

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

ARTHUR PEREIRA DA FONSECA

**ANÁLISE *IN SILICO* DA LIGAÇÃO DO SÍTIO POLIBÁSICO DE SARS-COV-2 A
SERINA PROTEASES RELACIONADAS A FURINA ATRAVÉS DE DOCAGEM E
DINÂMICA MOLECULAR, INDICANDO CONTATOS POSSIVELMENTE
RELEVANTES**

BELO HORIZONTE
2022

ARTHUR PEREIRA DA FONSECA

**ANÁLISE *IN SILICO* DA LIGAÇÃO DO SÍTIO POLIBÁSICO DE SARS-COV-2 A
SERINA PROTEASES RELACIONADAS A FURINA ATRAVÉS DE DOCAGEM E
DINÂMICA MOLECULAR, INDICANDO CONTATOS POSSIVELMENTE
RELEVANTES**

Versão final

Dissertação apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática da Universidade Federal de Minas Gerais, como requisito parcial para a obtenção do título de Mestre em Bioinformática.

Orientador: Dr. José Miguel Ortega

Coorientadora: Dra. Lucianna Helene Silva dos Santos

BELO HORIZONTE

2022

043

Fonseca, Arthur Pereira da.

Análise in silico da ligação do sítio polibásico de SARS-COV-2 a serina proteases relacionadas a furina através de docagem e dinâmica molecular, indicando contatos possivelmente relevantes [manuscrito] / Arthur Pereira da Fonseca. – 2022.

100 f. : il. ; 29,5 cm.

Orientador: Dr. José Miguel Ortega. Coorientadora: Dra. Lucianna Helene Silva dos Santos.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. Covid-19. 3. Furina. 4. Pró-Proteína Convertase. I. Ortega, José Miguel. II. Santos, Lucianna Helene Silva dos. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS INSTITUTO DE CIÊNCIAS BIOLÓGICAS
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA
FOLHA DE APROVAÇÃO

"Análise in silico da ligação do sítio polibásico de SARS-CoV-2 a serina proteases relacionadas a Furina através de docking e dinâmica molecular, indicando contatos possivelmente relevantes"

Arthur Pereira da Fonseca

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Prof. José Miguel Ortega - Orientador
Universidade Federal de Minas Gerais

Profa. Lucianna Helene Silva dos Santos – Coorientadora
Institut Pasteur de Montevideo

Prof. Lucas Bleicher
Universidade Federal de Minas Gerais

Prof. Rodrigo Bentes Kato
Universidade Federal de Minas Gerais

Belo Horizonte, 21 de dezembro de 2022.



Documento assinado eletronicamente por **Lucianna Helene Silva dos Santos, Usuário Externo**, em 21/12/2022, às 11:25, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rodrigo Bentes Kato, Usuário Externo**, em 21/12/2022, às 11:38, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por Jose Miguel Ortega, Servidor(a), em 23/12/2022, às 11:46, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Lucas Bleicher, Professor do Magistério Superior**, em 16/01/2023, às 12:24, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **1970921** eo código CRC **A97E5445**.

AGRADECIMENTOS

Agradeço em primeiro lugar aos meus pais. À minha mãe, quem eu mais admiro e minha maior inspiração, quem eu sempre pude contar e que sempre sabe como me acalmar. Ao meu pai, que me apoiou e incentivou desde o início dessa caminhada.

Agradeço também à minha namorada, que foi quase uma segunda coorientadora para mim, com quem as conversas e conselhos me ajudaram a resolver os problemas que foram aparecendo. Agradeço também aos meus irmãos e meu padrasto, por toda a ajuda, apoio e compreensão nesse momento.

Um agradecimento em especial ao meu orientador Miguelito, que foi mais do que um professor pra mim. Sem o qual talvez eu nem teria começado essa jornada. Ele que me ajudou desde antes de iniciar o mestrado, me dando aulas para prova. Que correu atrás para garantir minha bolsa, e que até me ofereceu hospedagem quando estive em Belo Horizonte. E cuja orientação não poderia ser melhor nessa jornada, sendo super paciente mesmo passando pelos momentos difíceis que passamos nessa pandemia.

À minha coorientadora Lucianna Helene, que aceitou participar desse processo, e foi fundamental para conclusão do mesmo. Que teve toda a paciência para ficar horas em reuniões, quase que semanais, para discutir o trabalho, dando os melhores conselhos e sugestões.

Agradeço à equipe do Biodados, por toda ajuda, em especial ao André, que praticamente acumulou a função de técnico em informática para nos salvar. E também ao pessoal do laboratório das professoras Liza e Rafaela, cujo auxílio com os nossos computadores foi fundamental para que eu pudesse concluir essa jornada de forma remota, especialmente nos momentos de quarentena que estive em outra cidade.

E agradeço também a secretaria do curso, sempre muito disposta a nos atender. E aos vários professores que deram dicas e sugestões no decorrer desse trabalho. A todo mundo que de alguma forma participou desse processo: meu muito obrigado!

“So there you sit. And how much blood was shed
That you might sit there. Do such stories bore you?
Well, don't forget that others sat before you
who later sat on people. Keep your head!
Your science will be valueless, you'll find
And learning will be sterile, if inviting
Unless you pledge your intellect to fighting
Against all enemies of all mankind.
Never forget that men like you got hurt
That you might sit here, not the other lot.
And now don't shut your eyes, and don't desert
But learn to learn, and try to learn for what.”

To the Students of the Workers' and Peasants' Faculty

Bertolt Brecht

RESUMO

Em dezembro de 2019 foi identificado um novo vírus da família dos *Coronaviridae*, chamado de SARS-CoV-2. Esse vírus, responsável por uma síndrome respiratória aguda grave, causou a pandemia de COVID-19 e milhões de mortes em todo o mundo. Uma das adaptações desse vírus que o permite infectar humanos é a inserção de um sítio de clivagem polibásico, formando a sequência PRRARS, na sua proteína da espícula. Esse sítio polibásico é reconhecido e clivado pela enzima Furina, separando as subunidades S1 e S2 da proteína. Após a clivagem pela Furina, a espícula assume uma conformação ativa que pode potencializar a infecção viral através da fusão vírus-célula, e, principalmente, através da fusão célula-a-célula em um processo causador de sincícios. Nesse estudo investigamos *in silico* se podem ocorrer interações entre a alça que contém o sítio polibásico e a enzima Furina, estudando as interações que podem ser relevantes para o encaixe e seus comportamentos durante simulações de dinâmica molecular. A enzima Furina faz parte de uma família de pró-proteína convertases subtilisina kexina, as PCSKs, e, por isso, também foram investigadas as enzimas PCSK 1 a 9 para verificar se poderiam, da mesma forma, estar atuando na ativação da proteína viral. Para isso, modelamos um peptídeo cobrindo a sequência do sítio de clivagem polibásico que foi docado sobre o sítio ativo das enzimas. Em seguida, fizemos simulações de dinâmica molecular para estudar e caracterizar a ligação entre o peptídeo e o bolso de ligação de cada enzima. Verificamos que a enzima Furina realiza interações de pontes salinas entre dois aspartatos do seu bolso de ligação S1, ASP258 e ASP306, e a arginina P1 do peptídeo e também interações de hidrogênio entre P1 e PRO256. Além disso, verificamos uma ligação entre a arginina P3 do peptídeo e o aspartato ASP191 do bolso de ligação S2 da Furina. Para as demais PCSKs, observamos uma fraca afinidade entre a PCSK1 e a PCSK5 com o peptídeo, sendo necessários mais estudos para comprovar se essas enzimas poderiam clivar o sítio. Já a PCSK2 apresentou interações muito similares às aquelas apresentadas pela Furina, podendo realizar interações através de dois aspartatos e uma prolina do bolso de ligação S1 ao resíduo P1, e uma ligação de um aspartato próximo ao subsítio S2 com P3. Dessa forma, propomos que a PCSK2, possivelmente, pode estar participando na clivagem e ativação da espícula, potencializando a infecção pelo vírus SARS-CoV-2.

Palavras-chave: COVID, COVID-19, coronavírus, SARS-CoV-2, Furina, PCSK.

ABSTRACT

In December 2019, a new virus from the *Coronaviridae* family, called SARS-CoV-2, was identified. This virus, responsible for a severe acute respiratory syndrome, caused the COVID-19 pandemic and millions of deaths worldwide. One of the adaptations of this virus that allows it to infect humans is the insertion of a polybasic cleavage site, forming the PRRARS sequence, in its spike protein. This polybasic site is recognized and cleaved by the Furina enzyme, separating the S1 and S2 subunits of the protein. After cleavage by Furina, the spike assumes an active conformation that can potentiate the viral infection through virus-cell fusion, and, mainly, through cell-cell fusion in a process that causes syncytia. In this study, we investigated *in silico* the interactions that can occur between the loop that contains the polybasic site and the enzyme Furina, studying the interactions that may be relevant to the fit and their behavior during molecular dynamics simulations. The Furina enzyme is part of a family of proprotein convertase subtilisin/kexin, the PCSKs, and therefore, PCSK enzymes 1 to 9 were also investigated to verify whether they could, in the same way, be acting in the activation of the viral protein. For this, we modeled a peptide covering the sequence of the polybasic cleavage site and it was docked over the active site of the enzymes. Then, we performed molecular dynamics simulations to study and characterize the interactions between the peptide and the binding pocket of each enzyme. Thus, we verified that the Furina enzyme performs salt bridge interactions between P1 arginine and two aspartates from its S1 binding pocket, ASP258 and ASP306, and also hydrogen interactions between P1 and PRO256. In addition, we found a link between P3 arginine and aspartate ASP191 from Furina S2 binding pocket. For the other PCSKs, we observed a weak affinity between PCSK1 and PCSK5 to the peptide, requiring further studies to prove whether these enzymes could cleave the site. PCSK2, on the other hand, presented interactions very similar to those presented by Furina, being able to perform interactions through two aspartates and a proline from the S1 binding pocket to the P1 residue, and a binding of an aspartate close to the S2 subsite with P3. Thus, we propose that PCSK2 may possibly be participating in the cleavage and activation of the spike, enhancing infection by the SARS-CoV-2 virus.

Keywords: COVID, COVID-19, coronavirus, SARS-CoV-2, Furina, PCSK

LISTA DE FIGURAS

Figura 1.1 - Esquema da estrutura do SARS-CoV-2 mostrando a proteína espícula na membrana.	19
Figura 1.2 – Alinhamento de trecho das sequências de espícula de diferentes coronavírus do tipo SARS.	20
Figura 1.3 – Estados da proteína espícula antes e após a clivagem por Furina.	21
Figura 1.4 – Árvore filogenética da família das PCSKs humanas.	23
Figura 1.5 - Formação do íon alcóxido nas serina proteases.	24
Figura 1.6 - Mecanismo de catálise nas serina proteases.	25
Figura 1.7 – Alinhamento das sequências proteicas de todas as PCSKs, com exceção da PCSK8.	28
Figura 4.1 - Fluxograma da metodologia	31
Figura 5.1 – Gráfico de potencial de clivagem por resíduo.	39
Figura 5.2 – Sequência dos peptídeos criados para teste.	41
Figura 5.3 – Modelo dos peptídeos de dez e treze aminoácidos.	41
Figura 5.4 – Modelos das docagens com os peptídeos de dez e treze aminoácidos.	42
Figura 5.5 – Sequência e nomenclatura do peptídeo criado.	42
Figura 5.6 – Modelo dos peptídeos após docagem com Haddock (ciano), selecionado por melhor <i>score</i> , e com CABSdock (rosa), selecionado pela proximidade ao bolso de ligação.	44
Figura 5.7 – <i>No fit</i> RMSD durante a simulação do peptídeo docado com Haddock à enzima Furina.	46
Figura 5.8 – Modelo ao final da simulação do peptídeo docado na Furina usando Haddock.	47
Figura 5.9 – Gráfico de porcentagem de BSA de P1 durante a simulação com a enzima Furina a partir da docagem com Haddock.	48
Figura 5.10 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima Furina durante a simulação realizada a partir da docagem com Haddock.	48
Figura 5.11 – Interações entre P1 e os resíduos do bolso de ligação S1 ao final da simulação a partir da docagem com Haddock.	49
Figura 5.12 – Mapa de possíveis contatos não-covalentes entre o peptídeo e a enzima Furina durante a simulação realizada a partir da docagem com Haddock.	50

Figura 5.13 - Interações entre P3 e ASP191 (rosa) ao final da simulação com a Furina a partir da docagem com Haddock.....	51
Figura 5.14 - RMSF do peptídeo por resíduo, medido na simulação realizada após a docagem à enzima Furina com Haddock.	52
Figura 5.15 – Gráfico de energia de van der Waals entre o peptídeo e a enzima Furina durante a simulação a partir da docagem com Haddock.	52
Figura 5.16 – Interação entre o peptídeo e sítio catalítico da Furina ao final da simulação rodada a partir da docagem com Haddock.	53
Figura 5.17 – Posição final da simulação a partir da docagem do peptídeo à Furina usando CABSdock.	54
Figura 5.18 - Mapa de possíveis pontes salinas entre o peptídeo e a enzima Furina durante a dinâmica realizada a partir da docagem com CABSdock.....	54
Figura 5.19 – <i>No fit</i> RMSD do peptídeo docado com Haddock na enzima PCSK2 durante a simulação.	55
Figura 5.20 - Modelo ao final da simulação do peptídeo docado na PCSK2 usando Haddock.	56
Figura 5.21 – Gráfico de BSA de P1 durante a dinâmica com a enzima PCSK2, docagem com Haddock.....	56
Figura 5.22 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK2 durante a simulação realizada a partir da docagem com Haddock.	57
Figura 5.23 – Interações entre P1 e os resíduos do bolso de ligação S1 ao final da simulação a partir da docagem com Haddock.....	58
Figura 5.24 – Mapa de possíveis contatos não-covalentes entre o peptídeo e a enzima PCSK2 durante a simulação realizada a partir da docagem com Haddock.	59
Figura 5.25 - Gráfico de distâncias em Å entre o oxigênio de PRO271 e um hidrogênio da cadeia lateral de P1, durante a simulação com PCSK2 a partir da docagem com Haddock.	59
Figura 5.26 - Interações entre P3 e ASP203 (rosa) ao final da simulação a partir da docagem com Haddock.....	60
Figura 5.27 - RMSF do peptídeo por resíduo, medido na simulação realizada após a docagem à enzima PCSK2 com Haddock.	60
Figura 5.28 – Gráfico de energia de van der Waals entre o peptídeo e a enzima PCSK2 durante a dinâmica a partir da docagem com Haddock.....	61

Figura 5.29 – Figura mostrando o peptídeo e sítio catalítico no final da simulação rodada a partir da docagem com Haddock.....	62
Figura 5.30 – Distância entre o nitrogênio da cadeia lateral de P1 e o oxigênio de ASP273, localizado na borda do bolso de ligação, durante a simulação com a enzima PCSK2, a partir da docagem com CABSdock.....	62
Figura 5.31 - Interações entre P1 e P3 e os resíduos do bolso de ligação S1 na simulação a partir da docagem com CABSdock na enzima PCSK2.	63
Figura 5.32 - Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK2 durante a simulação realizada a partir da docagem com CABSdock.....	63
Figura 5.33 – Posição final do peptídeo na simulação a partir da docagem à PCSK1 usando Haddock.	64
Figura 5.34 – RMSD do peptídeo docado com Haddock na enzima PCSK1 durante a simulação.	65
Figura 5.35 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK1 durante a simulação realizada a partir da docagem com Haddock.	66
Figura 5.36 - Interações entre P1 e ASP272 do bolso de ligação S1 da PCSK1, ao final da simulação a partir da docagem com CABSdock.....	67
Figura 5.37 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK1 durante a simulação realizada a partir da docagem com CABSdock.....	67
Figura 5.38 – Posição final do peptídeo na simulação a partir da docagem à PCSK4 usando Haddock.	68
Figura 5.39 – Gráfico de energia de van der Waals entre o peptídeo e a enzima PCSK1 durante a simulação a partir da docagem com Haddock.....	69
Figura 5.40 – RMSD do peptídeo docado com Haddock na enzima PCSK5 durante a simulação.	69
Figura 5.41 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK5 durante a simulação realizada a partir da docagem com Haddock.	70
Figura 5.42 - Interações entre P1 e a enzima PCSK5, ao final da simulação a partir da docagem com CABSdock.....	71
Figura 5.43 – Posição final do peptídeo na simulação a partir da docagem à PCSK6 usando Haddock.	71
Figura 5.44 – Mapa de possíveis contatos não-covalentes entre o peptídeo e a enzima PCSK7 durante a simulação realizada a partir da docagem com Haddock.	72

Figura 5.45 – Posição final do peptídeo na simulação a partir da docagem à PCSK9 usando Haddock.	73
Figura 5.46 - Gráfico de distâncias em Å entre P1 e PRO237 do bolso de ligação S1 da enzima PCSK9, durante a simulação de dinâmica a partir da docagem com Haddock.	74
Figura 5.47 - Figura mostrando uma molécula corrompida ao tentar retirar PBC utilizando o VMD sem informações iniciais da trajetória.	78

LISTA DE TABELAS

Tabela 1.1 – Resíduos mais comuns nos subsítios do bolso de ligação em Furina humana.	27
Tabela 5.1 – RMSD do sítio catalítico e de ligação medido entre modelos das PCSKs criados usando diferentes metodologias.....	40
Tabela 5.2 – Tabela de score obtido nas docagens do peptídeo às enzimas utilizando Haddock.	44
Tabela 6.1 – Porcentagem de <i>frames</i> em que P1 está até 3,3 Å de cada resíduo do bolso de ligação S1 para todas as simulações.....	84

LISTA DE ABREVIATURAS E SIGLAS

Å: Ångström

BSA: Buried Surface Area

COVID-19: Doença por Coronavírus 2019

ECA2: Enzima Conversora de Angiotensina 2

FS: Femtossegundo

K: Kelvin

NS: Nanossegundo

PBC: Periodic Boundary Conditions

PCSK: Proprotein Convertase Subtilisin/Kexin

PME: Particle Mesh Ewald

PS: Picossegundo

RBD: Receptor-binding Domain

RMS: Root Mean Square

RMSD: Root Mean Square Deviation

RMSF: Root Mean Square Fluctuation

SARS: Severe Acute Respiratory Syndrome

SARS-CoV-2: Severe Acute Respiratory Syndrome Coronavirus 2

SASA: Solvent Accessible Surface Area

TMPRSS2: Transmembrane Protease Serine 2

SUMÁRIO

1	INTRODUÇÃO	18
1.1	Pandemia de coronavírus.....	18
1.2	SARS-CoV-2	18
1.2.1	Sítio de clivagem polibásico.....	19
1.2.2	Infecção viral.....	20
1.3	PCSKs.....	22
1.3.1	Mecanismo de clivagem.....	23
1.3.2	Bolso de ligação.....	26
2	JUSTIFICATIVA	29
3	OBJETIVOS	30
3.1	Objetivos gerais.....	30
3.2	Objetivos específicos.....	30
4	MATERIAIS E MÉTODOS.....	31
4.1	Estruturas proteicas.....	32
4.2	Preparo das estruturas para docagem	32
4.2.1	Preparo das enzimas para docagem	32
4.2.2	Preparo dos peptídeos para docagem	33
4.3	Docagem	33
4.3.1	Haddock.....	33
4.3.2	CABSdock.....	34
4.4	Simulações de dinâmica molecular	34
4.5	Análises.....	35
4.5.1	RMS.....	35
4.5.2	Contatos.....	36
4.5.3	Distâncias	37
4.5.4	SASA e BSA	37
4.5.5	Energias.....	37
5	RESULTADOS	38
5.1	Definição da ligação suscetível à hidrólise	38
5.2	Modelos.....	39
5.2.1	Modelos das enzimas	39
5.2.2	Modelos dos peptídeos	40
5.3	Docagem	43
5.4	Simulações de dinâmica molecular	45
5.4.1	Furina.....	45
5.4.2	PCSK2	55

5.4.3	PCSK1	64
5.4.4	PCSK4	68
5.4.5	PCSK5	69
5.4.6	PCSK6	71
5.4.7	PCSK7	72
5.4.8	PCSK9	73
5.5	Aplicação computacional Pamda	75
5.5.1	RMS	76
5.5.2	Contatos.....	76
5.5.3	Distâncias	76
5.5.4	SASA e BSA	77
5.5.5	Energias.....	77
5.5.6	Gráficos.....	77
5.5.7	Análise de simulações muito grandes.....	77
6	DISCUSSÃO	79
6.1	Furina	79
6.2	PCSK2.....	80
6.3	Demais PCSKs.....	83
6.4	Sequência do sítio polibásico e variantes.....	86
7	CONCLUSÃO.....	88
	REFERÊNCIAS.....	90
	APÊNDICE A – PARÂMETROS DE CONFIGURAÇÃO NAMD.....	98
	APÊNDICE B – RESULTADOS DAS ANÁLISES	100
	APÊNDICE C – PONTE SALINAS AO LONGO DA SIMULAÇÃO COM PCSK7..	100
	APÊNDICE D – DOCUMENTAÇÃO PAMDA	100
	APÊNDICE E – CÓDIGO FONTE PAMDA.....	100

1 INTRODUÇÃO

1.1 Pandemia de coronavírus

Em dezembro de 2019, um grupo de pacientes em Wuhan, China, apresentou sintomas de pneumonia causada por um agente não identificado. Um vírus do gênero dos *Betacoronavirus*, até então desconhecido, foi apontado como causador da doença. Este vírus foi inicialmente chamado de 2019-nCoV, e depois como coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2) (WORLD HEALTH ORGANIZATION, 2020a; ZHU, N. *et al.*, 2020).

Em 11 de março de 2020, o surto da doença causada pelo vírus, conhecida como doença de coronavírus 2019 (COVID-19), foi classificado como pandemia pela Organização Mundial da Saúde (WORLD HEALTH ORGANIZATION, 2020b).

1.2 SARS-CoV-2

O SARS-CoV-2 pertence à família dos *Coronaviridae*, e ao subgênero dos vírus tipo SARS: *Sarbecovirus*. A síndrome respiratória aguda grave (SARS) causada pelos vírus desse gênero emergiu por volta de 2002, e desde então diversas linhagens foram descobertas em morcegos, seu repositório natural (LI, W. *et al.*, 2005; YANG, L. *et al.*, 2013). Acredita-se, portanto, que o SARS-CoV-2 também teve sua origem nos morcegos, passando a infectar o ser humano através de transmissão zoonótica (WONG *et al.*, 2020; ZHOU, P. *et al.*, 2020).

O vírus possui na sua superfície uma glicoproteína que medeia a entrada na célula hospedeira. A proteína espícula (do inglês *Spike Protein*), como é conhecida, é inserida em múltiplas cópias na membrana do vírus, dando-lhe a aparência de coroa, o que confere o nome à família (**Figura 1.1**). Essa proteína possui duas subunidades: S1 e S2 (JACKSON, C. B. *et al.*, 2021).

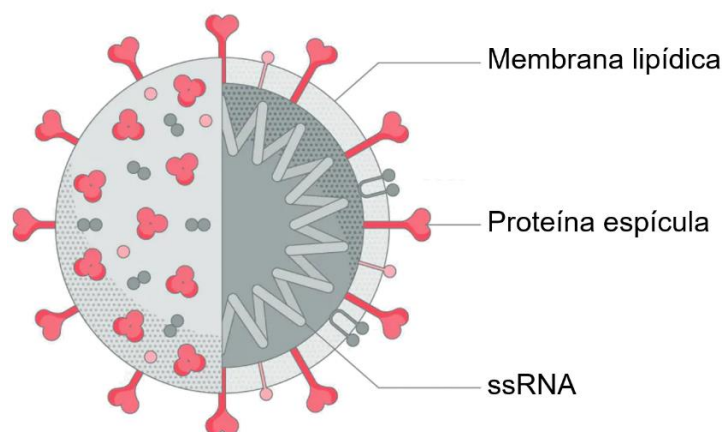


Figura 1.1 - Esquema da estrutura do SARS-CoV-2 mostrando a proteína espícula na membrana. Fonte: adaptado de ABCAM, 2022.

Duas características exclusivas do genoma desse novo coronavírus são importantes para que ele infecte humanos (ANDERSEN *et al.*, 2020). Uma que permite que o patógeno use de forma eficiente como receptor a enzima conversora de angiotensina 2 (ECA2) para entrar nas células hospedeiras (WU, F. *et al.*, 2020). E a outra, uma inserção de um sítio de clivagem polibásico, entre as subunidades S1 e S2 da proteína espícula, reconhecido pela enzima Furina humana (VILLOUTREIX *et al.*, 2020; WALLS *et al.*, 2020).

1.2.1 Sítio de clivagem polibásico

O sítio de clivagem polibásico ocorre na junção entre as subunidades S1 e S2 da glicoproteína espícula do SARS-CoV-2, e é processado durante sua biogênese ou quando ele chega a uma nova célula (JACKSON, C. B. *et al.*, 2021). Este sítio é uma característica exclusiva desse novo vírus, diferenciando-o dos demais vírus do tipo SARS (WALLS *et al.*, 2020).

A **Figura 1.2** mostra a inserção que ocorreu no genoma desse vírus, de quatro aminoácidos, começando com prolina. A origem mais provável dessa inserção foi através de seleção natural em humanos após transferência zoonótica, ou seleção em um animal hospedeiro intermediário antes da transferência em humanos, sendo improvável que tenha sido feita de forma artificial em um laboratório (GARRY, 2022).

	*	*	*	*	*	*	*	*		*								*	*						*		
Human – SARS-CoV-2	G	A	G	I	C	A	S	Y	Q	T	Q	T	N	S	P	R	R	A	R	S	V	A	S	Q	S	I	I
Bat – SARS-CoV RaTG13	G	A	G	I	C	A	S	Y	Q	T	Q	T	N	S	-	-	-	-	R	S	V	A	S	Q	S	I	I
Pangolin – CoV MP789	G	A	G	I	C	A	S	Y	Q	T	Q	T	N	S	-	-	-	-	R	S	V	S	S	X	A	I	I
Human – SARS-CoV	G	A	G	I	C	A	S	Y	H	T	V	S	L	L	-	-	-	-	R	S	T	S	Q	K	S	I	V
Bat – SARS-CoV-related	G	A	G	I	C	A	S	Y	H	T	V	S	S	L	-	-	-	-	R	S	T	S	Q	K	S	I	V

Figura 1.2 – Alinhamento de trecho das sequências de espícula de diferentes coronavírus do tipo SARS. Em destaque o sítio polibásico que é reconhecido pela enzima Furina, mostrando a inserção de quatro aminoácidos em SARS-CoV-2.

A formação desse sítio torna a proteína espícula um potencial substrato a ser clivado por pró-convertases (WALLS *et al.*, 2020), por meio de uma clivagem proteolítica que é amplamente utilizada para ativar a maquinaria de fusão de glicoproteínas virais, e é essencial para o processo de infecção viral (PEACOCK *et al.*, 2021; WHITTAKER, G. R., 2021).

1.2.2 Infecção viral

Para entrada na célula hospedeira, a subunidade S1 da proteína espícula madura do vírus se liga ao receptor humano ECA2 através do seu domínio de ligação ao receptor (RBD, na sigla em inglês). Enquanto a subunidade S2 facilita a fusão entre membranas através do peptídeo de fusão (PAPA *et al.*, 2021). A entrada do vírus na célula, portanto, é potencializada pela clivagem da espícula nessas subunidades por enzimas da célula hospedeira (HOFFMANN; KLEINE-WEBER; PÖHLMANN, 2020; JACKSON, C. B. *et al.*, 2021).

A clivagem e ativação da proteína espícula do coronavírus ocorre em duas etapas. Primeiro a enzima Furina cliva a proteína no sítio de clivagem polibásico, dividindo as unidades S1 e S2. E depois a proteína é ativada mediante clivagem do sítio S2' pela serina protease transmembrana 2 (TMPRSS2, na sigla em inglês) da célula hospedeira, expondo o peptídeo de fusão (HOFFMANN; KLEINE-WEBER; PÖHLMANN, 2020; JACKSON, C. B. *et al.*, 2021; PAPA *et al.*, 2021). O peptídeo de fusão da espícula é responsável por ativar a fusão da membrana do vírus com a da célula hospedeira, e também a fusão entre células no processo de formação de sincícios (PAPA *et al.*, 2021).

A TM serina protease 2 (TMPRSS2), é uma protease presente na membrana da célula humana. A ativação da espícula do vírus pela clivagem dessa proteína é importante, embora não seja essencial, para o reconhecimento do RBD pela ECA2 e posterior entrada do vírus na célula (HOFFMANN; KLEINE-WEBER; PÖHLMANN, 2020). Em células que não expressam essa enzima, a ativação da proteína espícula pode ser realizada via catepsina B/L (SIMMONS *et al.*, 2005).

A enzima Furina, é uma serina protease expressa diferencialmente nos pulmões. Essa protease celular pode clivar a proteína espícula do coronavírus no sítio polibásico, separando as unidades S1 e S2 e expondo o RBD para ligação ao ECA2 (HOFFMANN; KLEINE-WEBER; PÖHLMANN, 2020; WHITTAKER, G. R., 2021). A **Figura 1.3** mostra os estados da proteína espícula antes e depois da clivagem por Furina.

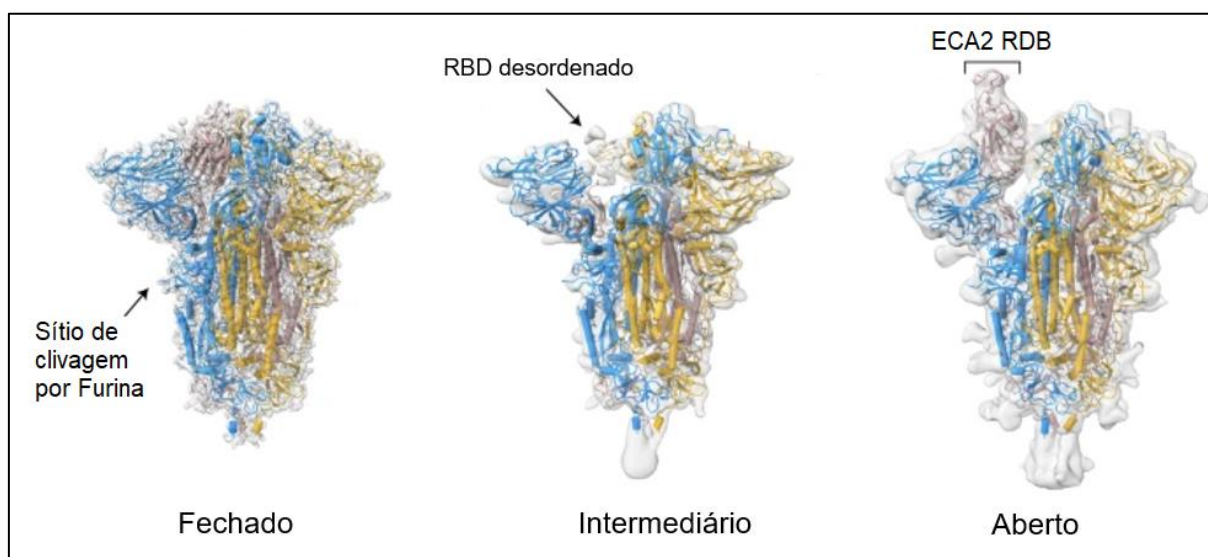


Figura 1.3 – Estados da proteína espícula antes e após a clivagem por Furina. Após a clivagem, a proteína assume um estado transitório, que culmina na sua forma madura, com o RBD exposto para reconhecimento por ECA2. Fonte: adaptado de WROBEL *et al.*, 2020.

