

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

ANA LUÍSA RODRIGUES DE ÁVILA

**Análise do comportamento dinâmico na interface entre variantes
do hospedeiro e do SARS-CoV-2**

Ribeirão Preto

2023

ANA LUÍSA RODRIGUES DE ÁVILA

Análise do comportamento dinâmico na interface entre variantes do hospedeiro e do SARS-CoV-2

Versão Original

Dissertação apresentada à Faculdade de
Medicina de Ribeirão Preto para obtenção
do título de Mestre em Ciências.

Área de concentração: Genética

Orientadora: Prof^a Dr^a Silvana Giuliatti

Ribeirão Preto

2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo de publicação preparada pela biblioteca do serviço de documentação da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

de Ávila, Ana Luísa Rodrigues

Análise do comportamento dinâmico na interface entre variantes do hospedeiro e do SARS-CoV-2. Ribeirão Preto, 2023.

Nº77f

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto para obtenção do título de Mestre em Ciências.

Área de concentração: Genética

Orientadora: Profª Drª Silvana Giuliatti

Palavras-chave: SARS-CoV-2, COVID-19, variantes, interação vírus-hospedeiro, bioinformática

Nome: Ana Luísa Rodrigues de Ávila

Título: Análise do comportamento dinâmico na interface entre variantes do hospedeiro e do SARS-CoV-2

Dissertação apresentada à Faculdade de Medicina de Ribeirão Preto para obtenção do título de Mestre em Ciências.

Aprovada em:

Banca Examinadora:

Prof. Dr.

Instituição:

Profa. Dra.

Instituição:

Prof. Dr.

Instituição:

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e com apoio do CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil.

Dedico

Aos meus pais, Lúcia R. de Ávila e Sidnei de Ávila.

Agradecimentos

À Prof^a. Dr^a. Silvana Giuliatti, pela orientação, apoio, paciência, confiança e dedicação. Agradeço pela oportunidade, a qual contribuiu para a minha formação.

Aos meus colegas do Grupo de Bioinformática (GBi) do Departamento de Genética da FMRP-USP, pela troca de experiências e momentos compartilhados.

À minha família, por todo o apoio e incentivo, pela compreensão e confiança nas minhas decisões.

Aos membros da banca examinadora, por terem atendido ao convite para avaliar este trabalho, contribuindo com os seus conhecimentos e para a melhoria do mesmo.

À Superintendência da Tecnologia e Informação da USP, pelos recursos computacionais de *High Performance Computing* (HPC) utilizados para o desenvolvimento deste trabalho.

Às agências de fomento, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq), pelo auxílio financeiro na realização do presente trabalho.

À USP e FMRP pela oportunidade de realização da pós-graduação.

Resumo

DE ÁVILA, Ana Luísa Rodrigues. Análise do comportamento dinâmico na interface entre variantes do hospedeiro e do SARS-CoV-2. Dissertação (Mestrado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2023.

A proteína Spike do SARS-CoV-2 (do inglês, *Severe Acute Respiratory Syndrome Coronavirus-2*) interage com os receptores da enzima conversora de angiotensina 2 (ACE2) e com a protease transmembrana de serina II (TMPRSS2) para entrar na célula humana. Substituições de aminoácidos em posições-chave no domínio de ligação ao receptor (RBD) da proteína Spike alteram a afinidade de ligação do SARS-CoV-2 com a ACE2 e TMPRSS2, podendo influenciar na taxa de infecção do vírus. Mutações específicas da ACE2 têm sido associados ao aumento da suscetibilidade ao SARS-CoV-2, e o mesmo pode ocorrer com variantes da TMPRSS2, afetando o risco de infecção e a gravidade da COVID-19. Além disso, o SARS-CoV-2 tem uma alta probabilidade de sofrer mutação e se adaptar ao ambiente. No entanto, o efeito dessas variações genéticas sobre a estabilidade e afinidade da interação Spike-ACE2 e Spike-TMPRSS2 ainda é pouco compreendido. Para avaliar o efeito de mutações e comparar mudanças estruturais entre os complexos, foi utilizado o método da simulação de dinâmica molecular (DM). A energia livre de ligação e a força de interação biomolecular foram estimadas pelo cálculo da área de superfície de Poisson-Boltzman da mecânica molecular MM/PBSA (do inglês, *Molecular Mechanics Poisson–Boltzmann Surface Area*). As simulações de dinâmica molecular revelaram uma trajetória estável com variações sutis, enfatizando alguns resíduos-chave na região de interação entre os complexos. A combinação de simulações de DM e métodos de aprendizado de máquina permitiu uma compreensão mais profunda sobre como as variações genéticas tanto no vírus quanto no receptor ACE2 podem impactar a região de interação entre essas proteínas essenciais. Os resultados deste estudo contribuem para o entendimento sobre como as variações genéticas no vírus e no receptor do hospedeiro influenciam a região de interação dessas proteínas, tendo implicações importantes para estratégias de tratamento, especialmente no campo da medicina de precisão.

Palavras-chave: SARS-CoV-2, COVID-19, variantes, interação vírus-hospedeiro, bioinformática

Abstract

DE ÁVILA, Ana Luísa Rodrigues. Analysis of dynamic behavior at the interface between host and SARS-CoV-2 variants. Dissertation (Master's degree) – Ribeirão Preto Medical School, University of São Paulo, 2023.

The Spike protein of SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus-2) interacts with the angiotensin-converting enzyme 2 (ACE2) receptors and transmembrane serine protease II (TMPRSS2) to enter the human cell. Amino acid substitutions in key positions in the receptor-binding domain (RBD) of the Spike protein alter the binding affinity of SARS-CoV-2 with ACE2 and TMPRSS2, potentially influencing the virus's infection rate. Specific ACE2 mutations have been associated with increased susceptibility to SARS-CoV-2, and the same may occur with TMPRSS2 variants, affecting the risk of infection and severity of COVID-19. Furthermore, SARS-CoV-2 has a high probability of mutating and adapting to the environment. However, the effect of these genetic variations on the stability and affinity of the Spike-ACE2 and Spike-TMPRSS2 interaction is still poorly understood. To evaluate the effect of mutations and compare structural changes between the complexes, the molecular dynamics (MD) simulation method was used. The binding free energy and biomolecular interaction strength were estimated by calculating the Molecular Mechanics Poisson–Boltzmann Surface Area (MM/PBSA). Molecular dynamics simulations revealed a stable trajectory with subtle variations, emphasizing some key residues in the interaction region between the complexes. Combining MD simulations and machine learning methods allowed to gain deeper insights into how genetic variations in both the virus and the ACE2 receptor can impact the region of interaction between these essential proteins. The results of this study contribute to the understanding of how genetic variations in the virus and host receptor influence the interaction region of these proteins, having important implications for treatment strategies, especially in the field of precision medicine.

Keywords: SARS-CoV-2, COVID-19, variants, virus-host interaction, bioinformatics

Lista de figuras

Figura 1. Estrutura do vírus SARS-CoV-2.....	17
Figura 2. Estrutura da proteína Spike trimérica.....	17
Figura 3. Representação esquemática dos domínios proteína Spike.....	18
Figura 4. Representação estrutural da proteína Spike com o domínio RBD.....	18
Figura 5. Representação das etapas do mecanismo de entrada do SARS-CoV-2 na célula do hospedeiro.....	19
Figura 6. Representação da ativação do domínio S2.....	19
Figura 7. Ativação e entrada do SARS-CoV-2 na célula hospedeira.....	20
Figura 8. Representação esquemática da estrutura da ACE2.....	21
Figura 9. Mapa linear dos domínios da proteína TMPRSS2.....	22
Figura 10. Representação dos diferentes locais de mutação de resíduos de aminoácidos localizados da proteína Spike das cinco variantes de SARS-CoV-2.....	24
Figura 11. Variantes do SARS-CoV-2 e mutações localizadas do RBD da proteína Spike.....	29
Figura 12. Estrutura terciária do complexo ACE2-RBD-Spike.....	30
Figura 13. Estrutura terciária da proteína Spike.....	31
Figura 14. Estrutura completa da TMPRSS2 por modelagem molecular.....	31
Figura 15. Fluxograma dos métodos utilizados no aprendizado de máquinas.....	36
Figura 16. Fluxograma geral do trabalho.....	37
Figura 17. Gráfico RMSD das trajetórias da cadeia ACE2 (A) e Spike-RBD (B).....	38
Figura 18. RMSF dos resíduos da cadeia ACE2, definida pelos resíduos 19 a 614.....	38
Figura 19. Resíduos com maiores flutuações na ACE2 selvagem estão localizados em regiões de loop.....	39
Figura 20. RMSF dos resíduos da cadeia Spike, definida pelos resíduos 333 a 527.....	40
Figura 21. Localização dos resíduos Arg408, Ile434 e Lys444 na Spike.....	40

Figura 22. Raio de giração das cadeias ACE2 (A) e Spike (B).....	41
Figura 23. SASA das trajetórias dos complexos ACE2 e variantes Spike-RBD.....	42
Figura 24. Ilustração gráfica do DSSP dos complexos com ACE2 selvagem e Spike-RBD das variantes.	43
Figura 25. Decomposição de energia por resíduo MM/PBSA para as trajetórias obtidas nas simulações de DM.	45
Figura 26. Pares de resíduos determinados como mais importantes para distinguir a ligação entre as variantes ACE2 WT e Spike-RBD.	49
Figura 27. Gráfico RMSD das trajetórias das proteínas ACE2 (A) e Spike (B).	50
Figura 28. RMSF dos resíduos ACE2-Spike.	52
Figura 29. Localização dos resíduos Arg341, Trp328, Gln325, Arg582, Arg211, Cys219, Asn338 na ACE2.	52
Figura 30. RMSF dos resíduos ACE2-Spike.	53
Figura 31. Localização dos resíduos Arg26 e Arg211 na ACE2 e dos resíduos Trp436 e Ser443 na Spike.	53
Figura 32. Raio de giração da cadeia ACE2 (A) e Spike (B).	53
Figura 33. SASA das trajetórias dos complexos com as variantes da ACE2 e Spike selvagem.	54
Figura 34. Ilustração gráfica do DSSP dos complexos com variantes da ACE2 e Spike-RBD canônica.	55
Figura 35. Decomposição de energia por resíduo MM/PBSA para as trajetórias obtidas nas simulações de DM.	56
Figura 36. Pares de resíduos determinados como mais importantes para diferenciar a ligação entre a Spike-RBD WT e variantes da ACE2.	58
Figura 37. Gráfico RMSD das trajetórias da cadeia Spike (A) e TMPRSS2 (B).....	59
Figura 38. RMSF dos resíduos da cadeia Spike, definida pelos resíduos 27 a 1146.....	61
Figura 39. (A) Resíduos 748-756 e 985-995 (destacados em roxo) obtiveram alta flutuação e estão localizados em uma alfa-hélice da Spike, porém encontram-se distantes da interação com a TMPRSS2. (B) Localização do resíduo P793 na Spike. (C) Localização do resíduo K811 na Spike. (D) Localização do resíduo K921 na Spike.	61

Figura 40. RMSF dos resíduos da cadeiaTMPRSS2, definida pelos resíduos 106 a 491.....	62
Figura 41. Localização dos resíduos K390, Q438, G464 e C465 na TMPRSS2.....	62
Figura 42. Raio de Giração da cadeia Spike (A) e TMPRSS2 (B).	63
Figura 43. SASA das trajetórias dos complexos com as variantes da TMPRSS2 e Spike selvagem.	64

Lista de tabelas

Tabela 1. Decomposição residual e energia de ligação da interação proteína-proteína (dado em kJ.mol ⁻¹).	44
Tabela 2. Resíduos importantes nas variantes da Spike interagindo com ACE2 WT. obtidos por MLP e RF.	47
Tabela 3. Decomposição residual e energia de ligação da interação proteína-proteína (dado em kJ.mol ⁻¹).	56
Tabela 4. Resíduos importantes obtidos por MLP e RF para as variantes de ACE2 interagindo com a Spike-RBD. Os pares destacados em negrito e marcados com asteriscos apareceram tanto em análises de MLP como de RF.....	57

Lista de abreviaturas e siglas

Sigla	Nome
ABraOM	Arquivo Brasileiro Online de Mutações
ACE2	<i>Angiotensin Converting Enzyme 2</i>
AF	<i>Allele Frequency</i>
CLD	<i>Collectrin-Like Domain</i>
DM	Dinâmica Molecular
DSSP	<i>Dictionary of Secondary Structure of Proteins</i>
Eele	Energia eletrostática
EvdW	Energia de Van der Waals
GnomAD	<i>Genome Aggregation Database</i>
HR	<i>Heptad Repeat</i>
MLP	<i>Multilayer Perceptron</i>
MM/PBSA	<i>Molecular Mechanics Poisson–Boltzmann Surface Area</i>
NTD	<i>N-terminal Domain</i>
PD	<i>Peptide Domain</i>
PDB	<i>Protein Data Bank</i>
RAAS	<i>Renin-Angiotensin-Aldosterone System</i>
RBD	<i>Receptor Binding Domain</i>
RF	<i>Random Forest</i>
Rg	<i>Radius of Gyration</i>
RMSD	<i>Root Mean Square Deviation</i>
RMSF	<i>Root Mean Square Fluctuation</i>
S1	Subunidade 1
S2	Subunidade 2
SARS-CoV-2	<i>Severe Acute Respiratory Syndrome Coronavirus-2</i>
SASA	<i>Solvent Accessible Surface Area</i>
TM	<i>Transmembranar Domain</i>
TMPRSS2	<i>Transmembrane protease serine 2</i>
VOC	<i>Variants of Concern</i>
WT	<i>Wild Type</i>

Sumário

1.	INTRODUÇÃO E JUSTIFICATIVA	16
1.1	COVID-19.....	16
1.2	SARS-CoV-2.....	16
1.3	ACE2.....	20
1.4	TMPRSS2.....	21
1.5	Resíduos envolvidos na interação Vírus-Hospedeiro.....	22
1.6	Variabilidade do Vírus.....	23
1.7	Variantes da ACE2 e TMPRSS2.....	25
1.7.1	Mutações na ACE2.....	25
1.7.2	Mutações na TMPRSS2.....	26
1.8	Justificativa.....	27
2.	OBJETIVO	28
3.	MATERIAL E MÉTODOS	28
3.1	Obtenção dos Dados.....	28
3.1.1	Variantes da Spike.....	28
3.1.2	Variantes da ACE2 e TMPRSS2.....	29
3.1.2	Estruturas Terciárias das Proteínas.....	30
3.2	<i>Docking</i> entre Spike e TMPRSS2.....	32
3.3	Construção dos complexos com as variantes da Spike e polimorfismos da ACE2 e TMPRSS2.....	32
3.4	Dinâmica Molecular (DM).....	32
3.5	Análises das simulações de Dinâmica Molecular.....	34
3.6	Identificação de resíduos-chave na interação da proteína ACE2 com a Spike por aprendizagem de máquinas.....	35
4.1	Complexo ACE2 selvagem associada a variantes da Spike.....	37
4.1.1	Raiz do Desvio Quadrado Médio (RMSD).....	37

4.1.2	Flutuação da Raiz Quadrada Média (RMSF)	38
4.1.3	Raio de Giração (Rg)	41
4.1.4	Área Acessível ao Solvente (SASA)	41
4.1.5	Dicionário de Estrutura Secundária de Proteínas (DSSP)	42
4.1.6	Área de Superfície de Poisson-Boltzmann da Mecânica Molecular (MM/PBSA)	44
4.1.7	Identificação de resíduos-chave na Spike selvagem e variantes por aprendizagem de máquinas	47
4.2	Complexo Spike selvagem associado a variantes da ACE2	50
4.2.1	Raiz do Desvio Quadrado Médio (RMSD)	50
4.2.2	Flutuação da Raiz Quadrada Média (RMSF)	51
4.2.3	Raio de Giração (Rg)	53
4.2.4	Área Acessível ao Solvente (SASA)	54
4.2.5	Dicionário de Estrutura Secundária de Proteínas (DSSP)	54
4.2.6	Área de Superfície de Poisson-Boltzmann da Mecânica Molecular (MM/PBSA)	55
4.2.7	Identificação de resíduos-chave na ACE2 selvagem e variantes por aprendizagem de máquinas	57
4.3	Complexo Spike selvagem e variantes Tmprss2	59
4.3.1	Raiz do Desvio Quadrado Médio (RMSD)	59
4.3.2	Flutuação da Raiz Quadrada Média (RMSF)	60
4.3.3	Raio de Giração (Rg)	63
4.3.4	Área Acessível ao Solvente (SASA)	64
5.	CONCLUSÃO	64

1. INTRODUÇÃO E JUSTIFICATIVA

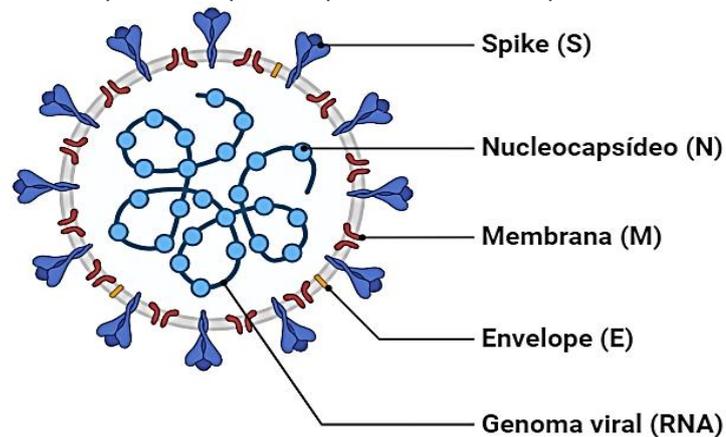
1.1 COVID-19

Desde a declaração da Organização Mundial de Saúde (OMS), em março de 2020, a contaminação global devido à COVID-19, doença infecciosa causada pela Síndrome Respiratória Aguda Grave do Coronavírus-2 (do inglês Severe Acute Respiratory Syndrome Coronavirus-2, SARS-CoV-2), registrou mais de 700 milhões de casos confirmados, incluindo mais de 6 milhões de mortes (WORLD HEALTH ORGANIZATION, 2023). O vírus SARS-CoV-2 foi isolado inicialmente em dezembro de 2019 em Wuhan, província de Hubei na China, após originar um grande surto da doença respiratória aguda (JAIMES *et al.*, 2020). A COVID-19 é uma doença respiratória transmitida por meio de aerossóis, que pode causar perda de olfato ou de paladar, febre, tosse seca, dores de cabeça, cansaço e dificuldade para respirar. Entretanto, não está limitada aos sintomas respiratórios e pode manifestar complicações inflamatórias, cardíacas, cardiovasculares, tromboembólicas e neurológicas (CHOUDHARY *et al.*, 2021).

1.2 SARS-CoV-2

SARS-CoV-2 é um vírus de RNA de fita simples positiva de aproximadamente 30.000 bases, do gênero *Betacoronavirus*, pertencente à família *Coronaviridae*, e codifica um polipeptídeo de aproximadamente 9860 aminoácidos de comprimento (POTDAR *et al.*, 2020; SALLEH; DERIS, 2022). É constituído por quatro proteínas estruturais, as proteínas Spike, envelope, membrana e nucleocapsídeo (Figura 1), juntamente com dezesseis não estruturais e cinco a oito proteínas acessórias (DU *et al.*, 2009).

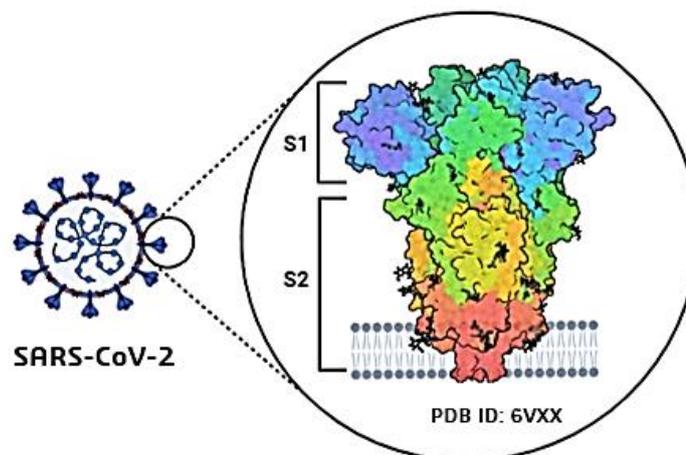
Figura 1. Estrutura do vírus SARS-CoV-2. M: proteína da membrana, N: proteína do nucleocapsídeo, S: proteína Spike, E: proteína do envelope.



Fonte: da autora (2023).

A proteína Spike (Figura 2) é uma glicoproteína transmembranar trimérica de 1273 aminoácidos, onde cada um dos três monômeros é composto por duas subunidades funcionalmente distintas, S1 (aminoácidos 14-685) e S2 (aminoácidos 686-1273), precedidas por um peptídeo sinal (aminoácidos 1-13) (SAKKIAH *et al.*, 2021; XIA *et al.*, 2020a). A subunidade S1 se liga ao receptor enzima conversora de angiotensina 2 (ACE2) do hospedeiro e a subunidade S2 medeia a fusão da membrana celular para a entrada da partícula viral (DU *et al.*, 2017).

Figura 2. Estrutura da proteína Spike trimérica. S1: subunidade 1, S2: subunidade 2.

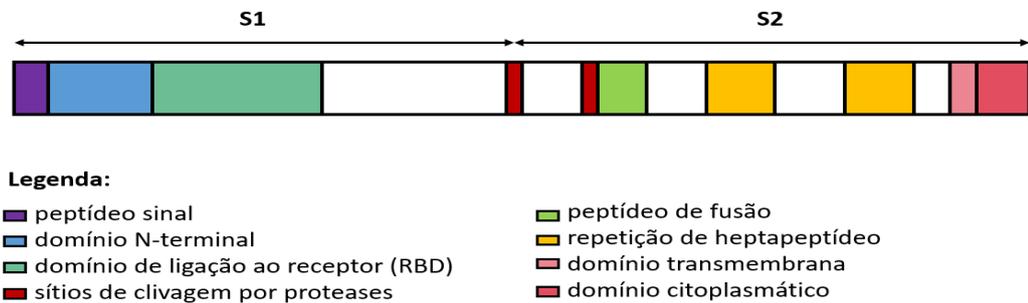


Fonte: da autora (2023).

A Figura 3 mostra a representação esquemática dos domínios da proteína Spike. A subunidade S1 é dividida em domínio N-terminal (aminoácidos 14-305), domínio de ligação ao receptor (RBD; aminoácidos 319-541) e um domínio carboxi-terminal (XIA *et al.*, 2020). A subunidade S2 consiste em um peptídeo de fusão (FP; aminoácidos 788-806) composto por

resíduos hidrofóbicos, repetição de heptapeptídeo 1 (HR1; aminoácidos 912-984), repetição de heptapeptídeo 2 (HR2; aminoácidos 1163-1213), um domínio transmembrana (aminoácidos 1213-1237) e um domínio citoplasmático (aminoácidos 1237-1273) (SAKKIAH *et al.*, 2021).

Figura 3. Representação esquemática dos domínios proteína Spike.



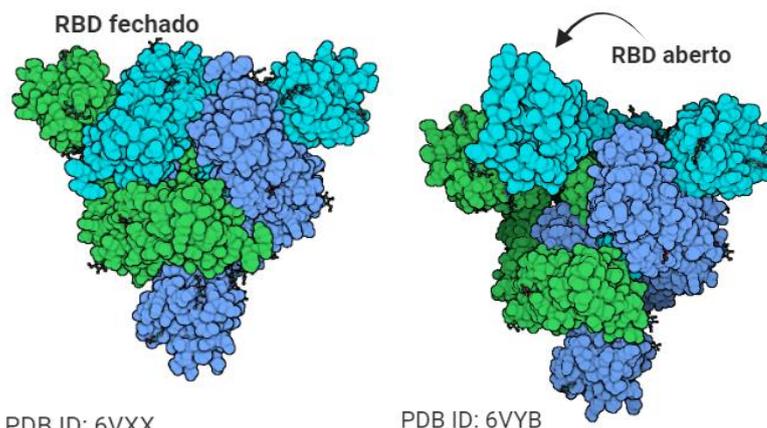
Fonte: da autora (2023).

Mecanismo de entrada do vírus na célula hospedeira

A enzima conversora de angiotensina humana 2 (ACE2), serina protease transmembrana 2 (TMPRSS2) e proteína Spike são os principais fatores que influenciam na infecção por SARS-CoV-2 (SENAPATI *et al.*, 2021).

O domínio de ligação ao receptor (RBD) da proteína Spike é flexível e sofre alterações estruturais entre uma conformação “fechada/para baixo” e uma conformação “aberta/para cima” menos estável, como representado na Figura 4 (WALLS *et al.*, 2020). Na conformação fechada, o RBD do domínio S1 tem sua superfície de ligação ao receptor enterrada, e não pode interagir com o receptor. Na posição aberta, a superfície de ligação ao receptor está totalmente exposta, podendo interagir com o ACE2 (WRAPP *et al.*, 2020).

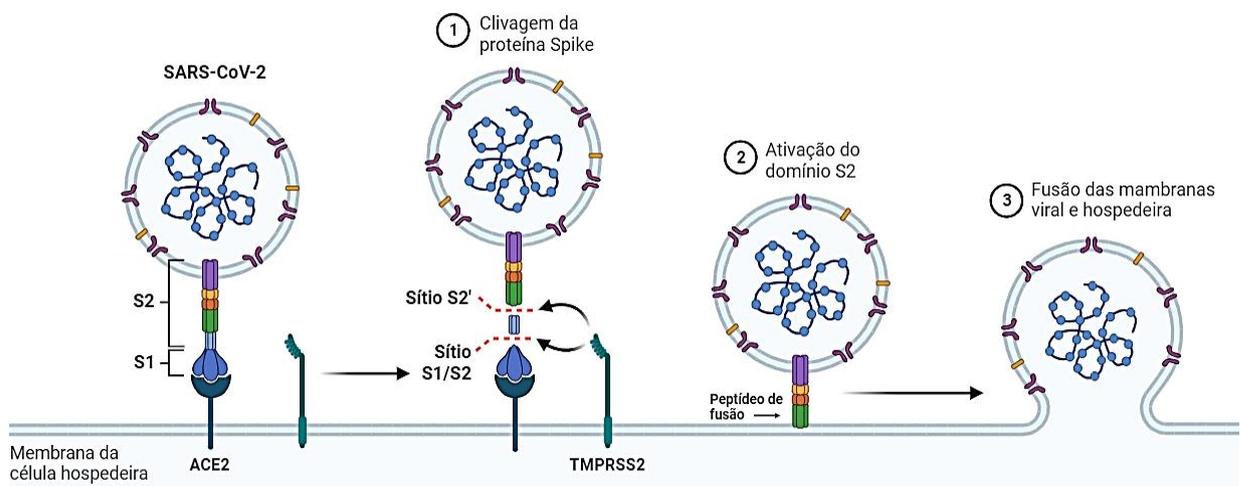
Figura 4. À esquerda, representação estrutural da proteína Spike com o domínio RBD da subunidade S1 na conformação fechada. À direita, RBD na conformação aberta.



Fonte: da autora (2023).

