

UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL  
PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIA DOS MATERIAIS

JÉSSICA DE MORAES CARRETONE

**EFEITOS DAS MUTAÇÕES E484A, E484K e E484Q NA PROTEÍNA SPIKE DO  
Sars-Cov-2: UM ESTUDO *in silico***

CAMPO GRANDE-MS

2022

JÉSSICA DE MORAES CARRETONE

**EFEITOS DAS MUTAÇÕES E484A, E484K e E484Q NA PROTEÍNA SPIKE DO  
Sars-Cov-2: UM ESTUDO *in silico***

Dissertação apresentada no curso de mestrado em  
Ciências dos Materiais do programa de pós-  
graduação Ciência dos Materiais - PPGCM do  
Instituto de Física da Universidade Federal do Mato  
Grosso do Sul

Orientador: Prof. Dr. Marcos Serrou do Amaral

CAMPO GRANDE – MS

2022

## Sumário

1. INTRODUÇÃO .....	13
1.1. Covid-19 e SARS-CoV-2.....	13
1.2. Genoma Viral.....	15
1.3. Proteína Spike .....	16
1.4. Mutações e Variantes.....	19
2. OBJETIVOS .....	21
3. METODOLOGIA.....	21
3.1. Dinâmica Molecular .....	21
3.2. AMBER.....	21
3.3. Obtenção da estrutura proteica experimental .....	22
3.4. Mutações.....	23
3.5. Refinamento do sistema .....	24
3.6. Simulação de Dinâmica Molecular .....	24
4. RESULTADOS e DISCUSSÃO .....	25
4.1. Interação com receptor ACE-2 .....	33
5. CONCLUSÃO .....	40
6. REFERÊNCIAS.....	41

## **AGRADECIMENTOS**

Primeiramente a meus pais, sem os quais nunca teria chegado até aqui. Por sempre apoiarem um sonho infantil, mesmo não sabendo muito bem do que se tratava. Ao meu namorado, Stéphan, por nunca me deixar desistir e acreditar em mim, mesmo quando eu mesma estava em dúvida. À minha afilhada Hella Sophia, que, inconscientemente, me motivou continuar no caminho que eu sempre quis trilhar.

Aos colegas e parceiros do grupo LAB2M, por serem solícitos o tempo todo, pela amizade e conversas. Ao técnico do laboratório, Rodrigo, pelo suporte incrível no período remoto, me atendendo até fora do horário de trabalho. À UFMS, minha segunda casa desde a graduação.

E por último, mas de maneira nenhuma menos importante, ao meu orientador Marcos Serrou do Amaral por ter me acolhido no grupo, mesmo sendo de outro curso e o começo da jornada ter sido com outro orientador. Obrigada pela paciência e por todos os ensinamentos. Guardarei esse período com muito carinho ao longo de toda minha carreira acadêmica.

Enfim, registro meu obrigada a todos que de uma forma ou de outra me incentivaram a chegar até aqui.

## RESUMO

CARRETONE, Jéssica. Efeitos das mutações E484A, E484K e E484Q na proteína Spike do Sars-Cov-2: Um estudo *in silico*. Campo Grande, 2022. Dissertação (Mestrado em Ciência dos Materiais) – Programa de Pós-Graduação em Ciência dos Materiais, Instituto de Física, Universidade Federal de Mato Grosso do Sul, Campo Grande, 2022.

A COVID-19, causada pelo vírus SARS-CoV-2 se tornou uma pandemia em março de 2020 e desde então tem causado inúmeros prejuízos, tanto materiais quanto em vidas. Como todos os vírus de RNA, possui alta capacidade de mutação, principalmente na proteína Spike. O objetivo é analisar os efeitos das mutações E484A, E484K e E484Q, em comparação com a estrutura nativa, através de Dinâmica Molecular. Os dados de RMSD e RMSF mostraram que a variante E484K foi a que obteve maior variação entre todas as variantes e a nativa. As variantes E484A e E484Q mostraram comportamento semelhante à nativa. A região do laço que se liga ao receptor ACE-2 mostrou menores valores de RMSD e RMSF em todas as variantes e a nativa.

Palavras-chave: SARS-CoV-2, COVID-19, Spike

## ABSTRACT

CARRETONE, Jéssica. Effects of E484A, E484K and E484Q mutations on the SARS-CoV-2 Spike protein: An *in silico* study. Campo Grande, 2022. Dissertação (Mestrado em Ciência dos Materiais) – Programa de Pós-Graduação em Ciência dos Materiais, Instituto de Física, Universidade Federal de Mato Grosso do Sul, Campo Grande, 2022.

COVID-19, caused by the SARS-CoV-2 virus, became a pandemic in March 2020 and since then has caused numerous damages, both material, and lives. Like all RNA viruses, it has a high capacity for mutation, mainly in the Spike protein. The objective is to analyze the effects of mutations E484A, E484K and E484Q, in comparison with the native structure, through Molecular Dynamics. The RMSD and RMSF data showed that the E484K variant had the greatest variation between all variants and the native one. The E484A and E484Q variants showed similar behavior to the native one. The loop region that binds to the ACE-2 receptor showed lower values of RMSD and RMSF in all variants and the native one.

*“Todas as vitórias ocultam uma abdicação”*

Simone de Beauvoir

## LISTA DE FIGURAS

Figura 1 – Cladograma da família Nidovirales. Da esquerda para a direita: ordem, subordem, família, subfamília, gênero e espécie.....	13
Figura 2 – Vírus Sars-Cov-2, causador da doença Covid-19. Imagem de CDC/Unsplash ...	15
Figura 3 – Componentes do SARS-CoV-2 (+ssRNA: do inglês, single-stranded positive-sense RNA). Adaptado da referência [16].....	16
Figura 4 - A estrutura da proteína spike (S) do SARS-CoV-2. Adaptado da referência [16]	17
Figura 5 – Estrutura em 3D do receptor ACE2 ligado à proteína Spike do Sars-Cov-2. Disponível em Protein Data Bank sob o código 7T9L. ....	18
Figura 6 – Interação entre o Sars-Cov-2 e o receptor ACE-2.....	19
Figura 7 – Informações básicas sobre o software AMBER. Disponível em Manual do Amber .....	22
Figura 8 – RBD da Spike, com o laço 9 em vermelho, as alfas hélices em verde e as folhas beta em amarelo.....	23
Figura 9 – RBD da Spike, com o laço 455-491 em vermelho, as alfas hélices em verde e as folhas beta em amarelo e o resíduo 484 destacado de roxo. ....	24
Figura 10 RMSD de todos os resíduos. A) Refere-se ao sistema Nativa, B) refere-se ao sistema E484K (Gama), C) refere-se a E484Q (Delta) e D) ao sistema E484A (Ômicron)...	26
Figura 11 – RMSD de todos os resíduos, com o sistema Nativa representado em preto, E484K (Gama) em vermelho, E484Q (Delta) em verde e E484A (Ômicron) em azul.....	27
Figura 12 – RMSF de todos os resíduos de todos os sistemas. Os laços estão destacados em marrom, as alfa hélices em preto e folha beta em alaranjado. A) refere-se ao sistema Nativa, B) E484K (Gama), C) E484Q (Delta) e D) ao sistema (E484A Ômicron). ....	29
Figura 13 – RMSF de todos os resíduos, com os sistemas Nativa representado em preto, E484K (Gama) representado em vermelho, E484Q (Delta) em verde e E484A (Ômicron) em azul.....	30
Figura 14 – Raio de Giro. A) representa o sistema Nativa, B) o sistema E484K (Gama), C) o sistema E484Q (Delta) e D) E484A (Ômicron).....	32
Figura 15 – Figura 15 – RMSD dos resíduos do laço 9. A) representa o sistema Nativa, B) E484K (Gama), C) E484Q (Delta) e E484A (Ômicron). ....	34
Figura 16 – RMSD dos resíduos 455-491, com os sistemas Nativa, representado em preto, E484K (Gama), representado em vermelho, E484Q (Delta), em verde e E484A (Ômicron) em azul.....	36



Figura 17 – RMSF dos resíduos 455-491, onde A) representa o sistema Nativa, B) E484K (Gama), C) E484Q (Delta) e D) E484A (Ômicron). ..... 37

Figura 18 – RMSF do laço 9, com os sistemas Nativa, destacado em preto, E484K (Gama), destacado em vermelho, E484Q (Delta), destacado em verde e E484A (Ômicron) em azul. .... 38

## LISTA DE TABELAS

Tabela 1 – Média aritmética, em nm, de todas as estruturas secundárias e laços de todos os sistemas. ....	31
Tabela 2 – Comparativo entre médias, em nm, de RMSD e RMSF da proteína toda e do laço 9. ....	36

## LISTA DE ABREVIATURAS E SIGLAS

COVID-19 – do inglês, *Coronavirus disease*

OMS – Organização Mundial da Saúde

CoV – Coronavírus

SARS – Síndrome Respiratória Aguda Grave

MERS-CoV – Síndrome Respiratória do Oriente Médio (do inglês, *Middle East Respiratory Syndrome*)

RNA – Ácido Ribonucleico

ORF – Fases de leitura aberta (do inglês, *Open Reading Frame*)

NSP – Proteínas não estruturais (do inglês, *Non structural protein*)

Spike – S

Envelope – E

Membrana – M

Nucleocapsídeo – N

NTD – RNA-terminal

CTD – C-terminal

S1 – Subunidade da Spike de ligação ao receptor

S2 – Subunidade da Spike responsável por fusão de membrana

S1-NTD – Domínio N-terminal da S1

S1-CTD – Domínio C-terminal da S1

ACE2 – Enzima conversora da angiotensina 2 (ECA2) (do inglês, *angiotensin-converting enzyme 2*)

DPP4 – Dipeptil peptidase 4

APN – Aminopeptidase

RBD – Domínio de ligação ao receptor (do inglês, *Receptor-Biding Domain*)

RAS – Resina-angiotensina-aldosterona

RBM – Motivo de ligação ao receptor (do inglês, *Receptor-Biding Motif*)

VOHC – Variantes de alta consequência (do inglês, *Variants of High Consequence*)

VOC – Variantes de Preocupação (do inglês, *Variants of Concern*)

Amber – (*Assisted Model Building Energy Refinement*)

PDB – Protein Data Bank

E – Ácido Glutâmico

A – Alanina

K – Lisina

Q – Glutamina

NVT – Número de partículas, Volume e Temperatura constantes

NPT – Número de partículas, Pressão e Temperatura constantes

RMSD - Desvio Quadrático Médio (do inglês, *Root Mean Square Deviation*)

RMSF - Flutuação Quadrática Média (do inglês, *Root Mean Square Fluctuation*)

Rg – Raio de giro (do inglês, *Radio of gyration*)

# 1. INTRODUÇÃO

## 1.1. Covid-19 e SARS-CoV-2

A COVID-19 (*Coronavirus disease*, em inglês) é definida como uma doença inflamatória sistêmica desencadeada por infecção viral SARS-CoV2[1]. Em março de 2020 foi declarada como pandemia pela OMS (Organização Mundial da Saúde). Estima-se que mais de 422 milhões de casos foram confirmados em 266 países, em todos os continentes, e mais de 5.8 milhões de mortes. No Brasil, foram 28 milhões de casos positivos e 640.000 mortes até fevereiro de 2022[2].

Os coronavírus (CoV) são vírus pertencentes à família *Coronaviridae*, ordem *Nidovirales* e subordem *Cornidovirineae*. São classificados em quatro gêneros: alfa, beta, gama e delta (Figura 1). Os coronavírus responsáveis pela infecção em humanos e outros mamíferos são dos gêneros alfa e beta. Os gêneros gama e delta infectam apenas aves[3, 4].

