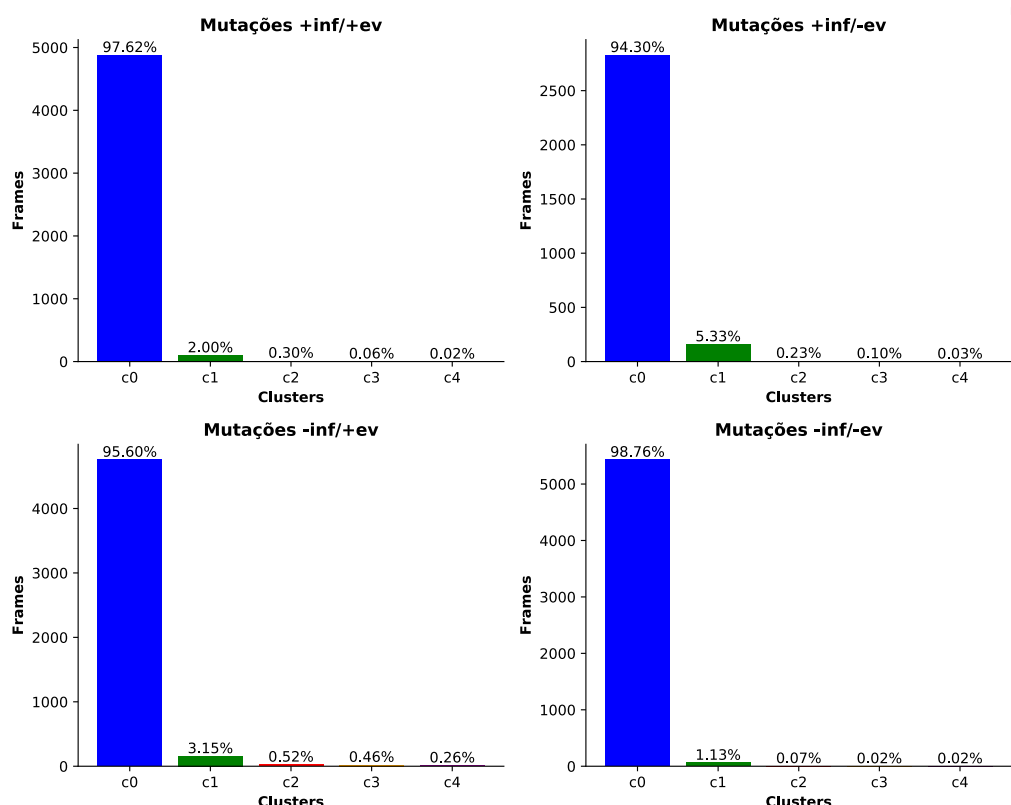


**Figura 5.21:** Análise de componentes principais para cada fenotípico na região 360-374 para os PCs 1, 2, e 3. Os círculos sobrepostos à superfície de energia livre correspondem à projeção dos frames das trajetórias de DM para cada grupo fenotípico (+inf/+ev (A), +inf/-ev (B), -inf/+ev (C) e -inf/-ev (D)).

Como mostrado na Figura 5.22, as conformações c0 apresentaram a maior representatividade: 97,62% para as mutações +inf/+ev, 94,93% para as +inf/-ev, 95,60% para as mutações -inf/+ev e 98,76% para as mutações -inf/-ev. As demais conformações possuem representatividade inferior a 6% para todas as mutantes.

Os *clusters* foram projetados nas superfícies de energia livre, onde é possível observar que as projeções das mutantes com aumento na infectividade ao receptor (+inf/+ev e

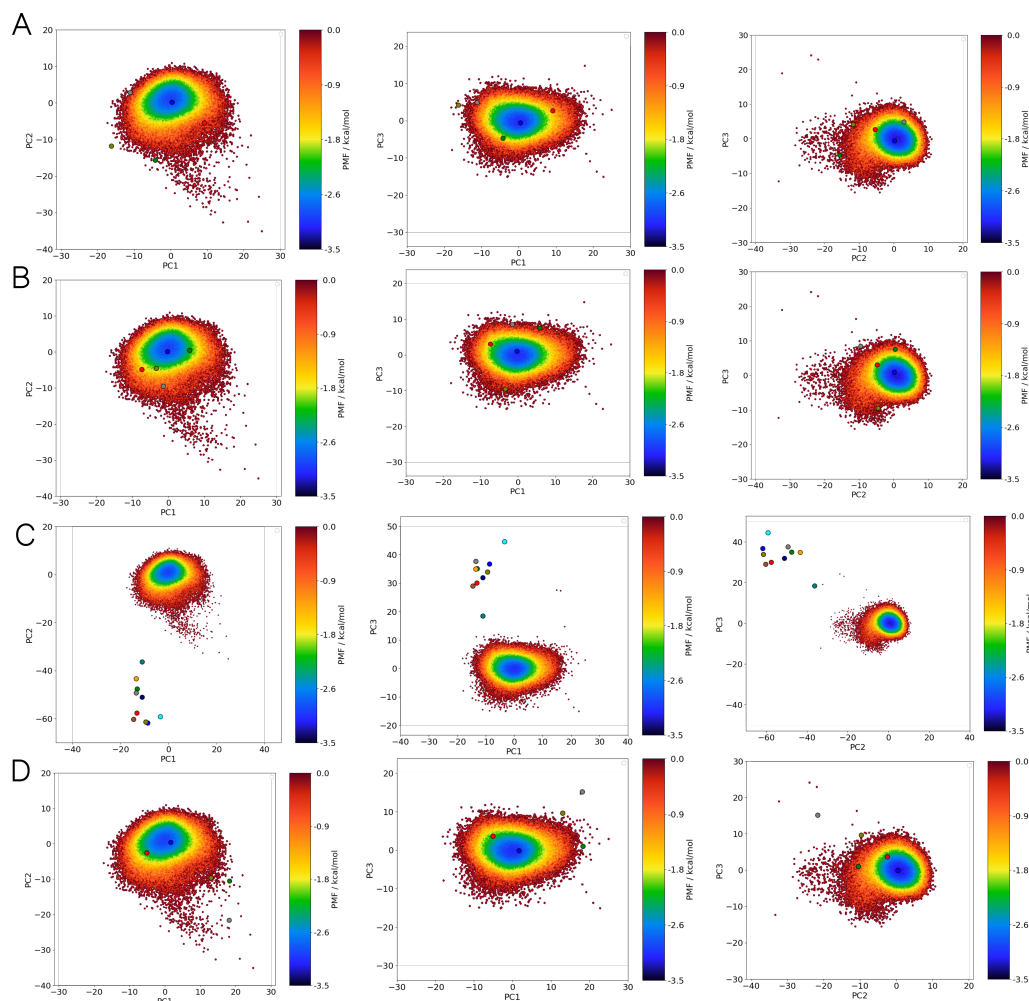




**Figura 5.22:** Os 5 clusters (c0-azul, c1-verde, c2-vermelho, c3-cinza e c4-oliva) mais populosos de cada trajetória com suas respectivas representatividade em porcentagem para cada sistema +inf/+ev, +inf/-ev, -inf/+ev e -inf/-ev. Esses *clusters* serão utilizados em análises posteriores.

+inf/-ev) estão localizadas nas regiões de maior frequência em comparação com as demais (Figura 5.23A-B). Por outro lado, para as mutações que aumentaram a evasividade ao sistema imune (Figura 5.23C), observa-se que estão situadas em regiões adjacentes às de menor frequência, indicando uma maior mobilidade. Isso pode ocorrer devido à evasividade ao sistema imune, especialmente na região do loop  $\beta_{1'}/\beta_{2'}$  – C. As mutações -inf/-ev apresentaram projeções tanto na região de maior quanto na de menor frequência.

A partir das análises realizadas RMSF e PCA para cada fenotípico, foram selecionadas as conformações mais representativas (c0 e c1) para serem utilizados nos cálculos de descritores quânticos. Esses cálculos visam identificar padrões entre as diferentes variantes, como dureza local, reatividade química, nucleofilicidade e eletrofilicidade, e serão realizados utilizando o software PRIMoRDIA.<sup>[105]</sup> Esse software permite calcular uma série de descritores de reatividade para sistemas de muitos átomos, como as proteínas.



**Figura 5.23:** Análise de componentes principais para cada fenotípico na região 360-374 para os PCs 1, 2, e 3. Os círculos sobrepostos à superfície de energia livre correspondem à projeção dos frames das trajetórias de DM para cada grupo fenotípico (+inf/+ev (A), +inf/-ev (B), -inf/+ev (C) e -inf/-ev (D)).

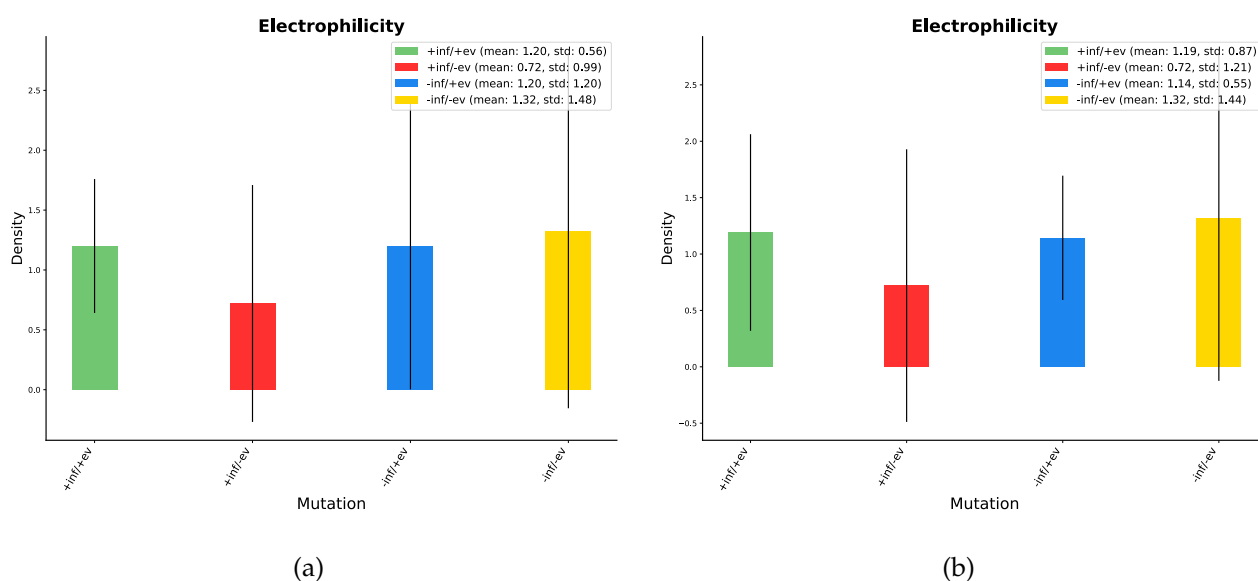
### 5.1.2.2 Descritores de Reatividade no Domínio de Ligação ao Receptor

Os DR para os resíduos de aminoácidos no domínio de ligação ao receptor para as conformações alinhadas pelas regiões 360-374 e RBM, foram calculados e usados na análise gráfica com a finalidade de encontrar padrões para as mutações e variantes do SARS-CoV-2 e assim compreendermos melhor o impacto dessas mutações na estrutura eletrônica do RBD dessa proteína.

A radicalidade, a carga parcial, a dureza local (potencial elétron-elétron) e o potencial eletrostático molecular são descritores que apresentaram a mesma distribuição de valores em todas as mutações quando comparamos as conformações alinhadas pelos resíduos 360-374 e pelo RDM. Contudo, mesmo apresentando valores similares, as mutações -inf/+ev apresentaram média e desvio padrão maiores para os descritores de radicalidade e de

potencial eletrostático molecular, conforme mostra a Figura 7.5 no apêndice.

A Figura 5.24 apresenta a média e o desvio padrão para a eletrofilicidade, que é a susceptibilidade ao ataque nucleofílico. É possível observar que as mutações +inf/-ev ou -inf/+ev são as que sofrem as maiores variações de acordo com o desvio padrão. As mutações -inf/+ev apresentam os maiores valores de eletrofilicidade nas conformações alinhadas pela região dos resíduos 360-374. Já quando alinharmos pelos resíduos do RBM, ocorreu uma diminuição da eletrofilicidade para essas mutações, ou seja, diminuindo a capacidade dessas mutações de serem atacadas por um nucleófilo.

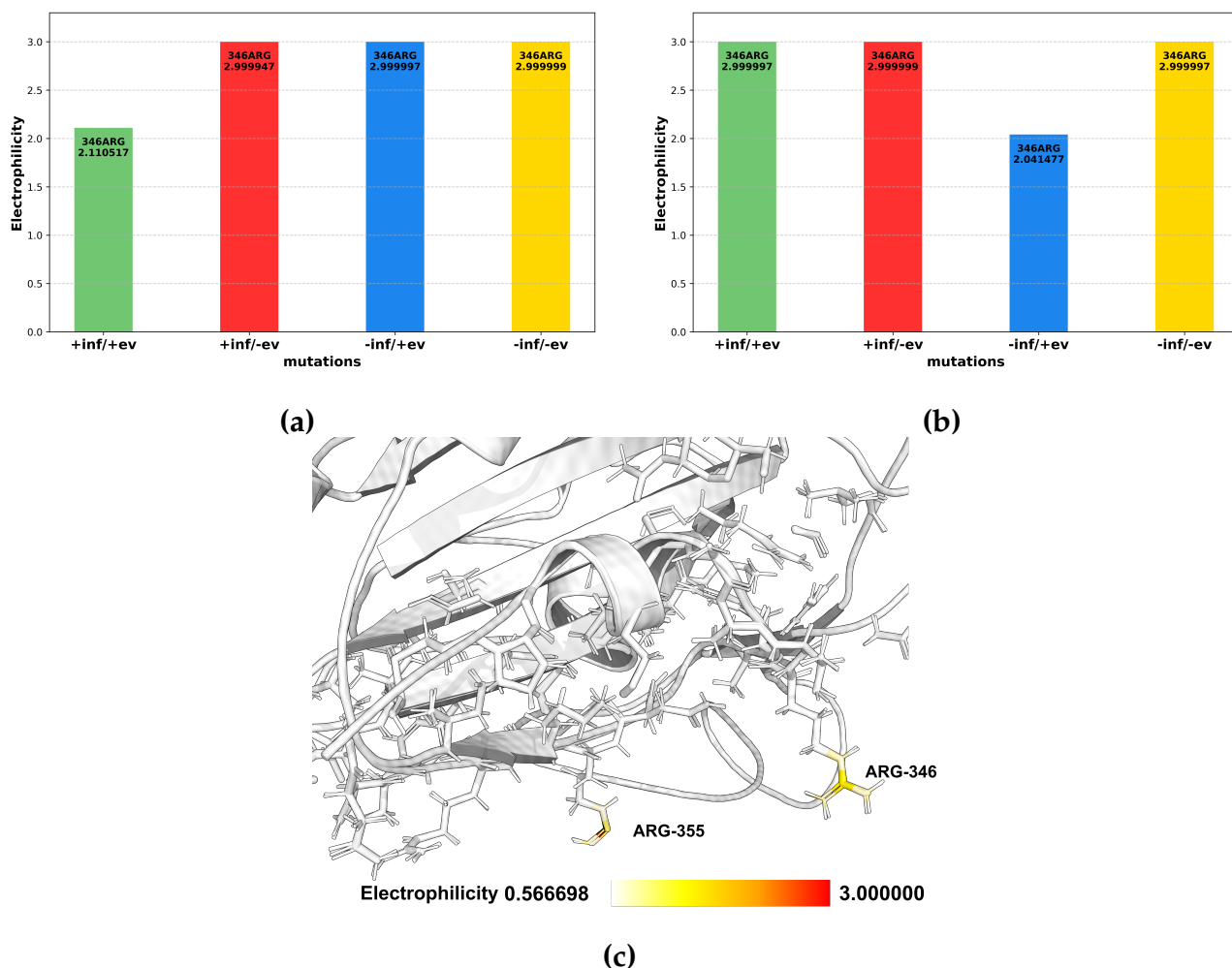


**Figura 5.24:** Media e desvio padrão da eletrofilicidade (a) região 360-374, (b) região do RBM.

As mutações +inf/+ev apresentaram o oposto do que foi observado para as mutações -inf/+ev, exibindo valores maiores de desvio padrão para as conformações alinhadas pelos resíduos RBM e menores quando alinhadas pelos resíduos 360-374, conforme mostrado na Figura 5.24. Para as mutações +inf/-ev observamos uma variação maior nas conformações alinhadas pelo RBM e em -inf/-ev o valor da eletrofilicidade é praticamente o mesmo para as duas regiões.

