



**UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”**

**FACULDADE DE CIÊNCIAS FARMACÊUTICAS DE ARARAQUARA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOCÊNCIAS E
BIOTECNOLOGIA APLICADAS À FARMÁCIA
MESTRADO ACADÊMICO**

ANDERSON AMENDOLA PINHEIRO

**DETERMINAÇÃO, SÍNTESE E APLICAÇÃO DE PEPTÍDEOS IMUNOGÊNICOS DA
PROTEÍNA ESTRUTURAL S DO VÍRUS SARS-COV-2**

**ARARAQUARA, SP
JUNHO DE 2022**

ANDERSON AMENDOLA PINHEIRO

DETERMINAÇÃO, SÍNTESE E APLICAÇÃO DE PEPTÍDEOS IMUNOGÊNICOS DA
PROTEÍNA ESTRUTURAL S DO VÍRUS SARS-COV-2

Dissertação apresentada ao Programa de Pós-Graduação em Biociências e Biotecnologias Aplicadas à Farmácia, da Faculdade de Ciências Farmacêuticas da UNESP, como parte dos requisitos para a obtenção do título de Mestre em Ciências.

Área de Concentração: Imunologia

Orientador: Prof. Dr. Paulo Inácio da Costa

Coorientador: Prof. Dr. Gustavo Troiano Feliciano

ARARAQUARA, SP

JUNHO DE 2022

ESTE TRABALHO FOI FINANCEIRAMENTE APOIADO COM UMA BOLSA DE ESTUDOS DO CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq) DURANTE 24 MESES ATRAVÉS DO PROCESSO 131680/2020-0.

AUTORIZO A REPRODUÇÃO E/OU DIVULGAÇÃO, SEJA ELA TOTAL OU PARCIAL, DESTE DOCUMENTO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

P654d Pinheiro, Anderson Amendola.
Determinação, síntese e aplicação de peptídeos imunogênicos da proteína estrutural s do vírus SARS-COV-2 / Anderson Amendola Pinheiro. – Araraquara: [S.n.], 2022.
99 f. : il.

Dissertação (Mestrado) – Universidade Estadual Paulista. “Júlio de Mesquita Filho”. Faculdade de Ciências Farmacêuticas. Programa de Pós Graduação em Biociências e Biotecnologia Aplicadas à Farmácia. Área Imunologia.

Orientador: Paulo Inácio da Costa.
Coorientador: Gustavo Troiano Feliciano.

1. SARS-CoV-2. 2. COVID-19. 3. Glicoproteína Spike. 4. Epítomos. 5. Imunossensores. I. Costa, Paulo Inácio, orient. II. Feliciano, Gustavo Troiano, coorient. III. Título.

Diretoria do Serviço Técnico de Biblioteca e Documentação - Faculdade de Ciências Farmacêuticas
UNESP - Campus de Araraquara

CAPES: 33004030081P7
Esta ficha não pode ser modificada



UNIVERSIDADE ESTADUAL PAULISTA

Câmpus de Araraquara



CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: Determinação, síntese e aplicação de peptídeos imunogênicos da proteína estrutural S do vírus SARS-CoV-2

AUTOR: ANDERSON AMENDOLA PINHEIRO
ORIENTADOR: PAULO INACIO DA COSTA
COORIENTADOR: GUSTAVO TROIANO FELICIANO

Aprovado como parte das exigências para obtenção do Título de Mestre em BIOCÊNCIAS E BIOTECNOLOGIA APLICADAS À FARMÁCIA, área de conhecimento: Imunologia pela Comissão Examinadora:

Prof. Dr. PAULO INACIO DA COSTA (Participação Virtual)
Departamento de Análises Clínicas / Faculdade de Ciências Farmacêuticas - UNESP - Araraquara

Prof. Dr. ALEX JOSÉ LEITE TORRES (Participação Virtual)
Departamento de Biomorfologia / Universidade Federal da Bahia

Profa. Dra. PAULA RAHAL (Participação Virtual)
Departamento de Engenharia de Bioprocessos e Biotecnologia / Instituto de Biociências, Letras e Ciências Exatas de São José do Rio Preto - Unesp

Araraquara, 21 de junho de 2022

DEDICATÓRIA

Para Aline, Álvaro, Sueli e Ernesta.