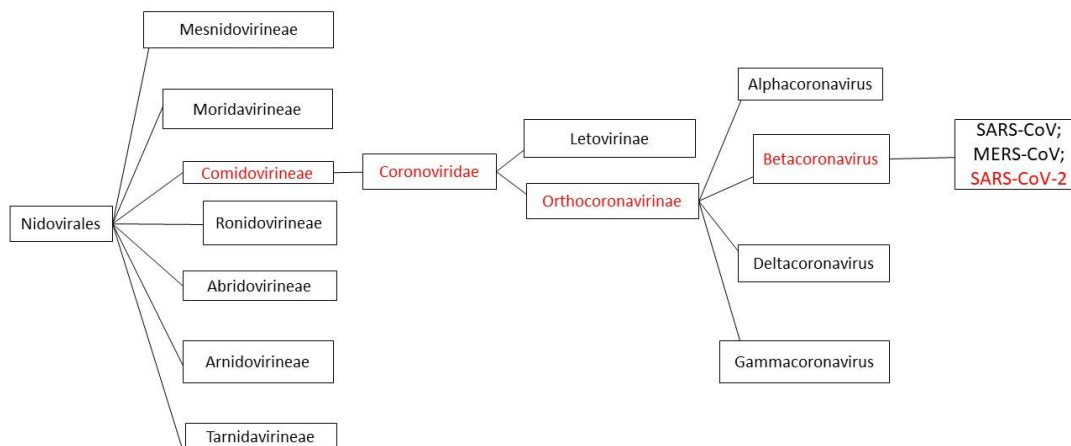


Figura 1 - Cladograma da família Nidovirales. Da esquerda para a direita: ordem, subordem, família, subfamília, gênero e espécie

Os primeiros relatos relacionando os CoV com síndromes respiratórias em humanos foram na China, no início dos anos 2000. Em 2002 e 2003 foram descritos

mais de 8 mil casos da Síndrome Respiratória Aguda Grave (SARS) com 774 mortes principalmente na Ásia, África do Sul, América Central e Europa[4].

Em meados de 2012 uma nova epidemia de SARS foi relatada, como consequência de outro vírus, diferente do primeiro, conhecido como MERS-CoV. Ficou conhecida como Síndrome Respiratória do Oriente Médio (do inglês *Middle East Respiratory Syndrome*) devido sua origem na Arábia Saudita e outros países do Oriente Médio. Houve 2.494 casos confirmados de MERS e 858 óbitos, mostrando uma letalidade alta, em torno de 35%[5].

Tanto o SARS quanto o MERS-CoV possuem origem zoonótica, ou seja, pela transmissão de origem animal para humanos. A hipótese mais aceita é a de morcegos como hospedeiros primários e posterior infecção em humanos mediada pelo civeta (*Paguma larvata*) para SARS e camelos dromedários para MERS[6-8].

Em 2019 foi identificado um novo tipo de coronavírus em Wuhan, China. Inicialmente nomeado de 2019-nCoV e foi posteriormente batizado como SARS-CoV-2, agente causador da COVID-19[9]. De forma semelhante ao SARS-CoV e MERS-CoV, o SARS-CoV-2 pode causar doença respiratória aguda grave[10].

Quanto a origem do SARS-CoV-2, estudos apontam uma hipótese zoonótica. Possui 96% de similaridade com o genoma do CoV-RaTG13, um CoV que infecta o morcego *Rhinolophus affinis*, com livre comércio no mercado de Wuhan, local onde os primeiros casos de COVID-19 foram descobertos[11-13].

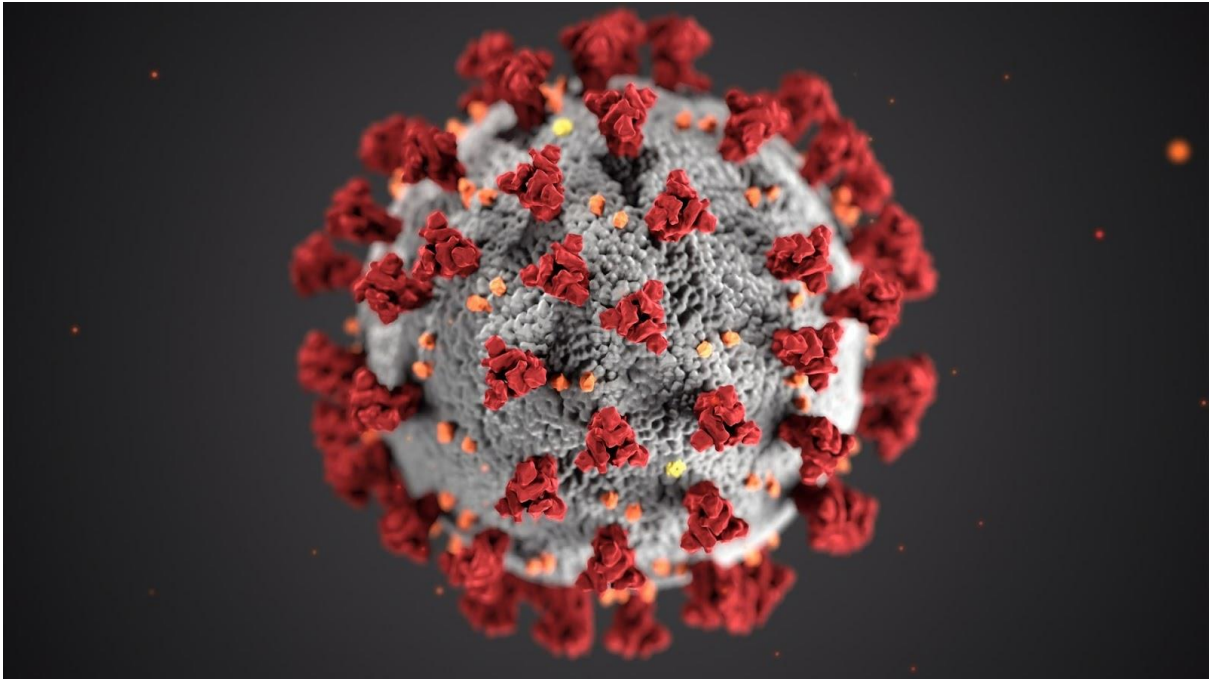


Figura 2 - Vírus Sars-Cov-2, causador da doença Covid-19. Imagem de CDC/Unsplash

## 1.2. Genoma Viral

O SARS-CoV-2, como outros vírus pertencentes à família *Coronaviridae*, apresenta os maiores genomas de ácido ribonucleico (RNA) conhecidos e possuem material genético constituído por RNA de cadeia simples de senso positivo. Possui um genoma com média de 29 mil e novecentas bases nitrogenadas, que são traduzidas para vinte e nove proteínas. Essas proteínas são divididas em: Proteínas não estruturais; Proteínas estruturais e Proteínas acessórias, que se encontram em ORFs (*Open Reading Frame*, ou Fases de leitura aberta). São eles: ORF1a; ORF1b; ORF3; ORF6; ORF7a; ORF7b; ORF8; ORF9b e ORF10[14].

**Proteínas não estruturais:** São proteínas envolvidas na transcrição e replicação do vírus. Ao todo são 16 proteínas cujos genes estão localizados no ORF1ab, na região 5' do RNA. São nomeadas de NSP (Proteínas não estruturais) e numeradas de 1 a 16. Ocupam a maior parte do genoma viral.

**Proteínas acessórias:** Proteínas que compreendem as 3a; 3b; 6; 7a; 7b; 8; 9b; 9c e 10 e embora não tenham participação na replicação viral, são extremamente

importantes para a interação vírus-hospedeiro. Seus genes estão localizados na porção 3' do RNA viral.

**Proteínas estruturais:** Compõe as proteínas Spike (S), Envelope (E), Proteínas de Membrana (M) e Nucleocapsídeo (N). O Nucleocapsídeo é uma proteína reguladora da replicação viral. Possui dois domínios de ligação ao RNA: RNA-terminal (NTD) e C-terminal (CTD). Participa dos processos de tradução e replicação viral e adesão do capsídeo ao envelope. O Envelope e Membrana juntas formam uma membrana lipídica, originada através de brotamento quando o vírus deixa a célula hospedeira. Delimitam o conteúdo viral e apresenta em sua superfície proteínas de membrana[15].

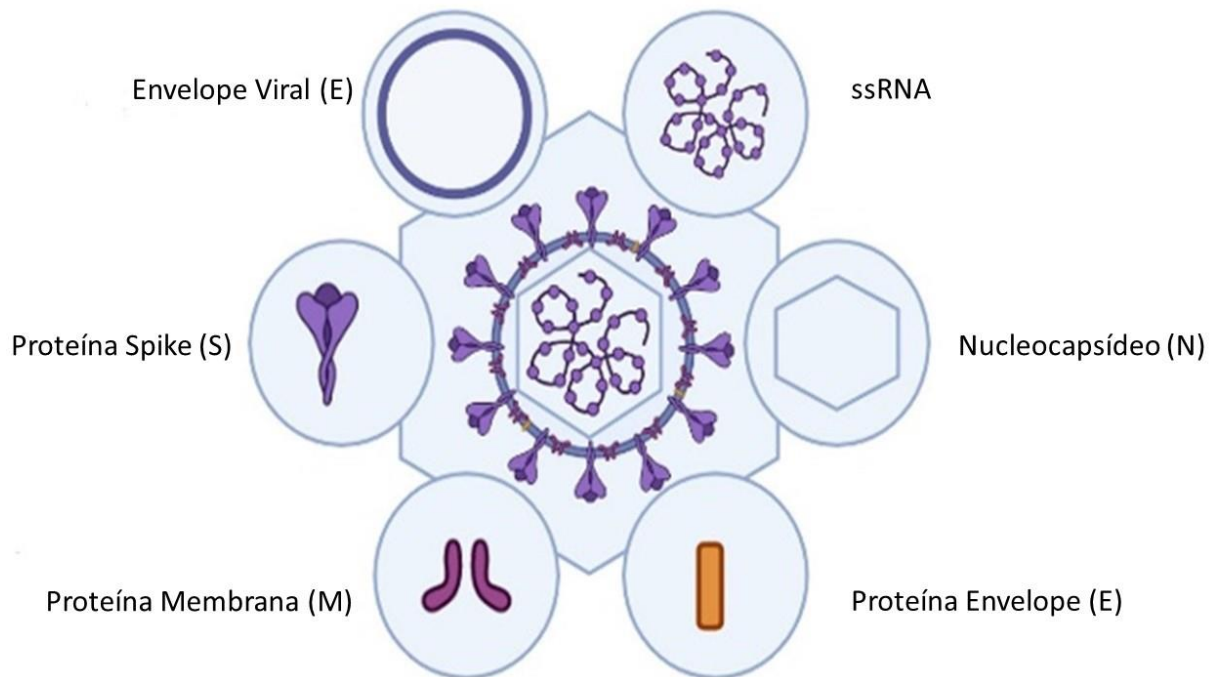


Figura 3 - Componentes do SARS-CoV-2 (+ssRNA: do inglês, single-stranded positive-sense RNA). Adaptado da referência [16]

### 1.3. Proteína Spike

A proteína Spike é uma proteína de membrana, possui formato de espinho, que fornece a característica conformacional em coroa da família *Coronaviridae*. É a responsável pela interação com os receptores da superfície celular do hospedeiro[17]. Está localizada no envelope viral. Consiste em três segmentos, sendo eles: ectodomínio, âncora transmembrana de passagem única e uma cauda intracelular



curta[18]. O ectodomínio é dividido em duas subunidades S1 de ligação ao receptor e uma subunidade S2 de fusão de membrana. A subunidade S1 possui dois domínios, um domínio N-terminal (S1-NTD) e um domínio C-terminal (S1-CTD). CTD liga-se a diferentes receptores, como ACE-2, dipeptil peptidase 4 (DPP4) e aminopeptidase N (APN). Na subunidade S2 acontece fusão da membrana[19, 20].