A clivagem da espícula pela Furina potencializa a infecção viral, apesar de também não ser essencial para a mesma (PAPA *et al.*, 2021). No entanto, o aumento da infecção está mais associado com o aumento da fusão célula-a-célula do que a fusão vírus-célula (HOFFMANN; KLEINE-WEBER; PÖHLMANN, 2020). Estudos sugerem que a Furina, enzima da família das pro-proteíno convertases

subtilisina/kexina (PCSKs), pode aumentar em até 200% a fusão célula-a-célula em infecções por SARS-CoV-2, permitindo que o genoma do vírus se espalhe independentemente do vírion, a partícula viral (AMIDEI; DOBROVOLNY, 2022; HOFFMANN; KLEINE-WEBER; PÖHLMANN, 2020; PAPA *et al.*, 2021).

1.3 PCSKs

A enzima Furina é membra do grupo de pró-proteíno convertases e pertence à família das serina proteases semelhantes à subtilisina/kexina, as PCSKs. As pró-proteíno convertases atuam principalmente na via secretora, onde diversas proteínas secretadas ou proteínas transmembrana são inicialmente sintetizadas como percussores de proteínas e subsequentemente clivadas por essas convertases, antes de se tornarem totalmente funcionais (TIAN, S.; JIANHUA, 2010). Além disso, a clivagem de percussores de proteínas pró-convertases está envolvida na patogênese molecular de diversas doenças humanas, incluindo câncer, doenças autoimunes e várias infecções virais (THOMAS, 2002). A sequência mais comum dos substratos das PCSKs é dada por R-X-[K/R]-R↓, em que a seta vertical indica a ligação que será clivada (TIAN, S.; JIANHUA, 2010).

Ao todo, nove enzimas fazem parte dessa família: PCSK1, PCSK2, PCSK3/Furina, PCSK4, PCSK5, PCSK6, PCSK7, PCSK8/S1P e PCSK9. Enquanto as primeiras sete PCSKs clivam resíduos básicos, as duas últimas clivam em resíduos não básicos, sendo que a PCSK9 possui apenas um substrato conhecido, ela mesma, em uma reação de clivagem auto catalítica para sua ativação (SEIDAH, Nabil G., 2011). A **Figura 1.4** mostra a relação filogenética entre as PCSKs humanas.

O mecanismo de atuação dessas serina proteases ocorre através de um sítio catalítico que possui uma tríade de aspartato, histidina e serina. A especificidade com o substrato se dá pela topologia do bolso de ligação próximo à tríade catalítica (HEDSTROM, 2002).

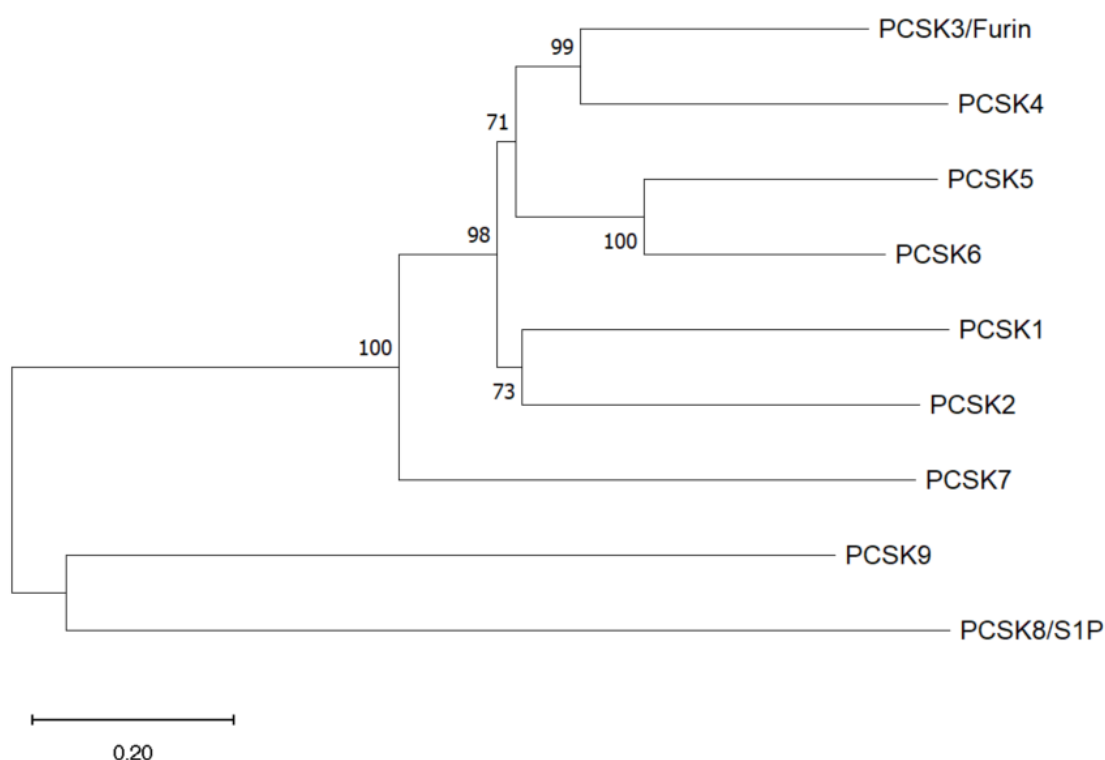


Figura 1.4 – Árvore filogenética da família das PCSKs humanas. A barra de escala refere-se a uma distância filogenética de 0,2 substituições de nucleotídeos por sítio, e os números nas ramificações indicam a porcentagem de *bootstrap* após 1.000 replicações na construção da árvore. Árvore construída utilizando o método estatístico *Neighbor-joining*.

1.3.1 Mecanismo de clivagem

As extensas homologias de sítio ativo entre várias serina proteases indicam que todas elas têm o mesmo mecanismo catalítico (VOET, D.; VOET, J., 2010). O mecanismo catalítico nas serina proteases acontece, principalmente, através da tríade catalítica. A tríade está localizada no sítio ativo da enzima, onde ocorre a catalise da hidrólise da ligação peptídica. A tríade é composta por três aminoácidos: aspartato, histidina e serina. A serina da tríade é quem desempenha o ataque nucleofílico à ligação peptídica, conferindo o nome dessa classe de enzimas de serina proteases (HEDSTROM, 2002; KRAUT, 1977; RADISKY *et al.*, 2006).

Quando um substrato adequado se liga à enzima, há um pequeno ajuste induzido que posiciona a tríade catalítica próxima à ligação peptídica que será clivada (WAHLGREN *et al.*, 2011). Esta mudança conformacional provoca o deslocamento do ácido aspártico, carregado negativamente, em direção ao anel da histidina, rico em elétrons. A carboxila ionizada do aspartato interage por ligação de hidrogênio ao anel da histidina. Esta interação entre o aspartato e a histidina favorece a abstração de um

próton do grupo hidroxila na cadeia lateral da serina, pela histidina. Este processo resulta na produção de um íon alcóxido muito reativo no sítio ativo (AHERN; RAJAGOPAL; TAN, T., 2018; RADISKY *et al.*, 2006; VOET, D.; VOET, J., 2010). A **Figura 1.5** ilustra esse processo.

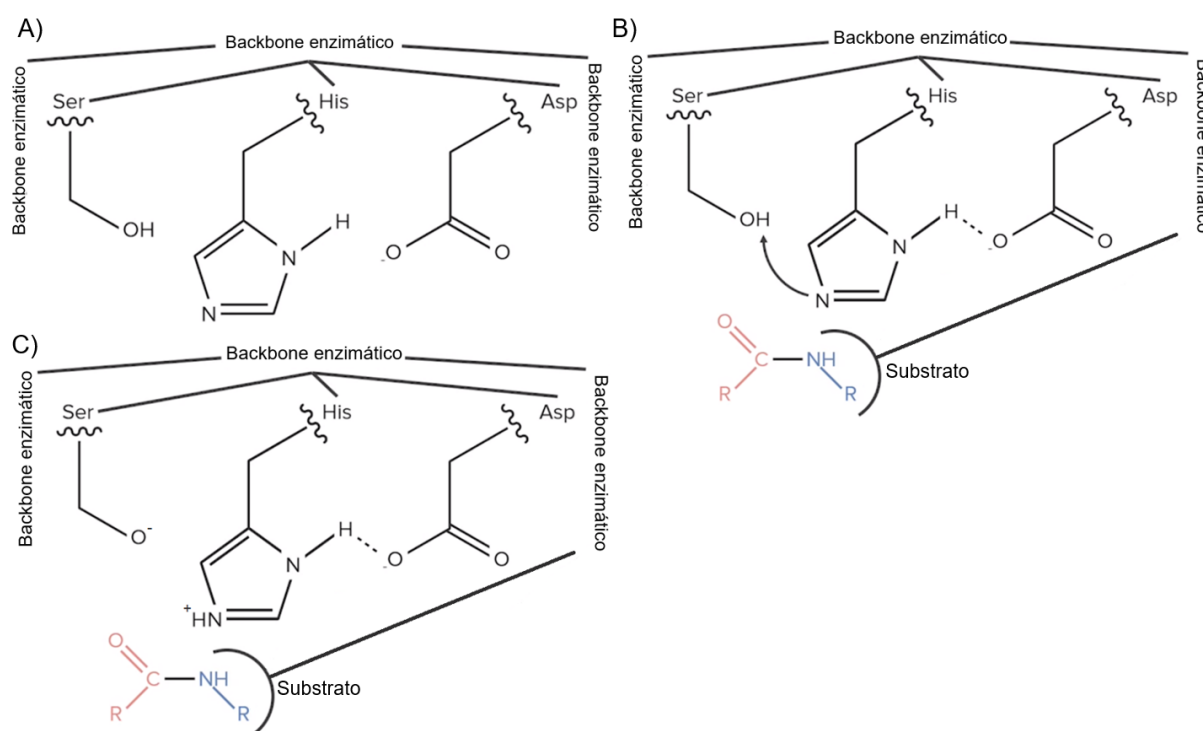


Figura 1.5 - Formação do íon alcóxido nas serina proteases. **A)** Estado inativo da enzima. **B)** Após a ligação do substrato ao bolso de ligação, a carboxila ionizada do aspartato interage por ligação de hidrogênio ao anel da histidina, que então pode aceitar um próton do grupo hidroxila na cadeia lateral da serina. **C)** Enzima ativa, com a formação de um íon alcóxido na cadeia lateral da serina catalítica. Fonte: adaptado de AHERN; RAJAGOPAL; TAN, T., 2018.

As serina proteases clivam ligações peptídicas em duas etapas principais: acilação e desacilação. Na primeira etapa, acilação, a serina nucleofílica ataca a ligação peptídica suscetível de hidrólise, formando um intermediário tetraédrico. Em seguida há a liberação do fragmento C-terminal, culminando na formação de uma acil-enzima covalente. Na segunda etapa, desacilação, uma molécula de água ataca a acil-enzima, levando a um segundo intermediário tetraédrico. Após, há a liberação do fragmento N-terminal e a enzima volta ao seu estado inicial (RADISKY *et al.*, 2006; VOET, D.; VOET, J., 2010; WAHLGREN *et al.*, 2011). A **Figura 1.6** mostra o mecanismo de clivagem.

O sítio ativo conta ainda com um bolsão do oxiânion, que nas PCSKs é formado por uma asparagina, exceto na PCSK2 em que um aspartato desempenha essa função (HEDSTROM, 2002; SEIDAH, Nabil G., 2013). Esse átomo forma um bolsão de carga positiva que ativa a carbonila da ligação peptídica suscetível à hidrólise, e estabiliza o oxiânion carregado negativamente do intermediário tetraédrico (HEDSTROM, 2002; VOET, D.; VOET, J., 2010).

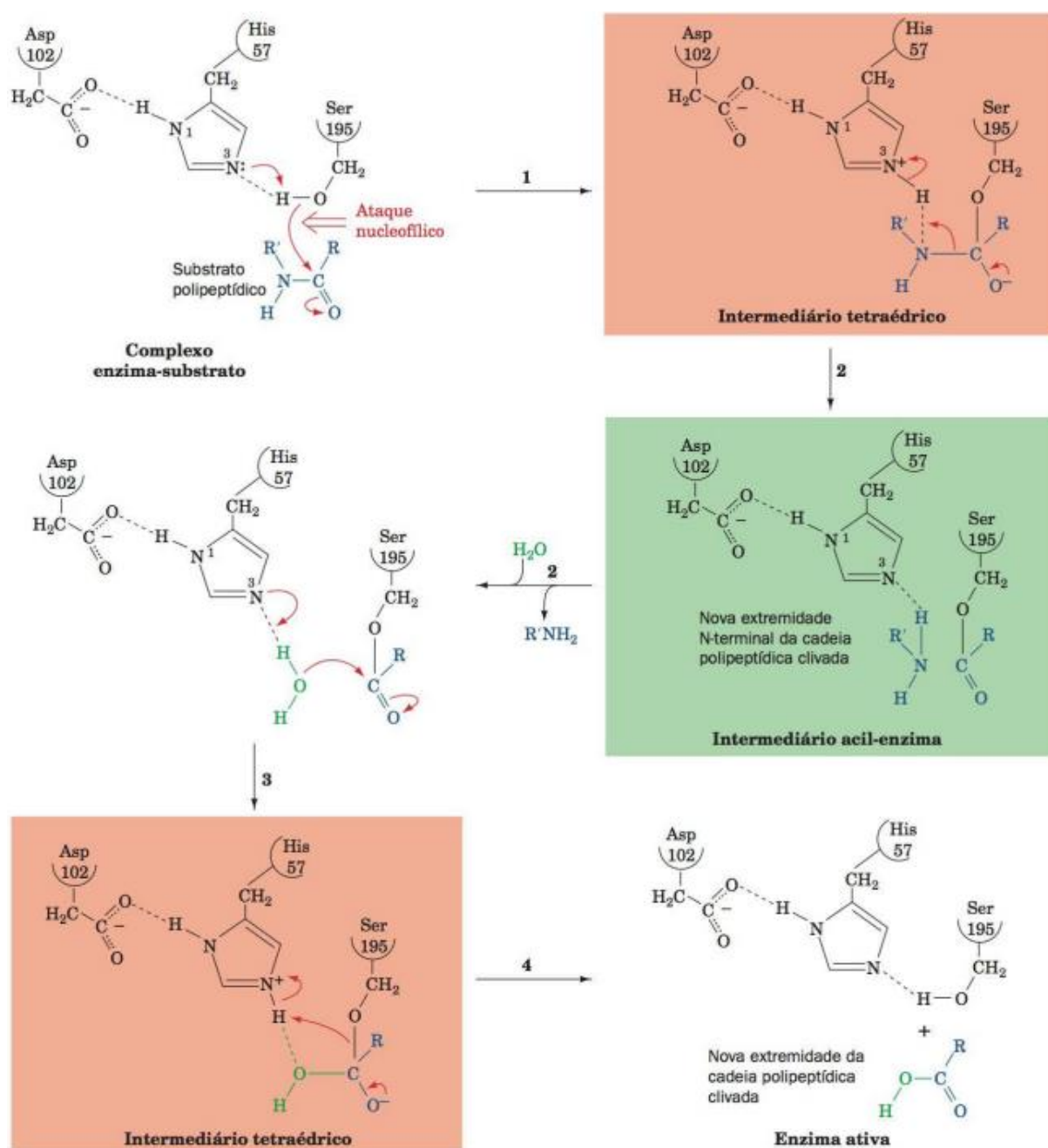


Figura 1.6 - Mecanismo de catálise nas serina proteases. Em (1), ocorre o ataque nucleofílico da serina ionizada sobre o átomo de carbono da carbonila da ligação peptídica que será clivada, levando à

formação de um intermediário tetraédrico. Em **(2)**, a decomposição do intermediário tetraédrico ao intermediário acil-enzima, seguido pela liberação do C-terminal, que será substituído por uma molécula de água. Em **(3)**, a formação de um segundo intermediário tetraédrico. E em **(4)**, a liberação do fragmento N-terminal, com a enzima voltando ao seu estado inicial. Fonte: Figura 15.23 em VOET, D.; VOET, J., 2010.

1.3.2 Bolso de ligação

A especificidade do substrato às enzimas se dá através do bolso de ligação. Os bolsos de ligação das PCSKs, em especial das PCSKs um a sete, são conservados evolutivamente e altamente similares. Dentre as PCSKs, a que possui o bolso mais bem estudado e caracterizado é a Furina (TIAN, S.; JIANHUA, 2010).

A nomenclatura usada para se referir ao bolso de ligação é a nomenclatura proposta por Schechter e Berger, na qual P1 se refere ao aminoácido que está a N-terminal da ligação suscetível à hidrólise e P1' ao aminoácido que está a C-terminal. Os resíduos adjacentes recebem a mesma nomenclatura, apenas incrementando o número a cada resíduo que se distancia da ligação. A nomenclatura dos subsítios do bolso é dada por qual resíduo se liga à cada subsítio. Por exemplo, o resíduo P1 se liga ao subsítio S1, P1' a S1', e assim por diante (SCHECHTER; BERGER, 1967).

Estudos por Becker *et al.* indicam que P1, que usualmente é um resíduo básico, pode realizar ligações com as cadeias laterais dos ácidos aspárticos 306 e 258 da enzima Furina humana. Além disso, P1 pode formar ligações de hidrogênio com a prolina 256 e contactar as carbonilas dos grupos laterais do ácido glutâmico 331 e do ácido aspártico 301 (BECKER *et al.*, 2010). Estudos realizados por Tian mostram que P1 pode interagir com os radicais negativamente carregados dos ácidos glutâmicos 257 e 331 e dos ácidos aspárticos 258, 301 e 306 (TIAN, S., 2009).

A posição P2 também costuma ser ocupada por um resíduo básico, arginina ou lisina, e, portanto, o subsítio de ligação S2, no qual se liga, correlaciona-se com suas características (HEDSTROM, 2002), sendo que, de acordo com Tian, os resíduos mais comuns onde se dá essa interação são os ácidos aspárticos 153, 154 e 191. Ainda segundo Tian, apenas 20% dos substratos das PCSKs possuem resíduos diferentes de arginina ou lisina na posição P2. Nesses casos, no entanto, observa-se que resíduos pequenos como glicina ou alanina ocupam essa posição (TIAN, S., 2009).

Em P3 há menor especificidade quanto ao seu subsítio de ligação, embora haja uma preferência por resíduos carregados positivamente ou apolares como a glicina 255. P4, que também costuma ser um resíduo básico, pode realizar pontes salinas com o ácido glutâmico 236, ácido aspártico 264 e interação com tirosina 308 (BECKER *et al.*, 2010; TIAN, S., 2009).

Os subsítios S1'-Sn' são menos estudados, porque usualmente esses subsítios são inferidos a partir do estudo das interações dessas enzimas com inibidores, que não contêm os resíduos P1'-Pn' ótimos. Portanto, as interações precisas das cadeias laterais P' não se encontram bem definidas (HEDSTROM, 2002). No entanto, Tian indica que P1' pode interagir com as serinas 368 e 253, e também com a asparagina 295 (TIAN, S., 2009).

A **Tabela 1.1**, resume as os resíduos mais comuns de cada subsítio, e que foram citados nesse estudo.

Tabela 1.1 – Resíduos mais comuns nos subsítios do bolso de ligação em Furina humana.

Subsítio	Resíduos em Furina
S1	PRO256, GLU257, ASP258, ASP301, ASP306, GLU331
S2	ASP153, ASP154, ASP191
S3	GLY255
S4	GLU236, ASP264, TYR308
S1'	SER253, ASN295, SER368

A **Figura 1.7** apresenta um alinhamento entre as sequências das PCSKs, destacando os resíduos apresentados na tabela. Como a sequência da PCSK8 é muito diferente das demais, ela foi excluída do alinhamento para fim de simplificação.

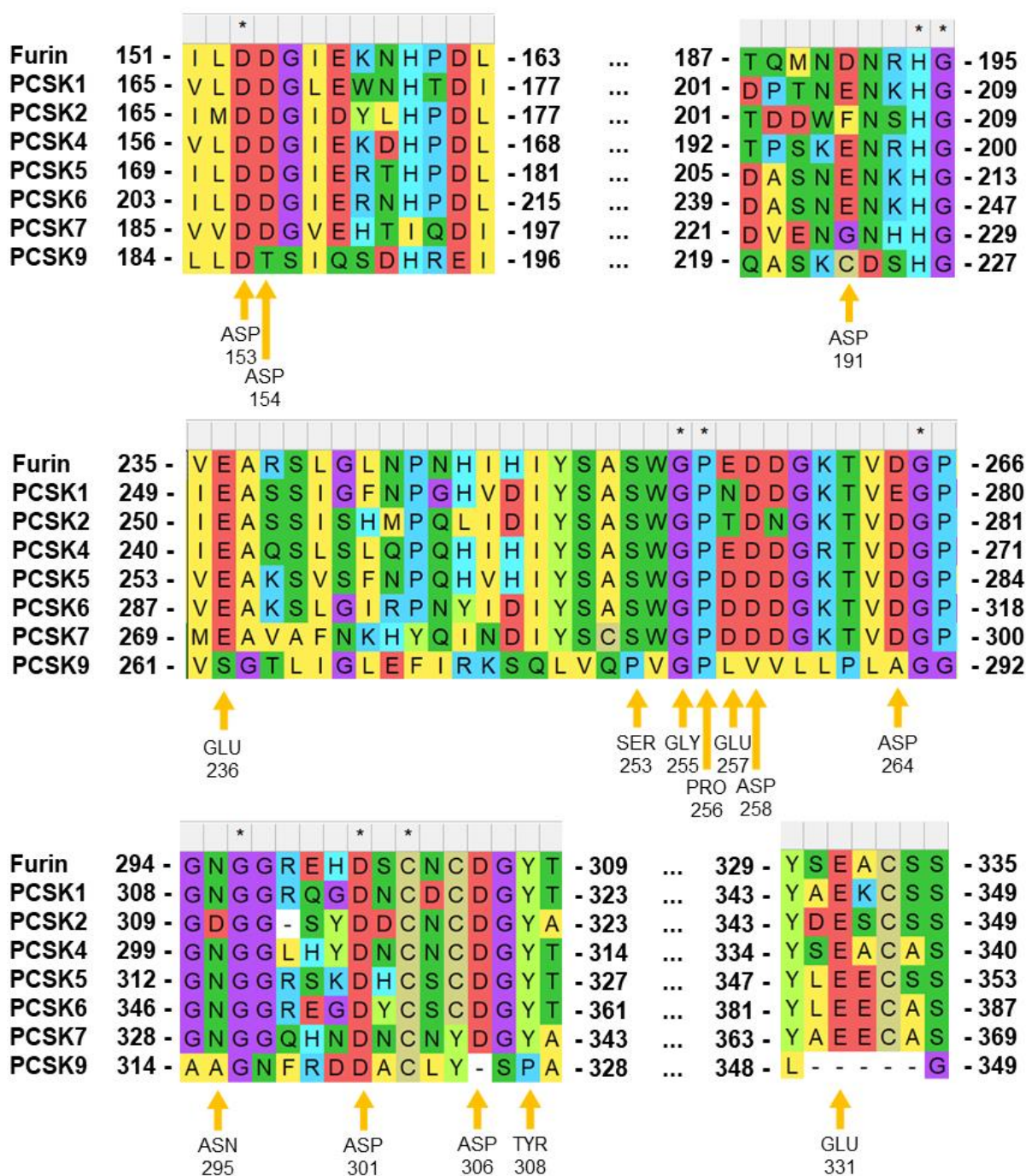


Figura 1.7 – Alinhamento das sequências proteicas de todas as PCSKs, com exceção da PCSK8. As setas destacam os resíduos que, em Furina, são reconhecidos por participar do bolso de ligação.

2 JUSTIFICATIVA

A doença por coronavírus, COVID-19, causou e ainda causa milhares de mortes. Segundo dados da plataforma *Our World In Data* (Our World in Data, 2022), consultados em novembro de 2022, mais de 630 milhões de casos foram confirmados em todo o mundo, e mais de 6,5 milhões de mortes foram causadas pela doença. Só no Brasil foram cerca de 35 milhões de casos e mais de 680 mil mortes. Estes números podem ser ainda maiores devidos há uma subnotificação dos casos e óbitos relacionados à COVID-19 (FRANÇA *et al.*, 2022; WHITTAKER, C. *et al.*, 2021).

Apesar de haver diversas vacinas aprovadas e eficientes, para a doença (Anvisa, 2022), apenas cerca de 65% da população mundial, e 80% da população brasileira, foi vacinada, segundo dados da plataforma *Our World In Data*. A baixa cobertura vacinal na população mundial somada ao fato de que novas variantes do vírus surgem a todo momento, infectando inclusive pessoas vacinadas (LEÓN *et al.*, 2022; VAIDYANATHAN, 2021), atrapalham o combate da doença. Só na primeira quinzena de novembro de 2022 foram confirmados cerca de 150 mil casos da doença só no Brasil, segundo dados da mesma plataforma, e 5 milhões em todo o mundo.

Dessa forma a doença está longe de ser um problema superado, e estudos que visem entender o comportamento do vírus e abrir caminho para novos tratamentos e terapias são de suma importância para mitigar os danos causados pelo patógeno.

Tendo em vista que a enzima Furina humana pode atuar clivando a proteína espícula do vírus, aumentando a infecciosidade viral (ADU-AGYEIWAH; GRANT, M. B.; OBUKHOV, 2020; ESSALMANI *et al.*, 2022; PEACOCK *et al.*, 2021), é essencial que entendamos a nível molecular como essa serina protease se liga ao vírus. Além disso, como a Furina faz parte de uma família de serina proteases, as PCSKs, que muitas vezes compartilham função (REMACLE *et al.*, 2008; SEIDAH, N. G.; CHRÉTIEN; DAY, R., 1994), é necessário saber se alguma outra enzima dessa família pode estar atuando sobre a proteína espícula. Pois, a partir dessas informações, podemos desenvolver pesquisas que visem o desenvolvimento de medicamentos e tratamentos para a doença, tendo como alvo a inibição ou bloqueio da enzima Furina, e de outras PCSKs que podem também estar atuando na clivagem da espícula.

3 OBJETIVOS

3.1 Objetivos gerais

Caracterizar e descrever, a nível molecular, a interação da proteína espícula do novo coronavírus à enzima Furina humana, utilizando metodologias de bioinformática para realizar o acoplamento das duas estruturas e simular o comportamento dinâmico entre elas. Analisar e descrever, da mesma forma, se a proteína viral seria associável a outras serina proteases da família das PCSKs, de modo a se constituir em um potencial substrato.

3.2 Objetivos específicos

- Definir um peptídeo que contenha o sitio polibásico presente na sequência da proteína espícula do SARS-CoV-2, para ser usado em testes computacionais representando a proteína;
- Docar o peptídeo na Furina e nas demais PCSKs;
- Simular o comportamento dinâmico da interação entre as enzimas e o peptídeo;
- Descrever a nível molecular as interações específicas que ocorrem entre o peptídeo e as enzimas durante as simulações de dinâmica molecular;
- Avaliar quais outras PCSKs, além da Furina, têm potencial para atuar clivando a proteína espícula do coronavírus.

4 MATERIAIS E MÉTODOS

Nesta seção apresentaremos a metodologia utilizada nesse estudo, que consiste em: obter as estruturas moleculares das PCSKs, e de um peptídeo que simule a região próxima ao sítio polibásico da espícula do SARS-CoV-2; preparar as estruturas para docagem e simulações de dinâmica molecular; realizar a docagem do peptídeo nas enzimas utilizando diferentes metodologias; realizar as simulações de dinâmica molecular a partir dos resultados da docagem; analisar o comportamento e interações das estruturas durante as simulações. O fluxograma da **Figura 4.1** apresenta essa metodologia.

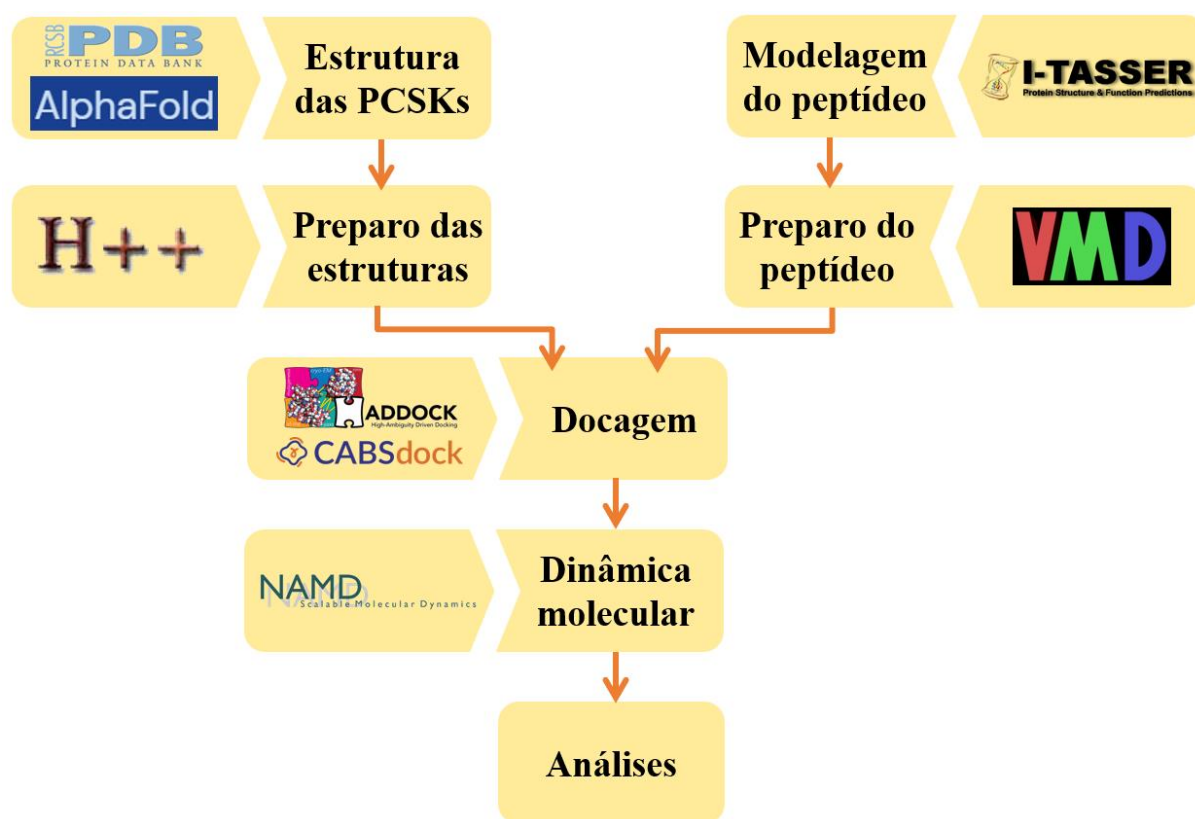


Figura 4.1 - Fluxograma da metodologia

4.1 Estruturas proteicas

As estruturas proteicas das enzimas, para as quais foram encontrados dados cristalográficos, foram baixadas do banco de dados de proteína RCSB PDB (BERMAN et al., 2000; BURLEY et al., 2021; “RCSB PDB: Homepage”). Para as enzimas que não possuíam esses dados, foram baixados modelos obtidos em dois bancos de dados diferentes, e realizada a comparação entre os modelos. Os bancos usados foram o AlphaFold Protein Structure Database do EBI, onde as proteínas foram modeladas com AlphaFold2 (AlphaFold, 2022; JUMPER et al., 2021; VARADI et al., 2022) e o SWISS-MODEL (BIENERT et al., 2017; SWISS-MODEL, 2022; WATERHOUSE et al., 2018).

