



**UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE BIOCÊNCIAS  
BACHARELADO EM BIOMEDICINA**

FRANCISCO AGENOR DE OLIVEIRA NETO

***CLEAVE ME OR LEAVE ME:***

CARACTERIZAÇÃO DA VARIAÇÃO DAS SEQUÊNCIAS DOS SÍTIOS  
DE CLIVAGEM HUMANOS DA PROTEASE NSP5 DE SARS-COV-2

Recife  
2024

FRANCISCO AGENOR DE OLIVEIRA NETO

***CLEAVE ME OR LEAVE ME:***

CARACTERIZAÇÃO DA VARIAÇÃO DAS SEQUÊNCIAS DOS SÍTIOS  
DE CLIVAGEM HUMANOS DA PROTEASE NSP5 DE SARS-COV-2

Trabalho de Conclusão de Curso  
apresentado ao Curso de Graduação em  
Biomedicina da Universidade Federal de  
Pernambuco, como pré-requisito à  
obtenção do título de Bacharel em  
Biomedicina.

Orientador: Profa. Dra. Danyelly Brunaska  
Gondim Martins

Coorientador: Dr. Carlos Henrique  
Madeiros Castelletti

Recife  
2024

Ficha de identificação da obra elaborada pelo autor,  
através do programa de geração automática do SIB/UFPE

Oliveira Neto, Francisco Agenor de.

Cleave me or leave me: caracterização da variação das sequências dos sítios de clivagem humanos da protease nsp5 de SARS-CoV-2 / Francisco Agenor de Oliveira Neto. - Recife, 2024.

56 p. : il., tab.

Orientador(a): Danyelly Bruneka Gondim Martins

Coorientador(a): Carlos Henrique Madeiros Castelletti

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Biociências, Biomedicina, 2024.

1. Virologia. 2. Genética. 3. Bioinformática. 4. Evolução. 5. Interação patógeno-hospedeiro. I. Martins, Danyelly Bruneka Gondim. (Orientação). II. Castelletti, Carlos Henrique Madeiros. (Coorientação). IV. Título.

570 CDD (22.ed.)

FRANCISCO AGENOR DE OLIVEIRA NETO

***CLEAVE ME OR LEAVE ME:***

CARACTERIZAÇÃO DA VARIAÇÃO DAS SEQUÊNCIAS DOS SÍTIOS  
DE CLIVAGEM HUMANOS DA PROTEASE NSP5 DE SARS-COV-2

Trabalho de Conclusão de Curso  
apresentado ao Curso de  
Graduação em Biomedicina da  
Universidade Federal de  
Pernambuco, como pré-requisito à  
obtenção do título de Bacharel em  
Biomedicina.

Aprovada em: \_\_\_/\_\_\_/\_\_\_

**BANCA EXAMINADORA**

---

Orientador: Profa. Dra. Danyelly Brunaska Gondim Martins  
Instituto Keizo Asami/Universidade Federal de Pernambuco

---

Prof. Dr. Lucas André Cavalcanti Brandão  
Universidade Federal de Pernambuco

---

Me. George de Vasconcelos Carvalho  
UCLA - University of California Los Angeles

Dedico este trabalho à minha mãe,  
Maria de Jesus (que me incentivou) e a  
todos futuros jovens cientistas  
brasileiros (quero ser um incentivo  
para vocês).

## AGRADECIMENTOS

Eu poderia iniciar este texto agradecendo a Deus primeiramente. Porém, como um primata que, com seu cérebro, busca e constrói significado, planeja e almeja, dúvida e acredita e, acima de tudo, tem fé, sei que em todos as pessoas que citarei abaixo, a presença e o agir de Deus estavam ali, construindo significado e dando sentido. E eu sou grato a Ele por isso.

Agradeço aos meus dois orientadores, Henrique Castelletti e Danyelly Brunaska, por terem me acompanhado, desde a seleção para a iniciação científica até o presente momento, com paciência e incentivo à curiosidade. A Henrique, eu agradeço por me acompanhar na pandemia, sempre pacientemente. À professora Brunaska, sou grato pelo brilho no olhar que tem pela Ciência e por repassar isso aos seus alunos.

Agradeço a CHESF/Fade pelo suporte financeiro na realização desse projeto.

Não poderia deixar de destacar a importância dos laços sociais para um homínido como eu na realização desse projeto e, para mais do que isso, para minha jornada como cientista. É por isso, que agradeço a minha mãe, Maria de Jesus (De Jé), pelo amor e incessantes incentivos para que eu não desistisse. Agradeço também ao meu pai, Francisco, pelos esforços feitos para me ajudar.

Não bastasse a expressão desse amor pelos meus pais, também fui agraciado com uma rede de mulheres que me deram grande suporte nesses anos: tia Fátima, tia Edeusa, tia Mila, tia Vaneide, tia Belinha, vovó Joana e vovó Zezinha obrigado por me ensinarem sobre alteridade e amor ao próximo.

Agradeço às famílias que não compartilham nenhum parentesco comigo, mas, com humanidade, fé e amor, me adotaram em Recife: Mirela, Silvia e seu Luís, ali no Derby, vocês cuidaram de mim de formas que nunca imaginei que poderia ser cuidado. Mais à frente, próximo à orla de Olinda, sou grato também à família de um grande amigo, Gabriel, que juntamente com Edson, Sandra e Sanmara, se fizeram lar para mim. Já no Engenho do Meio, próximo ao riacho do Cavouco, agradeço à família de Raquel pela hospitalidade. Finalmente, em Gaibu, no Cabo de Santo Agostinho, sou grato a Zeni, Jeovana e Gilmar, por terem sido a primeira casa a me receber quando cheguei em Pernambuco.

A Academia também me reservou motivos de gratidão: sou grato por

ter tido contato com docentes que me incentivaram a prática do exercício científico: Rafael, Paula Sandrin e Jaqueline, como professores da disciplina de Genética, foram responsáveis pelos meus primeiros êxtases contemplativos ao entender como a informação biológica se comporta nos seres vivos. Agradeço também a Tiago Salazar, que também foi uma grande inspiração para meu interesse pela Ciência.

Agradeço imensamente às minhas amigas de graduação: Bárbara e Sharla, que, de maneira muito especial, compartilharam comigo a mesma experiência da saída de suas casas e vinda para Recife, sendo parcerias singulares nesse processo. Aos “NONers”, grupo de amigos composto por Anndressa, Lucas, Zion, Marco, Almir, Eduardo e o responsável pelo nome, Rimaud, sou grato pelo companheirismo.

De forma muito inusitada, mas creio que providencial, do Recife e de vários locais do País, recebi vários amigos que me ajudaram a lembrar da minha dignidade e do valor das minhas especificidades. Aos amigos do ONG/Confessionário, minha grande gratidão pela vida de vocês: Estevão, Carlos Macena, Carlos Henrique, Vinícius, Natã, Fabrício, Kalebe e Moisés. Também sou grato pela vida de Ithalo e Casemiro que, apesar de terem chegado em minha jornada de outras formas, cumpriram o mesmo papel que citei anteriormente.

Também sou grato pelos meus irmãos da minha comunidade de fé, a Primeira Igreja Batista em Engenho do Meio. Mirelle, Micael, Andressa e Mateus, sou extremamente grato por serem canais de graça na minha vida, andando ao meu lado até a Eternidade.

Por último, agradeço a todos os que a minha memória não alcançou, mas que, em algum momento, fizeram parte da construção coletiva desse trabalho e da minha jornada. Ubuntu!

*“Now, here, you see, it takes all the running you  
can do, to keep in the same place”.*

*(From Through the Looking-Glass, by Lewis  
Carroll)*

## RESUMO

O vírus causador da COVID-19, SARS-CoV-2, assim como os outros vírus de RNA de fita simples de sentido positivo, possui um genoma compacto que codifica grandes poliproteínas, as quais são processadas por proteases virais a fim de gerarem proteínas maduras importantes para o ciclo viral. A protease nsp5 é responsável pela clivagem de 11 sítios de SARS-CoV-2, além de ser capaz também de clivar proteínas humanas. Nestes casos, a nsp5 pode servir como uma “arma molecular” contra o sistema imune do hospedeiro, mas também pode clivar proteínas do hospedeiro que servem como “isca”, ativando vias antivirais. Sendo assim, substituições nesses sítios de clivagem podem interferir nesse processo de clivagem e, conseqüentemente, impactar o conflito entre vírus e hospedeiro, como já foi visto para outros vírus que se utilizam desse mesmo mecanismo. Assim, o presente estudo objetivou caracterizar a variação das sequências humanas alvo de clivagem da nsp5. Para isso, inicialmente, foi realizada uma revisão de literatura extensa a fim de identificar os sítios de clivagem humanos da nsp5 já comprovados experimentalmente. Em seguida, com os dados dessa etapa inicial, foi construído um perfil de preferência físico-química, ou especificidade, da protease. Com essas informações, utilizando-se de *scripts* em Python, foram recuperadas variantes genéticas *missense* presentes nessas regiões a partir do maior banco de dados de variantes genéticas da atualidade, o gnomAD. Esses SNPs foram analisados quanto a sua frequência alélica e impacto funcional *in silico*. A variante mais frequente também foi analisada estruturalmente, por meio do software PyMol, quanto ao seu impacto. Dessa maneira, foram identificados 100 sítios na revisão de literatura e 242 variantes genéticas que, no geral, impactam negativamente a capacidade clivagem. Além disso, a mutação com maior frequência ocorre no gene *DDX58*, codificante do sensor de RNAs RIG-I e, *in silico*, parecer ser benéfica para a protease haja vista que, estruturalmente, foi visto uma mudança no perfil de hidrofobicidade e eletrostático. Portanto, este trabalho demonstrou que há variação nos sítios de clivagem humanos da nsp5 e que esta pode ter impactos no conflito entre vírus e hospedeiro, apontando essas interface de interação entre a protease viral e alvos humanos como regiões de interesse para busca de variação genética e conseqüente entendimento da genética da susceptibilidade e gravidade da COVID-19.

**Palavras-chave:** Protease. Variação genética. Evolução molecular. Virologia. Bioinformática

## ABSTRACT

The causative agent of COVID-19, SARS-CoV-2, like other positive-sense single-stranded RNA viruses, has a compact genome that encodes large polyproteins, which are processed by viral proteases to generate mature proteins that are important for the viral cycle. The nsp5 protease is responsible for the cleavage of 11 SARS-CoV-2 sites and is also capable of cleaving human proteins. In these cases, nsp5 can serve as a “molecular weapon” against the host’s immune system, but it can also cleave host proteins that serve as “tripwires”, activating antiviral pathways. Therefore, substitutions in these cleavage sites can interfere with this cleavage process and, consequently, impact the conflict between virus and host, as has already been seen for other viruses that use this same mechanism. Therefore, the present study aimed to characterize the variation existing in the human cleavage targets of nsp5. To this end, initially, an extensive literature review was carried out in order to identify the human nsp5 cleavage sites already experimentally proven. Then, with the data from this initial step, a physicochemical preference profile, or specificity profile, of the protease was constructed. With this information, using Python scripts, missense genetic variants present in these regions were retrieved from the largest database of genetic variants currently, gnomAD. These SNPs were analyzed for their allelic frequency and functional impact *in silico*. The most frequent variant was analyzed structurally in the software PyMol. In this way, 100 sites were identified in the literature review and 242 missense variants that, in general, negatively impact the cleavage capacity. Furthermore, the most frequent mutation occurs in the *DDX58* gene, which encodes the RIG-I RNA sensor and, *in silico*, it seems to be beneficial for the protease since structurally there was a change in the hydrophobicity and charges profiles. Therefore, this work demonstrated that there is variation in the human cleavage sites of nsp5 and that it can have impacts on the conflict between virus and host, pointing out these interaction interfaces between the viral protease and human targets as regions of interest for searching for genetic variation and consequent understanding of the genetics of susceptibility and severity of COVID-19.

**Key words:** Protease. Genetic variation. Molecular evolution. Virology. Bioinformatics.

## LISTA DE ILUSTRAÇÕES

|  |    |
|--|----|
| <b>Figura 1</b> – Tradução do RNA e processamento da poliproteína do HCV.....  | 13 |
| <b>Figura 2</b> – Representação gráfica da nomenclatura de Schechter e Berger.....   | 14 |
| <b>Figura 3</b> – Estrutura e funcionamento geral dos RLRs.....  | 17 |
| <b>Figura 4</b> – Conflitos evolucionários no contexto das proteases e suas proteínas-alvo.....  | 19 |
| <b>Figura 5</b> – Etapas do projeto de pesquisa.....   | 19 |
| <b>Figura 6</b> – Matriz de frequência absoluta.....   | 25 |
| <b>Figura 7</b> – Matriz de probabilidades.....  | 26 |
| <b>Figura 8</b> – Matriz de pontuação posição-específica.....  | 26 |
| <b>Figura 9</b> – <i>Boxplot</i> dos volumes dos grupo R laterais dos aminoácidos dos sítios de clivagem humanos da nsp5.....  | 38 |
| <b>Figura 10</b> – Perfil de preferência físico-química dos sítios da protease nsp5, a partir dos 111 sítios obtidos na revisão de literatura.....   | 39 |
| <b>Figura 11</b> – <i>Boxplot</i> dos valores de pLDDT dos aminoácidos dos sítios de clivagem humanos da nsp5.....   | 40 |
| <b>Figura 12</b> – Gráficos de densidade das distribuições das pontuações dos sítios de clivagem das sequências referência e dos sítios com as substituições provenientes das variantes..... | 41 |
| <b>Figura 13</b> – Visualização tridimensional do sítio de clivagem em RIG-I.....  | 43 |
| <b>Figura 14</b> – Visualização tridimensional do perfil de hidrofobicidade resíduo P4 com o resíduo de arginina e o de cisteína.....  | 44 |
| <b>Figura 15</b> – Visualização tridimensional do perfil eletrostático resíduo P4 com o resíduo de arginina e o de cisteína.....   | 44 |

## LISTA DE TABELAS

|  |    |
|--|----|
| <b>Tabela 1</b> – Resultados dos algoritmos de predição de estabilidade..... | 40 |
|--|----|

## LISTA DE ABREVIATURAS E SIGLAS

|          |  |
|----------|--|
| (+)ssRNA | RNA de fita simples de sentido positivo                      |
| 3CLpro   | Protease do tipo 3C  |
| ABraOM   | Arquivo Brasileiro Online de Mutações                        |
| ASA      | Superfície Acessível Absoluta                                |
| CARD8    | Domínio de recrutamento da caspase 8                         |
| CHIKV    | Vírus da Chikungunya   |
| COVID-19 | Doença por coronavírus 2019                                  |
| CTD      | Domínio C-terminal   |
| DENV     | Vírus da Dengue  |
| DNA      | Ácido desoxirribonucleico                                    |
| dsRNA    | Ácido ribonucleico de fita dupla                             |
| EV71     | Enterovírus 71   |
| FA       | Frequência alélica   |
| HCV      | Vírus da Hepatite C  |
| HIV-1    | Vírus da imunodeficiência humana 1                           |
| HRV      | Rinovírus humano   |
| IFN      | Interferon   |
| LGP2     | Laboratório de Genética e Fisiologia 2                       |
| MANE     | Anotações combinadas do NCBI e EBML-EBI                      |
| MAVS     | Proteína de sinalização antiviral mitocondrial               |
| MDA5     | Proteína 5 associado a diferenciação do melanoma             |
| MFN1     | Mitofusina 1   |
| MFN2     | Mitofusina 2   |
| mRNA     | Ácido ribonucleico mensageiro                                |
| NLRP1    | Proteína 1 contendo os domínio NACTH, LRR, FIIND, CARD E PYD |
| nsp3     | Proteína não-estrutural 3                                    |
| nsp5     | Proteína não-estrutural 5                                    |
| nsp5     | Proteínas não-estruturais                                    |
| OMS      | Organização Mundial da Saúde                                 |
| ORF      | Fase de leitura aberta                                       |
| RIG-I    | Gene I indizível pelo ácido retinóico                        |
| RLR      | Receptores semelhantes à RIG-I                               |

|            |  |
|------------|--|
| RNA        | Ácido ribonucleico                                 |
| RSA        | Superfície Acessível Relativa                      |
| SARS-CoV-2 | Coronavírus da síndrome respiratória aguda grave 2 |
| STING      | Estimulador dos genes do interferon                |
| YFV        | Vírus da Febre Amarela                             |
| ZIKV       | Vírus da Zika                                      |

## SUMÁRIO

|   |           |
|---|-----------|
| <b>1 INTRODUÇÃO</b> .....   | <b>15</b> |
| <b>2 REVISÃO BIBLIOGRÁFICA</b> .....                                      | <b>16</b> |
| 2.1 VÍRUS RNA DE FITA SIMPLES DE SENTIDO POSITIVO E PROTEASES VIRAIS..... | 16        |
| 2.1.1 SARS-COV-2 e a protease nsp5  |           |
| .....   | 18        |
| 2.2 RIG-I, UM IMPORTANTE SENSOR DE RNAs VIRAIS.....                       | 19        |
| 2.3 CONFLITOS EVOLUCIONÁRIOS ENTRE VÍRUS E HOSPEDEIRO.....                | 20        |
| 2.3.1 Proteases Virais e Conflitos  |           |
| Evolucionários.....   | 21        |
| <b>3 OBJETIVOS</b> .....  | <b>24</b> |
| 3.1 OBJETIVO GERAL .....  | 24        |
| 3.2 OBJETIVOS ESPECÍFICOS .....   | 24        |
| <b>4 METODOLOGIA</b> .....  | <b>25</b> |
| 4.1 MINERAÇÃO DE DADOS E SELEÇÃO DOS SÍTIOS DE CLIVAGEM .....             | 25        |
| 4.2 CARACTERIZAÇÃO FÍSICO-QUÍMICA DOS SÍTIOS DE CLIVAGEM .....            | 26        |
| 4.3 PESQUISA POR VARIANTES <i>MISSENSE</i> .....                          | 27        |
| 4.4 ANÁLISE E INTERPRETAÇÃO DAS VARIANTES .....                           | 28        |
| 4.5 ANÁLISE ESTRUTURAL.....   | 31        |
| <b>5 RESULTADOS</b> .....   | <b>32</b> |
| 5.1 COLETA DE DADOS .....   | 32        |
| 5.2 CARACTERIZAÇÃO FÍSICO-QUÍMICA DOS SÍTIOS DE CLIVAGEM .....            | 39        |
| 5.3 PESQUISA POR VARIANTES <i>MISSENSE</i> .....                          | 41        |
| 5.4 ANÁLISE E INTERPRETAÇÃO DAS VARIANTES .....                           | 43        |
| 5.5 ANÁLISE ESTRUTURAL.....   | 43        |
| <b>6 DISCUSSÃO</b> .....  | <b>46</b> |
| <b>7 CONCLUSÃO</b> .....  | <b>49</b> |
| <b>REFERÊNCIAS</b> .....  | <b>50</b> |

## 1 INTRODUÇÃO

Os vírus de RNA de fita simples de sentido positivo (*positive-sense single stranded RNA* ou ((+)ssRNA, em inglês) incluem patógenos de grande importância para a saúde pública, como o poliovírus, vírus da Zika (ZIKV), o vírus chikungunya (CHIKV) e o SARS-CoV-2. Esses patógenos se utilizam de um mecanismo comum em seus ciclos virais: traduzir uma ou mais grandes poliproteínas que são subsequentemente clivadas por proteases humanas ou virais (Modrow *et al.*, 2013). Portanto, esse mecanismo de clivagem se configura como um passo importantíssimo para a replicação e transmissão viral.