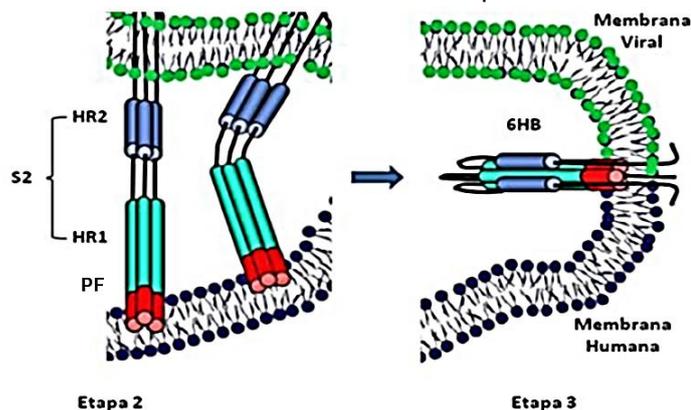
A ligação do RBD da proteína Spike ao receptor ACE2 ativa a clivagem proteolítica por proteases da célula hospedeira próxima da junção entre as subunidades S1 e S2, gerando dois fragmentos distintos da proteína Spike, S1/S2 e S2' (HOFFMANN *et al.*, 2020). A clivagem no local S1/S2 é realizada pela enzima proteolítica furina, e a clivagem S2' é realizada pela protease TMPRSS2 (HUANG *et al.*, 2020). Após a clivagem da proteína Spike, o peptídeo de fusão da Spike fica exposto, ocorrendo uma interação subsequente entre os domínios de repetição do heptapeptídeo (HR1 e HR2) da subunidade S2, resultando em um feixe helicoidal de seis hélices estável (6-HB) que aproxima as membranas viral e celular, promovendo a entrada do vírus (XIA *et al.*, 2020b). As etapas do mecanismo de entrada do vírus na célula hospedeira estão esquematizadas nas Figuras 5 e 6.

Figura 5. Representação das etapas do mecanismo de entrada do SARS-CoV-2 na célula do hospedeiro após a ligação da Spike ao receptor ACE2.



Fonte: da autora (2023).

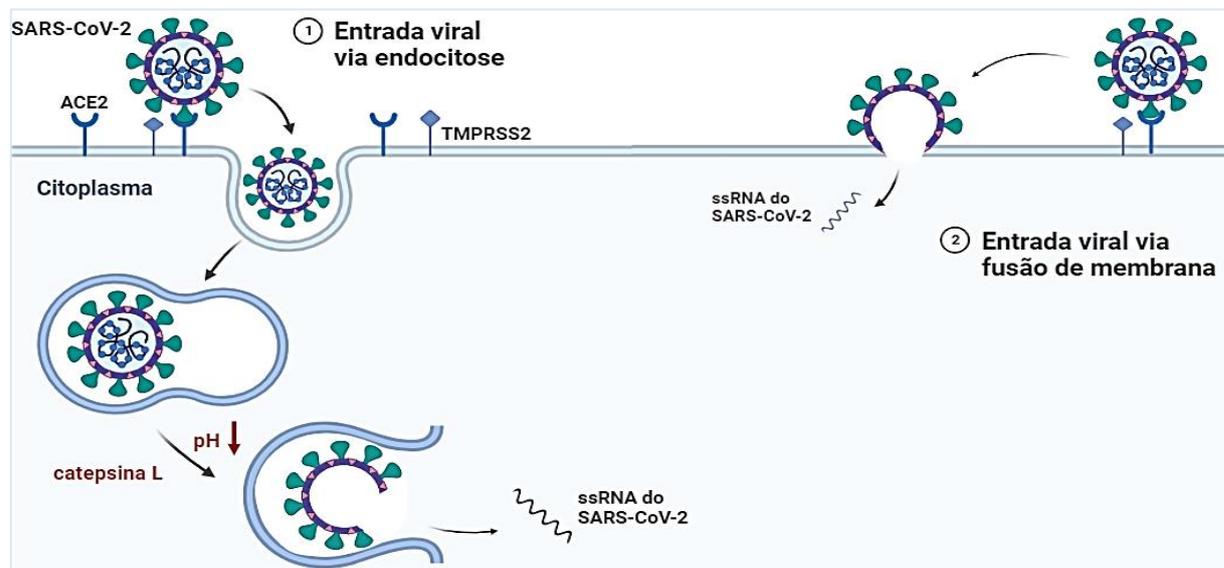
Figura 6. Representação da ativação do domínio S2, com subsequente interação entre os domínios de repetição do heptapeptídeo HR1 e HR2 e formação do feixe 6-HB, promovendo a aproximação das membranas viral e do hospedeiro.



Fonte: adaptado de LIMA *et al.* (2020).

Após a ligação do SARS-CoV-2 à superfície celular de uma célula hospedeira, o TMPRSS2 pode induzir a entrada viral na célula por outro mecanismo além da clivagem da glicoproteína Spike, que faz a fusão direta da membrana viral e do hospedeiro, levando à liberação do ssRNA viral no citoplasma. Alternativamente, o TMPRSS2 pode clivar o receptor ACE2, que então aumenta a captação viral através da endocitose. Assim, o SARS-CoV-2 entra nos compartimentos endossomais, onde uma diminuição no pH endossomal estimula enzimas catepsina L que clivam e ativam a proteína Spike, facilitando a liberação do ssRNA viral no citosol. As duas vias de entrada do SARS-CoV-2 na célula estão representadas na Figura 7. (MAHMOUD; JARRAR, 2021).

Figura 7. Ativação e entrada do SARS-CoV-2 na célula hospedeira por duas vias: endocitose e fusão direta de membrana.



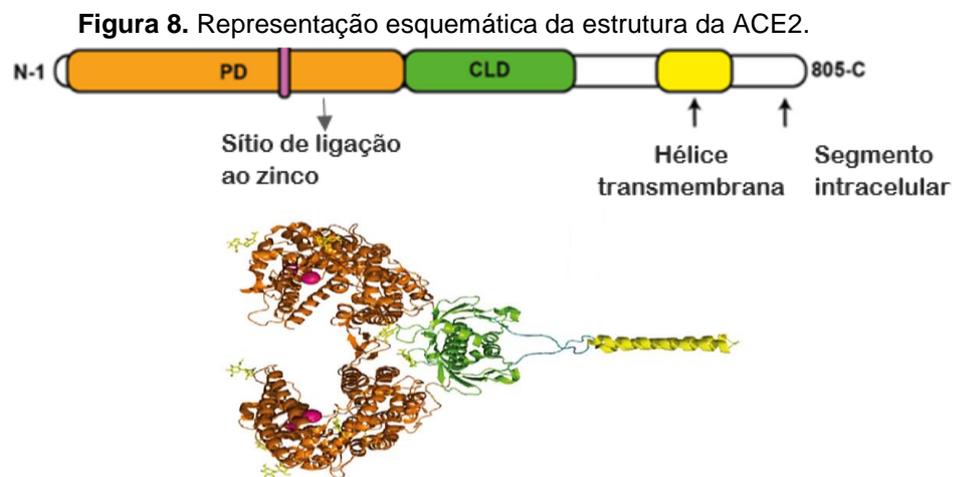
Fonte: da autora (2023).

1.3 ACE2

A enzima conversora de angiotensina 2 é uma proteína longa de 805 aminoácidos, codificada no cromossomo X, pertencente à família da enzima de conversão de angiotensina. A ACE2 é um receptor funcional para a glicoproteína Spike do SARS-CoV-2 (KOWALCZUK *et al.*, 2008).

A ACE2 é amplamente expressa no coração, vasos, intestino, pulmão, rim, testículo e cérebro (VERDECCHIA *et al.*, 2020). Pertence ao sistema renina angiotensina aldosterona (RAAS) e catalisa a conversão da angiotensina I e a angiotensina II em angiotensina 1-9 e angiotensina 1-7, respectivamente (HOU *et al.*, 2020).

A estrutura da ACE2, representada na Figura 8, é formada por um domínio peptidase (aminoácidos 19-615), dividido em dois subdomínios: subdomínio I (aminoácidos 19-102, 290-397 e 417-430) e subdomínio II (aminoácidos 103-289, 398-416 e 431-615). Após o domínio peptidase (PD), existe um domínio semelhante à colectrina (CLD) (aminoácidos 616 a 768). A região transmembrana é formada por uma hélice (740-768 aa) e o segmento intracelular é composto por aproximadamente 40 aminoácidos. A ACE2 está disponível como um homodímero, estabilizado por interações polares entre os resíduos PD e possui um sítio de ligação ao zinco coordenado pelos aminoácidos 374-HEMGH-378 pertencentes ao subdomínio I (ZEPEDA-CERVANTES *et al.*, 2022). Além disso, o domínio C-terminal transmembranar (resíduo 697 a 716) da ACE2 é clivado pela TMPRSS2, o que resulta em mudanças conformacionais no complexo S1-ACE2, facilitando a entrada viral no hospedeiro (SENAPATI *et al.*, 2021; SHULLA *et al.*, 2011).



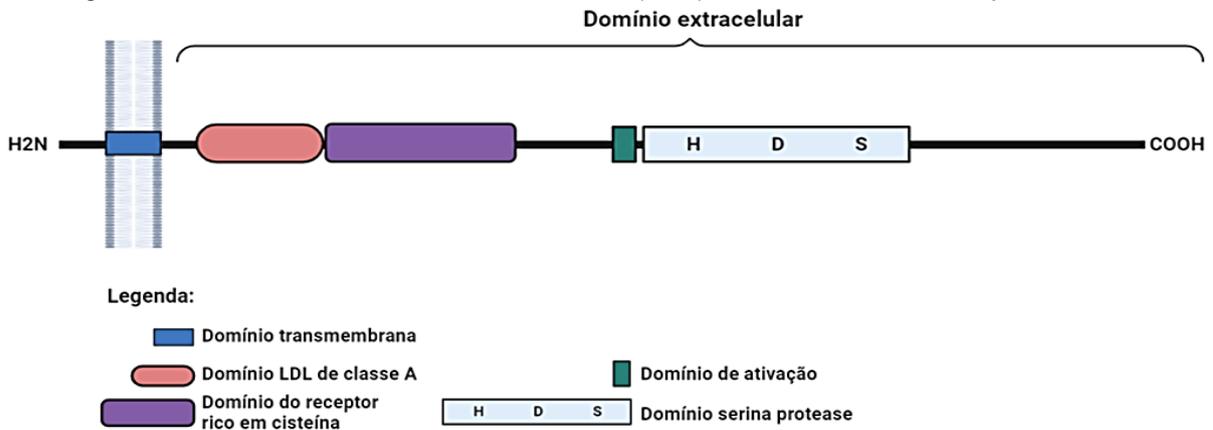
Fonte: adaptado de ZEPEDA-CERVANTES *et al.* (2022).

1.4 TMPRSS2

A protease transmembrana serina 2 (Figura 9) é uma protease ancorada na membrana de ~70kDa, codificada por um gene localizado no cromossomo 21, e é composta de 492 aminoácidos com os seguintes domínios: um domínio citoplasmático intracelular N-terminal (aminoácidos 1-84), uma região transmembranar (aminoácidos 85-105) e uma região extracelular C-terminal (aminoácidos 106-492) que contém um domínio do receptor de LDL do tipo A (aminoácidos 112-148; representa um sítio de ligação para o cálcio), um domínio do receptor rico em cisteína (aminoácidos 153-244; envolvido na ligação a moléculas extracelulares) e um domínio C-terminal da peptidase S1 que cliva em arginina (Arg) ou lisina

(Lys) (aminoácidos 256-489) (HUSSAIN *et al.*, 2020, MAHMOUD; JARRAR, 2021). O domínio catalítico contém uma tríade catalítica constituída pelos resíduos de aminoácidos His296, Asp345 e Ser44 (SHEN *et al.*, 2017).

Figura 9. Mapa linear dos domínios da proteína TMPRSS2. O terminal C (COOH) contém os domínios necessários para a ligação das moléculas extracelulares, além do domínio transmembrana para ancoragem de membrana. A extremidade N-terminal (H2N) contém uma cauda citoplasmática.



Fonte: da autora (2023).

A TMPRSS2 está envolvida nas ligações peptídicas de clivagem de proteínas que possuem serina no sítio ativo. Sua função biológica exata é desconhecida, porém, pesquisas demonstram que está associada ao câncer de próstata (SHABIR, 2021) e a algumas infecções virais, incluindo influenza, SARS e MERS (SARKER *et al.*, 2021).

A TMPRSS2 é expressa principalmente na próstata, cólon, estômago, intestino delgado, pulmão e glândula salivar (SINGH *et al.*, 2021b). O gene TMPRSS2 tem sua expressão dependente de andrógenos e encontra-se mais significativamente em tecidos sensíveis a andrógenos, como próstata e testículos, do que em outros tecidos humanos (LIN *et al.*, 1999). O andrógeno regula os níveis de expressão de TMPRSS2 ligando-se ao elemento de resposta androgênica, que está localizado no promotor de TMPRSS2 (CLINCKEMALIE *et al.*, 2013). A expressão de TMPRSS2 pode ser maior no pulmão dos homens, o que pode aumentar a capacidade viral de entrar nas células hospedeiras. Esses podem ser fatores que estão associados à predominância masculina da infecção por SARS-CoV-2 (SINGH *et al.*, 2021b).

1.5 Resíduos envolvidos na interação Vírus-Hospedeiro

Os resíduos de aminoácidos entre N437 e T508 do RBD da proteína Spike interagem com o receptor ACE2 (WANG, Q. *et al.*, 2020). Os resíduos do RBD K417, G446, T449, N487,

Q493, Q498, T500, N501, G502 e T505 fazem interações polares com ACE2 (SIRONI *et al.*, 2020). Os principais resíduos na interface ACE2 incluem S19, Q24, T27, F28, D30, K31, H34, E35, E37, D38, Y41, Q42, L45, L79, M82, Y83, T324, Q325, G326, E329, N330, K353, G354, D355, R357, P389 e R393 (LI *et al.*, 2005). Os resíduos de ACE2 humano K31, E35, D38, T41, M82 e K352, são aminoácidos interativos chave (LI, 2013). O resíduo de aminoácido de ACE2 M82 foi identificado por possuir forte interação com o resíduo F486 da proteína Spike (LUAN *et al.*, 2020). Os resíduos Y489 e L455 foram identificados como resíduos chave no RBD de SARS-CoV-2. Na ACE2, os resíduos chave identificados foram Q24, T27, D30, H34, Y83, K353 e D355 (JAWAD *et al.*, 2021).

Os resíduos Q498, T500 e N501 do RBD viral interagem com Q42 e T41 da ACE2 por meio de ligações de hidrogênio. Outras interações chave incluem os resíduos K417 e T453 do RBD com resíduos D30 e H34 da ACE2, respectivamente (YAN *et al.*, 2020).

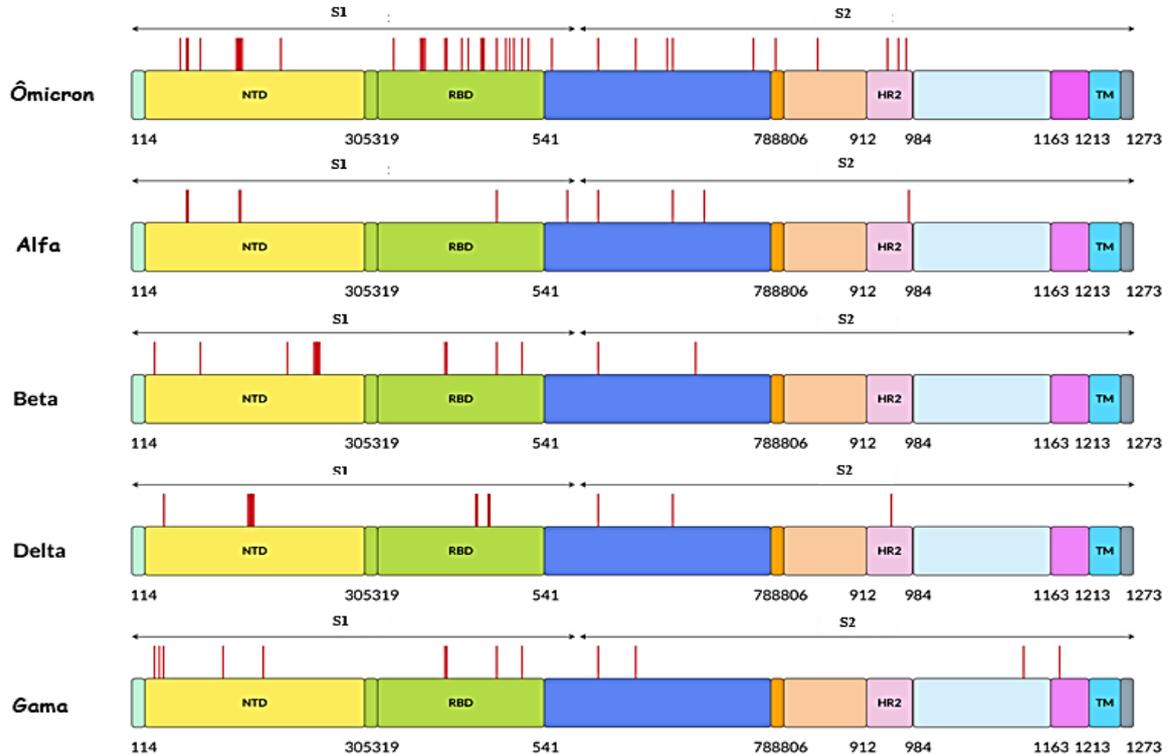
Os resíduos de aminoácidos T64, H66, F79, D80, T95, K187, V213, R214, D215, L242, H245, R246, D614, G261, A262 e R667 da proteína Spike estão envolvidos na interação com TMPRSS2 e são importantes para a estabilização da ligação (SENAPATI *et al.*, 2021). Os principais resíduos de aminoácidos do sítio ativo de TMPRSS2 que fazem contato com a proteína Spike são Q276, E299, K300, P301, K340, K342, E389, K390, L419, Q438, S441 e W461. Os resíduos H296, S441, K432, W461 e Q438 são altamente conservados entre outras serina proteases e são importantes na atividade catalítica da proteína (VANKADARI *et al.*, 2022).

1.6 Variabilidade do Vírus

O SARS-CoV-2 é um vírus de RNA que possui alta taxa de mutação, a qual pode alterar a estrutura proteica da glicoproteína Spike e, dessa forma, conferir vantagens seletivas, como maior transmissibilidade e infecciosidade (SALLEH; DERIS, 2022).

As variantes virais com mutações no domínio de ligação ao receptor (RBD) da Spike e que possuem maior afinidade de ligação do complexo RBD-ACE2, com alta taxa de transmissão e disseminação nas populações humanas são designadas variantes de preocupação (do inglês Variants of Concern, VOC) (WORLD HEALTH ORGANIZATION, 2022; CDC, 2022). As variantes de SARS-CoV-2 consideradas VOCs são Alfa (B.1.1.7), Beta (B.1.351), Gama (P.1), Delta (B.1.617.2) e Ômicron (B.1.1.529) (TAO *et al.*, 2021) e estão esquematizadas na Figura 10.

Figura 10. Representação dos diferentes locais de mutação de resíduos de aminoácidos localizados da proteína Spike das cinco variantes de SARS-CoV-2. As linhas vermelhas indicam os locais de mutação na proteína Spike do SARS-CoV-2 nas posições especificadas. NTD: domínio N-terminal, RBD: domínio de ligação ao receptor, HR2: repetição de heptapeptídeo 2, TM: domínio transmembranar, S1: subunidade 1, S2: subunidade 2.



Fonte: adaptado de ARAF et al. (2022).

A variante alfa (B.1.1.7) foi detectada no Reino Unido em setembro de 2020 e o seu aumento de transmissibilidade foi relacionado com a mutação N501Y no RBD da proteína Spike (LA ROSA., 2021). Esta linhagem possui 23 mutações (14 não sinônimas; 6 sinônimas) e 3 deleções. As mutações mais significativas são H69-V70del, N501Y e P681H (MOHAMMADI *et al.*, 2021).

A variante beta (B.1.351) foi detectada em outubro de 2020 na população sul-africana e apresenta 12 mutações não sinônimas e uma deleção (TAO *et al.*, 2021). Dez dessas mutações, incluindo a deleção, encontram-se na Spike (L18F, D80A, D215G, LAL 242-244 DEL, R246I, K417N, E484K, N501Y, D614G e A701V), sendo N501Y, K417N e E484K mutações no RBD da proteína Spike (TANG *et al.*, 2021).

A linhagem P.1 (variante gama) foi detectada em janeiro de 2021 no Brasil e apresentou três mutações preocupantes no RBD da Spike: N501Y, K417T e E484K (SING *et al.*, 2021a). Nesta variante há 17 mutações não sinônimas, 4 mutações sinônimas e 1 deleção. Na proteína

Spike encontram-se 12 mutações (L18F, T20N, P26S, D138Y, R190S, K417T, E484T, N501Y, D614G, H655Y, T1027I e V1176F) (MOHAMMADI *et al.*, 2021). A variante gama, também denominada B.1.1.28.1, pertence à linhagem B.1.1.28. Conjuntamente, se encontrou em circulação no país a variante zeta, da linhagem P.2, também denominada B.1.1.28.2 que, assim como a variante P.1, possui a mutação E484K (WORLD HEALTH ORGANIZATION, 2022).

A variante delta (B.1.617.2) foi detectada na Índia em dezembro de 2020 e possui 17 mutações, todas localizadas na Spike (T19R, V70F, T95I, G142D, E156-, F157-, R158G, A222V, W258L, K417N, L452R, 7478K, D614G, P681R, D950N, E484 e QL452R). Apenas as mutações K417N, L452R e T478K estão localizadas no RBD (MOHAMMADI *et al.*, 2021). A substituição de leucina por arginina na posição 452 está presente apenas na variante Delta, e é conhecida por aumentar a afinidade pelos receptores ACE2 (ARAF *et al.*, 2022).

A variante ômicron (B.1.1.529) foi anunciada como uma VOC pela Organização mundial de saúde (OMS) em 26 de novembro de 2021 (KUMAR *et al.*, 2021). Essa variante possui mais de 2,9 mil deleções e mais de 11 mil mutações não sinônimas. Na proteína Spike, são 32 mutações, sendo 16 substituições de aminoácidos chave no RBD da proteína Spike (G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H, T547K) (ARAF *et al.*, 2022; CDC, 2021). Foi revelado que a ômicron pode ser dez vezes mais contagiosa que o SARS-CoV-2 original e cerca de duas vezes mais infecciosa que a variante delta, principalmente devido às suas mutações RBD N440K, T478K e N501Y (CHEN *et al.*, 2021).

1.7 Variantes da ACE2 e TMPRSS2

Resíduos específicos são responsáveis pelas interações moleculares entre a Spike com ACE2 e TMPRSS2. Essas interações podem ser alteradas por mutações, afetando a estabilidade estrutural, a afinidade de ligação e o nível de expressão dessas três proteínas, sendo fatores para a suscetibilidade e heterogeneidade clínica para a COVID-19 (SENAPATI *et al.*, 2021).

1.7.1 Mutações na ACE2

O ACE2 pode desempenhar o papel de um receptor eficiente ou ineficiente para SARS-CoV-2 caso determinados aminoácidos sejam alterados. Assim, variantes de ACE2 podem influenciar na associação entre ACE2 e proteína Spike do SARS-CoV-2 (SEYEDALINAGHI *et al.*, 2021). A mutação pode ser um dos fatores que justifica as diferenças na suscetibilidade

e gravidade da infecção por SARS-CoV-2 nos indivíduos (FIJIKURA *et al.*, 2020). Os resíduos K26, T27, E37, K68, D206, G211, R219, G326, K341, G352, V447, I468, F486, R559 de ACE2 são importantes para determinar a força da interação Spike-ACE2 (SENEPATI *et al.*, 2021).

As variantes não sinônimas da ACE2 K31R, H34R, E35K, E37K, D38V, Y50F, N51S, K68E, F72V, Y83H, G326E, N33I, G352V, D355N e Q388L foram relacionadas com a redução da interação da ACE2 com a proteína Spike. Por outro lado, as variantes S19P, I21V, E23K, K26R, K26E, T27A, N64K, T92I, Q102P, H378R e M383T podem ser responsáveis pelo aumento na afinidade entre ACE2 e SARS-CoV2 (SURYAMOHAN *et al.*, 2021). Foi relatado que três variantes da ACE2 – K26R (rs4646116), M82I (rs267606406) e E329G (rs143936283) – estão associadas a maior afinidade de ligação com o domínio RBD da proteína Spike em comparação com a ACE2 de tipo selvagem. Isso pode resultar em maior suscetibilidade à infecção por SARS-CoV-2 (WANG, J. *et al.*, 2020). K26 interage com o glicano ligado a N90 da ACE2 por meio de contatos polares. Este glicano possui um obstáculo estérico, o qual pode diminuir a afinidade da ACE2 pelo RBD da Spike. A afinidade aumentada para o R26 da ACE2 com a Spike-WT ocorre pela anulação da estabilização dos contatos polares com o N90, alterando a conformação do glicano (RODRIGUEZ *et al.*, 2022). Além disso, o R26 estabelece interações de cadeia principal e lateral com o resíduo D30 da ACE2, que faz uma ponte salina com o resíduo K417 do RBD. Dessa forma, R26 estabiliza as alfa-hélices centrais, aumentando a afinidade de ligação do ACE2 ao RBD (SURYAMOHAN *et al.*, 2021; AL-MULLA *et al.*, 2021).

A variante G211R demonstrou menor estabilidade em relação à variante selvagem, conferindo uma ampla flexibilidade devido à capacidade de envolver diferentes interações com os resíduos de aminoácidos próximos (BENETTI *et al.*, 2020).

1.7.2 Mutações na TMPRSS2

Mutações no gene TMPRSS2 têm sido associadas à suscetibilidade de infecção por SARS-CoV-2 (SCHÖNFELDER *et al.*, 2021). Um estudo mostrou que a variante não sinônima V160M da TMPRSS2 foi a mais provável de afetar a função e estabilidade da proteína TMPRSS2 (PARINI, HOSSEINI, AKHAVAN-NIAKI, 2020). Foi demonstrado que a clivagem da proteína Spike por TMPRSS2 reduziu aproximadamente 2,4 vezes em células que expressam a variante V160M em pacientes da Índia (RAVIKANTH *et al.*, 2021). Em outro estudo, não foi encontrada interação direta da variante V160M com a Spike, mas potencialmente influenciando a conformação estrutural (SENEPATI *et al.*, 2020). Além disso,

pacientes com COVID-19 portadores do alelo selvagem (do inglês *Wild-Type*, WT) da TMPRSS2 portaram uma taxa de infecção maior em comparação aos pacientes com a mutação V160M, sugerindo que a mutação reduz a funcionalidade de TMPRSS2 (RAVIKANTH *et al.*, 2021). No estudo de Duman *et al.* (2022), foi demonstrado que na variante V160M a rede hidrofóbica não pode ser mantida e há um choque estérico entre a Metionina e aminoácidos próximos, sugerindo que a mutação pode desestabilizar a proteína.

1.8 Justificativa

As variantes de SARS-CoV-2 e mutações encontradas na ACE2 e TMPRSS2 podem ser fatores importantes para o entendimento da COVID-19. Compreender a complexa interação entre mutações virais e variações genéticas do hospedeiro é crucial para desvendar os fatores que influenciam o desenvolvimento da doença. Uma das áreas de investigação é o entendimento de como essas variações impactam a estabilidade e afinidade dos complexos Spike-ACE2 e Spike-TMPRSS2, que são críticos para a entrada viral nas células hospedeiras. Certas mutações na Spike podem aumentar a capacidade do vírus de interagir mais fortemente com o receptor ACE2, levando potencialmente ao aumento da replicação viral e da infecciosidade. Entretanto, como essas mutações contribuem para melhorar a estabilidade e afinidade de interação entre os complexos SARS-CoV-2-ACE2 e SARS-CoV-2-TMPRSS2 não é totalmente compreendido.

A análise dos movimentos conformacionais das estruturas, assim como dinâmica molecular são exemplos de abordagens empregadas na tentativa de alcançar total compreensão do processo. As simulações de dinâmica molecular são utilizadas para avaliar o efeito de uma mutação ou comparar mudanças estruturais entre complexos, gerando grandes quantidades de dados de milhares de átomos a cada intervalo de tempo. A análise de trajetórias complexas pode ser realizada através de diversas abordagens, incluindo análise da estabilidade e afinidade entre as estruturas, avaliação de propriedades termodinâmicas e investigação das ligações e interações presentes. Assim, neste trabalho, simulações de dinâmica molecular foram realizadas a fim de revelar as diferenças em linhagens de SARS-CoV-2 e investigar o impacto da variabilidade genética do SARS-CoV-2 e de variantes da ACE2 e TMPRSS2 na região de interação entre vírus e hospedeiro.