Para a proteína espícula do SARS-CoV-2, foram utilizados peptídeos contendo a sequência do sítio de clivagem polibásico, e os aminoácidos do entorno. As sequências utilizadas serão apresentadas na seção 5.2.2. A modelagem das estruturas dos peptídeos foi feita utilizando a ferramenta I-TASSER através do seu servidor web (I-TASSER, 2022; YANG et al., 2014; ZHENG et al., 2021), usando como entrada a sequência a ser modelada.

4.2 Preparo das estruturas para docagem

4.2.1 Preparo das enzimas para docagem

Antes de realizar a docagem foi feita a protonação dos modelos e das estruturas cristalográficas das enzimas. Essa protonação tem como objetivo ajustar os valores de pKa dos grupos ionizáveis da molécula, de acordo com um pH preestabelecido. Para este estudo usamos o pH sanguíneo humano, considerado igual a 7,35 (HOPKINS; SANVICTORES; SHARMA, 2022).

Para fazer essa protonação usamos a ferramenta H++ versão 4.0 disponível em endereço web (GORDON et al., 2005; H++, 2022). Para os parâmetros, utilizamos pH igual a 7,35, salinidade igual a 15% (CHERNECKY CC; BERGER BJ, 2013), e contínuo dielétrico igual a 80 (DEMCHUK; WADE, 1996).

4.2.2 Preparo dos peptídeos para docagem

A preparação das estruturas do peptídeo consiste em neutralizar a carga dos terminais do modelo. Como esse modelo de peptídeo deve representar um trecho da proteína, é importante que as cargas dos resíduos nas pontas do mesmo correspondam às aquelas encontradas na proteína usada como referência.

Para neutralizar essas cargas, usamos o VMD (HUMPHREY; DALKE; SCHULTEN, 1996) e rescrevemos a estrutura do peptídeo adicionando um *patch* aos resíduos terminais, que faz a adição e/ou modificação de átomos para alterar as cargas. No N-terminal usamos o *patch* ACE, enquanto que no C-terminal foi usado o CTER.

4.3 Docagem

Para o encaixe do peptídeo às PCSKs foram utilizadas duas metodologias diferentes de docagem. Uma utilizando o programa Haddock, e a outra utilizando o CABSdock.

4.3.1 Haddock

Haddock é uma aplicação de docagem orientada por informações para a modelagem de complexos biomoleculares (HONORATO *et al.*, 2021; ZUNDERT, VAN *et al.*, 2016). Foi utilizada a versão 2.4 disponibilizada através do servidor web (HADDOCK, 2022).

Para cada enzima, efetuamos a docagem usando como entrada o seu modelo ou estrutura cristalográfica, e o modelo do peptídeo. Ao carregar as estruturas no servidor, selecionamos a opção que correspondesse ao tipo de estrutura submetida: proteína ou peptídeo. Nos parâmetros de entrada, definimos como resíduos ativos os resíduos que compõem o sítio de clivagem e o bolso de ligação, para as enzimas, e todos os resíduos, para o peptídeo. Os resíduos passivos foram definidos como os resíduos até 6,5 Å dos ativos. Os parâmetros de docagem e de análise utilizados foram os padrões do servidor.

4.3.2 CABSdock

CABSdock é uma aplicação para acoplamento molecular de peptídeos a proteínas, que permite docagem flexível (BLASZCZYK *et al.*, 2019; KURCINSKI *et al.*, 2020). Para realizar os experimentos, utilizamos a versão 0.9.18 em formato *standalone* (KURCINSKI *et al.*, 2019), ou seja, o programa foi instalado e executado em um computador local.

Para cada PCSK, a docagem foi realizada usando como entrada o modelo ou a estrutura cristalográfica da mesma, e o modelo do peptídeo. Para os parâmetros de entrada, definimos a conformação do peptídeo como aleatória, permitindo que o programa gerasse conformações diferentes para a estrutura durante a docagem. E para o local de docagem, utilizamos um centro geométrico formado entre os resíduos serina, aspartato e histidina do sítio catalítico mais o bolsão de oxiânion de cada enzima.

4.4 Simulações de dinâmica molecular

Para preparar os complexos moleculares, escolhidos após a docagem, para a simulação de dinâmica molecular utilizamos o VMD (HUMPHREY; DALKE; SCHULTEN, 1996). Primeiro geramos os arquivos de estrutura proteica usando os parâmetros de topologia CHARMM 36 (BEST *et al.*, 2012). Em seguida, fizemos a solvatação do complexo. Para representar as moléculas de água foi usado o modelo TIP3W (MACKERELL, A. D. *et al.*, 1998).

Para realizar as simulações de dinâmica molecular utilizamos o programa NAMD (PHILLIPS *et al.*, 2020). As simulações de dinâmica foram executadas usando os parâmetros de topologia CHARMM 36 e temperatura de 310 K. Para controle de temperatura e equilíbrio foi usado termostato de Langevin, na qual amortecimento adicional e forças aleatórias são introduzidos no sistema. O controle de pressão foi feito utilizando o barostato Nose-Hoover Langevin (FELLER *et al.*, 1998), empregando uma pressão de 1 atm. O campo eletrostático usado foi malha de partículas de Ewald (PME) (EWALD, 1921). Foi empregado condições de fronteira periódicas (PBC) considerando o tamanho de cada complexo.

Primeiro foram executados 1000 passos de minimização de energias. Em seguida uma etapa de equilíbrio de temperatura e pressão por 2ns. A etapa de

produção foi executada por um período de 300ns, com um integrador de tempo de 2fs e gravada a cada 10ps.

As simulações foram executadas em três diferentes servidores, todos equipados com processador Ryzen 7 de 16 threads e acelerados por GPU RTX 2060 ou RTX 3060 NVIDIA. O tempo de execução observado foi de 0,0016 segundos por passo no mínimo, e um tempo médio de 0,0024 segundos por passo, totalizando um tempo de no mínimo 2,88 dias, e uma média de 4,21 dias por simulação.

O arquivo completo com as configurações do NAMD utilizadas encontra-se disponível no **APÊNDICE A**.

4.5 Análises

4.5.1 RMS

A raiz quadrada do desvio quadrático médio (RMSD, na sigla em inglês) é a medida do desvio médio entre átomos correspondentes de duas estruturas (KUFAREVA; ABAGYAN, 2012). O RMSD *Fit* foi medido através da superimposição das estruturas que serão comparadas, desconsiderando sua movimentação espacial, de forma a obter o menor valor possível. Enquanto que para o RMSD *no Fit*, fizemos um alinhamento do complexo molecular inteiro, medindo também a movimentação espacial no complexo da estrutura do peptídeo em cada intervalo em relação à pose inicial da simulação.

As análises de RMSD foram feitas utilizando o programa VMD (HUMPHREY; DALKE; SCHULTEN, 1996), através do comando *measure rmsd* (VMD measure documentation, 2022). O cálculo de RMSD é feito conforme a **Equação 4.1**. Para os cálculos foram considerados apenas os átomos da cadeia principal dos resíduos.

$$\sqrt{\frac{\sum_{i=1}^N (r_i(t_1) - r_i(t_2))^2}{N}}$$

Equação 4.1 – Fórmula do RMSD, na qual N é o número de átomos que estão sendo comparados e $r_i(t)$ é a posição do átomo i no tempo t.

A raiz quadrada da flutuação quadrática média (RMSF, na sigla em inglês) é a média do RMSD por tempo de uma estrutura. Enquanto o RMSD mede o quanto uma estrutura diverge de uma referência ao longo do tempo, o RMSF pode revelar quais dos seus resíduos são mais móveis. O RMSD é calculado com relação a um estado inicial, já o RMSF é calculado para a estrutura média da simulação.

Os cálculos RMSF foram realizados conforme representado pela **Equação 4.2**. Para medir a distância entre as posições de um resíduo, foi considerado o centro de massa entre os átomos de sua cadeia principal.

$$\sqrt{\frac{\sum_{i=1}^N (r_i(t_1) - \langle r_i \rangle)^2}{N}}$$

Equação 4.2 – Fórmula do RMSF, na qual N é o número de posições que estão sendo comparadas, $r_i(t)$ é a posição do átomo i no tempo t e $\langle r_i \rangle$ é a posição média do átomo i.

4.5.2 Contatos

Para medir todos os contatos não-covalentes entre átomos foi usado o comando *measure contacts* do VMD (VMD measure documentation, 2022). São considerados contatos desse tipo todos os pares de átomos que se encontram a uma distância menor do que o *cutoff* estabelecido. A distância de *cutoff* usada foi igual a 3,3 Å.

Para medir as interações não-covalentes do tipo pontes salinas também usamos o comando *measure contacts* com a mesma distância de *cutoff*. Porém, foram filtrados apenas os pares de átomos pesados e que estejam em resíduos com cargas opostas.

Os contatos não-covalentes do tipo interações de hidrogênio foram calculados utilizando o comando *measure hbonds*. Para se considerar essas interações é preciso que pares de átomos eletronegativos estejam a uma distância menor do que o *cutoff* estabelecido, e o ângulo formado entre eles seja menor do que o valor especificado. Usamos uma distância de *cutoff* igual a 3,3 Å, e o ângulo igual a 30°.

4.5.3 Distâncias

As distâncias entre átomos foram medidas conforme a **Equação 4.3**.

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2}$$

Equação 4.3 – Fórmula de distâncias entre dois átomos i e j, onde x, y e z representam suas coordenadas cartesianas.

4.5.4 SASA e BSA

A área de superfície acessível ao solvente (SASA, na sigla em inglês) é medida utilizando o comando *measure sasa* do VMD (VMD measure documentation, 2022). O cálculo da área acessível ao solvente é realizado estendendo um raio a partir de cada átomo para encontrar os pontos que estão expostos em uma esfera. O raio utilizado nesse estudo foi igual a 1,4 Å.

O cálculo da área de superfície encoberta (BSA, na sigla em inglês) é o complementar do SASA. Para essa medida, calculamos a área total da estrutura e depois subtraímos a área exposta. Para medir o quanto do peptídeo está encoberto exclusivamente pela enzima na simulação, subtraímos do valor total de BSA o valor de BSA do peptídeo desconsiderando a presença enzima.

4.5.5 Energias

As energias do complexo e energias de interação entre duas seleções foram calculadas usando o plugin *namdenergy* (NAMD, 2022) do VMD. Para calcular as energias de interação de cada complexo foi utilizado uma seleção contendo a enzima e outra com o peptídeo. Os parâmetros de topologia utilizados foram CHARMM 36 (BEST *et al.*, 2012).

5 RESULTADOS

No presente estudo, definimos e modelamos um peptídeo que cobrisse a região da proteína espícula do SARS-CoV-2 onde houve a inserção do sítio polibásico, que tem ocorrência exclusiva no coronavírus que infecta humanos e que pode ser substrato da enzima Furina. E testamos, computacionalmente, a afinidade desse peptídeo a essa enzima, e às demais enzimas da família das PCSKs. Inicialmente descreveremos os resultados obtidos desde a definição do sítio de clivagem, passando pela modelagem das estruturas de interesse, e a docagem do peptídeo às enzimas. Posteriormente serão apresentados os resultados das simulações de dinâmica molecular analisadas programaticamente através de uma ferramenta desenvolvida nesse trabalho.

5.1 Definição da ligação suscetível à hidrólise

Para podermos estudar a afinidade das enzimas PCSKs à proteína espícula do coronavírus, é fundamental sabermos em qual ligação peptídica do sítio polibásico deve ocorrer a clivagem. Isto porque, conforme descrito na seção 1.3.2, cada resíduo da proteína interage com resíduos específicos do bolso de ligação de acordo com sua posição em relação à ligação suscetível à hidrólise.

Com base nos estudos de Dahms *et al.* e Remacle *et al.*, e principalmente no estudo de Tian, que estabelecem a sequência padrão de clivagem dos substratos das PCSKs como sendo R-X-(R/K/X)-R↓-S (DAHMS *et al.*, 2014; REMACLE *et al.*, 2008; TIAN, S., 2009), podemos inferir que a ligação peptídica mais provável de ocorrer a catalise é a ligação a C-terminal da arginina 704 (ARG704). Esta arginina está presente no sítio de clivagem polibásico exclusivo de SARS-CoV-2, podendo formar a sequência RRAR↓S. Sequência que é a que mais se assemelha ao padrão de clivagem dos substratos descrito por esses autores.

Essa previsão foi ainda reforçada utilizando a ferramenta web PROP 1.0 (DUCKERT; BRUNAK; BLOM, 2004; ProP, 2022), que emprega aprendizado de máquina para prever o local de clivagem por Furina em uma dada sequência. As análises da região de interesse com o programa indicam um alto potencial de clivagem na ARG704, e indica também um baixo potencial nas argininas 702 e 701 (**Figura 5.1**).

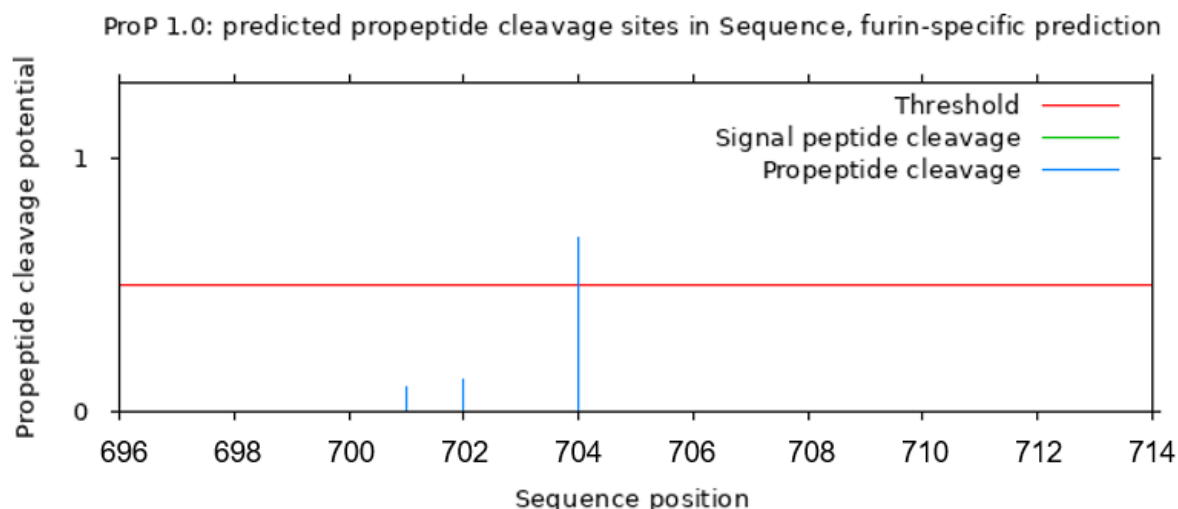


Figura 5.1 – Gráfico de potencial de clivagem por resíduo. Gráfico gerado pela ferramenta web PROP 1.0 mostrando um pico de potencial acima do *threshold* para ARG704 e um baixo potencial para as argininas 702 e 701. Fonte: Adaptado do resultado da ferramenta ProP, 2022

Por fim, artigos mais recentes, como o publicado em abril de 2021 por Peacock *et al.* também definiram o mesmo resíduo como sendo o alvo da clivagem (PEACOCK *et al.*, 2021), reforçando os resultados apresentados aqui.

5.2 Modelos

5.2.1 Modelos das enzimas

Para as enzimas Furina e PCSK9 foram encontrados dados cristalográficos definindo suas formas 3D no banco de dados proteicos RCSB PDB (BURLEY *et al.*, 2021; RCSB PDB, 2022). Para a enzima Furina foi utilizada a entrada 4OMC (DAHMS *et al.*, 2014), com resolução de 2,3 Å. E para a PCSK9 foi utilizada a entrada 2PMW (PIPER *et al.*, 2007), também com uma resolução de 2,3 Å.

Para as demais PCSKs não foram encontradas estruturas cristalográficas e, portanto, foi necessário recorrer à modelagem computacional dessas enzimas. Ao consultar o banco de dados UniProt (BATEMAN *et al.*, 2021) para essas enzimas, são listados dois algoritmos de modelagem diferentes, o AlphaFold2 (JUMPER *et al.*, 2021; VARADI *et al.*, 2022) e o SWISS-MODEL (BIENERT *et al.*, 2017; WATERHOUSE *et al.*, 2018). A enzima PCSK8, por ser muito diferente das demais, incluindo a relação com o tipo de substrato clivado (SEIDAH, Nabil G., 2011), e por

não terem sido encontrados modelos satisfatórios da mesma, não foi estudada no presente trabalho.

Para escolher entre essas duas ferramentas calculamos o RMSD dos resíduos do sítio catalítico e do bolso de ligação entre modelos gerados pelas duas metodologias. Notamos que o RMSD entre os modelos fica abaixo de 0,5 Å para todas as enzimas, conforme pode ser observado na **Tabela 5.1**.

Tabela 5.1 – RMSD do sítio catalítico e de ligação medido entre modelos das PCSKs criados usando diferentes metodologias.

Enzima	RMSD (Å)
PCSK1	0,45
PCSK2	0,41
PCSK4	0,47
PCSK5	0,32
PCSK6	0,47
PCSK7	0,30

Tendo em vista que a diferença estrutural da região de interesse é baixa, e que o AlphaFold2 é uma metodologia mais recente e que obteve o melhor score de predição de proteína no congresso de avaliação crítica da previsão de estrutura (CASP) (Prediction Center, 2022; PEREIRA et al., 2021; SIMPKIN et al., 2021), optamos por usar os modelos gerados pelo AlphaFold2.

5.2.2 Modelos dos peptídeos

Para testar a ligação das PCSKs ao sítio de clivagem da proteína espícula do coronavírus preparamos um peptídeo contendo o sítio de clivagem polibásico onde deve ocorrer a clivagem. Por se tratar de um trecho que não possui estrutura tridimensional cristalográfica, foi preciso criar os modelos computacionalmente. A modelagem do peptídeo foi realizada utilizando I-Tasser (YANG, J. et al., 2014; ZHENG, W. et al., 2021).

Com base na sequência da região sem estrutura cristalográfica da proteína foram testados peptídeos de dois tamanhos diferentes: com dez e com treze resíduos (**Figura 5.2**). O peptídeo de treze resíduos cobre a maior parte dessa região, e sua predição aponta para a formação de uma alfa-hélice em sua estrutura, com mais de uma volta completa, o que não ocorre com o modelo de apenas dez resíduos (**Figura 5.3**).

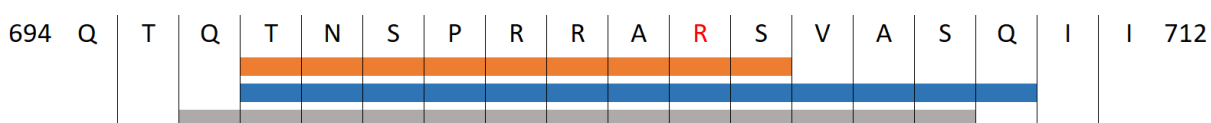


Figura 5.2 – Sequência dos peptídeos criados para teste. A área em laranja representa o peptídeo de dez aminoácidos. Em azul, o peptídeo de treze aminoácidos. E em cinza, a região da proteína que não possui estrutura tridimensional cristalográfica. A arginina que está a N-terminal da ligação que provavelmente será clivada está em vermelho.

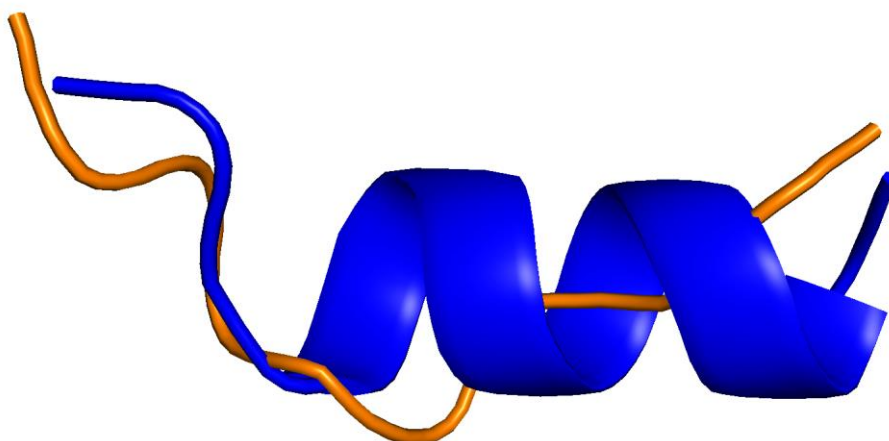


Figura 5.3 – Modelo dos peptídeos de dez e treze aminoácidos. Em laranja, o peptídeo de dez, e em azul o de treze. A predição de uma alfa-hélice ocorre apenas no peptídeo de treze aminoácidos.

Para a escolha do melhor modelo foi feita uma docagem preliminar de ambos, utilizando o programa Haddock (ZUNDERT, VAN *et al.*, 2016), na enzima Furina para observar o possível encaixe entre as moléculas. No modelo de dez resíduos o *score* de energia medido pelo Haddock foi de -65,0 e a ARG704 (P1) não ficou posicionada dentro do bolso de ligação S1. Já no modelo de treze resíduos o *score* medido pelo

Haddock foi de -83,7 e a ARG704 ficou posicionada dentro do bolso de ligação S1 podendo interagir com os resíduos do mesmo (**Figura 5.4**).

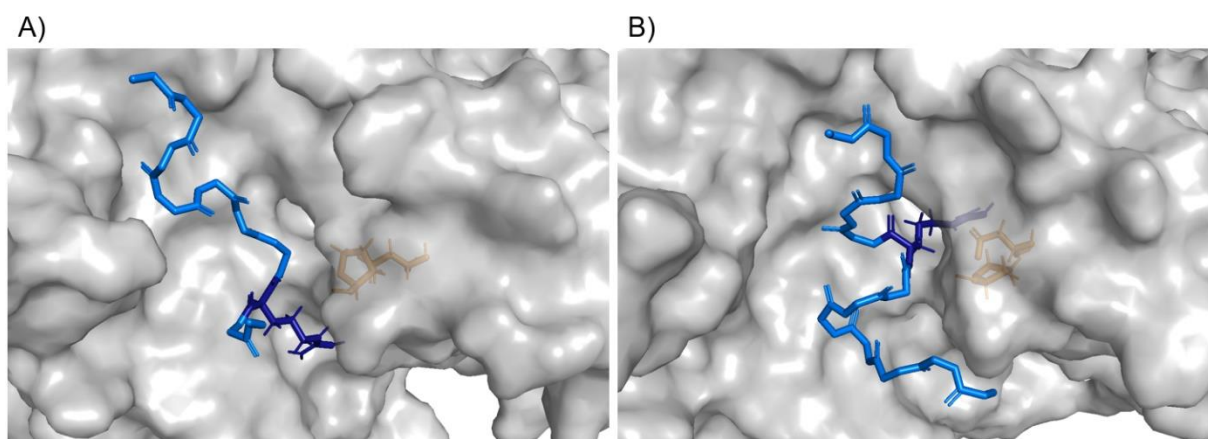


Figura 5.4 – Modelos das docagens com os peptídeos de dez e treze aminoácidos. A) Modelo de dez resíduos. B) Modelo de treze resíduos. A arginina que está a N-terminal da ligação que possivelmente sofre hidrólise (azul escuro) fica posicionada interagindo com alguns dos resíduos do bolso de ligação S1 (laranja) apenas na docagem com o modelo de treze resíduos (B).

Considerando que o modelo de treze resíduos obteve um score no Haddock mais negativo e, portanto, melhor, teve um posicionamento mais favorável da ARG704, e o modelo mostra uma estruturação em forma de alfa hélice; optamos por usar o modelo de treze resíduos.

Tendo estabelecido a sequência que será utilizada, deste ponto em diante usaremos a nomenclatura apresentada na seção 1.3.2 para nos referirmos aos resíduos do peptídeo. Nela, P1 representa a ARG704, ou seja, o resíduo na qual esperamos que ocorra o reconhecimento para a clivagem. A **Figura 5.5** mostra a sequência do peptídeo utilizando a nomenclatura de Schechter e Berger (SCHECHTER; BERGER, 1967).

T	N	S	P	R	R	A	R	S	V	A	S	Q
P8	P7	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'

Figura 5.5 – Sequência e nomenclatura do peptídeo criado. A arginina que está a N-terminal da ligação que possivelmente será clivada é chamada de P1 e está representada em vermelho.

5.3 Docagem

Para a docagem, a preparação das estruturas cristalográficas ou modelos preditos das enzimas foi feita conforme descrito na seção 4.2.1, nela fazemos a protonação da molécula e ajustamos os valores de pKa de grupos ionizáveis de acordo com o pH sanguíneo. Para o peptídeo, a preparação foi feita conforme descrito na seção 4.2.2, com o objetivo de neutralizar as cargas nos terminais do modelo. Dessa forma garantimos uma simulação mais fidedigna de um trecho da proteína, no qual essas cargas terminais não estariam presentes.

A docagem do peptídeo à enzima foi feita utilizando duas metodologias diferentes: com os programas Haddock (ZUNDERT, VAN *et al.*, 2016) e CABSdock (BLASZCZYK *et al.*, 2019; KURCINSKI *et al.*, 2020). Nas docagens utilizando o Haddock o peptídeo foi docado de forma rígida, isto é, mantendo em maior parte a conformação estrutural gerada na modelagem com o I-Tasser, embora possam ocorrer algumas adaptações estruturais no encaixe à enzima. Já na docagem com CABSdock, optamos por uma docagem flexível, permitindo que o peptídeo assumisse uma posição aleatória durante a docagem. Como consequência, o peptídeo pôde adotar conformações diferentes entre as duas ferramentas de docagem, conforme podemos observar na **Figura 5.6**. A escolha de duas metodologias diferentes foi feita para tentar maximizar nossas chances de obter a melhor docagem do peptídeo ao sítio catalítico da enzima.

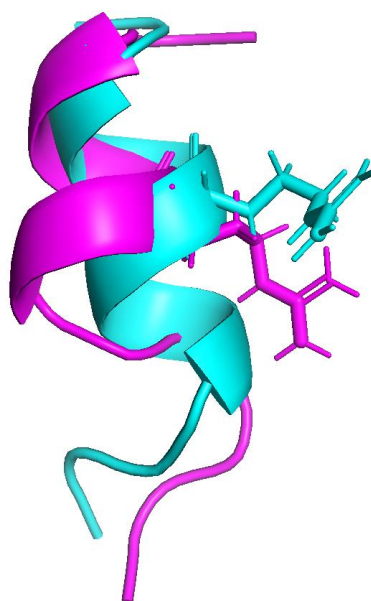


Figura 5.6 – Modelo dos peptídeos após docagem com Haddock (ciano), selecionado por melhor *score*, e com CABSdock (rosa), selecionado pela proximidade ao bolso de ligação. A cadeia lateral de P1 está representada em *sticks*. Nota-se diferenças estruturais entre os resultados após docagem com ferramentas diferentes.

Nas docagens realizadas com Haddock, foi utilizado o *score* gerado pela própria ferramenta para escolhermos os complexos. O valor de *score* consiste em uma combinação linear de várias energias e área de superfície enterrada (BSA), e valores mais negativos são considerados melhores (VANGONE *et al.*, 2017). Os valores de *score* obtidos em todas as docagens podem ser conferidos na **Tabela 5.2**.

Tabela 5.2 – Tabela de *score* obtido nas docagens do peptídeo às enzimas utilizando Haddock.

Enzima	Haddock Score
Furina	-91,2
PCSK1	-77,1
PCSK2	-110,1
PCSK4	-82,1
PCSK5	-77,1
PCSK6	-82,8
PCSK7	-90,6
PCSK9	-74,0

Já nas docagens realizadas com CABSdock, como não há um *score* definido pela ferramenta, buscamos pelos modelos no qual P1 mais se aproximou do bolso de ligação S1, para escolhermos os complexos.

5.4 Simulações de dinâmica molecular

Para aprofundar o estudo da interação entre a enzima e o peptídeo, realizamos simulações de dinâmica molecular a partir dos modelos obtidos com a docagem realizada tanto pelo Haddock quanto pelo CABSdock. As simulações foram rodadas por um tempo de 300ns utilizando o programa NAMD, conforme a metodologia apresentada na seção 4.4.

O objetivo dessas simulações é avaliar a estabilidade do complexo formado a partir da docagem, na presença de solvente, tentando simular uma situação mais próxima daquela que seria observada *in vivo*. Os contatos do bolso de ligação com o peptídeo foram analisados computacionalmente e também manualmente ao longo das simulações.

Os resultados das análises das simulações de dinâmica molecular serão apresentados a seguir. De maneira geral, as simulações com a Furina e PCSK2 a partir da docagem com Haddock foram as únicas que aparentaram uma estabilidade do peptídeo e interações de P1 com diversos resíduos do bolso de ligação S1, em especial com o aspartato no fundo do subsítio. As análises dessas simulações serão apresentadas de forma mais detalhada, enquanto para as outras nas quais não observamos interações interessantes serão mostrados de maneira resumida, apresentando apenas os resultados que julgamos mais relevantes para a conclusão. Todos os resultados das análises de todas as simulações estão disponíveis no **APÊNDICE B**.

5.4.1 Furina

Analisando os gráficos de RMSD do peptídeo nas simulações com a enzima Furina, notamos que houve uma estabilidade na sua movimentação após algum tempo de simulação. Na simulação rodada a partir da docagem com Haddock, por exemplo, essa estabilidade ocorre após cerca de 100ns, sendo observável pela

formação de um platô dos valores de RMSD em torno de 7 Å (**Figura 5.7**). A formação desse platô por cerca de 200ns até o final da simulação, nos indica que o peptídeo se acomodou em uma posição possivelmente estável.

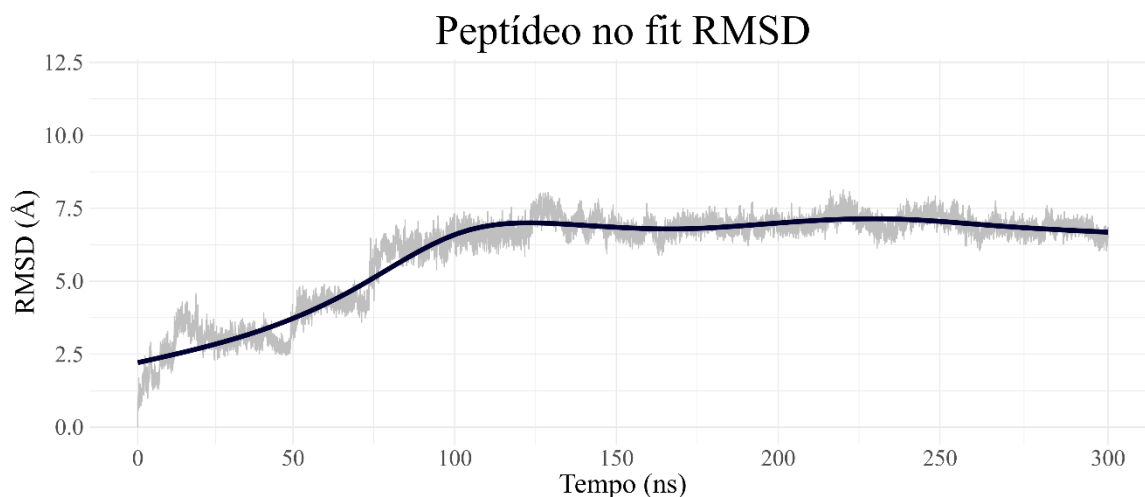


Figura 5.7 – *No fit* RMSD durante a simulação do peptídeo docado com Haddock à enzima Furina. Nota-se uma estabilidade em 7 Å a partir dos 100ns.