No caso do SARS-CoV-2, a protease nsp5 é responsável pela clivagem do maior número de sítios, totalizando um total de 11. Tal é a importância dela no ciclo viral, que alguns fármacos já aprovados como terapia para a COVID-19 tem como alvo justamente a inibição dela (Najjar-Debbiny, 2022). Além disso, inibidores de protease já são também utilizados na terapia antirretroviral contra o HIV (Wang; Lv; Chu, 2015).

Entretanto, hoje é sabido que a nsp5 não cliva somente proteínas virais, mas também proteínas do hospedeiro humano, manipulando vias bioquímicas de forma a favorecer o vírus (Liu *et al.*, 2021; Zhang *et al.*, 2022). Esses cenários colocam o vírus e o hospedeiro num ambiente de “conflito evolucionário” dado que mutações nos sítios de clivagem podem impactar a eficiência de clivagem da protease, ajudando o hospedeiro a evadir-se da “arma” viral. Similarmente, o vírus também pode sofrer mutações na sequência da protease de forma a mudar sua especificidade e reconhecer novos sítios.

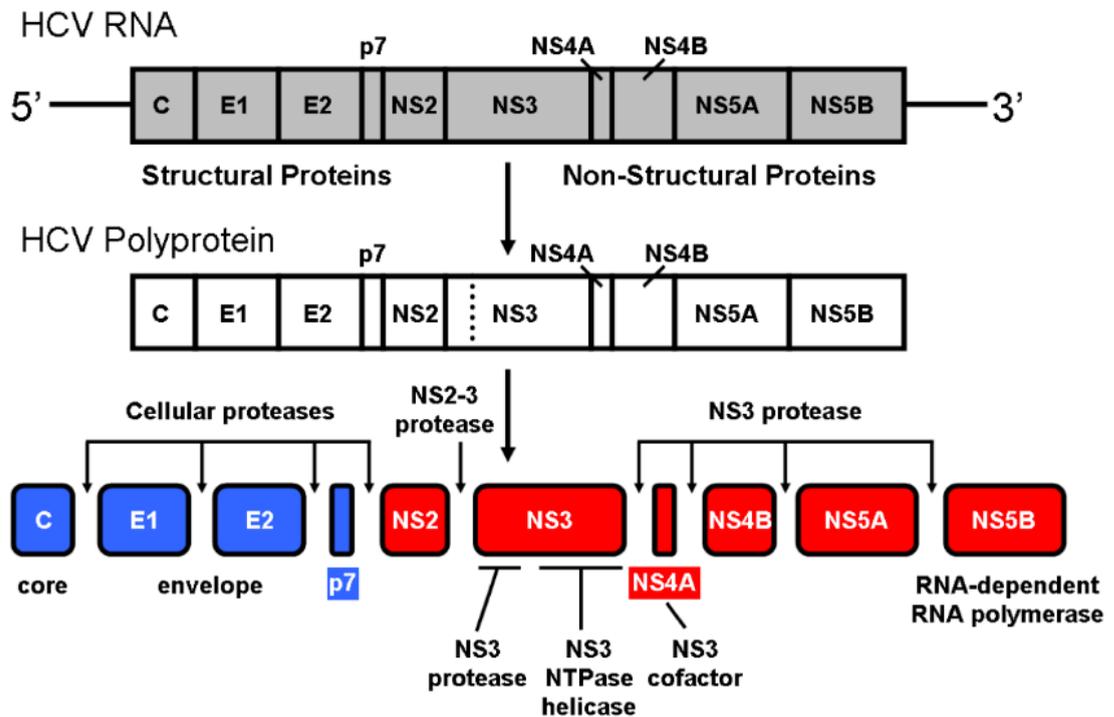
Sendo assim, estudar a variação genética nessas interfaces de interação vírus-hospedeiro é de grande importância para o melhor entendimento da interação entre os dois e de como variações genéticas podem ser a base de diferentes respostas frente a infecção pelo SARS-CoV-2. Portanto, este trabalho *in silico* buscou identificar e caracterizar as variações presentes nas sequências humanas alvo de clivagem da protease nsp5, buscando também mensurar o impacto dessas variações sobre a capacidade de clivagem da protease.

## 2 REVISÃO BIBLIOGRÁFICA

### 2.1 VÍRUS DE RNA DE FITA SIMPLES DE SENTIDO POSITIVO E PROTEASES VIRAIS

Os vírus (+)ssRNA são o maior grupo de vírus de RNA. Uma propriedade comum entre eles é a de utilizar seu genoma como um RNA mensageiro (*messenger RNA* ou mRNA, em inglês), a partir do qual sintetizam uma ou várias poliproteínas que são subsequentemente clivadas por proteases celulares ou virais (Modrow *et al.*, 2013), gerando proteínas individuais e funcionais como pode ser visto na Figura 1. Dessa forma, esse processo é crítico para uma replicação e propagação viral de sucesso, o que faz com que as proteases virais sejam um alvo atrativo para o desenvolvimento de fármacos antivirais (Lee *et al.*, 2020)

**Figura 1** - Tradução do RNA e processamento da poliproteína do HCV



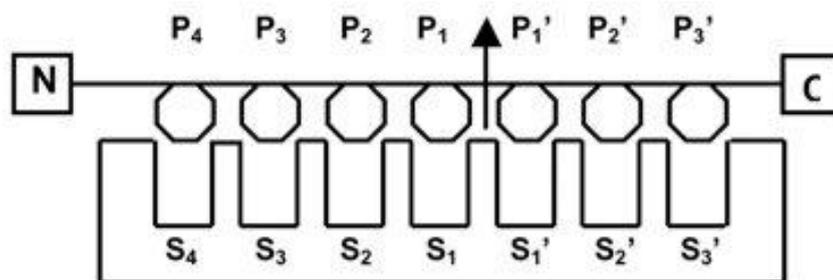
Fonte: Vega *et al.*, 2013

Esse mecanismo de clivagem já é conhecido em vários vírus de interesse para a espécie humana. Nos flavivírus, o vírus da hepatite C (*Hepatitis C virus* ou HCV, no inglês), o vírus da febre amarela (*yellow fever virus* ou YFV, em inglês), o

vírus da dengue (DENV) e o vírus da Zika (ZIKV) possuem uma serino-protease, NS3, responsável pela clivagem de suas poliproteínas (Modrow *et al.*, 2013). Na família *Togaviridae*, o vírus da Chikungunya (CHIKV) utiliza uma cisteína protease (Narwal *et al.*, 2018), assim como na família *Coronaviridae*, os vírus SARS-CoV e SARS-CoV-2 também se utilizam desse mesmo mecanismo.

Cabe salientar também que as proteases possuem especificidades individuais quanto a quais sequências de aminoácidos são melhor reconhecidas e clivadas pelos seus sítios de ligação. A fim de identificar de forma mais fácil os aminoácidos de preferência, foi desenvolvida uma nomenclatura por Schechter e Berger (Schechter e Berger, 1967), representada graficamente na Figura 2. Nela, o resíduo na direção N-terminal imediatamente ao lado da ligação peptídica que será quebrada é denominado P1 e assim progressivamente. Já o primeiro resíduo na direção C-terminal após a ligação que será clivada é denominado P1' e assim sucessivamente. Posto isso, é sabido que as proteases do ZIKV e DENV reconhecem sequências com resíduos carregados positivamente, como arginina e lisina, nas posições P1 e P2 (Li e Kang, 2021). Já para o HCV, essa preferência inclui resíduos acídicos (ácido glutâmico e ácido aspártico) na posição P6, resíduos polares (cisteína ou treonina) na posição P1 e resíduos com cadeias laterais pequenas (serina ou alanina) na posição P1' (Grakoui *et al.*, 1993).

**Figura 2** - Representação gráfica da nomenclatura de Schechter e Berger.



**Fonte:** Duschak e Couto, 2009

Porém, além desse papel essencial no ciclo viral, as proteases de vírus (+)ssRNA também são capazes de clivar proteínas do hospedeiro, incluindo proteínas do sistema imune inato, o que permite que esses patógenos possam manipular processos dentro da célula (Tsu *et al.*, 2021). A protease NS3/4A do HCV

e a protease 2A do Enterovírus 71 (EV71) são capazes de clivar a proteína MAVS (*mitochondrial antiviral-signaling protein*, em inglês) e, assim diminuir os níveis de interferon (IFN), contribuindo para a persistência do vírus no hospedeiro (Anggakusuma *et al.*, 2016; Wang *et al.*, 2013) Similarmente, a protease NS2B/NS3 do DENV cliva a proteína STING (*stimulator of interferon genes*, em inglês) e inibe a produção de interferon beta (IFN- $\beta$ ) (Yu *et al.*, 2015). Além disso, cliva as mitofusinas MFN1 e MFN2, responsáveis pela fusão mitocondrial, o que prejudica a produção de interferons (Lei e Hilgenfeld, 2017).

Ademais, é importante pontuar que esse processo de clivagem nem sempre funciona de forma a favorecer o hospedeiro. Recentemente, foi mostrado que a proteína CARD8 (*caspase recruitment domain family member 8*, em inglês) funciona como um sensor do HIV-1 (*human immunodeficiency virus 1*, em inglês) ao ser clivada pela protease do vírus (Kulsuptrakul *et al.*, 2023; Wang *et al.*, 2021). Outro exemplo é a clivagem pela protease 3C do rinovírus humano (HRV) que, ao clivar o fragmento N-terminal auto inibitório da proteína NLRP1, libera o fragmento C-terminal, que inicia a formação do inflamassoma (Wang *et al.*, 2021).

### 2.1.1 SARS-COV-2 e a protease nsp5

O vírus SARS-CoV-2 faz parte do grupo de vírus (+)ssRNA e é o agente causador da COVID-19 e da maior emergência em saúde pública do século XXI até o presente momento: segundo a Organização Mundial de Saúde, até fevereiro de 2024, foram mais 700 milhões casos reportados e 7 milhões de óbitos (Organização Mundial da Saúde, 2024).

O genoma do SARS-CoV-2 tem cerca de 30 kilobases (kb), com múltiplas *open reading frames* (ORFs). A ORF1a e 1ab produzem poliproteínas que, quando submetidas a clivagem, resultam em 16 proteínas não estruturais (*non-structural proteins* ou nsps, em inglês), enquanto outras ORFs codificam para proteínas como a spike (S) proteína de membrana (M), a do envelope (E) e a de nucleocapsídeo (N), por exemplo. A clivagem das poliproteínas, nesse caso, é realizada por duas cisteínas proteases: a nsp3 e a nsp5. Esta última, também chamada de 3CLpro (3 *chymotrypsin-like protease*, em inglês) ou protease principal (*main protease*, em inglês), é responsável pela clivagem em pelo menos 11 sítios virais (Zhang *et al.*, 2022).

Além desses sítios virais, vários sítios de clivagem já foram identificados em proteínas humanas, incluindo três alvos importantes: o receptor seletivo de autofagia p62 (Zhang *et al.*, 2022), que, quando degradado, ajuda o vírus a evitar a depuração autofágica; o sensor citosólico de RNA virais RIG-I (*retinoic acid-inducible gene 1*, em inglês) (Liu *et al.*, 2021), que, quando clivado, dificulta o reconhecimento e resposta imune ao vírus; e a proteína NLRP1 (*NACHT, LRR, FIIND, CARD domain and PYD domains-containing protein 1*, em inglês) (Planès *et al.*, 2022), que, quando clivada, ativa a formação do inflamassoma e a morte celular como forma de ativação da resposta imune.

## 2.2 RIG-I, UM IMPORTANTE SENSOR DE RNAs VIRAIS

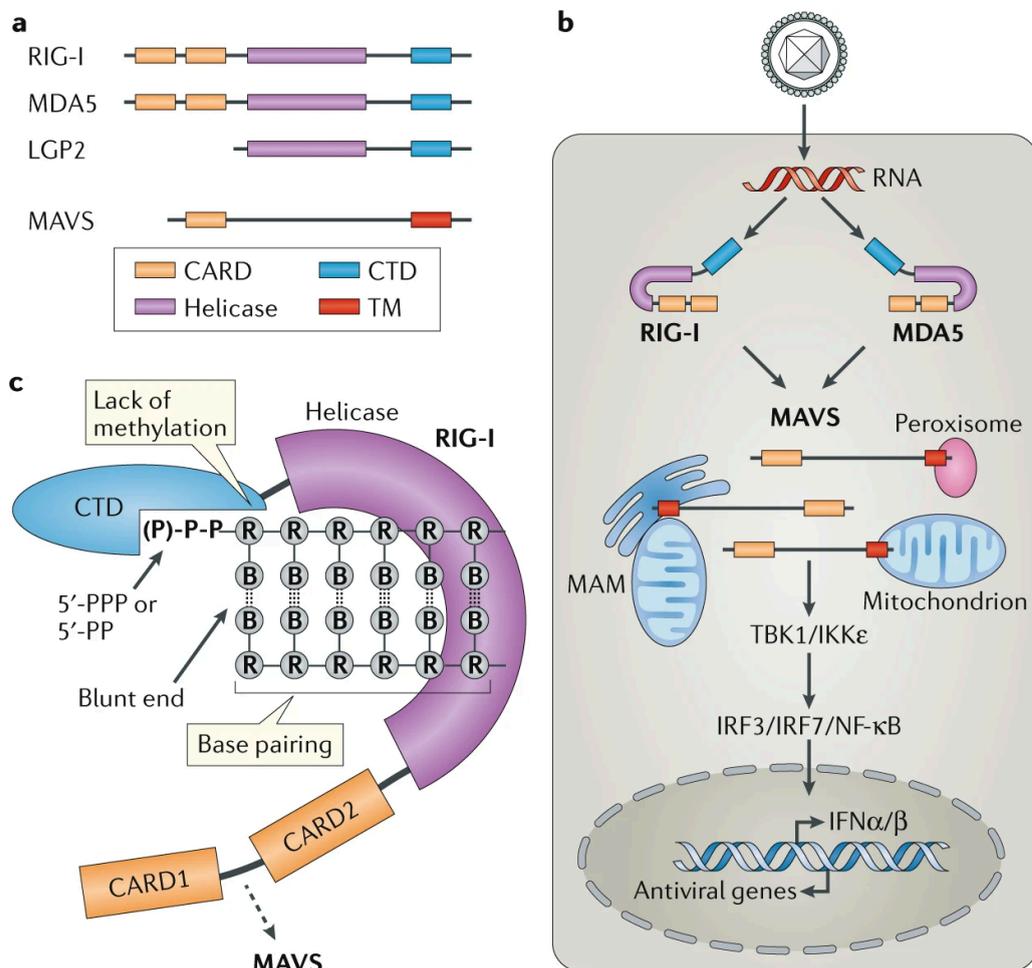
Nos ciclos virais, é essencial para o patógeno a introdução do genoma, seja de DNA ou RNA, na célula infectada. Nos seres humanos, os genomas codificam para várias famílias de sensores de ácidos nucleicos, dentre eles os receptores semelhantes ao gene induzível pelo ácido retinóico I (*retinoic acid-inducible gene I-like receptors* ou RLRs, em inglês). Ademais, hoje é sabido que os RLRs estão envolvidos no reconhecimento de quase todas as principais famílias de vírus. (Rehwinkel e Gack, 2020).