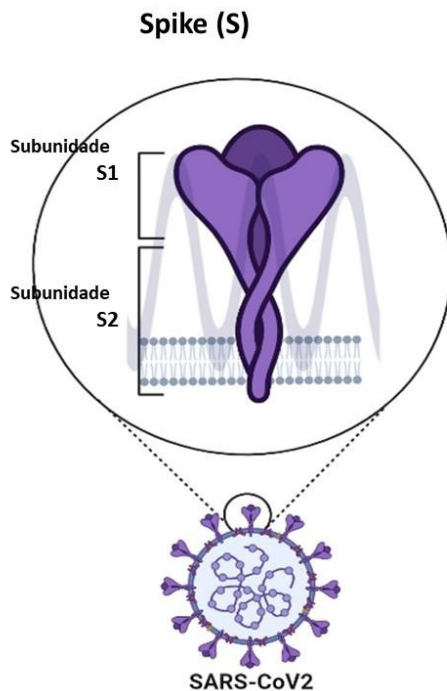


Figura 4 - A estrutura da proteína spike (S) do SARS-CoV-2. Adaptado da referência [16]

O domínio de ligação ao receptor (RBD), contido na região S1 da Spike, interage com o receptor celular ECA-2 (do inglês *aniotensin-converting enzyme 2*, ACE-2) uma enzima do sistema renina-angiotensina-aldosterona (RAS). Está localizado na superfície da célula alvo. A infecção viral inicia-se com a penetração das partículas virais no citosol[21, 22]. Ocorre clivagem do RBD da proteína S por protease furina e depois a interação com ACE-2, a porta de entrada do vírus para as células humanas, onde irá ocorrer posterior tradução e transcrição do RNA viral, completando assim seu ciclo de vida.



*Figura 5 - Estrutura em 3D do receptor ACE2 ligado à proteína Spike do Sars-Cov-2. Disponível em Protein Data Bank sob o código 7T9L.*

A região S1 é uma parte da Spike que fica mais externa em relação ao interior do vírus. Se compararmos a proteína S com um prego, a subunidade S1 seria a base do prego, enquanto a S2 seria a ponta, inserido no envelope viral. O RBD é um fragmento da proteína S que possui 195 resíduos (333-527) e uma região conhecida como motivo de ligação ao receptor (RBM), onde se encontra os resíduos ativos de ligação ao receptor. Esse RBM compreende os resíduos 424-494.

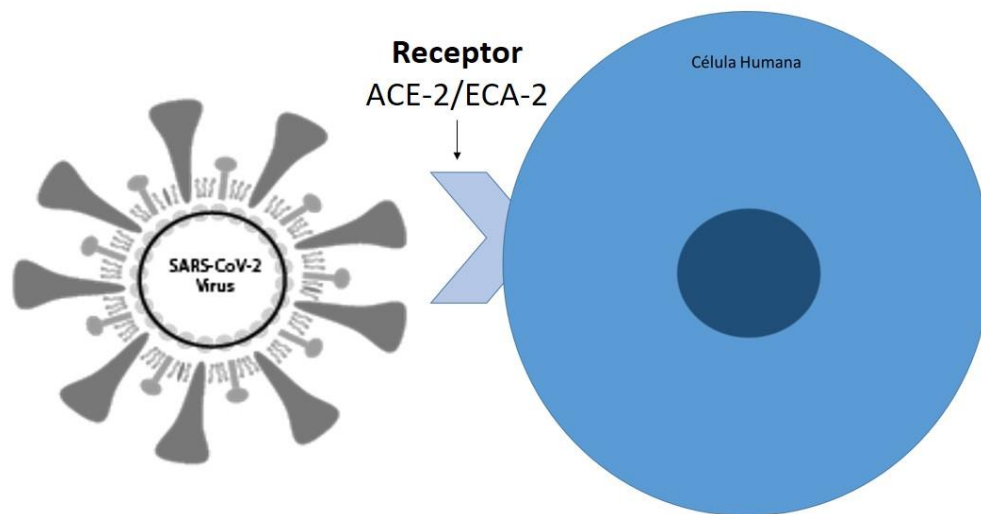


Figura 6 - Interação entre o Sars-Cov-2 e o receptor ACE-2

#### 1.4. Mutações e Variantes

Vírus de RNA possuem alta capacidade para sofrer mutações e se adaptar a seus hospedeiros. Com o SARS-CoV-2 não é diferente[23]. Existem milhares de variações no genoma do SARS-CoV-2 nativo, mas a maioria são silenciosas, ou seja, não alteram estruturas de proteínas[24]. No início da pandemia, as variantes recebiam o nome da região onde haviam sido descobertas. Mas para evitar discriminação e notificações erradas, a OMS decidiu alterar a nomenclatura para letras do alfabeto grego[25].

As variantes foram divididas em três categorias;

Variantes de alta consequência – VOHC (*variants of high consequence*) são variantes que tem maior escape a imunizantes e podem ocasionar infecções mais graves. Até o momento, não foi identificada nenhuma VOHC.

Variantes de preocupação – VOC (*variants of concern*) são mais transmissíveis e provocam infecções mais graves. São elas:

Alfa (B.1.17) Originária do Reino Unido, é considerada 43% mais perigosa que a nativa SARS-CoV-2. A incidência da alfa corresponde hoje a 2,07%.

Beta (B.1.351) foi identificada pela primeira vez na África do Sul. Bastante transmissível, quando surgiu a incidência ficava em torno de 3%, mas hoje corresponde a menos de 0,1% dos casos.

Gama (P.1) identificada em Manaus, Amazonas, tornou-se a principal variante no Brasil em poucos meses. A incidência em São Paulo fica entre 91% e 85% do total de amostras. Possui a mutação E484K.

Delta (B.1.617.2) identificada na Índia, já está em 163 países, com incidência de 23% dos casos quando surgiu. Possui a mutação E484Q.

Ômicron (B.1.1.529) identificada na África, hoje é a variante mais dominante. O boletim epidemiológico da Rede de Alerta das Variantes do SARS-CoV-2 mostrou que entre os dias 5 e 12 de fevereiro de 2022 a variante Ômicron foi responsável por 100% das amostras positivas sequenciadas em São Paulo[26]. Possui a mutação E484A.

A pandemia do Covid-19 já atingiu mais de 210 países, com aproximadamente 200 milhões de casos confirmados e quase 6 milhões de mortos, gerando diversos prejuízos na saúde física e psicológica, forçando bilhões de pessoas a mudarem suas rotinas, fazendo com que os governos desenvolvam estratégias para conter tanto o avanço do vírus como danos econômicos[2]. Devido a isso, muitos estudos foram conduzidos a fim de compreender melhor origem, estrutura viral e suas subunidades, bem como o desenvolvimento de vacinas e fármacos que possam ajudar no tratamento.

Apesar do avanço em pesquisas, ainda há muito que se descobrir acerca do SARS-CoV-2. Em geral, os vírus de RNA são vulneráveis a altas taxas de mutações, que pode ser correlacionado com a virulência específica da região geográfica da variante do vírus[27]. Como o cenário atual traz o surgimento de mutações do vírus SARS-CoV-2[28] que ameaçam a eficácia das vacinas disponíveis para a doença, torna-se relevante verificar se essas mudanças no material genético viral afetam o funcionamento da estrutura RBD, assim como suas relações com o ambiente em que atua.

## 2. OBJETIVOS

Analisar, por dinâmica molecular computacional, os efeitos estruturais das mutações E484A, E484K e E484Q, responsáveis pelas variantes Ômicron, Gama e Delta, respectivamente.

## 3. METODOLOGIA

### 3.1. Dinâmica Molecular

Simulação de Dinâmica Molecular é uma versátil técnica para o estudo de moléculas biológicas.[29] É um método de baixo custo, que fornece resultados de qualidade, permitindo assim estudos úteis em diversas áreas, como farmacoterapia e biotecnologia. Consiste em simular o comportamento molecular em função do tempo, através dos movimentos dos átomos individuais em uma molécula. Por não ser estática, essa técnica se aproxima muito de métodos experimentais.

### 3.2. AMBER

Amber (*Assisted Model Building Energy Refinement*) é um conjunto de programas que permite ao usuário realizar dinâmica molecular. Conta com um robusto conjunto de ferramentas que permitem parametrizar a caixa de simulação, como o *tleap*, *parmed* para manipulação do arquivo de topologia do sistema, *pdb4amber*, utilizado para reparos no sistema, entre outros.

Os principais softwares utilizados foram

*ANTECHAMBER*: ferramenta para preparação de arquivos de entrada;

*LEaP*: Cria e modifica sistemas

*SANDER*: Realiza as minimizações de energia, através dos algoritmos *steepest-descent* e *gradiente-conjugate*

*CPPTRAJ*: análise e processamento de trajetória.

*CPINUTIL.PY*: Cria uma ambientação de pH constante.

*PDB4AMBER*: Prepara os arquivos PDB para posterior uso no LEaP

*PMEMD.CUDA*: Acelera o processo de simulação através do uso de uma GPU

*CHIMERA*: Utilizado para visualização e manipulações de estruturas moleculares

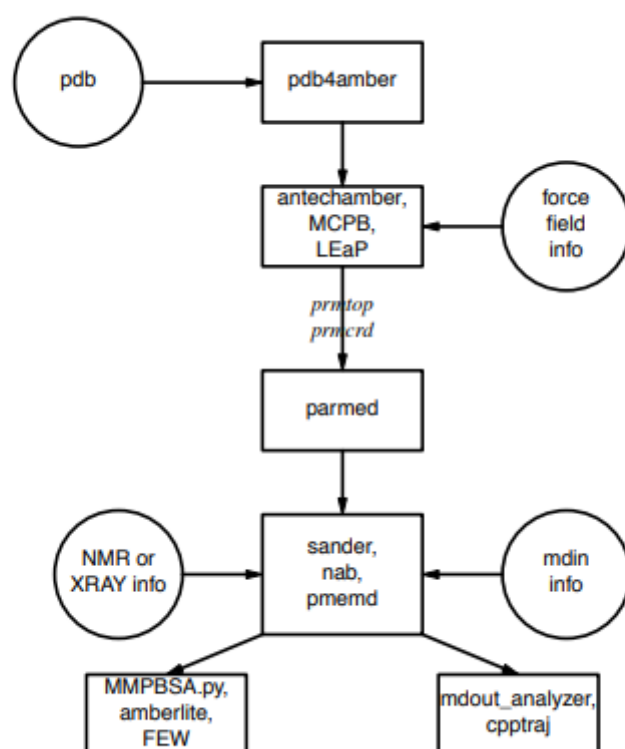
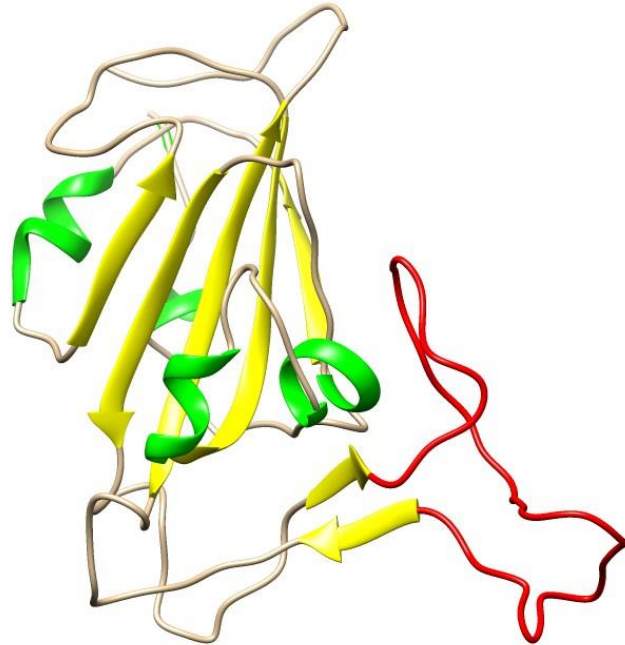


Figura 7 - Informações básicas sobre o software AMBER. Disponível em Manual do Amber

### 3.3. Obtenção da estrutura proteica experimental

A estrutura da RBD nativa foi obtida no banco de dados online Protein Data Bank (PDB) com o código de identificação 6VW1, com resolução de 2.68 Å e obtida através de Difração de Raio-X. A estrutura está complexada com o receptor ACE2,

portanto para ficar apenas o pdb do RBD a cadeia do receptor foi cortada através do software *PyMol*.