Vale ressaltar que no cálculo do *no fit* RMSD não é feita uma superimposição da estrutura do peptídeo em cada *frame* à sua conformação inicial, mas sim uma superimposição de todo o complexo à posição do primeiro *frame*. Dessa forma as distâncias calculadas refletem também a movimentação espacial do peptídeo com relação ao complexo, nesse caso, um valor de 7 Å até uma posição estável. Diferentemente do *fit* RMSD, no qual é feita essa superimposição da estrutura do peptídeo ao primeiro *frame*, permitindo-nos observar apenas a mudança conformacional da cadeia principal da estrutura desconsiderando sua movimentação espacial. Os dados de *fit* RMSD, nessa simulação, indicam que houve um ajuste da conformação da cadeia principal do peptídeo, que se estabiliza em torno de um valor 6 Å após cerca de 100ns de simulação.

Ao explorar os modelos ao final da simulação realizada a partir da docagem com Haddock, podemos ver que P1, a arginina do peptídeo que está a N-terminal da ligação que possivelmente sofrerá a ação da serina catalítica da enzima (SER368), ficou posicionada podendo interagir com os resíduos PRO256, ASP258, ASP306 e GLU331 do bolso de ligação S1 (**Figura 5.8**).

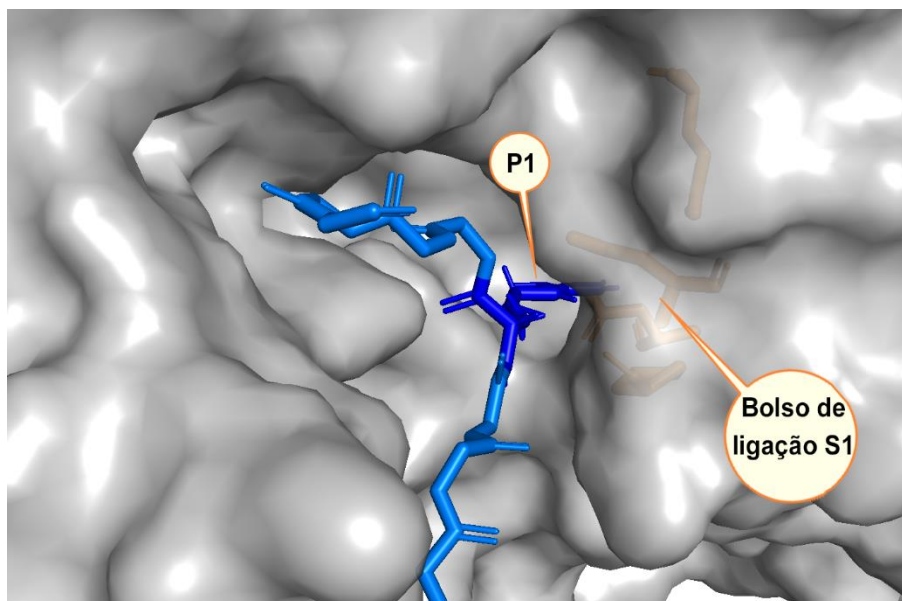


Figura 5.8 – Modelo ao final da simulação do peptídeo docado na Furina usando Haddock. Observe que P1 (azul escuro) encontra-se próximo aos resíduos do bolso de ligação S1 (laranja).

Para ter uma ideia do quanto P1 está inserido no subsítio S1 podemos medir a taxa de sua área de superfície que não está acessível ao solvente ao longo da simulação, a área de superfície encoberta (BSA). Ao analisar esses valores para P1 durante a simulação do peptídeo docado pelo Haddock, observa-se que cerca de 85% a 95% da sua área total está encoberta e inacessível ao solvente de forma constante, ocorrendo uma pequena modificação ao final da simulação. Ao repetir os cálculos desconsiderando a enzima, notamos que o valor de BSA cai para cerca de 35% da área total de P1, indicando que cerca de 60% da sua área está sendo encoberta exclusivamente pela Furina. Esse valor se mantém durante a maior parte da simulação, confirmando que P1 se alocou de forma estável no fundo do bolso de ligação S1 (**Figura 5.9**).

Ao analisar os contatos que podem ser realizados entre o peptídeo e a enzima, notamos que a cadeia lateral da arginina em P1 pode, possivelmente, formar pontes salinas com dois aspartatos (ASP258 e ASP306) do subsítio S1 do bolso de ligação (**Figura 5.10**). Pontes salinas são interações entre dois grupos de cargas opostas em que pelo menos um par de átomos pesados está dentro da distância de ligação de hidrogênio. Por ser uma interação entre átomos carregados é energeticamente mais alta do que interações entre átomos parcialmente carregados, tornando as pontes salinas mais fortes do que outras interações não-covalentes, como interações de

hidrogênio, interações eletrostáticas e interações polares (HENDSCH; TIDOR, 1994; KUMAR; NUSSINOV, 1999; SINDELAR; HENDSCH; TIDOR, 1998).

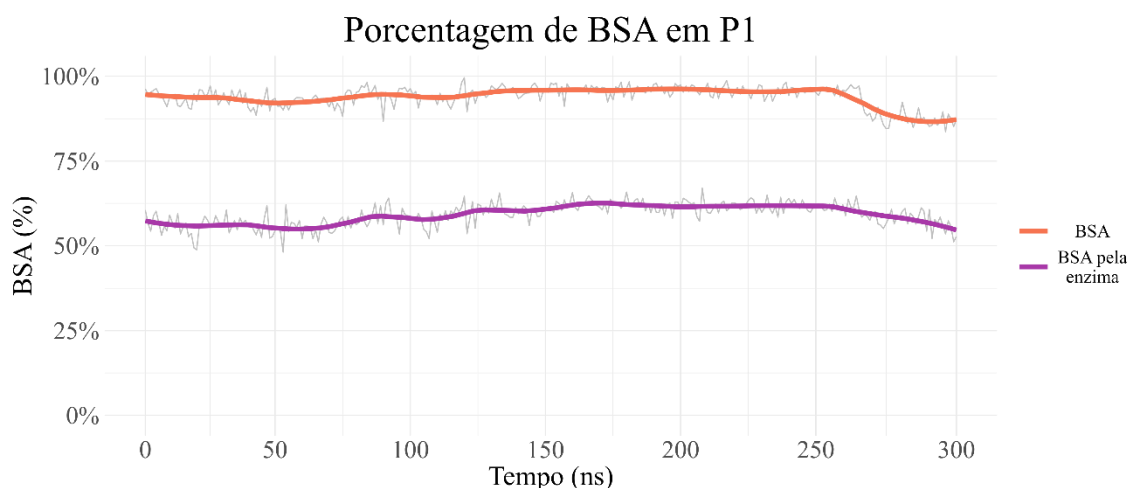


Figura 5.9 – Gráfico de porcentagem de BSA de P1 durante a simulação com a enzima Furina a partir da docagem com Haddock. Em laranja a porcentagem total de área encoberta de P1, e em roxo a porcentagem de área de P1 encoberta apenas pela Furina.

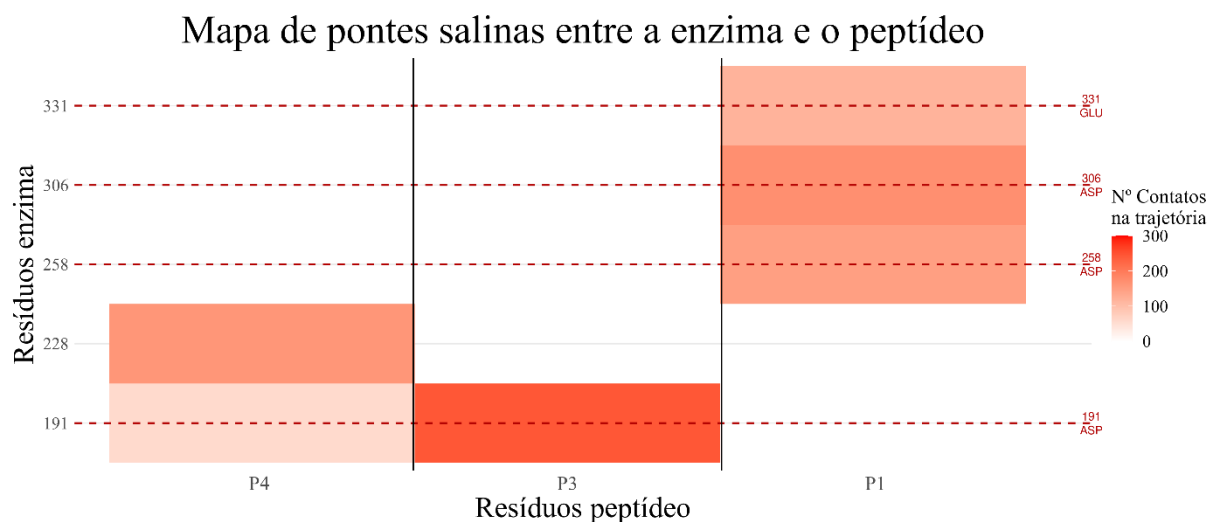


Figura 5.10 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima Furina durante a simulação realizada a partir da docagem com Haddock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Nota-se que P1 pode realizar interações de ponte salina com ASP258 e ASP306, e pontes menos frequentes com GLU331. Além disso também pode-se notar interações recorrentes entre P3 e ASP191.

Uma dessas possíveis pontes salinas se dá com entre a arginina P1 e o ASP306 da Furina. Esse aspartato faz parte do bolso de ligação S1 e está posicionado

no fundo do mesmo (**Figura 5.11**). Por se tratar de um aminoácido polar carregado negativamente, é possível que seus oxigênios interajam com os nitrogênios de P1 formando uma ponte salina. E, conforme demonstrado por Kumar e Nussinov, este tipo de interação pode contribuir para uma estabilização (KUMAR; NUSSINOV, 2002), o que pode estar causando a fixação de P1 no bolso de ligação.

As representações estruturais desse trabalho, como a apresentada na **Figura 5.11**, foram elaboradas usando PyMOL, nelas a carboxila ionizada dos aspartatos aparece com um dos oxigênios formando duas ligações com o carbono e o outro apenas uma. No entanto, entende-se que a carga negativa fica deslocalizada entre os átomos de oxigênio, assim como a positiva se deslocaliza dentre os três nitrogênios da cadeia lateral da arginina. Dessa forma, a arginina pode contatar pelos seus hidrogênios parcialmente positivos a carboxila ionizada do aspartato.

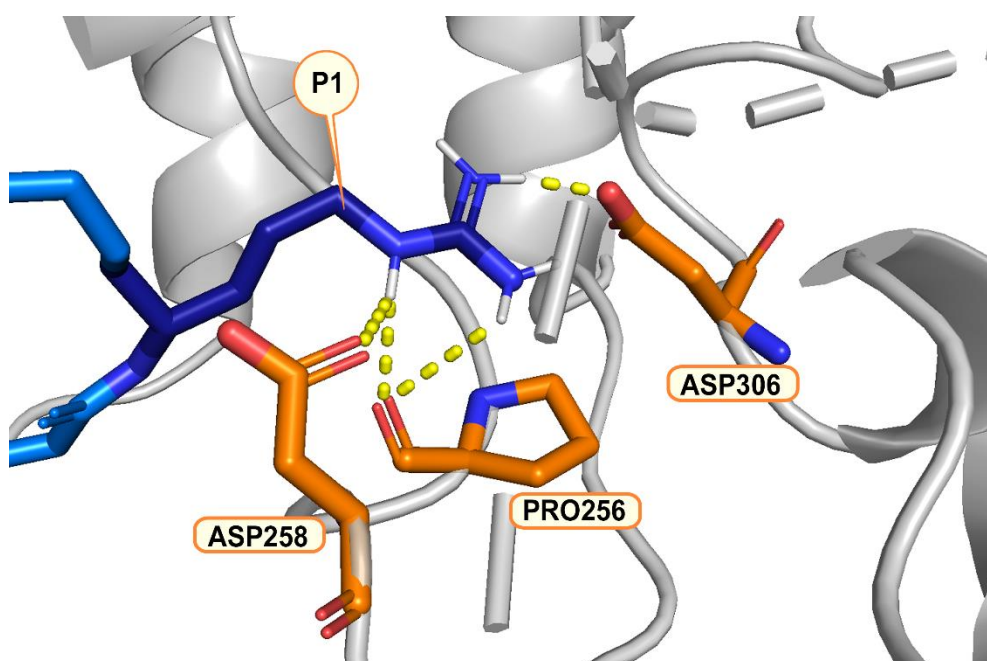


Figura 5.11 – Interações entre P1 e os resíduos do bolso de ligação S1 ao final da simulação a partir da docagem com Haddock. A cadeia lateral de P1 (azul escuro) pode interagir com os resíduos do bolso de ligação S1 (laranja).

Outra ponte salina possível de ocorrer se dá com a cadeia lateral do ASP258. Esse aspartato está localizado na entrada do subsítio S1 (**Figura 5.11**), e parece estabilizar P1 dentro do mesmo. As pontes salinas com ASP258 e ASP306 coincidem

com o reportado por Jawaid *et al.*, que também estudaram a ligação da Furina com a proteína espícula do coronavírus (JAWAID *et al.*, 2022).

Além dessas duas interações de pontes salinas que são mais recorrentes durante a simulação, P1 também parece estar realizando ponte salina com GLU331 (**Figura 5.10**). Essa interação, no entanto, ocorre apenas ao início da simulação, uma vez que, após cerca de 100ns, existe um distanciamento entre GLU331 e P1, mantendo uma distância de cerca de 4 Å no restante da trajetória.

Além das pontes salinas, a prolina 256 (PRO256), outro resíduo do S1, também frequentemente interage com P1 (**Figura 5.12**). O oxigênio da prolina encontra-se a uma distância menor que 3 Å de dois dos hidrogênios da cadeia lateral de P1, indicando uma possível interação com eles (**Figura 5.11**). Dessa forma, PRO256 pode também estar participando na fixação do peptídeo à enzima.

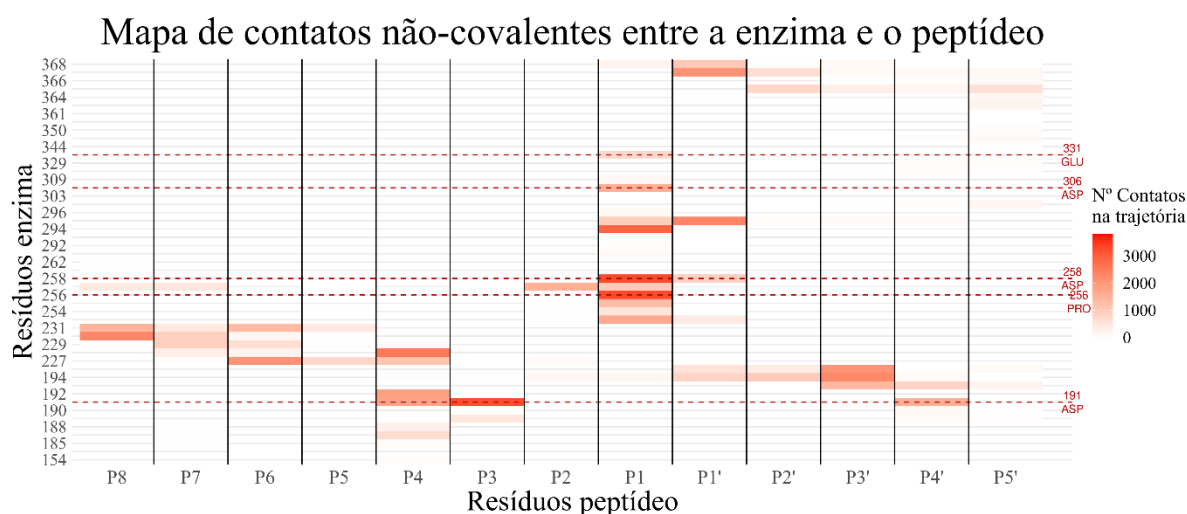


Figura 5.12 – Mapa de possíveis contatos não-covalentes entre o peptídeo e a enzima Furina durante a simulação realizada a partir da docagem com Haddock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Nota-se que podem ocorrer interações entre P1 e PRO256, além das pontes já citadas.

Além desses contatos no subsítio S1 do bolso de ligação, outro contato que parece ser determinante para a ligação e fixação do peptídeo ao bolso é uma possível ponte salina formada entre a arginina em P3 e o aspartato 191 (ASP191) do subsítio S2 da Furina (**Figura 5.10** e **Figura 5.13**). Essa interação se dá com uma arginina que

também foi adquirida pelo genoma desse novo coronavírus, e pode estar atuando na fixação do substrato à região, impedindo a movimentação e saída do mesmo.

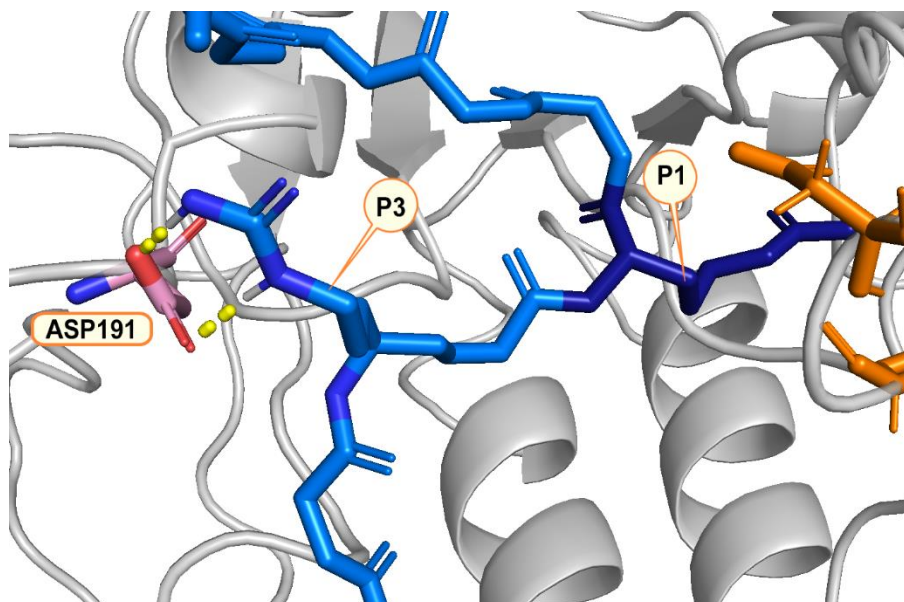


Figura 5.13 - Interações entre P3 e ASP191 (rosa) ao final da simulação com a Furina a partir da docagem com Haddock.

Essas pontes salinas que podem estar ocorrendo entre os resíduos P1 e P3 com a proteína podem atuar estabilizando esses resíduos no local, e, por conseguinte diminuir a movimentação destes. Isso fica mais claro ao analisarmos o gráfico de RMSF do peptídeo (**Figura 5.14**). Nele vemos que esses dois resíduos estão dentre os que possuem menor movimentação durante a simulação. A estabilização desses dois resíduos também pode estar fazendo com que seus vizinhos percam mobilidade, quando comparados com os resíduos mais afastados, contribuindo para a estabilização local do peptídeo.

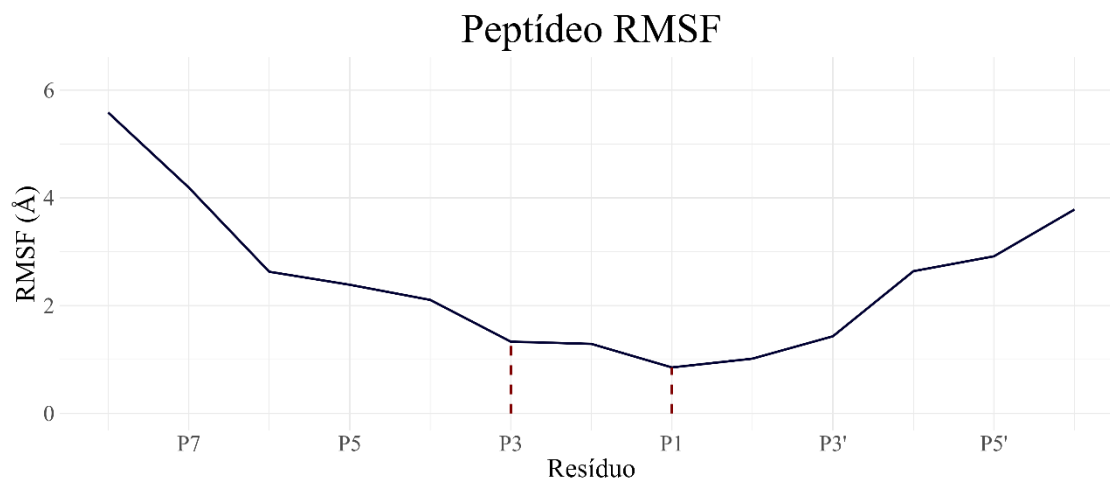


Figura 5.14 - RMSF do peptídeo por resíduo, medido na simulação realizada após a docagem à enzima Furina com Haddock.

A análise das energias de interação entre o peptídeo e a enzima também evidencia uma certa estabilidade na interação entre eles. Ao observar o termo de energia de van der Waals durante a simulação feita a partir do complexo da docagem com Haddock, podemos ver que o valor da energia tem ficado mais negativo nos primeiros 150ns de simulação, e depois mantém o mesmo patamar durante os 150ns finais (**Figura 5.15**). Valores de energias mais negativos indicam que é necessário fornecer mais energia para romper essa interação, e, portanto, podem indicar uma maior estabilidade (HANSSON; MARELIUS; ÅQVIST, 1998; JAIN, T.; JAYARAM, 2005).

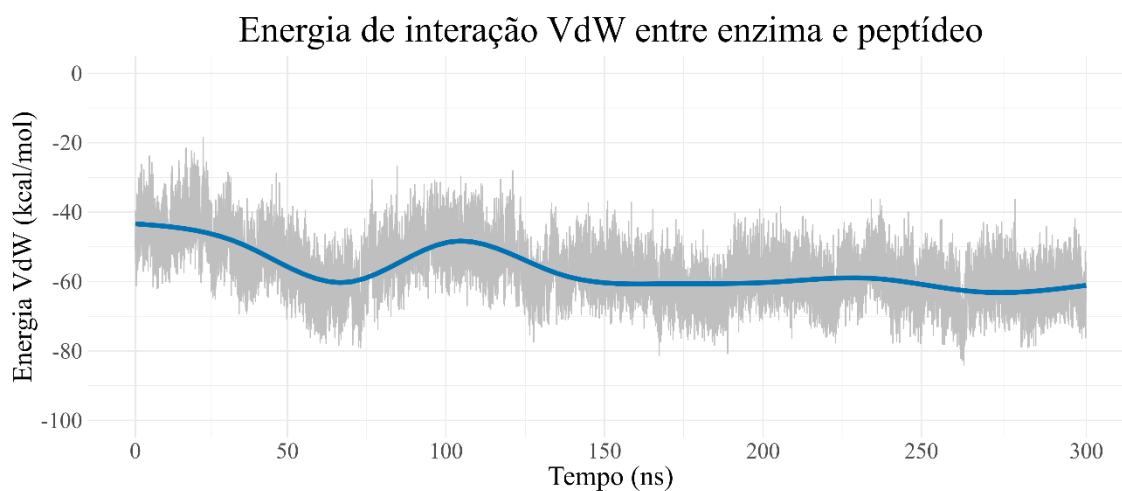


Figura 5.15 – Gráfico de energia de van der Waals entre o peptídeo e a enzima Furina durante a simulação a partir da docagem com Haddock.

Analisando o modelo ao final da simulação rodada a partir da docagem com Haddock, vemos que ASN295, que compõe o bolsão de oxiânion, descrito na seção 1.3.1, se encontra muito próximo da ligação que será atacada (**Figura 5.16**). Isto pode indicar um posicionamento já favorável para que esse resíduo haja neutralizando os radicais após a catálise pela serina, conforme também descrito na seção 1.3.1.

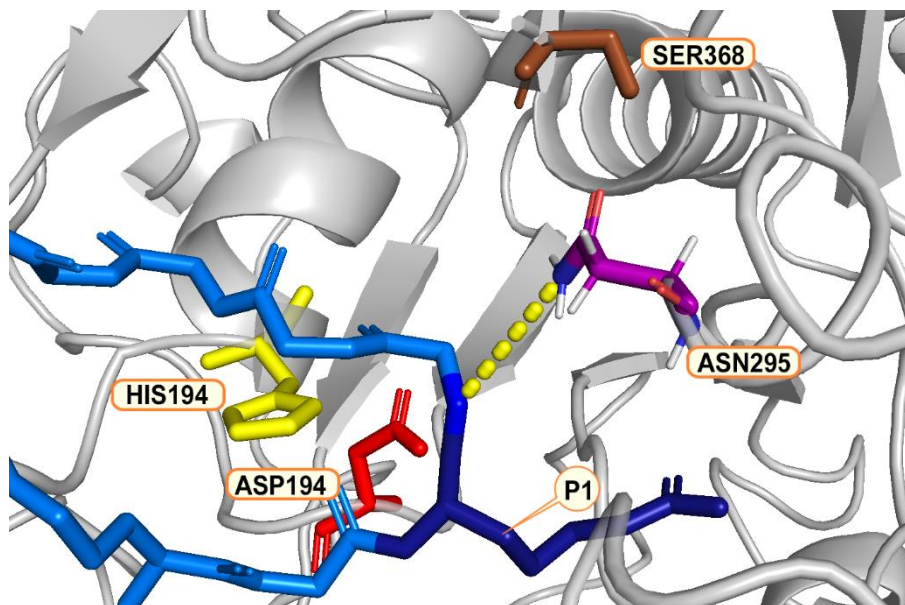


Figura 5.16 – Interação entre o peptídeo e sítio catalítico da Furina ao final da simulação rodada a partir da docagem com Haddock. P1 (azul escuro) pode interagir com ASP295 do bolsão de oxiânion (roxo), porém se encontra distante dos outros resíduos do sítio: ASP153 (vermelho), HIS194 (amarelo) e SER368 (marrom).

Observando a **Figura 5.16**, temos que a serina catalítica (SER368), ficou distante de P1 e da histidina catalítica (HIS194). Isto pode ocorrer por diversos fatores, como por exemplo, a aproximação entre P1 e a tríade catalítica pode acontecer em uma escala de tempo maior do que a utilizada aqui. Ou, como essa aproximação envolve trocas de carga, conforme descrito na seção 1.3.1, e as simulações convencionais não realizam as trocas de cargas necessárias, pode ser que não tenham sido formadas as condições necessárias para tal ajuste induzido.

Já nas simulações a partir da docagem com CABSdock, P1 fica voltado em sentido contrário ao bolso de ligação S1 (**Figura 5.17**). Isso ocorre desde a docagem, e se mantém durante toda a dinâmica molecular. Como P1 não se encontra inserido, ou mesmo voltado para o subsítio S1, as interações entre P1 e os resíduos do bolso,

e que foram apresentadas nas simulações a partir do Haddock, não se repetem (**Figura 5.18**). Por isso, nós não descreveremos em detalhes essa simulação.

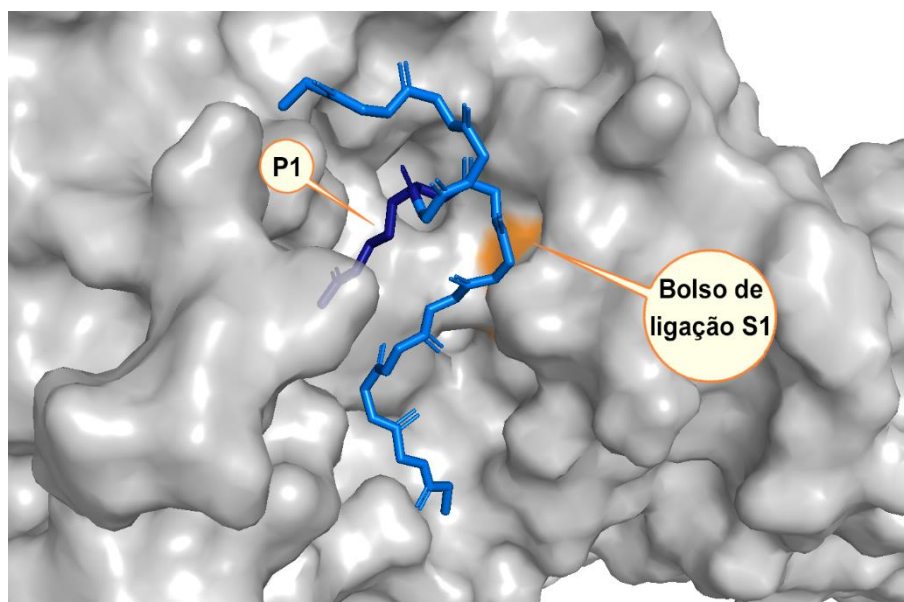


Figura 5.17 – Posição final da simulação a partir da docagem do peptídeo à Furina usando CABSdock. Observa-se que P1 (azul escuro) fica voltado em sentido contrário ao bolso de ligação S1 (laranja).

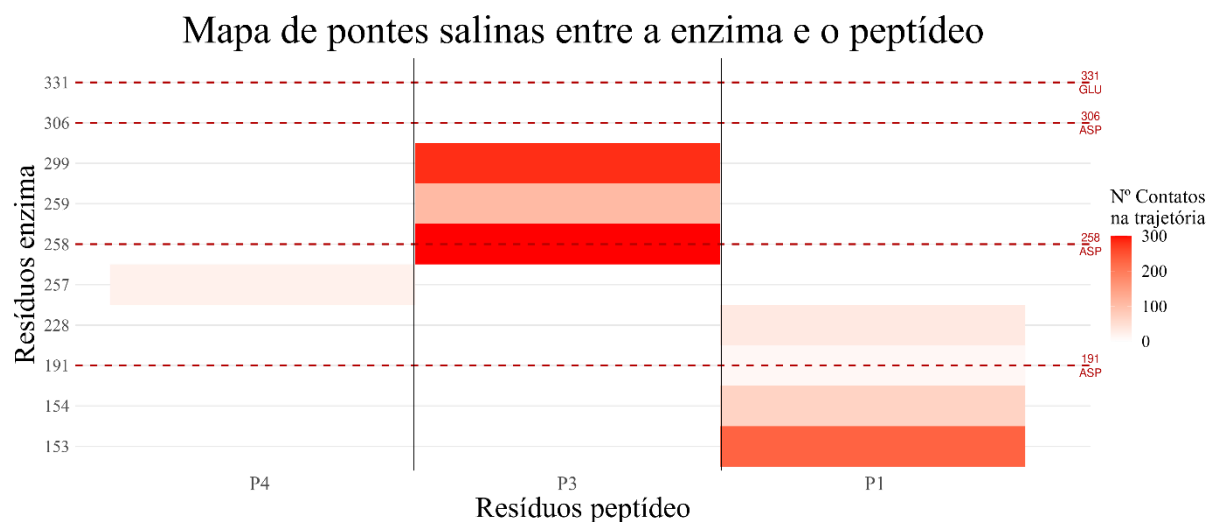


Figura 5.18 - Mapa de possíveis pontes salinas entre o peptídeo e a enzima Furina durante a dinâmica realizada a partir da docagem com CABSdock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Nota-se que P1 não realiza interações de ponte salina com nenhum dos resíduos do bolso de ligação S1.

5.4.2 PCSK2

Da mesma forma como ocorre na Furina, ao analisarmos os gráficos de RMSD do peptídeo na simulação com a enzima PCSK2, notamos que há uma certa estabilidade após algum tempo. Na simulação de dinâmica molecular rodada a partir da docagem com Haddock, o RMSD atinge o patamar de 5 Å por volta dos 100ns de simulação, e oscila em torno desse valor durante os 200ns restantes (**Figura 5.19**). O que indica que, com a PCSK2, da mesma forma que com a Furina, o peptídeo encontrou uma posição possivelmente estável.