São três os membros desse grupo: MDA5 (*melanoma differentiation-associated protein 5*, em inglês), LGP2 (*laboratory of genetics and physiology 2*, em inglês) e RIG-I. A organização dos seus domínios protéicos é mostrada na Figura 3A. Todos possuem um domínio helicase e um domínio carboxi-terminal (*carboxy-terminal domain* ou CTD, em inglês), os quais funcionam como reconhecedores dos RNAs imunoestimuladores. RIG-I e MDA5 possuem, além disso, dois domínios CARD, que servem para mediar a transdução dos sinais posteriores da via de reconhecimento.

Quando não ligado a um RNA de fita dupla (*double strand RNA* ou dsRNA, em inglês), RIG-I se encontra numa conformação autoinibitória onde o domínio CARD e o helicase interagem, não permitindo sinalização. Dessa forma, quando há a ligação ao dsRNA, o domínio helicase se liga fortemente ao redor do RNA, o que permite uma mudança conformacional que expõe os CARDS para a sinalização (Figura 3C). Essa liberação é importante dado que os CARDS de RIG-I interagem com CARDS de MAVS, uma proteína adaptadora importante que induz uma via para

transcrição de genes codificantes de interferon tipo I e outros genes antivirais e imunoregulatórios (Figura 3B) (Zerbe; Mouser; Cole, 2019)

**Figura 3** - Estrutura e funcionamento geral dos RLRs. **A)** Representação esquemática dos domínios proteínas dos membros dos RLRs. **B)** Via de sinalização de RIG-I e MDA5. **C)** Modelo de ligação e ativação da sinalização de RIG-I.



Fonte: Rehwinkel e Gack. 2020

### 2.3 CONFLITOS EVOLUCIONÁRIOS ENTRE VÍRUS E HOSPEDEIRO

Ao interagirem, vírus e hospedeiro competem pela sua sobrevivência ou para que possam coexistir no mesmo espaço. De fato, esse cenário se enquadra no que

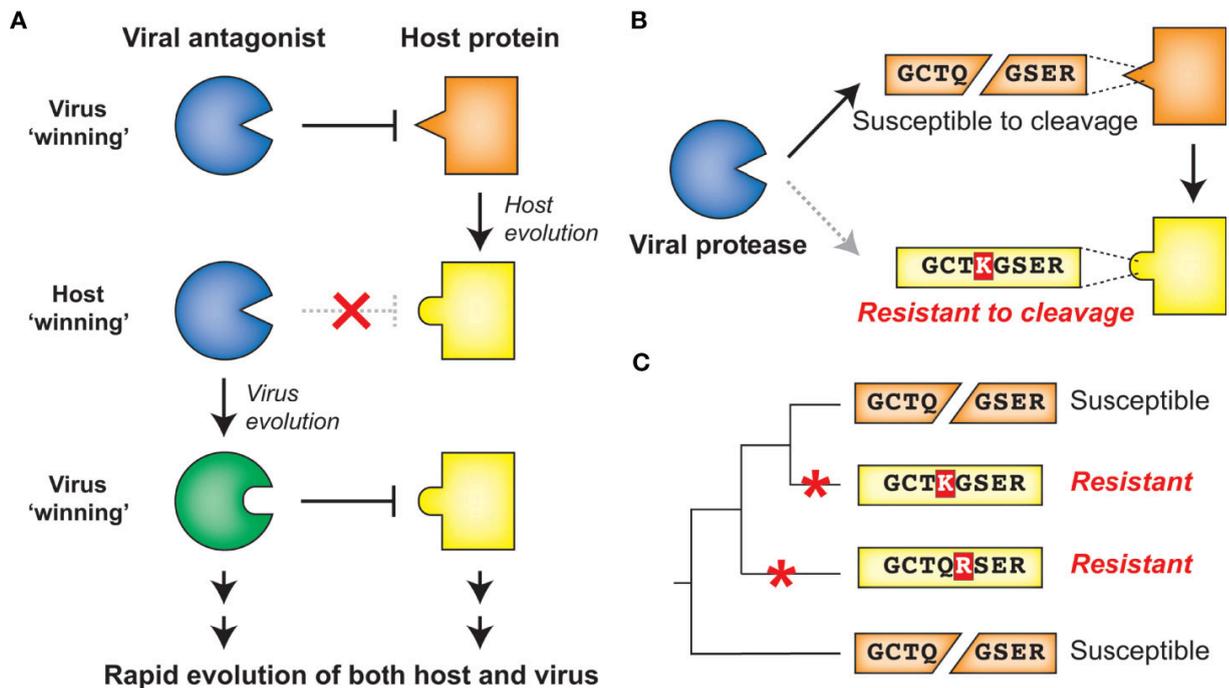
Van Valen propôs em 1973 (Van Valen, 1973), ao publicar a hipótese da Rainha Vermelha (ou *Red Queen hypothesis*, em inglês). Sob essa hipótese, o progresso evolutivo de uma espécie ocorre em detrimento de outras espécies coexistentes, mas, porque estas espécies coexistentes também podem evoluir, não há um ganho de aptidão para nenhuma espécie a longo prazo. Como a Rainha Vermelha disse a Alice no livro *“Through the Looking-Glass”*, de Lewis Carroll, *“Now, here, you see, it takes all the running you can do, to keep in the same place”* (Carroll, 2006).

No caso específico das interações vírus-hospedeiro, adaptações no vírus que favoreçam mecanismos de evasão às estratégias antivirais do hospedeiro são favorecidas e, no hospedeiro, interações que favoreçam a inibição viral ou que ativem defesas antivirais serão preferidas (Tsu *et al.*, 2021). Todavia, ambas as adaptações estão sujeitas a fortes pressões evolutivas, uma vez que o outro lado tentará se adaptar. Dessa forma, um conflito evolucionário passa a existir. As marcas desses conflitos podem ser detectadas por meio de sinais de seleção positiva, seleção onde alelos benéficos têm sua frequência aumentada (Zhang, 2008). Por fim, cabe salientar que essas mudanças moleculares determinam a amplitude de hospedeiros que podem ser infectados e a patogênese viral (Rothenburg e Brennan, 2020)

### 2.3.1 Proteases Virais e Conflitos Evolucionários

Dado que uma especificidade de sequência é requerida para a ocorrência do processo de clivagem das proteínas humanas pelas proteases virais, espera-se que conflitos evolucionários ocorram nessas interfaces. Sendo assim, uma substituição de um resíduo num determinado motivo de clivagem pode alterar a susceptibilidade de clivagem daquele sítio. Essa dinâmica pode ser visualizada de forma gráfica na Figura 4.

**Figura 4** - Conflitos evolucionários no contexto das proteases e suas proteínas-alvo. **A)** A evolução da proteína do hospedeiro que favorece sua aptidão (*fitness*) resulta na evolução da protease viral, o que configura um quadro de conflito entre os progressos evolutivos do patógeno e hospedeiro. **B)** Tendo a especificidade das proteases virais, substituições nos sítios de clivagem podem acarretar na susceptibilidade ou resistência à clivagem. **C)** Numa árvore filogenética, essas substituições podem ocorrer recorrentemente, o que pode resultar na resistência à clivagem entre espécies próximas.



Fonte: Tsu *et al.*, 2021

Existem evidências que dão suporte a essa hipótese. Num nível interespecífico, já foi demonstrado que, em primatas, as MAVS evoluíram sobre uma forte evolução positiva, acarretando substituições que tornam os sítios de clivagem de algumas espécies resistentes à clivagem pela protease NS3/4A do HCV. Consequentemente, essas substituições permitem uma melhor restrição à replicação viral (Patel *et al.*, 2012). Ademais, num nível intraespecífico, na população humana, um SNP que acarreta na substituição de uma serina por uma prolina na posição 39 (posição P2' do sítio de clivagem) de CARD8 diminui o processamento pela nsp5. Destarte, a formação do inflamassoma e o mecanismo de detecção do vírus são afetados (Tsu *et al.*, 2023).

Portanto, a análise das variações de sequência nessas interfaces de

interações entre a nsp5 e seus alvos humanos fornece oportunidades de entender onde e como, molecularmente, esses conflitos evolucionários ocorrem. Conseqüentemente, é possível também achar nessas variações possíveis respostas para a diversidade de fenótipos em relação à fisiopatologia da COVID-19.

### 3 OBJETIVOS

#### 3.1 OBJETIVO GERAL

Caracterizar a variação nas sequências dos sítios de clivagem humanos da protease nsp5 de SARS-CoV-2.

#### 3.1 OBJETIVOS ESPECÍFICOS

Identificar todos os sítios de clivagem da protease nsp5 nos seres humanos comprovados experimentalmente.

Identificar o perfil físico-químico das preferências de clivagem da protease nsp5.

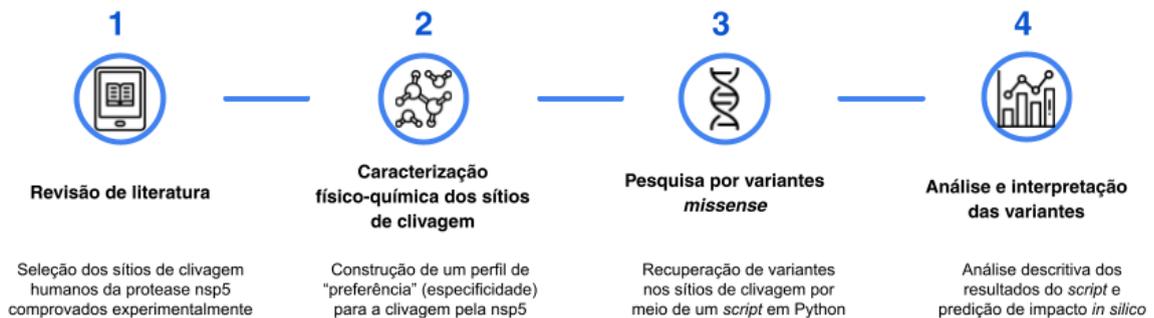
Construir um algoritmo capaz de recuperar informações sobre SNPs *missense* que ocorrem nas regiões codificantes dos aminoácidos existentes nos sítios de clivagem da protease nsp5.

Caracterizar os SNPs *missense* que ocorrem nas regiões codificantes dos aminoácidos existentes nos sítios de clivagem da protease nsp5 quanto a sua frequência alélica e seu impacto funcional *in silico*.

## 4 METODOLOGIA

As etapas desenvolvidas neste projeto podem ser sumarizadas na Figura 5. Abaixo, cada uma dessas etapas será apresentada de forma mais detalhada.

**Figura 5** - Etapas do projeto de pesquisa



**Fonte:** Autor, 2024

### 4.1 MINERAÇÃO DE DADOS E SELEÇÃO DOS SÍTIOS DE CLIVAGEM

Primeiramente, foi realizada uma extensa revisão de literatura e mineração de dados em busca de sítios de clivagem da protease nsp5 de SARS-CoV-2 no proteoma humano, considerando-se trabalhos publicados até dezembro de 2022. Nesta busca foram considerados somente trabalhos que comprovaram a clivagem experimentalmente (*in vitro*), excluindo-se trabalhos *in silico*. No que tange às posições dos sítios de clivagem que foram consideradas, somente os resíduos de P4 à P1', seguindo a nomenclatura de Schechter e Berger (Schechter e Berger, 1967) foram recuperados. Isso porque, evidências apontam que, tanto para SARS-CoV como para SARS-CoV-2, a maior especificidade parece estar restrita a esses resíduos (Chuck *et al.*, 2010; Goetz *et al.*, 2007; Jin *et al.*, 2020). Em seguida, foi feita uma curadoria a fim de identificar e excluir sítios duplicados (identificados em mais de um trabalho) ou aqueles onde a amplitude considerada anteriormente não era possível de ser recuperada (os cinco resíduos

necessários não estavam disponíveis).

#### 4.2 CARACTERIZAÇÃO FÍSICO-QUÍMICA DOS SÍTIOS DE CLIVAGEM

Para a interpretação e análise do impacto das mutações nos sítios de clivagem é necessário conhecer quais características dotam um sítio de ser clivável ou não. Dessa forma, foi feita uma caracterização físico-química a partir dos sítios obtidos anteriormente para determinar os aminoácidos mais frequentes em cada posição de clivagem, além de características químicas ou físicas: hidrofobicidade, tamanho, tamanho do grupo R, carga, estrutura secundária e desordem.

Primeiramente, a fim de realizar a caracterização da conservação de aminoácidos em cada uma das posições, foi construído, com a ferramenta WebLogo (Crooks *et al.*, 2004) um *sequence logo*, uma representação gráfica dos padrões vistos no alinhamento múltiplo de sequências, as quais são, nesse caso, os sítios de clivagem.

Em relação a características tridimensionais (3D), foi analisada a distribuição dos volume dos grupos R dos aminoácidos com dados de Chuck e colaboradores (Chuck *et al.*, 2010). Além disso, foram utilizados *scripts* próprios em Python para a extração e análise dos perfis de estrutura secundária e de exposição a partir de modelos 3D preditos pelo AlphaFold. O AlphaFold é um algoritmo de inteligência artificial capaz de prever estruturas protéicas com alta acurácia e velocidade (Jumper *et al.*, 2021). Para a classificação do perfil de exposição, foi utilizada a medida relativa de acessibilidade ao solvente (*relative solvent accessibility* ou RSA, em inglês), de forma que

$$RSA = ASA \div MaxASA$$

onde ASA é a área de superfície acessível ao solvente (*accessible surface area*, em inglês) do resíduo e MaxASA é o máximo de ASA para aquele resíduo. Os valores de ASA foram recuperados dos modelos preditos e os de MaxASA a partir do trabalho de Tien e colaboradores (Tien *et al.*, 2013). Resíduos com valores de RSA maiores que 0,25 foram classificados como expostos e menores que 0,25 como enterrados (ou *buried*, em inglês) (Zhang; Li; Lü, 2018). Para a classificação da estrutura secundária, foi utilizada a de três classes:

folha-beta (E), alfa-hélice (H) e outras regiões (C).

Além disso, as estruturas preditas pelo AlphaFold possuem uma métrica para cada resíduo que estima a confiança da predição, numa escala de 0 a 100, chamada de pLDDT. Os autores da ferramenta apontam que regiões com valores menores que 50 servem razoavelmente como preditores de desordem, ou seja, regiões que não possuem uma estrutura tridimensional estável em condições fisiológicas. Sendo assim, a distribuição dos valores de pLDDT também foi analisada.

#### 4.3 PESQUISA POR VARIANTES *MISSENSE*

Utilizando-se de um *script* próprio em Python, foi realizada a pesquisa por SNPs (*single nucleotide polymorphisms*, em inglês) do tipo *missense* para as posições dos sítios de clivagem obtidos anteriormente. O *script* funciona da seguinte forma: primeiramente, ele recebe como entrada (*input*) um arquivo .csv (*comma separated value*) com a sequência do sítio de clivagem, o Uniprot ID (identificador de uma sequência proteica no banco de dados Uniprot, <https://www.uniprot.org/>) e o nome do gene. Além disso, é necessário fornecer também a sequência FASTA para que o programa mapeie a posição dos sítios na proteína como um todo. Haja vista que as proteínas podem ter diferentes isoformas e isso poderia resultar na recuperação de variantes que não impactam os sítios de clivagem, foram utilizadas sequências proteicas consenso obtidas a partir do projeto MANE (*Matched Annotation from the NCBI and EMBL-EBI*, em inglês). Esse projeto gerou o conjunto de dados MANE Select, que consiste num conjunto de transcritos (e suas proteínas correspondentes) representativos para cada locus obtido por meio de métodos computacionais e revisão manual e discussão (Morales *et al.*, 2022).

Com essas informações, foram recuperados os SNPs *missense* a partir de um banco de dados específico, o dbSNP (<https://www.ncbi.nlm.nih.gov/snp/>). Apesar desse banco de dados concentrar informações sobre variações genéticas humanas de vários bancos, neste trabalho foram utilizadas somente as informações do banco de dados gnomAD (Gudmundsson *et al.*, 2021). Atualmente, o gnomAD é o maior e mais utilizado banco de variação genética na população humana, sendo composto por mais de

70.000 genomas e mais de 700.000 exomas. O *script* deverá ter como resultado um arquivo .csv com informações sobre o rsID do SNP, o gene, a frequência alélica, o número de alelos e a posição de clivagem.

#### 4.4 ANÁLISE E INTERPRETAÇÃO DAS VARIANTES

Com os resultados obtidos a partir do *script* anterior, foi realizada uma análise descritiva para a distribuição das frequências alélicas (FA) dos SNPs presentes na população geral.