*Figura 8 - RBD da Spike, com o laço 9 em vermelho, as alfas hélices em verde e as folhas beta em amarelo.*

### **3.4. Mutações**

Em posse do pdb do RBD, foram realizadas substituições de aminoácidos para cada variante no software Chimera. As substituições foram:

Ácido Glutâmico (E) por Alanina (A) para a variante Ômicron;

Ácido Glutâmico (E) por Lisina (K) para a variante Gama;

Ácido Glutâmico (E) por Glutamina (Q) para a variante Delta .

Os dados de mutações foram obtidos pelo GenBank.

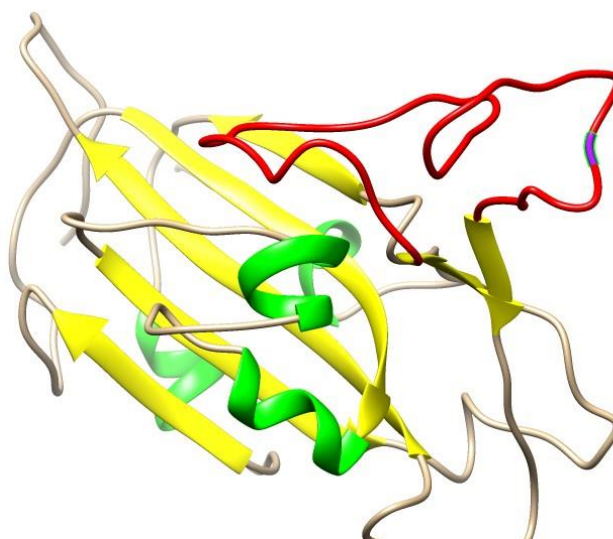


Figura 9 - RBD da Spike, com o laço 455-491 em vermelho, as alfas hélices em verde e as folhas beta em amarelo e o resíduo 484 destacado de roxo.

### 3.5. Refinamento do sistema

A estrutura obtida (PDB ID, 6VW1) foi submetida ao servidor H++ para correções de hidrogênio de acordo com o pH especificado do sistema. Os estados de ionização (protonação) são extremamente importantes para a função e estrutura das macromoléculas. Simulações geralmente requerem a especificação do estado de protonação dos grupos tituláveis. Cristalografia de raios-X normalmente não fornecem posições dos átomos de hidrogênio e a lacuna precisa ser preenchida por métodos teóricos que preveem estados de protonação (pK) de grupos ionizáveis dentro da macromolécula baseada em sua estrutura de resolução atômica. O H++ fornece um conjunto de ferramentas que automatizam este processo.[30]

### 3.6. Simulação de Dinâmica Molecular

O pacote do AmberTools19 do Linux foi usado para realizar a dinâmica molecular, tanto da proteína nativa quanto as variantes selecionadas. Para descrever a estrutura e dinâmica da proteína corretamente, os campos de força ff14SB e constph



foam escolhidos[31]. As estruturas foram solvatadas em caixa octaédrica com água explícita TIP3P[32], e íons de Na<sup>+</sup>Cl<sup>-</sup> foram adicionados para neutralizar o sistema. Para a minimização de energia foram usados os métodos *Conjugate gradiente* e em seguida o *Steepest-descent*, com a finalidade de eliminar maus contatos nas estruturas. Para a etapa de aquecimento a estrutura foi aquecida de 0 a 310K. Para a equilibração foram estabelecidas as condições NVT (número de partícula, volume e temperatura constantes) com temperatura de 310K e as condições NPT (número de partículas, pressão e temperatura constantes) foram estabelecidas a 1 atm, com a mesma temperatura. As simulações então foram realizadas a 310K, com pH a 7,4 para mimetizar as condições fisiológicas do organismo humano. Todas as simulações foram feitas por 200ns.

Para estudar as diferenças estruturais entre a proteína nativa e as variantes, cada uma foi submetida a análises de estabilidade e flexibilidade obtidas pelos valores do RMSD (do inglês, *Root Mean Square deviation*), Raio de Giro e RMSF (do inglês, *Root Mean Fluctuation*). A partir dos resultados obtidos foram criados gráficos pela ferramenta XMGrace.

#### **4. RESULTADOS e DISCUSSÃO**

Os dados seguintes referem-se a todos os resíduos da estrutura RBD, que no PDB original compreende os resíduos 333 até o 527.

O RMSD mede a distância média entre átomos de estruturas a partir de uma estrutura de referência, desse modo é comum utilizar para comparação de similaridade. Os valores de RMSD também podem trazer informações sobre o estado de equilíbrio da estrutura. Como os valores são dados em função do tempo, essa análise permite observar o tempo que as estruturas levam para se estabilizar espacialmente. Para esses resultados, a diferença foi medida entre os átomos do *backbone* da estrutura de partida e cada estrutura obtida ao longo da trajetória da simulação.

Todos os sistemas foram submetidos a variação de temperatura de 0 a 310K e pH constante a 7,4.

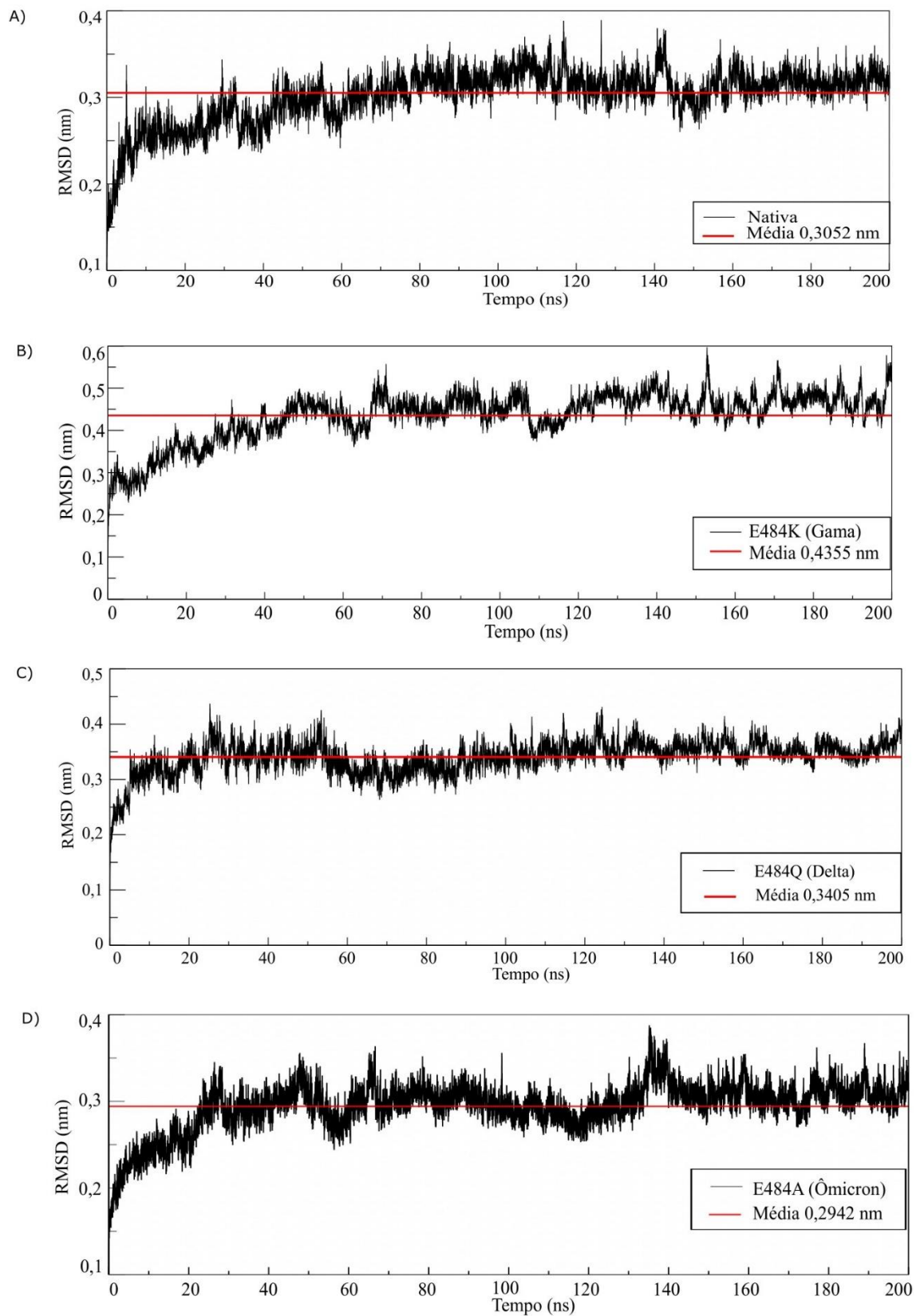


Figura 10 RMSD de todos os resíduos. A) Refere-se ao sistema Nativa, B) refere-se ao sistema E484K (Gama), C) refere-se a E484Q (Delta) e D) ao sistema E484A (Ômicron).

Para o sistema Nativa (Figura 10 A), aos primeiros 20 ns de simulação a proteína apresentou variação de 0,2 nm a 0,3 nm. Na maior parte do tempo a variação ocorre entre 0,3 a 0,4 nm, com um média de 0,3052 nm. Em comparação, o sistema Gama (Figura 10 B), variou entre 0,3 nm e 0,55 nm, durante os primeiros 60 nanossegundos de simulação. No restante da simulação o RMSD varia de ,045 nm a 0,60 nm, com uma média de 0,4355 nm, com vários picos no decorrer da simulação toda.

O sistema Delta (Figura 10 C) nos primeiros 50 nanossegundos o RMSD oscila de 0,2 a 0,4 nm. A partir dos 60 nanossegundos até o final da simulação há uma estabilidade em 0,35 nm, com uma média de 0,3405 nm. Já o sistema Ômicron (Figura 10 D) apresenta valores de RMSD relativamente menores aos encontrados no sistema Nativa. Durante quase toda a simulação os valores estão em torno de 0,3 nm a 0,35 nm, com um pico em 140 nanossegundos de 0,4 nm. A média para esse sistema é de 0,2942 nm, a menor média entre todos os sistemas.