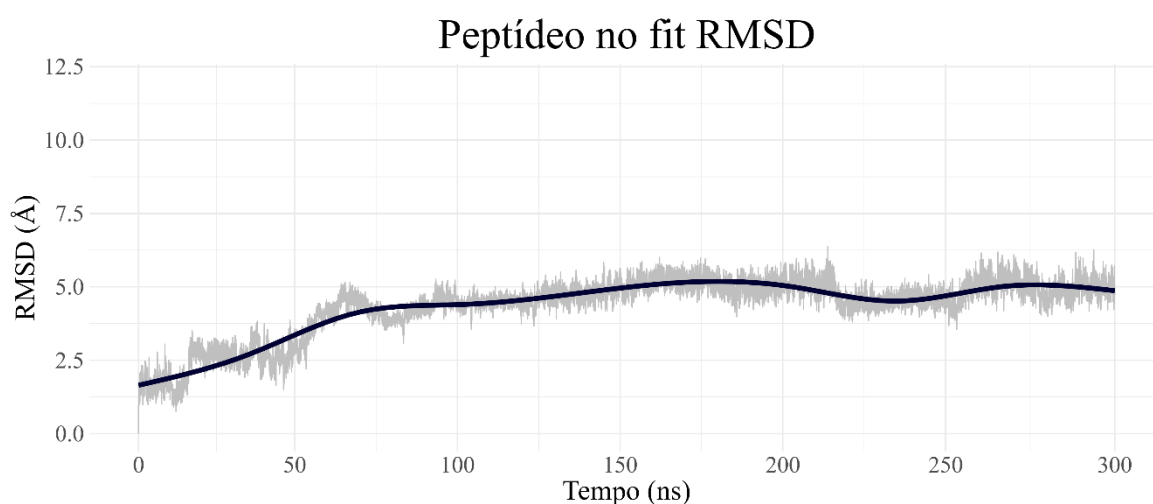


Figura 5.19 – *No fit* RMSD do peptídeo docado com Haddock na enzima PCSK2 durante a simulação. Nota-se uma estabilidade em 5 Å a partir dos 100ns.

Ao analisarmos os modelos ao final da simulação rodada a partir do posicionamento do peptídeo pela docagem com Haddock, notamos que também há uma inserção de P1 no bolso de ligação S1. Esse resíduo ficou, assim como na Furina, encoberto pela proteína e podendo contactar três dos resíduos do subsítio S1 (PRO271, ASP273 e ASP320) apresentados na seção 1.3.2 (**Figura 5.20**).

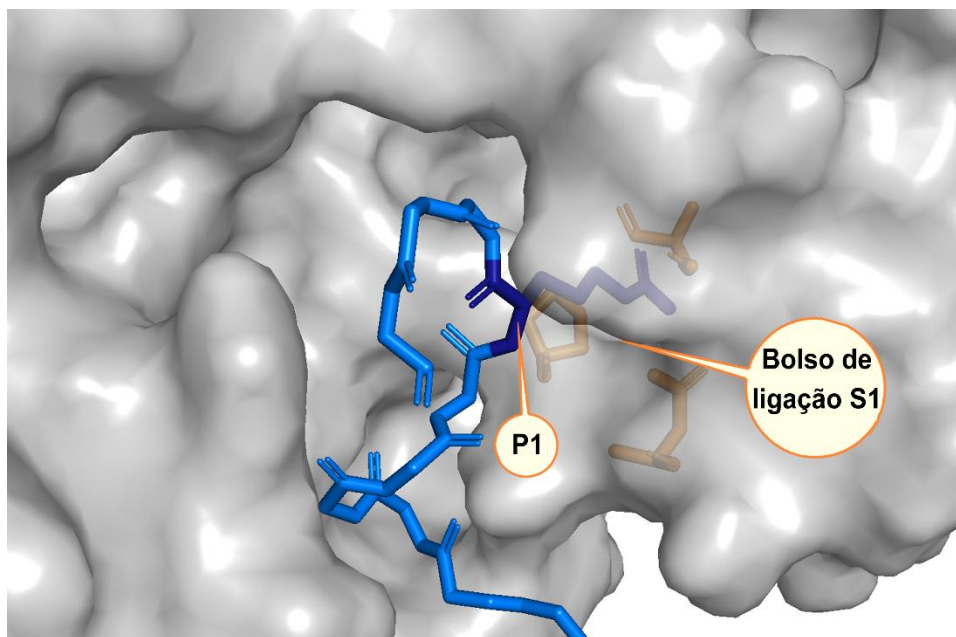


Figura 5.20 - Modelo ao final da simulação do peptídeo docado na PCSK2 usando Haddock. Observe que P1 (azul escuro) encontra-se próximo aos resíduos do bolso de ligação S1 (laranja).

Observando a taxa de superfície não acessível ao solvente de P1, podemos investigar o quanto este resíduo está encoberto pela enzima. Na simulação rodada a partir da docagem do peptídeo pelo Haddock, P1 manteve cerca de 80% a 90% de sua área total encoberta durante a simulação. Sendo que cerca de 60% de sua área total está sendo encoberta pela enzima (**Figura 5.21**), assim como observado para Furina, indicando uma boa alocação no fundo do bolso de ligação S1.

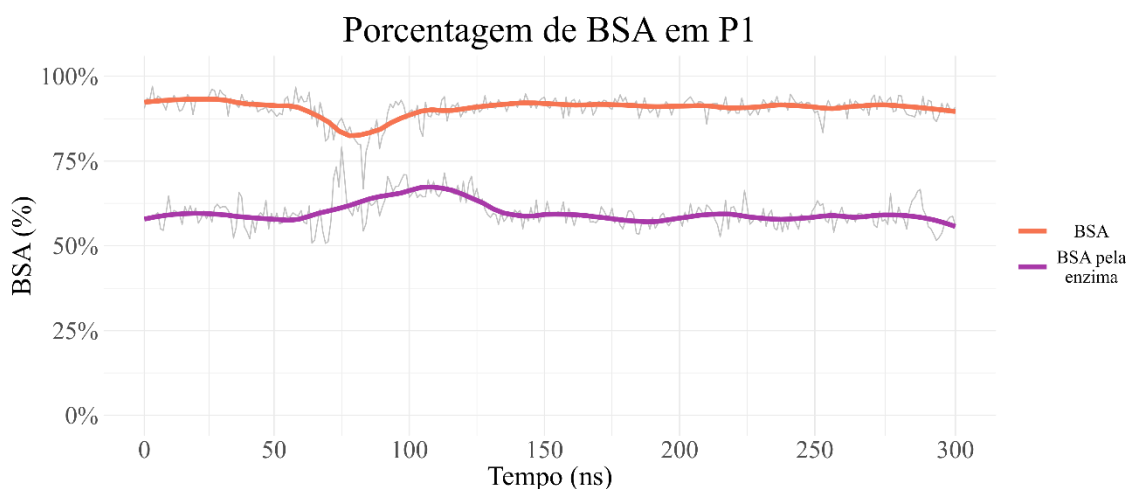


Figura 5.21 – Gráfico de BSA de P1 durante a dinâmica com a enzima PCSK2, docagem com Haddock. Em laranja a porcentagem total de área encoberta de P1, e em roxo a porcentagem de área de P1 encoberta apenas pela PCSK2.

Analisando os possíveis contatos entre a PCSK2 e o peptídeo na simulação feita a partir da docagem com Haddock, notamos interações semelhantes às aquelas apresentadas entre o peptídeo e a Furina (**Figura 5.10**), com a possível formação de duas pontes salinas entre a P1 e a enzima no subsítio S1, interagindo com ASP273 e ASP320 (**Figura 5.22**).

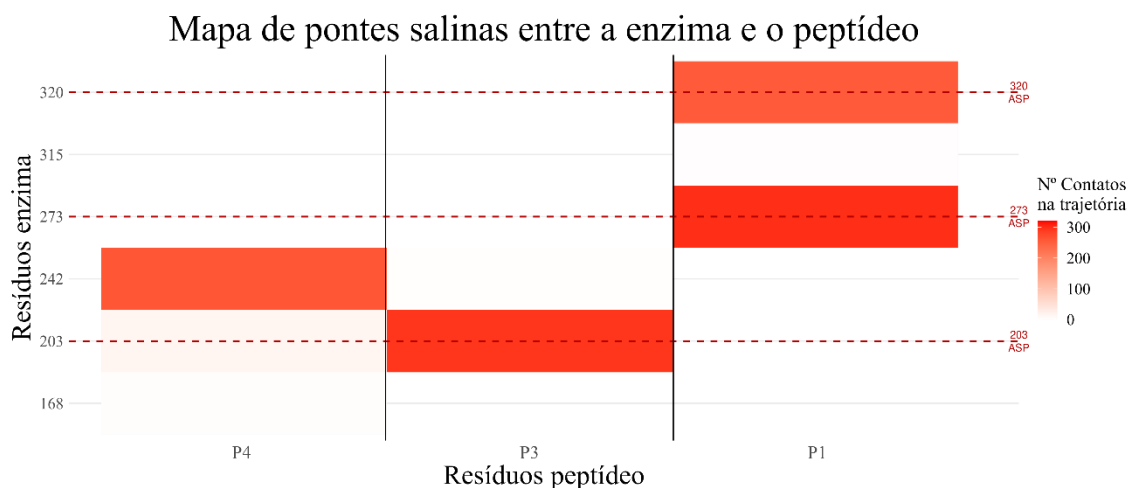


Figura 5.22 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK2 durante a simulação realizada a partir da docagem com Haddock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Nota-se que P1 pode realizar interações de ponte salina com ASP273 e ASP320 do S1. Além disso também pode-se notar interações recorrentes entre P3 e ASP203.

Uma dessas pontes salinas pode ocorrer entre P1 e ASP320 do bolso de ligação S1 (**Figura 5.23**). Assim como na enzima Furina, esse aspartato está localizado no fundo do subsítio S1, e a formação dessa interação pode influenciar na estabilidade local de P1, fazendo com que a arginina se mantenha inserida na cavidade durante toda a simulação. A outra possível ponte salina se dá com a cadeia lateral do ASP273, um ácido aspártico localizado na entrada do bolso de ligação S1, análogo ao ASP258 da Furina (**Figura 5.23**).

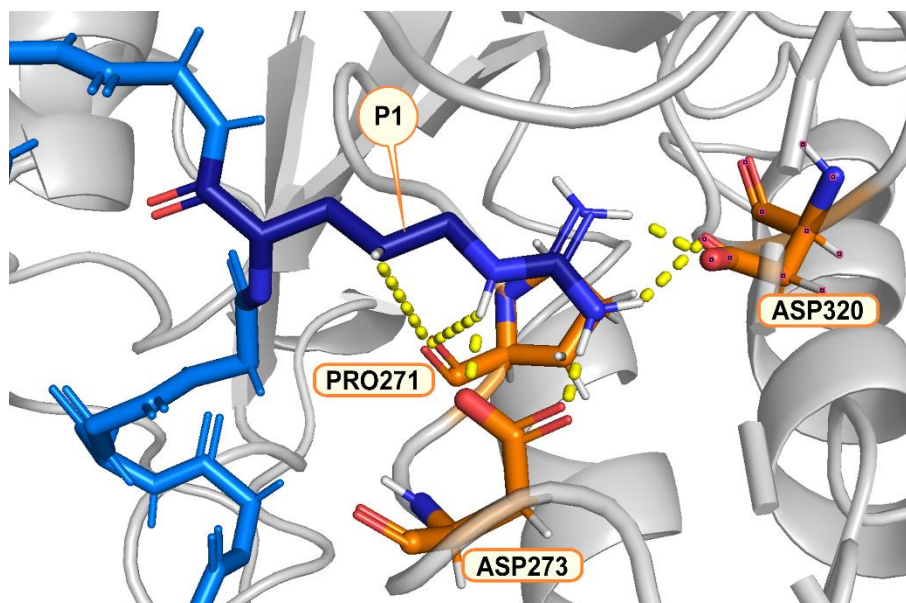


Figura 5.23 – Interações entre P1 e os resíduos do bolso de ligação S1 ao final da simulação a partir da docagem com Haddock. A cadeia lateral de P1 (azul escuro) pode interagir com os resíduos do bolso de ligação S1 (laranja).

Outro resíduo importante do bolso de ligação S1, a prolina (PRO271), também interage com P1 na simulação a partir da docagem com Haddock (**Figura 5.24**). No caso, o oxigênio da prolina se encontra a cerca de 3,5 Å de dois hidrogênios da arginina, a partir de 100ns de simulação, quando o peptídeo se estabiliza, indicando que pode haver interações não-covalentes entre P1 e PRO271. Nessa simulação, a aproximação se dá de maneira súbita e duradoura, indicando uma afinidade entre esses dois resíduos e mostrando a importância das simulações para análise das interações entre as estruturas (**Figura 5.25**).

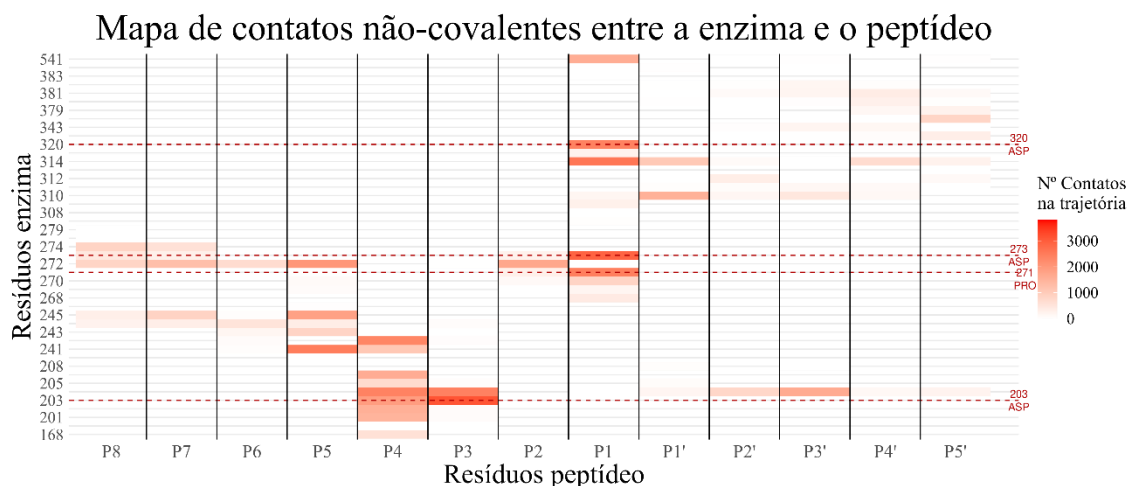


Figura 5.24 – Mapa de possíveis contatos não-covalentes entre o peptídeo e a enzima PCSK2 durante a simulação realizada a partir da docagem com Haddock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Destaca-se as interações que podem ocorrer entre P1 e PRO271.

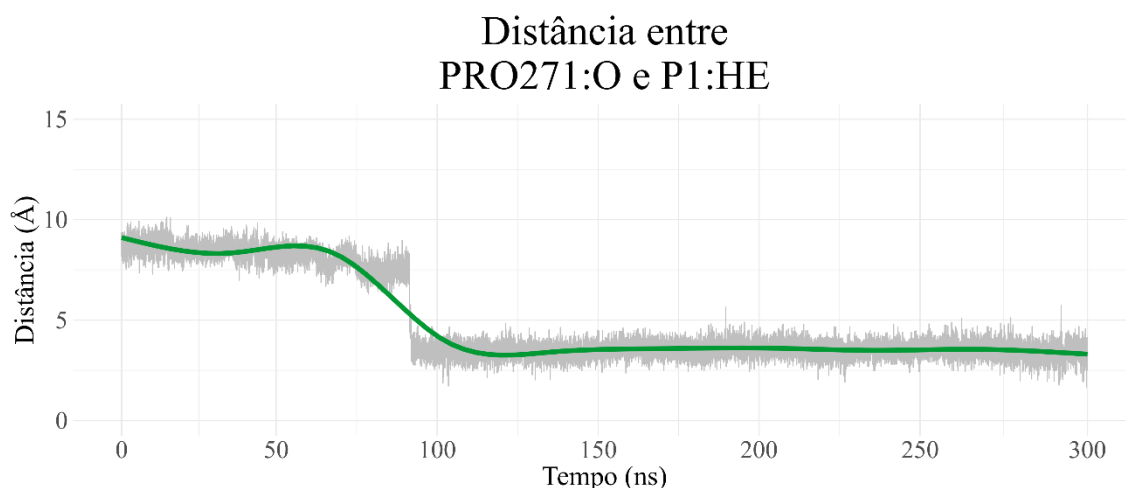


Figura 5.25 - Gráfico de distâncias em Å entre o oxigênio de PRO271 e um hidrogênio da cadeia lateral de P1 (HE), durante a simulação com PCSK2 a partir da docagem com Haddock.

Notamos que também pode haver uma interação entre P3 e um terceiro aspartato da enzima PCSK2 (**Figura 5.22** e **Figura 5.26**). Essa interação do tipo ponte salina ocorre com o aspartato ASP203, e, assim como na simulação com a Furina, pode estar contribuindo para a estabilização local do peptídeo. Analisando o alinhamento da **Figura 1.7**, vemos que esse aspartato, apesar de estar próximo do ASP191 da Furina, não corresponde ao mesmo aspartato do S2 nessa enzima. Embora, no alinhamento das estruturas tridimensionais, eles estejam em regiões muito semelhantes nas enzimas.

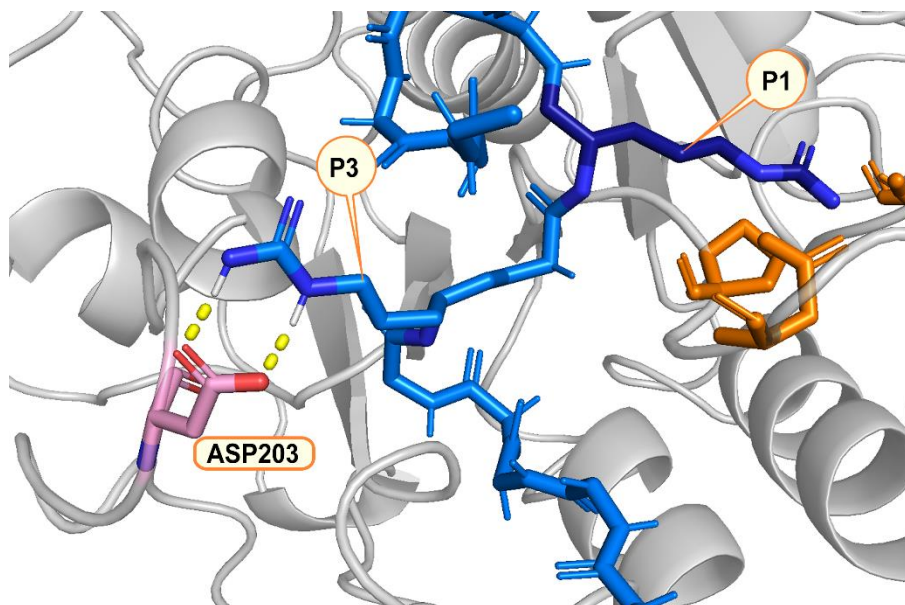


Figura 5.26 - Interações entre P3 e ASP203 (rosa) ao final da simulação a partir da docagem com Haddock.

Como P1 e P3 podem estar sendo estabilizados por pontes salinas, as mobilidades desses resíduos, medida através do RMSF, são uma das menores de toda a estrutura (**Figura 5.27**), indicando que essas interações podem estar atuando para uma maior fixação local desses resíduos.

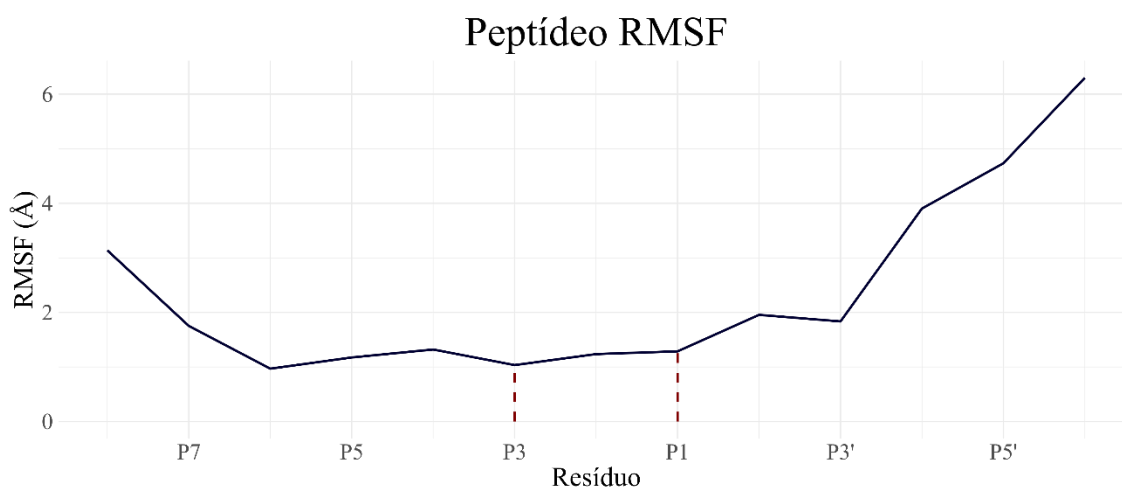


Figura 5.27 - RMSF do peptídeo por resíduo, medido na simulação realizada após a docagem à enzima PCSK2 com Haddock.

Analisando as energias de interação entre o peptídeo e a enzima durante a dinâmica a partir da docagem com Haddock, vemos que o termo de energia de van der Waals mantém uma média de valor constante (**Figura 5.28**). Apesar de não observarmos uma leve melhora dos valores de energia de interação como o mostrado na simulação com a Furina, o fato de não haver oscilações de energia pode também indicar a estabilidade da interação entre proteína e peptídeo. De qualquer forma, a energia de interação hidrofóbica chega a valores próximos da obtida com a Furina (**Figura 5.15**) e em ambos os casos se estabiliza em cerca de 150ns.

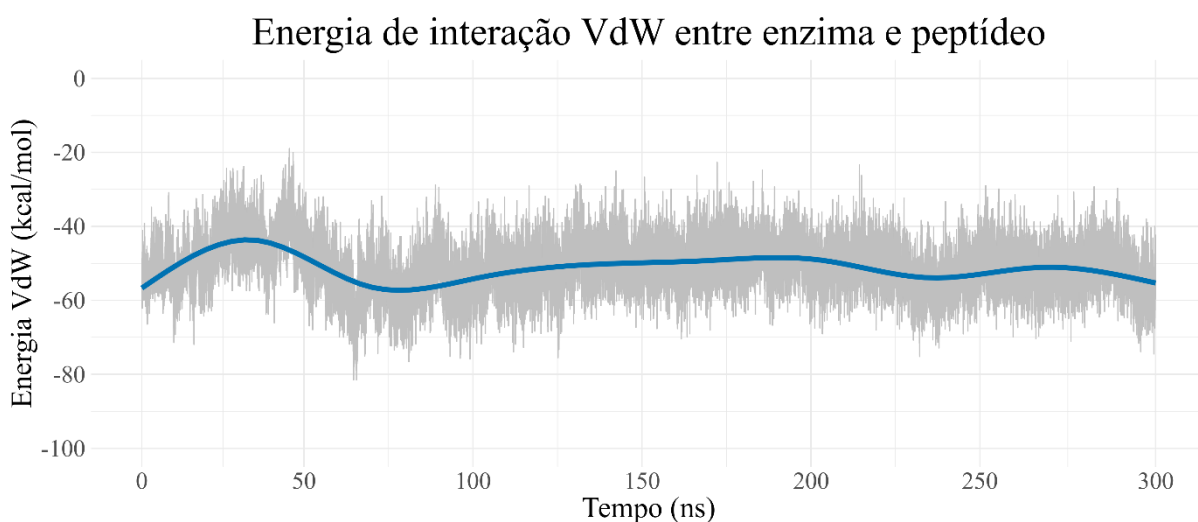


Figura 5.28 – Gráfico de energia de van der Waals entre o peptídeo e a enzima PCSK2 durante a dinâmica a partir da docagem com Haddock.

Em relação aos resíduos do sítio catalítico, diferente do que vimos na simulação com a Furina a partir da docagem com Haddock, nas simulações com a PCSK2 o resíduo do bolsão de oxianion, aspartato ASP310, não se encontra próximo da arginina que vai sofrer a clivagem. E, assim como na simulação com a Furina, a serina que realizará o ataque nucleofílico (SER384) se encontra distante de P1 (**Figura 5.29**) durante a docagem do possível substrato, talvez implicando que a movimentação para a catálise se faça com ajuste induzido posterior ao atracamento inicial.

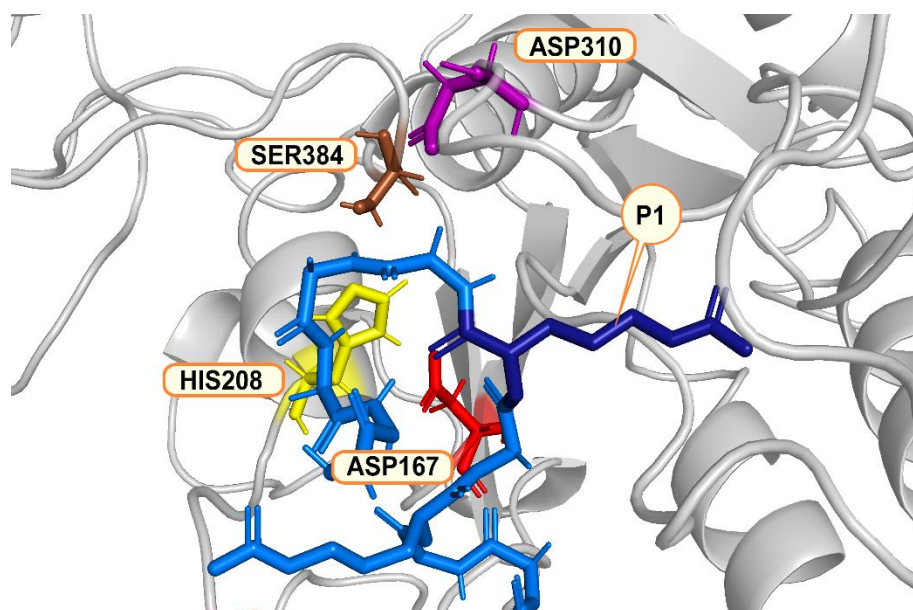


Figura 5.29 – Figura mostrando o peptídeo e sítio catalítico no final da simulação rodada a partir da docagem com Haddock. A cadeia lateral de P1 está representada em azul escuro, ASP295 do *oxyanion hole* em roxo, ASP153 em vermelho, HIS194 em amarelo e SER368 em marrom.

Nas simulações a partir da docagem realizada com o CABSdock, podemos ver que P1 começa a simulação afastado do bolso de ligação S1 e, ao percorrer da simulação, se aproxima dos resíduos da borda do subsítio (**Figura 5.30**), podendo até mesmo interagir com ASP273 ao final (**Figura 5.31**), o que não observamos na simulação com a Furina, feitas a partir da docagem com CABSdock. No entanto, P3 se encontra, durante toda a simulação, inserido no bolso de ligação S1, podendo formar ponte salina com ASP320 no fundo da cavidade (**Figura 5.31** e **Figura 5.32**).

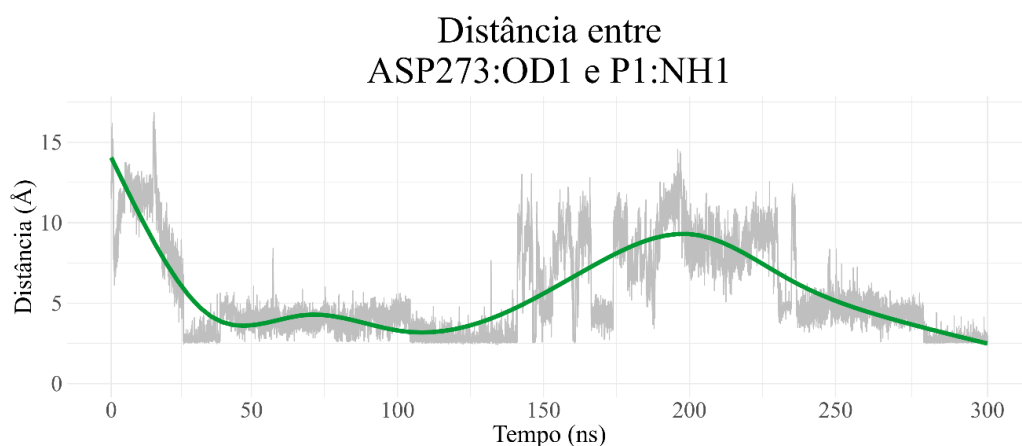


Figura 5.30 – Distância entre o nitrogênio da cadeia lateral de P1 e o oxigênio de ASP273, localizado na borda do bolso de ligação, durante a simulação com a enzima PCSK2, a partir da docagem com CABSdock.

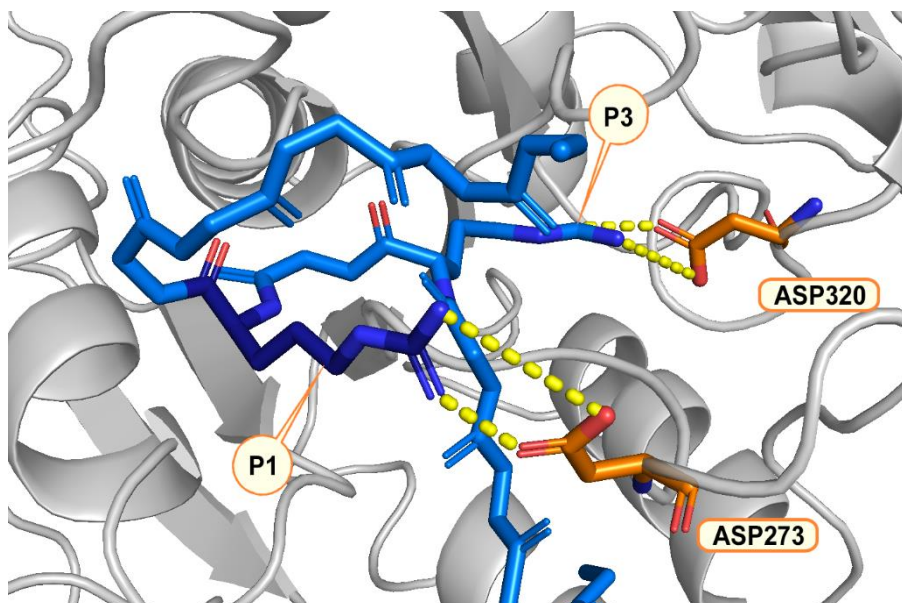


Figura 5.31 - Interações entre P1 e P3 e os resíduos do bolso de ligação S1 na simulação a partir da docagem com CABSdock na enzima PCSK2. A cadeia lateral de P1 (azul escuro) pode interagir com ASP273 na borda do bolso de ligação, e P3 se encontra inserido no bolso podendo interagir com ASP320.