Para você, caro colega cientista/acadêmico/curioso, que tirar proveito dos conhecimentos aqui estabelecidos.

AGRADECIMENTOS

Em primeiro lugar, aos meus pais, Álvaro e Sueli, por batalharem todos esses anos para que não me faltasse nada, em especial, a oportunidade de vir à Araraquara.

Aos meus avós, José e Josefa (*in memoriam*), Ernesta e Pedro (*in memoriam*) pela orientação, conselhos e criação.

À minha família, por todo suporte, apoio e momentos de descontração e reflexão.

À Aline, pelo amor, pelo carinho, pelo companheirismo essencial para que eu trilhasse esse caminho do começo até o fim.

Ao meu Orientador, Professor Paulo, pela oportunidade, amizade, pelos ensinamentos e orientação.

Ao meu Coorientador, Professor Gustavo, pela oportunidade, pela amizade, pelos ensinamentos e pela orientação. Ao Anderson, colaborador essencial em todos os aspectos relacionados aos experimentos de MD.

Aos amigos do “LICBM” Aline, Lorrane, Bruna, Rafa, Anna, Nicolý, Eli, Rualdinho, Helô e Felipinho, pelos momentos de descontração, pela parceria diária, pelos cafés, pelas risadas e conselhos indispensáveis nos bastidores da produção científica.

À UNESP, em nome da Faculdade de Ciências Farmacêuticas, pelo apoio institucional de infraestrutura e recursos humanos, em especial as meninas da Seção de Pós-Graduação (Claudinha, Ani e Denise), por todos os lembretes e pela ajuda em tudo que foi necessário. Ao professor André e ao Vinicius, por todo apoio, suporte e treinamento durante os experimentos com o cromatógrafo líquido. Ao Dr. Deivys, por todo suporte durante a implementação da síntese de peptídeos.

Ao CNPq pelo auxílio financeiro concedido e indispensável para realização do trabalho (Processo 131680/2020-0).

A big thanks to Alexandra Elbakyan for constantly removing barriers in the way of science.

Por fim, mas não menos importante, a Deus, que diariamente esteve presente, como guia dos meus passos.

*Though I can't believe it
Like master of the observation
Coming up with an explanation*

[...]

We can give you answers you are dying to receive

Keep on dreaming

Keep on screaming

[...]

Shine your light into the darkness

And let the storm descend upon you

Let the storm descend upon you

– Tobias Sammet's AVANTASIA

RESUMO

A Pandemia de COVID-19 já custou ao mundo incontáveis perdas humanas, sociais e econômicas. SARS-CoV-2, seu agente etiológico, é um *Betacoronavirus* envelopado com um genoma composto por RNA de fita simples e senso positivo que codifica quatro proteínas estruturais, sendo a glicoproteína *Spike* a mais importante devido a sua função de reconhecer o receptor ACE2 e ancorar o vírus na superfície da células hospedeira. Dado o *knowhow* de nosso grupo em desenvolver métodos diagnósticos inovadores para vírus de RNA, o presente trabalho propôs identificar e caracterizar na glicoproteína *Spike* alvos moleculares para serem aplicados em um sensor imunoeletroquímico. Para tal, um método combinatório baseado em inteligência artificial identificou 13 potenciais sítios imunodominantes espalhados por toda proteína S. Estes sítios variam de 12 a 22 aminoácidos de extensão. Através de experimentos de Dinâmica Molecular pudemos caracterizar tais epítomos e avaliar parâmetros bioquímicos e biofísicos importantes para o processo de interação antígeno-anticorpo. Foi constatado que tanto a proteína S quanto os nossos sítios apresentam boa estabilidade conformacional. Também pudemos avaliar a Área Superficial Acessível ao Solvente (SASA) em duas condições e determinar que, para os epítomos encontrados na região RBD, a mais relevante para o reconhecimento dos anticorpos, essa acessibilidade varia de acordo com a conformação adotada pelo RBD. Quando este encontra-se fechado, os epítomos estão mais inacessíveis, e quando ele está aberto, os epítomos são significativamente mais acessíveis. Por fim, após implementarmos e validarmos uma estratégia de Síntese de Peptídeos em Fase Sólida no laboratório, pudemos obter quatro epítomos a fim de os validarmos no sensor eletroquímico. Foram produzidos um epítomo Controle, publicado na literatura como sítio antigênico, e os epítomos E7, E8 e E9, alguns dos mais interessantes em termos de localização e propriedades bioquímicas. O epítomo E7 foi utilizado para validarmos o sistema sensor com bons resultados preliminares.