Para o descritor de eletrofilicidade, a arginina-346 foi o resíduo que apresentou os maiores valores em todas as mutações. Contudo, para aquelas que aumentam a evasividade ao sistema imune, houve uma variação maior na eletrofilicidade para a arginina. O estudo de Allison et al.<sup>[60]</sup> mostra que as mutações no local R346 afetam significativamente alguns anticorpos analisados no trabalho publicado, aumentando a evasividade a esse anticorpos. Nossos resultados indicam que outras mutações no RBD também podem influenciar a eletrofilicidade desse resíduo, sobretudo para as mutações com maior evasividade ao sistema

imunológico, que mostraram uma variação maior para esse descritor, conforme observado na Figura 5.25 e na tabela nos apêndices Tabela 7.1.

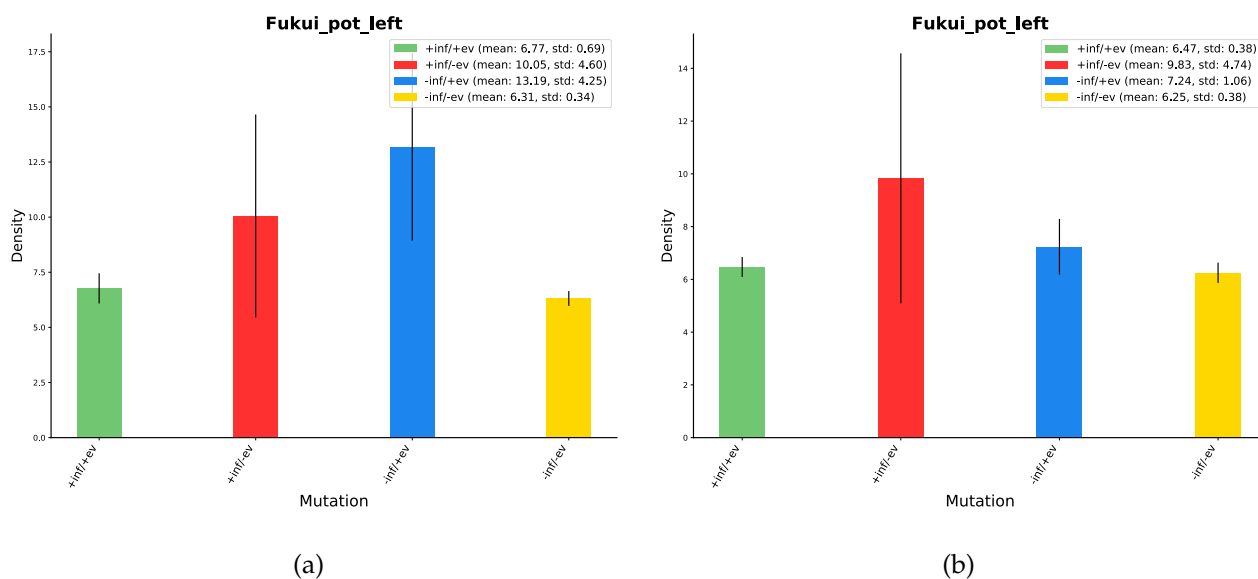


**Figura 5.25:** Eletrofilicidade por resíduos de aminoácidos (a) 360-374, (b) RBM, (c) Eletrofilicidade.

O descritor Dureza local (potencial Fukui) Figura 5.26 mostra que uma variação maior para as mutações +inf/-ev e -inf/+ev. As conformações alinhadas pela região 360-374 apresentaram valores mais elevados para as mutações -inf/+ev, enquanto que, para a região do RBM, os valores foram maiores para as mutações +inf/-ev. Isso indica que as mutações associadas à evasividade ao sistema imune são menos propensas a perder elétrons e, consequentemente, não são bons centros eletrofílicos.

Esse descritor, junto com outros que também caracterizam a dureza local, pode revelar informações importantes sobre a variação na reatividade local do RBD, mais precisamente do RBM. A análise combinada desses descritores pode ajudar a entender melhor como as mutações afetam a reatividade e a função da proteína, fornecendo *insights* sobre possíveis

mecanismos de evasão imunológica e/ou estabilidade estrutural da proteína.

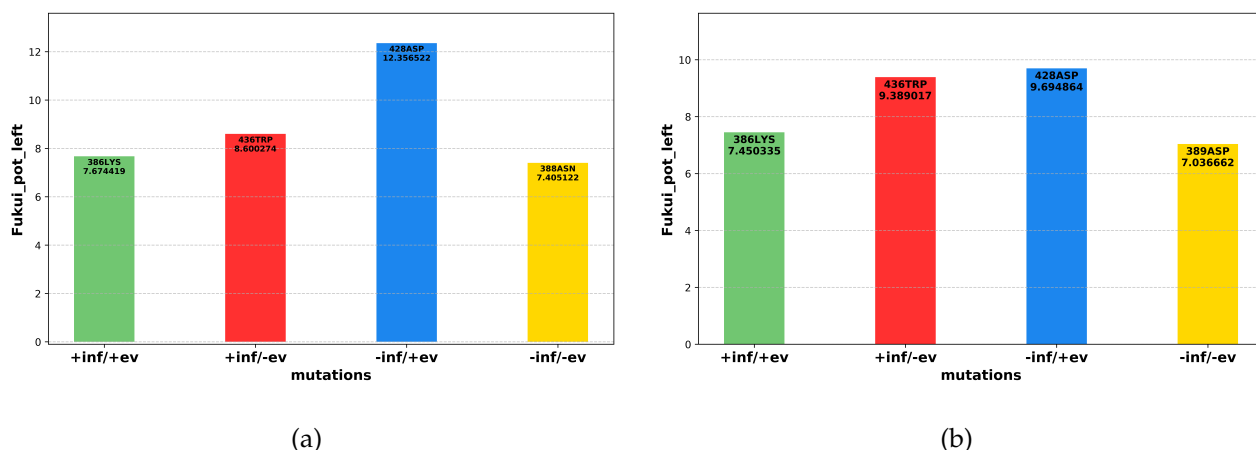


**Figura 5.26:** Media e desvio padrão para a Dureza local (potencial Fukui) (a) região 360-374, (b) região do RBM.

Os resíduos que apresentaram os maiores valores para esse descritor nas mutações +inf/-ev e -inf/+ev foram o triptofano-436 e o ácido aspártico-428, respectivamente. Esses dois resíduos podem desempenhar papéis diferentes no RBD. Enquanto o triptofano-436, que tem um papel chave na interação RBD com o ACE2,<sup>[160]</sup> possui valores mais altos para as mutações com maior infectividade. O ácido aspártico-428 está relacionado a interações eletrostáticas importantes para a conformação do RBD, conforme podemos observar na figura Figura 5.27, os valores são mais altos para a mutação com maior evasividade ao sistema imune. Esses aminoácidos são importantes no processo de infectividade e replicação do vírus.

Os demais descritores de dureza local mostram que as mutações -inf/+ev são menos reativas do que as demais mutações. Os descritores de dureza local indicam que o SARS-CoV-2 realiza essas mutações justamente para ganhar estabilidade eletrônica, sobretudo no RBM, e com isso, garantindo resistência e evasividade ao sistema imune, tornam-se menos propensas a reagir com outras moléculas ou inibidores.

As mutações +inf/-ev apresentam comportamento contrário ao observado para as mutações -inf/+ev, ou seja, são mais reativas, têm uma alta tendência a doar elétrons e, portanto, apresentam maior infectividade. De forma geral, temos que as mutações -inf/+ev e +inf/+ev são as mais duras, enquanto que as mutações +inf/-ev são mais moles. As mutações -inf/-ev não apresentam muita variação em relação as conformações estudadas.



**Figura 5.27:** Dureza local (potencial Fukui) por resíduos de aminoácidos (a) 360-374, (b) RBM.

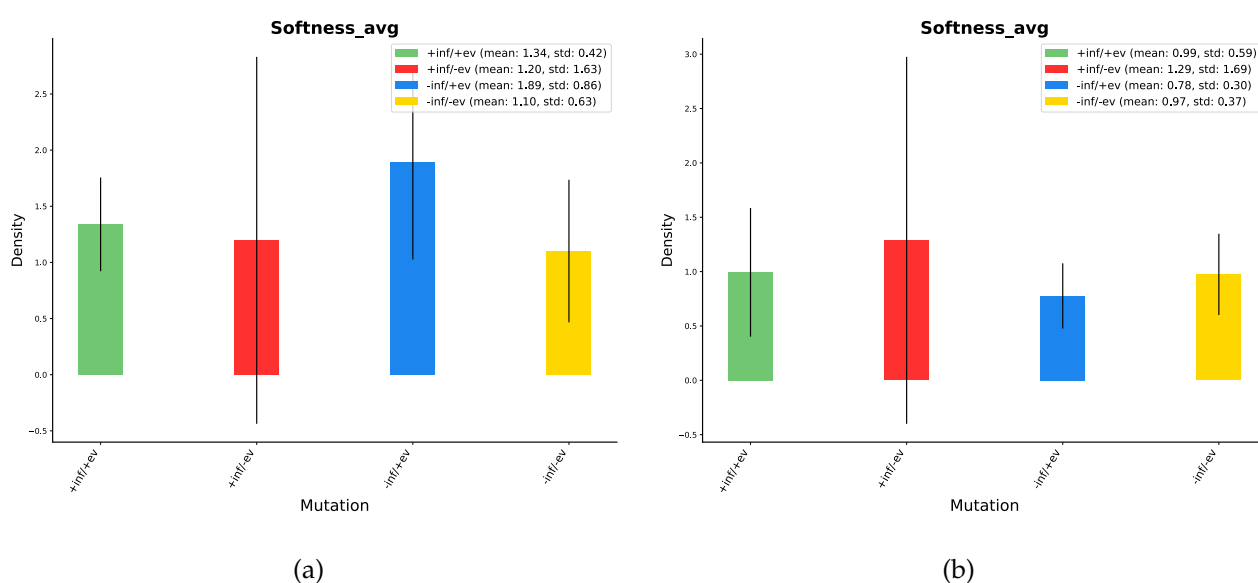
triptofano-436, conforme podemos observar seus respectivos valores na Tabela 5.1. Ao analisar cada mutação individualmente, nota-se que algumas apresentam valores maiores para a dureza local. As mutações -inf/+ev e +inf/+ev mostram valores mais altos para as conformações alinhadas pela região 360-374, sobretudo para as mutações R346K-E484K-N501Y em +inf/+ev e R346K em -inf/+ev. Os valores da dureza local aumentaram para todas essas mutações nas conformações alinhadas pelo RBM. Enquanto isso, a mutação N354D apresentou valores maiores nas condições +inf/-ev, possivelmente desfavorecendo a infectividade com o receptor, enquanto as mutações N501S, N501Y e S477N são as que mais favorecem o ganho de infectividade nas mutações +inf/-ev dessa protease.

**Tabela 5.1:** Valores da dureza local (potencial químico) para triptofano-436 nas diferentes mutações.

+inf/+ev			+inf/-ev			-inf/+ev			-inf/-ev		
Mutação	360-374	RBM	Mutação	360-374	RBM	Mutação	360-374	RBM	Mutação	360-374 (353-TRP)	RBM
E484K-N501Y	0.027991	0.049735	N354D	0.081667	0.056256	E484K	0.021276	0.048945	K417Y	0.050049	0.052293
E484Q	0.035137	0.051922	N501S	0.040624	0.036368	F490L	0.026078	0.053477	N439R	0.048821	0.044902
K417N-E484K-N501Y	0.034477	0.054159	N501Y	0.039487	0.035803	F490S	0.026160	0.054968	Q498I	0.053737	0.051514
K417T-E484K-N501Y	0.032565	0.054516	S477N	0.042188	0.034970	G446V	0.026947	0.055688	Q498L	0.054145	0.050380
L452Q	0.036552	0.055424	T478K	0.033100	0.052647	K417N	0.027983	0.057694	Q498V	0.053440	0.050266
L452R	0.032877	0.040062	V367F	0.041775	0.035796	K417T	0.034667	0.057816	S494K	0.047488	0.044890
L452R-E484Q	0.032954	0.044424				L452Q-F490S	0.025721	0.058339	S494Q	0.053577	0.050348
L452R-T478K	0.028498	0.042774				N439K	0.024066	0.051402	V445L	0.053420	0.056892
R346K-E484K-N501Y	0.044775	0.056185				R346K	0.047536	0.070283	wt	0.054266	0.050215
S477N-E484K	0.030480	0.047692				Y508H	0.035092	0.057162	Y489F	0.054211	0.053026
									Y505F	0.057513	0.055803

Para o descritor de moleza local média, foi possível observar pela Figura 5.28 que as

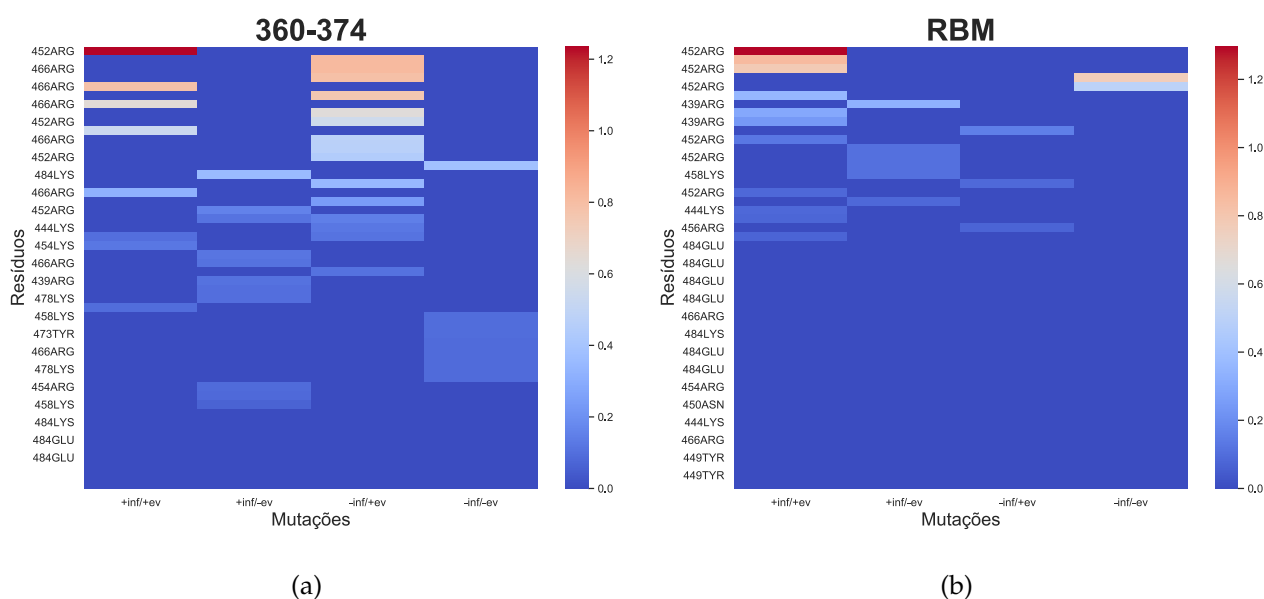
mutações  $-inf/+ev$  e  $+inf/+ev$  apresentam uma moleza local relativamente alta para as conformações alinhadas pela região 360-374. No entanto, a moleza local diminui significativamente para essas mutações nas conformações alinhadas pelo RBM, mostrando que essas mutações são mais duras, ou seja, menos polarizáveis e, portanto, apresentam resistência em variar o seu número de elétrons. Já, as mutações  $+inf/-ev$  são mais moles, mais polarizáveis e, portanto, mais suscetível a ataques nucleofílicos. A Moleza Dual Local e a Hiper Moleza Local que é a derivada da moleza local em relação ao número de elétrons apresentaram resultados similares aos da Moleza Local média, conforme mostra a Figura 7.6 nos apêndices.



**Figura 5.28:** Media e desvio padrão para a moleza local (a) região 360-374, (b) região do RBM.

Avaliar a moleza local dos resíduos do RBM pode nos fornecer informações importantes sobre a reatividade e interação do RBD com a ACE2. A Figura 5.29 mostra os resíduos que são mais moles e mais suscetíveis a ataques nucleofílicos. Resíduos como Arginina, Lisina, Glutamato, Tirosina e Asparagina apresentaram as maiores diferenças em sua distribuição ao comparar os quatro tipos de mutações. Esses resíduos apresentaram maior moleza local para as mutações  $+inf/+ev$ . Valores significativos também foram encontrados para as mutações  $-inf/+ev$ , mais precisamente para a Arginina-466. Esses resultados revelam que o tipo de resíduo pode desempenhar um papel importante, embora secundário, na descrição da variância dos dados.