Além disso, a fim de mensurar estatisticamente o impacto dessas substituições, foi construída, com os mesmos motivos de clivagem obtidos na revisão de literatura, uma matriz de pontuação posição-específica (*position-specific scoring matrix* ou PSSM, no inglês). As PSSMs, basicamente, são matrizes que especificam pontuações para a observação de um determinado aminoácido em determinada posição. Para sua construção, primeiramente, foi feito um alinhamento com todos os sítios de clivagem, a partir do qual foi construída uma matriz de frequência absoluta, onde cada linha representa um aminoácido e cada coluna representa uma posição de clivagem. Cada elemento dessa matriz é a frequência absoluta daquele aminoácido naquela posição. Na Figura 6 é possível observar essa matriz para um alinhamento de sequências de nucleotídeos.

**Figura 6** - Matriz de frequência absoluta

$$M = \begin{matrix} A \\ C \\ G \\ T \end{matrix} \begin{bmatrix} 3 & 6 & 1 & 0 & 0 & 6 & 7 & 2 & 1 \\ 2 & 2 & 1 & 0 & 0 & 2 & 1 & 1 & 2 \\ 1 & 1 & 7 & 10 & 0 & 1 & 1 & 5 & 1 \\ 4 & 1 & 1 & 0 & 10 & 1 & 1 & 2 & 6 \end{bmatrix}.$$

Fonte: Wikipédia, 2024

Em seguida, essa matriz de frequência foi transformada em uma matriz de probabilidades, na qual a probabilidade de observar um aminoácido em uma determinada posição é simplesmente o número de vezes que ele foi observado ali dividido pelo número de sequências no alinhamento (Figura 7). Com essa matriz, pode-se calcular a “probabilidade” de determinada sequência, multiplicando cada probabilidade de cada posição do aminoácido da sequência com os valores da

matriz.

**Figura 7** - Matriz de probabilidades

$$M = \begin{matrix} A \\ C \\ G \\ T \end{matrix} \begin{bmatrix} 0.3 & 0.6 & 0.1 & 0.0 & 0.0 & 0.6 & 0.7 & 0.2 & 0.1 \\ 0.2 & 0.2 & 0.1 & 0.0 & 0.0 & 0.2 & 0.1 & 0.1 & 0.2 \\ 0.1 & 0.1 & 0.7 & 1.0 & 0.0 & 0.1 & 0.1 & 0.5 & 0.1 \\ 0.4 & 0.1 & 0.1 & 0.0 & 1.0 & 0.1 & 0.1 & 0.2 & 0.6 \end{bmatrix}$$

Fonte: Wikipédia, 2024

Em seguida, essa matriz de probabilidade foi finalmente transformada em matriz de pontuação (Figura 8) através do cálculo da razão de chances (*odds ratio*, em inglês), que consiste na razão entre a frequência observada (calculada anteriormente) e a frequência esperada. Para aminoácidos, as frequências esperadas podem variar, uma vez que as frequências entre eles variam na natureza. Neste trabalho, foram utilizadas as frequências de Kyte e Doolittle (Kyte e Doolittle, 1982). A cada elemento da matriz de pontuação será aplicado um logaritmo de base 2 ( $\log_2$ ).

**Figura 8** - Matriz de pontuação posição-específica. Os valores infinitos nesse exemplo se devem ao fato de não se terem utilizado pseudocontagens, que são utilizados quando resíduos tem uma contagem 0 em uma posição. A utilização de pseudocontagens evita um sobreajuste da matriz e os valores infinitos.

$$M = \begin{matrix} A \\ C \\ G \\ T \end{matrix} \begin{bmatrix} 0.26 & 1.26 & -1.32 & -\infty & -\infty & 1.26 & 1.49 & -0.32 & -1.32 \\ -0.32 & -0.32 & -1.32 & -\infty & -\infty & -0.32 & -1.32 & -1.32 & -0.32 \\ -1.32 & -1.32 & 1.49 & 2.0 & -\infty & -1.32 & -1.32 & 1.0 & -1.32 \\ 0.68 & -1.32 & -1.32 & -\infty & 2.0 & -1.32 & -1.32 & -0.32 & 1.26 \end{bmatrix}$$

Fonte: Wikipédia, 2024.

Ao aplicar esse logaritmo, pode-se calcular, enfim, a pontuação (ou *score*, em inglês) de uma sequência somando cada valor de acordo com cada posição da matriz. O valor resultante dessa soma indica quão diferente aquela sequência é de uma sequência aleatória. Dessa forma, se o *score* de uma sequência é maior do que 0, ela tem maior probabilidade de ser um sítio funcional (nesse caso, um sítio de clivagem da nsp5) do que uma sequência aleatória. Se a pontuação de uma dada sequência for menor do que 0, é mais provável que ela seja um sítio aleatório e não um sítio funcional. Portanto, foi calculada a distribuição das

pontuações com os sítios selecionados e com suas respectivas substituições decorrentes das variantes *missense* a fim de avaliar o comportamento dessas pontuações.

Por último, foi realizada uma análise *in silico* mais aprofundada com a mutação mais frequente na população humana em geral. Essa mutação será avaliada por meio de alguns algoritmos de predição de impacto, descritos no Quadro 1.

**Quadro 1.** Algoritmos utilizados para predição de impacto da mutação

| Algoritmo | Funcionalidade  | Score de análise  | Referência                      |
|-----------|---|---|---------------------------------|
| NetCorona | Algoritmo para predição de sítios de clivagem da nsp5 de SARS-CoV por meio de uma rede neural   | < 0,5 indica que provavelmente não é clivado<br>0,5 a 0,8 indica possibilidade de ser clivado<br>> 0,8 indica que muito provavelmente é clivado | (Kierner <i>et al.</i> , 2004)  |
| mCSM      | Algoritmo baseado na distância entre os átomos e o resíduo sendo considerado  | $\Delta\Delta G$ positivo indica aumento da estabilidade<br>$\Delta\Delta G$ negativo indica uma substituição desestabilizadora                 | (Pires; Ascher; Blundell, 2013) |
| muPRO     | Algoritmo baseado em SVM ( <i>Support Vector Machine</i> ) que se utiliza de informações de sequência e estruturais para predição do impacto de substituições sobre a estabilidade proteica | $\Delta\Delta G$ positivo indica aumento da estabilidade<br>$\Delta\Delta G$ negativo indica uma substituição desestabilizadora                 | (Cheng; Randall; Baldi, 2005)   |
| DUET      | Algoritmo que se utiliza de um método integrado, gerando um consenso entre outros dois algoritmos (mCSM e SDM)  | $\Delta\Delta G$ positivo indica aumento da estabilidade e $\Delta\Delta G$ negativo indica uma substituição desestabilizadora                  | (Pires; Ascher; Blundell, 2014) |

#### 4.5 ANÁLISE ESTRUTURAL

Enfim, um estudo mutagênico com a variante mais frequente foi feito por meio do software PyMol (<https://pymol.org/2/>), onde foi possível visualizar a substituição, analisando-se a hidrofobicidade utilizando a escala de Eisenberg (Eisenberg *et al.*, 1984) e a carga potencial de contato eletrostático calculado pelo próprio programa.

## **5 RESULTADOS**

### **5.1 COLETA DE DADOS**

Os resultados da busca na literatura e da curadoria realizada estão apresentados no Quadro 2. Ao final, foram obtidos 111 sítios de clivagem (sendo 11 sítios das próprias poliproteínas do vírus), ou motivos de clivagem. Além disso, cabe salientar que houve sobreposição de resultados entre os trabalhos, com um total de 5 sítios de clivagem sendo identificados em dois trabalhos distintos.

**Quadro 2.** Sítios de clivagem da nsp5 comprovados experimentalmente até dezembro de 2022

| Gene           | Motivo de clivagem | Gene Ontology - Processo Biológico                 | Referência                  |
|----------------|--------------------|--|-----------------------------|
| <i>ALDOA</i>   | KRLQS              | actin filament organization                        | Pablos <i>et al.</i> , 2021 |
| <i>ARPC4</i>   | ATLQA              | actin filament polymerization                      |                             |
| <i>ATAD2</i>   | AVLQA              | positive regulation of DNA-templated transcription |                             |
| <i>ATL2</i>    | SMLQA              | endoplasmic reticulum organization                 |                             |
| <i>BTA1</i>    | PTLQA              | negative regulation of chromatin binding           |                             |
| <i>CARHSP1</i> | PTHQA              | intracellular signal transduction                  |                             |
| <i>CCT6A</i>   | ARAQA              | chaperone-mediated protein folding                 |                             |
| <i>CLSPN</i>   | LSLQS              | activation of protein kinase activity              |                             |
| <i>CLTB</i>    | AIAQA              | clathrin-dependent endocytosis                     |                             |
| <i>CREB1</i>   | VVVQA              | axonogenesis                                       |                             |
| <i>DYNC1H1</i> | ANLQM              | cell division                                      |                             |
| <i>EIF4G1</i>  | STPQA              | behavioral fear response                           |                             |
| <i>EIF4G2</i>  | SQLQG              | cell death   |                             |
| <i>FAF1</i>    | GILQS              | apoptotic process                                  |                             |
| <i>GOLGA2</i>  | GQLQA              | asymmetric cell division                           |                             |
| <i>HDLBP</i>   | PELQS              | cholesterol metabolic process                      |                             |

|                |       |   |
|----------------|-------|---|
| <i>HSPA1A</i>  | GVFQH | ATP metabolic process   |
| <i>HSPA1B</i>  | GVFQH | ATP metabolic process   |
| <i>KPNA3</i>   | AILQN | NLS-bearing protein import into nucleus                           |
| <i>KRT8</i>    | SRLQA | cell differentiation involved in embryonic placenta development   |
| <i>MCM4</i>    | PQMHS | DNA replication   |
| <i>MDC1</i>    | PELQA | DNA damage response   |
| <i>NACA</i>    | TTQQA | protein targeting to membrane                                     |
| <i>NUP107</i>  | VLLQA | female gonad development  |
| <i>PRRC2A</i>  | PTLQA | cell differentiation  |
| <i>PSMD8</i>   | AVLQA | proteasome-mediated ubiquitin-dependent protein catabolic process |
| <i>PTBP1</i>   | PNVHG | IRES-dependent viral translational initiation                     |
| <i>RABEP1</i>  | AVLQA | apoptotic process   |
| <i>RPAP1</i>   | ARLQA | transcription by RNA polymerase II                                |
| <i>RPS21</i>   | ASIQM | cytoplasmic translation   |
| <i>SEPTIN6</i> | VQLQS | cell differentiation  |
| <i>SEPTIN9</i> | SQLQS | cytoskeleton-dependent cytokinesis                                |
| <i>SERBP1</i>  | SAAQA | negative regulation of translation                                |
| <i>SF3B2</i>   | AQQQA | mRNA processing   |

Pablos *et al.*, 2021

|                 |       |   |                               |
|-----------------|-------|---|-------------------------------|
| <i>SRRM2</i>    | SRFQS | mRNA splicing, via spliceosome                                | Pablos <i>et al.</i> , 2021   |
| <i>STMN1</i>    | ASGQA | axonogenesis  |                               |
| <i>SUPT6H</i>   | HQLQA | blastocyst formation  |                               |
| <i>TSC22D2</i>  | AQLMG | negative regulation of cell cycle                             |                               |
| <i>WDR33</i>    | SSLQG | mRNA polyadenylation  |                               |
| <i>ZC3H4</i>    | AALQS | DNA-templated transcription termination                       |                               |
| <i>ZFYVE16</i>  | ILLQS | endosomal transport   |                               |
| <i>SPTBN1</i>   | ARLQA | actin cytoskeleton organization                               | Pablos <i>et al.</i> , 2021   |
|                 | SQAQA | actin cytoskeleton organization                               | Koudelka <i>et al.</i> , 2021 |
| <i>PNN</i>      | PVLQS | cell adhesion   | Pablos <i>et al.</i> , 2021   |
|                 | PALQS | cell adhesion   | Meyer <i>et al.</i> , 2021    |
| <i>ACTA2</i>    | PRHQG | cellular response to transforming growth factor beta stimulus | Koudelka <i>et al.</i> , 2021 |
| <i>ACTB</i>     | PRHQG | adherens junction assembly                                    |                               |
| <i>ACTG1</i>    | PRHQG | angiogenesis  |                               |
| <i>ALDH18A1</i> | SDLQG | citruilline biosynthetic process                              |                               |
| <i>ATIC</i>     | TEMQS | de novo' AMP biosynthetic process                             |                               |
| <i>CDC42BPB</i> | SIMQS | actin cytoskeleton organization                               |                               |
| <i>CDC5L</i>    | AILQT | cell cycle  |                               |
| <i>CORO7</i>    | RSLQS | actin filament polymerization                                 |                               |

|                |       |  |
|----------------|-------|--|
| <i>FLNA</i>    | PGIQS | actin crosslink formation  |
| <i>GRIPAP1</i> | LRLQA | negative regulation of receptor clustering                       |
| <i>GSN</i>     | ATLHT | actin filament capping   |
| <i>HNRNPF</i>  | LKFMS | mRNA splicing, via spliceosome                                   |
| <i>HSP90B1</i> | MKAQA | actin rod assembly   |
| <i>HSPA9</i>   | AVMEG | chaperone cofactor-dependent protein refolding                   |
| <i>HSPD1</i>   | LMLQG | de novo' protein folding   |
| <i>IQGAP1</i>  | ARPHY | cell migration   |
| <i>KARS1</i>   | AAVQA | basophil activation involved in immune response                  |
| <i>LARP1</i>   | TLLQA | cell population proliferation                                    |
| <i>LASP1</i>   | PVAQS | monoatomic ion transport   |
| <i>LIMA1</i>   | ASFQS | actin filament bundle assembly                                   |
| <i>LMNA</i>    | ASSHS | cellular response to hypoxia                                     |
|                | SGAQA | cellular response to hypoxia                                     |
|                | TRLQE | cellular response to hypoxia                                     |
| <i>OGDH</i>    | AHAQS | 2-oxoglutarate metabolic process                                 |
| <i>PAICS</i>   | VLLQS | de novo' AMP biosynthetic process                                |
| <i>PLEC</i>    | SLAQA | actomyosin contractile ring assembly actin filament organization |
|                | LRLQA | actomyosin contractile ring assembly actin                       |

Koudelka *et al.*, 2021

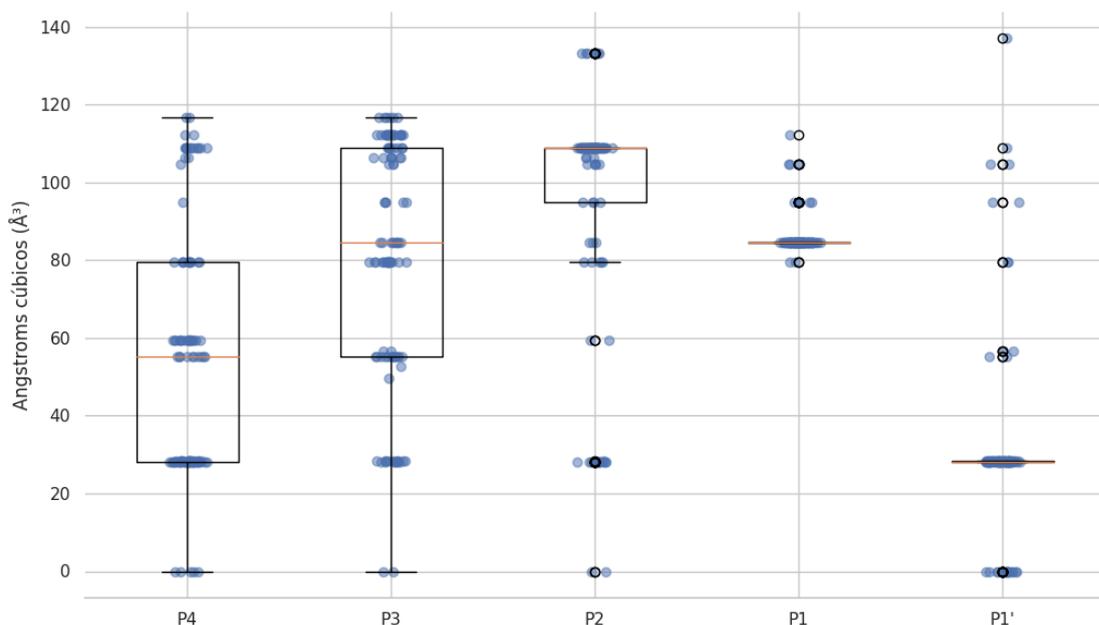
|                 |       |   |                               |
|-----------------|-------|---|-------------------------------|
|                 |       | filament organization   | Koudelka <i>et al.</i> , 2021 |
|                 | ARAQA | actomyosin contractile ring assembly actin filament organization    |                               |
|                 | TAQRS | actomyosin contractile ring assembly actin filament organization    |                               |
| <i>PRKCSH</i>   | IELQA | intracellular signal transduction                                   |                               |
| <i>PRPF4B</i>   | LILQG | mRNA cis splicing, via spliceosome                                  |                               |
| <i>RALY</i>     | LKLQA | cholesterol homeostasis   |                               |
| <i>RBM15</i>    | SRLHS | branching involved in blood vessel morphogenesis                    |                               |
| <i>SAFB</i>     | AHAQS | chromatin organization  |                               |
| <i>SAFB2</i>    | AHAQS | regulation of androgen receptor signaling pathway                   |                               |
| <i>SART1</i>    | LRLQA | maturation of 5S rRNA   |                               |
| <i>SERPINH1</i> | SALQS | chondrocyte development involved in endochondral bone morphogenesis |                               |
| <i>TRIM28</i>   | PVLMA | chromatin organization  |                               |
| <i>TUBB</i>     | VHIQA | cell division   |                               |
| <i>TUBB4B</i>   | VHLQA | microtubule cytoskeleton organization                               |                               |

|               |       |  |                                |
|---------------|-------|--|--------------------------------|
| <i>NLRP12</i> | KLFQG | activation of cysteine-type endopeptidase activity involved in apoptotic process | Moustaqil <i>et al.</i> , 2021 |
|               | VVLQA | activation of cysteine-type endopeptidase activity involved in apoptotic process |                                |
| <i>TAB1</i>   | ASLQS | aorta development  |                                |
|               | LTLQS | aorta development  |                                |
| <i>CTBP1</i>  | TRVQS | negative regulation of cell population proliferation                             |                                |
| <i>IRAK1</i>  | STLQA | activation of NF-kappaB-inducing kinase activity                                 |                                |
| <i>SQSTM1</i> | GELQS | aggrephagy   | Zhang <i>et al.</i> , 2022     |
| <i>RNF20</i>  | ALLQS | negative regulation of cell migration  | Zhang <i>et al.</i> , 2021     |
| <i>GOLGA3</i> | TKLQA | -  | Meyer <i>et al.</i> , 2021     |
| <i>GSDMD</i>  | TCLQG | cellular response to extracellular stimulus                                      | Shi <i>et al.</i> , 2022       |
| <i>IKBKG</i>  | AQLQV | anoikis  | Wenzel <i>et al.</i> , 2021    |
|               | LLGEL | anoikis  | Chen <i>et al.</i> , 2022      |
|               | LRMQG | anoikis  |                                |
| <i>NLRP1</i>  | VILQG | activation of cysteine-type endopeptidase activity involved in apoptotic process | Planès <i>et al.</i> , 2022    |
| <i>RIGI</i>   | RSLQA | antiviral innate immune response   | Liu <i>et al.</i> , 2021       |