Palavras-Chave: SARS-CoV-2, COVID-19, Glicoproteína *Spike*, Epítomos, Imunossensores

ABSTRACT

The COVID-19 Pandemic has already cost the world countless human, social and economic losses. SARS-CoV-2, its etiologic agent, is an enveloped Betacoronavirus with a genome composed of single-stranded, positive-sense RNA that encodes four structural proteins, with the Spike glycoprotein being the most important due to its function of recognizing the ACE2 receptor and anchoring the virus on the host cell surface. Given the know-how of our research group in developing innovative diagnostic methods for RNA viruses, the present work proposed to identify and characterize molecular targets in the Spike glycoprotein to be applied in an immunoelectrochemical sensor. To this end, a combinatorial method based on artificial intelligence identified 13 potential immunodominant sites scattered throughout the entire S protein. These sites range from 12 to 22 amino acids in length. Through Molecular Dynamics experiments we were able to characterize such epitopes and evaluate important biochemical and biophysical parameters for the antigen-antibody interaction process. It was found that both the S protein and our sites have good conformational stability. We were also able to evaluate the Solvent Accessible Surface Area (SASA) under two conditions and determinate that, for the epitopes found in the RBD region, the most relevant for antibody recognition, this accessibility varies according to the conformation adopted by the RBD. When it is closed, epitopes are more inaccessible, and when it is open, epitopes are significantly more accessible. Finally, after implementing and validating a Solid Phase Peptide Synthesis strategy in the laboratory, we were able to obtain four epitopes in order to validate them in the electrochemical sensor. Control, E7, E8 and E9 epitopes were produced, the most interesting in terms of location and biochemical properties. The epitope E7 was used to validate the sensor's platform with good preliminary results.

Keywords: SARS-CoV-2, COVID-19, Spike glycoprotein, Epitopes, Immunosensors

LISTA DE FIGURAS

Figura 1. Características gerais de diferentes espécies de coronavírus.....	16
Figura 2. Estrutura geral do genoma e do virion de SARS-CoV-2.....	20
Figura 3. Aspectos estruturais da glicoproteína <i>Spike</i> de SARS-CoV-2.....	22
Figura 4. Elementos básicos de um Biossensor.....	31
Figura 5. Estratégia de mapeamento dos potenciais epítomos lineares de célula B.....	51
Figura 6. Localização dos epítomos preditos na glicoproteína <i>Spike</i>	54
Figura 7. Aspectos dinâmicos da proteína S durante 100 ns de simulação.....	57
Figura 8. Evolução dinâmica dos epítomos preditos quando internos a estrutura da proteína S no sistema 6VXX.....	59
Figura 9. Valores de SASA dos epítomos mapeados no RBD da proteína S.....	61
Figura 10. Interações entre anticorpos e a região RBD da proteína <i>Spike</i>	62
Figura 11. RMSD dos epítomos E5 - E9 ao longo de 1 μ s de simulação.....	63
Figura 12. Evolução dinâmica e <i>folding</i> dos epítomos E5 - E9 quando isolados.....	65
Figura 13. Identificação do peptídeo Controle obtido através de SPPS.....	68
Figura 14. Identificação do peptídeo E7 obtido através de SPPS.....	70
Figura 15. Espectro de massas do peptídeo E8.....	71
Figura 16. Espectro de massas do peptídeo E9.....	71
Figura 17. Modelo do Sensor de Ouro Serigrafado Metrohm DropSens 220AT utilizado nos experimentos.....	73
Figura 18. Teste dos sensores.....	74
Figura 19. Alinhamento aplicado aos dados brutos da análise no servidor iBCE-EL.....	93
Figura 20. Alinhamento comparativo completo, com destaque para os epítomos imunodominantes identificados.....	96
Figura 21. Estrutura tridimensional da proteína <i>Spike</i> com destaque para a área coberta pelos epítomos identificados por aplicação individual.....	97