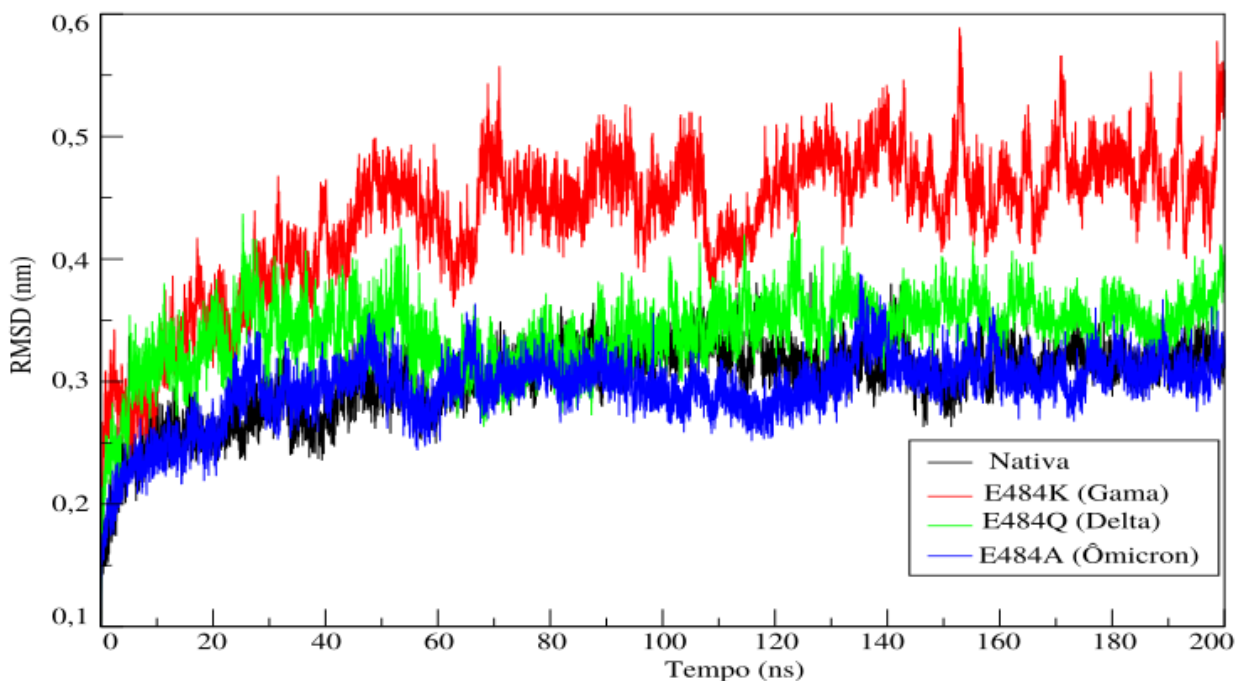


Figura 11 - RMSD de todos os resíduos, com o sistema Nativa representado em preto, E484K (Gama) em vermelho, E484Q (Delta) em verde e E484A (Ômicron) em azul.

Os sistemas Nativa e Ômicron, preto e azul respectivamente, foram os sistemas que apresentaram menores valores de RMSD no geral, conforme indica a Figura 11. A proteína Nativa tem como resíduo 484 o Ácido Glutâmico, estrutura primária que possui natureza ácida. Isso propicia a formação de ligações de hidrogênio, o que pode explicar a sua maior estabilidade estrutural em comparação com as outras variantes. A variante Ômicron, por outro lado, possui mutação no resíduo 484, onde foi substituído o ácido glutâmico por uma alanina. A alanina é um aminoácido hidrofóbico. A hidrofobicidade propicia formação de ligações covalentes, onde a energia de ligação é menor em comparação com uma ligação de hidrogênio.

A variante E484K, gama, representada em vermelho, possui mutação do tipo substituição entre o ácido glutâmico e a lisina. A lisina é um aminoácido básico, por consequência, há liberação de grupos carboxílicos, o que pode ter ocorrido rompimento das ligações de hidrogênio. Isso poderia explicar porque o sistema apresentou os maiores valores de RMSD.

A variante E484Q, delta, representada em verde (Figura 11), possui uma mutação no resíduo 484, onde foi substituído o ácido glutâmico por uma glutamina. A glutamina possui natureza polar não-carregada.

Os gráficos a seguir referem-se a resultados de RMSF para todos os resíduos de todos os sistemas. O RMSF afere uma variação de resíduos em comparação com uma estrutura de referência. Para esses dados, a diferença foi medida entre os átomos do *backbone* da estrutura de partida e cada resíduo obtido através da trajetória de simulação.

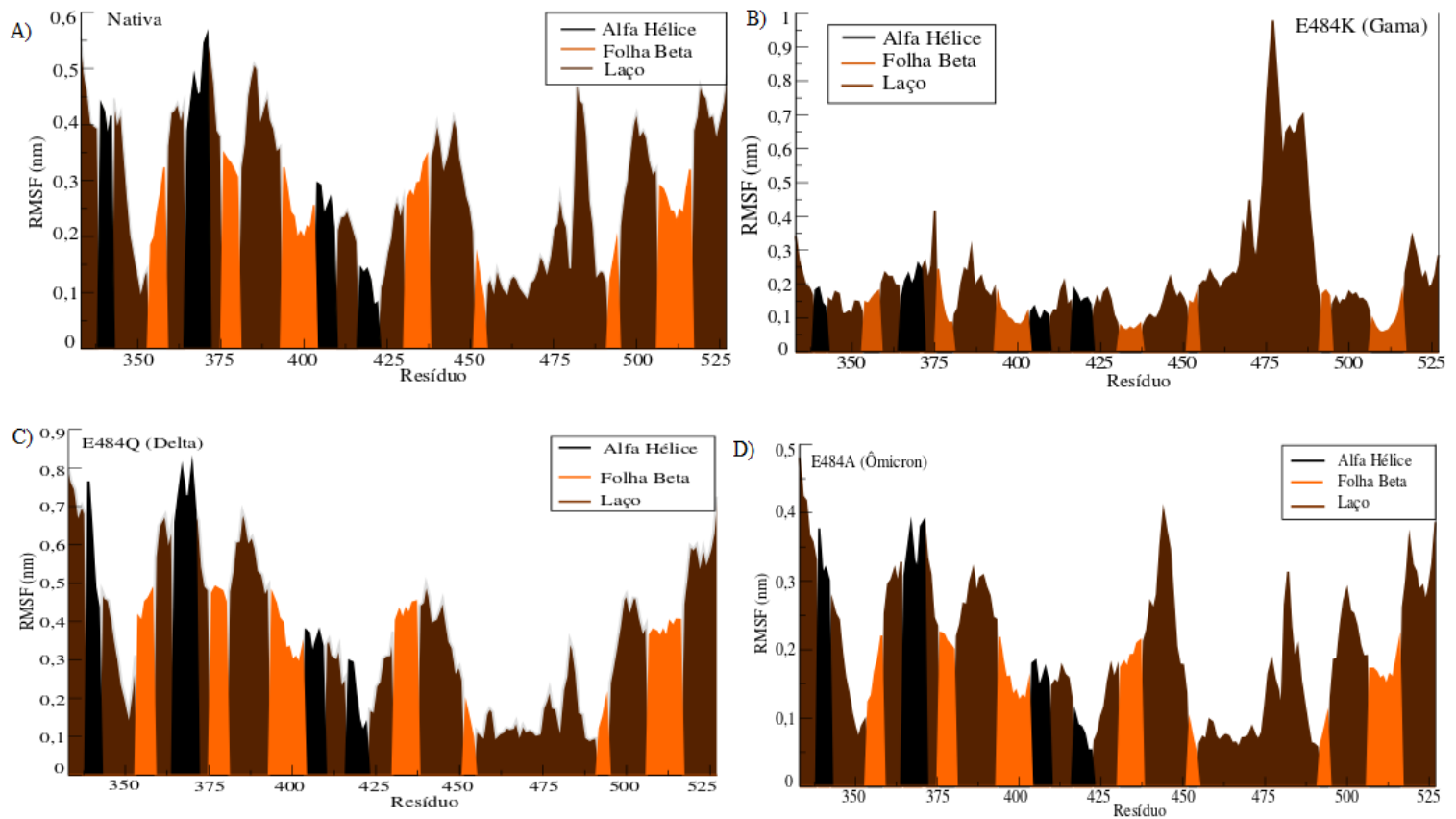


Figura 12 - RMSF de todos os resíduos de todos os sistemas. Os laços estão destacados em marrom, as alfas hélices em preto e folha beta em alaranjado. A) refere-se ao sistema Nativa, B) E484K (Gama), C) E484Q (Delta) e D) ao sistema (E484A Ômicron).

Em todos os sistemas os valores de RMSF são mais altos que os encontrados no RMSD. Para o sistema Nativa (Figura 12 A), os valores variam entre 0,1 para alguns laços e alfa hélice até 0,6 nm. A média total é de 0,2918 nm, enquanto nas alfas hélices é de 0,3168 nm, na folha beta, 0,2323 nm e nos laços 0,3366 nm. Para o sistema Gama, conforme a Figura 12 B), os valores de RMSF são mais altos em comparação ao sistema Nativa. Os picos são mais acentuados em torno de 1 nm, no laço que compreende os resíduos 455 a 491 e os menores valores ficam entre 0,2 e 0,3 nm. A média total é de 0,2817 nm, enquanto nas alfas hélices é 0,2374 nm, nas folhas beta, 0,2552 nm e nos laços, 0,2109 nm.

No sistema Delta (Figura 12 C) os resíduos apresentam grandes variações no RMSF. Os maiores picos compreendem o primeiro laço, em torno de 0,7 nm e nas alfas hélices 1 e 2, entre 0,75 e 0,8 nm. O menor valor está no laço 9, onde estão os resíduos 455 a 491. As médias são: 0,3821 nm para todos os resíduos, 0,4664 nm

para as alfas hélices, 0,3402 nm para as folhas beta e 0,4452 nm para os laços. Para o sistema Ômicron o RMSF apresenta picos em 0,4 nm no laço 1 e 8 e nas alfas hélices 1 e 2. Nas outras estruturas varia entre 0,2 e 0,3, como mostrado na Figura 12 D. As médias são: 0,1982nm para todo o RMSF, 0,1260 nm para as alfas hélices, 0,0604 nm para as folhas beta e 0,1336 nm para os laços.

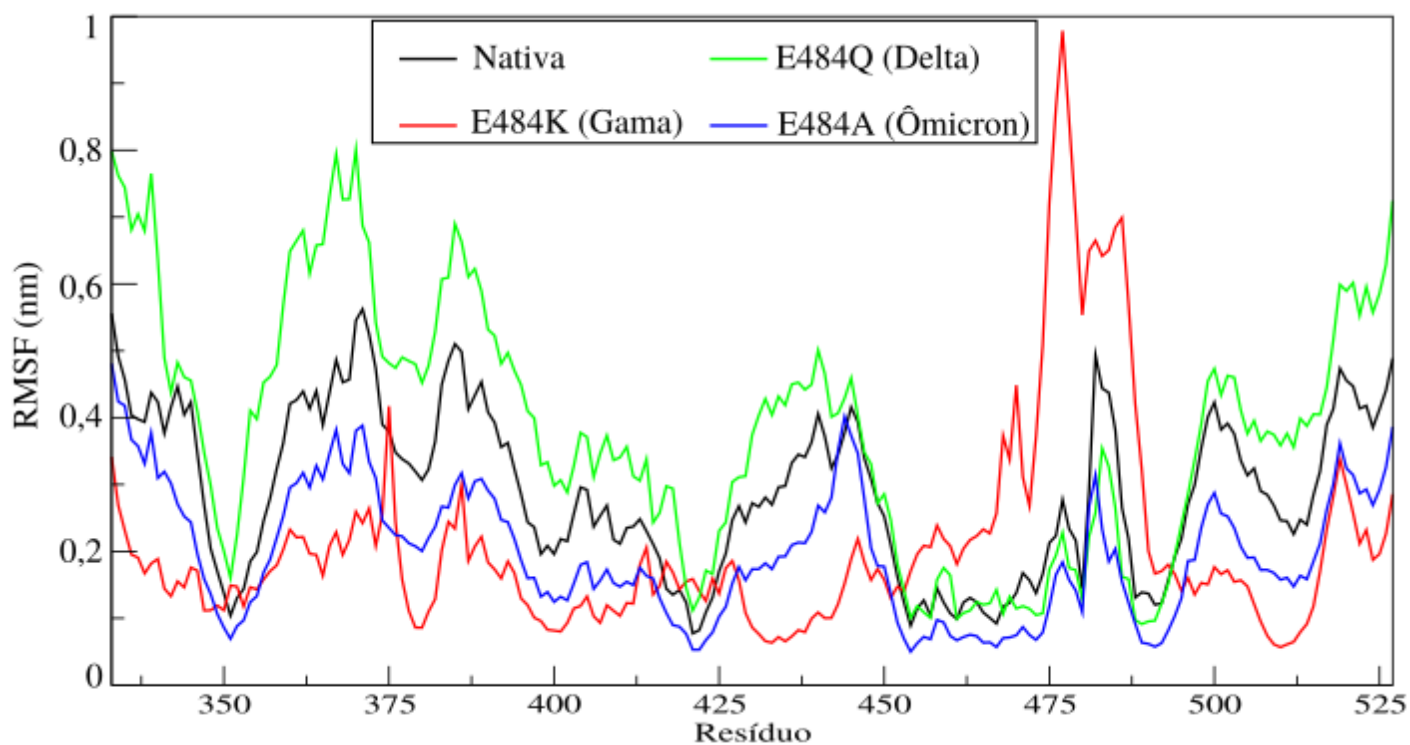


Figura 13 - RMSF de todos os resíduos, com os sistemas Nativa representado em preto, E484K (Gama) representado em vermelho, E484Q (Delta) em verde e E484A (Ômicron) em azul.