## 5.2 CARACTERIZAÇÃO FÍSICO-QUÍMICA DOS SÍTIOS DE CLIVAGEM

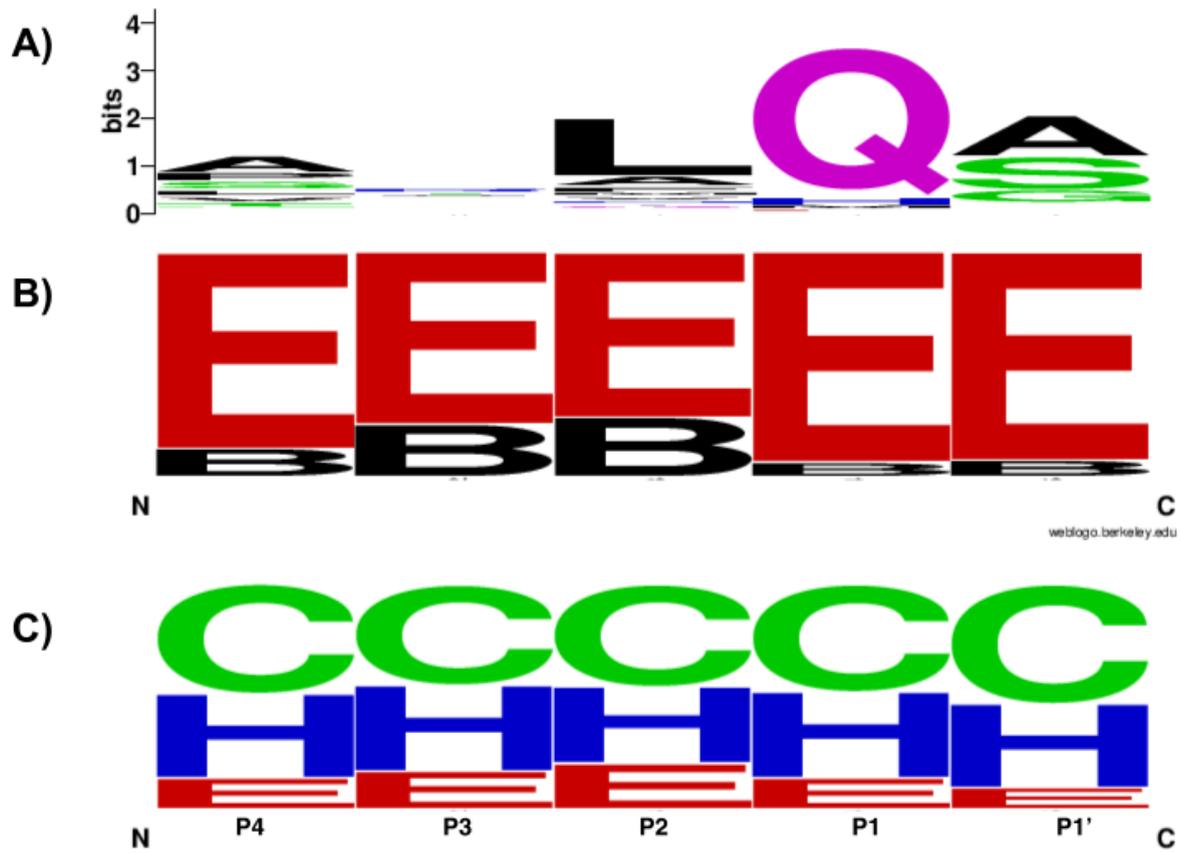
Com os 111 sítios foi realizada a caracterização físico-química. Excetuando-se a construção do *sequence logo*, todos os outros gráficos foram construídos a partir do processamento de dados por meio de *scripts* próprios, disponíveis em [https://github.com/agenorneto/cleave\\_me\\_or\\_leave\\_me](https://github.com/agenorneto/cleave_me_or_leave_me). A Figura 9 apresenta a distribuição dos volumes dos grupos R laterais em cada posição de clivagem. Primeiramente, na posição P1 há, como visto na Figura 10A, uma preferência quase absoluta por resíduos de glutamina (Q). Porém, algumas sequências apresentaram metionina, histidina e ácido glutâmico. Em relação à posição P1', foi visualizado uma preferência por resíduos com pequenos, com um volume do grupo R menor ( $\mu = 23,08 \text{ \AA}^3$ ), sendo três em maior frequência: alanina, glicina e serina. No que tange a posição P2, foi vista uma maior frequência de resíduos hidrofóbicos, sendo a leucina o mais presente. Em P3, não foi observado um padrão físico-químico específico, sendo os resíduos mais frequentes a arginina, a glutamina e a treonina e a valina. A posição P4 apresentou maior frequência de resíduos com volume de grupo R menor ( $\mu = 53,00 \text{ \AA}^3$ ) e hidrofóbicos, sendo a alanina o resíduo mais frequente.

**Figura 9** - *Boxplot* dos volumes dos grupo R laterais dos aminoácidos dos sítios de clivagem humanos da nsp5



Fonte: Autor, 2024

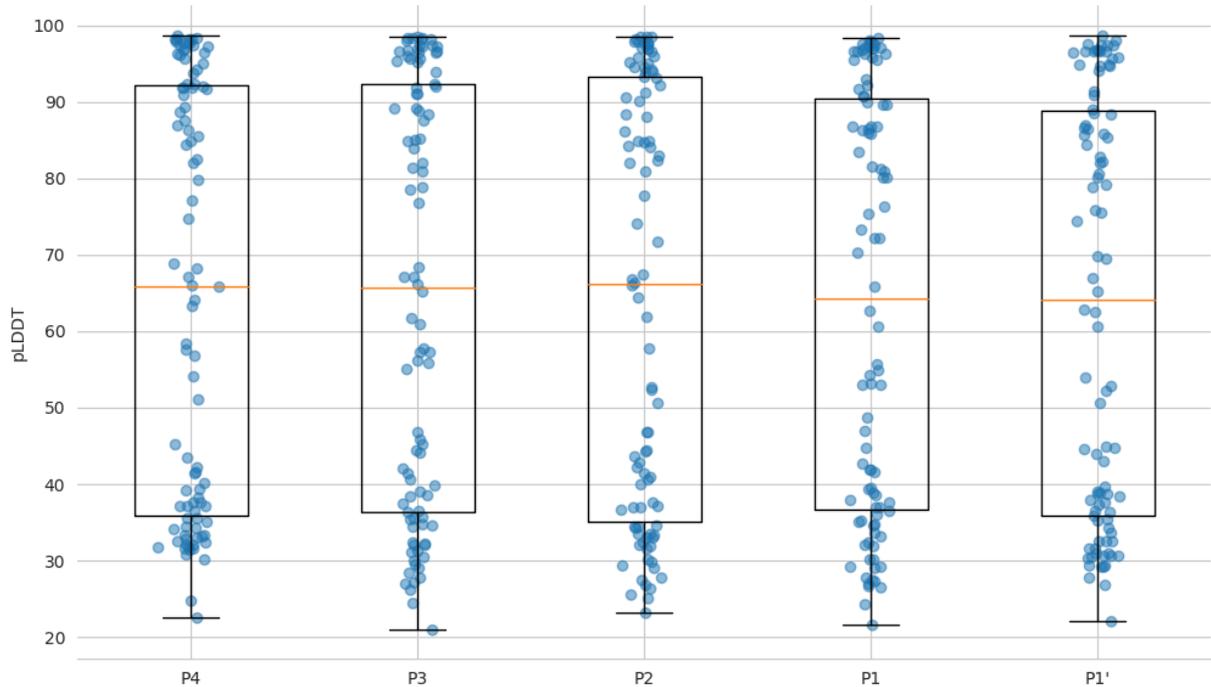
**Figura 10** - Perfil de preferência físico-química dos sítios da protease nsp5, a partir dos 111 sítios obtidos na revisão de literatura. **A)** *Sequence logo*. **B)** Perfil de exposição. **C)** Perfil de estrutura secundária



Fonte: Autor, 2024

Em relação ao perfil de exposição (Figura 10B), foi perceptível que a grande maioria dos resíduos se encontrava em exposição, principalmente nas posições P1 e P1'. Em relação a estrutura secundária (Figura 10C), foi possível observar que há uma predominância da ocorrência dos sítios de clivagem em regiões de *loop* e de alfa-hélices. Por fim, em relação a medida de pLDDT (Figura 11), é perceptível que para todos as posições de clivagem há uma distribuição mais concentrada nos dois extremos da escala desse valor. Isso converge com os resultados de estrutura secundária, uma vez que *loops* tendem a ter valores de pLDDT menores por serem mais desordenados e alfa-hélices, valores maiores, por serem estruturas mais ordenadas.

**Figura 11** - *Boxplot* dos valores de pLDDT dos aminoácidos dos sítios de clivagem humanos da nsp5.



Fonte: Autor, 2024

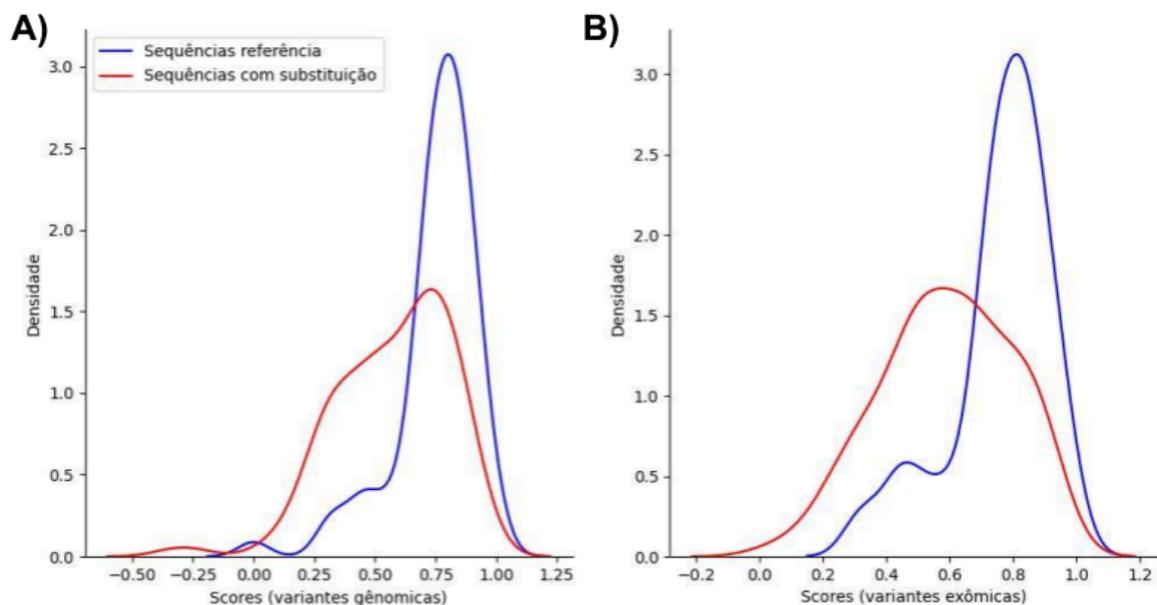
### 5.3 PESQUISA POR VARIANTES *MISSENSE*

Foram recuperadas 242 variantes a partir do *script* construído, sendo 143 provenientes de sequenciamentos genômicos e 167 de sequenciamentos exômicos. Além disso, em relação a frequência alélica (FA), no conjunto de variantes genômicas, 86 variantes foram *singletons*, isto é, só ocorrem uma única vez em todo o banco de dados, e 56 tiveram uma FA menor do que 0,01 (raras). Somente uma variante teve FA comum (FA > 0,05 na população): a variante 5177C>T (*rs10813831*) tem uma frequência de 0,23. No conjunto de variantes exômicas, esse mesmo comportamento ocorreu com 90, 76 e 1 (a mesma vista no outro conjunto) variantes nas classes de *singletons*, raras e comum. A única variante comum também foi a mesma nesse conjunto, com uma FA de 0,21.

No que tange ao impacto desses SNPs, após processamento dos resultados e cálculo das pontuações (*scores*) para sequências selvagens e sequências dos sítios mutados, foi possível observar o comportamento da distribuição desses

scores. Nos gráficos de densidade da Figura 12, é possível observar que a curva de densidade muda de forma quando se comparam os scores entre as sequências originais e as com substituições. Essa mudança, vista tanto no conjunto de variantes genômicas como exômicas, é para esquerda, denotando uma tendência de diminuição das pontuações.

**Figura 12** - Gráficos de densidade das distribuições das pontuações dos sítios de clivagem das sequências referência e dos sítios com as substituições provenientes das variantes, tanto A) genômicas como B) exômicas.



Fonte: Autor, 2024

A variante 5177C>T (rs10813831) ocorre no gene *DDX58*, que codifica a proteína RIG-I (*retinoic acid-inducible gene 1*), envolvida no reconhecimento de RNAs de dupla fita (*double strand RNA*, dsRNA) e a regulação da resposta imune inata contra vírus. Essa variante impacta a arginina (R) da posição 7, substituindo-a por uma cisteína (C). No que se refere a sua frequência alélica nas subpopulações humanas, existe uma heterogeneidade, a qual vai de 0,09 em indivíduos do Leste da Ásia até 0,26 em indivíduos da Europa (não-finlandeses). Essa mesma variabilidade é visível também nos dados do *Human Genome Diversity Project* (Bergström *et al.*, 2020), onde a frequência é ainda menor em nativos americanos (FA=0,057). A nível da população brasileira, essa variante apresenta uma FA de 0,24 no Arquivo Brasileiro Online de Mutações (ABraOM) (Naslavsky *et al.*, 2022), o qual possui

sequências exômicas e genômicas de 1.171 indivíduos idosos de São Paulo, a cidade mais populosa do país.

#### 5.4 ANÁLISE E INTERPRETAÇÃO DAS VARIANTES

As predições feitas no algoritmo NetCorona com a sequência selvagem da proteína RIG-I e com a substituição da arginina por uma cisteína na posição 7 foram comparadas. Na sequência original, nenhum motivo ultrapassa o limite de 0,5, indicando que não há nenhum sítio de clivagem predito. Por sua vez, a sequência com a substituição, resulta em somente um sítio de clivagem com valor acima de 0,5. Esse sítio é justamente o que envolve essa mudança (CSLQ↓A, pontuação: 0,514, onde ↓ indica o local da clivagem da ligação peptídica). É importante notar que esse sítio, na sequência original (RSLQ↓A) foi o que apresentou a maior pontuação, sendo próxima de 0,5 (0,478).

Em relação à estabilidade proteica, os três algoritmos, segundo os critérios estabelecidos anteriormente, predisseram que essa substituição é desestabilizante para a proteína (Tabela 1).