Mapa de pontes salinas entre a enzima e o peptídeo

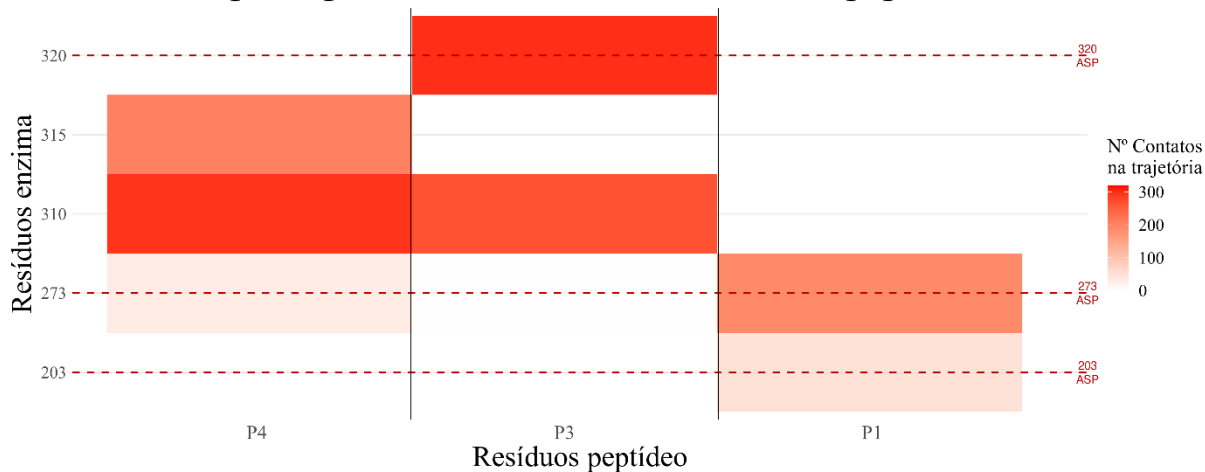


Figura 5.32 - Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK2 durante a simulação realizada a partir da docagem com CABSdock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Nota-se que P1 pode realizar interações de ponte salina com ASP273 da borda do bolso de ligação S1, e P3 interage com ASP320 no fundo do subsítio S1.

Essa possível interação entre P3 e ASP320 no fundo do bolso de ligação S1 estaria auxiliando na estabilização local do resíduo na cavidade. Dessa forma, P1, durante o tempo de simulação, não consegue entrar no subsítio que já está ocupado.

Isso pode ser um indício de que há uma competição entre P1 e P3 por essa posição, fazendo com que o primeiro não possa ocupar um posicionamento mais favorável à catálise enzimática. A entrada de P3 ao invés de P1 no sítio pode ter tido influência da estruturação do peptídeo necessária para a docagem com Haddock.

De forma geral, os resultados observados nas simulações de dinâmica molecular a partir do Haddock com as enzimas Furina e PCSK2, principalmente os contatos de P1 com os resíduos do bolso de ligação S1, não se repetiram nas simulações com as demais PCSKs. Apesar de, em alguns casos, o P1 se fixar próximo ao subsítio S1 do bolso de ligação, não foi observado em nenhuma outra enzima a inserção de P1 contactando o aspartato no fundo desse subsítio.

5.4.3 PCSK1

Nas simulações com a PCSK1 a partir da docagem do Haddock, P1 não ficou inserido dentro do bolso de ligação S1 (**Figura 5.33**). Além disso, o peptídeo não atingiu uma posição muito estável ao final dos 300ns de simulação, conforme podemos observar nos gráficos de RMSD do mesmo (**Figura 5.34**), sugerindo que não houveram interações que possivelmente estabilizariam o peptídeo em relação ao bolso de ligação da enzima.

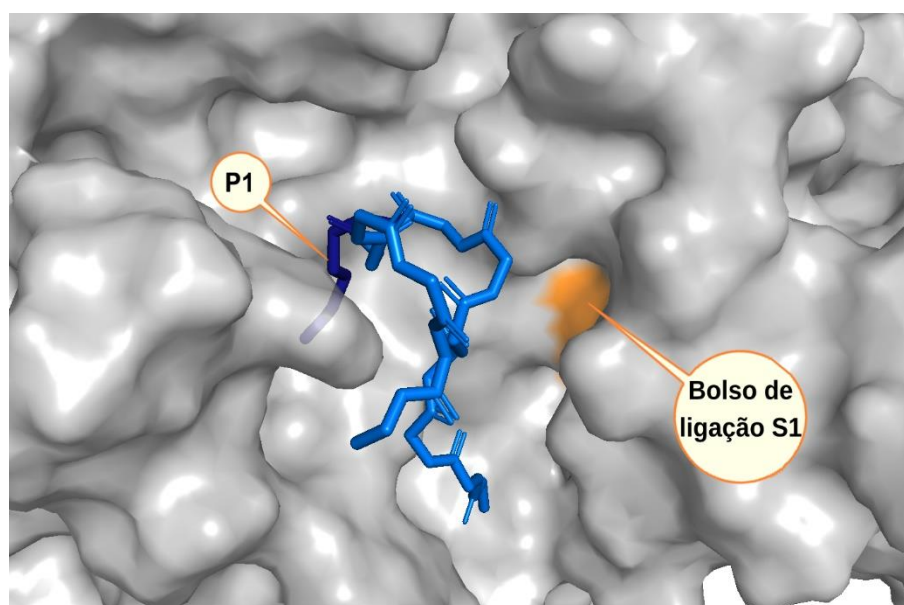


Figura 5.33 – Posição final do peptídeo na simulação a partir da docagem à PCSK1 usando Haddock. Observa-se que P1 (azul escuro) fica distante do bolso de ligação S1 (laranja).

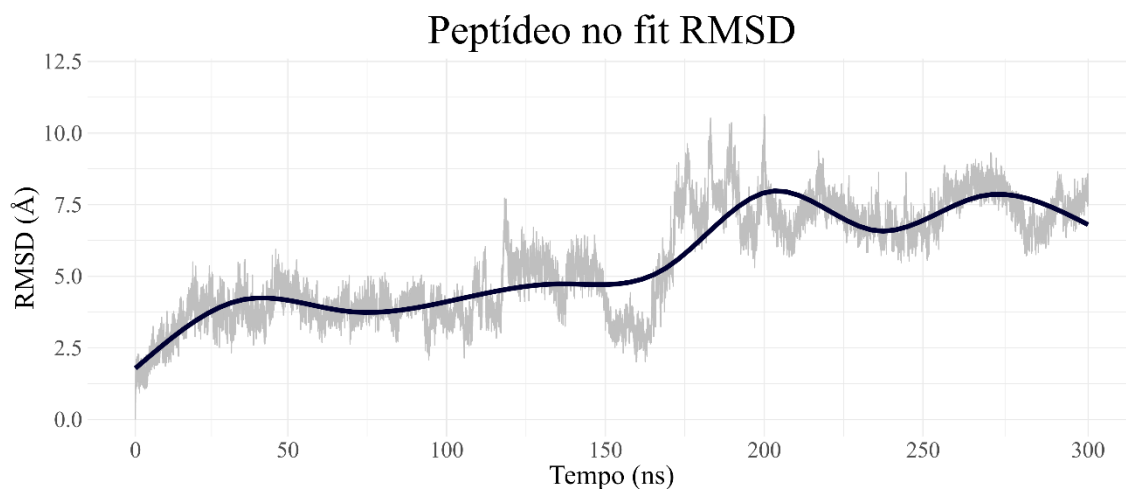


Figura 5.34 – RMSD do peptídeo docado com Haddock na enzima PCSK1 durante a simulação. Nota-se que o peptídeo se movimenta durante toda a simulação.

Foi observado nessa simulação que P3 pode formar ponte salina com o aspartato da borda do bolso de ligação S1 (ASP272). Da mesma forma, P4 pode interagir com um glutamato (GLU205) que, por alinhamento, corresponde ao (ASP191) do bolso de ligação S2 em Furina (**Figura 5.35**). Quando comparado às simulações a partir da docagem com Haddock com a Furina ou PCSK2, vemos que P3, neste caso, está ocupando a posição de P1, e P4 o lugar de P3. Novamente, sugerindo que pode haver uma competição entre esses resíduos, o que estaria interferindo na ligação de P1 ao bolso de ligação. Essa competição pode ser um dos limites para que a metodologia aplicada revele o potencial de PCSK1 em clivar o substrato.

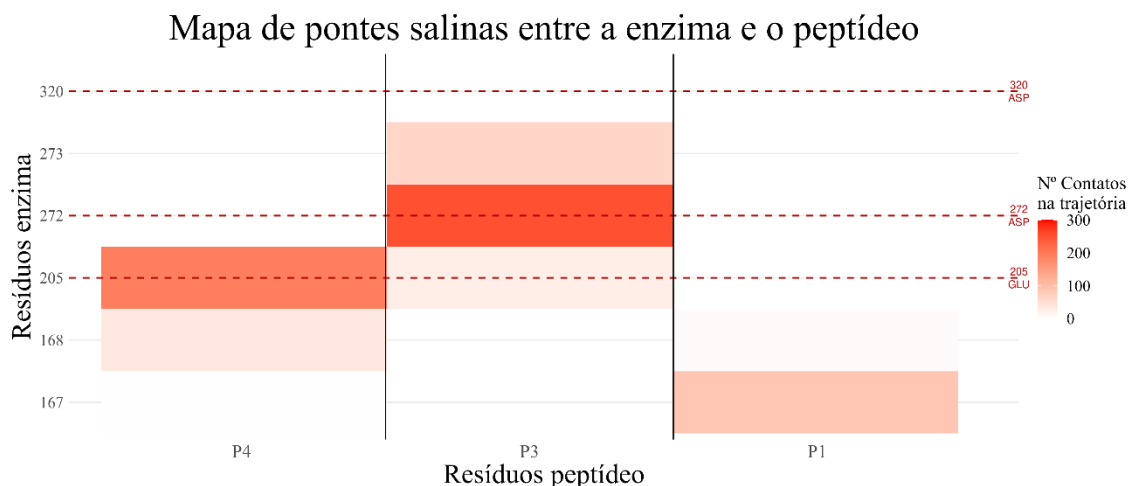


Figura 5.35 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK1 durante a simulação realizada a partir da docagem com Haddock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Nota-se que P3 pode realizar interações de ponte salina com ASP272, e P3 com GLU205.

Já na simulação a partir da docagem realizada com CABSdock, P1 ficou posicionado próximo do aspartato 272 (ASP272), que fica na superfície do bolso de ligação S1, podendo interagir através de uma ponte salina com o mesmo (**Figura 5.36** e **Figura 5.37**). Porém, como não ocorre a inserção de P1 no bolso de ligação, não há interação deste com os resíduos importantes no fundo do bolso (**Figura 5.36** e **Figura 5.37**). Além disso, P4 ficou próximo do GLU305 do bolso de ligação S2, podendo realizar algumas interações específicas (**Figura 5.37**). Desta forma não conseguimos evidenciar condições de associação eficiente do substrato com a PCSK1. Desta forma, não conseguimos evidenciar condições de associação eficiente do substrato com a PCSK1.

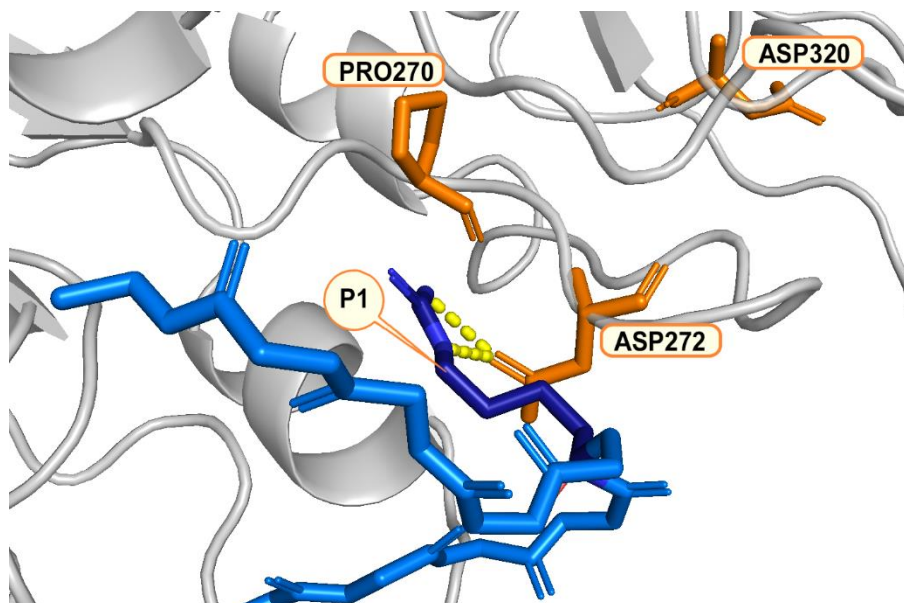


Figura 5.36 - Interações entre P1 e ASP272 do bolso de ligação S1 da PCSK1, ao final da simulação a partir da docagem com CABSdock. A cadeia lateral de P1 (azul escuro), no entanto, não interage com os outros resíduos do bolso de ligação S1.

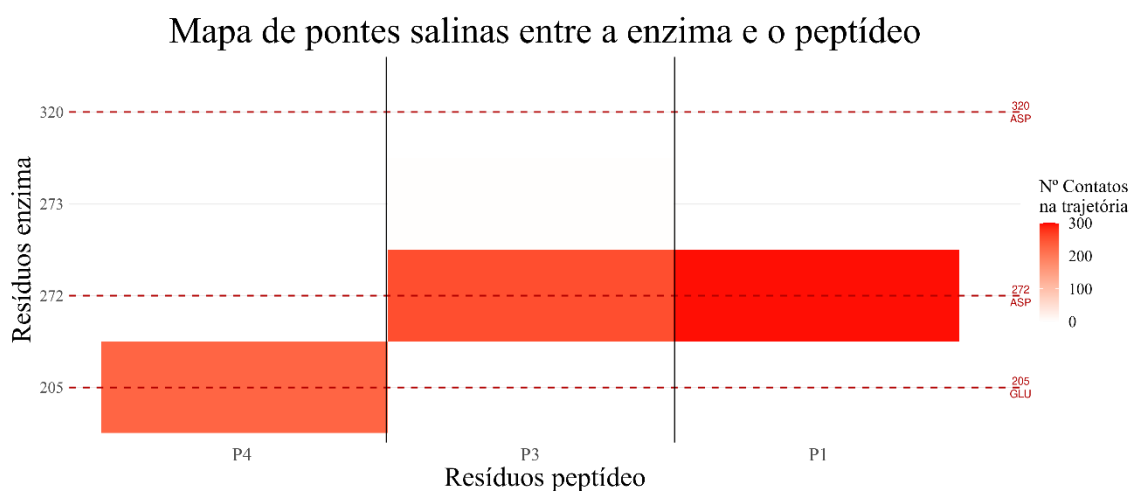


Figura 5.37 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK1 durante a simulação realizada a partir da docagem com CABSdock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Nota-se que P1 pode realizar interações de ponte salina apenas com ASP272 do S1. E P4 pode interagir com GLU205 do S2.

Estas interações de P1 com ASP272 na superfície do bolso de ligação S1, indica que a enzima apresentou uma certa afinidade, ainda que fraca, do P1 ao subsítio S1. Não ocorrendo, no entanto, de maneira ótima, contactando os outros resíduos que compõem o S1, apresentados na **Tabela 1.1**, e vistos interagindo com o P1 nas enzimas Furina e PCSK2.

5.4.4 PCSK4

Nas simulações com a enzima PCSK4, tanto a partir da docagem do Haddock, quanto do CABSdock, P1 não se posiciona próximo ao bolso de ligação S1 (**Figura 5.37**). Por isso, não observamos a ligação de P1 com nenhum dos resíduos apresentados na **Tabela 1.1**.

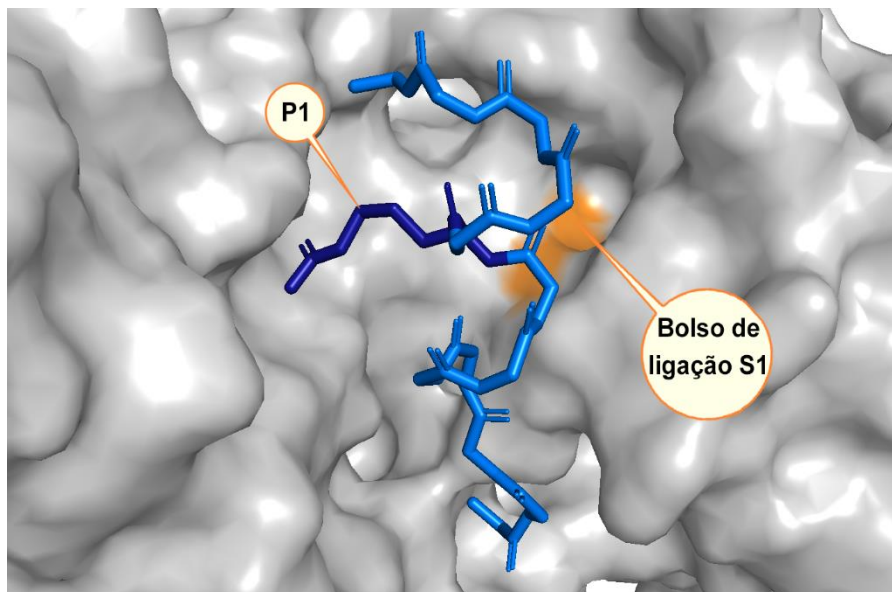


Figura 5.38 – Posição final do peptídeo na simulação a partir da docagem à PCSK4 usando Haddock. Observa-se que a cadeia lateral de P1 (azul escuro) fica posicionada oposta ao bolso de ligação S1 (laranja).

Analisando os gráficos de energia de interação, notamos ainda que os valores do termo de energia de van der Waals entre o peptídeo e a enzima diminuem em módulo durante a simulação, conforme pode ser observado na **Figura 5.39**. Indicando, que a ligação entre as duas cadeias ficou mais fraca e instável ao longo da simulação, mesmo partindo de valores que foram os obtidos no final da simulação com Furina e PCSK2.

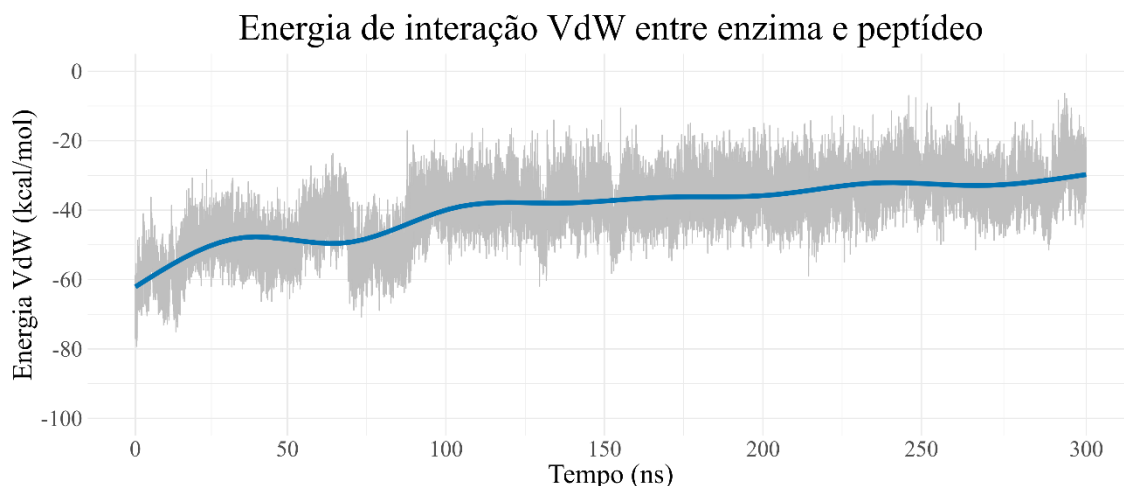


Figura 5.39 – Gráfico de energia de van der Waals entre o peptídeo e a enzima PCSK1 durante a simulação a partir da docagem com Haddock.

5.4.5 PCSK5

Na simulação com a enzima PCSK5, a partir da docagem com Haddock, não é possível notar um posicionamento muito estável do peptídeo, o que pode ser observado através do RMSD do mesmo ao longo da simulação (**Figura 5.40**), indicando que o peptídeo não se encaixa e interage com a enzima de uma forma estável quanto a observada nas simulações com Furin e PCSK2.

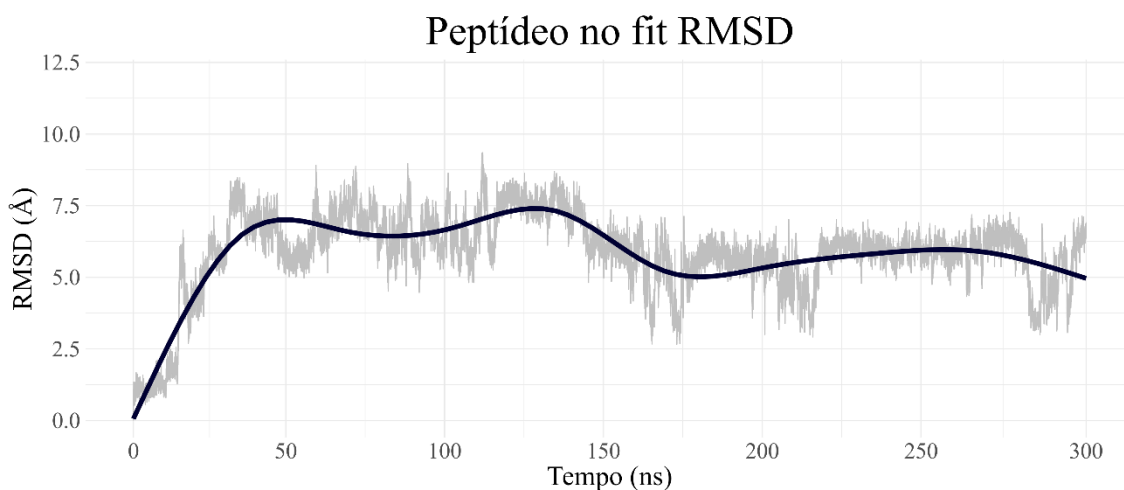


Figura 5.40 – RMSD do peptídeo docado com Haddock na enzima PCSK5 durante a simulação. Nota-se que o peptídeo se movimenta durante toda a simulação.

Além disso, P1 não contacta os resíduos do bolso de ligação S1 da forma esperada. Podendo haver apenas uma ponte salina entre P3 e o aspartato localizado na borda do subsítio (ASP276) (**Figura 5.41**).

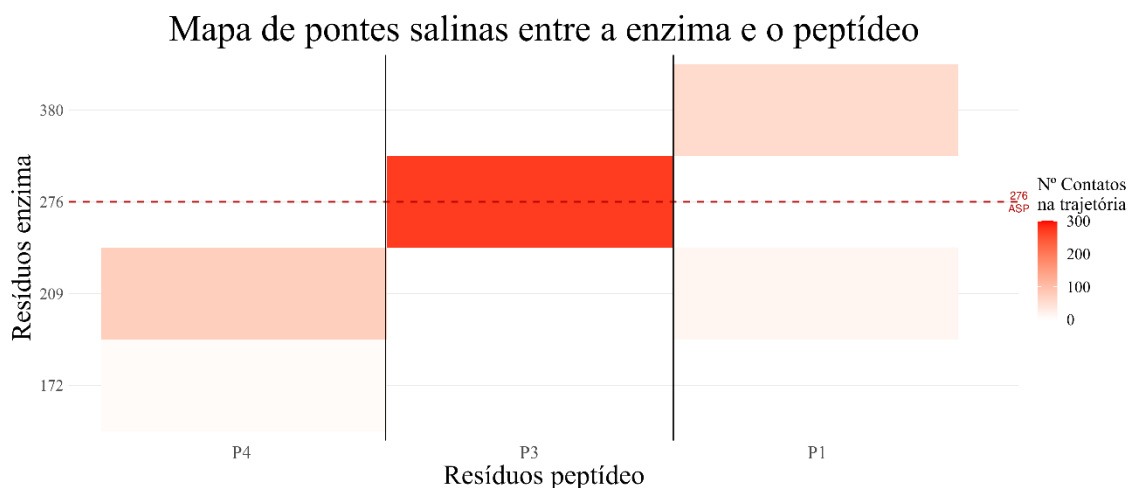


Figura 5.41 – Mapa de possíveis pontes salinas entre o peptídeo e a enzima PCSK5 durante a simulação realizada a partir da docagem com Haddock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Nota-se que P3 pode realizar interações de ponte salina com ASP276, enquanto P1 não parece interagir com nenhum dos resíduos do bolso S1.

Nas simulações a partir da docagem com CABSdock, P1 fica melhor posicionado em relação ao bolso de ligação S1. Conforme mostrado na **Figura 5.42**, P1 pode realizar pontes salinas com o aspartato na borda do subsítio S1 (ASP276), e também interagir com a prolina (PRO274) do mesmo sítio. Indicando, assim como na simulação com a PCSK1 a partir da docagem com CABSdock, uma afinidade entre a enzima e o peptídeo. Porém, não foram observados os contatos de P1 com S1 da mesma forma que nas simulações com a Furina e PCSK2. Assim não conseguimos listar PCSK5 como candidata à ativação do vírus.

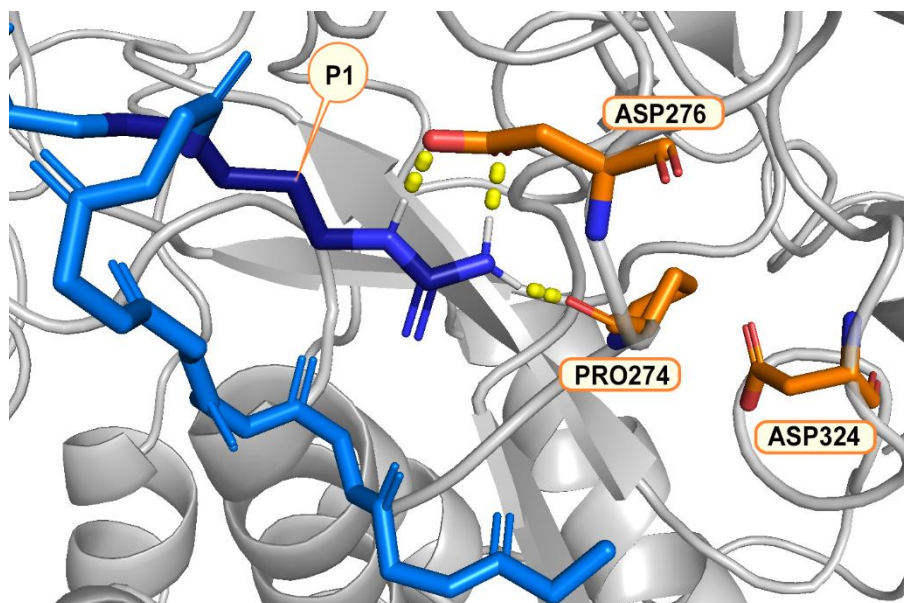


Figura 5.42 - Interações entre P1 e a enzima PCSK5, ao final da simulação a partir da docagem com CABSdock. A cadeia lateral de P1 (azul escuro) pode interagir com ASP276 na borda do bolso de ligação S1, e também com PRO274 do mesmo subsítio.

5.4.6 PCSK6

Em ambas as simulações com a enzima PCSK6, P1 não se posicionou da forma esperada para se ligar ao bolso de ligação S1. Ao término da simulação com Haddock, por exemplo, podemos ver que P1 ficou muito distante do subsítio (**Figura 5.43**).

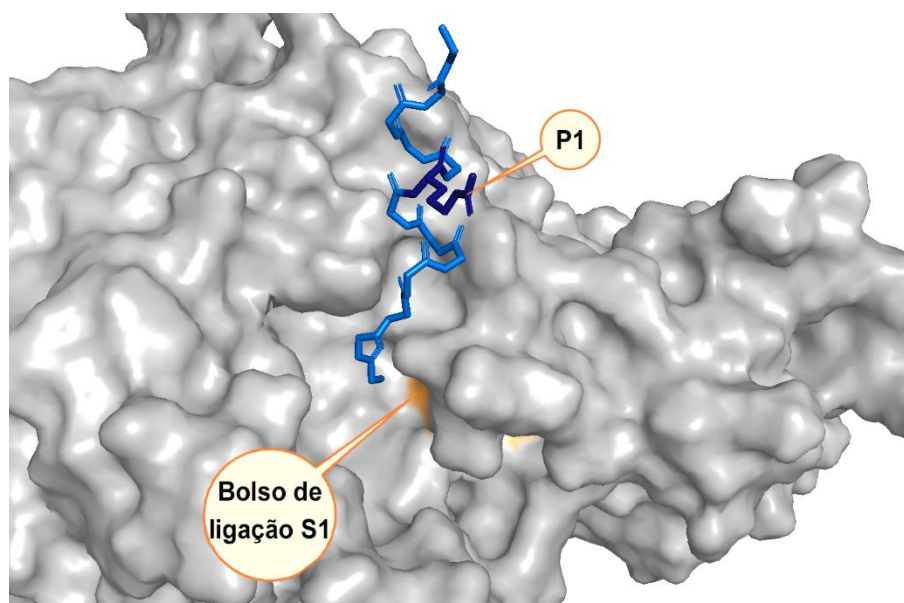


Figura 5.43 – Posição final do peptídeo na simulação a partir da docagem à PCSK6 usando Haddock. Observa-se que a cadeia lateral de P1 (azul escuro) fica bem distante do bolso de ligação S1 (laranja).

Dessa forma, P1, e até mesmo P3, ficaram a uma distância muito grande dos resíduos do bolso de ligação, sem realizar qualquer tipo de interação com esses.

5.4.7 PCSK7

Nas simulações com a enzima PCSK7, observamos que P1 também não está interagindo de com o bolso de ligação S1. Analisando o mapa de contatos entre a proteína e o peptídeo na simulação a partir da docagem com Haddock, podemos ver que P1, realiza poucos e fracos contatos distantes do bolso de ligação S1 (**Figura 5.44**).

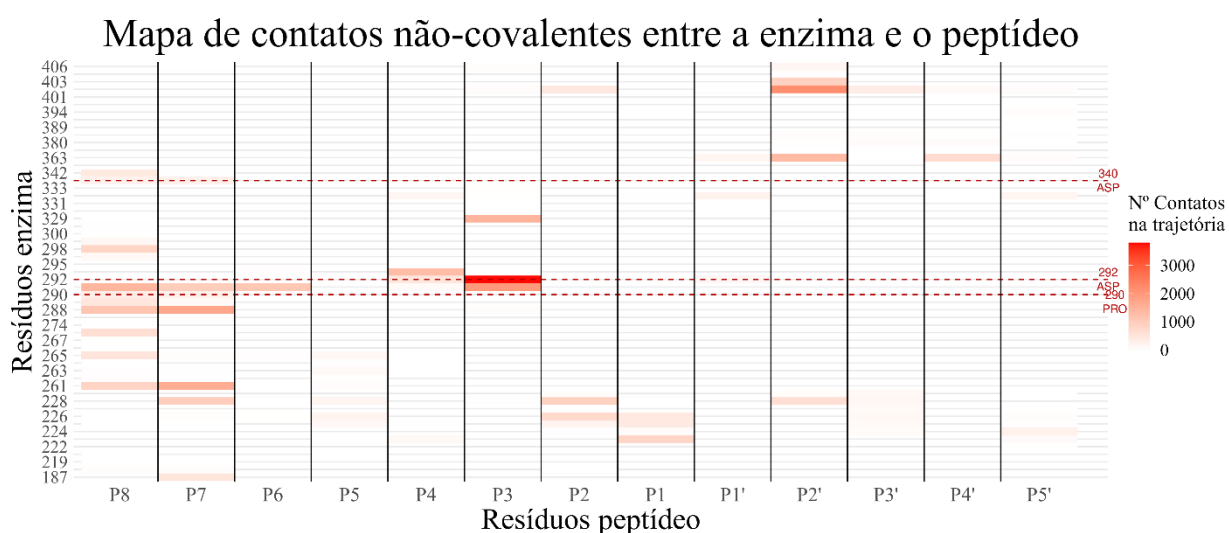


Figura 5.44 – Mapa de possíveis contatos não-covalentes entre o peptídeo e a enzima PCSK7 durante a simulação realizada a partir da docagem com Haddock. As cores do mapa estão relacionadas com a frequência das interações, mais claras reportam frequências mais baixas e mais escuras frequências mais altas. Observa-se que P1 não realiza nenhuma interação importante com o subsítio S1 do bolso de ligação, e podem ocorrer interações entre P3 e ASP292.