Da mesma forma que o RMSD, o sistema Ômicron apresentou os menores valores para RMSF (Figura 13). Os sistemas Nativa e Delta apresentam comportamento semelhante. O sistema que apresentou a maior variação foi o Gama, com picos chegando a 1 nm. As médias aritméticas também seguiram essa tendência, com a Ômicron sendo a menor com 0,1982nm, seguida de Gama 0,2817, Nativa 0,2918 nm e Delta com 0,3821 nm. Para as estruturas secundárias, no geral as alfa hélices apresentaram médias menores. O sistema Ômicron com 0,1260 nm, seguido de Gama 0,2374 nm, Nativa com 0,3168 nm e por último Delta com 0,4664 nm. As folhas betas seguiram essa tendência, com Ômicron a menor, 0,0604, seguida de Nativa, com 0,2323, Gama com 0,2552 e Delta com 0,3402nm. (Tabela 1)

Tabela 1 - Média aritmética, em nm, de todas as estruturas secundárias e laços de todos os sistemas.

	Alfa Hélice	Folha Beta	Laço	Média Geral
Nativa	0,3168	0,2323	0,3366	0,2918
Gama	0,2374	0,2552	0,2109	0,2817
Delta	0,4664	0,3402	0,4452	0,3821
Ômicron	0,1260	0,0604	0,1336	0,1982

O Raio de Giro ( $R_g$ ) é utilizado para obter dados sobre a compactação da estrutura proteica.

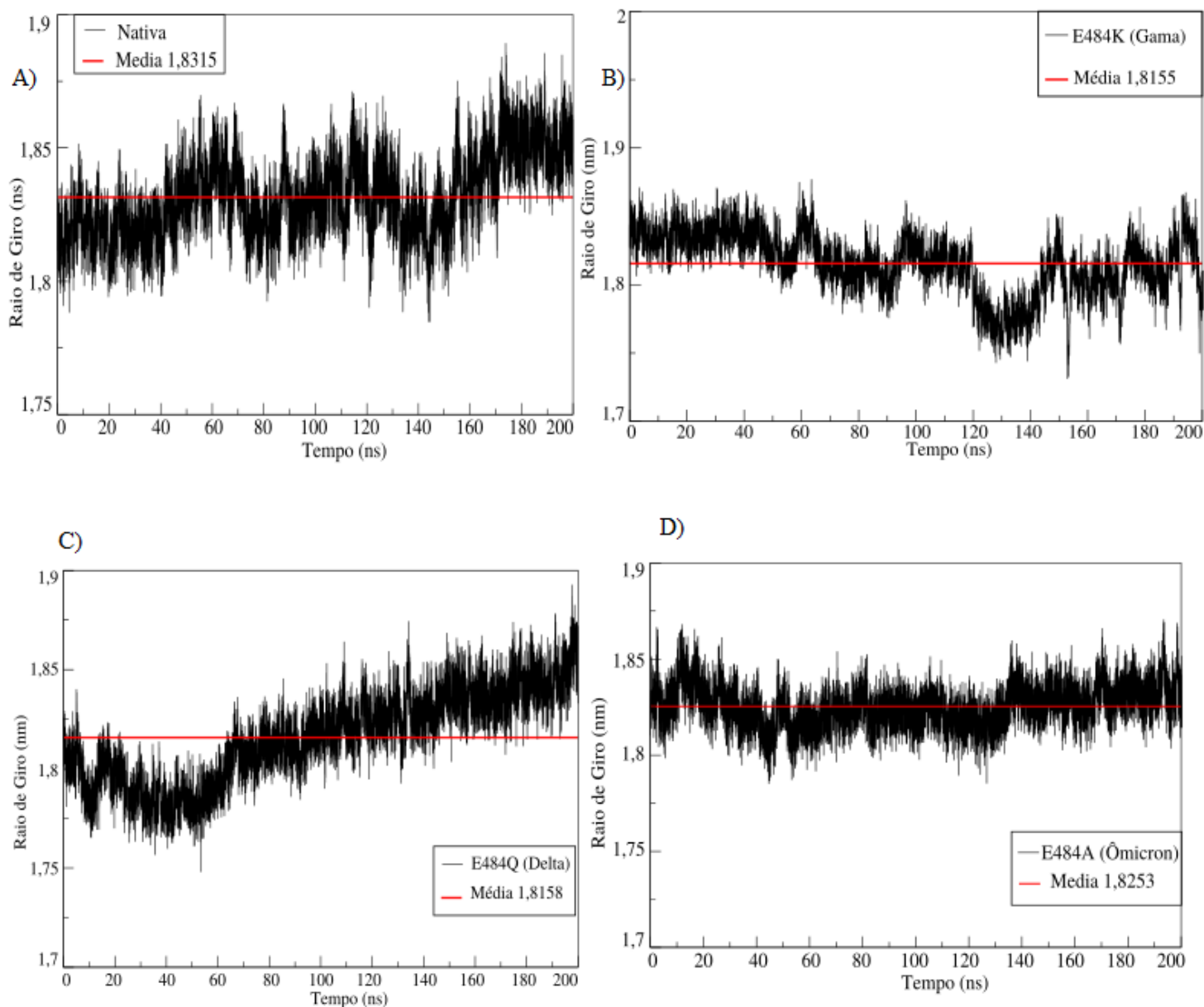


Figura 14 - Raio de Giro. A) representa o sistema Nativa, B) o sistema E484K (Gama), C) o sistema E484Q (Delta) e D) E484A (Ômicron).

No sistema Nativa, durante os primeiros 80 nanossegundos de simulação a estrutura oscila entre 1,8 nm e 1,85 nm, aparentando estar um pouco compactada. No restante da simulação fica entre 1,8 nm e 1,9 nm, onde há descompactação da proteína. A média é de 1,831 nm. Para o sistema Gama durante todo o tempo de simulação ocorre bastante oscilação, principalmente dos 120 nanossegundos até o final da simulação, como podemos observar na Figura 14 B). Os valores ficam entre



1,75 nm e 1,85 nm. A média é de 1,8155 nm, um pouco menor que a média encontrada na Nativa, o que indica uma leve compactação na estrutura em geral. Na Delta podemos observar poucas alterações ao longo da simulação, conforme a Figura 14 C). Nos primeiros 60 nanossegundos de simulação oscila entre 1,75 nm e estabiliza em 1,85 nm até o fim da simulação. A média é 1,8158 nm. A Ômicron é o sistema que aparenta ser o mais estável. Nos 10 primeiros nanossegundos de simulação há um aumento para 1,85 nm, indicando descompactação, e logo depois compacta novamente, estabilizando-se até o final da simulação.

#### **4.1. Interação com receptor ACE-2**

A subunidade S1 da proteína S é a responsável pela ligação ao receptor celular ACE-2. Essa subunidade é dividida em região interna e externa. A externa, que contém o RBD, possui dois laços expostos em sua superfície, o que permite a interação com ACE-2. Um desses laços compreende os resíduos 455 até 491. Por ser uma região importante para a interação vírus-hospedeiro e conter o resíduo que sofreu mutação, foram feitas análises de RMSD e RMSF dessa região, que serão mostradas a seguir.

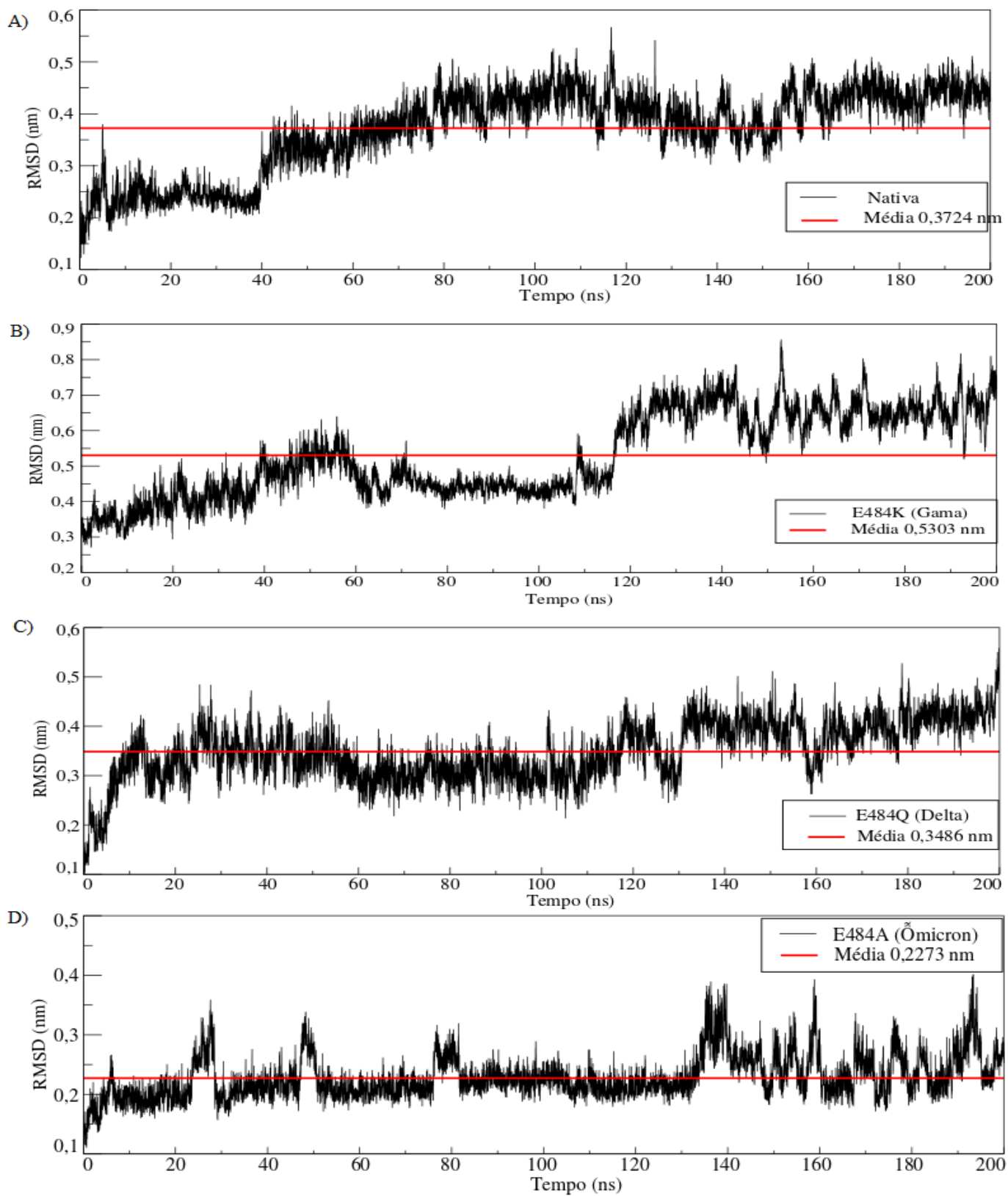


Figura 15 - Figura 15 - RMSD dos resíduos do laço 9. A) representa o sistema Nativa, B) E484K (Gama), C) E484Q (Delta) e E484A (Ômicron).