Outros descritores de nucleofilicidade, como a Localização de Banda de Fronteira e o Potencial Fukui Nucleofílico, apresentaram uma mudança sensível na densidade eletrônica



**Figura 5.29:** Resíduos do RBM com maior moleza local (a) região 360-374, (b) região do RBM.

das mutações +inf/-ev, indicando uma maior reatividade geral para essas mutações. Através dos resultados aqui apresentados, foi demonstrado que as propriedades eletrônicas e de reatividade calculadas para as mutações são significativamente diferentes para cada tipo de mutação e que mutações específicas no RBD podem ter efeitos variados, desde maior reatividade química até alteração na evasividade ao sistema imune.

As diferenças expostas por meio da investigação dos descritores de reatividade abrem várias questões de pesquisa sobre como essas mutações podem levar a mudanças em suas propriedades de reatividade nas mais diversas variantes do SARS-CoV-2.



A busca por inibidores das proteínas do SARS-CoV-2 com atividade biológica *in vivo* é um desafio contínuo. Embora haja um interesse crescente e pesquisas dedicadas a desvendar as interações proteína-inibidor, ainda existem lacunas a serem preenchidas no entendimento do comportamento conformacional das proteínas do vírus. Essas lacunas levantam questões que demandam uma compreensão mais aprofundada da dinâmica das principais proteínas do vírus. Mesmo com uma quantidade significativa de informações disponíveis, há a necessidade de avanços para compreender e esclarecer esses comportamentos, visando ao desenvolvimento racional de fármacos.

Para enfrentar esse desafio, as técnicas de simulação molecular surgem como uma ferramenta valiosa, proporcionando uma abordagem ágil e econômica para a compreensão da dinâmica conformacional da *Papain-Like protease* e do domínio RBD da proteína *Spike*.

As simulações de dinâmica molecular foram utilizadas para investigar e comparar a influência do zinco no comportamento dinâmico da PLpro. As análises comparativas da dinâmica das SCoV1-PLpro e SCoV2-PLpro, com e sem zinco, revelaram um aumento nas mudanças estruturais, principalmente no Ubl e no domínio do dedo de zinco. Observamos que o zinco desempenha um papel na estabilização do domínio do dedo de zinco, porém, ao mesmo tempo, aumenta a flutuação no Ubl. Essa dinâmica pode influenciar na capacidade de capturar substratos relacionados à resposta imunológica do nosso organismo. Além disso, notamos que quando a PLpro está associada ao zinco, a alça BL2, localizada no sítio ativo da proteína, apresenta conformações mais fechadas, o que pode afetar a interação entre a proteína e o inibidor.

Como parte do esforço para compreender o comportamento dinâmico da proteína e a influência dos domínios do dedo de zinco e do Ubl, este estudo também incluiu análises de Componentes Principais. Essas análises ofereceram uma explicação coerente para as

mudanças conformacionais, pois identificaram regiões com altas flutuações, principalmente nos resíduos do domínio do dedo de zinco e no Ubl ao longo das simulações de Dinâmica Molecular. Os resultados da PCA indicaram que a PLpro com zinco apresenta movimentos menos abrangentes durante a simulação, enquanto a PLpro sem zinco exibe movimentos mais amplos. Os descritores de reatividade mostraram que a PLpro do SARS-CoV-2 é mais reativa com bases nos descritores usados que descrevem interações de dureza local e o Potencial eletrostático molecular.

Paralelamente, através das simulações de dinâmica molecular com as diferentes mutações no RBD da proteína *Spike* verificou-se que existem regiões com maiores flutuações localizadas nas regiões dos resíduos 360 à 374 e na região do RBM. Na região dos resíduos 360 à 374 houve um desvio padrão maior para as mutações que possuem maior afinidade de interação com o receptor celular ACE2 e maior resistência aos anticorpos (+inf/+ev) e a maior afinidade de interação com o receptor celular ACE2 e menor resistência aos anticorpos (+inf/-ev). Enquanto, na região do RBM houve um desvio padrão maior para as mutações possuem menor afinidade de interação com o receptor celular ACE2 e maior resistência aos anticorpos (-inf/+ev).

Os descritores de reatividade mostraram que as reatividades calculadas para as mutações são significativamente diferentes para cada tipo de mutação. Além disso, as mutações +inf/-ev apresentam uma reatividade geral maior em relação aos descritores usados, que descrevem interações do tipo mole-mole (moleza local) ou duro-duro (dureza local), enquanto que, as mutação -inf/+ev são menos reativas.

Portanto, a incorporação de estudos que busquem compreender características chaves das duas enzimas e tratar as interações e estruturas eletrônicas pode ser um fator importante na compreensão da funcionalidade das proteínas fundamentais para a infecção e replicação do SARS-CoV-2. Nesse contexto, métodos e ferramentas computacionais que usem a mecânica molecular e mecânica quântica podem contribuir para avanços a fim de descobrir as implicações no comportamento conformacional dessas proteínas.

## Referências Bibliográficas

- [1] ZHANG, Q.; XIANG, R.; HUO, S.; ZHOU, Y.; JIANG, S.; WANG, Q.; YU, F. Molecular mechanism of interaction between SARS-CoV-2 and host cells and interventional therapy. *Signal Transduction and Targeted Therapy*, v. 6, n. 1, June 2021.
- [2] RUT, W.; LV, Z.; ZMUDZINSKI, M.; PATCHETT, S.; NAYAK, D.; SNIPAS, S. J.; OUALID, F. E.; HUANG, T. T.; BEKES, M.; DRAG, M.; OLSEN, S. K. Activity profiling and crystal structures of inhibitor-bound sars-cov-2 papain-like protease: A framework for anti-COVID-19 drug design. *Science Advances*, v. 6, n. 42, Oct. 2020.
- [3] WANG, Q.; ZHANG, Y.; WU, L.; NIU, S.; SONG, C.; ZHANG, Z.; LU, G.; QIAO, C.; HU, Y.; YUEN, K.-Y.; WANG, Q.; ZHOU, H.; YAN, J.; QI, J. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*, v. 181, n. 4, p. 894–904.e9, May 2020.
- [4] DA SILVA, S. J. R.; DA SILVA, C. T. A.; GUARINES, K. M.; MENDES, R. P. G.; PARDEE, K.; KOHL, A.; PENA, L. Clinical and laboratory diagnosis of SARS-CoV-2, the virus causing COVID-19. *ACS Infectious Diseases*, v. 6, n. 9, p. 2319–2336, Aug. 2020.
- [5] ROTH, G. A.; ABATE, D.; ET AL. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, v. 392, n. 10159, p. 1736–1788, Nov. 2018.
- [6] MORENS, D. M.; FAUCI, A. S. Emerging infectious diseases: Threats to human health and global stability. *PLoS Pathogens*, v. 9, n. 7, p. e1003467, July 2013.

- [7] VELAVAN, T. P.; MEYER, C. G. The COVID-19 epidemic. *Tropical Medicine & International Health*, v. 25, n. 3, p. 278–280, Feb. 2020.
- [8] DELIKHOON, M.; GUZMAN, M. I.; NABIZADEH, R.; BAGHANI, A. N. Modes of transmission of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) and factors influencing on the airborne transmission: A review. *International Journal of Environmental Research and Public Health*, v. 18, n. 2, p. 395, Jan. 2021.
- [9] WORLD HEALTH ORGANIZATION, O. Who coronavirus (covid-19), 2020. <https://www.who.int/> [Accessed: 2022-02-17].
- [10] SINGH, D.; YI, S. V. On the origin and evolution of SARS-CoV-2. *Experimental & Molecular Medicine*, v. 53, n. 4, p. 537–547, Apr. 2021.
- [11] KHAILANY, R. A.; SAFDAR, M.; OZASLAN, M. Genomic characterization of a novel SARS-CoV-2. *Gene Reports*, v. 19, p. 100682, June 2020.
- [12] SATARKER, S.; NAMPOOTHIRI, M. Structural proteins in severe acute respiratory syndrome coronavirus-2. *Archives of Medical Research*, v. 51, n. 6, p. 482–491, Aug. 2020.
- [13] YADAV, R.; CHAUDHARY, J. K.; JAIN, N.; CHAUDHARY, P. K.; KHANRA, S.; DHAMMIJA, P.; SHARMA, A.; KUMAR, A.; HANDU, S. Role of structural and non-structural proteins and therapeutic targets of SARS-CoV-2 for COVID-19. *Cells*, v. 10, n. 4, p. 821, Apr. 2021.
- [14] KUMAR, S.; SAXENA, S. K. Structural and molecular perspectives of SARS-CoV-2. *Methods*, v. 195, p. 23–28, Nov. 2021.
- [15] KHODADADI, E.; MAROUFI, P.; KHODADADI, E.; ESPOSITO, I.; GANBAROV, K.; ESPOSITO, S.; YOUSEFI, M.; ZEINALZADEH, E.; KAFIL, H. S. Study of combining virtual screening and antiviral treatments of the sars-CoV-2 (covid-19). *Microbial Pathogenesis*, v. 146, p. 104241, Sept. 2020.
- [16] BARRETTO, N.; JUKNELIENE, D.; RATIA, K.; CHEN, Z.; MESECAR, A. D.; BAKER, S. C. The papain-like protease of severe acute respiratory syndrome coronavirus has deubiquitinating activity. *Journal of Virology*, v. 79, n. 24, p. 15189–15198, Dec. 2005.

- [17] BÁEZ-SANTOS, Y. M.; JOHN, S. E. S.; MESECAR, A. D. The sars-coronavirus papain-like protease: Structure, function and inhibition by designed antiviral compounds. *Antiviral Research*, v. 115, p. 21–38, Mar. 2015.
- [18] YANG, H.; RAO, Z. Structural biology of SARS-CoV-2 and implications for therapeutic development. *Nature Reviews Microbiology*, v. 19, n. 11, p. 685–700, Sept. 2021.
- [19] HERSHKO, A.; CIECHANOVER, A. The ubiquitin system. *Annual Review of Biochemistry*, v. 67, n. 1, p. 425–479, June 1998.
- [20] BETT, J. S. Proteostasis regulation by the ubiquitin system. *Essays in Biochemistry*, v. 60, n. 2, p. 143–151, Oct. 2016.
- [21] LIU, W.; TANG, X.; QI, X.; FU, X.; GHIMIRE, S.; MA, R.; LI, S.; ZHANG, N.; SI, H. The ubiquitin conjugating enzyme: An important ubiquitin transfer platform in ubiquitin-proteasome system. *International Journal of Molecular Sciences*, v. 21, n. 8, p. 2894, Apr. 2020.
- [22] LIU, G.; LEE, J.-H.; PARKER, Z. M.; ACHARYA, D.; CHIANG, J. J.; VAN GENT, M.; RIEDL, W.; DAVIS-GARDNER, M. E.; WIES, E.; CHIANG, C.; GACK, M. U. ISG15-dependent activation of the sensor MDA5 is antagonized by the SARS-CoV-2 papain-like protease to evade host innate immunity. *Nature Microbiology*, v. 6, n. 4, p. 467–478, Mar. 2021.
- [23] LEITE, W. C.; WEISS, K. L.; PHILLIPS, G.; ZHANG, Q.; QIAN, S.; TSUTAKAWA, S. E.; COATES, L.; O'NEILL, H. Conformational dynamics in the interaction of SARS-CoV-2 papain-like protease with human interferon-stimulated gene 15 protein. *The Journal of Physical Chemistry Letters*, v. 12, n. 23, p. 5608–5615, June 2021.
- [24] PITSILLOU, E.; LIANG, J.; HUNG, A.; KARAGIANNIS, T. C. Inhibition of interferon-stimulated gene 15 and lysine 48-linked ubiquitin binding to the SARS-CoV-2 papain-like protease by small molecules: In silico studies. *Chemical Physics Letters*, Amsterdam, v. 771, p. 138468, May 2021.
- [25] SHIN, D.; MUKHERJEE, R.; GREWE, D.; BOJKOVA, D.; BAEK, K.; BHATTACHARYA, A.; SCHULZ, L.; WIDERA, M.; MEHDIPOUR, A. R.; TASCHER, G.; GEURINK, P. P.; WILHELM, A.; VAN DER HEDEN VAN NOORT, G. J.; OVAA, H.; MÜLLER, S.; KNOBELOCH, K.-P.; RAJALINGAM, K.; SCHULMAN, B. A.; CINATL, J.; HUMMER,

- G.; CIESEK, S.; DIKIC, I. Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature*, London, v. 587, n. 7835, p. 657–662, July 2020.
- [26] SHIN, D.; MUKHERJEE, R.; GREWE, D.; BOJKOVA, D.; BAEK, K.; BHATTACHARYA, A.; SCHULZ, L.; WIDERA, M.; MEHDIPOUR, A. R.; TASCHER, G.; GEURINK, P. P.; VAN DER HEDEN VAN NOORT, G. J.; OVAA, H.; KNOBELOCH, K.-P.; RAJALINGAM, K.; SCHULMAN, B. A.; CINATL, J.; HUMMER, G.; CIESEK, S.; DIKIC, I. Inhibition of papain-like protease PLpro blocks SARS-CoV-2 spread and promotes anti-viral immunity. May 2020.
- [27] GAO, X.; QIN, B.; CHEN, P.; ZHU, K.; HOU, P.; WOJDYLA, J. A.; WANG, M.; CUI, S. Crystal structure of SARS-CoV-2 papain-like protease. *Acta Pharmaceutica Sinica B*, v. 11, n. 1, p. 237–245, Jan. 2021.
- [28] KLEMM, T.; EBERT, G.; CALLEJA, D. J.; ALLISON, C. C.; RICHARDSON, L. W.; BERNARDINI, J. P.; LU, B. G.; KUCHEL, N. W.; GROHMANN, C.; SHIBATA, Y.; GAN, Z. Y.; COONEY, J. P.; DOERFLINGER, M.; AU, A. E.; BLACKMORE, T. R.; VAN DER HEDEN VAN NOORT, G. J.; GEURINK, P. P.; OVAA, H.; NEWMAN, J.; RIBOLDI-TUNNICLIFFE, A.; CZABOTAR, P. E.; MITCHELL, J. P.; FELTHAM, R.; LECHTENBERG, B. C.; LOWES, K. N.; DEWSON, G.; PELLEGRINI, M.; LESSENE, G.; KOMANDER, D. Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.*, v. 39, n. 18, p. e106275, Sept. 2020.
- [29] ARYA, R.; KUMARI, S.; PANDEY, B.; MISTRY, H.; BIHANI, S. C.; DAS, A.; PRASHAR, V.; GUPTA, G. D.; PANICKER, L.; KUMAR, M. Structural insights into SARS-CoV-2 proteins. *Journal of Molecular Biology*, v. 433, n. 2, p. 166725, Jan. 2021.
- [30] LAN, J.; GE, J.; YU, J.; SHAN, S.; ZHOU, H.; FAN, S.; ZHANG, Q.; SHI, X.; WANG, Q.; ZHANG, L.; WANG, X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, London, v. 581, n. 7807, p. 215–220, Mar. 2020.
- [31] FINKELSTEIN, M. T.; MERMELSTEIN, A. G.; MILLER, E. P.; SETH, P. C.; STANCOFSKI, E.-S. D.; FERA, D. Structural analysis of neutralizing epitopes of the SARS-CoV-2 spike to guide therapy and vaccine design strategies. *Viruses*, v. 13, n. 1, p. 134, Jan. 2021.
- [32] YANG, Y.; DU, L. SARS-CoV-2 spike protein: a key target for eliciting persistent neutralizing antibodies. *Signal Transduction and Targeted Therapy*, v. 6, n. 1, Feb. 2021.