LISTA DE TABELAS

Tabela 1. Composição dos sistemas parametrizados para os experimentos de Dinâmica Molecular.....	39
Tabela 2. Peptídeos possivelmente imunogênicos obtidos nas plataformas DBLEpitope, BepiPred 2.0 e iBCE-EL.....	48
Tabela 3. Epítomos de Célula B imunodominantes da proteína <i>Spike</i> encontrados na análise combinatória.....	52
Tabela 4. Evolução genética e a presença de mutações nos epítomos mapeados dentre as principais variantes de SARS-CoV-2.....	97

LISTA DE ABREVIATURAS E SIGLAS

ng	Nanograma
µg	Micrograma
mg	Miligrama
g	gramas
µL	Microlitro
mL	Mililitro
nm	Nanômetro
nM	Nanomolar
µM	Micromolar
mM	Milimolar
M	Molar
ps	Picossegundo
fs	Femtosegundo
ns	Nanosegundo
µs	Microsegundo
s	Segundo
min	Minuto
h	Hora
N-terminal	Extremidade amino-terminal
C-terminal	Extremidade carboxi-terminal
DNA	Ácido Desoxiribonucleico
RNA	Ácido Ribonucleico
mRNA	RNA mensageiro
S1	Subunidade 1 da proteína <i>Spike</i>
S2	Subunidade 2 da proteína <i>Spike</i>
NTD	Domínio N-terminal
RBD	Domínio de Ligação ao Receptor
RBM	<i>Motif</i> de Ligação ao Receptor
KCl	Cloreto de potássio
NaCl	Cloreto de sódio
HCl	Ácido clorídrico
EDTA	Ácido etilenodiamino tetra-acético
kDa	Quilodalton
pH	Potencial Hidrogeniônico
xg	aceleração

kb	Kilobase
DMF	Dimetilformamida
DCM	Diclorometano
PBS	Tampão salina com fosfato
BSA	Albumina do Soro Bovina
V	Volt
mV	Milivolt
rpm	Rotações por minuto
m/v	Relação massa / volume
° C	Graus Celsius
K	Graus Kelvin

LISTA DE SÍMBOLOS

α	Letra grega Alfa (minúscula)
β	Letra grega Beta (minúscula)
γ	Letra grega Gama (minúscula)
δ	Letra grega Delta (minúscula)
Δ	Letra grega Delta (maiúscula)
ε	Letra grega Épsilon (minúscula)
θ	Letra grega Theta (minúscula)
λ	Letra grega Lambda (minúscula)
μ	Letra grega Mi (minúscula)
π	Letra grega Pi (minúscula)
ϕ	Letra grega Phi (maiúscula)
∇	Operador Hamiltoniano Nabla
∂	Símbolo matemático para Derivada Parcial
Σ	Símbolo matemático para Somatória
$\vec{x}$	Símbolo matemático para Vetor (da variável genérica x)