**Tabela 1.** Resultados dos algoritmos de predição de estabilidade

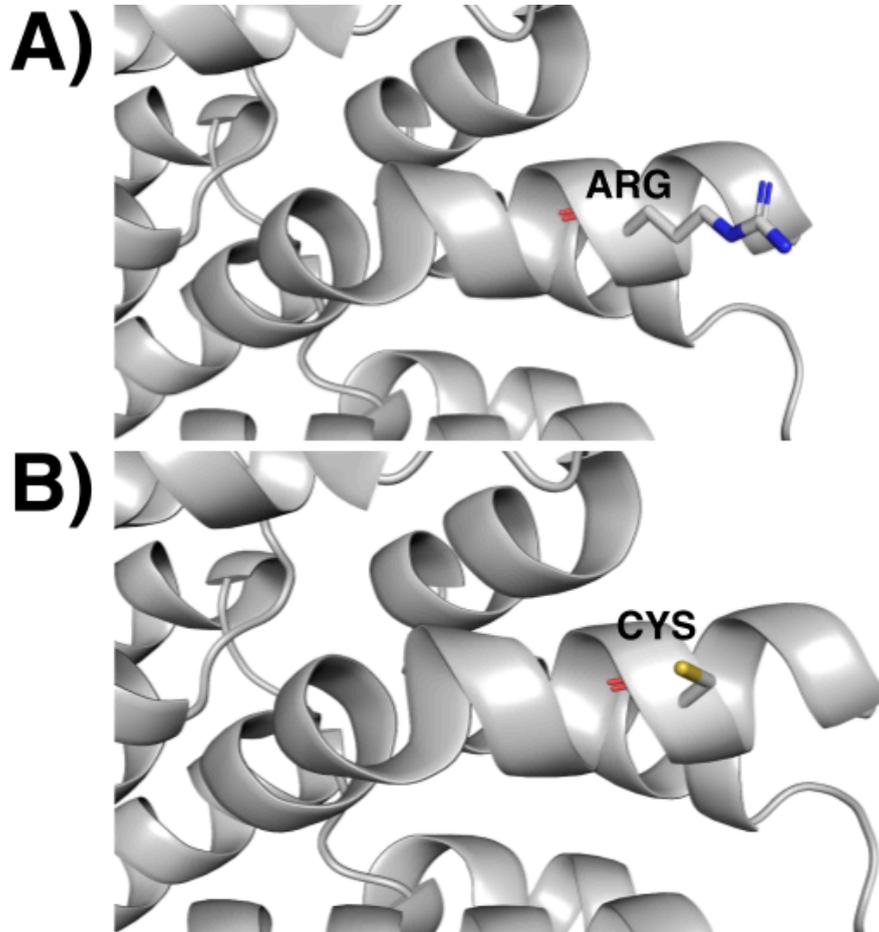
| Algoritmo | Mudança de estabilidade predita ( $\Delta\Delta G$ ) em kcal/mol |
|-----------|--|
| DUET      | -0,289   |
| mCSM      | -0,217   |
| muPRO     | -0,506   |

#### 5.5 ANÁLISE ESTRUTURAL

Além disso, quando visualizada tridimensionalmente (Figura 13), essa substituição revela uma mudança importante no tamanho do grupo R lateral na posição P4, indo de um com volume alto (arginina) para um grupo pequeno. Além disso, ao analisar a hidrofobicidade (Figura 14), é perceptível uma mudança pronunciada, dado que a cisteína é um aminoácido com maior hidrofobicidade (Eisenberg *et al.*, 1984). Por último, essa mudança também é pronunciada no

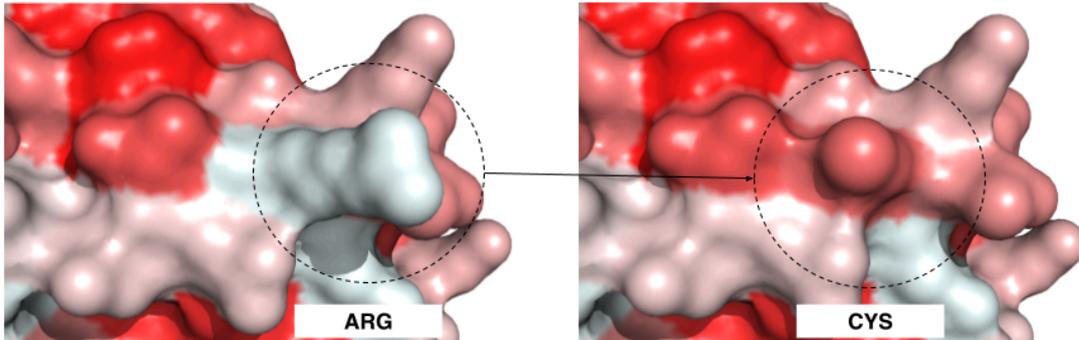
sentido de cargas eletrostáticas (Figura 15), havendo uma mudança do grupo guanidina carregado positivamente para o grupo tiol não carregado.

**Figura 13** - Visualização tridimensional do sítio de clivagem em RIG-I com **A)** o resíduo de arginina na posição P4 e **B)** com o resíduo de cisteína.



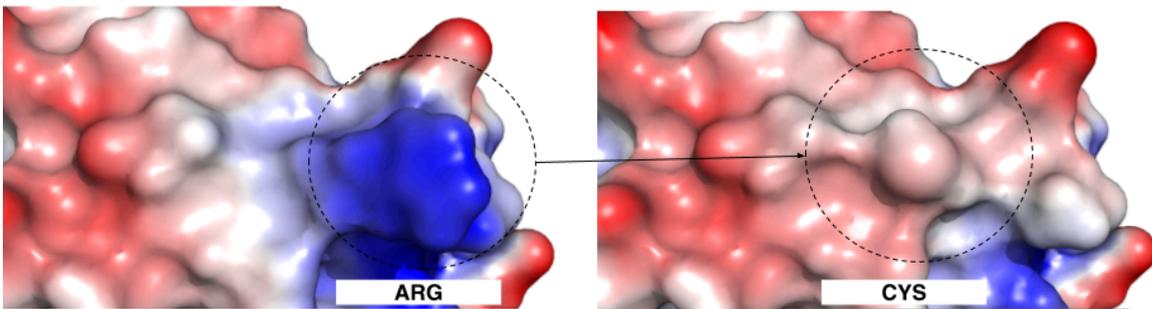
Fonte: Autor, 2024. Figura feita no PyMol.

**Figura 14** - Visualização tridimensional do perfil de hidrofobicidade resíduo P4 com o resíduo de arginina e o de cisteína.



Fonte: Autor, 2024. Figura feita no PyMol.

**Figura 15** - Visualização tridimensional do perfil eletrostático resíduo P4 com o resíduo de arginina e o de cisteína.



Fonte: Autor, 2024. Figura feita no PyMol.

## 6 DISCUSSÃO

O vírus SARS-CoV-2 causou a maior emergência em saúde pública já vista no século XXI até agora. Dentre as várias proteínas importantes para seu ciclo viral, a proteína não-estrutural 5, ou nsp5, apresenta-se como uma protease essencial pois participa da maturação de outras proteínas. Ademais, existem evidências que essa importância também é maior pelo fato da nsp5 não clivar somente as poliproteína virais, mas também proteínas do hospedeiro, que, nesse caso, é o ser humano. Esse mecanismo já é conhecido em outros vírus (Lei e Hilgenfeld, 2017; Yu *et al.*, 2015) e a evolução molecular nessas interfaces de interação entre protease virais e alvos humanos também já foi evidenciada (Patel *et al.*, 2012).

Sendo assim, de maneira geral, com efeito, a coleta de dados na literatura resultou num total de 111 sítios de clivagem e 242 variantes referentes a esses sítios, o que corrobora a existência dessa interação vírus-hospedeiro e a variação existente nessa interface.

Os resultados da caracterização físico-química dos sítios de clivagem corroboram a especificidade quase absoluta pela glutamina em P1 da nsp5 de SARS-CoV-2, a qual já foi comprovada em um recente estudo que determinou a estrutura cristalográfica da protease (Lee *et al.*, 2022). Da mesma forma, no mesmo estudo, foi observada uma maior prevalência da leucina em P2, o que permite a formação de ligações de van der Waals e interações hidrofóbicas e, para a posição P4, a predominância de resíduos pequenos que conseguem de encaixar na bolsa de ligação S4 (profunda e hidrofóbica) da protease. Por fim, é importante destacar que, em relação às proteases de outros vírus como SARS-CoV e outros coronavírus, esses mesmo perfil de preferência já foi comprovado experimentalmente, com destaque para o fato de que também foi observada a possibilidade de clivagem quando na presença de um resíduo de histidina em P1 (o que foi visto neste trabalho) e a preferência por resíduos pequenos em P1' (Chuck *et al.*, 2011; Chuck *et al.*, 2010; Goetz *et al.*, 2007). Portanto, a caracterização físico-química indica que a nsp5 de SARS-CoV-2 possui especificidades de ligação no sítio de clivagem similar a de outros coronavírus, o que é relevante para o entendimento das pressões evolutivas sob os alvos da nsp5, uma vez que a espécie humana já foi ou é exposta a outros coronavírus.

As preferências tridimensionais dos sítios observadas aqui também se

mostram similares com as de outras proteases (Barkan *et al.*, 2010). Além disso, é interessante notar que regiões desordenadas são mais propensas a estar sob seleção positiva do que regiões ordenadas e, assim, são locais onde a inovação genética pode ocorrer (Ridout *et al.*, 2010; Afanasyeva *et al.*, 2018). Essa evolução mais rápida pode ser explicada pela falta de restrição tridimensional nessas regiões. Coletivamente, esses fatos são interessantes porque mostram que, ao menos em parte, a nsp5 tem como alvo regiões que podem evoluir mais rápido, o que sugere que essas interfaces de interação protease-sítios são alvos interessantes para o estudo da adaptação do hospedeiro ao vírus e como esta impacta o curso da infecção. Finalmente, dado o fato da existência da circulação de outros coronavírus na população humana, pode-se hipotetizar também que estes e suas respectivas proteases já podem ter deixado marcas evolutivas nessas interfaces de interação.

No que tange a variação genética nos sítios de clivagem, o fato das 242 variantes *missense* se mostrarem no geral como “deletérias” e com frequência alélica rara, coloca-as como potenciais marcadores genéticos para o estudo da genética da susceptibilidade/resistência ao SARS-CoV-2.

Além disso, o algoritmo NetCorona mostrou que a substituição consequente da mutação mais frequente achada neste estudo, ao menos *in silico*, resulta numa maior possibilidade de clivagem pela nsp5, apesar de ser desestabilizante segundo os três algoritmos utilizados no trabalho. Esses resultados podem ser explicados pelo fato da cisteína ser um resíduo consideravelmente menor e mais hidrofóbico do que a arginina e isso ser uma preferência para resíduo na posição P4 dos sítios de clivagem da nsp5. Porém, em relação a RIG-I, essa substituição ocorre numa alfa-hélice e a cisteína demonstra ser um resíduo com baixa propensão de ocorrência em alfa-hélices, podendo ser um fator desestabilizador para essa estrutura secundária (Blaber e Zhang, 1993; Pace e Scholtz, 1998).

Cabe salientar também que essa variante está localizada no gene *DDX58*, que codifica RIG-I. A substituição ocorre num dos dois domínios CARD da proteína. Esses domínios são responsáveis pela interação com a proteína MAVS e, com isso, pela ativação de respostas *downstream*, incluindo a produção de interferons tipo I e outras proteínas antivirais (Kolakofsky; Kowalinki; Cusack, 2012) Na literatura, essa variante genética já foi associada ao *clearance* do vírus HCV (Wu *et al.*, 2019), maior risco de infecção por HCV após transplante de fígado (Biggings *et al.*, 2013) e a maiores níveis de expressão do próprio gene *DDX58* e também do gene *IFNB* (Hu *et*

*al.*, 2010).

É importante destacar também que RIG-I é comprovadamente um sensor do RNA viral de SARS-CoV-2, acarretando uma resposta inflamatória (Liu *et al.*, 2021; Thorne *et al.*, 2021). Foi também comprovado experimentalmente que o fragmento resultante da clivagem de RIG-I pela protease diminui essa ativação da imunidade inata. Esse fragmento conserva a capacidade de se ligar ao dsRNA, mas é incapaz de interagir e ativar MAVS (Liu *et al.*, 2021). Além disso, o gene que codifica RIG-I apresenta evidências de códons selecionados positivamente em mamíferos (Cagliani *et al.*, 2014; Matos; McFadden; Esteves, 2013), mas uma maior restrição evolucionária em humanos (Vasseur *et al.*, 2011).

Dessa maneira, esses fatos demonstram que essa variante hipoteticamente favorece o vírus no conflito entre vírus e hospedeiro. De fato, já foi proposto um modelo onde essas evasões da imunidade inata (aqui incluída a clivagem de proteínas) causam um atraso na supressão da replicação viral e que esse desbalanço na resposta contribui para o desenvolvimento da COVID-19 (Thorne *et al.*, 2021). Sendo assim, a variante aqui descrita se apresenta como uma candidata promissora para o estudo do impacto da variação genética na interação entre o SARS-CoV-2 com a imunidade inata e também na fisiopatologia da COVID-19.

## 7 CONCLUSÃO

O presente trabalho mostrou com os resultados que há variação nas sequências dos sítios de clivagem humanos da protease nsp5 de SARS-CoV-2. Essa variação mostrou-se no geral rara em relação às frequências alélicas e deletéria quando se considera o perfil de clivagem da protease, também delineado na pesquisa. Além disso, o achado da única variante comum ocorrer num gene envolvido no reconhecimento de RNAs virais revela que esse processo biológico é um alvo das armas virais e por isso pode apresentar variações genéticas. Essas variações podem ser importantes para o entendimento da fisiopatologia da COVID-19 e as diferentes respostas possíveis a ela. Além disso, este estudo mostra como análises detalhadas *in silico* desses conflitos moleculares são importantes e podem servir como primeiro passo para a realização de testes de associação e funcionais.

## REFERÊNCIAS

AFANASYEVA, Arina; BOCKWOLDT, Mathias; COONEY, Christopher R.; HEILAND, Ines; GOSSMANN, Toni I. Human long intrinsically disordered protein regions are frequent targets of positive selection. **Genome Research**, [S.L.], v. 28, n. 7, p. 975-982, 1 jun. 2018. Cold Spring Harbor Laboratory. <http://dx.doi.org/10.1101/gr.232645.117>. Disponível em: <https://genome.cshlp.org/content/28/7/975.full.pdf+html>. Acesso em: 4 mar. 2024.

ANGGAKUSUMA; BROWN, Richard J. P.; BANDA, Dominic H.; TODT, Daniel; VIEYRES, Gabrielle; STEINMANN, Eike; PIETSCHMANN, Thomas. Hepacivirus NS3/4A Proteases Interfere with MAVS Signaling in both Their Cognate Animal Hosts and Humans: implications for zoonotic transmission. **Journal Of Virology**, [S.L.], v. 90, n. 23, p. 10670-10681, dez. 2016. American Society for Microbiology. <http://dx.doi.org/10.1128/jvi.01634-16>. Disponível em: <https://journals.asm.org/doi/10.1128/jvi.01634-16>. Acesso em: 4 mar. 2024.

BARKAN, David T.; HOSTETTER, Daniel R.; MAHRUS, Sami; PIEPER, Ursula; WELLS, James A.; CRAIK, Charles S.; SALI, Andrej. Prediction of protease substrates using sequence and structure features. **Bioinformatics**, [S.L.], v. 26, n. 14, p. 1714-1722, 26 maio 2010. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btq267>. Disponível em: <https://academic.oup.com/bioinformatics/article/26/14/1714/178050>. Acesso em: 4 mar. 2024.

BERGSTRÖM, Anders; MCCARTHY, Shane A.; HUI, Ruoyun; ALMARRI, Mohamed A.; AYUB, Qasim; DANECEK, Petr; CHEN, Yuan; FELKEL, Sabine; HALLAST, Pille; KAMM, Jack. Insights into human genetic variation and population history from 929 diverse genomes. **Science**, [S.L.], v. 367, n. 6484, p. 0, 20 mar. 2020. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.aay5012>. Disponível em: <https://www.science.org/doi/10.1126/science.aay5012>. Acesso em: 6 mar. 2024.

BIGGINS, Scott W.; TROTTER, James; GRALLA, Jane; BURTON, James R.; BAMBHA, Kiran M.; DODGE, Jennifer; BROCATO, Megan; CHENG, Linling; MCQUEEN, Matt; FORMAN, Lisa. Differential effects of donor and recipient IL28B and DDX58 SNPs on severity of HCV after liver transplantation. **Journal Of Hepatology**, [S.L.], v. 58, n. 5, p. 969-976, maio 2013. Elsevier BV. <http://dx.doi.org/10.1016/j.jhep.2012.12.027>. Disponível em: [https://www.journal-of-hepatology.eu/article/S0168-8278\(13\)00015-9/abstract](https://www.journal-of-hepatology.eu/article/S0168-8278(13)00015-9/abstract). Acesso em: 4 mar. 2024.