Interpretar e extrair informações significativas destas trajetórias requer uma análise robusta e não é uma tarefa simples. Devido à natureza altamente dimensional das simulações, torna-se difícil a extração de algumas características das trajetórias e a compreensão mais

profunda dos processos moleculares, tais como regiões ou resíduos que podem contribuir sutilmente para as interações proteicas. Dessa forma, no laboratório da orientadora foram utilizadas técnicas de aprendizado de máquina para analisar os extensos conjuntos de dados obtidos nas simulações de DM, ajudando a identificar distinções cruciais entre as trajetórias, mesmo quando essas diferenças são sutis.

Desvendar os mecanismos pelos quais estas mutações influenciam a entrada e replicação viral pode abrir novos caminhos para intervenções terapêuticas. Os resultados podem ajudar no desenvolvimento e/ou triagem direcionada de moléculas disruptivas e/ou inibidoras, visando a entrada viral na célula hospedeira, sendo útil para o desenvolvimento de medicamentos personalizados para o tratamento de pacientes com COVID-19 e para o desenvolvimento de vacinas.

2. OBJETIVO

O presente trabalho teve como objetivo avaliar por meio de simulações de DM os efeitos da variabilidade genética do SARS-CoV-2 na interação com a ACE2 e TMPRSS2 de tipo selvagem, e os efeitos de mutações não sinônimas da ACE2 e TMPRSS2 na região de interação com a Spike selvagem. Além disso, identificar resíduos relevantes para a afinidade de ligação em variantes da ACE2 e TMPRSS2 presentes na população brasileira, uma vez que algumas variantes podem aumentar ou reduzir a afinidade da interação entre as proteínas.

3. MATERIAL E MÉTODOS

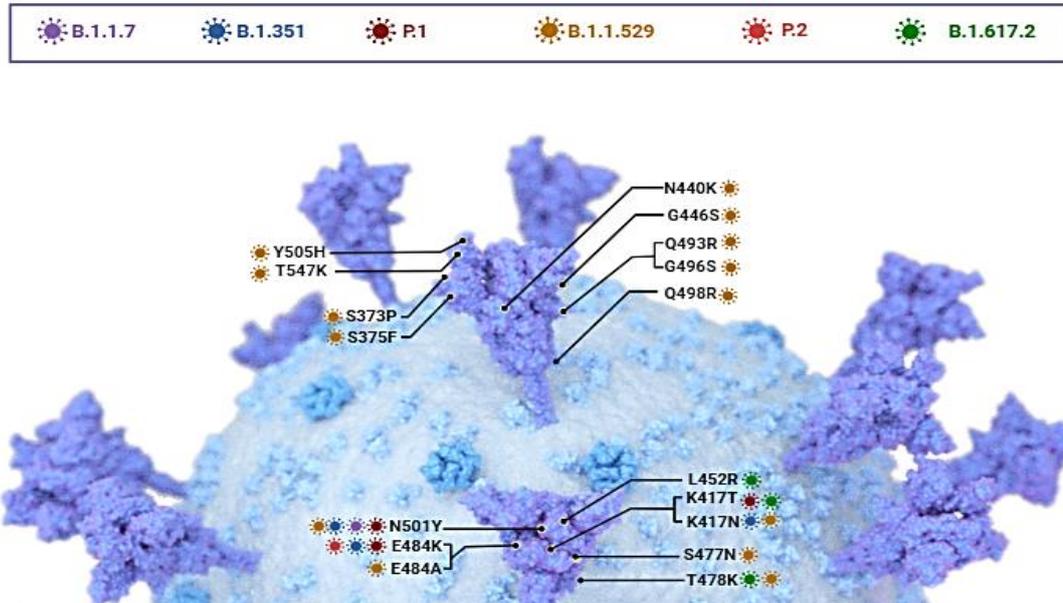
3.1 Obtenção dos Dados

3.1.1 Variantes da Spike

As variantes do SARS-CoV-2 foram selecionadas de acordo com as variantes de preocupação (VOC) durante o curso da pandemia, cada uma com mutações específicas que levantaram a preocupações de saúde globais: Ômicron (B.1.1.529 – vários países), Alfa (B.1.1.7 – Estados Unidos), Beta (B.1.351 – África do Sul), Gama (P.1 – Brasil) e Delta (1.617.2 - Índia). Além das VOCs, também foi selecionada a variante de interesse P2 (ou variante Zeta) (B.1.1.28.2), que foi detectada no Rio de Janeiro, Brasil, em outubro de 2020. (WHO - <https://www.who.int/pt>). As variantes selecionadas para este trabalho e suas respectivas

mutações na região do RBD da Spike estão ilustradas na Figura 11.

Figura 11. Variantes do SARS-CoV-2 e mutações localizadas do RBD da proteína Spike.



Fonte: da autora (2023).

3.1.2 Variantes da ACE2 e TMPRSS2

As variantes das proteínas ACE2 e TMPRSS2 foram selecionadas utilizando a base de dados *Genome Aggregation Database* (GnomAD – <https://gnomad.broad.institute.org>) e Arquivo Brasileiro Online de Mutações (ABraOM – <http://abraom.ib.usp.br>).

A seleção das mutações não sinônimas da ACE2 foi efetuada utilizando os seguintes critérios: resíduos de aminoácidos localizados na região do domínio peptidase da ACE2 (aminoácidos 19-614) e identificados como resíduos críticos na interação entre a ACE2 e o RBD da Spike segundo a literatura; variantes encontradas em amostras da população brasileira depositados na base de dados ABraOM e que estão em maior frequência alélica (do inglês *allele frequency*, AF) na população segundo dados do GnomAD. As variantes selecionadas foram K26R, G211R, R219C e K341R. Os resíduos Lys26, Gly211, Arg219, Lys341 da ACE2 são importantes para determinar a força da interação Spike-ACE2 (SENEPATI *et al.*, 2021).

A proteína ACE2 é altamente conservada e possui poucas variantes não sinônimas (LAN *et al.*, 2020). Apenas a variante K26R possui frequência alélica global acima de 0.1% (AF = 0.4%). G211R e K341R possuem AF de 0.1% e R219C AF < 0.1%.

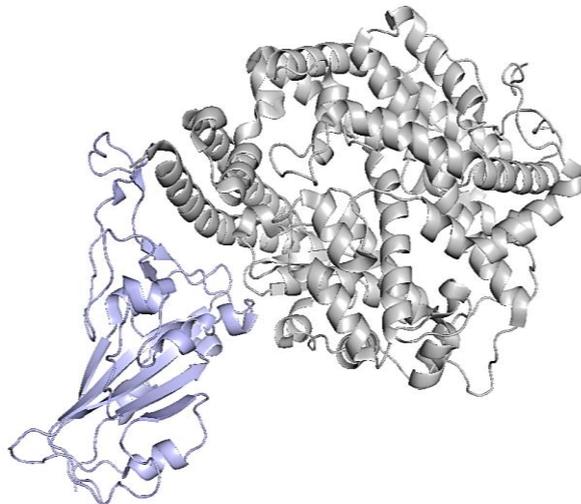
As variantes da TMPRSS2 foram selecionadas de acordo com amostras encontradas na população brasileira e depositadas na base de dados ABraOM e, destas, foram escolhidas as

com maior frequência alélica global segundo dados do GnomAD. As variantes selecionadas foram: V160M, F209I, P375S e V415I. Apenas V160M possui frequência alélica global significativa (AF = 25%). As demais são variantes raras, sendo P375S com AF = 0.1%, F209I e V415I (AF < 0.1%).

3.1.2 Estruturas Terciárias das Proteínas

A estrutura terciária do complexo formado entre a proteína ACE2 e o RBD da proteína Spike foi resolvida por meio de difração de raio X com fator de resolução de 2.50 Å e encontra-se depositada na base de dados *Protein Data Bank* (PDB - <https://www.rcsb.org/>) com o PDB ID 6LZG (Figura 12).

Figura 12. Estrutura terciária do complexo ACE2-RBD-Spike.

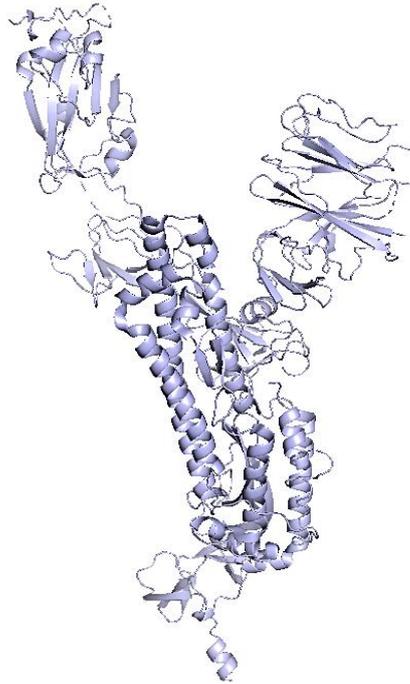


Fonte: da autora (2023).

O *software* Modeller v9.23 (SALI e BLUNDELL, 1993) foi utilizado para preencher os átomos e resíduos faltantes. Foram adicionados átomos de hidrogênio por meio do *software* UCSF CHIMERA versão 1.14 (PETTERSEN *et al.*, 2004).

A estrutura terciária da Spike (Figura 13) está depositada no PDB (ID 6VSB), resolvida por microscopia eletrônica com 3.46 Å.

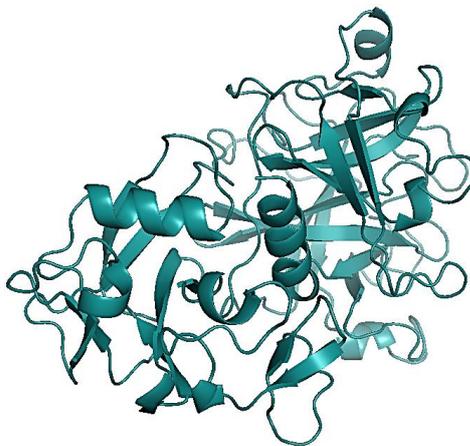
Figura 13. Estrutura terciária da proteína Spike.



Fonte: da autora (2023).

A parte extracelular da estrutura terciária da TMPRSS2 (Figura 14) foi resolvida por meio de difração de raio X com fator de resolução 1.95 Å e está depositada no PDB (PDB ID 7MEQ). Entretanto, a estrutura não está completamente resolvida, além de possuir resíduos faltantes no cristal. Portanto, no presente trabalho, foi utilizada a estrutura extracelular completa modelada por um dos membros do nosso laboratório, Wasim Syed (manuscrito em preparação).

Figura 14. Estrutura completa da TMPRSS2 por modelagem molecular.



Fonte: da autora (2023).

3.2 Docking entre Spike e TMPRSS2

As estruturas proteicas TMPRSS2 e Spike foram submetidas ao *docking* molecular usando o servidor web HADDOCK 2.2 (VAN ZUNDERT *et al.*, 2006). HADDOCK realiza *docking* direcionado. Assim, os sítios de clivagem na Spike foram definidos como resíduos ativos, e os sítios de ligação ao substrato e sítios ativos foram definidos como resíduos ativos na proteína TMPRSS2. O TMPRSS2 cliva a proteína Spike em dois locais potenciais, Arg685/Ser686 e Arg815/Ser816 (HUSSAIN *et al.*, 2020), e também em posições ricas em arginina (HOFFMANN *et al.*, 2020). Baseado no trabalho de Hussain e colaboradores (2020) e no trabalho de Vardhan e Sahoo (2022), os resíduos da Spike escolhidos para o *docking* foram os do sítio de clivagem na posição Arg815/Ser816, responsável pela formação dos fragmentos S1/S2 e S2'. Já para a TMPRSS2, foram utilizados os resíduos do sítio ativo HIS296, ASP345, ILE346, ASP435, SER441, SER460, GLY462, SER465 e LYS467.

Todas as simulações de *docking* foram realizadas usando parâmetros-padrão. O melhor modelo, com o menor valor de *Z score* foi selecionado para as etapas de dinâmica molecular.

3.3 Construção dos complexos com as variantes da Spike e polimorfismos da ACE2 e TMPRSS2

Os resíduos de aminoácidos da sequência ACE2 de tipo selvagem e RBD-Spike de tipo selvagem da SARS-CoV-2 foram substituídos pelos das variantes selecionadas por meio do *software* UCFS CHIMERA versão 1.14 (PETTERSEN *et al.*, 2004).

3.4 Dinâmica Molecular (DM)

A dinâmica molecular permite visualizar e avaliar a movimentação atômica das estruturas biológicas a partir da física que rege as interações atômicas, diferentemente das estruturas em modelos tridimensionais e o *docking* molecular, que fornecem apenas uma visão estática das estruturas (HUG, 2013). Este método permite explorar o comportamento da proteína sob certas condições que não são possíveis de estabelecer em laboratório ou que são muito custosas, além de obter uma visão dos mecanismos moleculares dessas proteínas (ISHAK *et al.*, 2017). As trajetórias da simulação podem ser aplicadas para avaliar e comparar proteínas de diversos

sistemas, fornecendo informações relevantes sobre o efeito de mutações no comportamento proteico (DE FREITAS, 2017).

Esse método é baseado nos princípios da mecânica clássica e consiste na simulação do movimento molecular em função do tempo por meio da análise do comportamento dinâmico dos átomos que compõem o sistema, a cada intervalo de tempo (ABRAHAM *et al.*, 2015). A simulação por dinâmica molecular está sendo cada vez mais aplicada para o entendimento de processos como o enovelamento e estabilidade proteica, determinação e refinamento de modelos estruturais, desenho de fármacos, entre outros, devido ao seu baixo custo e rapidez (DE FREITAS, 2017).

A DM utiliza a integração das leis de movimento de Newton para determinar a força resultante e a aceleração de cada átomo, fornecendo uma trajetória que especifica as posições e velocidades das partículas ao longo do tempo (SINGH *et al.*, 2017). Para a produção da trajetória de cada átomo do sistema, é utilizado um Campo de Força, que é um conjunto completo dos potenciais de interação entre as partículas. Dessa forma, a energia potencial total do sistema pode ser calculada a partir da estrutura tridimensional (3D) do sistema. A energia potencial é descrita como a soma de vários termos de energia, incluindo os termos para átomos ligados (comprimentos e ângulos de ligação, ângulos diedros) e os termos para átomos não ligados (interações de van der Waals e de Coulomb) (NAMBA *et al.*, 2008). Assim, o Campo de Força contém um conjunto de parâmetros que representam o comportamento da biomolécula, baseado em propriedades moleculares como comprimento e ângulo de ligações, rotações, torções, entre outras (HUANG *et al.*, 2020). A escolha do campo de força depende do sistema a ser estudado e das propriedades que serão investigadas. Neste trabalho, o campo de força utilizado foi o CHARMM36, que possui parâmetros compatíveis para a análise do sistema proteína-solvente (HUANG *et al.*, 2013).

A simulação da DM é realizada por meio da montagem e preparação do sistema a ser estudado. A estrutura de interesse é inserida em uma caixa cúbica e, posteriormente, solvatada e neutralizada. A energia do sistema é minimizada para que maus contatos entre as moléculas sejam eliminados através de ajustes nas posições atômicas, em um processo de relaxamento das distorções nas ligações químicas, nos ângulos entre ligações e nos contatos de van der Waals. Em seguida, é realizada a etapa de equilíbrio do sistema, onde é estabelecido que o número de partículas, volume e temperatura (NVT) esteja em condição constante, e condições nas quais a pressão também se estabiliza. Após tais etapas, o sistema está apto para a simulação ser realizada.

As simulações de dinâmica molecular dos complexos Spike-ACE2 e Spike-TMPRSS2 foram realizadas por meio do programa computacional GROMACS versão 2020.5 (SPOEL *et al.*, 2005). Onze complexos ACE2-Spike foram analisados: ACE2-Spike (complexo selvagem), ACE-Spike contendo as variantes da Spike (Alfa, Beta, Gama, Delta, Ômicron e P2 – Zeta) e complexos ACE-Spike com os polimorfismos da ACE2 (K26R, G211R, R219C, K341R).

Cinco complexos TMPRSS2-Spike foram analisados: TMPRSS2-Spike selvagem e TMPRSS2-Spike com os polimorfismos da TMPRSS2 (V160M, F209I, P375S e V415I). Os sistemas foram solvatados em uma caixa cúbica com distância mínima de 1,25 nm do soluto até a borda da caixa. O campo de força utilizado foi o CHARMM36 (HUANG, MACKERELL, 2013). As moléculas foram solvatadas com moléculas de água TIP3P (HARRACH; DROSSEL, 2014) e neutralizadas pela adição do número apropriado de íons Na+Cl considerando a concentração iônica de 0,15 M. A minimização de energia foi realizada usando o método de descida mais íngreme com uma força máxima de 1000 Kj/mol.nm. Após a minimização, os sistemas foram equilibrados em dois estágios: um conjunto canônico NVT seguido por um conjunto isotérmico-isobárico NPT. O equilíbrio NVT foi realizado com temperatura constante de 300 K por 500 ps. O equilíbrio NPT foi realizado com pressão constante de 1 bar e temperatura constante de 300 K por 500 ps. Os algoritmos v-rescale e Parrinello-Rahman foram utilizados para manter temperatura e pressão constantes. A etapa de produção foi conduzida a 300 K por 100 ns e as trajetórias foram salvas a cada 10 ps.

3.5 Análises das simulações de Dinâmica Molecular

As trajetórias obtidas a partir da DM foram utilizadas para calcular os valores da raiz do desvio quadrado médio - RMSD (do inglês, *Root Mean Square Deviation*) para analisar a estabilidade das estruturas ao longo do tempo (BRÜSCHWEILER, 2002). O RMSD mede a diferença entre as coordenadas atômicas do estado inicial da estrutura em relação aos estados simulados, sendo utilizado para analisar a estabilidade da proteína ao longo da trajetória da dinâmica molecular. Os valores RMSD são dados em função do tempo, o que permite a observação do período de tempo em que as estruturas se estabilizam (FREITAS, 2017). A flutuação da raiz quadrada média - RMSF (do inglês, *Root Mean Square Fluctuation*) também foi calculada, permitindo avaliar regiões das biomoléculas que são flexíveis (PITERA, 2014). O RMSF representa o grau de variação da posição de um determinado átomo ao longo do tempo, em relação à estrutura de referência. Valores mais altos de RMSF por resíduo caracterizam

maior flexibilidade do mesmo (MARTÍNEZ, 2015). A compactação dos 16 complexos foi analisada pelo cálculo do raio de giração (R_g), permitindo descrever o equilíbrio conformacional ao longo de uma simulação (LOBANOV *et al.*, 2008). Os valores da área acessível ao solvente SASA (do inglês, Solvent Accessible Surface Area) foram gerados para descrever o contato entre a biomolécula e o solvente (EISENHABER *et al.*, 1995). O programa DSSP (do inglês, Dictionary of Secondary Structure of Proteins) (KABSCH; SANDER, 1983) foi utilizado para a predição das estruturas secundárias formada pelos aminoácidos das proteínas, com base nas ligações intramoleculares de hidrogênio (PAULING; COREY, 1951a; PAULING *et al.* 1951b). A energia livre de ligação e a força de interação biomolecular foram estimadas pelo cálculo da área de superfície de Poisson-Boltzman da mecânica molecular MM/PBSA (do inglês, *Molecular Mechanics Poisson–Boltzmann Surface Area*) (KUMARI; KUMARK, 2014). O MM/PBSA é um método para avaliar a energia livre de ligação e a força das interações biomoleculares. Este método foi implementado ao software `g_mmpbsa` do pacote GROMACS, sendo utilizado para analisar a energia livre de ligação nas interações proteína-proteína (total e componentes – van der Waals, eletrostática, solvatação polar e energia de superfície), e também determinar as contribuições dos resíduos para essa energia de ligação (KUMARI, 2014). Para os cálculos de MM/PBSA, foram incluídos os quadros dos 10 ns finais da produção de cada complexo.

Todas as análises foram realizadas pelo serviço da Superintendência de Tecnologia da Informação (STI-USP), por meio do pacote GROMACS versão 2020.5 (SPOEL *et al.*, 2005).

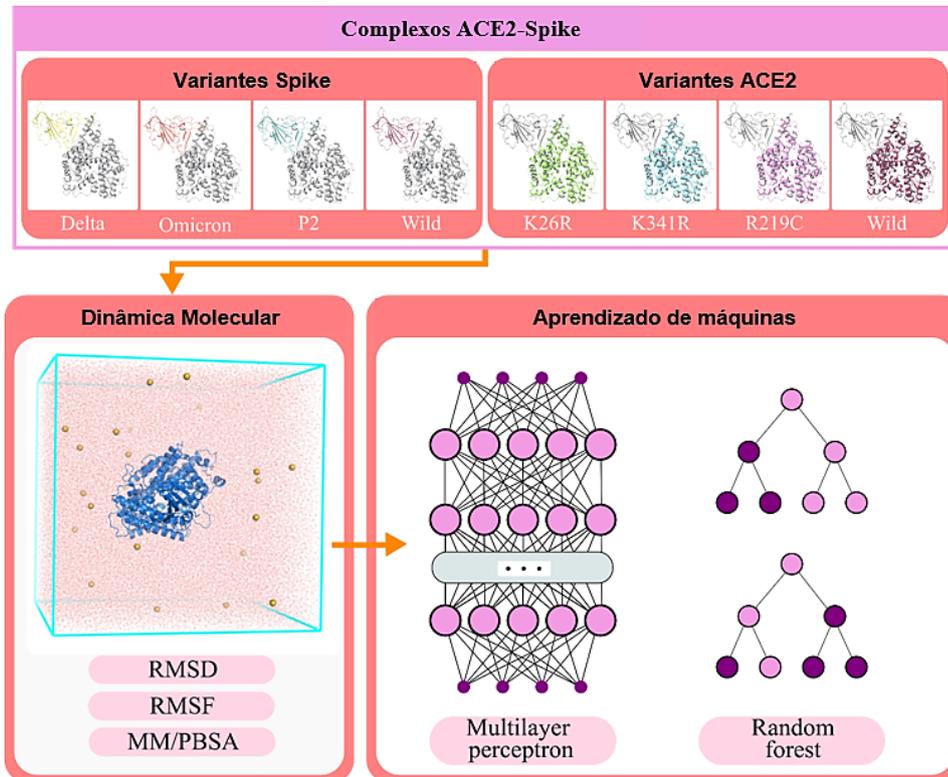
3.6 Identificação de resíduos-chave na interação da proteína ACE2 com a Spike por aprendizagem de máquinas

Com a finalidade de extrair dos dados das trajetórias de dinâmica molecular os resíduos-chave que possam contribuir significativamente, porém de forma sutil, na interação entre as proteínas ACE2 e Spike, os membros do laboratório Ana Carolina Damasceno Sanches e Arthur Scorsolini Fares utilizaram métodos de aprendizado de máquinas para analisar a grande quantidade de dados obtidos nas DM.

Baseado em Fleetwood *et al.*, matrizes de correlação de mapas de contato filtrados de trajetórias MD foram usadas como entradas para técnicas de MLP supervisionadas. Para reduzir a influência de um único modelo e aumentar a estabilidade dos resultados, foram utilizadas duas estratégias de classificação diferentes, *Multilayer Perceptron* (MLP) e *Random Forest* (RF).

Ambos os métodos foram utilizados para identificar resíduos que contribuem para a diferença no comportamento dinâmico entre os complexos (Figura 15).

Figura 15. Fluxograma dos métodos utilizados no aprendizado de máquinas.



Fonte: adaptada do manuscrito submetido (Ávila, A. L. et al, 2023).

O MLP é uma rede neural artificial com múltiplas camadas entre as camadas de entrada e saída, sendo adequado para capturar associações complexas e não lineares nos dados. Por outro lado, RF é um algoritmo de aprendizagem de máquina que constrói múltiplas árvores de decisões e combina seus resultados para chegar a um único resultado (SHICHKIN *et al.*, 2018). O desempenho e capacidade do RF de gerenciar dados ruidosos e incompletos é um recurso valioso em vários cenários (BREIMAN, 2001).

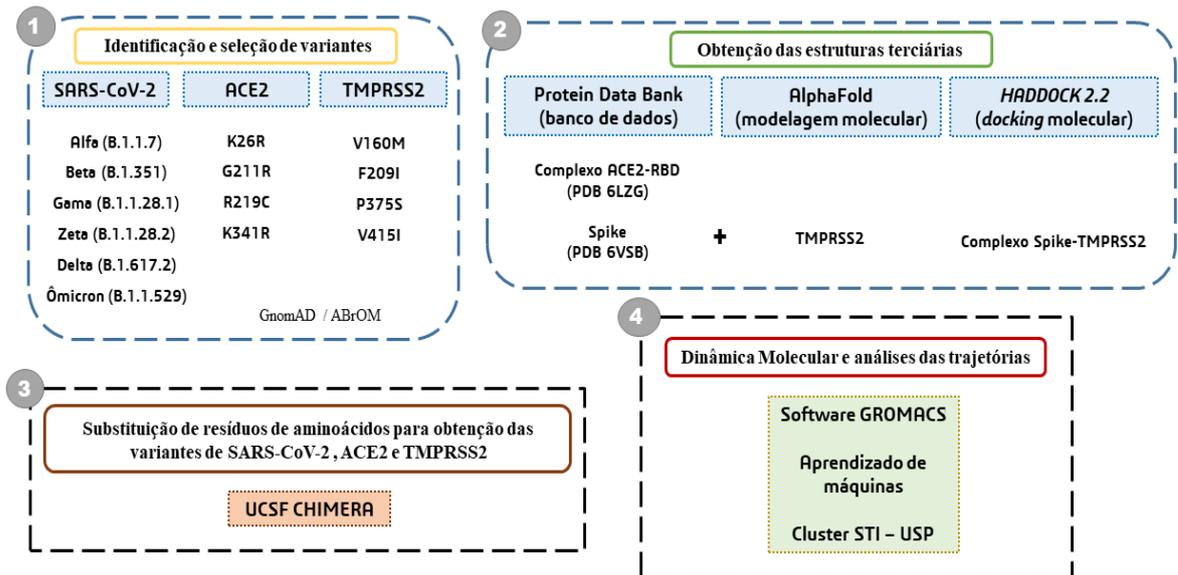
O que constitui os dados de entrada para o MLP e RF são as trajetórias obtidas na dinâmica molecular, a qual passam por uma etapa de pré-processamento antes de serem usadas para treinamento e teste.

Foram utilizadas matrizes de correlação obtidas a partir das distâncias de contato entre os resíduos da ACE2 e Spike-RBD. As distâncias foram calculadas considerando a distância mínima entre os átomos pesados de resíduos na região de interação e depois filtrados, deixando apenas as distâncias menores que 15 Å, a fim de estabelecer uma faixa de análise pré-determinada para as regiões estudadas.

Nesta etapa, foram analisados dez complexos no total: Spike-RBD (*Wild*) e variantes da Spike-RBD (Delta, Ômicron e P2) e ACE2 selvagem (*Wild*) e variantes (K26R, K341R e R219C).

A figura 16 ilustra o fluxo do trabalho contendo as principais etapas da metodologia.

Figura 16. Fluxograma geral do trabalho.



Fonte: da autora (2023).