SUMÁRIO

1 INTRODUÇÃO E JUSTIFICATIVA.....	15
1.1 Os Coronavírus.....	15
1.2 SARS-CoV-2 e a pandemia de COVID-19.....	17
1.2.1 Surto inicial, possíveis origens e os aspectos clínicos da COVID-19.....	17
1.2.2 SARS-CoV-2.....	19
1.2.3 A glicoproteína <i>Spike</i>	21
1.3 Epítomos de Célula B: estrutura, predição e aplicações.....	24
1.4 Estudos atomísticos de Dinâmica Molecular.....	26
1.5 Biossensores e suas aplicações.....	30
1.6 Justificativa.....	33
2 HIPÓTESE.....	34
3 OBJETIVOS.....	35
3.1 Geral.....	35
3.2 Específicos.....	35
4 MATERIAIS E MÉTODOS.....	36
4.1 Análises de bioinformática e imunoinformática da proteína <i>Spike</i>	36
4.1.1 Predição de epítomos lineares de célula B.....	36
4.1.2 Análise comparativa e identificação dos epítomos dominantes.....	37
4.1.3 Alinhamento, modelagem tridimensional e análises complementares.....	37
4.2 Simulações atomísticas via Dinâmica Molecular.....	38
4.2.1 Preparo do modelo da proteína <i>Spike</i>	38
4.2.2 Preparo dos modelos dos epítomos E5 – E9.....	39
4.2.3 Minimização da energia, termalização do sistema e produção das dinâmicas.....	39
4.2.4 Análise das trajetórias.....	40
4.3 Produção e caracterização dos epítomos.....	40
4.3.1 Reagentes.....	40
4.3.2 Síntese química de Peptídeos em Fase Sólida (SPPS).....	41
4.3.3 Clivagem.....	41
4.3.4 Análise da identidade molecular por Espectrometria de Massas (CID-MS2).....	42
4.3.5 Determinação do perfil cromatográfico produtos de síntese por HPLC.....	42
4.4 Desenvolvimento e teste de imunossensor eletroquímico para detecção de anticorpos anti-S.....	43
4.4.1 Sensores.....	43
4.4.2 Lavagem do sensor e imobilização dos epítomos.....	43
4.4.3 Incubação com os soros e leitura imunossensores.....	44
5 RESULTADOS E DISCUSSÃO.....	45
5.1 Mapeando os determinantes antigênicos da glicoproteína <i>Spike</i>	45
5.2 Características gerais dos epítomos preditos.....	53
5.3 Dinâmica Molecular com a proteína S e seus epítomos.....	56
5.3.1 O trímero S é estável e seu comportamento dinâmico é “sincronizado”.....	56
5.3.2 Acessibilidade superficial dos epítomos.....	60
5.3.3 Evolução dinâmica dos epítomos encontrados na região RBD.....	62
5.4 Obtenção e caracterização química dos epítomos.....	66
5.5 Teste dos imunossensores funcionalizados com E7 para detecção de anticorpos anti-S.....	72
6 CONCLUSÕES.....	75
7 PERSPECTIVAS.....	76
8 REFERÊNCIAS.....	77