Podemos ver também na **Figura 5.44** que P3 pode estar interagindo com o aspartato 292 (ASP292) que está localizado na borda do bolso de ligação S1. Dessa forma, P3 ocupa a entrada do subsítio S1, impedindo que P1 se posicione corretamente para a catálise. Indicando, mais uma vez, que pode haver uma competição entre esses dois resíduos *in silico*, a depender da estrutura do sítio catalítico da enzima, e talvez *in vivo* também.

Na simulação a partir da docagem com CABSdock, P1 chega a interagir por algum tempo com ASP292 na borda do bolso de ligação S1. Porém, essa interação não é duradoura, e se desfaz ao término da simulação. Esse comportamento das interações entre P1 e os resíduos do bolso ao longo da simulação pode ser observada no vídeo disponível no **APÊNDICE C**.

5.4.8 PCSK9

As simulações com a enzima PCSK9 foram as que apresentaram menor afinidade entre a enzima e o peptídeo. Observando os modelos finais provenientes da simulação é possível notar que o peptídeo se desloca pela superfície da proteína e termina a simulação bem distante do bolso de ligação (**Figura 5.45**).

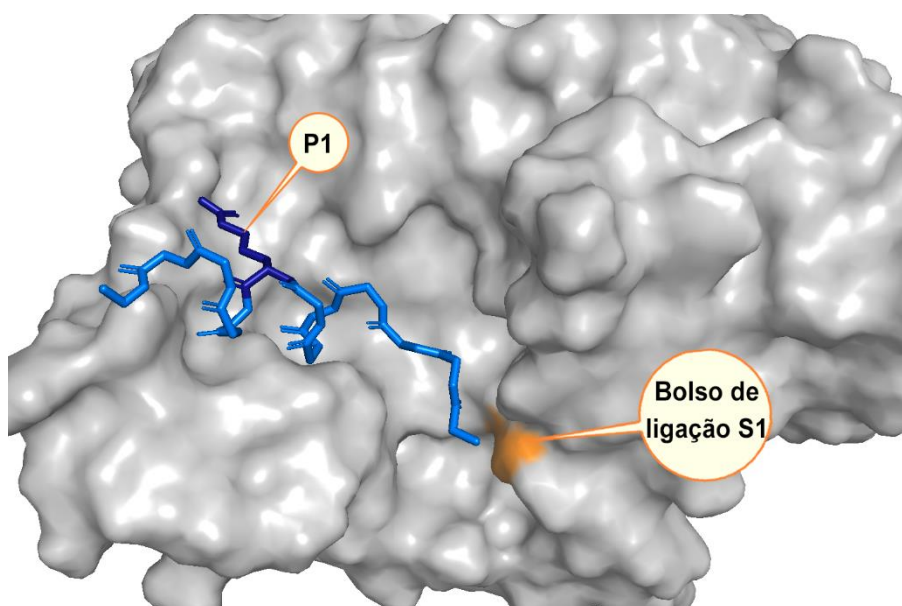


Figura 5.45 – Posição final do peptídeo na simulação a partir da docagem à PCSK9 usando Haddock. Observa-se que a maior parte do peptídeo, incluindo P1 (azul escuro), fica bem distante do bolso de ligação S1 (laranja).

Esse distanciamento também pode ser observado através do gráfico de distâncias entre P1 e os resíduos do bolso de ligação S1. Nota-se que P1 fica posicionado distante do bolso S1 desde a docagem, e, durante a simulação, essa distância aumenta (**Figura 5.46**). Sugerindo que a posição inicial do peptídeo obtida

na docagem não foi a ideal para alcançar a estabilidade local, e, por isso, ele se movimenta durante a simulação.

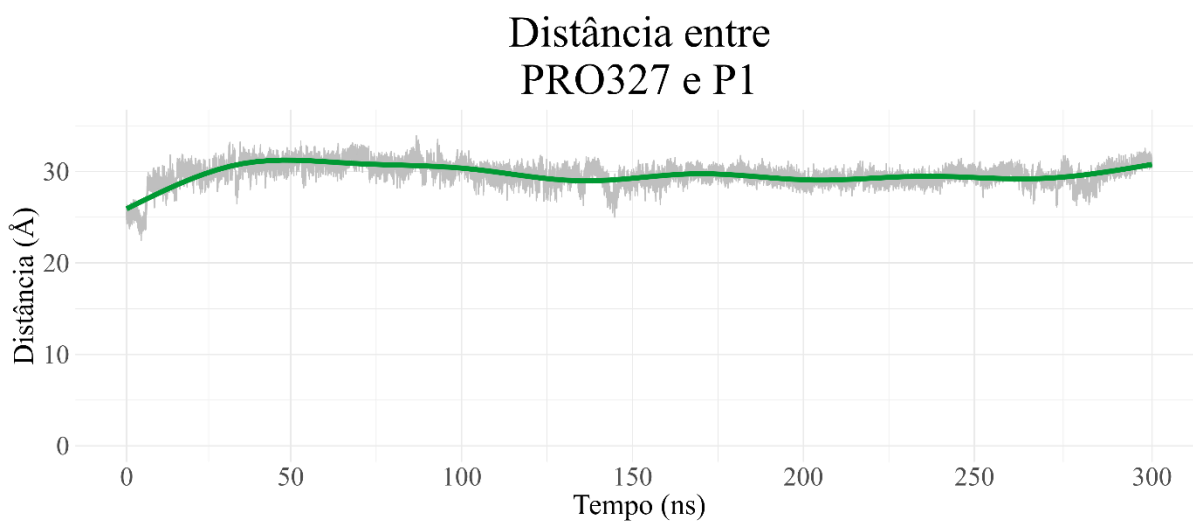


Figura 5.46 - Gráfico de distâncias em Å entre P1 e PRO237 do bolso de ligação S1 da enzima PCSK9, durante a simulação de dinâmica a partir da docagem com Haddock.

5.5 Aplicação computacional Pamda

Para realizar as análises apresentadas nesse trabalho, foi necessário o estudo e desenvolvimento na linguagem de escrita do VMD: *tcl*. E, devido ao grande número de análises realizadas em diversas simulações de dinâmica molecular, também se fez necessário automatizar esse processo, isto é, criar *pipelines* que executassem, de forma automática, todas as análises necessárias para cada simulação.

Para facilitar estudos futuros, reunimos todas essas análises em um único programa que pudesse ser livremente distribuído. O programa foi batizado de Pamda (do inglês: Practical Analysis of Molecular Dynamics). Foi realizada a sua *containerização* em Docker permitindo uma instalação mais acessível e amigável ao usuário, não requerendo a instalação de dependências ou configurações adicionais. No entanto, como a maioria das análises são feitas através do VMD, é necessário obter a licença deste e instalá-lo separadamente.

A aplicação Pamda deve ser executada através do terminal Linux por linha de comando. Além de possibilitar análises sem requerer conhecimentos específicos de programação pelo usuário, o programa também oferece a possibilidade de analisar dinâmicas moleculares muito grandes e que não podem ser carregadas inteiramente na memória RAM do computador.

O *download*, o guia de instalação completo e documentação do programa estão disponíveis no endereço web listado no **APÊNDICE D**. O código do programa se encontra disponível no repositório do GitHub indicado no **APÊNDICE E**.

O programa recebe como entrada um arquivo estrutural e um arquivo de trajetória de dinâmica molecular. O arquivo estrutural pode estar no formato *psf*, *pdb* ou *gro*. Já para o arquivo de trajetória de dinâmica molecular são aceitos os formatos *dcd* ou *xtc*. Desta forma, o programa processa resultados obtidos pela simulação com NAMD e também com Gromacs.

Ao todo, a aplicação é capaz de realizar cinco tipos de análises. São elas: RMS, contatos entre cadeias, distâncias, SASA e energias. Além de contar também com a opção de gerar gráficos automaticamente para cada uma das análises requisitadas.

5.5.1 RMS

Os cálculos de RMS são realizados conforme descrito na seção 4.5.1. Para cada arquivo de trajetória o programa gerará três relatórios:

- O primeiro contendo RMSD *Fit* e RMSD *no Fit* por *frame*, para o complexo inteiro e para cada cadeia separadamente;
- O segundo contendo RMSD por *frame* para cada resíduo individualmente;
- E o terceiro contendo o RMSF de cada resíduo.

5.5.2 Contatos

Os contatos são medidos conforme descrito na seção 4.5.2. Os contatos analisados pelo programa são sempre entre cadeias. Para cada par de cadeia, serão gerados relatórios de três tipos de contatos:

- Contatos não-covalentes de todos os tipos, medidos com base na distância de *cutoff* estabelecida pelo usuário;
- Interações de hidrogênio, medidas conforme a mesma distância de *cutoff* e o ângulo da ligação. No relatório também serão inclusos dados indicando os doadores e os aceptores de elétron em cada interação;
- Pontes salinas, medidas entre oxigênios de resíduos polares para nitrogênios de resíduos apolares. Como uma ponte salina entre dois resíduos pode apresentar diversas configurações geométricas, envolvendo múltiplos átomos (DONALD; KULP; DEGRADO, 2011; PYLAEVA; BREHM; SEBASTIANI, 2018), optamos por relatar todas possíveis formações ao usuário, cabendo a este uma triagem mais rigorosa para avaliar as pontes que realmente podem ocorrer na simulação.

5.5.3 Distâncias

O programa gera um relatório contendo as distâncias por *frame* para cada par de índices entrados pelo usuário. São aceitos tanto índices de átomos quanto de resíduos. No caso dos resíduos, as distâncias serão medidas entre os centros de massa de cada um. O cálculo de distância é feito conforme especificado na seção 4.5.3.

5.5.4 SASA e BSA

As análises de superfície acessível a solvente (SASA) são feitas conforme o descrito pela seção 4.5.4.

O programa gerará um relatório de SASA por *frame* para cada cadeia e também para resíduos especificados pelo usuário. Também serão inclusos dados de área de superfície encoberta (BSA) por *frame* para as mesmas seleções. O cálculo de BSA se encontra descrito na seção 4.5.4.

5.5.5 Energias

São medidas diversas energias por *frame* para cada molécula, e, quando mais de uma cadeia está presente, também serão medidas as energias de interação entre cada par de cadeias. A análise dessas energias é feita conforme descrito na seção 4.5.5.

As energias medidas para toda a molécula são energias de ligação, ângulos, diédricas, impróprias, eletrostáticas, van der Waals, conformacionais, *nonbond* e totais. Já as energias de interação medidas entre cadeias são energias eletrostáticas, van der Waals, *nonbond* e totais.

5.5.6 Gráficos

Adicionalmente, o usuário pode optar por gerar gráficos para todas as análises mencionadas acima. Os gráficos são gerados utilizando R. Exemplos dos gráficos gerados são aqueles disponíveis no endereço do **APÊNDICE B**.

5.5.7 Análise de simulações muito grandes

Um dos problemas que enfrentamos no desenvolvimento desse estudo foi a análise de simulações de dinâmica molecular muito grandes, que não cabem na memória RAM do computador para uma análise convencional. Para solucionar esse problema, a abordagem utilizada foi carregar um *frame* individualmente na memória, realizar as análises necessárias e depois deletar o *frame* antes de carregar o próximo.

Porém, isso acarretou um novo problema. Como as simulações foram rodadas usando condições de fronteira periódicas (PBC), através da abordagem de carregar cada frame de uma vez, o VMD não é capaz de retirar as fronteiras da dinâmica de forma correta para realizar as análises. Podendo gerar *frames* distorcidos, nos quais a molécula se encontra dividida e com ligações não naturais (**Figura 5.47**).

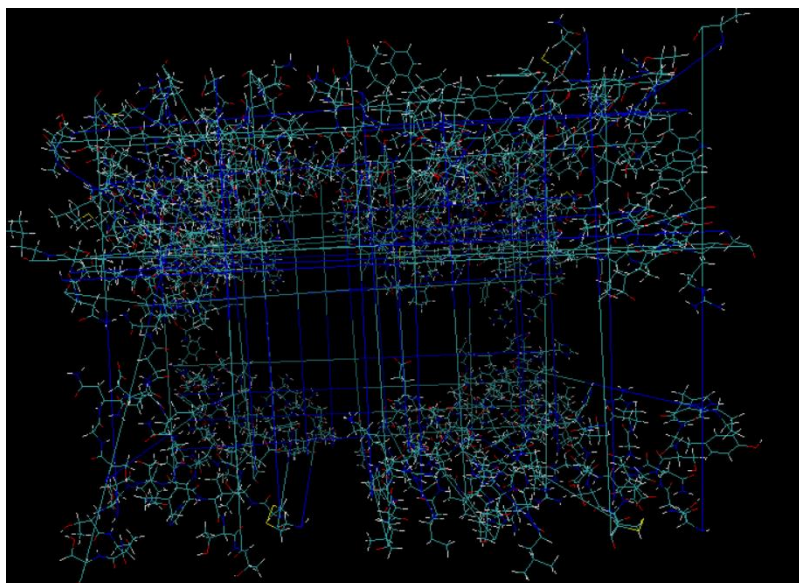


Figura 5.47 - Figura mostrando uma molécula corrompida ao tentar retirar PBC utilizando o VMD sem informações iniciais da trajetória.

Para resolver esse problema, modificamos o código para manter 1% da dinâmica na memória. Logo, ao invés de descartar todos os *frames*, nós mantemos o equivalente a um por cento do total para que o VMD consiga analisar a trajetória e retirar as condições de fronteiras corretamente antes das análises.

6 DISCUSSÃO

A partir dos resultados obtidos com as análises das simulações de dinâmica molecular entre o peptídeo e as PCSKs, podemos caracterizar e descrever as possíveis interações que ocorrem entre as estruturas. É interessante destacar a importância das simulações de dinâmica molecular para tal caracterização. Uma vez que, realizando estudos apenas a partir da docagem, nós não veríamos alguns dos contatos entre a enzima e o peptídeo, em especial entre P1 e os resíduos do fundo do bolso de ligação S1. Como exemplo, podemos citar a interação entre P1 e PRO271 da PCSK2 que, conforme observamos na **Figura 5.25**, estão distantes e não interagem logo após a docagem, mas se aproximam após cerca de 100ns de simulação e passam a poder interagir.

A simulação de dinâmica molecular implicou em um deslocamento de até sete angstroms na mensuração de *no Fit* RMSD após 100ns, seguindo de estabilização para Furina e PCSK2, exemplificando a relevância do procedimento. A obtenção de contatos esperados e/ou produtores de estabilidade envolvendo os resíduos mais importantes, principalmente o resíduo P1, suportam o uso da metodologia de docagem seguida de dinâmica molecular para a investigação realizada aqui. O trabalho também acrescenta uma contribuição em forma de aplicativo computacional para automatizar as análises aqui apresentadas, sendo que os resultados apresentados ilustram sua aplicação. Analisaremos a seguir por partes as contribuições do trabalho.

6.1 Furina

Na seção 5.4.1, mostramos que o peptídeo docado na Furina com Haddock se estabiliza após 100ns na simulação. Além disso, P1 fica inserido e encoberto pelo bolso de ligação S1, evidenciado pelos gráficos de BSA do resíduo, onde pode interagir com resíduos importantes para ligação à enzima. As interações entre P1 e o bolso de ligação S1 são essenciais para reconhecimento do substrato e início da ação catalítica da enzima (HEDSTROM, 2002; WAHLGREN *et al.*, 2011).

No subsítio S1 do bolso de ligação, P1 pode realizar pontes salinas com as cadeias laterais dos resíduos ASP258 e ASP306 da Furina, condizendo com aquilo que foi descrito nos estudos de Becker *et al.* e Tian (BECKER *et al.*, 2010; TIAN, S.,

2009), para uma ligação da enzima a um substrato com potencial catalítico. A participação desses dois aspartatos na ligação da Furina ao substrato, também foi descrita no artigo em *preprint* de 2022 de Jawaid *et al.* (JAWAID *et al.*, 2022), que, assim como o presente estudo, investigou a interação da Furina com o sítio de clivagem polibásico da proteína espícula do SARS-CoV-2. Os resultados descritos aqui, portanto, reforçam a hipótese de que a interação da arginina da proteína espícula com esses aspartatos é essencial para a ligação, e posterior clivagem, da espícula do coronavírus pela enzima Furina. Desta forma, essa interação poderia ser investigada para as demais enzimas relacionadas.

Além desses resíduos, P1 pode interagir com PRO256 do subsítio S1, que também é encontrada participando da ligação dos substratos conhecidos da enzima (BECKER *et al.*, 2010), indicando que é mais um resíduo que participa na fixação e estabilização do peptídeo à enzima para clivagem. Uma interação passageira também foi identificada entre P1 e GLU331 do bolso de ligação S1. Essa interação, que só foi vista em Furina, ocorre apenas nos primeiros momentos da simulação, com os resíduos se afastando ao decorrer da trajetória, e, portanto, não parece de grande importância na fixação de P1 em S1.

Conforme mostrado nas análises das simulações com Furina, a partir da docagem com Haddock, pode ocorrer ainda uma interação entre P3 e o resíduo ASP191 da enzima (**Figura 5.13**). Esse resíduo ASP191 é descrito por Tian como integrante do subsítio S2 do bolso de ligação (TIAN, S., 2009), e, portanto, espera-se que ele contacte P2. No entanto, como P2 na sequência do SARS-CoV-2 é uma alanina e não um resíduo básico, conforme a sequência padrão dos substratos das PCSKs (TIAN, S.; JIANHUA, 2010), a arginina P3 pode estar assumindo o papel esperado de P2 e contactando ASP191. Nossos dados suportam essa possibilidade.

6.2 PCSK2

Além da enzima Furina, também observamos uma afinidade entre o peptídeo e a enzima PCSK2, nas simulações de dinâmica molecular a partir da docagem com Haddock. Conforme exposto na seção 5.4.2, é possível observar uma estabilidade do peptídeo após 100ns, nessa simulação, através dos gráficos de RMSD. Além disso, vemos nos gráficos de BSA que P1 fica bem inserido na cavidade do subsítio S1, onde

notamos que a enzima apresenta interações com o peptídeo muito semelhantes às aquelas apresentadas entre a Furina e o peptídeo. Tornando-a uma possível candidata à clivagem da proteína espícula do coronavírus.

P1 pode interagir via pontes salinas com os ácidos aspárticos ASP273 e ASP320 do bolso de ligação S1 da proteína PCSK2. Observando o alinhamento da **Figura 1.7**, vemos que esses resíduos correspondem aos resíduos ASP258 e ASP306 da Furina, respectivamente. Resíduos cuja participação na ligação das PCSKs aos seus substratos é frequentemente descrita na literatura. Essas interações, portanto, possivelmente atuam na estabilização e fixação de P1 no bolso de ligação S1, criando as condições para que ocorra a catálise da hidrólise da ligação entre P1 e P1'.

P1 também parece realizar ligações do tipo não-covalentes com PRO271 do bolso de ligação S1 da enzima. Dando mais suporte à hipótese de que há uma afinidade entre a PCSK2 e o sítio de clivagem polibásico presente no genoma do SARS-CoV-2.

Além dessas interações com P1, a enzima PCSK2 parece interagir com o peptídeo através de uma interação entre ASP203 e P3. Com base nos estudos de Becker *et al.* e Tian (BECKER *et al.*, 2010; TIAN, S., 2009), e observando o alinhamento da **Figura 1.7**, esse aspartato não parece fazer parte de nenhum subsítio do bolso de ligação.

Conforme descrito por Tian e Jianhua (TIAN, S.; JIANHUA, 2010), e visto no alinhamento da **Figura 1.7**, o ASP191 que compõem o subsítio S2 da Furina sofreu uma mutação para uma fenilalanina (PHE205) em PCSK2. No entanto, dado à proximidade de ASP203 à PHE205, e à sua posição na estrutura da enzima, inferimos que ele está desempenhando o mesmo papel que ASP191 na Furina, agindo como um substituto para o ácido aspártico que foi perdido nessa posição da PCSK2. Logo, ASP203 estaria compondo o subsítio S2 nessa PCSK, participando da ligação do substrato à enzima. Nossos dados, portanto, suportam que a substituição de um resíduo conservado na família não desfavorece a participação da posição na associação do substrato.

Devido ao comportamento do peptídeo na simulação, e às interações observadas entre o mesmo e a PCSK2, podemos inferir que há uma afinidade entre

as duas estruturas, tornando a PCSK2 uma enzima com potencial de clivar a proteína espícula do SARS-CoV-2. Essa ação enzimática, caso comprovada, pode estar atuando no aumento da infecciosidade viral (AMIDEI; DOBROVOLNY, 2022; PAPA *et al.*, 2021; PEACOCK *et al.*, 2021), nos tecidos onde a PCSK2 é mais ativa.

A PCSK2, segundo dados da ferramenta humanExpProfiles 2.1 (ORSINE, 2020), consultada usando os bancos de dados GTEx e IBM, é expressa diferencialmente na glândula tireoide. O que significa que, caso confirmada sua participação na clivagem da proteína espícula do coronavírus, ela pode estar atuando na infecção viral do órgão.

Artigos recentes, como o de Muller *et al.*, mostram que infecções pelo novo coronavírus estão associadas a casos de tireoidite atípica causados pelo vírus SARS-CoV-2 infectando diretamente a glândula (MULLER *et al.*, 2020). Dando suporte à hipótese levantada aqui de que a PCSK2 pode estar atuando na catálise da proteína espícula, o que amplificaria a infecção viral no local.

Dessa forma, se fazem urgentes estudos mais aprofundados sobre a participação dessa PCSK na clivagem da proteína do SARS-CoV-2. Pois, caso confirmada sua participação, esta enzima deve ser alvo de pesquisas e estudos a fim de desenvolver inibidores para sua atividade, uma vez que não foram encontrados inibidores descritos para ela no banco de dados DrugBank (WISHART *et al.*, 2018).

Tanto nas simulações com a enzima Furina, quanto com a PCSK2, nas quais a interação entre proteína e peptídeo apresentou maior afinidade, não foi observada uma aproximação dos resíduos da tríade catalítica ao resíduo P1. O que pode ser explicado por diversos motivos. Como por exemplo, as simulações foram executadas por apenas 300ns, o que pode não ter sido tempo o suficiente para nos permitir observar tal aproximação. Ou ainda, como a aproximação da tríade também acompanha mudança de cargas nos resíduos da mesma, conforme mostrado na seção 1.3.1, e as simulações utilizadas não realizam essas trocas de cargas, podem não ter sido criadas as condições necessárias para um ajuste induzido da tríade catalítica. Portanto, se fazem necessários estudos utilizando metodologias de simulações de dinâmica molecular diferentes, ou executando as simulações por tempo maior, para podermos tentar verificar esse ajuste induzido.

Observamos resultados distintos com CABSdock, considerados desfavoráveis por não produzirem os contatos relevantes detectados com as docagens usando Haddock. Logo, o acompanhamento desses contatos pode ser produtivo para acompanhar o teste de outras estratégias.

6.3 Demais PCSKs

De forma geral, os resultados observados nas enzimas Furina e PCSK2, principalmente no que diz respeito às interações de P1 com S1, não se repetiram nas demais PCSKs. Em especial a interação de P1 com o aspartato no fundo do subsítio S1 (ASP306 na Furina), que só foi observada nas simulações com Furina e PCSK2. A **Tabela 6.1**, resume as interações entre P1 e os resíduos do S1 durante todas as simulações. Nela, estão representados a porcentagem de *frames* da trajetória em que P1 está a uma distância de contato (3,3 Å) de cada resíduo do subsítio.

Conforme podemos ver na tabela, as PCSKs 1 e 5 foram as que mais apresentaram afinidade entre P1 e o bolso de ligação S1, depois da Furina e PCSK2, apesar de que em nenhuma dessas duas simulações, P1 ficou completamente inserido no subsítio S1 contactando o aspartato no fundo do mesmo. Em ambas, P1 fica posicionado na entrada do subsítio, interagindo principalmente com o aspartato na borda do mesmo.

Tabela 6.1 – Porcentagem de *frames* em que P1 está até 3,3 Å de cada resíduo do bolso de ligação S1 para todas as simulações. A coluna “Docagem” refere-se à metodologia de docagem da qual partiu a simulação. E o ácido aspártico indicado com um asterisco (ASP*) refere-se ao aspartato no fundo do subsítio S1 (ASP306 na Furina). Uma escala de cor verde foi aplicada para destacar as interações que ocorrem com mais frequência.

PCSK	Docagem	PRO	ASP	ASP*
Furina	Haddock	100%	100%	87%
	CABSdock	0%	0%	0%
PCSK1	Haddock	0%	0%	0%
	CABSdock	9%	99%	0%
PCSK2	Haddock	78%	100%	86%
	CABSdock	0%	64%	0%
PCSK4	Haddock	0%	0%	0%
	CABSdock	0%	0%	0%
PCSK5	Haddock	9%	0%	0%
	CABSdock	41%	95%	0%
PCSK6	Haddock	0%	0%	0%
	CABSdock	0%	8%	0%
PCSK7	Haddock	0%	0%	0%
	CABSdock	0%	64%	0%
PCSK9	Haddock	0%	0%	0%
	CABSdock	0%	0%	0%

Apesar de as interações observadas com as PCSKs 1 e 5 não terem se assemelhado àquelas observas nas simulações com Furina e PCSK2, estudos desenvolvidos por Jaimes *et al.* sugerem que a PCSK1 pode clivar a espícula, embora com menor eficiência do que a Furina (JAIMES; MILLET; WHITTAKER, G. R., 2020). Logo, essas poucas interações observadas aqui podem indicar uma baixa afinidade à

espícula correspondendo a essa baixa eficiência de clivagem descrita por Jaimes et al. Mais estudos são necessários para confirmar tal afinidade, principalmente com respeito à PCSK5.

Por exemplo, tentativas de docagem com parâmetros diferentes, para tentar um melhor posicionamento inicial do peptídeo sobre a enzima, podem ser realizadas para tentar um melhor posicionamento de P1 em relação ao bolso de ligação S1. Uma perspectiva seria docar as conformações de peptídeo estabilizadas por Furina ou PCSK2 nas demais enzimas.

Além disso, simulações maiores do que 300ns talvez possam permitir que P1 se acomode de forma completa no subsítio S1, caracterizando um encaixe mais parecido com aquelas observadas na Furina e PCSK2, e que indicam maior afinidade com o possível substrato.

Para as demais PCSKs, não foram vistos indícios de estabilidade entre o peptídeo e as enzimas, conforme a descrita na seção 1.3.2, e que caracterizasse um modo de ligação para que ocorra a clivagem. Diversos fatores, no entanto, podem contribuir para que não observássemos esses elementos, não indicando necessariamente que elas não ocorram *in vivo*.

O uso de modelos para algumas PCSKs, por exemplo, pode ser um dos fatores que enviesasse os resultados obtidos por esse estudo. Como apenas a Furina e PCSK9 possuíam estrutura tridimensional cristalográfica, o fato de usarmos modelos, que não são representações perfeitas da enzima real, pode fazer com que os resultados obtidos pelas simulações também não sejam iguais àqueles que seriam observados *in vivo*.

Além disso, o tempo de simulação executado pode não ter sido o suficiente para permitir que P1 se direcionasse para o bolso de ligação S1. Em especial nas simulações em que P1 fica próximo do subsítio, como nas simulações com as enzimas PCSK1 ou PCSK5, já citadas, ou com as enzimas PCSK7 e PCSK4 em que o peptídeo fica estacionado sobre a enzima, tendo P1 próximo ao S1. Talvez essas simulações, se executadas por mais tempo, permitiriam que P1 se direcionasse e se acomodasse de forma satisfatória no bolso de ligação S1.

A docagem e posição inicial do peptídeo também podem interferir nos resultados observados com as simulações. Logo estratégias de docagem diferentes, usando outros métodos ou parâmetros, podem gerar resultados e interações diferentes entre o peptídeo e as enzimas.

6.4 Sequência do sítio polibásico e variantes

Por fim, é importante destacar que os resultados apresentados indicam uma possível competição entre duas argininas do sítio polibásico, P3 e P1, pela ligação ao bolso de ligação S1. O que foi mostrado nas seções 5.4.2, 5.4.3, 5.4.5 e 5.4.7, nas quais vimos que P3 pode ocupar o subsítio S1, deixando P1 mais distante e sem interagir com os resíduos do mesmo. Isto pode ocorrer porque a arginina P3 não foi inserida em um local ótimo para a clivagem, tendo como base a sequência típica dos substratos dessas enzimas: R-X-(R/K/X)-R↓-S (DAHMS *et al.*, 2014a; REMACLE *et al.*, 2008; TIAN, S., 2009).

O fato da sequência do sítio polibásico presente na espícula do SARS-CoV-2 não corresponder a essa sequência comum dos substratos já foi discutido por outros artigos. Garry, por exemplo, argumenta que esse sítio de clivagem seria “subótimo” para a ação enzimática (GARRY, 2022). E Jawaid discute que uma lisina na posição P2 teria maior potencial de realizar ligações de hidrogênio com os resíduos do bolso de ligação, o que, segundo ele, caracterizaria uma ligação melhor entre a enzima e a espícula (JAWAID *et al.*, 2022).

Segundo Tian, apenas 20% dos substratos conhecidos não possuem um resíduo básico na posição P2, nesses casos, são observados que resíduos pequenos ocupam esse local (TIAN, S., 2009). E é isso o que vemos na sequência do sítio polibásico presente em SARS-Co-V-2, onde o resíduo básico, arginina, ocupou a posição P3 e um resíduo pequeno, alanina, ocupa P2.

Nesse caso, P3 é quem está interagindo com os resíduos típicos do bolso de ligação S2, de cargas negativas, conforme vimos nas seções 5.4.1 e 5.4.2, com as enzimas Furina e PCSK2. E o fato de um resíduo pequeno ocupar P2 permite que P1 ainda realize as interações típicas com o bolso de ligação S1.

No entanto, essas divergências entre a sequência do sítio polibásico observada no SARS-CoV-2 e a sequência mais comum dos substratos, podem significar que o sítio de clivagem polibásico ainda não está otimizado para clivagem pela Furina e outras PCSKs. Dessa forma, mutações que envolvam P3 e P2 talvez possam tornar a clivagem da espícula mais eficaz.

Estudos *in silico* podem ser realizados, fazendo as mutações no peptídeo, em especial com a mutação de uma lisina em P2, para confirmar se essas trocas realmente acarretariam em uma melhora de ligação da espícula à Furina e às demais PCSKs. Caso confirmado, essas mutações em variantes do vírus podem representar um novo risco à saúde populacional, uma vez que poderiam acarretar em uma maior atividade enzimática, e, conseqüentemente, ao aumento da infecção viral.

7 CONCLUSÃO

Conforme discutido no capítulo anterior, os resultados obtidos por esse estudo nos permitiram caracterizar e descrever a nível molecular as interações que podem ocorrer entre a enzima Furina e o sítio de clivagem polibásico presente na proteína espícula do coronavírus. Interações que viabilizam a catálise da proteína pela enzima, separando as subunidades S1 e S2 da espícula, e contribuindo para o aumento da infecciosidade viral (AMIDEI; DOBROVOLNY, 2022; PAPA *et al.*, 2021; PEACOCK *et al.*, 2021).