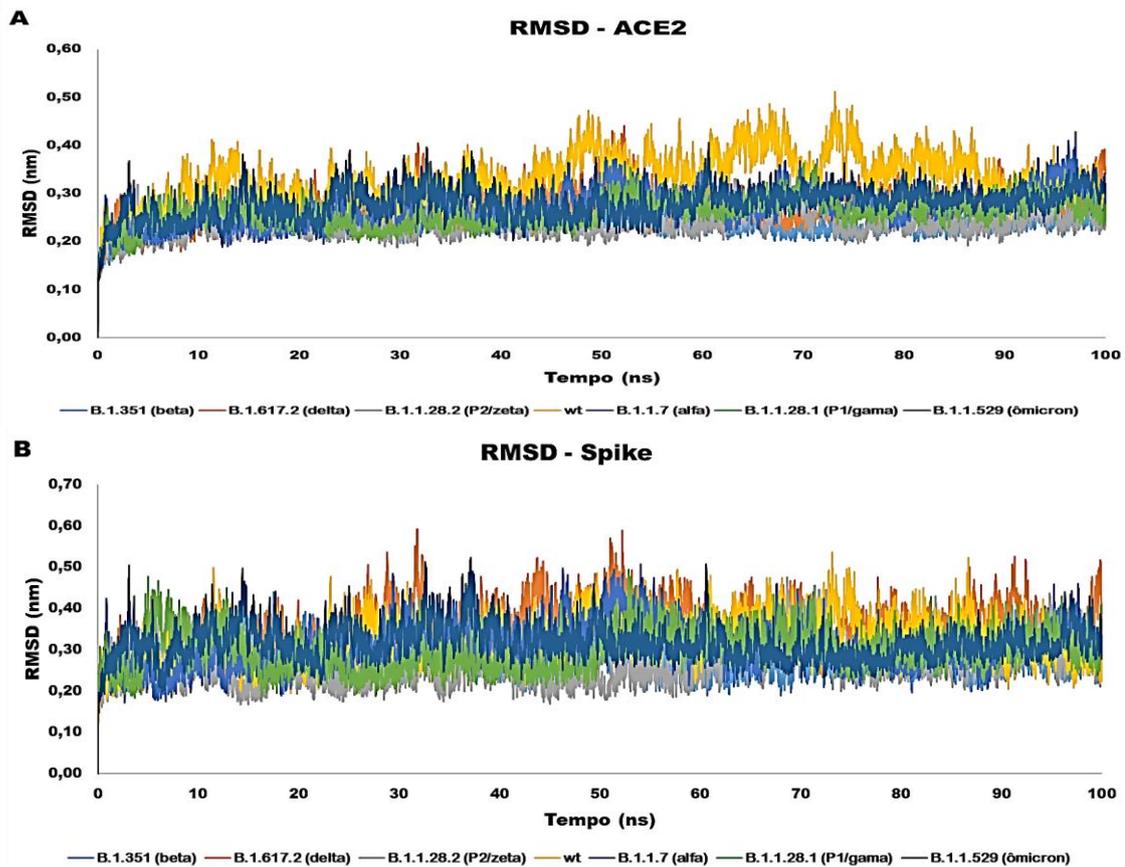
4. RESULTADOS E DISCUSSÃO

4.1 Complexo ACE2 selvagem associada a variantes da Spike

4.1.1 Raiz do Desvio Quadrado Médio (RMSD)

Os gráficos com os valores RMSD dos complexos contendo a ACE2 selvagem e variantes da Spike estão representados nas Figuras 17A e 17B, respectivamente. O gráfico da ACE2 mostra uma estabilização nas trajetórias analisadas para a cadeia ACE2, com valores RMSD similares (entre 0,2 nm e 0,4 nm) e desvio padrão baixo, significando que são estáveis. A trajetória do complexo selvagem da Spike (wt) obteve valor RMSD ligeiramente maior em relação aos demais complexos, revelando maior variação estrutural ao longo do tempo analisado. As trajetórias da proteína Spike de todos os complexos mantiveram-se em equilíbrio, com valores RMSD entre 0,2 e 0,6 nm.

Figura 17. Gráfico RMSD das trajetórias da cadeia ACE2 (A) e Spike-RBD (B).

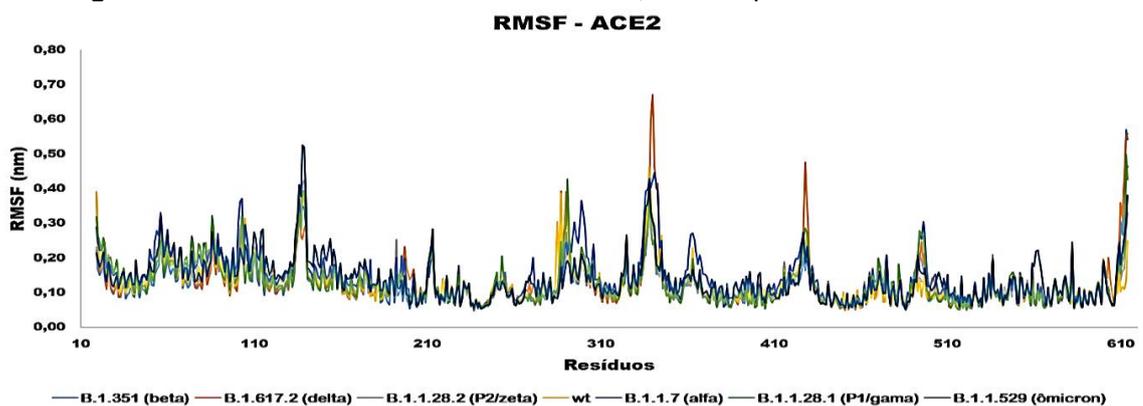


Fonte: da autora (2023).

4.1.2 Flutuação da Raiz Quadrada Média (RMSF)

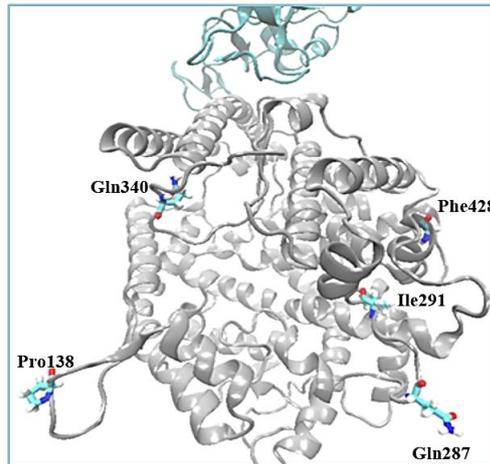
Os valores RMSF obtidos para as trajetórias da ACE2 (Figura 18) não apresentaram regiões de alta flutuação, apenas em regiões de *loop*, como em Pro138, Gln287, Ile291, Gln340 e Phe428 (Figura 19). Gln340 obteve o maior pico RMSF do gráfico na trajetória da variante delta, com uma diferença de aproximadamente 3 Å das demais trajetórias. Além disso, apenas o resíduo Gln340 está localizado em um *loop* próximo ao RBD da Spike.

Figura 18. RMSF dos resíduos da cadeia ACE2, definida pelos resíduos 19 a 614.



Fonte: da autora (2023).

Figura 19. Resíduos com maiores flutuações na ACE2 selvagem estão localizados em regiões de loop.



Fonte: da autora (2023).

No gráfico RMSF referente à Spike (Figura 20), o resíduo Arg408 da variante delta apresentou pico de flutuação ligeiramente menor (0,13 nm), seguido pela variante beta (0,16 nm), wt (0,21 nm), ômicron (0,25 nm), alfa (0,26 nm), gama (0,26 nm) e zeta (0,28 nm). Arg408 está localizado em uma alfa-hélice, próximo a interface de interação Spike-ACE2. A arginina é o aminoácido com menor índice de hidropatia, tendo como consequência uma cadeia lateral muito polar (DAVID; COX, 2013). O resíduo Arg408 está situado próximo ao resíduo 417, o qual sofreu mutação nas variantes beta (K417N), gama (K417T), e ômicron (K417N) (WORLD HEALTH ORGANIZATION, 2023). No tipo selvagem, Lys417 forma uma ponte salina muito estável com o aspartato do resíduo 30 do receptor ACE2. A troca da lisina por asparagina ou treonina interrompe amplamente a ligação nesta posição, pois induz uma perda da ponte salina nesta posição (SOCHER *et al.*, 2021).

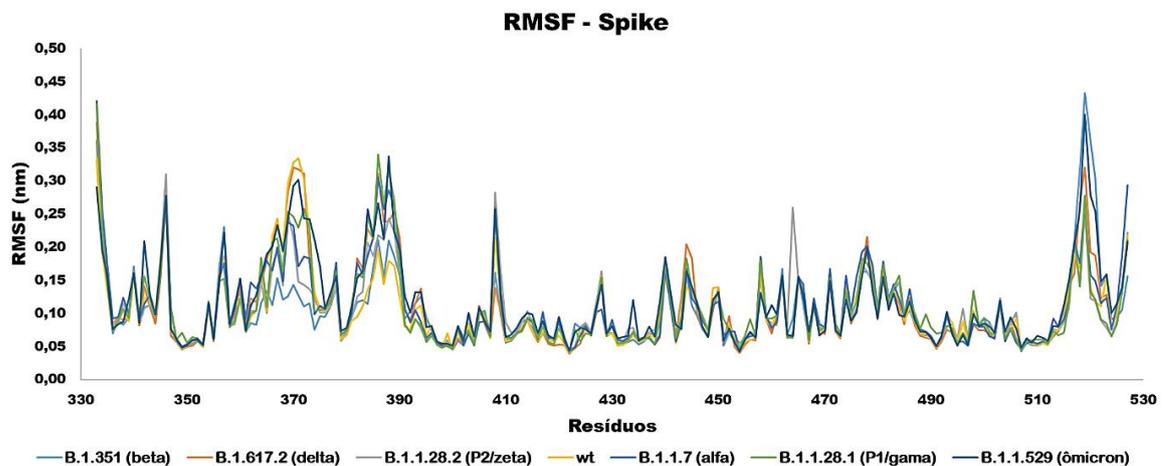
O resíduo Ile434 do complexo da variante ômicron obteve uma flutuação ligeiramente maior (0,12 nm) do que as variantes selvagem/wt (0,6 nm), delta (0,7 nm), beta (0,6 nm), zeta (0,7 nm), alfa (0,7 nm) e gama (0,7 nm). Ile434 está localizado em um folha-beta, relativamente distante da interface Spike-ACE2 e é um aminoácido apolar, importante para a estabilização estrutural das proteínas devido a interações hidrofóbicas em seu interior (STRYER, 1988). Além disso, na variante ômicron, o resíduo Ile434 está próximo das mutações Ser375Phe e Asn440Lys que, segundo Kumar *et al.* (2022), reduzem a estabilidade da proteína, o que pode explicar a maior flutuação observada nesta variante.

Os resíduos com maiores picos de flutuação estão em regiões de *loop* (Arg346, Ser371, Lys386, Asn388, Lys444, Phe464 e His519). Destes, apenas o resíduo Lys444 está localizado próximo da interface de interação com a ACE2. O resíduo Lys444 da variante delta teve o pico

de flutuação mais alto (0,20 nm), seguido pela variante gama (0,18 nm), alfa (0,16 nm), ômicron (0,16 nm), zeta (0,16 nm), beta (0,15 nm) e tipo selvagem (0,14 nm). Lys444 está localizado em um *loop* próximo aos resíduos Gly446, Tyr449, Gln498, Thr500 e Asn501 na Spike selvagem, que possuem interações polares com ACE2, de acordo com estudo de Sironi *et al.* (2020). A variante ômicron possui mutação no resíduo Gly446Ser, com uma glicina substituída por serina e outra mutação no resíduo Gln498Arg, onde a glutamina é substituída por uma arginina. As variantes alfa, beta, gama e ômicron possuem mutação no resíduo Asn501Tyr, havendo uma troca de asparagina por tirosina (WORLD HEALTH ORGANIZATION, 2023). A mutação L452R presente na variante delta, que está relativamente próxima do resíduo Lys444, reduz a estabilidade da proteína (KUMAR *et al.*, 2022).

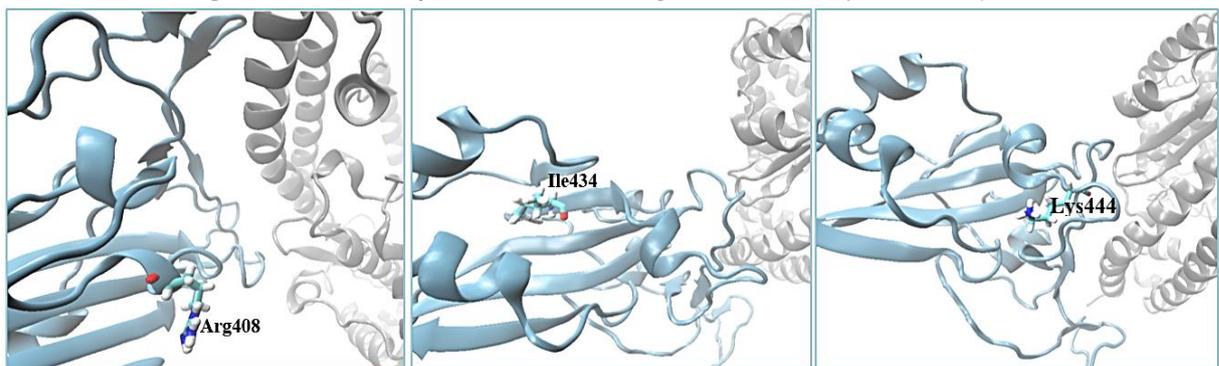
Os resíduos Arg408, Ile434 e Lys444 situados próximos da região de interação com a ACE2 estão ilustrados na Figura 21.

Figura 20. RMSF dos resíduos da cadeia Spike, definida pelos resíduos 333 a 527.



Fonte: da autora (2023).

Figura 21. Localização dos resíduos Arg408, Ile434 e Lys444 na Spike.



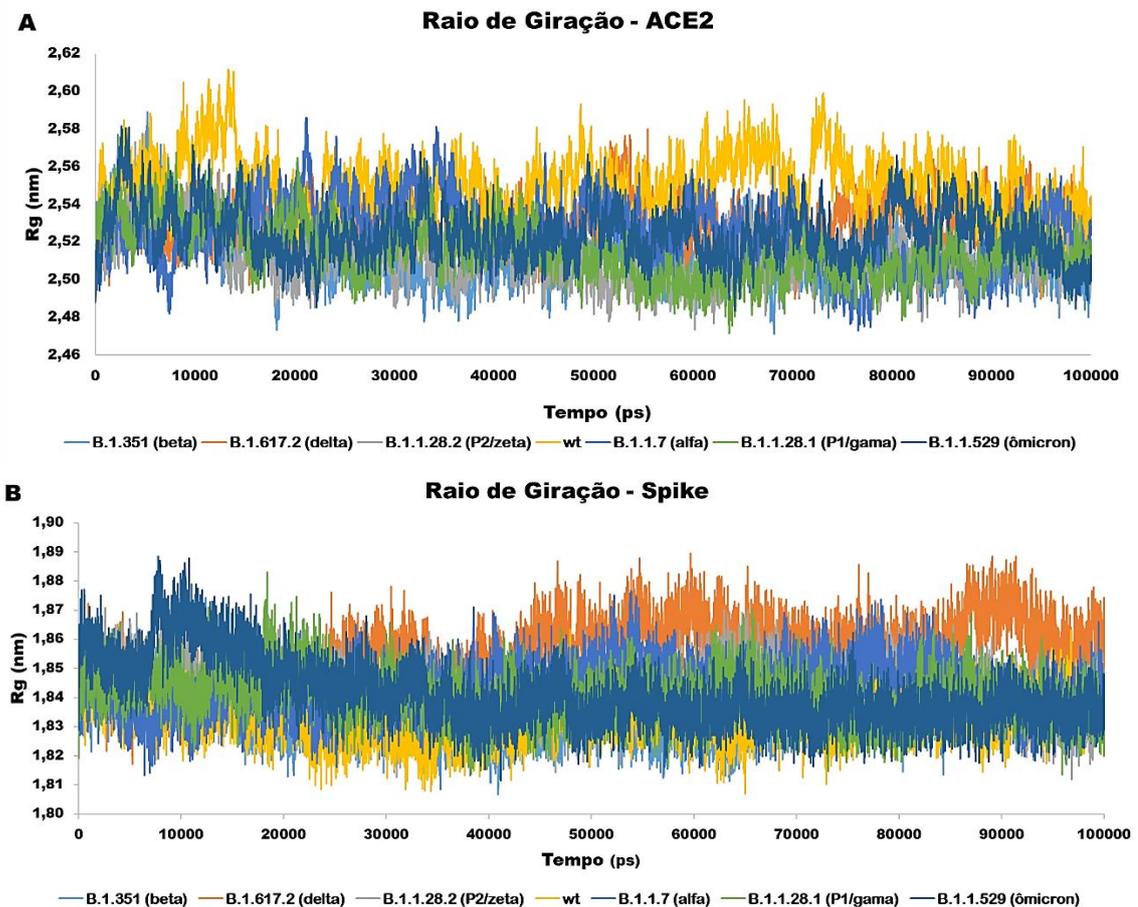
Fonte: da autora (2023).

4.1.3 Raio de Giração (R_g)

O raio de giração é uma medida da compactação de uma proteína. Se uma proteína for dobrada de forma estável, manterá um valor relativamente estável de R_g . Se a proteína está menos compactada, será mais flexível e se moverá mais rápido, obtendo valor de R_g maior. Já uma proteína compactada e com mais estabilidade terá um R_g menor (SNEHA, P.; DOSS, G., 2016).

Os gráficos de R_g dos complexos ACE2-RBD-Spike (Figura 22) mostram valores constantes, com baixo desvio padrão, indicando que as trajetórias permanecem em equilíbrio conformacional durante o período de tempo analisado (100000 ps).

Figura 22. Raio de giração das cadeias ACE2 (A) e Spike (B).



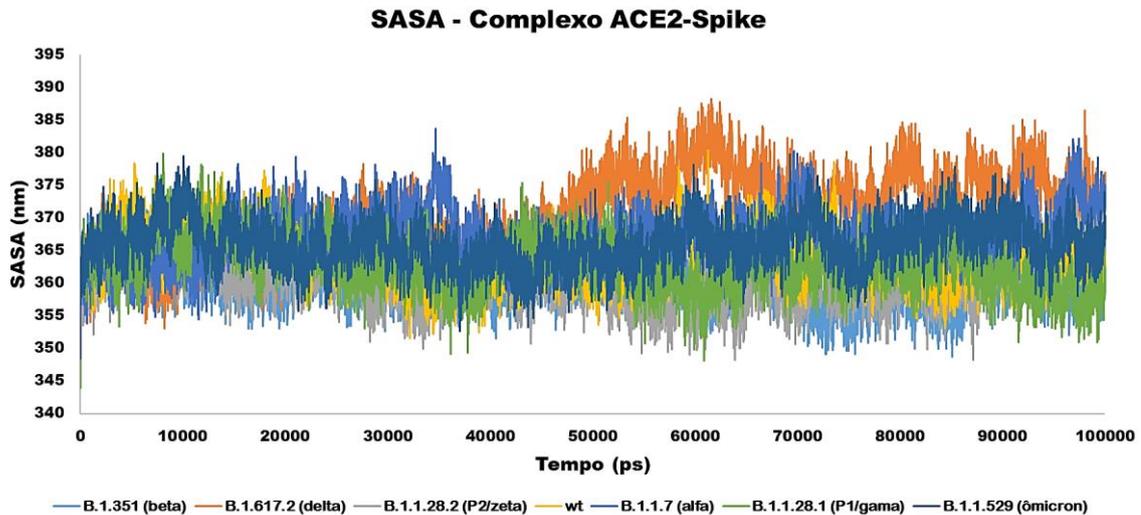
Fonte: da autora.

4.1.4 Área Acessível ao Solvente (SASA)

SASA é a área de superfície de uma biomolécula que é acessível a um solvente. Com base nos valores SASA, os resíduos de aminoácidos de uma proteína podem ser classificados como enterrados ou expostos (ALI *et al.*, 2014). Os complexos Spike-ACE2 mantiveram-se

relativamente estáveis ao longo do tempo analisado (Figura 23). A trajetória do complexo da variante Delta sofre um ligeiro aumento de SASA em relação aos outros complexos por volta de 50000 ps (50 ns), indicando que a proteína sofre uma desestabilização, podendo expor regiões hidrofóbicas ao solvente.

Figura 23. SASA das trajetórias dos complexos ACE2 e variantes Spike-RBD.

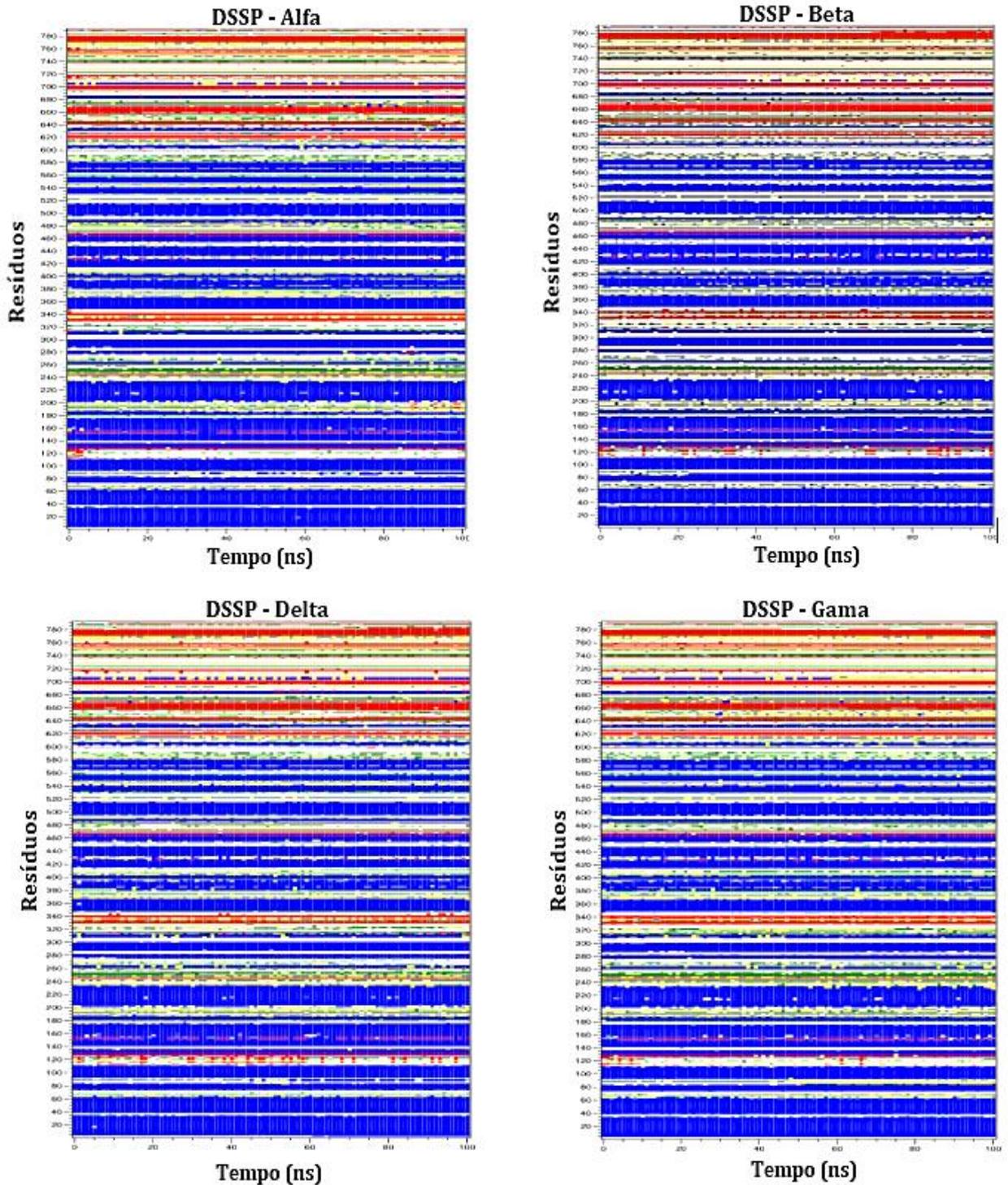


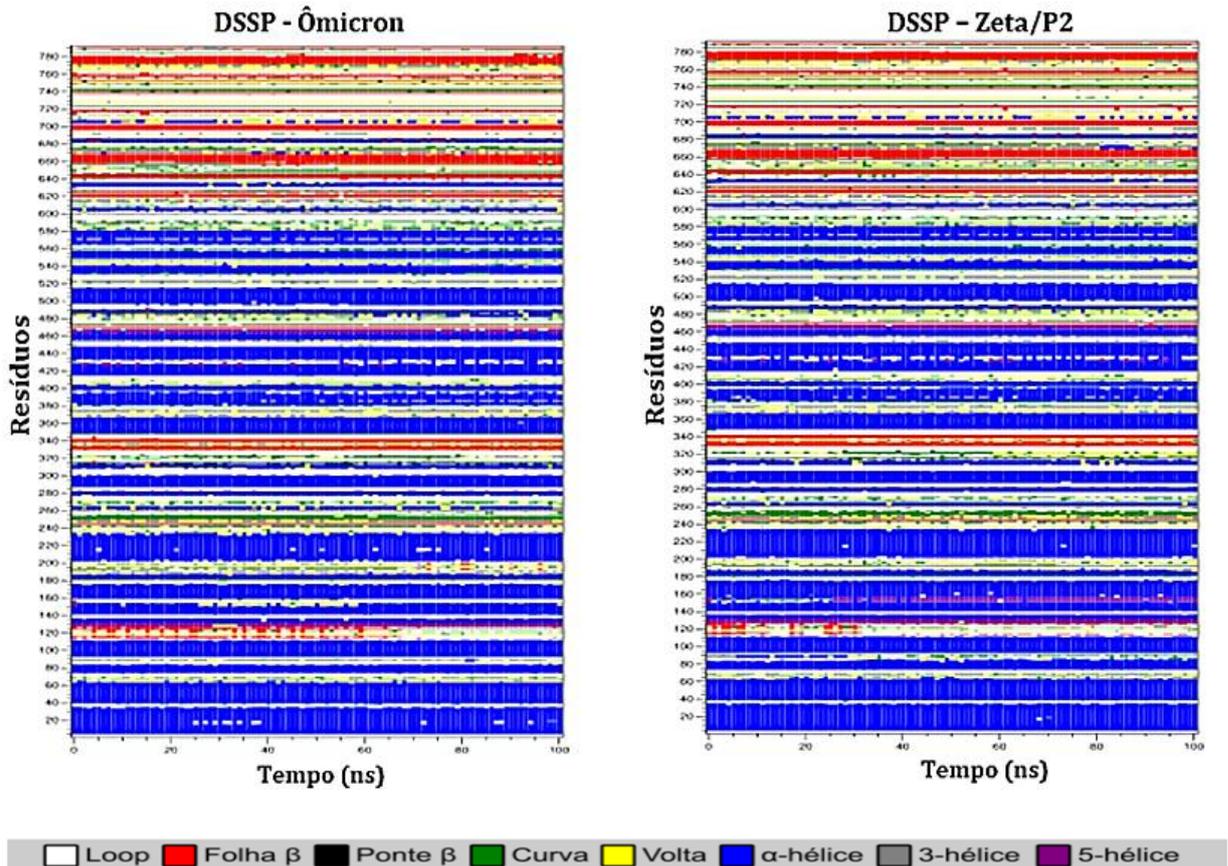
Fonte: da autora (2023).

4.1.5 Dicionário de Estrutura Secundária de Proteínas (DSSP)

Essa análise prediz o conteúdo da estrutura secundária de uma proteína em função do tempo. A partir do gráfico gerado, é possível visualizar a estabilidade (ou instabilidade) dos elementos da estrutura secundária em função do tempo (ANDERSEN *et al.*, 2002). A Figura 24 ilustra as variações estruturais dos resíduos da ACE2 WT e variantes da Spike-RBD em relação à evolução temporal. Não houve mudanças significativas na estrutura secundária dos complexos durante a simulação de dinâmica molecular. A figura mostra que houve ligeiras transições para estruturas desdobradas, como voltas e curvas, e também para 5-hélice e 3-hélice, que são variações da alfa-hélice. Assim, as mudanças de flexibilidade conformacional aparentam não induzir variações secundárias.