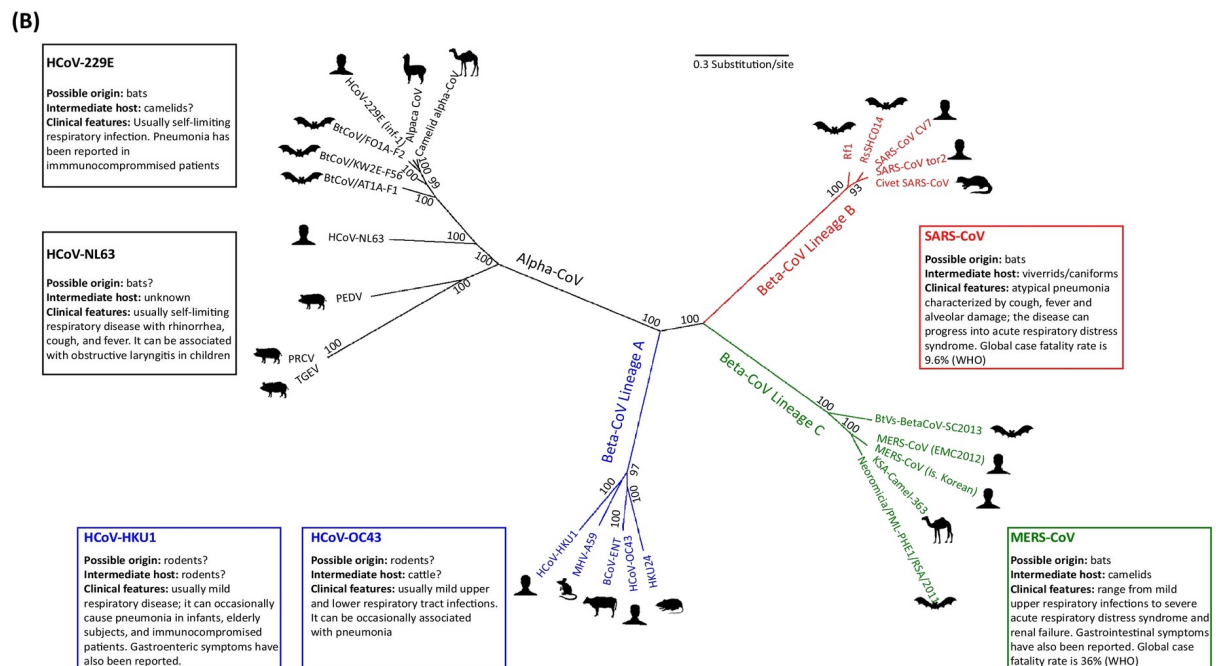
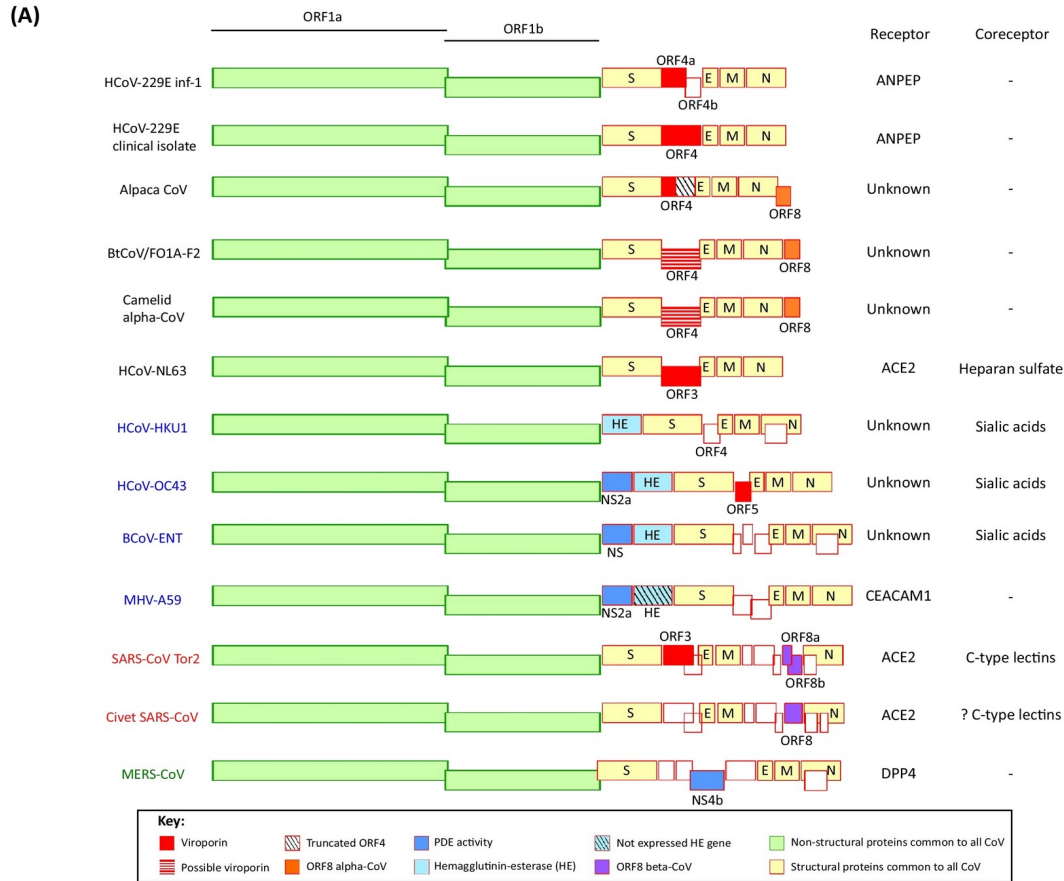
1 INTRODUÇÃO E JUSTIFICATIVA