- [33] STERNBERG, A.; NAUJOKAT, C. Structural features of coronavirus SARS-CoV-2 spike protein: Targets for vaccination. *Life Sciences*, v. 257, p. 118056, Sept. 2020.
- [34] BERGER, I.; SCHAFFITZEL, C. The SARS-CoV-2 spike protein: balancing stability and infectivity. *Cell Research*, v. 30, n. 12, p. 1059–1060, Nov. 2020.
- [35] SHANG, J.; YE, G.; SHI, K.; WAN, Y.; LUO, C.; AIHARA, H.; GENG, Q.; AUERBACH, A.; LI, F. Structural basis of receptor recognition by SARS-CoV-2. *Nature*, London, v. 581, n. 7807, p. 221–224, Mar. 2020.
- [36] YAN, R.; ZHANG, Y.; LI, Y.; XIA, L.; GUO, Y.; ZHOU, Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*, Washington, v. 367, n. 6485, p. 1444–1448, Mar. 2020.
- [37] GUPTA, D.; SHARMA, P.; SINGH, M.; KUMAR, M.; ETHAYATHULLA, A. S.; KAUR, P. Structural and functional insights into the spike protein mutations of emerging SARS-CoV-2 variants. *Cellular and Molecular Life Sciences*, v. 78, n. 24, p. 7967–7989, Nov. 2021.
- [38] NAINWAL, L. M.; SUMAN. FDA coronavirus treatment acceleration program: approved drugs and those in clinical trials. In: *Coronavirus Drug Discovery*. Elsevier, 2022. p. 249–264.
- [39] MACALINO, S. J. Y.; GOSU, V.; HONG, S.; CHOI, S. Role of computer-aided drug design in modern drug discovery. *Archives of Pharmacal Research*, v. 38, n. 9, p. 1686–1701, July 2015.
- [40] HADNI, H.; FITRI, A.; BENJELLOUN, A. T.; BENZAKOUR, M.; MCHARFI, M.; BENBRAHIM, M. Identification of terpenoids as potential inhibitors of SARS-CoV-2 (main protease) and spike (RBD) via computer-aided drug design. *Journal of Biomolecular Structure and Dynamics*, p. 1–14, Aug. 2023.
- [41] CHOWDHURY, S. M.; TALUKDER, S. A.; KHAN, A. M.; AFRIN, N.; ALI, M. A.; ISLAM, R.; PARVES, R.; MAMUN, A. A.; SUFIAN, M. A.; HOSSAIN, M. N.; HOSSAIN, M. A.; HALIM, M. A. Antiviral peptides as promising therapeutics against SARS-CoV-2. *The Journal of Physical Chemistry B*, v. 124, n. 44, p. 9785–9792, Oct. 2020.
- [42] KIM, S.; LIU, Y.; LEI, Z.; DICKER, J.; CAO, Y.; ZHANG, X. F.; IM, W. Differential

interactions between human ACE2 and spike RBD of SARS-CoV-2 variants of concern. *Journal of Chemical Theory and Computation*, v. 17, n. 12, p. 7972–7979, Dec. 2021.

- [43] PRAJAPATI, J.; PATEL, R.; RAO, P.; SARAF, M.; RAWAL, R.; GOSWAMI, D. Perceiving SARS-CoV-2 mpro and PLpro dual inhibitors from pool of recognized antiviral compounds of endophytic microbes: an in silico simulation study. *Structural Chemistry*, v. 33, n. 5, p. 1619–1643, Apr. 2022.
- [44] NOGARA, P. A.; OMAGE, F. B.; BOLZAN, G. R.; DELGADO, C. P.; ORIAN, L.; ROCHA, J. B. T. Reactivity and binding mode of disulfiram, its metabolites, and derivatives in SARS-CoV-2 PLpro: insights from computational chemistry studies. *Journal of Molecular Modeling*, v. 28, n. 11, Oct. 2022.
- [45] TREZZA, A.; IOVINELLI, D.; SANTUCCI, A.; PRISCHI, F.; SPIGA, O. An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors. *Scientific Reports*, v. 10, n. 1, Aug. 2020.
- [46] ROY, R.; JONNIYA, N. A.; PODDAR, S.; SK, M. F.; KAR, P. Unraveling the molecular mechanism of recognition of human interferon-stimulated gene product 15 by coronavirus papain-like proteases: A multiscale simulation study. *Journal of Chemical Information and Modeling*, v. 61, n. 12, p. 6038–6052, Nov. 2021.
- [47] ELSEGINY, S. A.; ANWAR, M. M. *In silico* analysis of SARS-CoV-2 papain-like protease potential inhibitors. *RSC Advances*, v. 11, n. 61, p. 38616–38631, 2021.
- [48] RAJPOOT, S.; ALAGUMUTHU, M.; BAIG, M. S. Dual targeting of 3clpro and PLpro of SARS-CoV-2: A novel structure-based design approach to treat COVID-19. *Current Research in Structural Biology*, v. 3, p. 9–18, 2021.
- [49] BOSKEN, Y. K.; CHOLKO, T.; LOU, Y.-C.; WU, K.-P.; EN A. CHANG, C. Insights into dynamics of inhibitor and ubiquitin-like protein binding in SARS-CoV-2 papain-like protease. *Frontiers in Molecular Biosciences*, v. 7, Aug. 2020.
- [50] ALAMRI, M. A.; UL QAMAR, M. T.; MIRZA, M. U.; ALQAHTANI, S. M.; FROEYEN, M.; CHEN, L.-L. Discovery of human coronaviruses pan-papain-like protease inhibitors using computational approaches. *Journal of Pharmaceutical Analysis*, v. 10, n. 6, p. 546–559, Dec. 2020.



- [51] CHEN, H.; KANG, Y.; DUAN, M.; HOU, T. Regulation mechanism for the binding between the SARS-CoV-2 spike protein and host angiotensin-converting enzyme II. *The Journal of Physical Chemistry Letters*, v. 12, n. 27, p. 6252–6261, July 2021.
- [52] SARTORE, G.; BASSANI, D.; RAGAZZI, E.; TRALDI, P.; LAPOLLA, A.; MORO, S. In silico evaluation of the interaction between ACE2 and SARS-CoV-2 spike protein in a hyperglycemic environment. *Scientific Reports*, v. 11, n. 1, Nov. 2021.
- [53] ALI, A.; VIJAYAN, R. Dynamics of the ACE2–SARS-CoV-2/SARS-CoV spike protein interface reveal unique mechanisms. *Scientific Reports*, v. 10, n. 1, Aug. 2020.
- [54] ZECH, F.; SCHNIERTSHAUER, D.; JUNG, C.; HERRMANN, A.; CORDSMEIER, A.; XIE, Q.; NCHIOUA, R.; BOZZO, C. P.; VOLCIC, M.; KOEPKE, L.; MÜLLER, J. A.; KRÜGER, J.; HELLER, S.; STENGER, S.; HOFFMANN, M.; PÖHLMANN, S.; KLEGER, A.; JACOB, T.; CONZELMANN, K.-K.; ENSER, A.; SPARRER, K. M. J.; KIRCHHOFF, F. Spike residue 403 affects binding of coronavirus spikes to human ACE2. *Nature Communications*, v. 12, n. 1, Nov. 2021.
- [55] RAHNAMA, S.; IRANI, M. A.; AMININASAB, M.; EJTEHADI, M. R. S494 o-glycosylation site on the SARS-CoV-2 RBD affects the virus affinity to ACE2 and its infectivity; a molecular dynamics study. *Scientific Reports*, v. 11, n. 1, July 2021.
- [56] FERNANDEZ-FUENTES, N.; MOLINA, R.; OLIVA, B. A collection of designed peptides to target SARS-CoV-2 spike RBD—ACE2 interaction. *International Journal of Molecular Sciences*, v. 22, n. 21, p. 11627, Oct. 2021.
- [57] CHEN, J.; WANG, R.; WANG, M.; WEI, G.-W. Mutations strengthened SARS-CoV-2 infectivity. *Journal of Molecular Biology*, v. 432, n. 19, p. 5212–5226, Sept. 2020.
- [58] GAN, H. H.; TWADDLE, A.; MARCHAND, B.; GUNSALUS, K. C. Structural modeling of the SARS-CoV-2 spike/human ACE2 complex interface can identify high-affinity variants associated with increased transmissibility. *Journal of Molecular Biology*, v. 433, n. 15, p. 167051, July 2021.
- [59] WANG, R.; CHEN, J.; GAO, K.; WEI, G.-W. Vaccine-escape and fast-growing mutations in the united kingdom, the united states, singapore, spain, india, and other COVID-19-devastated countries. *Genomics*, v. 113, n. 4, p. 2158–2170, July 2021.

- [60] GREANEY, A. J.; STARR, T. N.; GILCHUK, P.; ZOST, S. J.; BINSHTEIN, E.; LOES, A. N.; HILTON, S. K.; HUDDLESTON, J.; EGUIA, R.; CRAWFORD, K. H.; DINGENS, A. S.; NARGI, R. S.; SUTTON, R. E.; SURYADEVARA, N.; ROTHLAUF, P. W.; LIU, Z.; WHELAN, S. P.; CARNAHAN, R. H.; CROWE, J. E.; BLOOM, J. D. Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host & Microbe*, v. 29, n. 1, p. 44–57.e9, Jan. 2021.
- [61] NAMSANI, S.; PRAMANIK, D.; KHAN, M. A.; ROY, S.; SINGH, J. K. Metadynamics-based enhanced sampling protocol for virtual screening: case study for 3clpro protein for SARS-CoV-2. *Journal of Biomolecular Structure and Dynamics*, v. 40, n. 15, p. 7002–7017, Mar. 2021.
- [62] WATANABE, C.; OKIYAMA, Y.; TANAKA, S.; FUKUZAWA, K.; HONMA, T. Molecular recognition of SARS-CoV-2 spike glycoprotein: quantum chemical hot spot and epitope analyses. *Chemical Science*, v. 12, n. 13, p. 4722–4739, 2021.
- [63] KHELFAOUI, H.; HARKATI, D.; SALEH, B. A. Molecular docking, molecular dynamics simulations and reactivity, studies on approved drugs library targeting ACE2 and SARS-CoV-2 binding with ACE2. *Journal of Biomolecular Structure and Dynamics*, v. 39, n. 18, p. 7246–7262, Aug. 2020.
- [64] FERNANDES, H. S.; SOUSA, S. F.; CERQUEIRA, N. M. F. S. A. New insights into the catalytic mechanism of the SARS-CoV-2 main protease: an ONIOM QM/MM approach. *Molecular Diversity*, v. 26, n. 3, p. 1373–1381, June 2021.
- [65] ARAFET, K.; SERRANO-APARICIO, N.; LODOLA, A.; MULHOLLAND, A. J.; GONZÁLEZ, F. V.; ŚWIDEREK, K.; MOLINER, V. Mechanism of inhibition of SARS-CoV-2 m<sup>pro</sup> by n3 peptidyl michael acceptor explained by QM/MM simulations and design of new derivatives with tunable chemical reactivity. *Chemical Science*, v. 12, n. 4, p. 1433–1444, 2021.
- [66] SRIVASTAVA, A. K.; KUMAR, A.; SRIVASTAVA, H.; MISRA, N. The role of herbal plants in the inhibition of SARS-CoV-2 main protease: A computational approach. *Journal of the Indian Chemical Society*, v. 99, n. 9, p. 100640, Sept. 2022.
- [67] DE SÁ, É. R. A.; COSTA, A. N.; COSTA, R. K. M.; SOUZA, J. L.; RAMOS, R. M.; DAS C. A. LIMA, F. *In silico* study of the interactions of *pilocarpus microphyllus* imidazolic

- alkaloids with the main protease ( $m^{pro}$ ) of SARS-CoV-2. *Molecular Simulation*, v. 47, n. 1, p. 74–87, Jan. 2021.
- [68] HOGNON, C.; MARAZZI, M.; GARCÍA-IRIEPA, C. Atomistic-level description of the covalent inhibition of SARS-CoV-2 papain-like protease. *International Journal of Molecular Sciences*, v. 23, n. 10, p. 5855, May 2022.
- [69] ROCHA, R. E. O.; CHAVES, E. J. F.; FISCHER, P. H. C.; COSTA, L. S. C.; GRILLO, I. B.; DA CRUZ, L. E. G.; GUEDES, F. C.; DA SILVEIRA, C. H.; SCOTTI, M. T.; CAMARGO, A. D.; MACHADO, K. S.; WERHLI, A. V.; FERREIRA, R. S.; ROCHA, G. B.; DE LIMA, L. H. F. A higher flexibility at the SARS-CoV-2 main protease active site compared to SARS-CoV and its potentialities for new inhibitor virtual screening targeting multi-conformers. *Journal of Biomolecular Structure and Dynamics*, v. 40, n. 19, p. 9214–9234, May 2021.
- [70] GENHEDEN, S.; REYMER, A.; SAENZ-MÉNDEZ, P.; ERIKSSON, L. A. Computational chemistry and molecular modelling basics. In: *Computational Tools for Chemical Biology*. The Royal Society of Chemistry, Nov. 2017. p. 1–38.
- [71] SMITH, J. S.; NEBGEN, B. T.; ZUBATYUK, R.; LUBBERS, N.; DEVEREUX, C.; BARROS, K.; TRETIK, S.; ISAYEV, O.; ROITBERG, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nature Communications*, v. 10, n. 1, July 2019.
- [72] KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, v. 9, n. 9, p. 646–652, Sept. 2002.
- [73] NAMBA, A. M.; SILVA, V. B. D.; DE PAULA DA SILVA, C. H. T. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. *Ecletica Química*, v. 33, n. 4, p. 13–23, Dec. 2008.
- [74] VOLLMEYER-LEE, K. Introduction to molecular dynamics simulations. *American Journal of Physics*, Woodbury, v. 88, n. 5, p. 401–422, May 2020.
- [75] HOLLINGSWORTH, S. A.; DROR, R. O. Molecular dynamics simulation for all. *Neuron*, v. 99, n. 6, p. 1129–1143, Sept. 2018.
- [76] ALDER, B. J.; WAINWRIGHT, T. E. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, v. 27, n. 5, p. 1208–1209, Nov. 1957.

- [77] MCCAMMON, J. A.; GELIN, B. R.; KARPLUS, M. Dynamics of folded proteins. *Nature*, London, v. 267, n. 5612, p. 585–590, June 1977.
- [78] LIFSON, S.; WARSHHEL, A. Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and *n*-alkane molecules. *The Journal of Chemical Physics*, v. 49, n. 11, p. 5116–5129, Dec. 1968.
- [79] LEVITT, M.; LIFSON, S. Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of Molecular Biology*, v. 46, n. 2, p. 269–279, Dec. 1969.
- [80] DROR, R. O.; DIRKS, R. M.; GROSSMAN, J.; XU, H.; SHAW, D. E. Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, v. 41, n. 1, p. 429–452, June 2012.
- [81] HATA, H.; NISHIYAMA, M.; KITAO, A. Molecular dynamics simulation of proteins under high pressure: Structure, function and thermodynamics. *Biochimica et Biophysica Acta (BBA) - General Subjects*, v. 1864, n. 2, p. 129395, Feb. 2020.
- [82] GU, Y.; LI, D.-W.; BRÜSCHWEILER, R. NMR order parameter determination from long molecular dynamics trajectories for objective comparison with experiment. *Journal of Chemical Theory and Computation*, v. 10, n. 6, p. 2599–2607, May 2014.
- [83] LIU, X.; SHI, D.; ZHOU, S.; LIU, H.; LIU, H.; YAO, X. Molecular dynamics simulations and novel drug discovery. *Expert Opinion on Drug Discovery*, v. 13, n. 1, p. 23–37, Nov. 2017.
- [84] ALDER, B. J.; WAINWRIGHT, T. E. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, v. 31, n. 2, p. 459–466, Aug. 1959.
- [85] ROBUSTELLI, P.; PIANA, S.; SHAW, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences*, v. 115, n. 21, May 2018.
- [86] ALAVI, S. *The mechanics of molecular dynamics*. New York, N.Y: Wiley, jun 2020.
- [87] PONDER, J. W.; CASE, D. A. Force fields for protein simulations. In: *Protein Simulations*. Elsevier, 2003. p. 27–85.