Figura 24. Ilustração gráfica do DSSP dos complexos com ACE2 selvagem e Spike-RBD das variantes. As estruturas secundárias estão representadas conforme as cores indicadas na legenda da imagem.





Fonte: da autora (2023).

4.1.6 Área de Superfície de Poisson-Boltzmann da Mecânica Molecular (MM/PBSA)

O cálculo MM/PBSA foi realizado para estimar as energias de ligação entre as proteínas ACE2 e Spike, bem como para compreender os fatores que contribuem para a estabilidade ou instabilidade da interação. Os valores de energia de ligação estão resumidos na Tabela 1, revelando que a variante Ômicron demonstra uma afinidade mais forte com a ACE2, apresentando uma energia de ligação de $-2572,23 \pm 144,92$ KJ.mol⁻¹, seguida pelas variantes Delta e P2, que registram valores de $-1.875,71 \pm 132,97$ e $-1.837,21 \pm 119,867$ kJ.mol⁻¹, respectivamente. A análise da decomposição da energia residual do MM/PBSA revelou que são as interações eletrostáticas que influenciam predominantemente a estabilidade das variantes da proteína Spike na interação proteína-proteína. Isto contrasta com os valores obtidos para os componentes de energia de Van der Waals, energia polar de solvatação e SASA, que não apresentaram variações significativas.

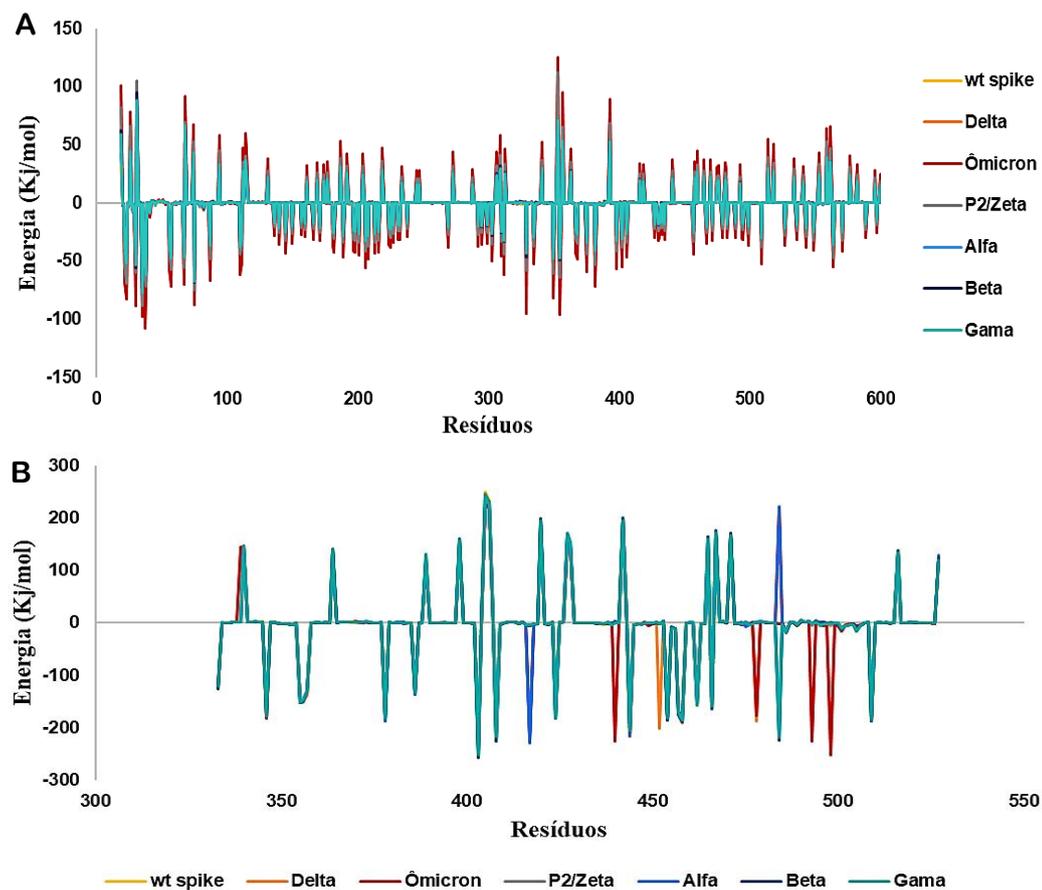
Tabela 1. Decomposição residual e energia de ligação da interação proteína-proteína (dado em kJ.mol⁻¹).

ACE2	SPIKE	ΔE_{vdW}	ΔE_{elet}	Energia Polar de Solvatação	SASA	Energia de Ligação
Wild	Wild	-321.597 ± 22.537	-1429.286 ± 59.435	785.306 ± 107.762	-40.912 ± 3.417	-1006.488 ± 115.085
Wild	Delta	-315.863 ± 19.370	-2046.183 ± 73.329	527.732 ± 131.845	-41.399 ± 3.557	-1875.713 ± 132.967
Wild	Ômicron	-328.232 ± 20.667	-2866.463 ± 96.486	664.301 ± 135.753	-41.834 ± 3.542	-2572.228 ± 144.922
Wild	P2/Zeta	-322.528 ± 24.992	-2190.347 ± 71.591	719.820 ± 115.046	-44.160 ± 3.584	-1837.214 ± 119.867
Wild	Alfa	-331.948 ± 16.908	-1202.110 ± 63.963	568.956 ± 167.500	-40.688 ± 3.737	-1005.789 ± 167.256
Wild	Beta	-326.388 ± 18.791	-1606.571 ± 70.629	484.462 ± 110.407	-40.675 ± 3.364	-1489.173 ± 109.571
Wild	Gama	-324.219 ± 18.860	-1566.002 ± 50.006	418.446 ± 138.536	-38.549 ± 3.452	-1510.324 ± 133.772

Fonte: da autora (2023).

A Figura 25 ilustra a decomposição de energia por resíduo através do MM/PBSA dos complexos analisados. Embora os resíduos da ACE2 mostrem variações sutis na interação com as variantes da Spike (Figura 25A), as variações dos resíduos na proteína Spike foram mais pronunciadas (Figura 25B). Especificamente, os resíduos nas posições 408, 417, 440, 452, 478, 484, 493 e 498 exibiram variações energéticas significativas.

Figura 25. Decomposição de energia por resíduo MM/PBSA para as trajetórias obtidas nas simulações de DM. (A) Energia de contribuição dos resíduos ACE2 WT em interação com variantes Spike-RBD; (B) Energia de contribuição dos resíduos das variantes Spike-RBD em interação com a ACE2 WT.



Fonte: da autora (2023).

Na cadeia Spike-RBD (Figura 25B), a variante Ômicron mostra uma energia de ligação ligeiramente menor no resíduo Arg408 em comparação com a wt e as outras variantes. Além disso, o resíduo Lys417 apresenta uma energia de ligação mais fraca nas variantes Gama (-3,24 KJ.mol⁻¹), Beta (-5 KJ.mol⁻¹) e Ômicron (-4,72 KJ.mol⁻¹) do que nos outros complexos, com uma discrepância superior a 200 kJ.mol⁻¹. Como mencionado anteriormente, a substituição de lisina por asparagina (variante Ômicron e Beta) ou treonina (variante Gama), rompe uma interação de ponte salina altamente estável entre Lys417 e o resíduo 30 da ACE2, consequentemente desestabilizando o complexo. Além disso, a mutação N440K presente na variante Ômicron registra um valor de -225,52 kJ.mol⁻¹, diferindo em mais de 200 kJ.mol⁻¹ dos demais complexos.

No complexo da variante Delta, a mutação L452R apresentou energia de ligação de -200,19 KJ.mol⁻¹, enquanto nos demais complexos o resíduo atingiu valores em torno de -1 KJ.mol⁻¹. O resíduo L452R está localizado na região hidrofóbica da proteína Spike e não interage com o receptor ACE2. No entanto, existe a possibilidade de que a mutação induza alterações estruturais que promovem sua interação com o receptor ACE2 (DENG *et al.*, 2021).

A mutação T478K, identificada nas variantes Delta e Ômicron, resultou em energias de ligação de -186,69 e -175,78 kJ.mol⁻¹, respectivamente. Em contrapartida, os complexos WT, P2/Zeta, Alfa, Beta e Gama, apresentaram energias de ligação em torno de -1 kJ.mol⁻¹. A substituição do resíduo polar e não carregado (treonina) pelo aminoácido básico e carregado positivamente (lisina) aumenta a contribuição do potencial eletrostático da Spike-RBD, promovendo uma afinidade mais forte com o receptor ACE2. Além disso, a cadeia lateral alongada da lisina poderia aumentar os efeitos estéricos da variante Delta, potencialmente elucidando o aumento da interação entre Spike-RBD e o receptor ACE2 (DI GIACOMO *et al.*, 2021).

O resíduo Glu484 sofre mutação na variante Ômicron, envolvendo a substituição da glutamina pela alanina, e da mesma forma nas variantes P2/Zeta, Gama e Beta, mas com a substituição da glutamina pela lisina. Notavelmente, a mutação E484K nas variantes P2/Zeta, Gama e Beta, produziu uma energia de ligação mais elevada (-211,87 KJ.mol⁻¹), contrastando com a mutação E484A observada na variante Ômicron (-1,76 KJ.mol⁻¹). A mutação E484A na variante Ômicron elimina a ligação fraca de Glu484 com a ACE2, ao mesmo tempo que atenua a desestabilização decorrente da repulsão eletrostática entre Glu484 de WT e Glu35 da ACE2 após a alteração para Ala484. Consequentemente, a mutação E484A não tem influência na energia livre de ligação, em contraste com a mutação E484K, que aumenta a interação (JAWAD

et al., 2022). Já a mutação do resíduo Glu484 carregado negativamente para o resíduo carregado positivamente Lys484 contribui para a energia livre de ligação negativa e, conseqüentemente, para a estabilização do complexo (ISTIFLI *et al.*, 2021).

Q493R e Q498R são mutações presentes na variante Ômicron que apresentaram aumento na energia de ligação com valores de -226 e -252 $\text{kJ}\cdot\text{mol}^{-1}$, respectivamente, com diferença de mais de 200 $\text{kJ}\cdot\text{mol}^{-1}$ das outras variantes. A combinação das mutações Q498R e N501Y aumenta significativamente a capacidade de ligação da ACE2 devido à formação de duas novas pontes salinas fortes entre Arg493 e Arg498 da Ômicron, e Glu35 e Asp38 na ACE2 (JAWAD *et al.*, 2022).

4.1.7 Identificação de resíduos-chave na Spike selvagem e variantes por aprendizagem de máquinas

A Tabela 2 mostra os cinco pares de resíduos mais significativos para cada complexo obtidos por MLP e RF. Resíduos-chave responsáveis por variações na ligação entre as variantes Spike-RBD e ACE2 WT foram identificados. Além dos pares de resíduos relatados pela literatura (destacados em negrito), os métodos foram capazes de identificar novos resíduos-chave importantes na interação entre as estruturas que formam os complexos estudados.

Tabela 2. Resíduos importantes nas variantes da Spike interagindo com ACE2 WT obtidos por MLP e RF.

MLP				RF			
Variante da Spike	ACE2	Spike	Importância	Variante da Spike	ACE2	Spike	Importância
Selvagem (Wild)	SER106	GLY485	1.00	Selvagem (Wild)	SER19	VAL483	1.00
	VAL107	PHE486	0.99		SER19	CYS488	0.95
	GLN89	SER477	0.98		SER19	CYS480	0.60
	SER19	PRO479	0.89		SER44	TYR505	0.52
	ALA71	GLU484	0.82		SER19	GLN474	0.52
Delta	ASP30	GLU484	1.00	Delta	ALA36	ASN501	1.00
	GLN24	LYS417	0.71		GLY66	ASN501	0.98
	GLY352	ARG408	0.68		ALA342	THR500	0.93
	ALA65	SER443	0.62		ASN103	TYR505	0.71
	ASN33	GLN498	0.60		LYS68	ASN501	0.64
Ômicron	GLU329	SER438	1.00	Ômicron	ALA25	ASN417	1.00
	GLN42	SER349	0.96		GLN24	ASN417	0.98
	TYR381	GLY502	0.92		ILE21	ASN417	0.97
	GLY352	ASN448	0.91		LYS353	ARG498	0.94
	GLY354	GLY504	0.90		THR27	ASN417	0.88
P2	PRO321	ARG403	1.00	P2	SER106	LYS484	1.00
	SER19	ASN477	0.92		SER19	CYS480	0.90
	SER19	PRO479	0.87		SER105	ASN487	0.87
	GLN325	SER371	0.84		GLY104	ASN487	0.80
	GLU37	THR415	0.81		SER105	LYS484	0.79

Fonte: da autora (2023).

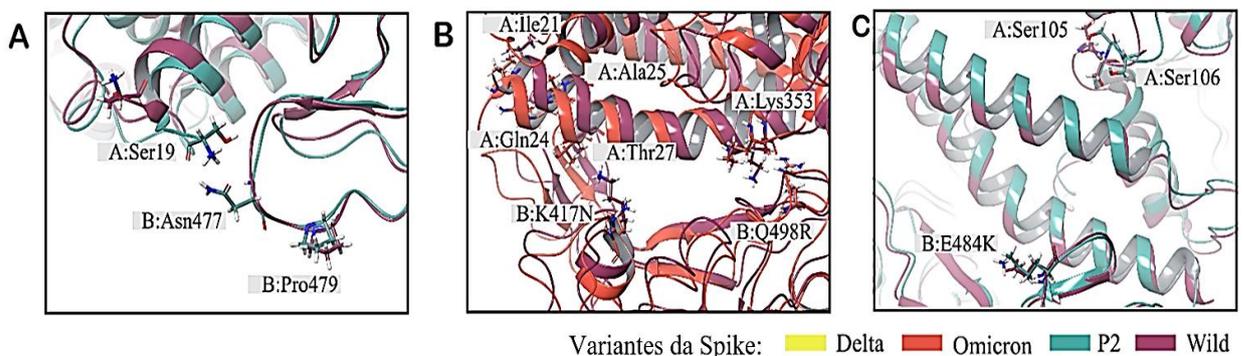
O modelo MPL identificou a interação entre Asp30 da ACE2 e Glu484 da Spike na variante Delta. Asp30 está localizado em uma alfa-hélice, próximo do resíduo Glu35, o qual interage com Glu484, promovendo uma repulsão eletrostática entre ambos os resíduos, como mencionado anteriormente. A interação entre os resíduos Gln24 da ACE2 e Lys417 da Spike identificada na variante Delta também é relevante. Como já citado no presente trabalho, a substituição de lisina por asparagina na mutação Lys417Asn da variante Delta rompe uma interação de ponte salina altamente estável entre Lys417 e o resíduo 30 da ACE2, o qual está próximo de Gln24, localizados na alfa-hélice.

Outros novos resíduos-chave relevantes foram identificados na variante Ômicron pelo modelo MLP. A interação com maior pontuação ocorreu entre o resíduo Glu329 da ACE2 e Ser438 da Spike. O resíduo Ser438 está próximo da mutação N440K da variante Ômicron, onde a troca da asparagina (polar não carregado) para lisina (polar carregada positivamente) pode causar uma mudança estrutural devido à remoção do grupo carbonila e da adição de grupos amina, diminuindo o valor da energia livre de ligação e favorecendo sua ligação ao receptor ACE2 (ALAM *et al.*, 2023). O receptor ACE2 é globalmente carregado negativamente, portanto, a substituição de N440 por lisina melhora as interações eletrostáticas atrativas, que favorecem a ligação do RBD da variante Ômicron ao receptor ACE2 (WANG *et al.*, 2023). Na interação Tyr381 da ACE2 com Gly502 da Spike, o resíduo Gly502 está localizado ao lado da mutação N501Y, que aumenta as interações eletrostáticas entre a proteína Spike e a ACE2 por meio de uma forte ligação de hidrogênio perto do local da mutação. O aminoácido asparagina, quando substituído por tirosina no resíduo 501, cria uma ligação de hidrogênio entre a cadeia lateral e K353, consequentemente aumentando as interações intermoleculares entre o RBD e o ACE2 (VERMA; SUBBARAO, 2021). Outro par de resíduos relevante é a interação entre Gly352 da ACE2 e Asn448 da Spike. O resíduo Asn448 está próximo da mutação G446S, a qual ocorre a troca do resíduo hidrofóbico glicina para o resíduo polar serina, aumentando a interação com o receptor ACE2 (ZHANG *et al.*, 2022). Os resíduos 446 e 448 estão localizados em um *loop*, na interface de ligação da Spike. O par de resíduos-chave Gly354 da ACE2 e Gly504 da Spike também é interessante. Gly354 da ACE2 está localizado em um *loop* ao lado de Lys353 e Gly504 está localizado em uma alfa-hélice da Spike, próximo das mutações N501Y e Y505H. Segundo Verma e Subbarao (2021), o resíduo Lys353 é um hotspot da ACE2 importante na interação com os resíduos 505 e 501 da Spike. A mutação de N501Y leva a um aumento do número de contatos no hotspot-353. O aminoácido N, quando substituído por Y em 501, cria uma ligação de hidrogênio entre cadeia lateral e K353. Consequentemente, aumentando as interações intermoleculares entre o RBD e o ACE2.

A Figura 26 ilustra alguns dos resíduos importantes que distinguem as variantes do WT, os quais estão destacados em negrito na Tabela 2 (Figuras 26A-C).

Os pares de resíduos identificados pelo modelo RF diferiram daqueles identificados pelo modelo MLP. No entanto, alguns dos resíduos foram identificados por ambos os métodos. As análises dos resultados enfatizam a importância dos principais resíduos de ACE2 na interação com Spike-RBD, conforme mencionado por Ali e Vijayan (2020), que incluem Gln24, Thr27, Asp30, Glu37, Gln42 e Lys353. Além disso, o resíduo Ser19 da proteína ACE2, que foi comumente visto entre os pares, também é importante. Ser19 participa de uma rede de interações de hidrogênio, particularmente interagindo com Pro462 (DE ANDRADE; GONÇALVES; NETZ, 2021).

Figura 26. Pares de resíduos determinados como mais importantes para distinguir a ligação entre as variantes ACE2 WT e Spike-RBD. (A) Pares de resíduos importantes encontrados na Spike-RBD WT e P2; (B) Pares de resíduos importantes encontrados na Ômicron; (C) Par Pares de resíduos importantes encontrados na P2. Resíduos importantes são compostos de pares altamente distintos de distâncias entre a ACE2 WT e as variantes Spike-RBD estudadas.



Fonte: adaptada do manuscrito submetido (Ávila, A. L. et al, 2023).

Os resultados obtidos pelo MLP sugerem que Ser19 pode interagir sutilmente com resíduos próximos a Pro462, formando pares com Pro479 na Spike-RBD WT e Asn477 e Pro479 na variante P2. A mutação S477N, presente na variante P2, pode favorecer uma maior interação com Ser19 da proteína ACE2. Além disso, a mutação S19P aumentou a interação entre ACE2 e Spike-RBD (SURYAMOHAN *et al.*, 2021). Isso sugere que o resíduo Ser19 desempenha um papel crítico na modulação da interação entre a ACE2 e a Spike, potencialmente influenciando a capacidade de infecção e a transmissibilidade do vírus na variante P2. Os pares mencionados podem ser importantes para distinguir a interação entre a Spike-RBD WT e P2 com a ACE2 (Fig. 26A).

No modelo RF, as mutações K417N e Q498R da variante Ômicron formaram pares de resíduos interessantes. A mutação K417N pode ter efeitos significativos nos resíduos Ile21, Gln24, Ala25 e The27, enquanto a mutação Q498R pode afetar Lys353 da proteína ACE2

(Figura 26B). Estudos indicam que a mutação K417N leva a uma redução na afinidade de ligação da Spike com a ACE2; no entanto, é uma mutação de escape imunológico, a qual ajuda o SARS-CoV-2 a escapar das defesas imunológicas naturais do hospedeiro, contribuindo para aumento da transmissibilidade viral. Por outro lado, Q498R é associada a um aumento na infecção viral. O par Lys353-Arg498 teve uma importância de 0,94. O estudo de Zhang *et al.* (2022) sugere que Q498R é estruturalmente incompatível com Lys353 na ACE2, mas está estruturalmente adaptado para Asp38 (ZHANG et al., 2022).

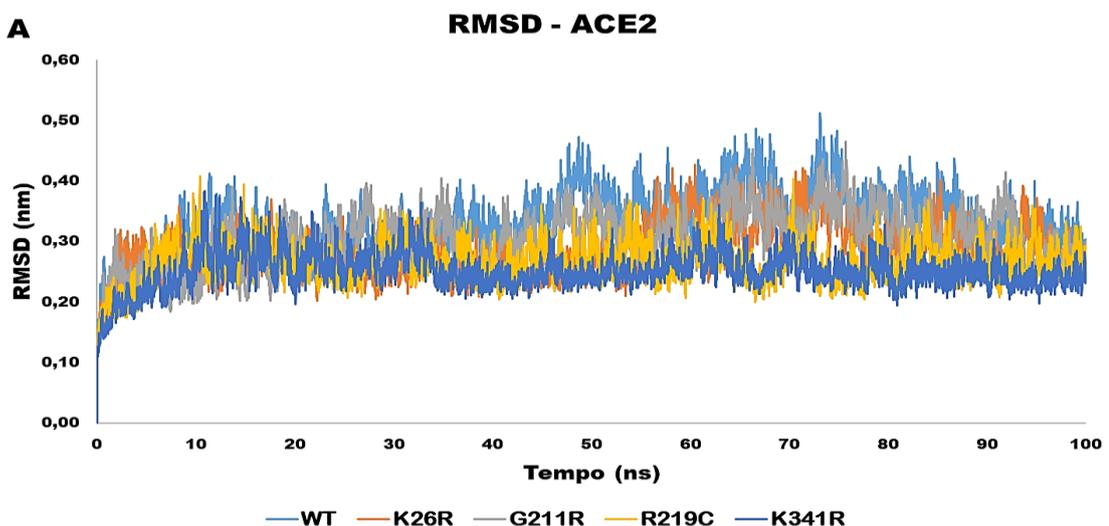
A mutação E484K, encontrada na variante P2, é associada a uma redução nos anticorpos neutralizantes. Essa mutação resulta em uma interface de ligação mais estreita entre a proteína Spike RBD e ACE2, contribuindo para um aumento na afinidade de ligação (WANG *et al.*, 2021). Os resultados indicam a presença de dois pares relacionados a este resíduo, nomeadamente Ser106-Lys484 com uma significância de 1,00 e Ser105-Lys484, com significância de 0,79 (Figura 26C).

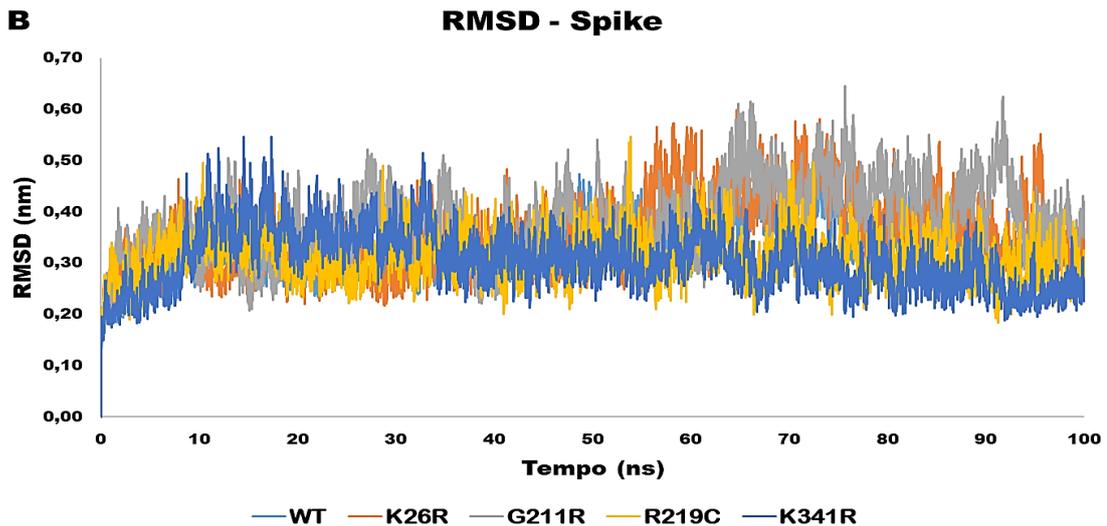
4.2 Complexo Spike selvagem associado a variantes da ACE2

4.2.1 Raiz do Desvio Quadrado Médio (RMSD)

Os gráficos com os valores RMSD dos complexos da Spike selvagem e variantes da ACE2 estão representados nas Figuras 27A e 27B, respectivamente. Todas as trajetórias analisadas dos complexos permaneceram estáveis durante o período de simulação. Todos os complexos obtiveram valores semelhantes, entre 0,2 nm a 0,4 nm, mantendo-se em equilíbrio.

Figura 27. Gráfico RMSD das trajetórias das proteínas ACE2 (A) e Spike (B).





Fonte: da autora (2023).

4.2.2 Flutuação da Raiz Quadrada Média (RMSF)

O gráfico RMSF com os valores das trajetórias dos resíduos da cadeia ACE2 está representado na Figura 28.

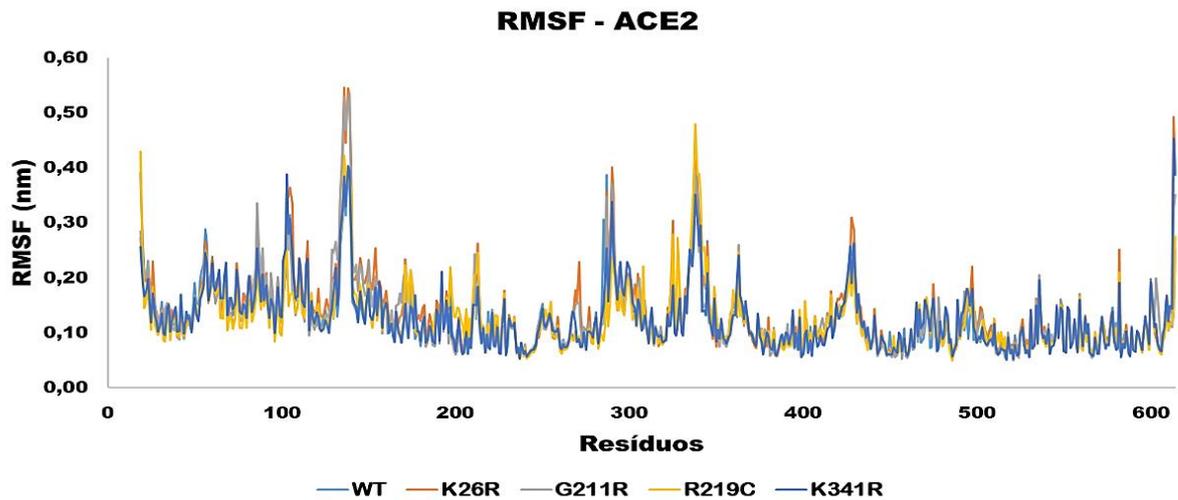
O resíduo Gln325 localizado em uma alfa-hélice da ACE2, próximo ao RBD da Spike, obteve flutuação ligeiramente menor na trajetória do complexo da variante K341R (0,18 nm), seguida por G211R (0,24 nm), WT (0,26 nm), R219C (0,27 nm) e a maior, K26R (0,30 nm). Da mesma forma, o resíduo Trp328 também obteve menor flutuação na trajetória do complexo K341R, com uma diferença de 1,6 nm comparado aos outros complexos. Os resíduos Triptofano e Glutamina estão localizados em uma alfa-hélice, próximo ao RBD da Spike. Ambos Gln325 e Trp328 estão situados próximo da variante K341R.

O resíduo Arg582 do complexo selvagem obteve valor RMSF ligeiramente inferior (0,07 nm) em comparação aos complexos com as variantes K26R (0,25 nm), G211R (0,19 nm), R219C (0,20 nm) e K341R (0,19 nm). O resíduo Arg582 está localizado em uma alfa-hélice próximo das variantes G211R e R219C.