No sistema Nativa, há vários picos ao longo de toda a simulação. Dos 40 ns ao fim da simulação varia entre 0,4 nm a 0,5 nm, valores maiores que o RMSD de todos os resíduos no geral. A média é de 0,3724 nm. (Figura 15 A). No sistema Gama, na região de “laço” que contém o resíduo de interesse podemos perceber maior instabilidade e valores mais altos de RMSD, chegando a 0,85 nm em comparação com todos os resíduos da proteína, onde o valor de RMSD estava em 0,45 a 0,6 nm. A média é de 0,5303 nm, superior à média encontrada na proteína toda, que é 0,4355 nm. Para o sistema Delta, os valores variam entre 0,2 e 0,5, com várias oscilações ao longo de toda a simulação. Esses valores são um pouco maiores que o encontrado no RMSD de todos os resíduos, que ficava em torno de 0,35 nm, como pode-se observar na figura 15. A média do RMSD é de 0,3486 nm, similar ao encontrado no RMSD de todos os resíduos deste sistema, 0,3405 nm.

Já no Ômicron, o sistema apresenta diversos picos oscilando entre 0,35 a 0,4 nm ao longo de toda a simulação, conforme a figura 15 D. A média de resíduos do laço 9 na Ômicron é 0,2273 nm, menor que a média observada no RMSD da proteína toda, 0,2942 nm (Tabela 2). Todos os sistemas, com exceção do Ômicron, mostraram médias de RMSD menores na proteína toda em comparação com o RMSD do laço 9.

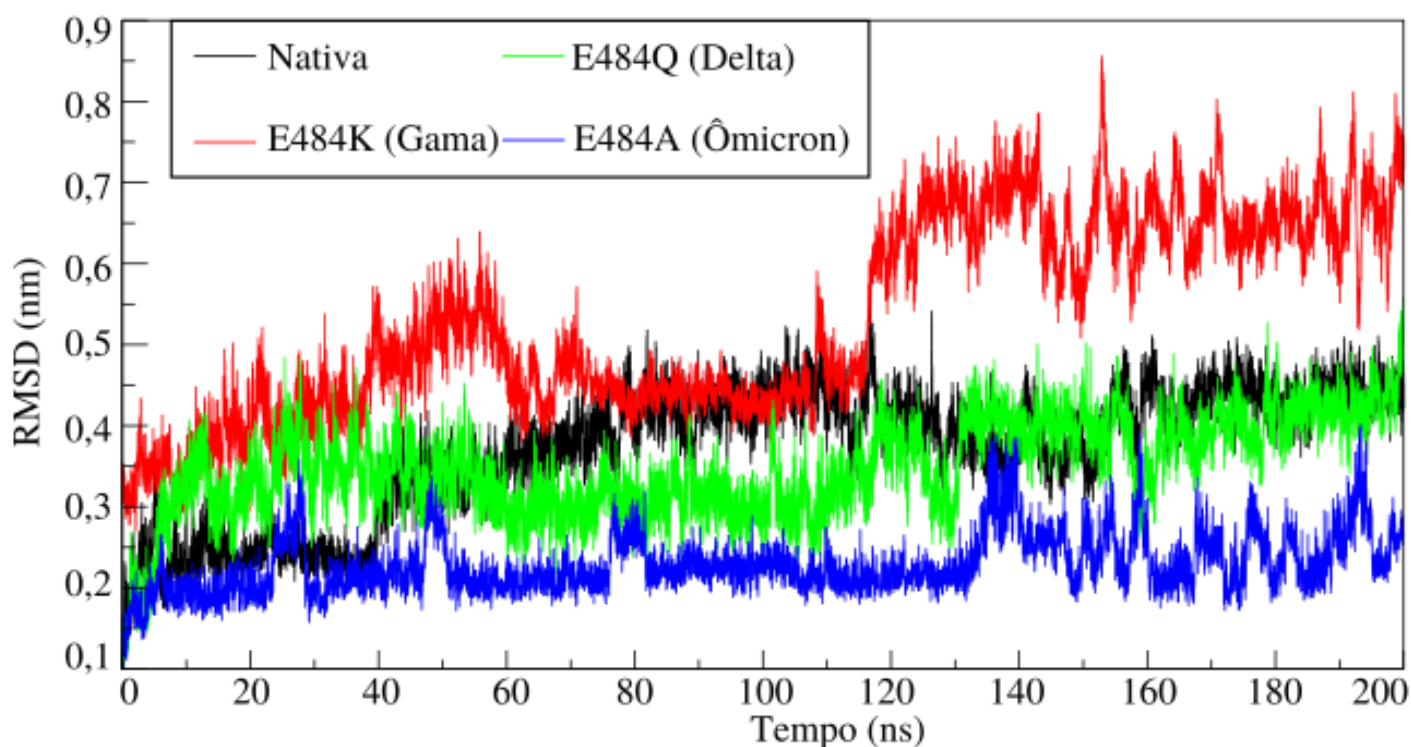


Figura 16 - RMSD dos resíduos 455-491, com os sistemas Nativa, representado em preto, E484K (Gama), representado em vermelho, E484Q (Delta), em verde e E484A (Ômicron) em azul.

Tabela 2 - Comparativo entre médias, em nm, de RMSD e RMSF da proteína toda e do laço 9.

	RMSD Todos os resíduos	RMSD Laço 9	RMSF Todos os resíduos	RMSF Laço 9
Nativa	0,3052	0,3724	0,2918	0,1856
Gama	0,4355	0,5303	0,2103	0,4262
Delta	0,3405	0,3486	0,3821	0,1526
Ômicron	0,2942	0,2273	0,1982	0,1324

Da mesma maneira que o RMSD de todos os resíduos (Fig. 10), o sistema Ômicron mostrou os menores valores para esta análise do laço 9, onde está o resíduo de interesse. Delta e Nativa mostraram comportamento semelhante, sobretudo ao final da simulação, e Gama mostrou a maior instabilidade entre todos os sistemas, conforme podemos observar na Figura 16.

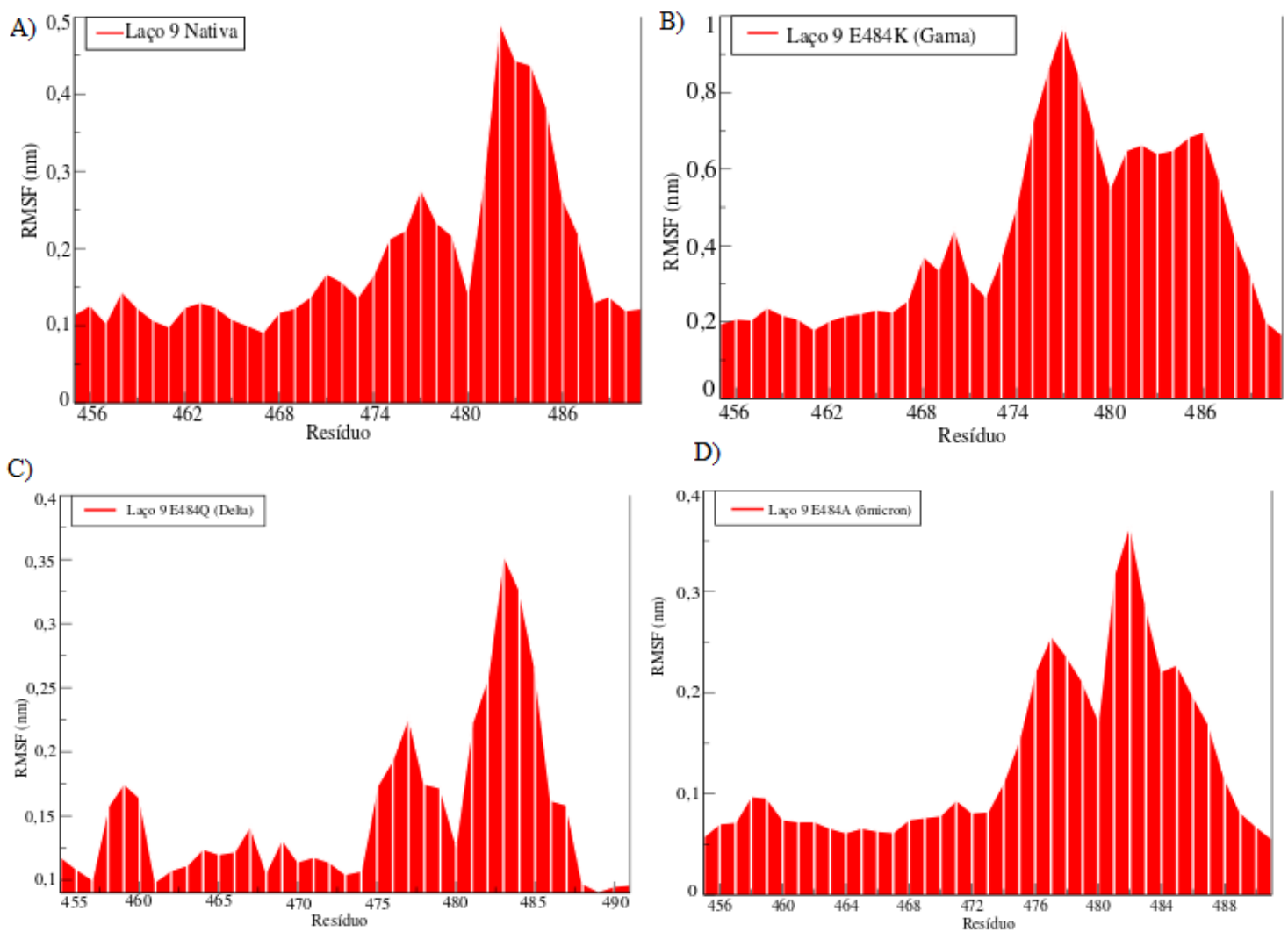


Figura 17 - RMSF dos resíduos 455-491, onde A) representa o sistema Nativa, B) E484K (Gama), C) E484Q (Delta) e D) E484A (Ômicron).

Com relação ao RMSF do laço 9 todos os sistemas, com exceção do Gama, mostraram valores menores do que em relação a proteína toda. O Nativa oscila entre 0,2 a 0,3 nm, com um pico acentuado em 0,5 nm. O sistema Gama possui variações entre 0,2 a 0,7 nm, com um pico alcançando os 1 nm. Já o sistema delta apresenta diversas oscilações, entre 0,1 a 0,4 nm. O sistema Ômicron entre os resíduos 455 ao 473 permaneceu em 0,1 nm, e entre os demais resíduos entre 0,3 a 0,4 nm (Figura 17).

O RBD é uma região da proteína Spike. Possui baixa quantidade de estruturas secundárias, com 66% da proteína sendo composta de laços, 12% de alfa hélices e 22% de folhas betas, sendo que as folhas beta 5 e 6 possuem apenas 3 resíduos

cada. O laço 9 é a estrutura que possui a maior quantidade de resíduos, 37, seguida pelo laço 8, com 14 resíduos. A maior folha hélice é a 2, com apenas 7 resíduos e entre folhas beta são as 3 e 7, com 10 resíduos cada. Isso pode ajudar a explicar porque o laço 9 tem valores menores de RMSF.