1.1 Os Coronavírus

Os coronavírus são um grupo de vírus de RNA envelopados, cosmopolitas entre os continentes habitados, conhecidos desde a primeira metade do século XX, que infectam mamíferos e aves e causam patologias respiratórias, neurológicas, hepáticas e entéricas, agudas e/ou crônicas (WEISS; LEIBOWITZ, 2011). O termo “corona” foi atribuído na década de 1960 quando estes vírus foram observados pela primeira vez através de microscopia eletrônica e refere-se a microestrutura do envelope viral que se assemelha a uma “coroa” devido a presença das espículas, ou proteínas *Spike*, as que o vírus usa para ligar-se aos receptores presentes em uma célula hospedeira (PEIRIS, 2016). Taxonomicamente, estes vírus pertencem à ordem Nidovirales, família Coronaviridae, subfamília *Orthocoronavirinae*. As diferentes espécies de coronavírus são distribuídas entre quatro gêneros: os *Alphacoronavirus* (α CoVs), *Betacoronavirus* (β CoVs), *Gammacoronavirus* (γ CoVs) e *Deltacoronavirus* (δ CoVs) (CUI; LI; SHI, 2018).

O material genômico dos coronavírus é composto por fitas simples de RNA de sentido positivo (+ssRNA), fica armazenado enrolado e associado à proteínas do nucleocapsídeo no interior do virion, e por ser de sentido positivo, já se apresentam na forma de mensageiros prontos para serem traduzidos pelos ribossomos nas células hospedeiras. Os coronavírus são os vírus de RNA com os maiores genomas conhecidos, que podem passar de 30 kb de extensão (PEIRIS, 2016). É interessante pontuar que a organização dos genes presentes no genoma desses vírus segue um padrão específico que pode ser observado em praticamente todas as espécies conhecidas, assim como destacado na Figura 1A. Nele, o genoma fica dividido em duas porções, uma compreendendo a extremidade 5' e a outra a extremidade 3'. Na extremidade 5' encontra-se a ORF1a/b que codifica proteínas não estruturais como as proteases (ORF1a) e maquinaria de replicação (polimerase/helicase, ORF1b), além de proteínas regulatórias/auxiliares. Já na extremidade 3' encontram-se as proteínas estruturais do vírus seguindo a ordem S (*Spike*) – E (envelope) – M (de membrana) – N (nucleocapsídeo), podendo conter a hemaglutinina (HE), além de proteínas regulatórias/auxiliares. No total, em torno de 30 proteínas são codificadas (FORNI et al., 2016; PEIRIS, 2016).

Figura 1. Características gerais de diferentes espécies de coronavírus. **a)** Organização genômica com destaque para a estrutura típica encontrada entre coronavírus. **b)** Relações taxonômicas e principais características dos coronavírus de importância médica antes do surgimento de SARS-CoV-2.



Fonte: adaptada de FORNI et al., 2017.

Até o fim do ano de 2019, eram conhecidas seis espécies de coronavírus com importância médica capazes de infectar e causar alguma doença em seres humanos, doenças estas que vão desde o resfriado comum (para os vírus 229E, OC43, NL63 e HKU1) à patologias respiratórias graves e muitas vezes fatais (para os vírus SARS-CoV-1 e MERS-CoV) (SU et al., 2016; CUI; LI; SHI, 2018). Todas estas seis espécies são pertencentes aos gêneros dos *Alphacoronavirus* e dos *Betacoronavirus*, que infectam aves e mamíferos (Figura 1B).

Uma característica importante dessas espécies é sua origem zoonótica, com hospedeiro natural e possíveis hospedeiros intermediários identificados. HCoV-229E, HCoV-NL36, MERS-CoV e SARS-CoV possuem variantes ancestrais e/ou muito próximas circulando em morcegos, seu principal reservatório e possivelmente hospedeiro natural. Ambos HCoV-HKU1 e HCoV-OC43 possuem possível origem em roedores. A partir de seus reservatórios naturais, esses vírus utilizam espécies próximas aos seres humanos para saltarem, como camelídeos, bovinos e mamíferos de pequeno porte, além disso, o salto direto entre o hospedeiro natural e o homem também pode ocorrer em regiões onde o contato é frequente (WEISS; LEIBOWITZ, 2011; PEIRIS, 2016 ; FORNI et al., 2017).