Os maiores picos estão localizados em loops, nos resíduos Asp136, Gln139, Gln287, Asn290 e Asn338. Destes, apenas o Asn338 está relativamente próximo do RBD da Spike. A asparagina 338 é um aminoácido polar e está próximo do polimorfismo K341R. A alteração de um aminoácido menor (lisina) por um maior (arginina) pode causar protuberâncias que aumentam a instabilidade da região.

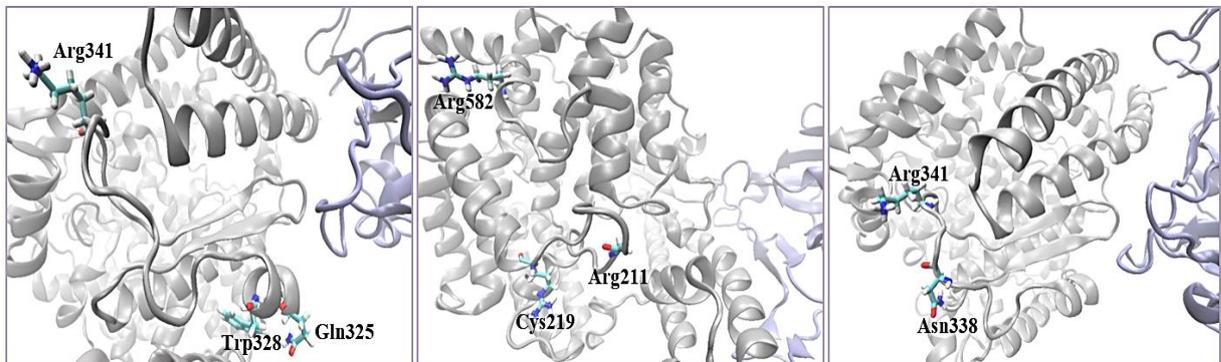
A localização dos resíduos Arg341, Trp328, Gln325, Arg582, Arg211, Cys219, Asn338 na ACE2 estão ilustrados na Figura 29.

Figura 28. RMSF dos resíduos ACE2-Spike. A cadeia ACE2 é definida pelos resíduos 19 a 614.



Fonte: da autora (2023).

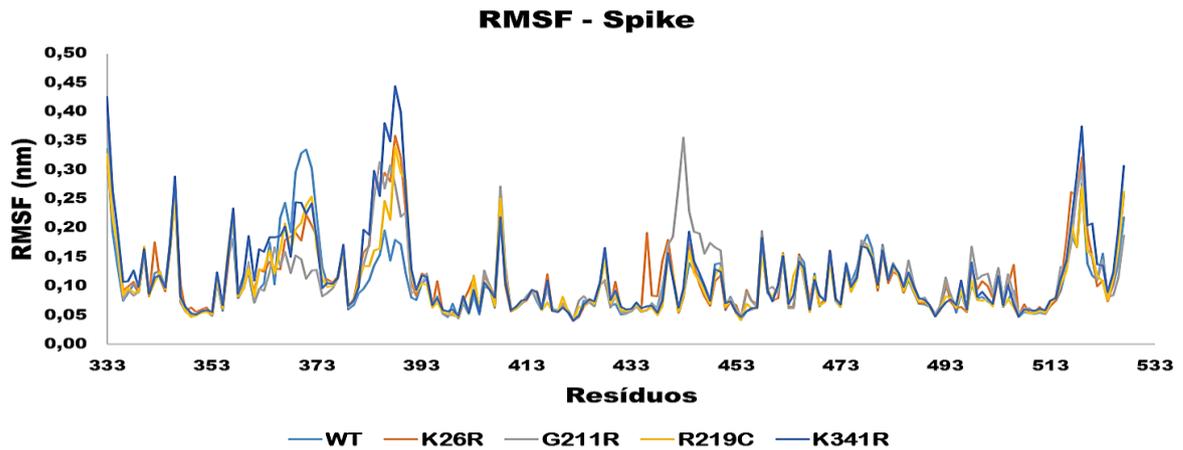
Figura 29. Localização dos resíduos Arg341, Trp328, Gln325, Arg582, Arg211, Cys219, Asn338 na ACE2.



Fonte: da autora (2023).

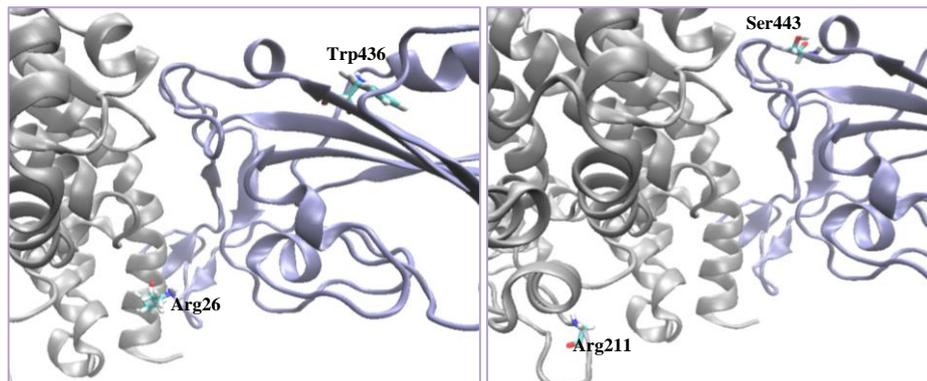
No gráfico da cadeia Spike (Figura 30), o complexo ACE2-RBD da variante K26R da ACE2 apresentou valor RMSF no resíduo Trp436 da proteína Spike ligeiramente maior (0,1 nm de diferença) em comparação com o complexo das demais variantes. A variante Arg26 é um resíduo polar, localizado em uma alfa-hélice, próximo ao RBD da proteína Spike, enquanto o resíduo Trp436 é um aminoácido aromático apolar, e está localizado em uma folha-beta do RBD-Spike. Além disso, o complexo com a variante G211R obteve valor RMSF relativamente maior no resíduo Ser443, apresentando uma diferença de 0,27 nm em relação às outras variantes. Ser443 está situado em uma alfa-hélice próximo ao RBD da Spike, e é um resíduo polar não carregado. A Figura 31 ilustra a localização dos resíduos Arg26, Arg211, Ser436 e Ser443.

Figura 30. RMSF dos resíduos ACE2-Spike. A cadeia Spike é definida pelos resíduos 333 a 527.



Fonte: da autora (2023).

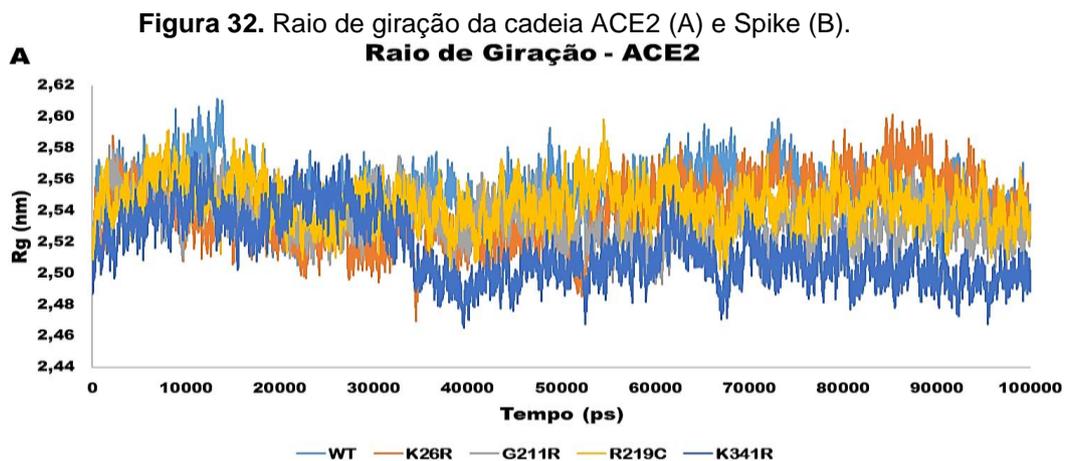
Figura 31. Localização dos resíduos Arg26 e Arg211 na ACE2 e dos resíduos Trp436 e Ser443 na Spike.

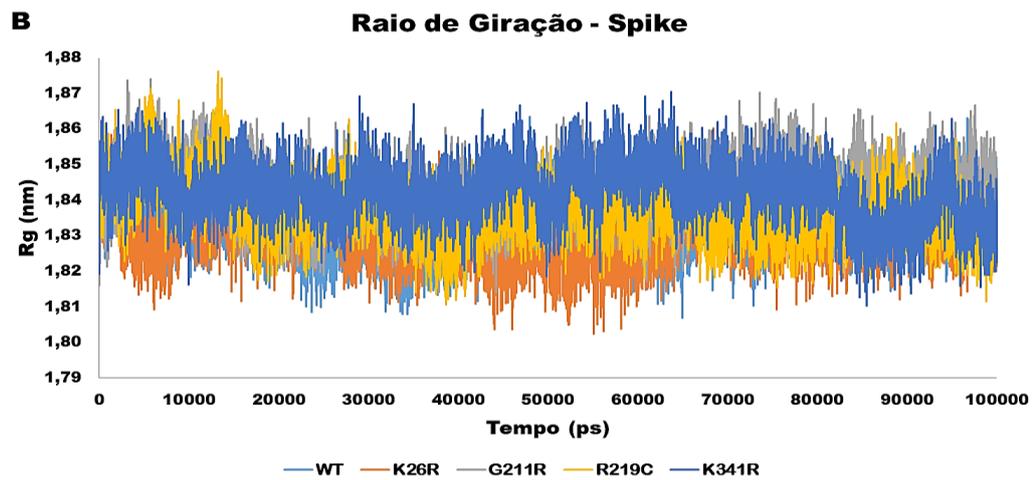


Fonte: da autora (2023).

4.2.3 Raio de Giração (R_g)

Os gráficos do Raio de Giração dos complexos ACE2-RBD-Spike contendo a cadeia ACE2 selvagem e variantes associadas à cadeia Spike selvagem (Figura 32) mostram que as trajetórias permaneceram em equilíbrio conformacional durante o período de tempo analisado, com valores constantes e baixo desvio padrão.



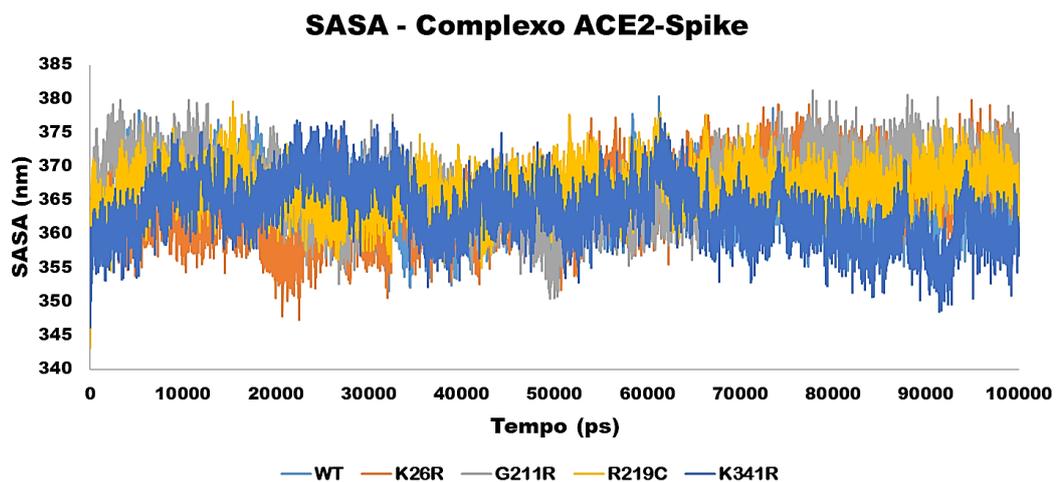


Fonte: da autora (2023).

4.2.4 Área Acessível ao Solvente (SASA)

Os complexos Spike-ACE2 mantiveram-se estáveis ao longo do tempo analisado (Figura 33). Todas as trajetórias obtiveram comportamento semelhante durante os 100000 ps da simulação, não apresentando diferenças conformacionais.

Figura 33. SASA das trajetórias dos complexos com as variantes da ACE2 e Spike selvagem.

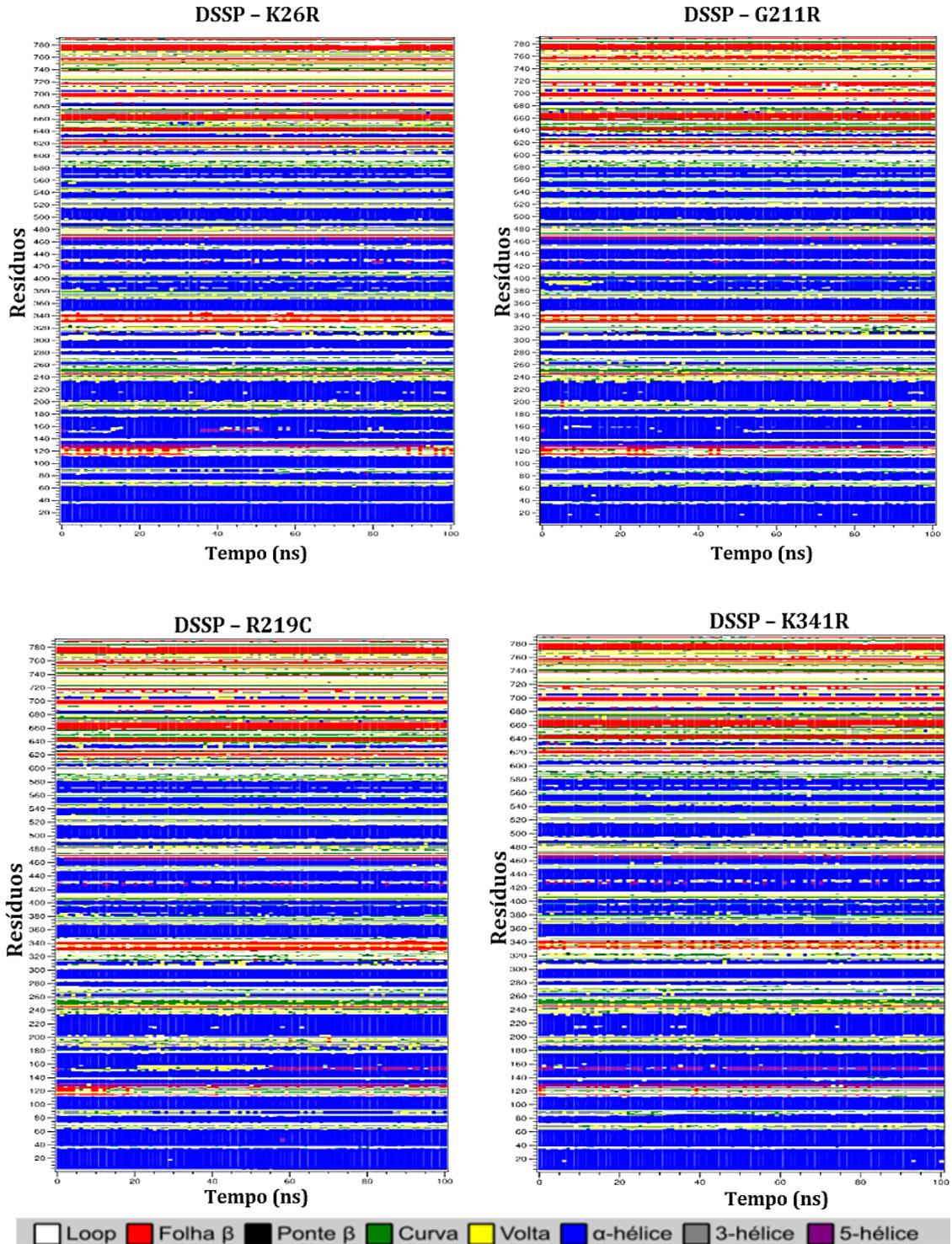


Fonte: da autora (2023).

4.2.5 Dicionário de Estrutura Secundária de Proteínas (DSSP)

As variações estruturais dos resíduos da Spike-RBD selvagem e das variantes da ACE2 em relação à evolução temporal estão representadas na Figura 34. Assim como nos complexos com as variantes da Spike, não houve mudanças significativas na estrutura secundária dos resíduos nos complexos contando as mutações da ACE2 durante a simulação de dinâmica molecular.

Figura 34. Ilustração gráfica do DSSP dos complexos com variantes da ACE2 e Spike-RBD canônica. As estruturas secundárias estão representadas conforme as cores indicadas na legenda da imagem.



Fonte: da autora (2023).

4.2.6 Área de Superfície de Poisson-Boltzmann da Mecânica Molecular (MM/PBSA)

Na interação da Spike-RBD WT com as variantes da ACE2, os valores estimados para

energia de ligação não apresentaram variações significativas (Tabela 3). Isto pode ser atribuído às alterações pontuais nas variantes da ACE2, que não foram suficientes para diferenciar a energia de interação de ligação. Como citado anteriormente, estudos anteriores relataram um aumento na afinidade de ligação da mutação K26R. Porém, a mudança na afinidade de ligação ocorre devido a participação de um glicano ligado a ACE2. No presente estudo, não foram adicionados glicanos junto as estruturas da Spike e ACE2, o que pode explicar o resultado obtido.

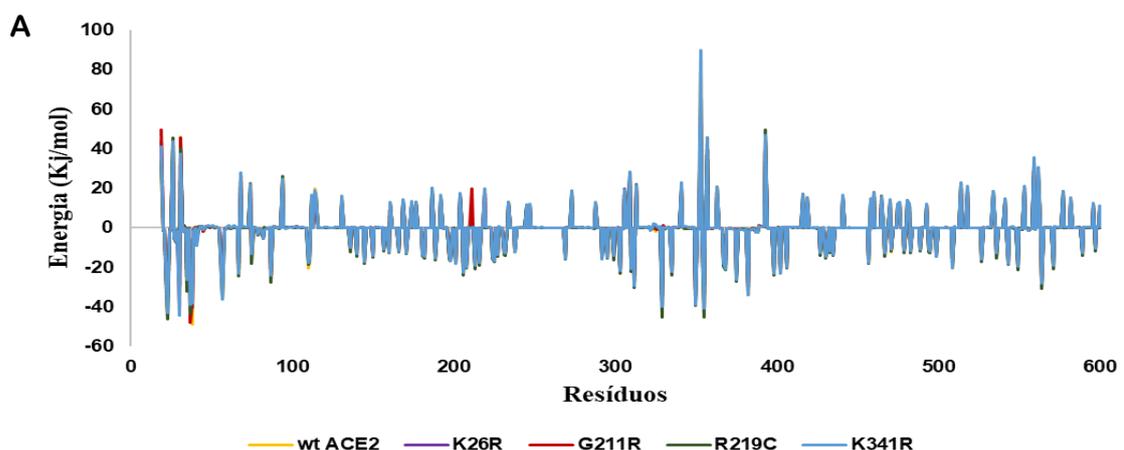
Tabela 3. Decomposição residual e energia de ligação da interação proteína-proteína (dado em kJ.mol⁻¹).

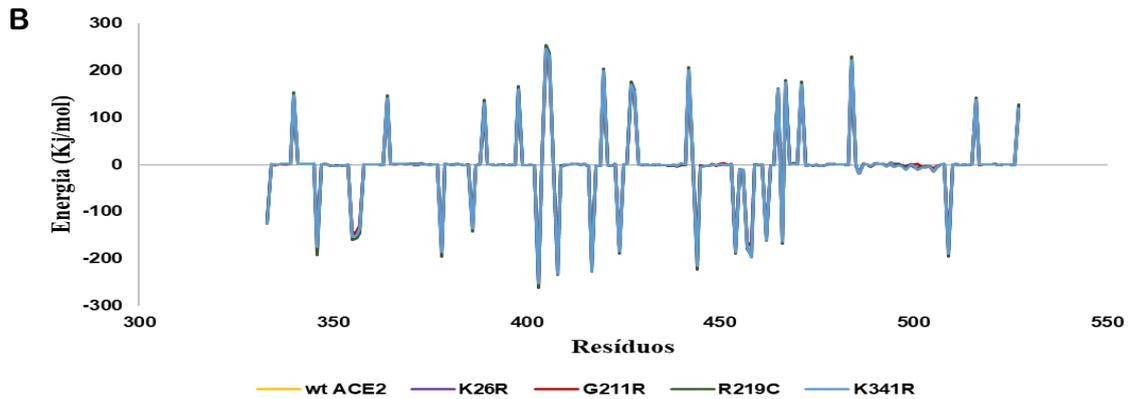
ACE2	SPIKE	ΔE_{vdW}	ΔE_{elet}	Energia Polar de Solvatação	SASA	Energia de Ligação
K26R	Wild	-323.334 ± 26.126	-1243.890 ± 76.663	612.919 ± 157.233	-41.409 ± 3.323	-995.714 ± 140.331
G211R	Wild	-289.217 ± 31.810	-1126.235 ± 91.138	608.183 ± 158.584	-39.983 ± 4.339	-847.252 ± 124.898
R219C	Wild	-337.366 ± 20.153	-1433.607 ± 65.647	669.348 ± 174.176	-41.373 ± 4.127	-1142.998 ± 161.385
K341R	Wild	-333.922 ± 21.914	-1394.377 ± 78.384	801.886 ± 159.180	-42.524 ± 3.730	-968.938 ± 171.393
Wild	Wild	-321.597 ± 22.537	-1429.286 ± 59.435	785.306 ± 107.762	-40.912 ± 3.417	-1006.488 ± 115.085

Fonte: da autora (2023).

Não foram observados valores significativos nos resultados MM/PBSA para os resíduos em questão (Figura 35A e 35B). Porém, os resíduos Val343 e His345, localizados na cadeia ACE2 do complexo contendo a mutação K341R, apresentaram energia de ligação ligeiramente mais fraca em comparação aos demais complexos (Figura 35A). Val343 e His345 estão situados próximos da mutação K341R, todos dentro da região do loop e relativamente próximos ao Spike-RBD.

Figura 35. Decomposição de energia por resíduo MM/PBSA para as trajetórias obtidas nas simulações de DM. (A) Energia de contribuição dos resíduos das variantes da ACE2 em interação com Spike-RBD WT; (B) Energia de contribuição dos resíduos Spike-RBD WT em interação com variantes da ACE2.





Fonte: da autora (2023).

4.2.7 Identificação de resíduos-chave na ACE2 selvagem e variantes por aprendizagem de máquinas

A Tabela 4. mostra os cinco pares responsáveis pelas variações da ligação entre as variantes da ACE2 e Spike-RBD. Ambos os modelos RF e MLP identificaram resíduos de pares idênticos, como Phe356-Tyr495 em K26R e Gly104-Phe486 e Ser105-Tyr489 em K341R. Para a ACE2 WT e a variante R219C, os pares de resíduos identificados pelo modelo RF diferiram daqueles identificados pelo modelo MLP. No entanto, o resíduo Phe486 (em R219C) e Gln325 (em WT) foram identificados em ambos os métodos. Notavelmente, o resíduo Ser19 na proteína ACE2 recebeu pontuações altas para todos os polimorfismos, incluindo K26R, K341R e R219C identificados por MLP, além do WT identificado pela RF.

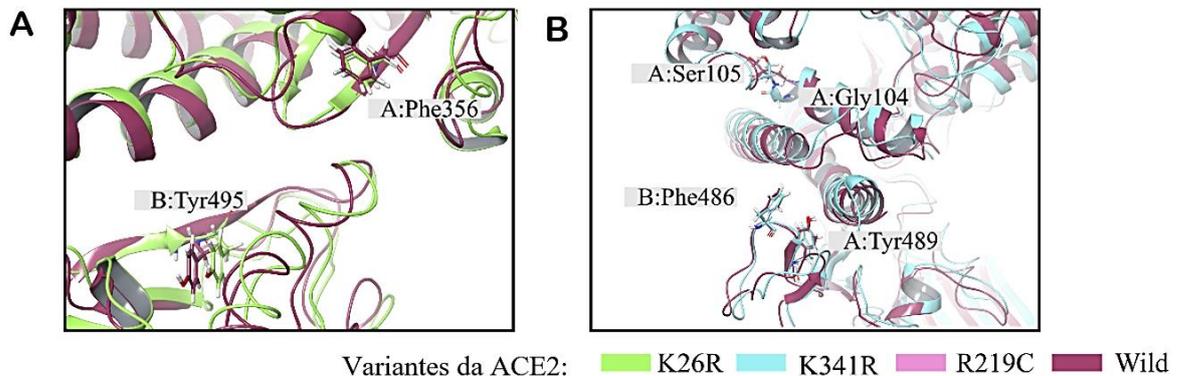
Tabela 4. Resíduos importantes obtidos por MLP e RF para as variantes de ACE2 interagindo com a Spike-RBD. Os pares destacados em negrito e marcados com asteriscos apareceram tanto em análises de MLP como de RF.

MLP				RF			
Variante da ACE2	ACE2	Spike	Importância	Variante da ACE2	ACE2	Spike	Importância
K26R	SER19	SER477	1.00	K26R	PHE356*	TYR495*	1.00*
	SER19	ASP467	0.91		LYS353	TYR495	0.94
	PHE356*	TYR495*	0.80*		GLY352	TYR495	0.94
	LEU351	TYR495	0.41		ASP355	ASN439	0.79
	ASN194	PHE486	0.39		ASP350	TYR495	0.66
K341R	GLY104*	PHE486*	1.00*	K341R	GLY104*	PHE486*	1.00*
	SER19	PRO479	1.00		GLN102	ASN487	0.79
	ALA193	PHE486	0.86		SER105*	TYR489*	0.78*
	SER105*	TYR489*	0.80*		LYS26	GLY504	0.73
	ASN103	TYR505	0.77		GLN102	TYR489	0.65
R219C	SER19	PRO479	1.00	R219C	SER105	TYR489	1.00
	ASP38	LYS444	0.54		GLY104	PHE486	0.98
	PHE32	GLY446	0.43		GLN102	GLY485	0.91
	PHE40	TYR495	0.43		ASN103	GLY485	0.86
	ALA193	PHE486	0.41		ASN103	TYR489	0.84
Selvagem (Wild)	ASN103	TYR505	1.00	Selvagem (Wild)	SER19	ALA475	1.00
	ASN103	TYR489	0.85		GLN325	THR500	0.96
	ANS103	GLY485	0.80		SER19	GLN474	0.87
	HIS401	TYR505	0.74		GLN325	GLN498	0.81
	GLN325	TYR449	0.54		GLN325	ASN448	0.76

Fonte: da autora (2023).

As Figuras 36A e 36B ilustra os principais resíduos que diferenciam as variantes do WT.

Figura 36. Pares de resíduos determinados como mais importantes para diferenciar a ligação entre a Spike-RBD WT e variantes da ACE2. (A) Pares de resíduos importantes encontrados para Spike-RBD WT e K26R da ACE2; (B) Pares de resíduos importantes encontrados para Spike-RBD WT e K341R da ACE2. Resíduos importantes são compostos por pares de distâncias altamente distintos entre as variantes da ACE2 e a Spike-RBD.



Fonte: adaptada do manuscrito submetido (Ávila, A. L. et al, 2023).

A variante K26R da ACE2 aumenta a suscetibilidade ao SARS-CoV-2 devido a uma maior afinidade de ligação com a proteína Spike (SURYAMOHAN *et al.*, 2021). Nesta variante, Phe356 forma um par com Tyr495. Phe356 está localizado em uma folha-beta, próximo aos resíduos Tyr41, Gln42, Lys353 e Arg357, que interage com o Spike-RBD, enquanto Tyr495 participa de uma rede de ligações de hidrogênio com a ACE2. No entanto, atualmente não há literatura disponível sobre a importância específica do Phe356 na variante K26R ao interagir com a Spike-RBD.