Nós mostramos que ASP258 e ASP306 da Furina podem participar estabilizando P1 no bolso de ligação S1, o que já foi sugerido por Jawaid (JAWAID *et al.*, 2022), e reforçado aqui. Além disso, mostramos que P1 também pode interagir com PRO256 do subsítio S1. E P3, um resíduo básico, interage com ASP191 que integra o S2. Essa interação de P3 ao S2 enquanto P1 se encaixa ao S1, conforme discutida na última seção, é possível porque na posição P2 do sítio polibásico encontra-se uma asparagina, que é um aminoácido pequeno

Com relação à sequência do sítio polibásico, apresentamos e discutimos a possibilidade de que o sítio presente no novo coronavírus se encontra numa forma “subótima”. Mais estudos podem ser realizados para testar e caracterizar a ligação das enzimas à proteína espícula para as variantes já conhecidas, e para possíveis variantes que, alterando a sequência do sítio polibásico, tentem melhorar a ligação entre o peptídeo e as PCSKs. Pois, caso ocorram essas variantes que potencializem a ação enzimática, elas podem representar sérios problemas à saúde pública, reforçando a necessidade de campanhas de vacinação preventivas, e desenvolvimento de medicamentos que realizem a inibição das enzimas que atuam no aumento da infecciosidade viral, e combatam a doença.

Mostramos que a PCSK1 e a PCSK5 foram outras serina proteases que apresentaram uma certa afinidade pelo peptídeo. Embora as ligações realizadas entre essas enzimas com o peptídeo não tenham sido iguais às apresentadas nas simulações com a Furina, pudemos notar uma proximidade de P1 ao bolso de ligação S1. Nesses casos, P1 parece interagir com o aspartato localizado na borda do bolso de ligação S1, ASP272 na PCSK1 e ASP276 na PCSK5. Por isso, são necessárias mais pesquisas para elucidar a atuação dessas PCSKs sobre a proteína espícula.

Para as demais PCSKs, com exceção da PCSK2, não foram encontradas evidências de sua participação no processo de catálise da proteína viral. Isso, no entanto, não significa necessariamente que elas não atuam sobre a espícula. Conforme discutido na seção 6.3, diversos aspectos da metodologia utilizada nesse estudo podem contribuir para esse resultado, não representando necessariamente o que seria observado *in vivo*.

Por fim, mostramos que a enzima PCSK2 apresentou possível afinidade pelo peptídeo simulando a proteína espícula, tornando-a uma candidata a realizar a clivagem da proteína viral, da mesma forma que a Furina. Assim como na enzima Furina, descrevemos a nível molecular as interações que podem ocorrer entre a enzima e a proteína viral. Mostrando que P1 pode interagir através de pontes salinas com ASP273 e ASP320, análogos a ASP258 e ASP306 da Furina, respectivamente. P1 também interage com PRO271. E P3 pode interagir com ASP203, que apesar de não ser análogo ao ASP191 em Furina, aparentemente desempenha a mesma função.

Dessa forma, mais estudos são necessários para confirmar a participação da PCSK2 na catálise da proteína espícula, pois isso significaria que a enzima PCSK2 estaria atuando na dispersão do vírus, em especial na glândula tireoide, onde essa enzima é mais expressa. Por se tratar de uma enzima que não possui inibidores conhecidos, há uma grande necessidade e potencial para o desenvolvimento de novos medicamentos que visem combater a sua ação catalítica, caso confirmada.

REFERÊNCIAS

- ABCAM. Structural and functional mechanism of SARS-CoV-2 cell entry. 19 maio. 2022. Disponível em: <<https://www.abcam.com/content/structural-and-functional-mechanism-of-sars-cov-2-cell-entry>>. Acesso em: 28 nov. 2022.
- ADU-AGYEIWAH, Y.; GRANT, M. B.; OBUKHOV, A. G. The Potential Role of Osteopontin and Furin in Worsening Disease Outcomes in COVID-19 Patients with Pre-Existing Diabetes. **Cells**, 23 nov. 2020. v. 9, n. 11. Disponível em: <<https://pmc/articles/PMC7700577/>>. Acesso em: 24 nov. 2022.
- AHERN, K.; RAJAGOPAL, I.; TAN, T. **BIOCHEMISTRY Free For All**. 3. ed. [S.l.]: NC - Creative Commons, 2018.
- AlphaFold Protein Structure Database. [s.d.]. Disponível em: <<https://alphafold.ebi.ac.uk/>>. Acesso em: 24 nov. 2022.
- AMIDEI, A.; DOBROVOLNY, H. M. Estimation of virus-mediated cell fusion rate of SARS-CoV-2. **Virology**, 1 out. 2022. v. 575, p. 91. Disponível em: <<https://pmc/articles/PMC9449781/>>. Acesso em: 28 nov. 2022.
- ANDERSEN, K. G. *et al.* The proximal origin of SARS-CoV-2. **Nature Medicine** 2020 **26:4**, 17 mar. 2020. v. 26, n. 4, p. 450–452. Disponível em: <<https://www.nature.com/articles/s41591-020-0820-9>>. Acesso em: 24 nov. 2022.
- BATEMAN, A. *et al.* UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, 8 jan. 2021. v. 49, n. D1, p. D480–D489. . Acesso em: 25 nov. 2022.
- BECKER, G. L. *et al.* Potent Inhibitors of Furin and Furin-like Proprotein Convertases Containing Decarboxylated P1 Arginine Mimetics. **Journal of Medicinal Chemistry**, 11 fev. 2010. v. 53, n. 3, p. 1067–1075.
- BERMAN, H. M. *et al.* The Protein Data Bank. **Nucleic Acids Research**, 1 jan. 2000. v. 28, n. 1, p. 235–242. . Acesso em: 24 nov. 2022.
- BEST, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 Dihedral Angles. **Journal of Chemical Theory and Computation**, 11 set. 2012. v. 8, n. 9, p. 3257–3273. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/ct300400x>>. Acesso em: 24 nov. 2022.
- BIENERT, S. *et al.* The SWISS-MODEL Repository-new features and functionality. **Nucleic Acids Research**, 1 jan. 2017. v. 45, n. D1, p. D313–D319. . Acesso em: 24 nov. 2022.
- BLASZCZYK, M. *et al.* Protein-peptide docking using CABS-dock and contact information. **Briefings in Bioinformatics**, 1 nov. 2019. v. 20, n. 6, p. 2299–2305. . Acesso em: 24 nov. 2022.
- BURLEY, S. K. *et al.* RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. **Nucleic Acids Research**, 2021. v. 49, n. 1, p. D437–D451. . Acesso em: 24 nov. 2022.

CHERNECKY CC; BERGER BJ. **Laboratory Tests and Diagnostic Procedures**. 6th. ed. [S.l.]: [s.n.], 2013.

COVID-19 Data Explorer - Our World in Data. [s.d.]. Disponível em: <<https://ourworldindata.org/explorers/coronavirus-data-explorer>>. Acesso em: 24 nov. 2022.

DAHMS, S. O. *et al.* X-ray structures of human furin in complex with competitive inhibitors. **ACS chemical biology**, 16 maio. 2014a. v. 9, n. 5, p. 1113–1118. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/24666235/>>. Acesso em: 24 nov. 2022.

DEMCHUK, E.; WADE, R. C. Improving the Continuum Dielectric Approach to Calculating pKas of Ionizable Groups in Proteins. **Journal of Physical Chemistry**, 24 out. 1996. v. 100, n. 43, p. 17373–17387. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/jp960111d>>. Acesso em: 24 nov. 2022.

DONALD, J. E.; KULP, D. W.; DEGRADO, W. F. Salt bridges: geometrically specific, designable interactions. **Proteins**, mar. 2011. v. 79, n. 3, p. 898–915. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/21287621/>>. Acesso em: 25 nov. 2022.

DUCKERT, P.; BRUNAK, S.; BLOM, N. Prediction of proprotein convertase cleavage sites. **Protein engineering, design & selection : PEDS**, jan. 2004. v. 17, n. 1, p. 107–112. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/14985543/>>. Acesso em: 24 nov. 2022.

ESSALMANI, R. *et al.* Distinctive Roles of Furin and TMPRSS2 in SARS-CoV-2 Infectivity. **Journal of Virology**, 27 abr. 2022. v. 96, n. 8. Disponível em: <<https://journals.asm.org/doi/10.1128/jvi.00128-22>>. Acesso em: 24 nov. 2022.

EWALD, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. **Annalen der Physik**, 1 jan. 1921. v. 369, n. 3, p. 253–287. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1002/andp.19213690304>>. Acesso em: 24 nov. 2022.

FELLER, S. E. *et al.* Constant pressure molecular dynamics simulation: The Langevin piston method. **The Journal of Chemical Physics**, 4 jun. 1998. v. 103, n. 11, p. 4613. Disponível em: <<https://aip.scitation.org/doi/abs/10.1063/1.470648>>. Acesso em: 24 nov. 2022.

FRANÇA, E. B. *et al.* Measuring misclassification of Covid-19 as garbage codes: Results of investigating 1,365 deaths and implications for vital statistics in Brazil. **PLOS Global Public Health**, 5 maio. 2022. v. 2, n. 5, p. e0000199. Disponível em: <<https://journals.plos.org/globalpublichealth/article?id=10.1371/journal.pgph.0000199>>. Acesso em: 1º fev. 2023.

GARRY, R. F. SARS-CoV-2 furin cleavage site was not engineered. **Proceedings of the National Academy of Sciences of the United States of America**, 4 out. 2022. v. 119, n. 40, p. e2211107119. Disponível em: <<https://www.pnas.org/doi/abs/10.1073/pnas.2211107119>>. Acesso em: 28 nov. 2022.

GORDON, J. C. *et al.* H++: a server for estimating pKas and adding missing hydrogens to macromolecules. **Nucleic acids research**, jul. 2005. v. 33, n. Web

Server issue. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/15980491/>>. Acesso em: 24 nov. 2022.

H++ (web-based computational prediction of protonation states and pK of ionizable groups in macromolecules). [s.d.]. Disponível em: <<http://newbiophysics.cs.vt.edu/H++/>>. Acesso em: 24 nov. 2022.

HADDOCK Web Server. [s.d.]. Disponível em: <<https://wenmr.science.uu.nl/haddock2.4/>>. Acesso em: 24 nov. 2022.

HANSSON, T.; MARELIUS, J.; ÅQVIST, J. Ligand binding affinity prediction by linear interaction energy methods. **Journal of computer-aided molecular design**, 1998. v. 12, n. 1, p. 27–35. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/9570087/>>. Acesso em: 25 nov. 2022.

HEDSTROM, L. Serine protease mechanism and specificity. **Chemical Reviews**, dez. 2002. v. 102, n. 12, p. 4501–4523. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/cr000033x>>. Acesso em: 29 nov. 2022.

HENDSCH, Z. S.; TIDOR, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. **Protein science : a publication of the Protein Society**, 1994. v. 3, n. 2, p. 211–226. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/8003958/>>. Acesso em: 25 nov. 2022.

HOFFMANN, M.; KLEINE-WEBER, H.; PÖHLMANN, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. **Molecular Cell**, 5 maio. 2020. v. 78, n. 4, p. 779. Disponível em: <<https://pmc/articles/PMC7194065/>>. Acesso em: 28 nov. 2022.

Home - Prediction Center. [s.d.]. Disponível em: <<https://www.predictioncenter.org/index.cgi>>. Acesso em: 25 nov. 2022.

HONORATO, R. V. *et al.* Structural Biology in the Clouds: The WeNMR-EOSC Ecosystem. **Frontiers in Molecular Biosciences**, 28 jul. 2021. v. 8. . Acesso em: 24 nov. 2022.

HOPKINS, E.; SANVICTORES, T.; SHARMA, S. Physiology, Acid Base Balance. **Urolithiasis**, 12 set. 2022. p. 19–22. Disponível em: <<https://www.ncbi.nlm.nih.gov/books/NBK507807/>>. Acesso em: 24 nov. 2022.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. **Journal of Molecular Graphics**, 1996. v. 14, n. 1, p. 33–38. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/8744570/>>. Acesso em: 24 nov. 2022.

I-TASSER server for protein structure and function prediction. [s.d.]. Disponível em: <<https://zhanggroup.org/I-TASSER/>>. Acesso em: 24 nov. 2022.

JACKSON, C. B. *et al.* Mechanisms of SARS-CoV-2 entry into cells. **Nature Reviews Molecular Cell Biology** 2021 23:1, 5 out. 2021. v. 23, n. 1, p. 3–20. Disponível em: <<https://www.nature.com/articles/s41580-021-00418-x>>. Acesso em: 28 nov. 2022.

JAIMES, J. A.; MILLET, J. K.; WHITTAKER, G. R. Proteolytic Cleavage of the SARS-CoV-2 Spike Protein and the Role of the Novel S1/S2 Site. **iScience**, 6 jun. 2020. v. 23, n. 6, p. 101212. Disponível em: <<https://pmc/articles/PMC7255728/>>. Acesso em: 12 dez. 2022.

JAIN, T.; JAYARAM, B. An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes. **FEBS letters**, 5 dez. 2005. v. 579, n. 29, p. 6659–6666. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/16307743/>>. Acesso em: 25 nov. 2022.

JAWAID, M. Z. *et al.* Computational study of the furin cleavage domain of SARS-CoV-2: delta binds strongest of extant variants. **bioRxiv**, 5 jan. 2022. p. 2022.01.04.475011. Disponível em: <<https://www.biorxiv.org/content/10.1101/2022.01.04.475011v1>>. Acesso em: 25 nov. 2022.

JUMPER, J. *et al.* Highly accurate protein structure prediction with AlphaFold. **Nature** **2021 596:7873**, 15 jul. 2021. v. 596, n. 7873, p. 583–589. Disponível em: <<https://www.nature.com/articles/s41586-021-03819-2>>. Acesso em: 24 nov. 2022.

KRAUT, J. Serine Proteases: Structure and Mechanism of Catalysis. **Annual Review of Biochemistry**, jun. 1977. v. 46, n. 1, p. 331–358.

KUFAREVA, I.; ABAGYAN, R. Methods of protein structure comparison. **Methods in molecular biology (Clifton, N.J.)**, 2012. v. 857, p. 231–257. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/22323224/>>. Acesso em: 24 nov. 2022.

KUMAR, S.; NUSSINOV, R. Salt bridge stability in monomeric proteins. **Journal of Molecular Biology**, 12 nov. 1999. v. 293, n. 5, p. 1241–1255. . Acesso em: 25 nov. 2022.

KUMAR, S.; NUSSINOV, R. Close-range electrostatic interactions in proteins. **Chembiochem : a European journal of chemical biology**, 2 jul. 2002. v. 3, n. 7, p. 604–17.

KURCINSKI, M. *et al.* CABS-dock standalone: a toolbox for flexible protein–peptide docking. **Bioinformatics**, 15 out. 2019. v. 35, n. 20, p. 4170–4172. Disponível em: <<https://academic.oup.com/bioinformatics/article/35/20/4170/5378705>>. Acesso em: 24 nov. 2022.

KURCINSKI, M. *et al.* Flexible docking of peptides to proteins using CABS-dock. **Protein Science**, 1 jan. 2020. v. 29, n. 1, p. 211–222. . Acesso em: 24 nov. 2022.

LEÓN, T. M. *et al.* COVID-19 Cases and Hospitalizations by COVID-19 Vaccination Status and Previous COVID-19 Diagnosis — California and New York, May–November 2021. **MMWR. Morbidity and Mortality Weekly Report**, 2022. v. 71, n. 4, p. 125–131. Disponível em: <<https://www.cdc.gov/mmwr/volumes/71/wr/mm7104e1.htm>>. Acesso em: 24 nov. 2022.

LI, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. **Science (New York, N.Y.)**, 28 out. 2005. v. 310, n. 5748, p. 676–679. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/16195424/>>. Acesso em: 27 nov. 2022.

MACKERELL, A. D. *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. **The journal of physical chemistry. B**, 30 abr. 1998. v. 102, n. 18, p. 3586–3616. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/24889800/>>. Acesso em: 8 dez. 2022.

MULLER, I. *et al.* SARS-CoV-2-related atypical thyroiditis. **The Lancet. Diabetes & Endocrinology**, 1 set. 2020. v. 8, n. 9, p. 739. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/3392564/>>. Acesso em: 2 dez. 2022.

NAMD Energy Plugin, Version 1.4. [s.d.]. Disponível em: <<https://www.ks.uiuc.edu/Research/vmd/plugins/namdenergy/>>. Acesso em: 24 nov. 2022.

ORSINE, L. humanExpProfiles. jun. 2020. Disponível em: <http://bioinfo.icb.ufmg.br:3838/humanExpProfiles_v2.1/>. Acesso em: 2 dez. 2022.

PAPA, G. *et al.* Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. **PLOS Pathogens**, 25 jan. 2021. v. 17, n. 1, p. e1009246.

PEACOCK, T. P. *et al.* The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. **Nature Microbiology** 2021 6:7, 27 abr. 2021. v. 6, n. 7, p. 899–909. Disponível em: <<https://www.nature.com/articles/s41564-021-00908-w>>. Acesso em: 24 nov. 2022.

PEREIRA, J. *et al.* High-accuracy protein structure prediction in CASP14. **Proteins: Structure, Function, and Bioinformatics**, 1 dez. 2021. v. 89, n. 12, p. 1687–1699. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1002/prot.26171>>. Acesso em: 25 nov. 2022.

PHILLIPS, J. C. *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. **The Journal of Chemical Physics**, 30 jul. 2020. v. 153, n. 4, p. 044130. Disponível em: <<https://aip.scitation.org/doi/abs/10.1063/5.0014475>>. Acesso em: 24 nov. 2022.

PIPER, D. E. *et al.* The Crystal Structure of PCSK9: A Regulator of Plasma LDL-Cholesterol. **Structure**, 16 maio. 2007. v. 15, n. 5, p. 545–552. . Acesso em: 25 nov. 2022.

ProP - 1.0 - Services - DTU Health Tech. [s.d.]. Disponível em: <<https://services.healthtech.dtu.dk/service.php?ProP-1.0>>. Acesso em: 24 nov. 2022.

PYLAEVA, S.; BREHM, M.; SEBASTIANI, D. Salt Bridge in Aqueous Solution: Strong Structural Motifs but Weak Enthalpic Effect. **Scientific Reports** 2018 8:1, 11 set. 2018. v. 8, n. 1, p. 1–7. Disponível em: <<https://www.nature.com/articles/s41598-018-31935-z>>. Acesso em: 25 nov. 2022.

RADISKY, E. S. *et al.* Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. **Proceedings of the National Academy of Sciences of the United States of America**, 2 maio. 2006. v. 103, n. 18, p. 6835–6840. Disponível em: <<https://www.pnas.org/doi/abs/10.1073/pnas.0601910103>>. Acesso em: 29 nov. 2022.

RCSB PDB: Homepage. [s.d.]. Disponível em: <<https://www.rcsb.org/>>. Acesso em: 24 nov. 2022.

REMACLE, A. G. *et al.* Substrate cleavage analysis of furin and related proprotein convertases. A comparative study. **The Journal of biological chemistry**, 25 jul.

2008. v. 283, n. 30, p. 20897–20906. Disponível em:

<<https://pubmed.ncbi.nlm.nih.gov/18505722/>>. Acesso em: 24 nov. 2022.

SCHECHTER, I.; BERGER, A. On the size of the active site in proteases. I. Papain.

Biochemical and biophysical research communications, 20 abr. 1967. v. 27, n. 2, p. 157–162. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/6035483/>>. Acesso em: 30 nov. 2022.

SEIDAH, N. G.; CHRÉTIEN, M.; DAY, R. The family of subtilisin/kexin like pro-protein and pro-hormone convertases: divergent or shared functions. **Biochimie**, 1994. v. 76, n. 3–4, p. 197–209. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/7819324/>>.

Acesso em: 24 nov. 2022.

SEIDAH, Nabil G. The proprotein convertases, 20 years later. **Methods in molecular biology (Clifton, N.J.)**, 2011. v. 768, p. 23–57. Disponível em:

<<https://pubmed.ncbi.nlm.nih.gov/21805237/>>. Acesso em: 28 nov. 2022.

SEIDAH, Nabil G. Proprotein Convertase 2. **Handbook of Proteolytic Enzymes**, 2013. v. 3, p. 3290–3295. . Acesso em: 11 dez. 2022.

SIMMONS, G. *et al.* Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. **Proceedings of the National Academy of Sciences of the United States of America**, 16 ago. 2005. v. 102, n. 33, p. 11876–11881.

Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/16081529/>>. Acesso em: 28 nov. 2022.

SIMPKIN, A. J. *et al.* Evaluation of model refinement in CASP14. **Proteins**, 1 dez. 2021. v. 89, n. 12, p. 1852–1869. Disponível em:

<<https://pubmed.ncbi.nlm.nih.gov/34288138/>>. Acesso em: 25 nov. 2022.

SINDELAR, C. V.; HENDSCH, Z. S.; TIDOR, B. Effects of salt bridges on protein structure and design. **Protein science : a publication of the Protein Society**, 1998. v. 7, n. 9, p. 1898–1914. Disponível em:

<<https://pubmed.ncbi.nlm.nih.gov/9761471/>>. Acesso em: 25 nov. 2022.

SWISS-MODEL. [s.d.]. Disponível em: <<https://swissmodel.expasy.org/>>. Acesso em: 24 nov. 2022.

THOMAS, G. Furin at the cutting edge: from protein traffic to embryogenesis and disease. **Nature reviews. Molecular cell biology**, 1 out. 2002. v. 3, n. 10, p. 753–766. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/12360192/>>. Acesso em: 28 nov. 2022.

TIAN, S. A 20 Residues Motif Delineates the Furin Cleavage Site and its Physical Properties May Influence Viral Fusion. **<https://doi.org/10.4137/BCI.S2049>**, 8 abr. 2009. v. 2, p. BCI.S2049. Disponível em:

<<https://journals.sagepub.com/doi/full/10.4137/BCI.S2049>>. Acesso em: 24 nov. 2022.

TIAN, S.; JIANHUA, W. Comparative study of the binding pockets of mammalian proprotein convertases and its implications for the design of specific small molecule inhibitors. **International Journal of Biological Sciences**, 2010. v. 6, n. 1, p. 89.

Disponível em: <pmc/articles/PMC2820236/>. Acesso em: 28 nov. 2022.

Vacinas — Anvisa. [s.d.]. Disponível em: <<https://www.gov.br/anvisa/pt-br/assuntos/paf/coronavirus/vacinas>>. Acesso em: 24 nov. 2022.

VAIDYANATHAN, G. Coronavirus variants are spreading in India - what scientists know so far. **Nature**, 1 maio. 2021. v. 593, n. 7859, p. 321–322. . Acesso em: 24 nov. 2022.

VANGONE, A. *et al.* Sense and simplicity in HADDOCK scoring: Lessons from CASP-CAPRI round 1. **Proteins**, 1 mar. 2017. v. 85, n. 3, p. 417. Disponível em: <[pmc/articles/PMC5324763/](https://pubmed.ncbi.nlm.nih.gov/2744763/)>. Acesso em: 25 nov. 2022.

VARADI, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. **Nucleic Acids Research**, 7 jan. 2022. v. 50, n. D1, p. D439–D444. Disponível em: <<https://academic.oup.com/nar/article/50/D1/D439/6430488>>. Acesso em: 24 nov. 2022.

VILLOUTREIX, B. O. *et al.* Targeting furin activity through in silico and in vitro drug repurposing strategy for SARS-CoV-2 spike glycoprotein cleavage repression. 2020. Disponível em: <<https://doi.org/10.21203/rs.3.rs-25856/v1>>. Acesso em: 28 nov. 2022.

VMD measure documentation. [s.d.]. Disponível em: <<http://www-s.ks.uiuc.edu/Research/vmd/vmd-1.9.1/ug/node136.html>>. Acesso em: 24 nov. 2022.

VOET, D.; VOET, J. **Biochemistry**. 4th. ed. [S.l.]: Kaye Pace, 2010.

WAHLGREN, W. Y. *et al.* The Catalytic Aspartate Is Protonated in the Michaelis Complex Formed between Trypsin and an in Vitro Evolved Substrate-like Inhibitor: A REFINED MECHANISM OF SERINE PROTEASE ACTION*. **The Journal of Biological Chemistry**, 2 fev. 2011. v. 286, n. 5, p. 3587. Disponível em: <[pmc/articles/PMC3030363/](https://pubmed.ncbi.nlm.nih.gov/2030363/)>. Acesso em: 29 nov. 2022.

WALLS, A. C. *et al.* Structure, function and antigenicity of the SARS-CoV-2 spike glycoprotein. **bioRxiv**, 20 fev. 2020. p. 2020.02.19.956581. Disponível em: <<https://www.biorxiv.org/content/10.1101/2020.02.19.956581v1>>. Acesso em: 28 nov. 2022.

WATERHOUSE, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. **Nucleic Acids Research**, 2 jul. 2018. v. 46, n. W1, p. W296–W303. . Acesso em: 24 nov. 2022.

WHITTAKER, C. *et al.* Under-reporting of deaths limits our understanding of true burden of covid-19. **BMJ**, 12 out. 2021. v. 375. Disponível em: <<https://www.bmj.com/content/375/bmj.n2239>>. Acesso em: 1º fev. 2023.

WHITTAKER, G. R. SARS-CoV-2 spike and its adaptable furin cleavage site. **The Lancet. Microbe**, 1 out. 2021. v. 2, n. 10, p. e488. Disponível em: <[pmc/articles/PMC8346238/](https://pubmed.ncbi.nlm.nih.gov/346238/)>. Acesso em: 28 nov. 2022.

WISHART, D. S. *et al.* DrugBank 5.0: A major update to the DrugBank database for 2018. **Nucleic Acids Research**, 1 jan. 2018. v. 46, n. D1, p. D1074–D1082. . Acesso em: 2 dez. 2022.

WONG, G. *et al.* Zoonotic origins of human coronavirus 2019 (HCoV-19 / SARS-CoV-2): why is this work important? **Zoological Research**, 5 maio. 2020. v. 41, n. 3, p. 213. Disponível em: <<http://pmc/articles/PMC7231470/>>. Acesso em: 27 nov. 2022.

WORLD HEALTH ORGANIZATION. Surveillance case definitions for human infection with novel coronavirus (nCoV). **Interim guidance**, 11 jan. 2020a. Disponível em: <<http://apps.who.int/iris/handle/10665/330376>>. Acesso em: 27 nov. 2022.

WORLD HEALTH ORGANIZATION. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. 11 mar. 2020b. Disponível em: <<https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>>. Acesso em: 27 nov. 2022.

WROBEL, A. G. *et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. **Nature Structural & Molecular Biology** 2020 27:8, 9 jul. 2020. v. 27, n. 8, p. 763–767. Disponível em: <<https://www.nature.com/articles/s41594-020-0468-7>>. Acesso em: 28 nov. 2022.

WU, F. *et al.* A new coronavirus associated with human respiratory disease in China. **Nature** 2020 579:7798, 3 fev. 2020. v. 579, n. 7798, p. 265–269. Disponível em: <<https://www.nature.com/articles/s41586-020-2008-3>>. Acesso em: 28 nov. 2022.

YANG, J. *et al.* The I-TASSER Suite: protein structure and function prediction. **Nature Methods** 2015 12:1, 30 dez. 2014. v. 12, n. 1, p. 7–8. Disponível em: <<https://www.nature.com/articles/nmeth.3213>>. Acesso em: 24 nov. 2022.

YANG, L. *et al.* Novel SARS-like betacoronaviruses in bats, China, 2011. **Emerging infectious diseases**, jun. 2013. v. 19, n. 6, p. 989–991. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/23739658/>>. Acesso em: 27 nov. 2022.

ZHENG, W. *et al.* Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. **Cell reports methods**, 26 jul. 2021. v. 1, n. 3. Disponível em: <<https://pubmed.ncbi.nlm.nih.gov/34355210/>>. Acesso em: 24 nov. 2022.

ZHOU, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. **Nature** 2020 579:7798, 3 fev. 2020. v. 579, n. 7798, p. 270–273. Disponível em: <<https://www.nature.com/articles/s41586-020-2012-7>>. Acesso em: 27 nov. 2022.

ZHU, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. **New England Journal of Medicine**, 20 fev. 2020. v. 382, n. 8, p. 727–733. Disponível em: <<https://www.nejm.org/doi/10.1056/NEJMoa2001017>>. Acesso em: 27 nov. 2022.

ZUNDERT, G. C. P. VAN *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. **Journal of Molecular Biology**, 22 fev. 2016. v. 428, n. 4, p. 720–725. . Acesso em: 24 nov. 2022.

APÊNDICE A – PARÂMETROS DE CONFIGURAÇÃO NAMD

```
#####
## ADJUSTABLE PARAMETERS ##
#####
structure      <structure.psf>
coordinates     <coordinates.pdb>

set temperature 310
set outputname  <outputname>

firsttimestep   0

#####
## SIMULATION PARAMETERS ##
#####
# Input
paraTypeCharmm  on
parameters      par_all36_prot.prm
parameters      par_all36_na.prm
parameters      toppar_water_ions.str
temperature     $temperature

# Force-Field Parameters
exclude         scaled1-4
1-4scaling      1.0
cutoff          12.0
switching       on
switchdist      10.0
pairlistdist    14.0

# Integrator Parameters
timestep        2.0  ;# 2fs/step
rigidBonds      all  ;# needed for 2fs steps
nonbondedFreq   1
fullElectFrequency 2
stepspercycle   10

# Constant Temperature Control
langevin        on  ;# do langevin dynamics
langevinDamping 1   ;# damping coefficient (gamma) of 1/ps
```

```

langevinTemp      $temperature
langevinHydrogen  off  ;# don't couple langevin bath to hydrogens

# Constant Pressure Control (variable volume)
useGroupPressure  yes ;# needed for rigidBonds
useFlexibleCell   no
useConstantArea   no

langevinPiston    on
langevinPistonTarget 1.01325 ;# pressure in bar -> 1 atm
langevinPistonPeriod 100. ;# oscillation period around 100 fs
langevinPistonDecay  50. ;# oscillation decay time of 50 fs
langevinPistonTemp  $temperature ;# coupled to heat bath

# Electrostatic Force Evaluation
PME               yes  #PME (for full-system periodic electrostatics)
PMEGridSpacing    1.0  # let NAMD determine grid

# Periodic Boundary conditions
wrapAll           on
cellBasisVector1  71    0.    0.  ;# cell dimensions
cellBasisVector2  0.    58    0.
cellBasisVector3  0.    0     51
cellOrigin        -1    2     -1  ;# the *center* of the cell

# Output
outputName        $outputname
restartfreq        500  ;# 500steps = every 1ps
dcdfreq           5000
outputEnergies     100
outputPressure     100

#####
## EXECUTION SCRIPT                                     ##
#####

# Minimization
minimize           1000
reinitvels         $temperature

run 151000000

```

APÊNDICE B – RESULTADOS DAS ANÁLISES

Os resultados das análises de todas as simulações foram disponibilizados através do endereço web:

<http://biodados.icb.ufmg.br/coronavirus/>.

APÊNDICE C – PONTE SALINAS AO LONGO DA SIMULAÇÃO COM PCSK7

O vídeo que mostra o mapa de pontes salina entre a enzima PCSK7 e o peptídeo ao longo da simulação a partir da docagem com Haddock encontra-se disponível no endereço web:

http://biodados.icb.ufmg.br/coronavirus/data_media/pcsk7_13/pcsk7_ap_13_c7_AHNS/contacts/pcsk7_ap_13_c7_AHNS_A-B_sbridges_contacts_map_steps.mp4.

APÊNDICE D – DOCUMENTAÇÃO PAMDA

O *download*, guia de instalação completo e documentação do programa Pamda estão disponíveis no endereço web:

<http://biodados.icb.ufmg.br/pamda/>.

APÊNDICE E – CÓDIGO FONTE PAMDA

O código do programa Pamda se encontra disponível no repositório do GitHub:

<https://github.com/Artotim/pamda>.