- [88] JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, Washington, v. 118, n. 45, p. 11225–11236, Nov. 1996.
- [89] VANOMMESLAEGHE, K.; HATCHER, E.; ACHARYA, C.; KUNDU, S.; ZHONG, S.; SHIM, J.; DARIAN, E.; GUVENCH, O.; LOPES, P.; VOROBYOV, I.; MACKERELL, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, p. NA–NA, 2009.
- [90] OOSTENBRINK, C.; VILLA, A.; MARK, A. E.; GUNSTEREN, W. F. V. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53a5 and 53a6. *Journal of Computational Chemistry*, v. 25, n. 13, p. 1656–1676, July 2004.
- [91] CHENG, X.; IVANOV, I. Molecular dynamics. In: *Methods in Molecular Biology*. Humana Press, 2012. p. 243–285.
- [92] MONTICELLI, L.; TIELEMAN, D. P. Force fields for classical molecular dynamics. In: *Methods in Molecular Biology*. Humana Press, Aug. 2012. p. 197–213.
- [93] KULAKOVA, L.; ARAMPATZIS, G.; ANGELIKOPOULOS, P.; HADJIDOUKAS, P.; PAPADIMITRIOU, C.; KOUMOUTSAKOS, P. Data driven inference for the repulsive exponent of the lennard-jones potential in molecular dynamics simulations. *Scientific Reports*, v. 7, n. 1, Nov. 2017.
- [94] TUCKERMAN, M. E. *Statistical mechanics: Theory and molecular simulation*. 2. ed., Oxford Graduate Texts. London, England: Oxford University Press, Aug. 2023.
- [95] BONOMI, M.; CAMILLONI, C. (Eds.). *Biomolecular simulations*. Springer New York, 2019.
- [96] LI, H.; CHOWDHARY, J.; HUANG, L.; HE, X.; MACKERELL, A. D.; ROUX, B. Drude polarizable force field for molecular dynamics simulations of saturated and unsaturated zwitterionic lipids. *Journal of Chemical Theory and Computation*, v. 13, n. 9, p. 4535–4552, Aug. 2017.

- [97] TIAN, C.; KASAVAJHALA, K.; BELFON, K. A. A.; RAGUETTE, L.; HUANG, H.; MIGUES, A. N.; BICKEL, J.; WANG, Y.; PINCAY, J.; WU, Q.; SIMMERLING, C. ff19sb: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *Journal of Chemical Theory and Computation*, v. 16, n. 1, p. 528–552, Nov. 2019.
- [98] 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, July 2015.
- [99] ABRIATA, L. A.; PERARO, M. D. Assessment of transferable forcefields for protein simulations attests improved description of disordered states and secondary structure propensities, and hints at multi-protein systems as the next challenge for optimization. *Computational and Structural Biotechnology Journal*, v. 19, p. 2626–2636, 2021.
- [100] PETERS, M. B.; YANG, Y.; WANG, B.; FÜSTI-MOLNÁR, L.; WEAVER, M. N.; MERZ, K. M. Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *Journal of Chemical Theory and Computation*, v. 6, n. 9, p. 2935–2947, Aug. 2010.
- [101] CORDOMÍ, A.; CALTABIANO, G.; PARDO, L. Membrane protein simulations using AMBER force field and berger lipid parameters. *Journal of Chemical Theory and Computation*, v. 8, n. 3, p. 948–958, Feb. 2012.
- [102] VIJAYARAJ, R.; SUBRAMANIAN, V.; CHATTARAJ, P. K. Comparison of global reactivity descriptors calculated using various density functionals: A QSAR perspective. *Journal of Chemical Theory and Computation*, v. 5, n. 10, p. 2744–2753, Sept. 2009.
- [103] GEERLINGS, P.; CHAMORRO, E.; CHATTARAJ, P. K.; PROFT, F. D.; GÁZQUEZ, J. L.; LIU, S.; MORELL, C.; TORO-LABBÉ, A.; VELA, A.; AYERS, P. Conceptual density functional theory: status, prospects, issues. *Theoretical Chemistry Accounts*, v. 139, n. 2, Jan. 2020.
- [104] GRILLO, I. B.; BACHEGA, J. F. R.; TIMMERS, L. F. S. M.; CACERES, R. A.; DE SOUZA, O. N.; FIELD, M. J.; ROCHA, G. B. Theoretical characterization of the shikimate 5-dehydrogenase reaction from mycobacterium tuberculosis by hybrid QC/MM si-

mulations and quantum chemical descriptors. *Journal of Molecular Modeling*, v. 26, n. 11, Oct. 2020.

- [105] GRILLO, I. B.; URQUIZA-CARVALHO, G. A.; ROCHA, G. B. PRIMoRDiA: A software to explore reactivity and electronic structure in large biomolecules. *Journal of Chemical Information and Modeling*, v. 60, n. 12, p. 5885–5890, Nov. 2020.
- [106] GRILLO, I. B.; URQUIZA-CARVALHO, G. A.; BACHEGA, J. F. R.; ROCHA, G. B. Elucidating enzymatic catalysis using fast quantum chemical descriptors. *Journal of Chemical Information and Modeling*, v. 60, n. 2, p. 578–591, Jan. 2020.
- [107] ROCHA-SANTOS, A.; CHAVES, E. J. F.; GRILLO, I. B.; DE FREITAS, A. S.; ARAÚJO, D. A. M.; ROCHA, G. B. Thermochemical and quantum descriptor calculations for gaining insight into ricin toxin a (RTA) inhibitors. *ACS Omega*, v. 6, n. 13, p. 8764–8777, Mar. 2021.
- [108] RATIA, K.; SAIKATENDU, K. S.; SANTARSIERO, B. D.; BARRETTO, N.; BAKER, S. C.; STEVENS, R. C.; MESECAR, A. D. Severe acute respiratory syndrome coronavirus papain-like protease: Structure of a viral deubiquitinating enzyme. *Proceedings of the National Academy of Sciences*, v. 103, n. 15, p. 5717–5722, Apr. 2006.
- [109] OSIPIUK, J.; AZIZI, S.-A.; DVORKIN, S.; ENDRES, M.; JEDRZEJCZAK, R.; JONES, K. A.; KANG, S.; KATHAYAT, R. S.; KIM, Y.; LISNYAK, V. G.; MAKI, S. L.; NICOLAESCU, V.; TAYLOR, C. A.; TESAR, C.; ZHANG, Y.-A.; ZHOU, Z.; RANDALL, G.; MICHALSKA, K.; SNYDER, S. A.; DICKINSON, B. C.; JOACHIMIAK, A. Structure of papain-like protease from SARS-CoV-2 and its complexes with non-covalent inhibitors. *Nature Communications*, v. 12, n. 1, Feb. 2021.
- [110] WATERHOUSE, A.; BERTONI, M.; BIENERT, S.; STUDER, G.; TAURIELLO, G.; GUMIENNY, R.; HEER, F. T.; DE BEER, T. A. P.; REMPFER, C.; BORDOLI, L.; LEPORE, R.; SCHWEDE, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, v. 46, n. W1, p. W296–W303, May 2018.
- [111] SCHRÖDINGER, L.; DELANO, W. Pymol.
- [112] LÓPEZ-CORTÉS, G. I.; PALACIOS-PÉREZ, M.; ZAMUDIO, G. S.; VELEDÍAZ, H. F.; ORTEGA, E.; JOSÉ, M. V. Neutral evolution test of the spike protein of SARS-CoV-2 and its implications in the binding to ACE2. *Scientific Reports*, v. 11, n. 1, Sept. 2021.

- [113] DENG, X.; GARCIA-KNIGHT, M. A.; KHALID, M. M.; SERVELLITA, V.; WANG, C.; MORRIS, M. K.; SOTOMAYOR-GONZÁLEZ, A.; GLASNER, D. R.; REYES, K. R.; GLIWA, A. S.; REDDY, N. P.; MARTIN, C. S. S.; FEDERMAN, S.; CHENG, J.; BALCEREK, J.; TAYLOR, J.; STREITHORST, J. A.; MILLER, S.; SREEKUMAR, B.; CHEN, P.-Y.; SCHULZE-GAHMEN, U.; TAHA, T. Y.; HAYASHI, J. M.; SIMONEAU, C. R.; KUMAR, G. R.; MCMAHON, S.; LIDSKY, P. V.; XIAO, Y.; HEMARAJATA, P.; GREEN, N. M.; ESPINOSA, A.; KATH, C.; HAW, M.; BELL, J.; HACKER, J. K.; HANSON, C.; WADFORD, D. A.; ANAYA, C.; FERGUSON, D.; FRANKINO, P. A.; SHIVRAM, H.; LAREAU, L. F.; WYMAN, S. K.; OTT, M.; ANDINO, R.; CHIU, C. Y. Transmission, infectivity, and neutralization of a spike l452r SARS-CoV-2 variant. *Cell*, v. 184, n. 13, p. 3426–3437.e8, June 2021.
- [114] DI GIACOMO, S.; MERCATELLI, D.; RAKHIMOV, A.; GIORGI, F. M. Preliminary report on severe acute respiratory syndrome coronavirus 2 (sars-cov-2) spike mutation t478k. *Journal of Medical Virology*, v. 93, n. 9, p. 5638–5643, May 2021.
- [115] LI, Q.; WU, J.; NIE, J.; ZHANG, L.; HAO, H.; LIU, S.; ZHAO, C.; ZHANG, Q.; LIU, H.; NIE, L.; QIN, H.; WANG, M.; LU, Q.; LI, X.; SUN, Q.; LIU, J.; ZHANG, L.; LI, X.; HUANG, W.; WANG, Y. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell*, v. 182, n. 5, p. 1284–1294.e9, Sept. 2020.
- [116] OU, J.; ZHOU, Z.; DAI, R.; ZHANG, J.; ZHAO, S.; WU, X.; LAN, W.; REN, Y.; CUI, L.; LAN, Q.; LU, L.; SETO, D.; CHODOSH, J.; WU, J.; ZHANG, G.; ZHANG, Q. V367f mutation in SARS-CoV-2 spike RBD emerging during the early transmission phase enhances viral infectivity through increased human ACE2 receptor binding affinity. *Journal of Virology*, v. 95, n. 16, July 2021.
- [117] WILHELM, A.; TOPTAN, T.; PALLAS, C.; WOLF, T.; GOETSCH, U.; GOTTSCHALK, R.; VEHRERESCHILD, M. J. G. T.; CIESEK, S.; WIDERA, M. Antibody-mediated neutralization of authentic SARS-CoV-2 b.1.617 variants harboring l452r and t478k/e484q. *Viruses*, v. 13, n. 9, p. 1693, Aug. 2021.
- [118] MCCALLUM, M.; WALLS, A. C.; SPROUSE, K. R.; BOWEN, J. E.; ROSEN, L. E.; DANG, H. V.; MARCO, A. D.; FRANKO, N.; TILLES, S. W.; LOGUE, J.; MIRANDA, M. C.; AHLRICHS, M.; CARTER, L.; SNELL, G.; PIZZUTO, M. S.; CHU, H. Y.; VOORHIS, W. C. V.; CORTI, D.; VEESLER, D. Molecular basis of immune evasion by



the delta and kappa SARS-CoV-2 variants. *Science*, Washington, v. 374, n. 6575, p. 1621–1626, Dec. 2021.

- [119] MLCOCHOVA, P.; KEMP, S. A.; DHAR, M. S.; PAPA, G.; MENG, B.; FERREIRA, I. A. T. M.; DATIR, R.; COLLIER, D. A.; ALBECKA, A.; SINGH, S.; PANDEY, R.; BROWN, J.; ZHOU, J.; GOONAWARDANE, N.; MISHRA, S.; WHITTAKER, C.; MELLAN, T.; MARWAL, R.; DATTA, M.; SENGUPTA, S.; PONNUSAMY, K.; RADHAKRISHNAN, V. S.; ABDULLAHI, A.; CHARLES, O.; CHATTOPADHYAY, P.; DEVI, P.; CAPUTO, D.; PEACOCK, T.; WATTAL, C.; GOEL, N.; SATWIK, A.; VAISHYA, R.; AGARWAL, M.; CHAUHAN, H.; DIKID, T.; GOGIA, H.; LALL, H.; VERMA, K.; DHAR, M. S.; SINGH, M. K.; SONI, N.; MEENA, N.; MADAN, P.; SINGH, P.; SHARMA, R.; SHARMA, R.; KABRA, S.; KUMAR, S.; KUMARI, S.; SHARMA, U.; CHAUDHARY, U.; SIVASUBBU, S.; SCARIA, V.; OBEROI, J. K.; RAVEENDRAN, R.; DATTA, S.; DAS, S.; MAITRA, A.; CHINNASWAMY, S.; BISWAS, N. K.; PARIDA, A.; RAGHAV, S. K.; PRASAD, P.; SARIN, A.; MAYOR, S.; RAMAKRISHNAN, U.; PALAKODETI, D.; SESHASAYEE, A. S. N.; THANGARAJ, K.; BASHYAM, M. D.; DALAL, A.; BHAT, M.; SHOUCHE, Y.; PILLAI, A.; ABRAHAM, P.; POTDAR, V. A.; CHERIAN, S. S.; DESAI, A. S.; PATTABIRAMAN, C.; MANJUNATHA, M. V.; MANI, R. S.; UDUPI, G. A.; NANDICOORI, V.; TALLAPAKA, K. B.; SOWPATI, D. T.; KAWABATA, R.; MORIZAKO, N.; SADAMASU, K.; ASAKURA, H.; NAGASHIMA, M.; YOSHIMURA, K.; ITO, J.; KIMURA, I.; URIU, K.; KOSUGI, Y.; SUGANAMI, M.; OIDE, A.; YOKOYAMA, M.; CHIBA, M.; SAITO, A.; BUTLERTANAKA, E. P.; TANAKA, Y. L.; IKEDA, T.; MOTOZONO, C.; NASSER, H.; SHIMIZU, R.; YUAN, Y.; KITAZATO, K.; HASEBE, H.; NAKAGAWA, S.; WU, J.; TAKAHASHI, M.; FUKUHARA, T.; SHIMIZU, K.; TSUSHIMA, K.; KUBO, H.; SHIRAKAWA, K.; KAZUMA, Y.; NOMURA, R.; HORISAWA, Y.; TAKAORI-KONDO, A.; TOKUNAGA, K.; OZONO, S.; BAKER, S.; DOUGAN, G.; HESS, C.; KINGSTON, N.; LEHNER, P. J.; LYONS, P. A.; MATHESON, N. J.; OWEHAND, W. H.; SAUNDERS, C.; SUMMERS, C.; THAVENTHIRAN, J. E. D.; TOSHNER, M.; WEEKES, M. P.; MAXWELL, P.; SHAW, A.; BUCKE, A.; CALDER, J.; CANNA, L.; DOMINGO, J.; ELMER, A.; FULLER, S.; HARRIS, J.; HEWITT, S.; KENNET, J.; JOSE, S.; KOURAMPA, J.; MEADOWS, A.; O'BRIEN, C.; PRICE, J.; PUBLICO, C.; RASTALL, R.; RIBEIRO, C.; ROWLANDS, J.; RUFFOLO, V.; TORDESILLAS, H.; BULLMAN, B.; DUNMORE, B. J.; FAWKE, S.; GRÄF, S.; HODGSON, J.; HUANG, C.; HUNTER, K.; JONES, E.; LEGCHENKO, E.; MATARA, C.; MARTIN, J.; MESCIA, F.; O'DONNELL, C.; POINTON,

L.; POND, N.; SHIH, J.; SUTCLIFFE, R.; TILLY, T.; TREACY, C.; TONG, Z.; WOOD, J.; WYLOT, M.; BERGAMASCHI, L.; BETANCOURT, A.; BOWER, G.; COSSETTI, C.; SA, A. D.; EPPING, M.; FAWKE, S.; GLEADALL, N.; GRENFELL, R.; HINCH, A.; HUH, O.; JACKSON, S.; JARVIS, I.; KRISHNA, B.; LEWIS, D.; MARSDEN, J.; NICE, F.; OKECHA, G.; OMARJEE, O.; PERERA, M.; POTTS, M.; RICHOS, N.; ROMASHOVA, V.; YARKONI, N. S.; SHARMA, R.; STEFANUCCI, L.; STEPHENS, J.; STREZLECKI, M.; TURNER, L.; BIE, E. M. D. D. D.; BUNCLARK, K.; JOSIPOVIC, M.; MACKAY, M.; ROSSI, S.; SELVAN, M.; SPENCER, S.; YONG, C.; ALLISON, J.; BUTCHER, H.; CAPUTO, D.; CLAPHAM-RILEY, D.; DEWHURST, E.; FURLONG, A.; GRAVES, B.; GRAY, J.; IVERS, T.; KASANICKI, M.; GRESLEY, E. L.; LINGER, R.; MELOY, S.; MULDOON, F.; OVINGTON, N.; PAPADIA, S.; PHELAN, I.; STARK, H.; STIRRUPS, K. E.; TOWNSEND, P.; WALKER, N.; WEBSTER, J.; SCHOLTES, I.; HEIN, S.; KING, R.; MAVOUSIAN, A.; LEE, J. H.; BASSI, J.; SILACCI-FEGNI, C.; SALIBA, C.; PINTO, D.; IRIE, T.; YOSHIDA, I.; HAMILTON, W. L.; SATO, K.; BHATT, S.; FLAXMAN, S.; JAMES, L. C.; CORTI, D.; PICCOLI, L.; BARCLAY, W. S.; RAKSHIT, P.; AGRAWAL, A.; GUPTA, R. K.; ; AND. SARS-CoV-2 b.1.617.2 delta variant replication and immune evasion. *Nature*, London, v. 599, n. 7883, p. 114–119, Sept. 2021.