A mutação K341R, que substitui a lisina por arginina, resulta em um resíduo mutante maior, que pode causar saliências (RODRIGUEZ *et al.*, 2022). Neste estudo, são destacadas variações nos pares Gly104-Phe486 e Ser105-Tyr489. Os papéis de alguns desses resíduos foram documentados por Ali e Vijayan (2020). Especificamente, Phe486 participa de importantes interações polares com Tyr83 e interações hidrofóbicas com Leu79 da ACE2. Tyr489 está envolvida em interações polares com Thr27 e Lys31, bem como interações hidrofóbicas com Phe28, Tyr83, Thr27, Phe32 e Phe72 da ACE2. É importante ressaltar que os resíduos Gly104 e Ser105 da ACE2 estão localizados próximos dos resíduos citados anteriormente.

Na variante R219C, o modelo MLP identificou a interação Ala193 da ACE2 e Phe486. O resíduo Ala193 é um aminoácido apolar e está localizado em um alfa-hélice próximo à mutação R219C, o qual possui o aminoácido polar cisteína.

Na ACE2 WT, o modelo RF identificou a interação de Gln325 da ACE2 com os resíduos Asn448, Gln498 e Thr500 da Spike. Os três resíduos da Spike estão localizados em uma região de loop próximos ao resíduo Gln325 da ACE2. Todos os resíduos envolvidos nesta interação são polares não carregados.

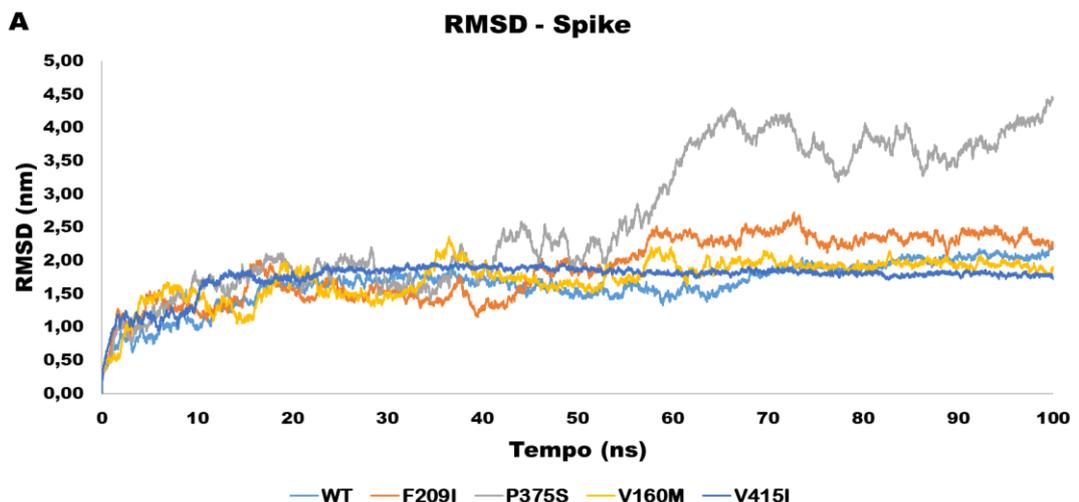
4.3 Complexo Spike selvagem e variantes TMPRSS2

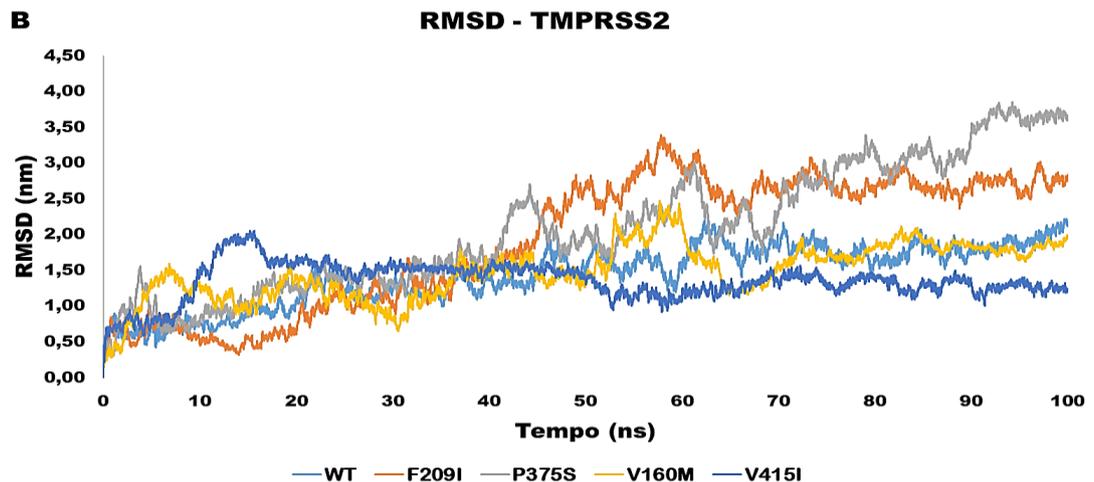
4.3.1 Raiz do Desvio Quadrado Médio (RMSD)

No gráfico com as trajetórias da cadeia Spike (Figura 37A), todas as trajetórias permaneceram estáveis durante os 100ns de simulação, com exceção da trajetória do complexo com a variante P375S, que obteve um aumento de aproximadamente 1,5 nm no valor RMSD partir de 55 ns. Também foi observado que a trajetória do complexo com a variante V415I demonstra ser a mais estável, permanecendo retilínea a partir de 25 ns.

As trajetórias da cadeia TMPRSS2 (Figura 37B) seguem o mesmo padrão observado na cadeia Spike. A trajetória da variante P375S demonstra um aumento crescente do RMSD, caracterizando uma desestabilização da proteína. A trajetória da variante V415I, ao contrário, demonstra estabilidade e permanece com o menor valor RMSD a partir de 50 ns, caracterizando menor flexibilidade.

Figura 37. Gráfico RMSD das trajetórias da cadeia Spike (A) e TMPRSS2 (B).





Fonte: da autora (2023).

4.3.2 Flutuação da Raiz Quadrada Média (RMSF)

A estrutura terciária da proteína Spike contém muitas regiões de loop, o que explica as diferentes flutuações entre as trajetórias em diversos pontos do gráfico contendo as trajetórias da cadeia Spike (Figura 38). Os picos com RMSF mais elevados estão em localizados em regiões de loop distantes da interação com a proteína Tmprss2.

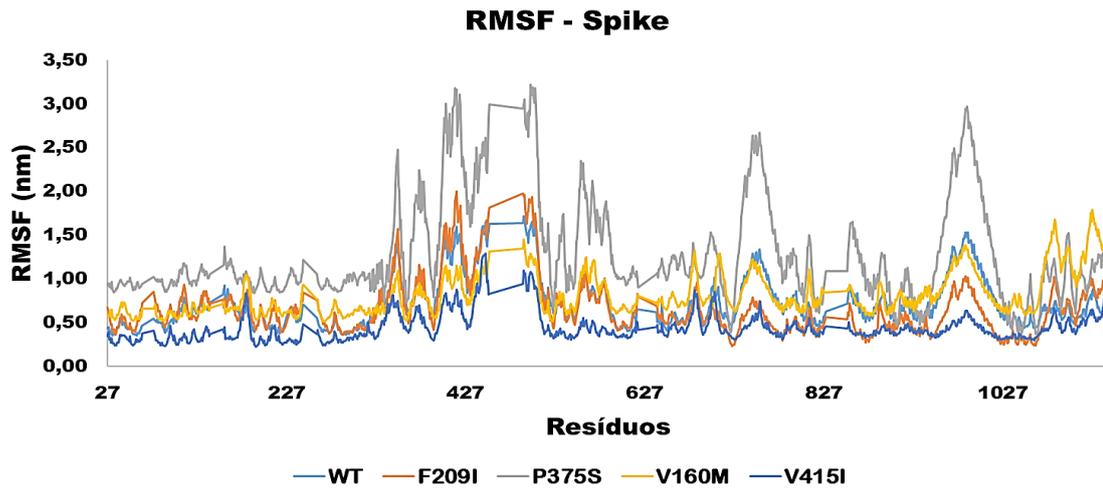
Os resíduos entre 748 e 756 da cadeia Spike da trajetória do complexo com a variante P375S (Figura 39A) estão localizados em uma alfa-hélice, com flutuação RMSF de aproximadamente 1,5 nm mais alta em relação às trajetórias dos outros complexos. O mesmo ocorre com os resíduos 985 a 995 (Figura 39A), também localizados na alfa-hélice. Os resíduos citados acima, porém, não estão próximos da região de interação entre a Spike com a Tmprss2.

O resíduo P793 da trajetória pertencente ao complexo com a variante P375S (Figura 39B) obteve maior flutuação, com valor RMSF de aproximadamente 0,5 nm acima das outras trajetórias. P793 está localizado em um loop próximo da Tmprss2.

O resíduo K811, também localizado em um loop próximo da Tmprss2 (Figura 39C), obteve maior flutuação na trajetória da variante P375S (1,5 nm), seguido pela variante V160M (1,11 nm), WT (0,70 nm), F209I (0,63 nm) e V415I (0,55 nm). K811 está localizado em um loop próximo da Tmprss2.

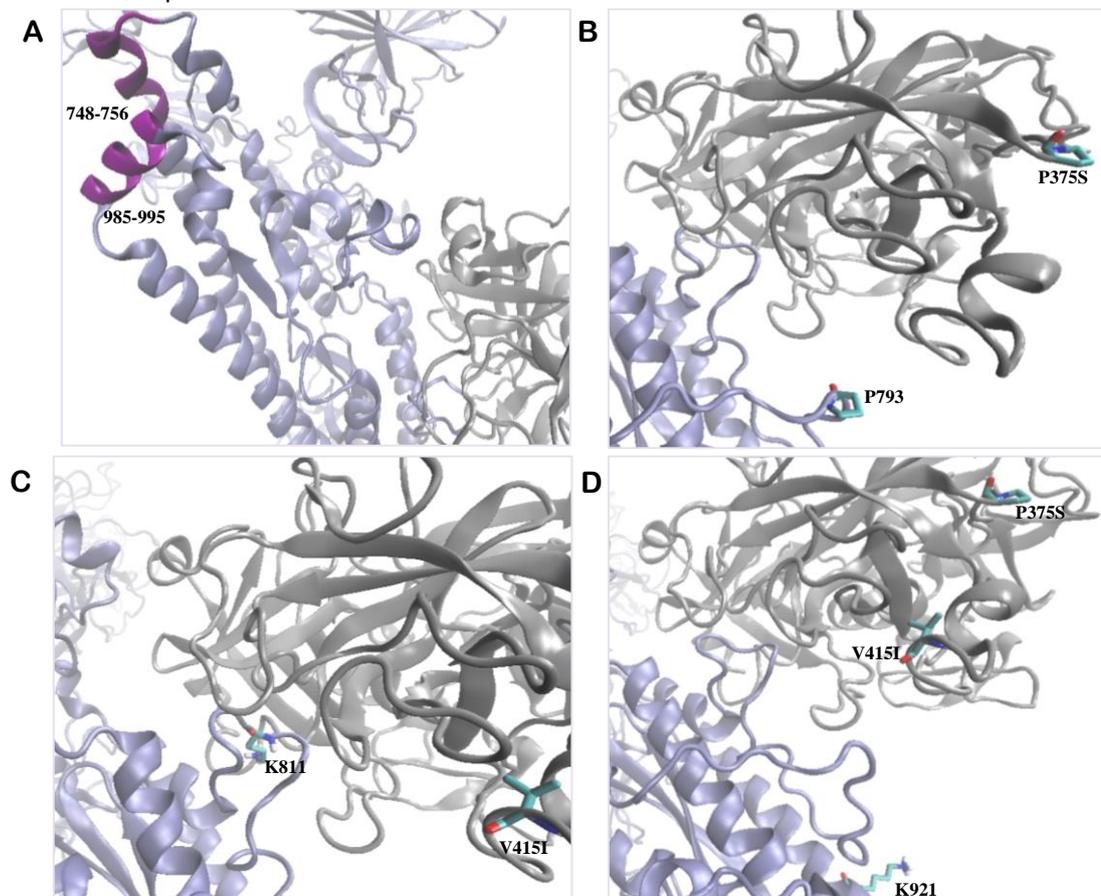
K921, localizado em uma alfa-hélice próximo da Tmprss2 (Figura 39D), obteve maior flutuação no complexo da variante P375S e menor flutuação na trajetória da variante V415I, com uma diferença de aproximadamente 0,6 nm entre ambas.

Figura 38. RMSF dos resíduos da cadeia Spike, definida pelos resíduos 27 a 1146.



Fonte: da autora (2023).

Figura 39. (A) Resíduos 748-756 e 985-995 (destacados em roxo) obtiveram alta flutuação e estão localizados em uma alfa-hélice da Spike, porém encontram-se distantes da interação com a TMPRSS2. (B) Localização do resíduo P793 na Spike. (C) Localização do resíduo K811 na Spike. (D) Localização do resíduo K921 na Spike.



No gráfico RMSF da cadeia TMPRSS2 (Figura 40), as flutuações com picos mais elevados estão localizadas em regiões de loop e distantes da região de interação com a Spike

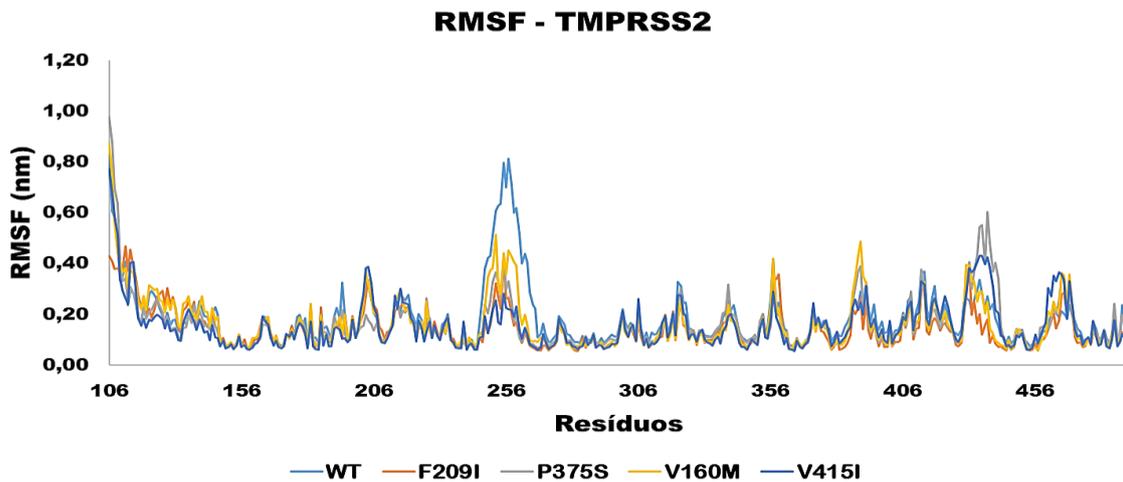
(Figura 41).

O resíduo K390 está localizado em um loop próximo da Spike e obteve maior flutuação no complexo com a variante V160M (0,48 nm), seguida do complexo com a variante P375S (0,39 nm), com diferença de aproximadamente 0,15 nm dos complexos WT, F209I e V415I.

O resíduo Q438, também localizado em um loop próximo da Spike, alcançou maior valor RMSF no complexo da variante P375S (0,60 nm). O segundo maior pico foi o do complexo com a variante V415I (0,42 nm), seguida da WT (0,27 nm), V160M (0,24 nm) e F209I (0,17 nm).

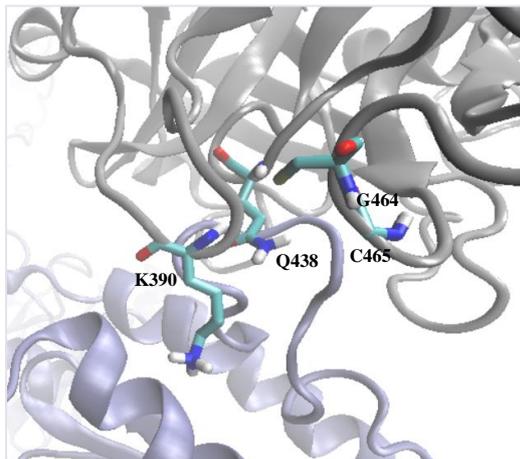
G464 e C465 obtiveram maior flutuação no complexo da variante V415I (0,36 nm), com uma diferença de 0,12 nm no valor RMSF dos outros complexos. Ambos os resíduos, G464 e C465 estão localizados em um loop próximo da Spike.

Figura 40. RMSF dos resíduos da cadeiaTMPRSS2, definida pelos resíduos 106 a 491.



Fonte: da autora (2023).

Figura 41. Localização dos resíduos K390, Q438, G464 e C465 na TMPRSS2.



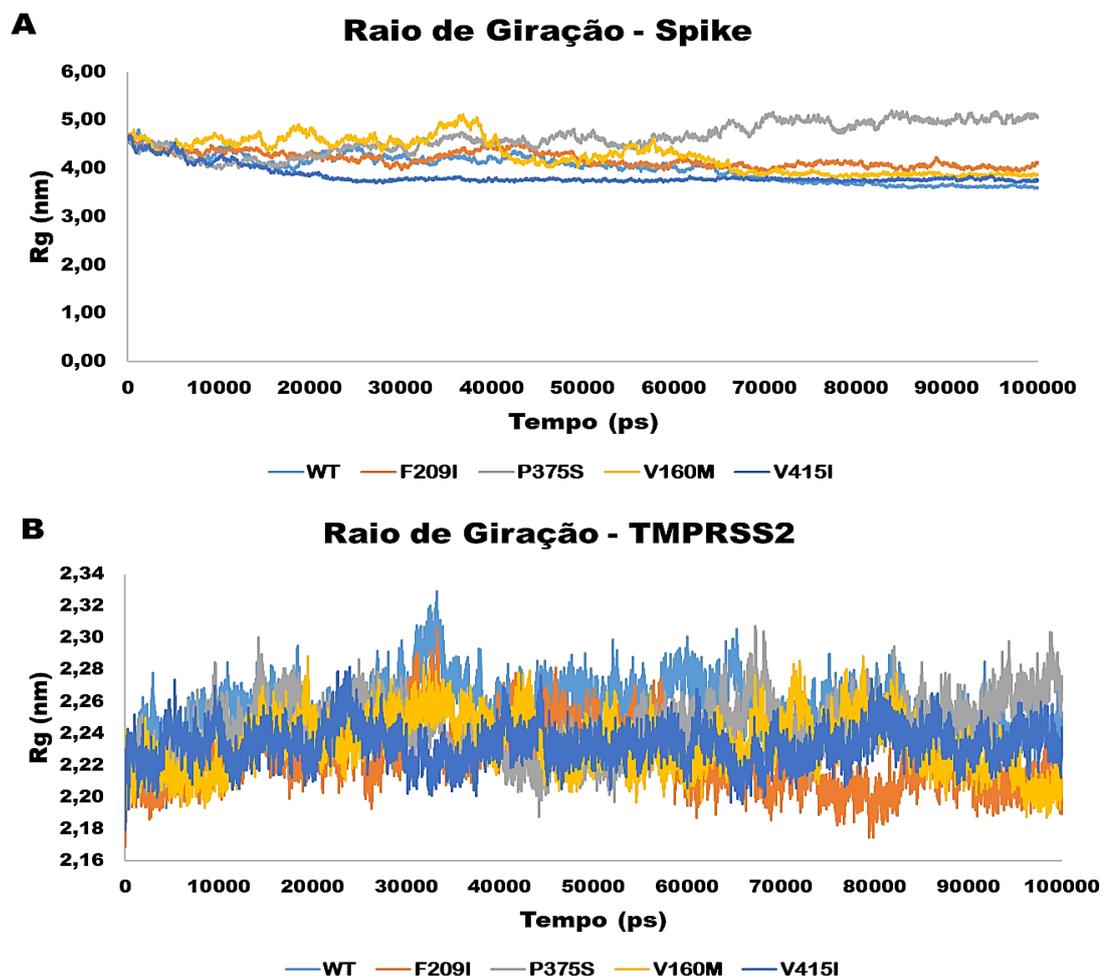
Fonte: da autora (2023).

4.3.3 Raio de Giração (R_g)

O gráfico com os valores do Raio de Giração referentes a cadeia Spike de cada complexo representado na Figura 42A mostra que a trajetória do complexo selvagem (WT) e as trajetórias das variantes V160M, F209I e V415I entram em equilíbrio a partir de 70000 ps, mantendo-se constantes e apresentando valores de R_g próximos. A única exceção é a trajetória do complexo contendo a variante P375S, a qual apresenta elevação no valor R_g de 1 nm a mais em relação às trajetórias dos outros complexos a partir de 60000 ps. Esta diferença indica que a cadeia Spike do complexo da variante P375S possui menor compactação e, portanto, é mais flexível e com menor estabilidade.

A Figura 42B apresenta o gráfico do Raio de Giração da cadeia TMPRSS2. Todos os complexos permanecem estáveis e com valores constantes durante todo o período de tempo analisado.

Figura 42. Raio de Giração da cadeia Spike (A) e TMPRSS2 (B).

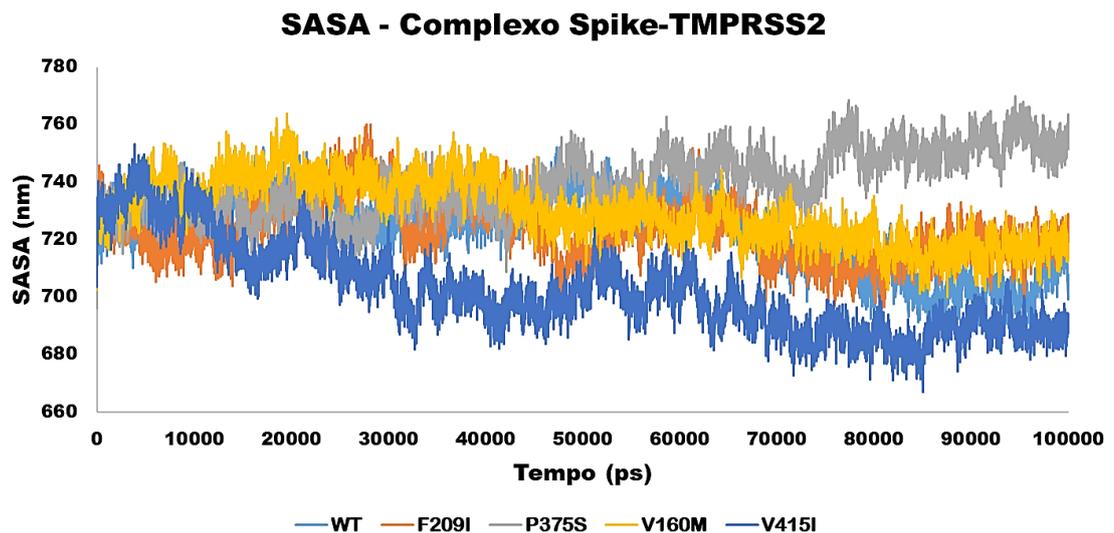


Fonte: da autora (2023).

4.3.4 Área Acessível ao Solvente (SASA)

O gráfico SASA das trajetórias dos complexos Spike-TMPRSS2 (Figura 43) mostra um aumento de aproximadamente 20 nm no valor de SASA do complexo com a variante P375S, caracterizando maior desestabilização deste complexo em relação aos demais. Além disso, nota-se que o complexo com a variante V415I apresenta SASA ligeiramente menor que os complexos WT, V160M e F209I, indicando maior estabilidade.

Figura 43. SASA das trajetórias dos complexos com as variantes da TMPRSS2 e Spike selvagem.



Fonte: da autora (2023).

5. CONCLUSÃO

A interação entre as proteínas Spike e ACE2, assim como Spike e TMPRSS2, desempenha um papel crucial na determinação da taxa de replicação do SARS-CoV-2 e tem implicações na progressão da doença em indivíduos infectados. O vírus apresenta uma pronunciada propensão para mutações, como evidenciado pelo surgimento de várias variantes nos últimos anos. Resíduos específicos desempenham um papel crucial na mediação das interações moleculares entre ACE2-Spike e TMPRSS2-Spike. A suscetibilidade à COVID-19 é influenciada pela estabilidade estrutural, afinidade de ligação e nível de expressão destas três proteínas em interação. Mutações genéticas de ocorrência natural podem levar a substituições de aminoácidos chave, potencialmente modificando interações específicas entre as proteínas e contribuindo para o aumento da heterogeneidade clínica na COVID-19. Assim, as mutações não sinônimas encontradas nas proteínas responsáveis pela interação do vírus com o hospedeiro

podem fornecer explicações potenciais para os diferentes graus de susceptibilidade ao COVID-19, sendo um dos fatores genéticos que contribuem para a diversidade de casos de infecção. Portanto, a identificação de assinaturas de variação em nível de populações específicas é um aspecto fundamental da doença e do controle global da pandemia. Além disso, compreender as variantes populacionais não apenas fornece informações sobre a suscetibilidade ao SARS-CoV-2, mas também pode ser utilizada para desenvolver inibidores especializados e de alta afinidade. No entanto, uma compreensão abrangente de como as mutações impactam a estabilidade e a dinâmica de interação dentro do complexo SARS-CoV-2-ACE2 e SARS-CoV-2-TMPRSS2 continua a ser um esforço contínuo.

Neste trabalho, foram utilizadas simulações de DM para explorar a interação entre diferentes variantes do SARS-CoV-2 e variantes da ACE2 e TMPRSS2. Através dessas simulações, foram obtidas informações significativas sobre a interação proteína-proteína. Ao mesmo tempo, o emprego de técnicas de aprendizado de máquina permitiu identificar resíduos críticos de aminoácidos dentro da região de ligação entre Spike e ACE2 que contribuem sutilmente para essa interação.

As simulações de DM revelaram padrões de estabilidade semelhantes entre os complexos estudados. Além disso, as trajetórias resultantes indicaram uma convergência das simulações para um estado de equilíbrio. A estabilidade do complexo proteico ACE2 com Spike-RBD WT foi ligeiramente diminuída, como evidenciado pelos valores RMSD, em contraste com os complexos das variantes do SARS-CoV-2. Esta observação é consistente com o efeito previsto de mutações na região de interação Spike, levando a uma maior estabilidade.

Embora as flutuações mais significativas tenham sido observadas nas regiões de *loop*, alguns resíduos próximos à interface de interação exibiram flutuações notáveis. Arg408 e Lys444 do Spike-RBD apresentaram valores de RMSF ligeiramente mais elevados na variante Delta. Os resíduos Gln325 e Trp328 da proteína ACE2 apresentaram menores flutuações nas trajetórias do polimorfismo K341R, enquanto Trp436 e Ser443 exibiram maiores flutuações para WT e G211R, respectivamente. Mudanças significativas no RMSF nessas regiões podem sugerir alterações conformacionais importantes para a atividade biológica na interação ACE2-Spike.

A variante Ômicron demonstra uma afinidade mais forte com ACE2, como evidenciado pelos valores MM/PBSA, onde as mutações Q493R e Q498R contribuíram mais significativamente para a energia de ligação. Em relação aos complexos formados entre as variantes da ACE2 e o Spike-RBD, não foram identificadas diferenças significativas nos

resultados de MM/PBSA para os resíduos considerados. Estas mutações pontuais não foram suficientes para gerar variações notáveis detectáveis usando o método MM/PBSA.