BLABER, Michael; ZHANG, Xue-Jun; MATTHEWS, Brian W.. Structural Basis of Amino Acid  $\alpha$  Helix Propensity. **Science**, [S.L.], v. 260, n. 5114, p. 1637-1640, 11 jun. 1993. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.8503008>. Disponível em: <https://www.science.org/doi/10.1126/science.8503008>. Acesso em: 5 mar. 2024.

CAGLIANI, Rachele; FORNI, Diego; TRESOLDI, Claudia; POZZOLI, Uberto; FILIPPI, Giulia; RAINONE, Veronica; GIOIA, Luca de; CLERICI, Mario; SIRONI, Manuela.

RIG-I-Like Receptors Evolved Adaptively in Mammals, with Parallel Evolution at LGP2 and RIG-I. **Journal Of Molecular Biology**, [S.L.], v. 426, n. 6, p. 1351-1365, mar. 2014. Elsevier BV. <http://dx.doi.org/10.1016/j.jmb.2013.10.040>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0022283613006980?via%3Dihub#bb0305>. Acesso em: 4 mar. 2024.

CARROLL, Lewis. **Alice's Adventures in Wonderland & Trough the Looking-Glass**. 1. ed. Nova Iorque: Bantam Dell, 2006. 236p.

CHENG, Jianlin; RANDALL, Arlo; BALDI, Pierre. Prediction of protein stability changes for single-site mutations using support vector machines. **Proteins: Structure, Function, and Bioinformatics**, [S.L.], v. 62, n. 4, p. 1125-1132, 21 dez. 2005. Wiley. <http://dx.doi.org/10.1002/prot.20810>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/prot.20810>. Acesso em: 14 dez. 2023.

CHUCK, Chi-Pang; CHONG, Lin-Tat; CHEN, Chao; CHOW, Hak-Fun; WAN, David Chi-Cheong; WONG, Kam-Bo. Profiling of Substrate Specificity of SARS-CoV 3CLpro. **Plos One**, [S.L.], v. 5, n. 10, e13197, 6 out. 2010. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0013197>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0013197>. Acesso em: 1 mar. 2024.

CHUCK, Chi-Pang; CHOW, Hak-Fun; WAN, David Chi-Cheong; WONG, Kam-Bo. Profiling of Substrate Specificities of 3C-Like Proteases from Group 1, 2a, 2b, and 3 Coronaviruses. **Plos One**, [S.L.], v. 6, n. 11, e27228, 2 nov. 2011. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0027228>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0027228>. Acesso em: 1 mar. 2024.

CROOKS, Gavin E.; HON, Gary; CHANDONIA, John-Marc; BRENNER, Steven E.. WebLogo: a sequence logo generator. **Genome Research**, [S.L.], v. 14, n. 6, p. 1188-1190, jun. 2004. Cold Spring Harbor Laboratory. <http://dx.doi.org/10.1101/gr.849004>. Disponível em: <https://genome.cshlp.org/content/14/6/1188>. Acesso em: 14 dez. 2023.

DUSCHAK, Vilma; COUTO, Alicia. Cruzipain, the Major Cysteine Protease of Trypanosoma cruzi: a sulfated glycoprotein antigen as relevant candidate for vaccine development and drug target. a review. **Current Medicinal Chemistry**, [S.L.], v. 16, n. 24, p. 3174-3202, 1 ago. 2009. Bentham Science Publishers Ltd.. <http://dx.doi.org/10.2174/092986709788802971>. Disponível em: <https://www.eurekaselect.com/article/14598>. Acesso em: 5 mar. 2024.

EISENBERG, D.; SCHWARZ, E.; KOMAROMY, M.; WALL, R.. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. **Journal Of Molecular Biology**, [S.L.], v. 179, n. 1, p. 125-142, out. 1984. Elsevier BV. [http://dx.doi.org/10.1016/0022-2836\(84\)90309-7](http://dx.doi.org/10.1016/0022-2836(84)90309-7). Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/0022283684903097?via%3Dihub>. Acesso em: 6 mar. 2024.

GOETZ, D. H.; CHOE, Y.; HANSELL, E.; CHEN, Y. T.; MCDOWELL, M.; JONSSON, C. B.; ROUSH, W. R.; MCKERROW, J.; CRAIK, C. S.. Substrate Specificity Profiling

and Identification of a New Class of Inhibitor for the Major Protease of the SARS Coronavirus., **Biochemistry**, [S.L.], v. 46, n. 30, p. 8744-8752, 1 jul. 2007. American Chemical Society (ACS). <http://dx.doi.org/10.1021/bi0621415>. Disponível em: <https://pubs.acs.org/doi/10.1021/bi0621415>. Acesso em: 4 mar. 2024.

GRAKOU, A; MCCOURT, D W; WYCHOWSKI, C; FEINSTONE, S M; RICE, C M. Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites. **Journal Of Virology**, [S.L.], v. 67, n. 5, p. 2832-2843, maio 1993. American Society for Microbiology. <http://dx.doi.org/10.1128/jvi.67.5.2832-2843.1993>. Disponível em: <https://journals.asm.org/doi/10.1128/jvi.67.5.2832-2843.1993>. Acesso em: 4 mar. 2024.

GUDMUNDSSON, Sanna; SINGER-BERK, Moriel; WATTS, Nicholas A.; PHU, William; GOODRICH, Julia K.; SOLOMONSON, Matthew; REHM, Heidi L.; MACARTHUR, Daniel G.; O'DONNELL-LURIA, Anne. Variant interpretation using population databases: lessons from gnomad. **Human Mutation**, [S.L.], v. 43, n. 8, p. 1012-1030, 16 dez. 2021. Hindawi Limited. <http://dx.doi.org/10.1002/humu.24309>. Disponível em: <https://onlinelibrary.wiley.com/doi/10.1002/humu.24309>. Acesso em: 14 dez. 2023.

HU, Jianzhong; NISTAL-VILLÁN, Estanislao; VOHO, Anu; GANEE, Arnold; KUMAR, Madhu; DING, Yaomei; GARCÍA-SASTRE, Adolfo; WETMUR, James G.. A Common Polymorphism in the Caspase Recruitment Domain of RIG-I Modifies the Innate Immune Response of Human Dendritic Cells. **The Journal Of Immunology**, [S.L.], v. 185, n. 1, p. 424-432, 1 jul. 2010. The American Association of Immunologists. <http://dx.doi.org/10.4049/jimmunol.0903291>. Disponível em: <https://journals.aai.org/jimmunol/article/185/1/424/82939/A-Common-Polymorphism-in-the-Caspase-Recruitment>. Acesso em: 4 mar. 2024.

JIN, Zhenming; DU, Xiaoyu; XU, Yechun; DENG, Yongqiang; LIU, Meiqin; ZHAO, Yao; ZHANG, Bing; LI, Xiaofeng; ZHANG, Leike; PENG, Chao. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. **Nature**, [S.L.], v. 582, n. 7811, p. 289-293, 9 abr. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41586-020-2223-y>. Disponível em: <https://www.nature.com/articles/s41586-020-2223-y>. Acesso em: 14 dez. 2023.

JUMPER, John; EVANS, Richard; PRITZEL, Alexander; GREEN, Tim; FIGURNOV, Michael; RONNEBERGER, Olaf; TUNYASUVUNAKOOL, Kathryn; BATES, Russ; ŠÍDEK, Augustin; POTAPENKO, Anna. Highly accurate protein structure prediction with AlphaFold. **Nature**, [S.L.], v. 596, n. 7873, p. 583-589, 15 jul. 2021. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41586-021-03819-2>. Disponível em: <https://www.nature.com/articles/s41586-021-03819-2>. Acesso em: 14 dez. 2023.

KIEMER, Lars; LUND, Ole; BRUNAK, Søren; BLOM, Nikolaj. Coronavirus 3CLproproteinase cleavage sites:: possible relevance to sars virus pathology. **Bmc Bioinformatics**, [S.L.], v. 5, n. 1, p. 72, 2004. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1471-2105-5-72>. Disponível em: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-5-72>. Acesso em: 14 dez. 2023.

KOLAKOFSKY, Daniel; KOWALINSKI, Eva; CUSACK, Stephen. A structure-based model of RIG-I activation. **Rna**, [S.L.], v. 18, n. 12, p. 2118-2127, 1 nov. 2012. Cold Spring Harbor Laboratory. <http://dx.doi.org/10.1261/rna.035949.112>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3504664/>. Acesso em: 7 mar. 2024.

KOUDELKA, Tomas; BOGER, Juliane; HENKEL, Alessandra; SCHÖNHERR, Robert; KRANTZ, Stefanie; FUCHS, Sabine; RODRÍGUEZ, Estefanía; REDECKE, Lars; THOLEY, Andreas. N-Terminomics for the Identification of In Vitro Substrates and Cleavage Site Specificity of the SARS-CoV-2 Main Protease. **Proteomics**, [S.L.], v. 21, n. 2, n. p., 17 nov. 2020. Wiley. <http://dx.doi.org/10.1002/pmic.202000246>. Disponível em: <https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/pmic.202000246>. Acesso em: 6 mar. 2024.

KULSUPTRAKUL, Jessie; A TURCOTTE, Elizabeth; EMERMAN, Michael; MITCHELL, Patrick s. A human-specific motif facilitates CARD8 inflammasome activation after HIV-1 infection. **Elife**, [S.L.], v. 12, n. p., 7 jul. 2023. ELife Sciences Publications, Ltd. <http://dx.doi.org/10.7554/elife.84108>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10359095/>. Acesso em: 7 mar. 2024.

KYTE, Jack; DOOLITTLE, Russell F.. A simple method for displaying the hydropathic character of a protein. **Journal Of Molecular Biology**, [S.L.], v. 157, n. 1, p. 105-132, maio 1982. Elsevier BV. [http://dx.doi.org/10.1016/0022-2836\(82\)90515-0](http://dx.doi.org/10.1016/0022-2836(82)90515-0). Disponível em: [https://linkinghub.elsevier.com/retrieve/pii/0022-2836\(82\)90515-0](https://linkinghub.elsevier.com/retrieve/pii/0022-2836(82)90515-0). Acesso em: 14 dez. 2023.

LEE, Jaeyong; KENWARD, Calem; WORRALL, Liam J.; VUCKOVIC, Marija; GENTILE, Francesco; TON, Anh-Tien; NG, Myles; CHERKASOV, Artem; STRYNADKA, Natalie C. J.; PAETZEL, Mark. X-ray crystallographic characterization of the SARS-CoV-2 main protease polyprotein cleavage sites essential for viral processing and maturation. **Nature Communications**, [S.L.], v. 13, n. 1, p. 0, 3 set. 2022. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41467-022-32854-4>. Disponível em: <https://www.nature.com/articles/s41467-022-32854-4>. Acesso em: 7 mar. 2024.

LEE, Jaeyong; WORRALL, Liam J.; VUCKOVIC, Marija; ROSELL, Federico I.; GENTILE, Francesco; TON, Anh-Tien; CAVENEY, Nathanael A.; BAN, Fuqiang; CHERKASOV, Artem; PAETZEL, Mark. Crystallographic structure of wild-type SARS-CoV-2 main protease acyl-enzyme intermediate with physiological C-terminal autoprocessing site. **Nature Communications**, [S.L.], v. 11, n. 1, n. p., 18 nov. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41467-020-19662-4>. Disponível em: <https://www.nature.com/articles/s41467-020-19662-4>. Acesso em: 7 mar. 2024.

LEI, Jian; HILGENFELD, Rolf. RNA-virus proteases counteracting host innate immunity. **Febs Letters**, [S.L.], v. 591, n. 20, p. 3190-3210, 15 set. 2017. Wiley. <http://dx.doi.org/10.1002/1873-3468.12827>. Disponível em: <https://febs.onlinelibrary.wiley.com/doi/10.1002/1873-3468.12827>. Acesso em: 4 mar. 2024.

LI, Qingxin; KANG, Congbao. Structure and Dynamics of Zika Virus Protease and Its Insights into Inhibitor Design. **Biomedicines**, [S.L.], v. 9, n. 8, p. 1044-0, 19 ago. 2021. MDPI AG. <http://dx.doi.org/10.3390/biomedicines9081044>. Disponível em: <https://www.mdpi.com/2227-9059/9/8/1044>. Acesso em: 4 mar. 2024.

LIU, Yongzhen; QIN, Chao; RAO, Youliang; NGO, Chau; FENG, Joshua J.; ZHAO, Jun; ZHANG, Shu; WANG, Ting-Yu; CARRIERE, Jessica; SAVAS, Ali Can. SARS-CoV-2 Nsp5 Demonstrates Two Distinct Mechanisms Targeting RIG-I and MAVS To Evade the Innate Immune Response. **Mbio**, [S.L.], v. 12, n. 5, n.p., 26 out. 2021. American Society for Microbiology. <http://dx.doi.org/10.1128/mbio.02335-21>. Disponível em: <https://journals.asm.org/doi/10.1128/mbio.02335-21>. Acesso em: 14 dez. 2023.

MATOS, Ana Lemos de; MCFADDEN, Grant; ESTEVES, Pedro J.. Positive Evolutionary Selection On the RIG-I-Like Receptor Genes in Mammals. **Plos One**, [S.L.], v. 8, n. 11, e81864, 27 nov. 2013. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0081864>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0081864>. Acesso em: 4 mar. 2024.

MICZI, Márió; GOLDA, Mária; KUNKLI, Balázs; NAGY, Tibor; TÖZSÉR, József; MÓTYÁN, János András. Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease. **International Journal Of Molecular Sciences**, [S.L.], v. 21, n. 24, p. 9523, 15 dez. 2020. MDPI AG. <http://dx.doi.org/10.3390/ijms21249523>. Disponível em: <https://www.mdpi.com/1422-0067/21/24/9523>. Acesso em: 6 mar. 2024.

MODROW, Susanne; FALKE, Dietrich; TRUYEN, Uwe; SCHÄTZL, Hermann. Viruses with Single-Stranded, Positive-Sense RNA Genomes. **Molecular Virology**, [S.L.], p. 185-349, [2013]. Springer Berlin Heidelberg. [http://dx.doi.org/10.1007/978-3-642-20718-1\\_14](http://dx.doi.org/10.1007/978-3-642-20718-1_14). Disponível em: [https://link.springer.com/referenceworkentry/10.1007/978-3-642-20718-1\\_14](https://link.springer.com/referenceworkentry/10.1007/978-3-642-20718-1_14). Acesso em: 4 mar. 2024.

MORALES, Joannella; PUJAR, Shashikant; LOVELAND, Jane E.; ASTASHYN, Alex; BENNETT, Ruth; BERRY, Andrew; COX, Eric; DAVIDSON, Claire; ERMOLAEVA, Olga; FARRELL, Catherine M.. A joint NCBI and EMBL-EBI transcript set for clinical genomics and research. **Nature**, [S.L.], v. 604, n. 7905, p. 310-315, 6 abr. 2022. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41586-022-04558-8>. Disponível em: <https://www.nature.com/articles/s41586-022-04558-8>. Acesso em: 5 mar. 2024.

MOUSTAQIL, Mehdi; OLLIVIER, Emma; CHIU, Hsin-Ping; VAN TOL, Sarah; RUDOLFFI-SOTO, Paulina; STEVENS, Christian; BHUMKAR, Akshay; HUNTER, Dominic J. B.; FREIBERG, Alexander N.; JACQUES, David. SARS-CoV-2 proteases PL<sub>pro</sub> and 3CL<sub>pro</sub> cleave IRF3 and critical modulators of inflammatory pathways (NLRP12 and TAB1): implications for disease presentation across species. **Emerging Microbes & Infections**, [S.L.], v. 10, n. 1, p. 178-195, 1 jan. 2021. Informa UK Limited. <http://dx.doi.org/10.1080/22221751.2020.1870414>. Disponível em: <https://www.tandfonline.com/doi/full/10.1080/22221751.2020.1870414>. Acesso em: 6 mar. 2024.

NAJJAR-DEBBINY, Ronza; GRONICH, Naomi; WEBER, Gabriel; KHOURY, Johad; AMAR, Maisam; STEIN, Nili; GOLDSTEIN, Lee Hilary; SALIBA, Walid. Effectiveness of Paxlovid in Reducing Severe Coronavirus Disease 2019 and Mortality in High-Risk Patients. **Clinical Infectious Diseases**, [S.L.], v. 76, n. 3, p. 342-349, 6 jul. 2022. Oxford University Press (OUP). <http://dx.doi.org/10.1093/cid/ciac443>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9214014/>. Acesso em: 6 mar. 2024.

NARWAL, Manju; SINGH, Harvijay; PRATAP, Shivendra; MALIK, Anjali; KUHN, Richard J.; KUMAR, Pravindra; TOMAR, Shailly. Crystal structure of chikungunya virus nsP2 cysteine protease reveals a putative flexible loop blocking its active site. **International Journal Of Biological Macromolecules**, [S.L.], v. 116, p. 451-462, set. 2018. Elsevier BV. <http://dx.doi.org/10.1016/j.ijbiomac.2018.05.007>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0141813018300114?via%3Dihub>. Acesso em: 4 mar. 2024.