[120] POUELS, K. B.; PRITCHARD, E.; MATTHEWS, P. C.; STOESEER, N.; EYRE, D. W.; VIHTA, K.-D.; HOUSE, T.; HAY, J.; BELL, J. I.; NEWTON, J. N.; FARRAR, J.; CROOK, D.; COOK, D.; ROURKE, E.; STUDLEY, R.; PETO, T. E. A.; DIAMOND, I.; WALKER, A. S. Effect of delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nature Medicine*, v. 27, n. 12, p. 2127–2135, Oct. 2021.

[121] KIMURA, I.; KOSUGI, Y.; WU, J.; YAMASOBA, D.; BUTLERTANAKA, E. P.; TANAKA, Y. L.; LIU, Y.; SHIRAKAWA, K.; KAZUMA, Y.; NOMURA, R.; HORISAWA, Y.; TOKUNAGA, K.; TAKAORI-KONDO, A.; ARASE, H.; SAITO, A.; NAKAGAWA, S.; AND, K. S. SARS-CoV-2 lambda variant exhibits higher infectivity and immune resistance. July 2021.

[122] FARIA, N. R.; MELLAN, T. A.; WHITTAKER, C.; CLARO, I. M.; DA S. CANDIDO, D.; MISHRA, S.; CRISPIM, M. A. E.; SALES, F. C. S.; HAWRYLUK, I.; MCCRONE, J. T.; HULSWIT, R. J. G.; FRANCO, L. A. M.; RAMUNDO, M. S.; DE JESUS, J. G.; ANDRADE, P. S.; COLETTI, T. M.; FERREIRA, G. M.; SILVA, C. A. M.; MANULI, E. R.;

PEREIRA, R. H. M.; PEIXOTO, P. S.; KRAEMER, M. U. G.; GABURO, N.; DA C. CAMILO, C.; HOELTGEBAUM, H.; SOUZA, W. M.; ROCHA, E. C.; DE SOUZA, L. M.; DE PINHO, M. C.; ARAUJO, L. J. T.; MALTA, F. S. V.; DE LIMA, A. B.; DO P. SILVA, J.; ZAULI, D. A. G.; DE S. FERREIRA, A. C.; SCHNEKENBERG, R. P.; LAYDON, D. J.; WALKER, P. G. T.; SCHLÜTER, H. M.; DOS SANTOS, A. L. P.; VIDAL, M. S.; CARO, V. S. D.; FILHO, R. M. F.; DOS SANTOS, H. M.; AGUIAR, R. S.; PROENÇA-MODENA, J. L.; NELSON, B.; HAY, J. A.; MONOD, M.; MISCOURIDOU, X.; COUPLAND, H.; SONABEND, R.; VOLLMER, M.; GANDY, A.; PRETE, C. A.; NASCIMENTO, V. H.; SUCHARD, M. A.; BOWDEN, T. A.; POND, S. L. K.; WU, C.-H.; RATMANN, O.; FERGUSON, N. M.; DYE, C.; LOMAN, N. J.; LEMEY, P.; RAMBAUT, A.; FRAIJL, N. A.; DO P. S. S. CARVALHO, M.; PYBUS, O. G.; FLAXMAN, S.; BHATT, S.; SABINO, E. C. Genomics and epidemiology of the p.1 SARS-CoV-2 lineage in manaus, brazil. *Science*, Washington, v. 372, n. 6544, p. 815–821, May 2021.

- [123] NAVECA, F. G.; NASCIMENTO, V.; DE SOUZA, V. C.; DE LIMA CORADO, A.; NASCIMENTO, F.; SILVA, G.; COSTA, Á.; DUARTE, D.; PESSOA, K.; MEJÍA, M.; BRANDÃO, M. J.; JESUS, M.; GONÇALVES, L.; DA COSTA, C. F.; SAMPAIO, V.; BARROS, D.; SILVA, M.; MATTOS, T.; PONTES, G.; ABDALLA, L.; SANTOS, J. H.; ARANTES, I.; DEZORDI, F. Z.; SIQUEIRA, M. M.; WALLAU, G. L.; RESENDE, P. C.; DELATORRE, E.; GRÄF, T.; BELLO, G. COVID-19 in amazonas, brazil, was driven by the persistence of endemic lineages and p.1 emergence. *Nature Medicine*, v. 27, n. 7, p. 1230–1238, May 2021.
- [124] FUJINO, T.; NOMOTO, H.; KUTSUNA, S.; UJIE, M.; SUZUKI, T.; SATO, R.; FUJIMOTO, T.; KURODA, M.; WAKITA, T.; OHMAGARI, N. Novel SARS-CoV-2 variant in travelers from brazil to japan. *Emerging Infectious Diseases*, v. 27, n. 4, Apr. 2021.
- [125] BASCOS, N. A. D.; MIRANO-BASCOS, D.; SALOMA, C. P. Structural analysis of spike protein mutations in the SARS-CoV-2 p.3 variant. Mar. 2021.
- [126] HAW, N. J.; CAÑAL, E. M.; JR., J. Z.; LOECHE, M. J.; BERNADAS, J. Epidemiological characteristics of the SARS-CoV-2 theta variant (p.3) in the central visayas region, philippines, 30 october 2020–16 february 2021. *Western Pacific Surveillance and Response Journal*, v. 13, n. 1, p. 60–62, Jan. 2022.
- [127] HALFMANN, P. J.; KURODA, M.; ARMBRUST, T.; THEILER, J.; BALARAM, A.; MORENO, G. K.; ACCOLA, M. A.; IWATSUKI-HORIMOTO, K.; VALDEZ, R.; STO-

- NEMAN, E.; BRAUN, K.; YAMAYOSHI, S.; SOMSEN, E.; BACZENAS, J. J.; MITA-MURA, K.; HAGIHARA, M.; ADACHI, E.; KOGA, M.; MCLAUGHLIN, M.; REHRAUER, W.; IMAI, M.; YAMAMOTO, S.; TSUTSUMI, T.; SAITO, M.; FRIEDRICH, T. C.; O'CONNOR, S. L.; O'CONNOR, D. H.; GORDON, A.; KORBER, B.; KAWAOKA, Y. Characterization of the SARS-CoV-2 b.1.621 (mu) variant. *Science Translational Medicine*, v. 14, n. 657, Aug. 2022.
- [128] LAITON-DONATO, K.; FRANCO-MUÑOZ, C.; ÁLVAREZ-DÍAZ, D. A.; RUIZ-MORENO, H. A.; USME-CIRO, J. A.; PRADA, D. A.; REALES-GONZÁLEZ, J.; CORCHUELO, S.; HERRERA-SEPÚLVEDA, M. T.; NAIZAQUE, J.; SANTAMARÍA, G.; RIVERA, J.; ROJAS, P.; ORTIZ, J. H.; CARDONA, A.; MALO, D.; PRIETO-ALVARADO, F.; GÓMEZ, F. R.; WIESNER, M.; MARTÍNEZ, M. L. O.; MERCADO-REYES, M. Characterization of the emerging b.1.621 variant of interest of SARS-CoV-2. *Infection, Genetics and Evolution*, v. 95, p. 105038, Nov. 2021.
- [129] ANNAVAJHALA, M. K.; MOHRI, H.; WANG, P.; NAIR, M.; ZUCKER, J. E.; SHENG, Z.; GOMEZ-SIMMONDS, A.; KELLEY, A. L.; TAGLIAVIA, M.; HUANG, Y.; BEDFORD, T.; HO, D. D.; UHLEMANN, A.-C. Emergence and expansion of SARS-CoV-2 b.1.526 after identification in new york. *Nature*, London, v. 597, n. 7878, p. 703–708, Aug. 2021.
- [130] YANG, W.-T.; HUANG, W.-H.; LIAO, T.-L.; HSIAO, T.-H.; CHUANG, H.-N.; LIU, P.-Y. SARS-CoV-2 e484k mutation narrative review: Epidemiology, immune escape, clinical implications, and future considerations. *Infection and Drug Resistance*, v. Volume 15, p. 373–385, Feb. 2022.
- [131] RADVAK, P.; KWON, H.-J.; KOSIKOVA, M.; ORTEGA-RODRIGUEZ, U.; XIANG, R.; PHUE, J.-N.; SHEN, R.-F.; ROZZELLE, J.; KAPOOR, N.; RABARA, T.; FAIRMAN, J.; XIE, H. SARS-CoV-2 b.1.1.7 (alpha) and b.1.351 (beta) variants induce pathogenic patterns in k18-hACE2 transgenic mice distinct from early strains. *Nature Communications*, v. 12, n. 1, Nov. 2021.
- [132] ANANDAKRISHNAN, R.; AGUILAR, B.; ONUFRIEV, A. V. H+ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research*, v. 40, n. W1, p. W537–W541, May 2012.

- [133] SALOMON-FERRER, R.; CASE, D. A.; WALKER, R. C. An overview of the amber biomolecular simulation package. *WIREs Computational Molecular Science*, v. 3, n. 2, p. 198–210, Sept. 2012.
- [134] PRICE, D. J.; BROOKS, C. L. A modified TIP3p water potential for simulation with ewald summation. *The Journal of Chemical Physics*, v. 121, n. 20, p. 10096–10103, Nov. 2004.
- [135] KIRSCHNER, K. N.; YONGYE, A. B.; TSCHAMPEL, S. M.; GONZÁLEZ-OUTEIRIÑO, J.; DANIELS, C. R.; FOLEY, B. L.; WOODS, R. J. GLYCAM06: A generalizable biomolecular force field. carbohydrates. *Journal of Computational Chemistry*, v. 29, n. 4, p. 622–655, Sept. 2007.
- [136] PHILLIPS, J. C.; BRAUN, R.; WANG, W.; GUMBART, J.; TAJKHORSHID, E.; VILLA, E.; CHIPOT, C.; SKEEL, R. D.; KALÉ, L.; SCHULTEN, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, v. 26, n. 16, p. 1781–1802, 2005.
- [137] ROE, D. R.; CHEATHAM, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, v. 9, n. 7, p. 3084–3095, June 2013.
- [138] SINKO, W.; MIAO, Y.; DE OLIVEIRA, C. A. F.; MCCAMMON, J. A. Population based reweighting of scaled molecular dynamics. *The Journal of Physical Chemistry B*, v. 117, n. 42, p. 12759–12768, July 2013.
- [139] MIAO, Y.; SINKO, W.; PIERCE, L.; BUCHER, D.; WALKER, R. C.; MCCAMMON, J. A. Improved reweighting of accelerated molecular dynamics simulations for free energy calculation. *Journal of Chemical Theory and Computation*, v. 10, n. 7, p. 2677–2689, Jul 2014.
- [140] STEWART, J. J. P. Mopac: A semiempirical molecular orbital program. *Journal of Computer-Aided Molecular Design*, v. 4, n. 1, p. 1–103, Mar. 1990.
- [141] STEWART, J. J. P. Optimization of parameters for semiempirical methods vi: more modifications to the nddo approximations and re-optimization of parameters. *Journal of Molecular Modeling*, v. 19, n. 1, p. 1–32, Nov. 2012.

- [142] STEWART, J. J. P. Application of localized molecular orbitals to the solution of semi-empirical self-consistent field equations. *International Journal of Quantum Chemistry*, v. 58, n. 2, p. 133–146, 1996.
- [143] GILES, N. M.; WATTS, A. B.; GILES, G. I.; FRY, F. H.; LITTLECHILD, J. A.; JACOB, C. Metal and redox modulation of cysteine protein function. *Chemistry & Biology*, v. 10, n. 8, p. 677–693, Aug. 2003.
- [144] HübNER, C.; HAASE, H. Interactions of zinc- and redox-signaling pathways. *Redox Biology*, v. 41, p. 101916, May 2021.
- [145] PATCHETT, S.; LV, Z.; RUT, W.; BÉKÉS, M.; DRAG, M.; OLSEN, S. K.; HUANG, T. T. A molecular sensor determines the ubiquitin substrate specificity of SARS-CoV-2 papain-like protease. *Cell Reports*, v. 36, n. 13, p. 109754, Sept. 2021.
- [146] VAN HUIZEN, M.; BLOEME TER HORST, J. R.; DE GRUYTER, H. L. M.; GEURINK, P. P.; VAN DER HEDEN VAN NOORT, G. J.; KNAAP, R. C. M.; NELEMANS, T.; OGANDO, N. S.; LEIJS, A. A.; URAKOVA, N.; MARK, B. L.; SNIJDER, E. J.; MYENI, S. K.; KIKKERT, M. Deubiquitinating activity of sars-cov-2 papain-like protease does not influence virus replication or innate immune responses in vivo. *PLOS Pathogens*, v. 20, n. 3, p. e1012100, Mar. 2024.
- [147] LEE, H.; LEI, H.; SANTARSIERO, B. D.; GATUZ, J. L.; CAO, S.; RICE, A. J.; PATEL, K.; SZYPULINSKI, M. Z.; OJEDA, I.; GHOSH, A. K.; JOHNSON, M. E. Inhibitor recognition specificity of MERS-CoV papain-like protease may differ from that of SARS-CoV. *ACS Chemical Biology*, v. 10, n. 6, p. 1456–1465, Mar. 2015.
- [148] SHEN, Z.; RATIA, K.; COOPER, L.; KONG, D.; LEE, H.; KWON, Y.; LI, Y.; ALQARNI, S.; HUANG, F.; DUBROVSKYI, O.; RONG, L.; THATCHER, G.; XIONG, R. Potent, novel SARS-CoV-2 PLpro inhibitors block viral replication in monkey and human cell cultures. Feb. 2021.
- [149] KRISHNA, S. S. Structural classification of zinc fingers: SURVEY AND SUMMARY. *Nucleic Acids Research*, v. 31, n. 2, p. 532–550, Jan. 2003.
- [150] TAN, H.; HU, Y.; JADHAV, P.; TAN, B.; WANG, J. Progress and challenges in targeting the SARS-CoV-2 papain-like protease. *Journal of Medicinal Chemistry*, v. 65, n. 11, p. 7561–7580, May 2022.

- [151] MA, C.; WANG, J. Validation and invalidation of SARS-CoV-2 papain-like protease inhibitors. *ACS Pharmacology & Translational Science*, v. 5, n. 2, p. 102–109, Jan. 2022.
- [152] DUDEV, T.; LIM, C. Modeling  $\text{zn}^{2+}$ -cysteinate complexes in proteins. *The Journal of Physical Chemistry B*, v. 105, n. 43, p. 10709–10714, Sept. 2001.
- [153] ULLRICH, S.; NITSCHKE, C. SARS-CoV-2 papain-like protease: Structure, function and inhibition. *ChemBioChem*, v. 23, n. 19, Sept. 2022.
- [154] WYDORSKI, P. M.; OSIPIUK, J.; LANHAM, B. T.; TESAR, C.; ENDRES, M.; ENGLE, E.; JEDRZEJCZAK, R.; MULLAPUDI, V.; MICHALSKA, K.; FIDELIS, K.; FUSHMAN, D.; JOACHIMIAK, A.; JOACHIMIAK, L. A. Dual domain recognition determines SARS-CoV-2 PLpro selectivity for human ISG15 and k48-linked di-ubiquitin. Sept. 2021.
- [155] GOLD, I. M.; REIS, N.; GLASER, F.; GLICKMAN, M. H. Coronaviral PLpro proteases and the immunomodulatory roles of conjugated versus free interferon stimulated gene product-15 (ISG15). *Seminars in Cell & Developmental Biology*, v. 132, p. 16–26, Dec. 2022.
- [156] SANTOS, V. P. D.; RODRIGUES, A.; DUTRA, G.; BASTOS, L.; MARIANO, D.; MENDONÇA, J. G.; LOBO, Y. J. G.; MENDES, E.; MAIA, G.; DOS SANTOS MACHADO, K.; WERHLI, A. V.; ROCHA, G.; DE LIMA, L. H. F.; DE MELO-MINARDI, R. Evolve: understanding the impact of mutations in SARS-CoV-2 variants spike protein on antibodies and ACE2 affinity through patterns of chemical interactions at protein interfaces. *PeerJ*, v. 10, p. e13099, Mar. 2022.
- [157] DAI, L.; GAO, G. F. Viral targets for vaccines against COVID-19. *Nature Reviews Immunology*, v. 21, n. 2, p. 73–82, Dec. 2020.
- [158] DE LIMA, L. H. F.; FERNANDEZ-QUINTÉRO, M. L.; ROCHA, R. E. O.; MARIANO, D. C. B.; DE MELO-MINARDI, R. C.; LIEDL, K. R. Conformational flexibility correlates with glucose tolerance for point mutations in  $\beta$ -glucosidases – a computational study. *Journal of Biomolecular Structure and Dynamics*, v. 39, n. 5, p. 1621–1634, Mar. 2020.
- [159] HARVEY, W. T.; CARABELLI, A. M.; JACKSON, B.; GUPTA, R. K.; THOMSON, E. C.; HARRISON, E. M.; LUDDEN, C.; REEVE, R.; RAMBAUT, A.; PEACOCK, S. J.; AND,

D. L. R. SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, v. 19, n. 7, p. 409–424, June 2021.

- [160] CHEN, Y.; LIU, Q.; GUO, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *Journal of Medical Virology*, v. 92, n. 4, p. 418–423, Feb. 2020.