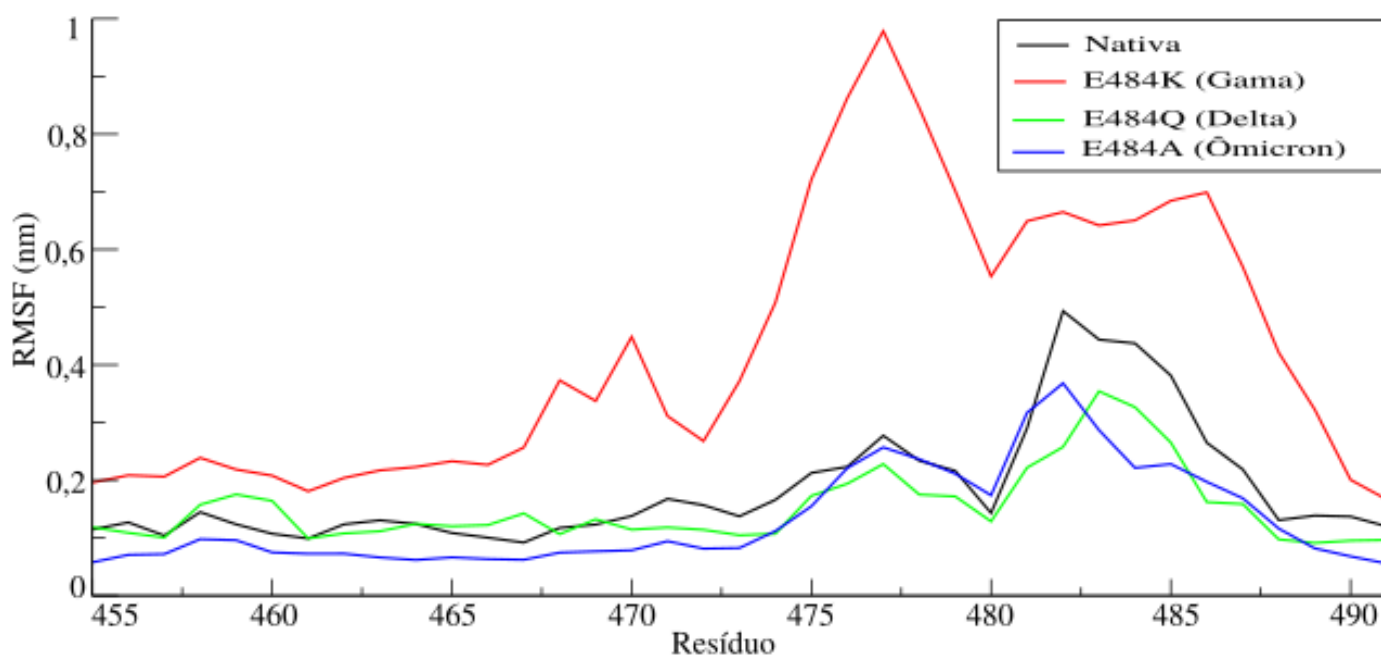


Figura 18 - RMSF do laço 9, com os sistemas Nativa, destacado em preto, E484K (Gama), destacado em vermelho, E484Q (Delta), destacado em verde e E484A (Ômicron) em azul.

Os sistemas Nativa, Ômicron e Delta mostraram comportamentos semelhantes, assim como os valores de RMSF. Quem mais se distancia é o sistema Gama. Mutações virais em regiões de pontos de acesso de ligação do vírus ao hospedeiro geram impactos significativos na afinidade de ligação, estrutura e resposta imune[33]. A mutação K479N no RBM, por exemplo, facilitou a transmissão de SARS-CoV de humanos para humanos[34]. Durante a pandemia de SARS-CoV-2, pudemos observar diversas variantes se sobressaírem umas às outras, conforme mutações foram ocorrendo. Até agora, a variante com maior sucesso evolutivo é a Ômicron, cujo genoma já ultrapassou a Delta em diagnósticos positivos no mundo todo. O resíduo 484 interage com o resíduo K31 da ACE-2. Alguns estudos mostraram que a variante E484K quebra uma ponte salina que existe entre a E484 nativa e o resíduo

E31 da ACE-2. [35] Isso pode explicar os valores muito acima da variante E484K em relação à nativa.

Para o SARS-CoV, a existência de um bolso hidrofóbico formado pelas sequências YGFY (481-484) e YGF (495-497) é fundamental para a interação com o receptor[36]. SARS-CoV e SARS-CoV-2 possuem interações semelhantes com a ACE-2[37]. Para a variante Ômicron ocorreu uma substituição por um aminoácido não polar hidrofóbico, o que poderia aumentar a afinidade com essa região hidrofóbica do receptor. Isso pode ajudar a explicar porque a variante Ômicron tem uma taxa de transmissão maior em relação às outras variantes.

## **5. CONCLUSÃO**

O sistema E484A (Ômicron) apresentou os menores valores de RMSD em comparação com as demais variantes e nativa. A E484K (Gama) apresentou os maiores valores.

No RMSF todos os valores mostraram grande instabilidade. A E484A apresentou os menores valores em quase toda a estrutura, enquanto que na região onde está o laço que contém o resíduo 484 todos os sistemas mostraram valores muito abaixo se comparados ao RMSF de toda a estrutura.



## 6. REFERÊNCIAS

1. Zhu, N., et al., *A Novel Coronavirus from Patients with Pneumonia in China, 2019*. New England Journal of Medicine, 2020. **382**(8): p. 727-733.
2. Saúde, O.M.d. *Coronavirus disease*. 2022 [cited 2022 February 3, 2022]; Available from: <https://www.who.int/>.
3. Corman, V.M., et al., *Chapter Eight - Hosts and Sources of Endemic Human Coronaviruses*, in *Advances in Virus Research*, M. Kielian, T.C. Mettenleiter, and M.J. Roossinck, Editors. 2018, Academic Press. p. 163-188.
4. Cheng Vincent, C.C., et al., *Severe Acute Respiratory Syndrome Coronavirus as an Agent of Emerging and Reemerging Infection*. Clinical Microbiology Reviews, 2007. **20**(4): p. 660-694.
5. Memish Za Fau - Cotten, M., et al., *Human infection with MERS coronavirus after exposure to infected camels, Saudi Arabia, 2013*. (1080-6059 (Electronic)).
6. Saqib, M., et al., *Serologic Evidence for MERS-CoV Infection in Dromedary Camels, Punjab, Pakistan, 2012–2015*. Emerging Infectious Disease journal, 2017. **23**(3): p. 550.
7. Drosten, C., et al., *Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome*. New England Journal of Medicine, 2003. **348**(20): p. 1967-1976.
8. Reusken, C.B.E.M., et al., *Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study*. The Lancet Infectious Diseases, 2013. **13**(10): p. 859-866.
9. Gorbalenya, A.E., et al., *The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2*. Nature Microbiology, 2020. **5**(4): p. 536-544.
10. Wu, F., et al., *A new coronavirus associated with human respiratory disease in China*. Nature, 2020. **579**(7798): p. 265-269.
11. Andersen, K.G., et al., *The proximal origin of SARS-CoV-2*. Nature Medicine, 2020. **26**(4): p. 450-452.
12. Tang, X., et al., *On the origin and continuing evolution of SARS-CoV-2*. National Science Review, 2020. **7**(6): p. 1012-1023.
13. Zhang, T., Q. Wu, and Z. Zhang, *Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak*. Current Biology, 2020. **30**(7): p. 1346-1351.e2.
14. Wu, A., et al., *Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China*. Cell Host & Microbe, 2020. **27**(3): p. 325-328.
15. Malik, Y.A., *Properties of Coronavirus and SARS-CoV-2*. (0126-8635 (Print)).
16. Cosar, B., et al., *SARS-CoV-2 Mutations and their Viral Variants*. Cytokine & Growth Factor Reviews, 2022. **63**: p. 10-22.
17. Liu, Z., et al., *Composition and divergence of coronavirus spike proteins and host ACE2 receptors predict potential intermediate hosts of SARS-CoV-2*. Journal of Medical Virology, 2020. **92**(6): p. 595-601.

18. Wrapp, D., et al., *Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation*. Science, 2020. **367**(6483): p. 1260-1263.
19. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein*. Cell, 2020. **181**(2): p. 281-292.e6.
20. Beniac, D.R., et al., *Architecture of the SARS coronavirus prefusion spike*. (1545-9993 (Print)).
21. Shang, J., et al., *Cell entry mechanisms of SARS-CoV-2*. Proceedings of the National Academy of Sciences, 2020. **117**(21): p. 11727-11734.
22. Alexandre, J., et al., *Renin-angiotensin-aldosterone system and COVID-19 infection*. Annales d'Endocrinologie, 2020. **81**(2): p. 63-67.
23. Claverie, J.M., *Viruses take center stage in cellular evolution*. (1474-760X (Electronic)).
24. Grubaugh, N.D., et al., *Public health actions to control new SARS-CoV-2 variants*. Cell, 2021. **184**(5): p. 1127-1132.
25. Saúde, O.P.-A.d. *OMS anuncia nomenclaturas simples e fáceis de pronunciar para variantes de interesse e de preocupação do SARS-CoV-2*. 2021 [cited 2022 5 February,2022]; Available from: <https://www.paho.org/pt/noticias/1-6-2021-oms-anuncia-nomenclaturas-simples-e-faceis-pronunciar-para-variantes-interesse-e>.
26. Benito, L.A.O., et al., *Variantes de atenção e/ou preocupação por sequenciamento genômico do COVID-19 no Brasil*. Revista de Divulgação Científica Sena Aires, 2021. **10**: p. 783-787.
27. Alexander, H.K., S.I. Mayer, and S. Bonhoeffer, *Population Heterogeneity in Mutation Rate Increases the Frequency of Higher-Order Mutants and Reduces Long-Term Mutational Load*. (1537-1719 (Electronic)).
28. Rehman, S.A.-O., et al., *Identification of novel mutations in SARS-COV-2 isolates from Turkey*. (1432-8798 (Electronic)).
29. Namba, A., V. Silva, and C. Silva, *Dinâmica molecular: teoria e aplicações em planejamento de fármacos*. Ectética Química - ECLET QUIM, 2008. **33**.
30. Anandakrishnan, R., A.V. Aguilar B Fau - Onufriev, and A.V. Onufriev, *H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations*. (1362-4962 (Electronic)).
31. Maier, J.A., et al., *ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB*. Journal of chemical theory and computation, 2015. **11**(8): p. 3696-3713.
32. Mark, P. and L. Nilsson, *Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K*. The Journal of Physical Chemistry A, 2001. **105**(43): p. 9954-9960.
33. Sanjuán, R. and P. Domingo-Calap, *Mechanisms of viral mutation*. Cellular and molecular life sciences : CMLS, 2016. **73**(23): p. 4433-4448.
34. Li, W., et al., *Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2*. The EMBO Journal, 2005. **24**(8): p. 1634-1643.
35. Ibrahim, I.M., A.A. Elfiky, and A.M. Elgohary, *Recognition through GRP78 is enhanced in the UK, South African, and Brazilian variants of SARS-CoV-2; An in silico perspective*. Biochemical and biophysical research communications, 2021. **562**: p. 89-93.

36. Li, F., et al., *Structure of SARS coronavirus spike receptor-binding domain complexed with receptor*. (1095-9203 (Electronic)).
37. Lan, J., et al., *Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor*. *Nature*, 2020. **581**(7807): p. 215-220.