Em relação aos métodos de aprendizado de máquina, as abordagens ML e RF identificaram os principais resíduos de ambas as proteínas responsáveis pelas diferenças na região de ligação, algumas das quais foram relatadas anteriormente na literatura. Isso demonstra que o método foi capaz de identificar resíduos que contribuam significativamente para a distinção entre interação entre vírus e hospedeiro devido às variantes da Spike e ACE2, identificando até mesmo pares de resíduos que não foram documentados anteriormente na literatura existente. Dentre os pares de resíduos mais significativos para o complexo das variantes da Spike com a ACE2 WT identificados pelo modelo MLP, destacam-se: Ser19-Pro479 em WT, Asp30-Glu484 e Gln24-Lys417 na variante Delta, Glu329-Ser438, Tyr381-Gly502, Gly352-Asn448, Gly504-Asn448 na variante Ômicron e Ser19-Pro479 na variante P2. Nos pares identificados pelo modelo RF, destacam-se Ala25/Gln24/Ile21/Thr27-Asn417 e Lys353-Arg498 na variante Ômicron e Ser106-Lys484 e Ser105-Lys484 na variante P2. Para o complexo com a Spike WT e variantes da ACE2, os pares de resíduos mais significativos identificados foram Phe356-Tyr495 na variante K26R, Gly104-Phe486 e Ser105-Tyr489 na variante K341R, em ambos os modelos MLP e RF. No modelo MLP, Ala193-Phe486 na variante R219C e Gln325-Tyr449 na ACE2 WT. No modelo RF, Gln325-Thr500/Gln498/Asn448 na ACE2 WT.

Os valores RMSD mostraram uma diminuição significativa na estabilidade do complexo Spike com a variante P375S da TMPRSS2. Em contraste, o complexo Spike com a variante V415I da TMPRSS2 apresentou a maior estabilidade.

Como esperado, as flutuações mais significativas aparecem no complexo Spike com a variante P375S da TMPRSS2, com destaque para o resíduo P793 na cadeia Spike, localizado em um loop próximo da interação com a TMPRSS2. Na cadeia TMPRSS2, destaca-se o resíduo Q438, também localizado em um *loop* próximo da região de interação com a Spike.

É importante ressaltar que a variante P375S possui a substituição de um resíduo de aminoácido de estrutura cíclica e, portanto, mais rígido (prolina) por um aminoácido flexível devido à presença de uma cadeia lateral relativamente simples (serina). Esta substituição pode levar a alterações na estrutura proteica local e global, podendo introduzir maior flexibilidade. A prolina é frequentemente encontrada em regiões de *loop* de proteínas, e sua substituição por serina pode afetar a estabilidade e a formação dessas estruturas secundárias. Além disso, a prolina desempenha um papel nas interações proteína-proteína e a sua substituição pode afetar

a afinidade de ligação da proteína aos seus resíduos de interação.

Os resultados do presente estudo indicam que podem haver resíduos importantes além daqueles já relatados na literatura que podem impactar na interação entre as proteínas. Esses resíduos podem ser responsáveis por diferenças na estabilidade da proteína e na afinidade de ligação, levando a níveis variados de suscetibilidade ao SARS-CoV-2 e resultando em vários graus de gravidade da doença. Assim, este trabalho contribui para o estudo da suscetibilidade genética do COVID-19 e no desenvolvimento de medicamentos novos/reaproveitados contra o SARS-CoV-2, fornecendo dados para a compreensão mais profunda da relação entre mutações e a afinidade entre Spike-ACE2 e Spike-TMPRSS2.

REFERÊNCIAS

- ABRAHAM, M. J. et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, p. 19-25, 2015.
- ALAM, M. et al. Beta, Delta, and Omicron, Deadliest Among SARS-CoV-2 Variants: A Computational Repurposing Approach. **Evolutionary Bioinformatics**, 2023;19. doi:10.1177/11769343231182258
- ALI, S. A. et al. A review of methods available to estimate solvent-accessible surface areas of soluble proteins in the folded and unfolded states. **Current Protein & Peptide Science**, v. 15, n. 5, p. 456–476, 2014.
- ALI, A.; VIJAYAN, R., Dynamics of the Ace2–SARS-CoV2/SARS-CoV Spike protein interface reveal unique mechanisms. **Sci Rep**, vol. 10(14214), pp. 1-12, Aug. 2020, doi: <https://doi.org/10.1038/s41598-020-71188-3>
- ANDERSEN, C. A. F. et al. Continuum Secondary Structure Captures Protein Flexibility. **Structure**, v. 10, n. 2, p. 175–184, 1 fev. 2002.
- ARAF, Y. et al. Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. **Journal of Medical Virology**, v. 94, n. 5, p. 1825–1832, 2022. CDC. **Coronavirus Disease 2019 (COVID-19)**.
- ARQUIVO BRASILEIRO ONLINE DE MUTAÇÕES. Disponível em: <http://abraom.ib.usp.br>. Acesso em: 02 out. 2023.
- ÁVILA, A. L.; SANCHES, A. C.; SCORSOLINI, A.; ALVES, L.; GIULIATTI, S. Comprehensive Machine Learning Analysis of Key Residues in Variants and Polymorphisms for Ace2-Spike Interaction in SARS-CoV-2. Manuscrito submetido.
- BENETTI, E. ACE2 gene variants may underlie interindividual variability and susceptibility to COVID-19 in the Italian population. **Eur J Hum Genet**, 2020 Nov;28(11):1602-1614. doi: 10.1038/s41431-020-0691-z. Epub 2020 Jul 17. PMID: 32681121; PMCID: PMC7366459.
- BERMAN, H. et al. *The Protein Data Bank*. (2000). **Nucleic Acids Research**, 28: 235-242. Doi: <https://doi.org/10.1093/nar/28.1.235>. Disponível em: www.rcsb.org. Acesso em: 02 out. 2023.
- BREIMAN, L. Random Forests, **Machine Learning**, vol.45, pp. 5–32, 2001.
- BRÜSCHWILER, R. Efficient RMSD measures for the comparison of two molecular ensembles. **Proteins**, [S. l.], v. 50, n. 1, p. 26-34, jan. 2003.
- CHEN, J. et al. Omicron (B.1.1.529): Infectivity, vaccine breakthrough, and antibody resistance. **ArXiv**, p. arXiv:2112.01318v1, 1 dez. 2021.

CHOUDHARY, S. et al. Role of Genetic Variants and Gene Expression in the Susceptibility and Severity of COVID-19. **Annals of Laboratory Medicine**, v. 41, n. 2, p. 129–138, 1 mar. 2021.

CLINCKEMALIE, L. et al. Androgen Regulation of the TMPRSS2 Gene and the Effect of a SNP in an Androgen Response Element. **Molecular Endocrinology**, v. 27, n. 12, p. 2028–2040, dez. 2013.

DAVID, L.; COX, M. **Princípios de Bioquímica de Lehninger**. 6. ed. Artmed. 2013.

DE FREITAS, M. **Dinâmica Molecular e Análise In Silico da Proteína Triptofano Hidroxilase 2 (TPH2)**. Dissertação (Mestrado), Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Genética e Biologia Molecular, Rio de Janeiro, 2017.

DENG, X. et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. **Cell**. 2021 Jun 24;184(13):3426-3437.e8. doi: 10.1016/j.cell.2021.04.025. Epub 2021 Apr 20. PMID: 33991487; PMCID: PMC8057738.

DEVAUX, C. A.; ROLAIN, J.-M.; RAOULT, D. ACE2 receptor polymorphism: Susceptibility to SARS-CoV-2, hypertension, multi-organ failure, and COVID-19 disease outcome. **Journal of Microbiology, Immunology, and Infection**, v. 53, n. 3, p. 425–435, jun. 2020.

DI GIACOMO, S. et al. Preliminary report on severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Spike mutation T478K. **J Med Virol**. 2021 Sep;93(9):5638-5643. doi: 10.1002/jmv.27062. Epub 2021 May 15. PMID: 33951211; PMCID: PMC8242375.

DE ANDRADE, J.; GONÇALVES, P.F.B.; NETZ, P.A. Why Does the Novel Coronavirus Spike Protein Interact so Strongly with the Human ACE2? A Thermodynamic Answer. **ChemBiochem**, vol. 22(5), pp. 865-875, Mar. 2021, doi: 10.1002/cbic.202000455

DU, L. et al. The spike protein of SARS-CoV — a target for vaccine and therapeutic development. **Nature Reviews. Microbiology**, v. 7, n. 3, p. 226–236, 2009.

DU, L. et al. MERS-CoV spike protein: a key target for antivirals. **Expert opinion on therapeutic targets**, v. 21, n. 2, p. 131–143, fev. 2017.

DUMAN, N. Analysis of ACE2 and TMPRSS2 coding variants as a risk factor for SARS-CoV-2 from 946 whole-exome sequencing data in the Turkish population. **J Med Virol**. Jul, 2022. doi: 10.1002/jmv.27976.

EISENHABER, F. et al. The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. **Journal of Computational Chemistry**, [S. l.], v. 16, n. 3, p. 273-284, mar. 1995.

FLEETWOOD, O. et al. Molecular Insights from Conformational Ensembles via Machine Learning. **Biophysical Journal**, vol. 118(3), pp. 765–780, Feb. 2020. doi:10.1016/j.bpj.2019.12.016.

FUJIKURA, K.; UESAKA, K. Genetic variations in the human severe acute respiratory syndrome coronavirus receptor ACE2 and serine protease TMPRSS2. **Journal of Clinical Pathology**, p. jclinpath-2020-206867, jul. 2020.

GENOME AGGREGATION DATABASE. Disponível em: <https://gnomad.broad.institute.org>. Acesso em: 02 out. 2023.

HARRACH, M. F.; DROSSEL, B. Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity. **The Journal of Chemical Physics**, v. 140, n. 17, p. 174501, 7 maio 2014.

HEURICH, A. et al. TMPRSS2 and ADAM17 Cleave ACE2 Differentially and Only Proteolysis by TMPRSS2 Augments Entry Driven by the Severe Acute Respiratory Syndrome Coronavirus Spike Protein. **Journal of Virology**, v. 88, n. 2, p. 1293–1307, jan. 2014.

HOFFMANN, M. et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. **Cell**, v. 181, n. 2, p. 271- 280.e8, 16 abr. 2020.

HOU, Y. et al. New insights into genetic susceptibility of COVID-19: an ACE2 and TMPRSS2 polymorphism analysis. **BMC Medicine**, v. 18, p. 216, 15 jul. 2020.

HUANG, J.; MACKERELL, A. D. and A. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. **J. Comput. Chem.**, vol. 34(1), pp. 2135- 2145, Sep. 2013, doi:10.1002/jcc.23354.

HUANG, S. et al. PMFF: Development of a Physics-Based Molecular Force Field for Protein Simulation and Ligand Docking. **J. Phys. Chem. B**, v. 6, p. 974–989, 2020.

HUANG, Y. et al. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. **Acta Pharmacologica Sinica**, v. 41, n. 9, p. 1141–1149, set. 2020.

HUG, S. Classical Molecular Dynamics in a Nutshell. Em: MONTICELLI, L.; SALONEN, E. (Eds.). **Biomolecular Simulations: Methods and Protocols. Methods in Molecular Biology**. Totowa, NJ: Humana Press, 2013. p. 127–152.

HUSSAIN, M. et al. Molecular docking between human TMPRSS2 and SARS-CoV-2 spike protein: conformation and intermolecular interactions. **AIMS Microbiology**, v. 6, n. 3, p. 350–360, 24 set. 2020.

ISHAK, S. et al. Molecular dynamic simulation of space and earth grown crystal structures of thermostable T1 lipase *Geobacillus zalihae* revealed a better structure. **Molecules**, v. 22, n. 10, p. 1574, 2017.

ISTIFLI, E. S. et al. Understanding the molecular interaction of SARS-CoV-2 spike mutants with ACE2 (angiotensin converting enzyme 2). **J Biomol Struct Dyn**. 2022;40(23):12760-12771. doi: 10.1080/07391102.2021.1975569. Epub 2021 Sep 8. PMID: 34495817; PMCID: PMC8442754.

JAIMES, J.; MILLET, J.; WHITTAKER, G. Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. **IScience**, v. 23, ed. 6, 2020.

JAWAD, B. et al. Key Interacting Residues between RBD of SARS-CoV-2 and ACE2 Receptor: Combination of Molecular Dynamics Simulation and Density Functional Calculation. **Journal of Chemical Information and Modeling**, v. 61, n. 9, p. 4425–4441, 27 set. 2021.

JAWAD, B. et al. Binding Interactions between RBD of Spike-Protein and Human ACE2 in Omicron variant. **BioRxiv**. 2022.02.10.480009; doi: <https://doi.org/10.1101/2022.02.10.480009>

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

KOWALCZUK, S. et al. A protein complex in the brush-border membrane explains a Hartnup disorder allele. **The FASEB Journal**, v. 22, n. 8, p. 2880–2887, 2008.

KUMAR, S. et al. Omicron and Delta variant of SARS-CoV-2: A comparative computational study of spike protein. **Journal of Medical Virology**, v. 94, n. 4, p. 1641–1649, 2022.

KUMARI, R. et al. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. **Journal of Chemical Information and Modeling**, v. 54, n. 7, p. 1951–1962, 28 jul. 2014.

KUZMANIC, A.; ZAGROVIC, B. Determination of Ensemble-Average Pairwise Root Mean-Square Deviation from Experimental B-Factors. **Biophysical Journal**, v. 98, n. 5, p. 861–871, 3 mar. 2010.

LA ROSA, G. et al. Rapid screening for SARS-CoV-2 variants of concern in clinical and environmental samples using nested RT-PCR assays targeting key mutations of the spike protein. **Water Research**, v. 197, p. 117104, 1 jun. 2021.

LAN, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. **Nature**, 581, 215–220. 2020.

LI, F. Receptor recognition and cross-species infections of SARS coronavirus. **Antiviral Research**, v. 100, n. 1, p. 246–254, out. 2013.

LI, W. et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. **The EMBO Journal**, v. 24, n. 8, p. 1634–1643, 20 abr. 2005.

LIMA, L. et al. As descobertas genômicas do SARS-CoV-2 e suas implicações na pandemia de COVID-19. **Journal of Health & Biological Sciences**. 8. 1. 10.12662/2317-3076jhbs.v8i1.3232.p1-9. 2020.

LIN, B. et al. Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2. **Cancer Research**, v. 59, n. 17, p. 4180–4184, 1 set. 1999.

- LOBANOV, M. Y. et al. Radius of gyration as an indicator of protein structure compactness. **Molecular Biological**, [S. l.], v. 42, p. 623-628, 2008.
- LUAN, J. et al. Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. **Biochemical and Biophysical Research Communications**, v. 526, n. 1, p. 165–169, 21 maio 2020.
- MAHMOUD, I. S.; JARRAR, Y. B. Targeting the intestinal TMPRSS2 protease to prevent SARS-CoV-2 entry into enterocytes-prospects and challenges. **Molecular Biology Reports**, v. 48, n. 5, p. 4667–4675, 1 maio 2021.
- MARTÍNEZ, L. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. **PLoS One**. 2015 Mar 27;10(3):e0119264. doi: 10.1371/journal.pone.0119264. PMID: 25816325; PMCID: PMC4376797.
- MOHAMMADI, M.; SHAYESTEHPOUR, M.; MIRZAEI, H. The impact of spike mutated variants of SARS-CoV2 [Alpha, Beta, Gamma, Delta, and Lambda] on the efficacy of subunit recombinant vaccines. **The Brazilian Journal of Infectious Diseases**, v. 25, n. 4, p. 101606, 17 ago. 2021.
- MONTICELLI, L.; TIELEMAN, D. P. Force Fields for Classical Molecular Dynamics. Em: MONTICELLI, L.; SALONEN, E. (Eds.). *Biomolecular Simulations: Methods and Protocols. Methods in Molecular Biology*. Totowa, NJ: Humana Press, 2013. p. 197–213.
- NAMBA, A. M.; SILVA, V. B. DA; SILVA, C. H. T. P. DA. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. **Eclética Química**, v. 33, p. 13–24, dez. 2008.
- NATIONAL CENTER FOR IMMUNIZATION AND RESPIRATORY DISEASES (NCIRD), DIVISION OF VIRAL DISEASES. Science Brief: Omicron (B.1.1.529) Variant. Em: **CDC COVID-19 Science Briefs**. Atlanta (GA): Centers for Disease Control and Prevention (US), 2020.
- PANIRI, A.; HOSSEINI, M. M.; AKHAVAN-NIAKI, H. First comprehensive computational analysis of functional consequences of TMPRSS2 SNPs in susceptibility to SARS-CoV-2 among different populations. **Journal of Biomolecular Structure & Dynamics**, p. 1–18, 1 jun. 2020.
- PAULING, L.; COREY, R. B. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds. **Proceedings of the National Academy of Sciences of the United States of America**, v. 37, n. 11, p. 729–740, 1951a.
- PAULING, L.; COREY, R. B.; BRANSON, H. R. The Structure of Proteins. **Proceedings of the National Academy of Sciences of the United States of America**, v. 37, n. 4, p. 205–211, 1951b.
- PDB-101: **Educational resources supporting molecular explorations through biology and medicine**. Christine Zardecki, Shuchismita Dutta, David S. Goodsell, Robert Lowe, Maria Voigt, Stephen K. Burley. (2022) *Protein Science* 31: 129-140 <https://doi.org/10.1002/pro.4200>

PETTERSEN, E. et al. UCSF Chimera--a visualization system for exploratory research and analysis. **J Comput Chem.** 2004 Oct;25(13):1605-12. doi: 10.1002/jcc.20084. PMID: 15264254.

PITERA, J. W. Expected Distributions of Root-Mean-Square Positional Deviations in Proteins. **The Journal of Physical Chemistry B**, [S. l.], v. 118, n. 24, p. 6526-6530, 2014.

POTDAR, V. et al. Genomic analysis of SARS-CoV-2 strains among Indians returning from Italy, Iran & China, & Italian tourists in India. **The Indian Journal of Medical Research**, v. 151, n. 2-3, p. 255-260, 2020.

RAVIKANTH, V. et al. A variant in TMPRSS2 is associated with decreased disease severity in COVID-19. **Meta Gene**, v. 29, p. 100930, 1 set. 2021.

RAY, D.; LE, L.; ANDRICIOAEI, I. Distant residues modulate conformational opening in SARS-CoV-2 spike protein. **Proceedings of the National Academy of Sciences of the United States of America**, v. 118, n. 43, p. e2100943118, 26 out. 2021.

RODRIGUEZ, J. A. et al. Computational modeling of the effect of five mutations on the structure of the ACE2 receptor and their correlation with infectivity and virulence of some emerged variants of SARS-CoV-2 suggests mechanisms of binding affinity dysregulation. **Chemico-Biological Interactions**. Vol. 368, 2022.

SAKKIAH, S. et al. Elucidating Interactions Between SARS-CoV-2 Trimeric Spike Protein and ACE2 Using Homology Modeling and Molecular Dynamics Simulations. **Frontiers in Chemistry**, v. 8, p. 622632, 5 jan. 2021.

SALLEH, M. Z.; DERIS, Z. Z. In Silico Molecular Characterization of Human TMPRSS2 Protease Polymorphic Variants and Associated SARS-CoV-2 Susceptibility. **Life**, v. 12, n. 2, p. 231, fev. 2022.

SARKER, J. et al. A Review on Expression, Pathological Roles, and Inhibition of TMPRSS2, the Serine Protease Responsible for SARS-CoV-2 Spike Protein Activation. **Scientifica**, v. 2021, p. 2706789, 24 jul. 2021.

SCHÖNFELDER, K. et al. Transmembrane serine protease 2 Polymorphisms and Susceptibility to Severe Acute Respiratory Syndrome Coronavirus Type 2 Infection: A German Case-Control Study. **Frontiers in Genetics**, v. 12, p. 667231, 21 abr. 2021.

SECOLIN, R. et al. Genetic variability in COVID-19-related genes in the Brazilian population. **Human Genome Variation**, v. 8, p. 15, 2 abr. 2021.

SENAPATI, S. et al. Assessment of risk conferred by coding and regulatory variations of TMPRSS2 and CD26 in susceptibility to SARS-CoV-2 infection in human. **Journal of Genetics**, v. 99, n. 1, p. 53, 2020.

SENAPATI, S. et al. Contributions of human ACE2 and TMPRSS2 in determining host-pathogen interaction of COVID-19. **Journal of Genetics**, v. 100, n. 1, p. 12, 2021.

SEYEDALINAGHI, S. et al. Genetic susceptibility of COVID-19: a systematic review of current evidence. **European Journal of Medical Research**, v. 26, p. 46, 20 maio 2021.

SHABIR, O. "What is TMPRSS2?". **News-Medical**. <https://www.news-medical.net/health/What-is-TMPRSS2.aspx> (accessed November 21, 2022).

SHARMA, S. et al. ACE2 Homo-dimerization, Human Genomic variants and Interaction of Host Proteins Explain High Population Specific Differences in Outcomes of COVID19. **bioRxiv**, 19 maio 2020.

SHEN, L. W. et al. TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. **Biochimie**, v. 142, p. 1–10, 1 nov. 2017.

SHICHKIN, A.; BUEVICH, A.; SERGEEY, A. Comparison of artificial neural network, random forest and random perceptron forest for forecasting the spatial impurity distribution. **AIP Conf. Proc.** 27 July 2018; 1982 (1): 020005. <https://doi.org/10.1063/1.5045411>

SHULLA, A. et al. A Transmembrane Serine Protease Is Linked to the Severe Acute Respiratory Syndrome Coronavirus Receptor and Activates Virus Entry. **Journal of Virology**, v. 85, n. 2, p. 873–882, jan. 2011.

SINGH, H. et al. ACE2 and TMPRSS2 polymorphisms in various diseases with special reference to its impact on COVID-19 disease. **Microbial Pathogenesis**, v. 150, p. 104621, 2021b.

SINGH, J. et al. Structure-Function Analyses of New SARS-CoV-2 Variants B.1.1.7, B.1.351 and B.1.1.28.1: Clinical, Diagnostic, Therapeutic and Public Health Implications. **Viruses**, v. 13, n. 3, p. 439, 2021a.

SIRONI, M. et al. SARS-CoV-2 and COVID-19: A genetic, epidemiological, and evolutionary perspective. **Infection, Genetics and Evolution**, v. 84, p. 104384, out. 2020.

SNEHA, P.; DOSS, G. Molecular Dynamics: New Frontier in Personalized Medicine, Advances in Protein Chemistry and Structural Biology. **Academic Press**, v. 102, p. 181-224, 2016.

SOCHER, E.; CONRAD, M, HEGER, L, et al. Computational decomposition reveals reshaping of the SARS-CoV-2–ACE2 interface among viral variants expressing the N501Y mutation. **J Cell Biochem.** 2021; 122: 1863-1872. <https://doi.org/10.1002/jcb.30142>

STRYER, L. **Biochimistry**. New York, Frerman, p. 1089. 1988.

SURYAMOHAN, K. et al. Human ACE2 receptor polymorphisms and altered susceptibility to SARS-CoV-2. **Communications Biology**, v. 4, p. 475, 12 abr. 2021.

TANG, J. W. et al. Introduction of the South African SARS-CoV-2 variant 501Y.V2 into the UK. **The Journal of Infection**, v. 82, n. 4, p. e8–e10, abr. 2021.

TAO, K. et al. The biological and clinical significance of emerging SARS-CoV-2 variants. **Nature Reviews. Genetics**, v. 22, n. 12, p. 757–773, 2021.

Tracking SARS-CoV-2 variants. Disponível em: <<https://www.who.int/activities/tracking-SARS-CoV-2-variants>>. Acesso em set. 2022.

UZUNIAN, A. Coronavirus SARS-CoV-2 and Covid-19. **J. Bras. Patol. Med. Lab.** 56, 2020. <https://doi.org/10.5935/1676-2444.20200053>

VAN DER SPOEL, D. et al. GROMACS: Fast, flexible, and free. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1701–1718, 2005.

VANKADARI, N. et al. Structure of Human TMPRSS2 in Complex with SARS-CoV-2 Spike Glycoprotein and Implications for Potential Therapeutics. **The Journal of Physical Chemistry Letters**, v. 13, n. 23, p. 5324–5333, 16 jun. 2022.

VARDHAN, S.; SAHOO, S. K. Virtual screening by targeting proteolytic sites of furin and TMPRSS2 to propose potential compounds obstructing the entry of SARS-CoV-2 virus into human host cells. **Journal of Traditional and Complementary Medicine, SARS-CoV-2 and Coronavirus Disease19 (COVID-19)**. v. 12, n. 1, p. 6–15, 1 jan. 2022.

VERDECCHIA, P. et al. The pivotal link between ACE2 deficiency and SARS-CoV-2 infection. **European Journal of Internal Medicine**, v. 76, p. 14–20, jun. 2020.

VERLI, H. Bioinformática: da biologia à flexibilidade molecular. **Sociedade Brasileira de Bioquímica e Biologia Molecular (SBBq)**. São Paulo, 2014.

VERMA, J.; SUBBARAO, N. In silico study on the effect of SARS-CoV-2 RBD hotspot mutant's interaction with ACE2 to understand the binding affinity and stability, **Virology**, Volume 561, 2021, P. 107-116, ISSN 0042-6822, <https://doi.org/10.1016/j.virol.2021.06.009>.

VINAY KUMAR, C. et al. Computational Analysis Reveals the Association of Threonine 118 Methionine Mutation in PMP22 Resulting in CMT-1A. **Advances in bioinformatics**, v. 2014, p. 502618, 20 out. 2014.

WALLS, A. C. et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. **Cell**, v. 181, n. 2, p. 281- 292.e6, 16 abr. 2020.

WANG, J. et al. Molecular simulation of SARS-CoV-2 spike protein binding to pangolin ACE2 or human ACE2 natural variants reveals altered susceptibility to infection. **The Journal of General Virology**, v. 101, n. 9, p. 921–924, set. 2020.

WANG, Q. et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. **Cell**, v. 181, n. 4, p. 894- 904.e9, 14 maio 2020.

WANG, W. et al. E484K mutation in SARS-CoV-2 RBD enhances binding affinity with hAce2 but reduces interactions with neutralizing antibodies and nanobodies: Binding free energy calculation studies. **J Mol Graph Model**, vol. 109(108035), pp. 1-15, Dec. 2021, doi: 10.1016/j.jmglm.2021.108035

WANG, W. et al. Identification of key mutations responsible for the enhancement of receptor-binding affinity and immune escape of SARS-CoV-2 Omicron variant. **Journal of Molecular Graphics and Modelling**, Volume 124, 2023, 108540, ISSN 1093-3263, <https://doi.org/10.1016/j.jmgm.2023.108540>.

SYED, W. et al. Modelagem molecular da TMPRSS2. Manuscrito submetido.

WORLD HEALTH ORGANIZATION. **WHO Director-General's Opening Remarks at the Media Briefing on COVID-19**. 11 mar. 2020. [Online]. Available at: <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid19--11-march-2020> (Acesso: 02 de outubro 2023).

WRAPP, D. et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. **Science (New York, N.y.)**, v. 367, n. 6483, p. 1260–1263, 13 mar. 2020.

XIA, S. et al. Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. **Cellular and Molecular Immunology**, v. 17, n. 7, p. 765–767, 2020a.

XIA, S. et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. **Cell Research**, v. 30, n. 4, p. 343–355, 2020b.

YAN, R. et al. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. **Science (New York, N.y.)**, v. 367, n. 6485, p. 1444–1448, 27 mar. 2020.

ZHANG, W. et al. Structural basis for mouse receptor recognition by SARS-CoV-2 omicron variant. **PNAS**, vol. 119(44), pp. e2206509119, Oct. 2022. doi: <https://doi.org/10.1073/pnas.2206509119>

ZHANG, Z.; ZHANG, J.; WANG, J. Surface charge changes in spike RBD mutations of SARS-CoV-2 and its variant strains alter the virus evasiveness via HSPGs: A review and mechanistic hypothesis, **Frontiers in Public Health**, v. 10, 2022. ISSN=2296-2565. doi: [10.3389/fpubh.2022.952916](https://doi.org/10.3389/fpubh.2022.952916)

ZEPEDA-CERVANTES, J. et al. Implications of the Immune Polymorphisms of the Host and the Genetic Variability of SARS-CoV-2 in the Development of COVID-19. **Viruses**, v. 14, n. 1, p. 94, jan. 2022.

ZUNDERT, G. et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. **Journal of molecular biology**, v. 428, 27 set. 2015.