NASLAVSKY, Michel S.; SCLIAR, Marilia O.; YAMAMOTO, Guilherme L.; WANG, Jaqueline Yu Ting; ZVERINOVA, Stepanka; KARP, Tatiana; NUNES, Kelly; CERONI, José Ricardo Magliocco; CARVALHO, Diego Lima de; SIMÕES, Carlos Eduardo da Silva. Whole-genome sequencing of 1,171 elderly admixed individuals from Brazil. **Nature Communications**, [S.L.], v. 13, n. 1, p. 0, 4 mar. 2022. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41467-022-28648-3>. Disponível em: <https://www.nature.com/articles/s41467-022-28648-3>. Acesso em: 6 mar. 2024.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **WHO Coronavirus (COVID-19) Dashboard**, 2023. Disponível em: <https://covid19.who.int/>. Acesso em: 7 mar. 2024.

PABLOS, Isabel; MACHADO, Yoan; JESUS, Hugo C. Ramos de; MOHAMUD, Yasir; KAPPELHOFF, Reinhild; LINDSKOG, Cecilia; VLOK, Marli; BELL, Peter A.; BUTLER, Georgina S.; GRIN, Peter M.. Mechanistic insights into COVID-19 by global analysis of the SARS-CoV-2 3CLpro substrate degradome. **Cell Reports**, [S.L.], v. 37, n. 4, p. 109892, out. 2021. Elsevier BV. <http://dx.doi.org/10.1016/j.celrep.2021.109892>. Disponível em: [https://www.cell.com/cell-reports/fulltext/S2211-1247\(21\)01362-0](https://www.cell.com/cell-reports/fulltext/S2211-1247(21)01362-0). Acesso em: 6 mar. 2024.

PACE, C. Nick; SCHOLTZ, J. Martin. A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins. **Biophysical Journal**, [S.L.], v. 75, n. 1, p. 422-427, jul. 1998. Elsevier BV. [http://dx.doi.org/10.1016/s0006-3495\(98\)77529-0](http://dx.doi.org/10.1016/s0006-3495(98)77529-0). Disponível em: <https://www.sciencedirect.com/science/article/pii/S0006349598775290>. Acesso em: 5 mar. 2024.

PATEL, Maulik R.; LOO, Yueh-Ming; HORNER, Stacy M.; GALE, Michael; MALIK, Harmit S.. Convergent Evolution of Escape from Hepaciviral Antagonism in Primates. **Plos Biology**, [S.L.], v. 10, n. 3, e1001282, 13 mar. 2012. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pbio.1001282>. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001282>. Acesso em: 5 mar. 2024.

PIRES, Douglas E. V.; ASCHER, David B.; BLUNDELL, Tom L.. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. **Nucleic Acids Research**, [S./], v. 42, n. 1, p. 314-319, 14 maio 2014. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gku411>. Disponível em: <https://academic.oup.com/nar/article/42/W1/W314/2437397>. Acesso em: 14 dez. 2023.

PIRES, Douglas E. V.; ASCHER, David B.; BLUNDELL, Tom L.. MCSM: predicting the effects of mutations in proteins using graph-based signatures. **Bioinformatics**, [S./], v. 30, n. 3, p. 335-342, 26 nov. 2013. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btt691>. Disponível em: <https://academic.oup.com/bioinformatics/article/30/3/335/228906>. Acesso em: 14 dez. 2023.

PLANÈS, Rémi; PINILLA, Miriam; SANTONI, Karin; HESSEL, Audrey; PASSEMAR, Charlotte; LAY, Kenneth; PAILLETTE, Perrine; VALADÃO, Ana-Luiza Chaves; ROBINSON, Kim Samirah; BASTARD, Paul. Human NLRP1 is a sensor of pathogenic coronavirus 3CL proteases in lung epithelial cells. **Molecular Cell**, [S./], v. 82, n. 13, p. 2385-2400, jul. 2022. Elsevier BV. <http://dx.doi.org/10.1016/j.molcel.2022.04.033>. Disponível em: [https://www.cell.com/molecular-cell/fulltext/S1097-2765\(22\)00433-6?\\_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1097276522004336%3Fshowall%3Dtrue](https://www.cell.com/molecular-cell/fulltext/S1097-2765(22)00433-6?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS1097276522004336%3Fshowall%3Dtrue). Acesso em: 14 dez. 2023.

POSITION WEIGHT MATRIX. In: WIKIPÉDIA: a enciclopédia livre. [São Francisco, CA: Fundação Wikimedia], 2024. Disponível em: [https://en.wikipedia.org/wiki/Position\\_weight\\_matrix](https://en.wikipedia.org/wiki/Position_weight_matrix). Acesso em: 6 mar. 2024.

REHWINKEL, Jan; GACK, Michaela U.. RIG-I-like receptors: their regulation and roles in rna sensing. **Nature Reviews Immunology**, [S.L.], v. 20, n. 9, p. 537-551, 13 mar. 2020. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/s41577-020-0288-3>. Disponível em: <https://www.nature.com/articles/s41577-020-0288-3>. Acesso em: 6 mar. 2024.

RIDOUT, Kate E.; DIXON, Christopher J.; FILATOV, Dmitry A.. Positive Selection Differs between Protein Secondary Structure Elements in Drosophila. **Genome Biology And Evolution**, [S.L.], v. 2, p. 166-179, 1 jan. 2010. Oxford University Press (OUP). <http://dx.doi.org/10.1093/gbe/evq008>. Disponível em: <https://academic.oup.com/gbe/article/doi/10.1093/gbe/evq008/569192>. Acesso em: 4 mar. 2024.

ROBINSON, Kim S.; TEO, Daniel Eng Thiam; TAN, Kai Sen; TOH, Gee Ann; ONG, Hsiao Hui; LIM, Chrissie Kaishi; LAY, Kenneth; AU, Bijin Veonice; LEW, Tian Sheng; CHU, Justin Jang Hann. Enteroviral 3C protease activates the human NLRP1 inflammasome in airway epithelia. **Science**, [S.L.], v. 370, n. 6521, n. p., 4 dez. 2020. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.aay2002>. Disponível em: [https://www.science.org/doi/10.1126/science.aay2002?url\\_ver=Z39.88-2003&rfr\\_id=ori:rid:crossref.org&rfr\\_dat=cr\\_pub%20%20pubmed](https://www.science.org/doi/10.1126/science.aay2002?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%20%20pubmed). Acesso em: 5 mar. 2024.

ROTHENBURG, Stefan; BRENNAN, Greg. Species-Specific Host–Virus Interactions: implications for viral host range and virulence. **Trends In Microbiology**, [S.L.], v. 28, n. 1, p. 46-56, jan. 2020. Elsevier BV. <http://dx.doi.org/10.1016/j.tim.2019.08.007>. Disponível em: [https://www.cell.com/trends/microbiology/abstract/S0966-842X\(19\)30220-3?](https://www.cell.com/trends/microbiology/abstract/S0966-842X(19)30220-3?). Acesso em: 7 mar. 2024.

SCHECHTER, Israel; BERGER, Arieh. On the size of the active site in proteases. I. Papain. **Biochemical And Biophysical Research Communications**, [S.], v. 27, n. 2, p. 157-162, abr. 1967. Elsevier BV. [http://dx.doi.org/10.1016/s0006-291x\(67\)80055-x](http://dx.doi.org/10.1016/s0006-291x(67)80055-x). Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0006291X6780055X?via%3Dihub>. Acesso em: 14 dez. 2023.

SHI, Fushan; LV, Qian; WANG, Tingjun; XU, Jidong; XU, Wei; SHI, Yuhua; FU, Xinyu; YANG, Tianming; YANG, Yang; ZHUANG, Lenan. Coronaviruses Nsp5 Antagonizes Porcine Gasdermin D-Mediated Pyroptosis by Cleaving Pore-Forming p30 Fragment. **Mbio**, [S.L.], v. 13, n. 1, p. 0, 22 fev. 2022. American Society for Microbiology. <http://dx.doi.org/10.1128/mbio.02739-21>. Disponível em: <https://journals.asm.org/doi/10.1128/mbio.02739-21>. Acesso em: 6 mar. 2024.

SONG, Jiangning; TAN, Hao; PERRY, Andrew J.; AKUTSU, Tatsuya; WEBB, Geoffrey I.; WHISSTOCK, James C.; PIKE, Robert N.. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. **Plos One**, [S.L.], v. 7, n. 11, e50300, 29 nov. 2012. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0050300>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0050300>. Acesso em: 4 mar. 2024.

THORNE, Lucy G; REUSCHL, Ann-Kathrin; ZULIANI-ALVAREZ, Lorena; WHELAN, Matthew V X; TURNER, Jane; NOURSADEGHI, Mahdad; JOLLY, Clare; TOWERS, Greg J. SARS-CoV-2 sensing by RIG-I and MDA5 links epithelial infection to macrophage inflammation. **The Embo Journal**, [S.L.], v. 40, n. 15, n. p., 2 jul. 2021. Springer Science and Business Media LLC. <http://dx.doi.org/10.15252/emboj.2021107826>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8209947/>. Acesso em: 7 mar. 2024.

TIEN, Matthew Z.; MEYER, Austin G.; SYDYKOVA, Dariya K.; SPIELMAN, Stephanie J.; WILKE, Claus O.. Maximum Allowed Solvent Accessibilities of Residues in Proteins. **Plos One**, [S.L.], v. 8, n. 11, e80635, 21 nov. 2013. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pone.0080635>. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0080635>. Acesso em: 5 mar. 2024.

TIMMER, John C; ZHU, Wenhong; POP, Cristina; REGAN, Tim; SNIPAS, Scott J; EROSHKIN, Alexey M; RIEDL, Stefan J; SALVESEN, Guy s. Structural and kinetic determinants of protease substrates. **Nature Structural & Molecular Biology**, [S.L.], v. 16, n. 10, p. 1101-1108, 20 set. 2009. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/nsmb.1668>. Disponível em: <https://www.nature.com/articles/nsmb.1668>. Acesso em: 4 mar. 2024.

TSU, Brian V.; AGARWAL, Rimjhim; GOKHALE, Nandan S.; KULSUPTRAKUL, Jessie; RYAN, Andrew P.; FAY, Elizabeth J.; CASTRO, Lennice K.; BEIERSCHMITT, Christopher; YAP, Christina; TURCOTTE, Elizabeth A.. Host-specific sensing of coronaviruses and picornaviruses by the CARD8 inflammasome. **Plos Biology**, [S.L.], v. 21, n. 6, e3002144, 8 jun. 2023. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pbio.3002144>. Disponível em: <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002144>. Acesso em: 5 mar. 2024.

TSU, Brian V.; FAY, Elizabeth J.; NGUYEN, Katelyn T.; CORLEY, Miles R.; HOSURU, Bindhu; DOMINGUEZ, Viviana A.; DAUGHERTY, Matthew D.. Running With Scissors: evolutionary conflicts between viral proteases and the host immune system. **Frontiers In Immunology**, [S.L.], v. 12, n.p., 1 nov. 2021. Frontiers Media SA. <http://dx.doi.org/10.3389/fimmu.2021.769543>. Disponível em: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.769543/full>. Acesso em: 14 dez. 2023.

VAN VALEN, Leigh. A new evolutionary Law. **Evol. Theory**, [S.L.], v. 1, n.1, p. 1-30. BIBL. DISSEM. Disponível em: <https://www.mn.uio.no/cees/english/services/van-valen/evolutionary-theory/volume-1/vol-1-no-1-pages-1-30-l-van-valen-a-new-evolutionary-law.pdf>. Acesso em: 7 mar. 2024.

VASSEUR, Estelle; PATIN, Etienne; LAVAL, Guillaume; PAJON, Sandra; FORNARINO, Simona; CROUAU-ROY, Brigitte; QUINTANA-MURCI, Lluís. The selective footprints of viral pressures at the human RIG-I-like receptor family. **Human Molecular Genetics**, [S.L.], v. 20, n. 22, p. 4462-4474, 24 ago. 2011. Oxford University Press (OUP). <http://dx.doi.org/10.1093/hmg/ddr377>. Disponível em: <https://academic.oup.com/hmg/article/20/22/4462/639888?login=false>. Acesso em: 4 mar. 2024.

VEGA, Sonia; NEIRA, Jose; MARCUELLO, Carlos; LOSTAO, Anabel; ABIAN, Olga; VELAZQUEZ-CAMPOY, Adrian. NS3 Protease from Hepatitis C Virus: biophysical studies on an intrinsically disordered protein domain. **International Journal Of Molecular Sciences**, [S.L.], v. 14, n. 7, p. 13282-13306, 26 jun. 2013. MDPI AG. <http://dx.doi.org/10.3390/ijms140713282>. Disponível em: <https://www.mdpi.com/1422-0067/14/7/13282>. Acesso em: 5 mar. 2024.

WANG, Bei; XI, Xueyan; LEI, Xiaobo; ZHANG, Xiaoyan; CUI, Sheng; WANG, Jianwei; JIN, Qi; ZHAO, Zhendong. Enterovirus 71 Protease 2Apro Targets MAVS to Inhibit Anti-Viral Type I Interferon Responses. **Plos Pathogens**, [S.L.], v. 9, n. 3, e1003231, 21 mar. 2013. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.ppat.1003231>. Disponível em: <https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1003231>. Acesso em: 4 mar. 2024.

WANG, Qiankun; GAO, Hongbo; CLARK, Kolin M.; MUGISHA, Christian Shema; DAVIS, Keanu; TANG, Jack P.; HARLAN, Gray H.; DESELM, Carl J.; PRESTI, Rachel M.; KUTLUAY, Sebla B.. CARD8 is an inflammasome sensor for HIV-1 protease activity. **Science**, [S.L.], v. 371, n. 6535, p. 0-0, 19 mar. 2021. American Association for the Advancement of Science (AAAS).

<http://dx.doi.org/10.1126/science.abe1707>. Disponível em: <https://www.science.org/doi/10.1126/science.abe1707>. Acesso em: 4 mar. 2024.

WANG, Yong; LV, Zhengtong; CHU, Yuan. HIV protease inhibitors: a review of molecular selectivity and toxicity. **Hiv/Aids - Research And Palliative Care**, [S.L.], p. 95, abr. 2015. Informa UK Limited. <http://dx.doi.org/10.2147/hiv.s79956>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4396582/>. Acesso em: 6 mar. 2024.

WU, Xinyu; ZANG, Feng; LIU, Mei; ZHUO, Lingyun; WU, Jingjing; XIA, Xueshan; FENG, Yue; YU, Rongbin; HUANG, Peng; YANG, Sheng. Genetic variants in RIG-I-like receptor influences HCV clearance in Chinese Han population. **Epidemiology And Infection**, [S.L.], v. 147, n. p., [2019]. Cambridge University Press (CUP). <http://dx.doi.org/10.1017/s0950268819000827>. Disponível em: <https://www.cambridge.org/core/journals/epidemiology-and-infection/article/genetic-variants-in-rigilike-receptor-influences-hcv-clearance-in-chinese-han-population/432912EAD0B77214CEB861ABC8BF73BF>. Acesso em: 4 mar. 2024.

YU, Chia-Yi; LIANG, Jian-Jong; LI, Jin-Kun; LEE, Yi-Ling; CHANG, Bi-Lan; SU, Chan-I; HUANG, Wei-Jheng; LAI, Michael M. C.; LIN, Yi-Ling. Dengue Virus Impairs Mitochondrial Fusion by Cleaving Mitofusins. **Plos Pathogens**, [S.L.], v. 11, n. 12, e1005350, 30 dez. 2015. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.ppat.1005350>. Disponível em: <https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1005350>. Acesso em: 4 mar. 2024.

ZERBE, Cassie M.; MOUSER, David J.; COLE, James L.. Oligomerization of RIG-I and MDA5 2CARD domains. **Protein Science**, [S.L.], v. 29, n. 2, p. 521-526, 20 nov. 2019. Wiley. <http://dx.doi.org/10.1002/pro.3776>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6954692/>. Acesso em: 6 mar. 2024.

ZHANG, Buzhong; LI, Linqing; LÜ, Qiang. Protein Solvent-Accessibility Prediction by a Stacked Deep Bidirectional Recurrent Neural Network. **Biomolecules**, [S.L.], v. 8, n. 2, p. 33, 25 maio 2018. MDPI AG. <http://dx.doi.org/10.3390/biom8020033>. Disponível em: <https://www.mdpi.com/2218-273X/8/2/33>. Acesso em: 7 mar. 2024.

ZHANG, Jianzhi. Positive selection, not negative selection, in the pseudogenization of rcsA in *Yersinia pestis*. **Proceedings Of The National Academy Of Sciences**, [S.L.], v. 105, n. 42, p. 0-0, 21 out. 2008. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.0806419105>. Disponível em: <https://www.pnas.org/doi/full/10.1073/pnas.0806419105>. Acesso em: 5 mar. 2024.

ZHANG, Yabin; LIU, Shiyan; XU, Qingjia; LI, Huihui; LU, Kefeng. Cleavage of the selective autophagy receptor SQSTM1/p62 by the SARS-CoV-2 main protease NSP5 prevents the autophagic degradation of viral membrane proteins. **Molecular Biomedicine**, [S.L.], v. 3, n. 1, n. p., 3 jun. 2022. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/s43556-022-00083-2>. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9162485/>. Acesso em: 5 mar. 2024.