## 7.1 Diferentes versões do script escrito em Python para plotar os gráficos de RMSF

### RMSF.py

```
1 # -*- coding: utf-8 -*-
2 #!/usr/bin/env python3
3 # -*- coding: utf-8 -*-
4 """
5 Created on Wed Jul 27 19:18:57 2022
6
7 @author: gutembergue
8 """
9
10 import pandas as pd
11 import numpy as np
12 from matplotlib import pyplot as pl
13 import seaborn as sns
14
15 df = pd.read_csv('rmsfrbm-+.csv')
16 df.head()
17 x = df['resnumber']
18 med1 = df['Med-+']
19 stdev1 = df['Stdev-+']
20
21 fig,ax = pl.subplots()
22 line_c, = ax.plot(x,med1,'#1E90FF', label='- +')
23 ax.fill_between(x,med1-stdev1, med1+stdev1,
24 color=line_c.get_color(),alpha=.3)
25 pl.ylabel('RMSF($\AA$)', fontsize=14, fontweight='bold')
26 pl.xlabel('Resíduos', fontsize=14, fontweight='bold')
27 pl.title('RBM(-+)', fontsize=30, color='#C2C2C2',
28 fontweight='bold')
29 pl.ylim(0,5.0)
30 pl.xlim(438,508)
```

```

31 pl.tight_layout()
32 pl.legend(title='Mutacao')
33 sns.despine()
34 pl.savefig('rmsfrbm-+1.svg', dpi= 600,  bbox_inches='tight')
35
36 pl.show()

```

## RMSF.py

```

1
2 #!/usr/bin/env python3
3 # -*- coding: utf-8 -*-
4 """
5 Created on Wed Jul 27 14:31:04 2022
6
7 @author: gutembergue
8 """
9
10 import pandas as pd
11 import matplotlib.pyplot as plt
12 import seaborn as sns
13
14 df = pd.read_csv('rmsfrbm.csv')
15 df.head()
16 x = df['resnumber']
17 med1 = df['Med++']
18 stdev1 = df['Stdev++']
19 med2 = df['Med+-']
20 stdev2 = df['Stdev+-']
21 med3 = df['Med-+']
22 stdev3 = df['Stdev-+']
23 med4 = df['Med--']
24 stdev4 = df['Stdev--']
25
26 plt.plot(x, med1, '#228B22', label='+ +')
27 plt.fill_between(x, med1-stdev1, med1+stdev1, color='#228B22',
28 alpha=0.3)
29
30 plt.plot(x, med2, '#FF0000', label='+ -')
31 plt.fill_between(x, med2-stdev2, med2+stdev2, color='#FF0000',
32 alpha=0.3)
33
34 plt.plot(x, med3, '#1E90FF', label='- +')
35 plt.fill_between(x, med3-stdev3, med3+stdev3, color='#1E90FF',
36 alpha=0.3)
37
38 plt.plot(x, med4, 'gold', label='- -')
39 plt.fill_between(x, med4-stdev4, med4+stdev4, color='gold',
40 alpha=0.3)
41
42 plt.ylabel('RMSF($\AA$)', fontsize=14, fontweight='bold')
43 plt.xlabel('Resíduos', fontsize=14, fontweight='bold')
44 plt.title('RBD', fontsize=32, color='#C2C2C2', fontweight='bold')
45 plt.ylim(0.0,5.0)

```

```

46 plt.xlim(438,508)
47 plt.tight_layout()
48 #display plot
49 plt.legend(title='Mutacao')
50 plt.savefig('rmsfrbm1.svg', dpi= 600,  bbox_inches='tight')
51
52 plt.show()

```

## RMSF.py

```

1  #!/usr/bin/env python3
2  # -*- coding: utf-8 -*-
3  """
4  Created on Mon Nov 29 15:35:52 2021
5
6  @author: gutembergue
7  """
8  import pandas as pd
9  import matplotlib.pyplot as plt
10 df = pd.read_csv('rmsf_all.csv', index_col='Residuos')
11 df.head()
12 plt.plot(plotsize=(20,30), dpi = 600)
13 plt.plot(df['PLpro_SARS1'], label='PLpro_SARS1', color='#FF3030',
14 linewidth=1)
15 plt.plot(df['PLpro_SARS2'], label='PLpro_SARS2', color='#FFD700',
16 linewidth=1)
17 plt.plot(df['PLpro_SARS1_Zn'], label='PLpro_SARS1_Zn',
18 color='#1C86EE', linewidth=1)
19 plt.plot(df['PLpro_SARS2_Zn'], label='PLpro_SARS2_Zn',
20 color='#71C671', linewidth=1)
21 plt.legend()
22
23 plt.annotate("", xy=(40,3.8), xytext=(40.0,5.0),
24               arrowprops=dict(arrowstyle="->",
25                               connectionstyle="arc3"))
26 plt.text(40.0, 5.0, r'Ubl',
27          {'color': 'green', 'fontsize': 8, 'ha': 'center',
28            'va': 'center', 'bbox': dict(boxstyle="square",
29                                          fc="white", ec="black", pad=0.2,)}}
30 plt.annotate("", xy=(191,3.7622), xytext=(191,4.7),
31               arrowprops=dict(arrowstyle="->",
32                               connectionstyle="arc3"))
33 plt.text(191, 4.7, r'Zn-($\bet4\beta5$)',
34          {'color': 'blue', 'fontsize': 8, 'ha': 'center',
35            'va': 'center', 'bbox': dict(boxstyle="square",
36                                          fc="white", ec="black", pad=0.2,)}}
37 plt.annotate("", xy=(225,5.7237), xytext=(225,6.8),
38               arrowprops=dict(arrowstyle="->",
39                               connectionstyle="arc3"))
40 plt.text(225, 6.8, r'Zn-($\bet6\beta7$)',
41          {'color': 'red', 'fontsize': 8, 'ha': 'center',
42            'va': 'center', 'bbox': dict(boxstyle="square",
43                                          fc="white", ec="black", pad=0.2,)}}
44 plt.annotate("", xy=(268,4.1432), xytext=(268,5.1),

```

```

45         arrowprops=dict(arrowstyle="->",
46                           connectionstyle="arc3"))
47 plt.text(268, 5.1, r'Alca BL2',
48         {'color': 'gold', 'fontsize': 8, 'ha': 'center',
49          'va': 'center', 'bbox': dict(boxstyle="square",
50                                       fc="white", ec="black", pad=0.2,)})
51 #Nomes da colunas e titulo(layout)
52 plt.ylabel('RMSF($\AA$)', fontsize=14, fontweight='bold')
53 plt.xlabel('Resíduos', fontsize=14, fontweight='bold')
54 plt.title('RMSF SARS-CoV-1 e SARS-CoV-2', fontsize=16,
55 color='#030303')
56 plt.ylim(0,12)
57 plt.xlim(0,315)
58 plt.tight_layout()
59 #salvar plot
60 plt.savefig('rmsf_all1.svg', dpi= 600,  bbox_inches='tight')
61 plt.show()

```

## RMSF.py

```

1  #!/usr/bin/env python3
2  # -*- coding: utf-8 -*-
3  """
4  Created on Wed Jul 27 19:18:57 2022
5
6  @author: gutembergue
7  """
8
9  import pandas as pd
10 import numpy as np
11 from matplotlib import pyplot as pl
12 import seaborn as sns
13
14 df = pd.read_csv('rmsfrbm-+.csv')
15 df.head()
16 x = df['resnumber']
17 med1 = df['Med-+']
18 stdev1 = df['Stdev-+']
19
20 fig,ax = pl.subplots()
21 line_c, = ax.plot(x,med1,'#1E90FF', label='- +')
22 ax.fill_between(x,med1-stdev1, med1+stdev1,
23 color=line_c.get_color(),alpha=.3)
24 pl.ylabel('RMSF($\AA$)', fontsize=14, fontweight='bold')
25 pl.xlabel('Resíduos', fontsize=14, fontweight='bold')
26 pl.title('RBM(-+)', fontsize=30, color='#C2C2C2',
27 fontweight='bold')
28 pl.ylim(0,5.0)
29 pl.xlim(438,508)
30 pl.tight_layout()
31 pl.legend(title='Mutacoes')
32 sns.despine()
33 pl.savefig('rmsfrbm-+1.svg', dpi= 600,  bbox_inches='tight')
34

```

```
35 pl.show()
```

## RMSF.py

```

1  #!/usr/bin/env python3
2  # -*- coding: utf-8 -*-
3  """
4  Created on Wed Jul 27 14:31:04 2022
5
6  @author: gutemberg
7  """
8
9  import pandas as pd
10 import matplotlib.pyplot as plt
11 import seaborn as sns
12
13 df = pd.read_csv('rmsfrbm.csv')
14 df.head()
15 x = df['resnumber']
16 med1 = df['Med++']
17 stdev1 = df['Stdev++']
18 med2 = df['Med+-']
19 stdev2 = df['Stdev+-']
20 med3 = df['Med-+']
21 stdev3 = df['Stdev-+']
22 med4 = df['Med--']
23 stdev4 = df['Stdev--']
24
25 plt.plot(x, med1, '#228B22', label='+ +')
26 plt.fill_between(x, med1-stdev1, med1+stdev1, color='#228B22',
27 alpha=0.3)
28
29 plt.plot(x, med2, '#FF0000', label='+ -')
30 plt.fill_between(x, med2-stdev2, med2+stdev2, color='#FF0000',
31 alpha=0.3)
32
33 plt.plot(x, med3, '#1E90FF', label='- +')
34 plt.fill_between(x, med3-stdev3, med3+stdev3, color='#1E90FF',
35 alpha=0.3)
36
37 plt.plot(x, med4, 'gold', label='- -')
38 plt.fill_between(x, med4-stdev4, med4+stdev4, color='gold',
39 alpha=0.3)
40
41 plt.ylabel('RMSF($\AA$)', fontsize=14, fontweight='bold')
42 plt.xlabel('Residuos', fontsize=14, fontweight='bold')
43 plt.title('RBD', fontsize=32, color='#C2C2C2', fontweight='bold')
44 plt.ylim(0.0,5.0)
45 plt.xlim(438,508)
46 plt.tight_layout()
47 #display plot
48 plt.legend(title='Mutacoes')
49 plt.savefig('rmsfrbm1.svg', dpi= 600, bbox_inches='tight')
50

```

```
51 plt.show()
```

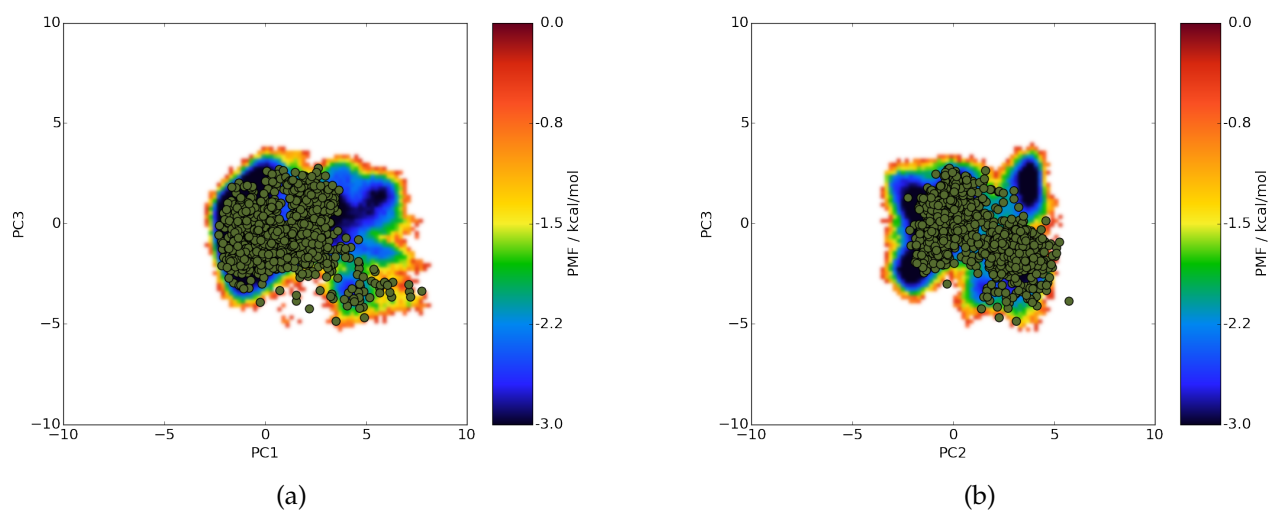
## RMSD.py

```

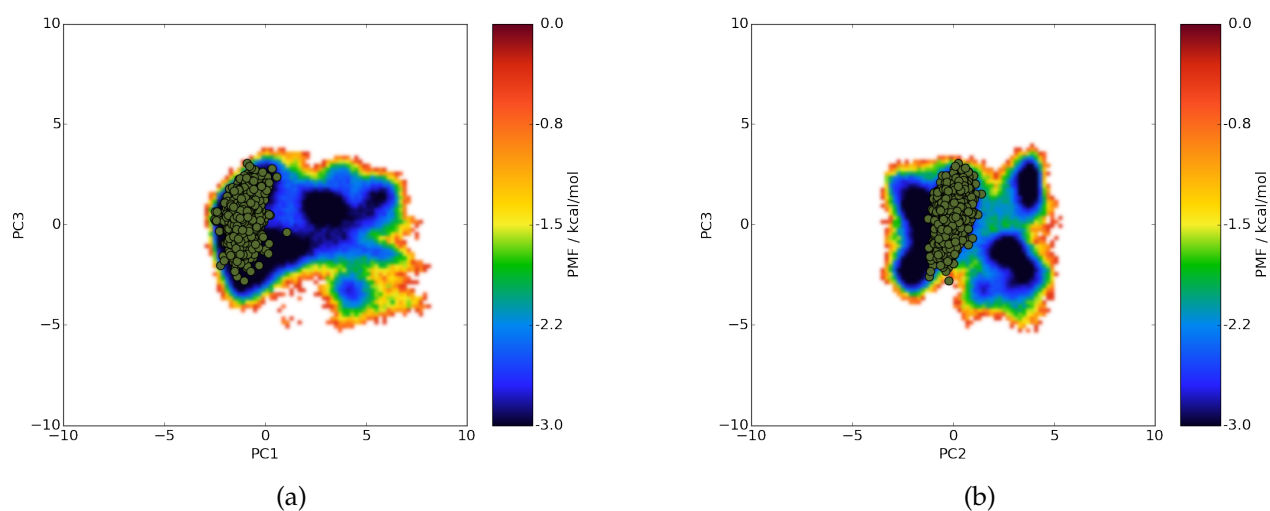
1
2 #!/usr/bin/env python3
3 # -*- coding: utf-8 -*-
4 """
5 Created on Mon Nov 29 15:35:52 2021
6
7 @author: gutembergue
8 """
9 import pandas as pd
10 import matplotlib.pyplot as plt
11 df = pd.read_csv('rmsd_all.csv', index_col='Tempo(ns)')
12 df.head()
13 plt.plot(figsize=(20,30), dpi = 600)
14 plt.plot(df['PLpro_SARS1'], label='PLpro_SARS1',
15 color='#FF3030', linewidth=1)
16 plt.plot(df['PLpro_SARS2'], label='PLpro_SARS2',
17 color='#FFD700', linewidth=1)
18 plt.plot(df['PLpro_SARS1_Zn'], label='PLpro_SARS1_Zn',
19 color='#1C86EE', linewidth=1)
20 plt.plot(df['PLpro_SARS2_Zn'], label='PLpro_SARS2_Zn',
21 color='#71C671', linewidth=1)
22 plt.legend()
23 #add axis labels and a title
24 plt.ylabel('RMSD($\AA$)', fontsize=14, fontweight='bold')
25 plt.xlabel('Tempo(ns)', fontsize=14, fontweight='bold')
26 plt.title('RMSD PLpro SARS-CoV-1 e SARS-CoV-2', fontsize=16,
27 fontweight='bold')
28 plt.tight_layout()
29 #display plot
30 plt.savefig('rmsd_all.svg', dpi= 600, bbox_inches='tight')
31 plt.show()

```

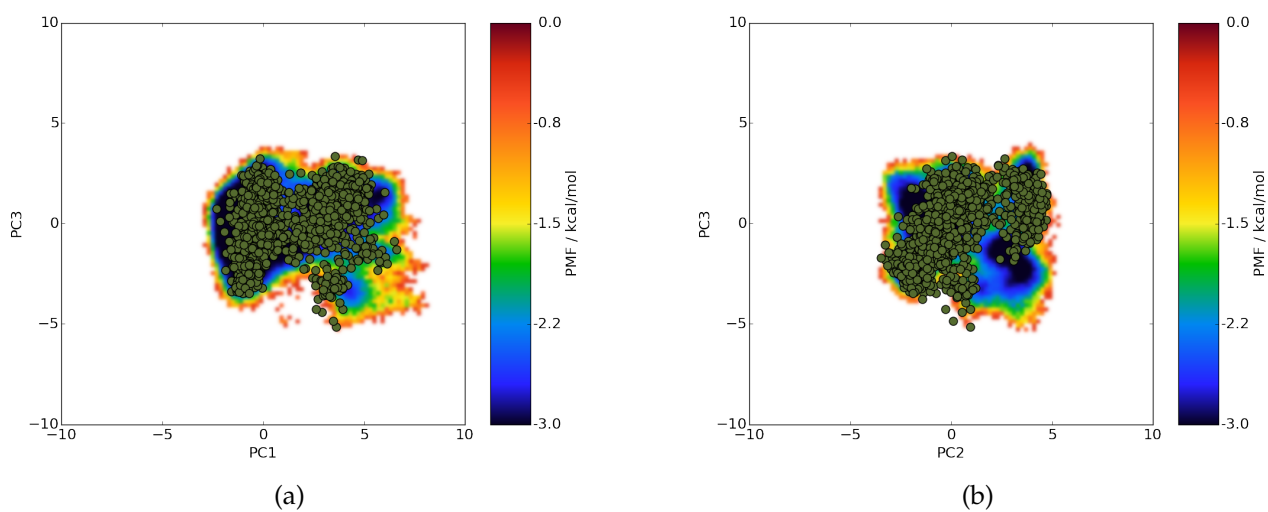
## 7.2 Perfis de energia livre PCA



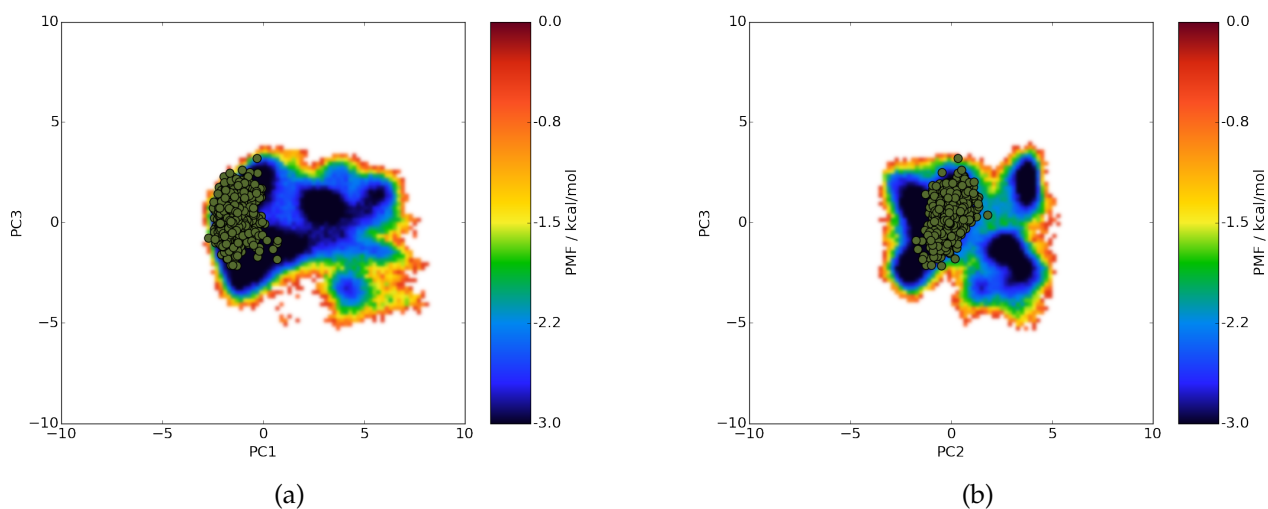
**Figura 7.1:** Perfis de energia livre para a PLpro SCoV-1 sem zinco. Os círculos ao longo dos gráficos representam as projeções da trajetória ao longo de (a) PC1 *vs.* PC3 e (b) PC2 *vs.* PC3.



**Figura 7.2:** Perfis de energia livre para a PLpro SCoV-1 com zinco. Os círculos ao longo dos gráficos representam as projeções da trajetória ao longo de (a) PC1 *vs.* PC3 e (b) PC2 *vs.* PC3.



**Figura 7.3:** Perfis de energia livre para a PLpro SCoV-2 sem zinco. Os círculos ao longo dos gráficos representam as projeções da trajetória ao longo de (a) PC1 *vs.* PC3 e (b) PC2 *vs.* PC3.



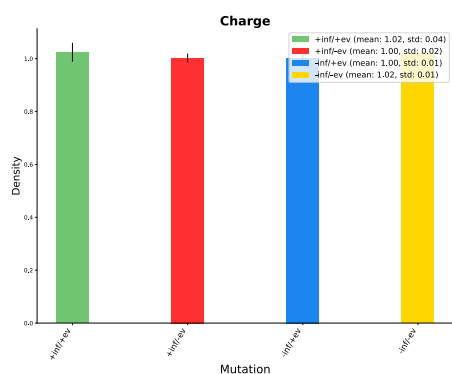
**Figura 7.4:** Perfis de energia livre para a PLpro SCoV-2 sem zinco. Os círculos ao longo dos gráficos representam as projeções da trajetória ao longo de (b) PC1 *vs.* PC3 e (a) PC2 *vs.* PC3.



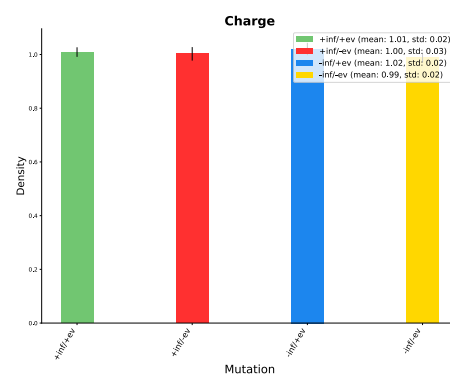
### 7.3 Descritores de Reatividade

**Tabela 7.1:** Valores da Eletrofilicidade para a arginina-346 na diferentes mutações.

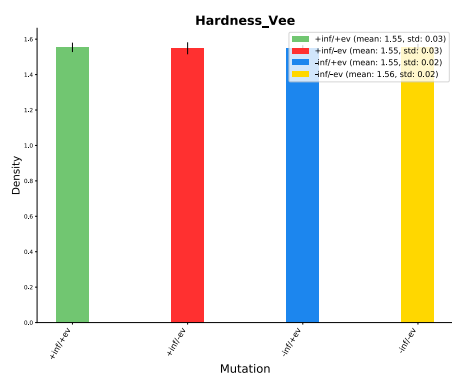
+inf/+ev			+inf/-ev			-inf/+ev			-inf/-ev		
Mutação	360-374	RBM	Mutação	360-374	RBM	Mutação	360-374	RBM	Mutação	360-374	RBM
E484K-N501Y	1.834980	2.279729	N354D	0.717770	1.827783	E484K	1.915893	1.989777	K417Y	2.999995	2.999983
E484Q	1.888844	1.975217	N501S	2.567328	2.999993	F490L	2.562971	1.569482	N439R	2.423379	1.791554
K417N-E484K-N501Y	1.928988	2.432253	N501Y	2.569800	2.999993	F490S	2.589467	1.931312	Q498I	2.999995	2.999985
K417T-E484K-N501Y	1.981176	2.436129	S477N	2.568276	2.999993	G446V	2.591805	1.476425	Q498L	2.999995	2.999983
L452Q	1.857328	2.080030	T478K	2.303422	2.524978	K417N	2.999968	1.868912	Q498V	2.999995	2.999983
L452R	1.347987	1.490374	V367F	2.561468	2.999993	K417T	2.999984	1.675920	S494K	2.999995	2.999985
L452R-E484Q	1.595461	1.669666				L452Q-F490S	2.587072	1.743562	S494Q	2.999995	2.999985
L452R-T478K	1.621682	1.613773				N439K	2.999962	1.448901	V445L	2.999995	2.999983
R346K-E484K-N501Y	0.632015	2.999996				R346K	1.231340	0.640296	wt	2.999995	2.999983
S477N-E484K	1.919801	2.229430				Y508H	2.505917	1.521774	Y489F	2.999995	2.999981
									Y505F	2.999993	2.999982



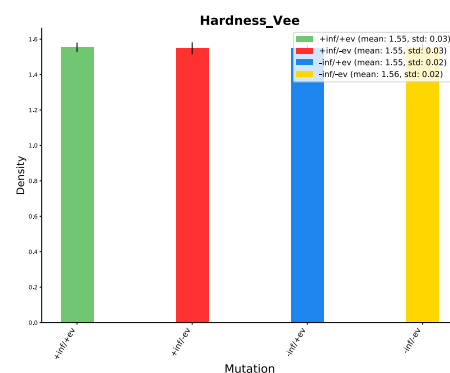
(a)



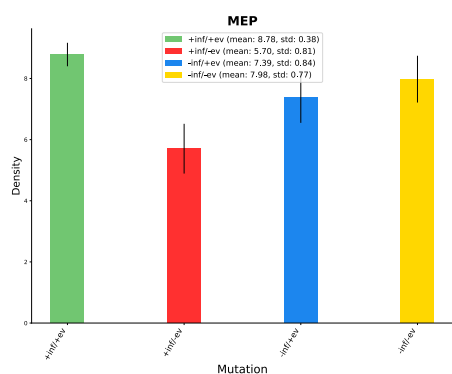
(b)



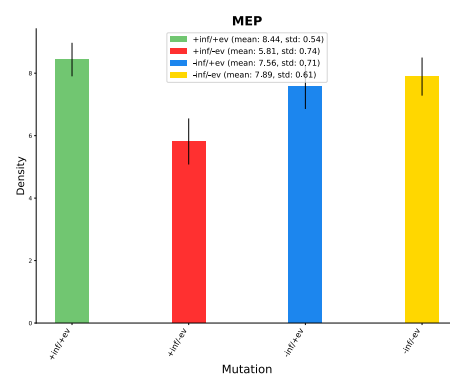
(c)



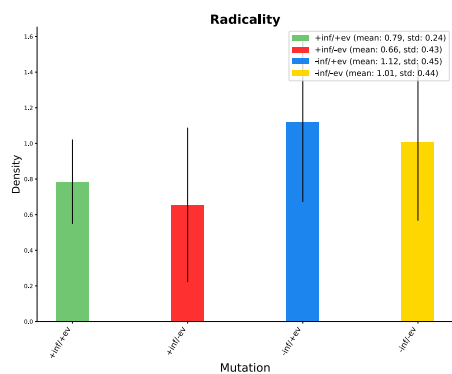
(d)



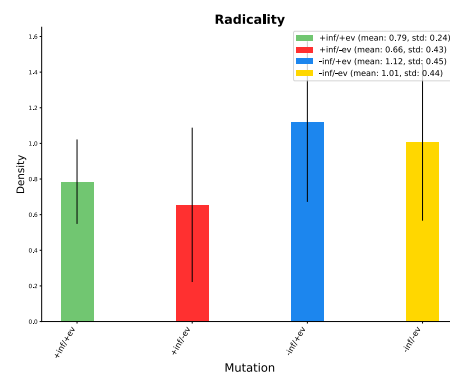
(e)



(f)

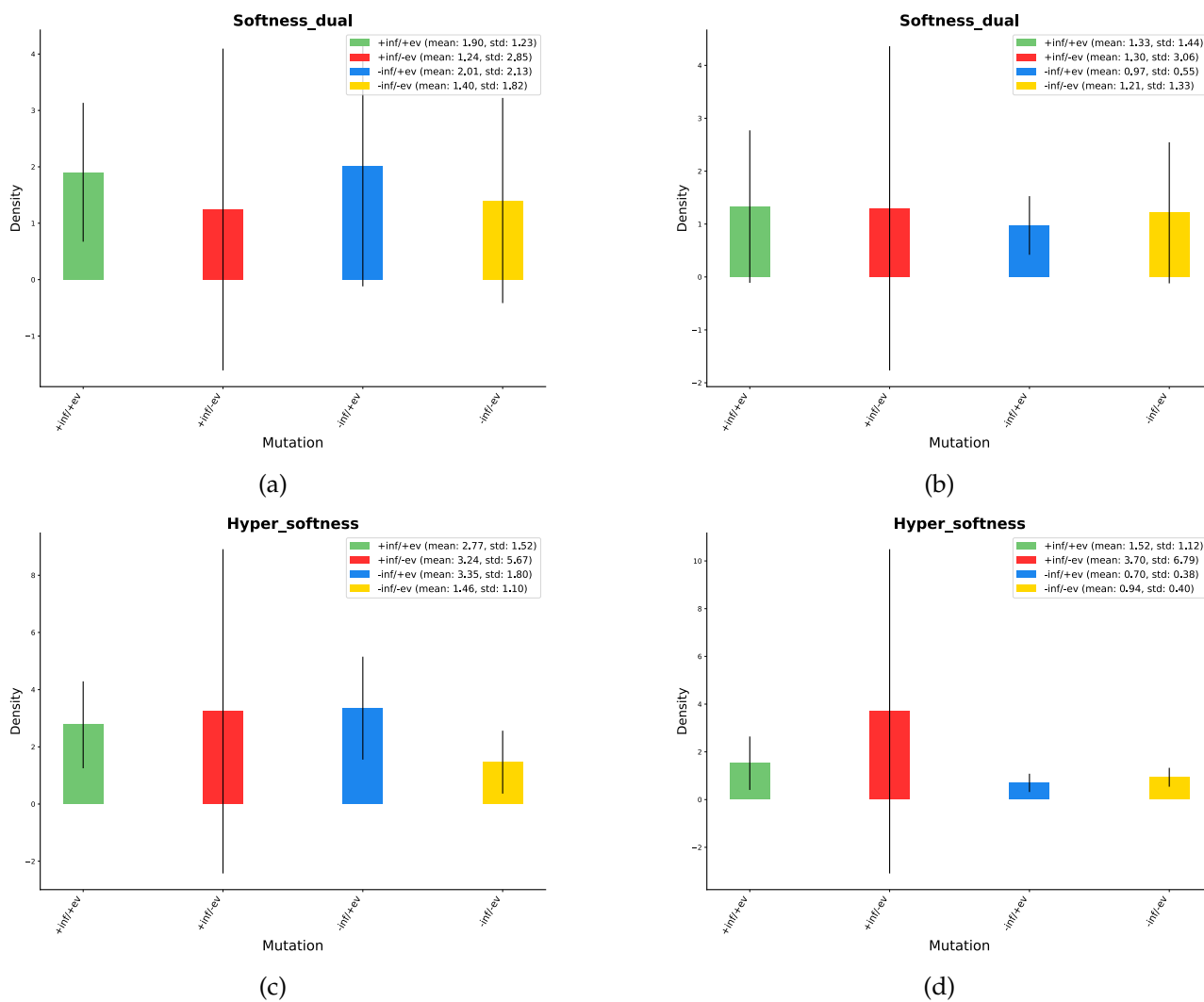


(g)



(h)

**Figura 7.5:** Descritores de reatividade (a) e (b) Carga parcial, (c) e (d) Dureza local (potencial elétron-elétron), (e) e (f) Potencial eletrostático molecular e (g) e (h) Radicalidade.



**Figura 7.6:** Descritores para moleza local (a) Moleza Dual Local (360-374), (b) Moleza Dual Local (RBM), (c) Hiper Moleza Local(360-374) e (d) Hiper Moleza Local(RBM).