1.2 SARS-CoV-2 e a pandemia de COVID-19

1.2.1 Surto inicial, possíveis origens e os aspectos clínicos da COVID-19

Em dezembro de 2019, na cidade de Wuhan, capital da província de Hubei, China, foi constatado um surto de pneumonia de causa até então desconhecida, e em 7 de Janeiro de 2020, autoridades sanitárias chinesas isolaram, a partir de pacientes em Wuhan, uma nova variante de coronavírus diretamente relacionada com tais casos (ZHU et al., 2020). A rápida identificação e sequenciamento do novo vírus permitiram o desenvolvimento de métodos diagnósticos baseados em PCR em tempo real (RT-qPCR), fato que auxiliou pesquisadores e autoridades em saúde a mapear a veloz expansão do número de casos e da área afetada, e ao final de Janeiro de 2020, quase 10 mil casos já haviam sido identificados em 20 países, sendo a maioria ainda concentrada em território Chinês (WHO, 2020a; CORMAN et al., 2020). Devido ao cenário, a Organização Mundial da Saúde (OMS) reconheceu a epidemia uma Emergência de Saúde Pública de Âmbito Internacional (WHO, 2020b).

O primeiro caso do novo coronavírus confirmado no Brasil se deu em 27 de fevereiro de 2020, importado da Itália. Neste ponto, o vírus já era denominado 2019n-CoV. Durante a escrita do projeto que deu origem ao presente trabalho o país possuía quase 1.2 milhão de casos, com ~54 mil mortes confirmadas (WHO, 2020c; BRASIL, 2020a). Durante a escrita do presente documento, o Brasil soma quase 29 milhões de casos e quase 700 mil óbitos (BRASIL, 2020b). A OMS reconheceu a pandemia do novo vírus no dia 11 de março de 2020 (WHO, 2020d). Por recomendação do Grupo de Estudos dos Coronaviridae, parte do Comitê Internacional para Taxonomia de Vírus, 2019n-CoV foi oficialmente reconhecido como SARS-CoV-2 devida sua similaridade com o vírus SARS-CoV original, agora SARS-CoV-1, com quem compartilha cerca de 80% de similaridade genômica (GORBALENYA et al., 2020).

A origem do novo vírus ainda é uma incógnita em aberto, entretanto, é interessante salientar que, como já comentado, todos os coronavírus conhecidos por infectar humanos possuem origem zoonótica, que também pode ser o caso de SARS-CoV-2 (CUI; LI; SHI et al., 2019). As análises das primeiras sequências publicadas indicam uma provável origem em morcegos, uma vez que seu genoma apresenta 96,2% de identidade com a espécie BatCoV RaTG13, também identificada na China e derivada de SARS-CoV (ZHOU et al., 2020), hipótese que é sustentada pelo fato de que ambos SARS-CoV e MERS-CoV, de maior importância epidemiológica, possuem origem e reservatório em morcegos (LI et al., 2005; MOHD; ALTAWFIQ; MEMISH, 2016; ZHOU et al., 2020).

Se levarmos em consideração os surtos anteriores de *Betacoronavírus*, como o de 2002-2003 para SARS-CoV e os de 2013-2016 para MERS-CoV, além de se originarem em morcegos, estes vírus contaram com outros mamíferos como intermediários (civetas e dromedários, respectivamente) entre estes e os seres humanos, o que levou diversos grupos a propor os mais prováveis intermediários entre o novo coronavírus e o homem (RABI et al., 2020). Três estudos mais recentes também buscaram aprofundar a discussão em relação à origem de SARS-CoV-2 e demonstraram, através de análises filogenéticas detalhadas, que sua origem pode ser reconstruída a partir de variantes que circulam em espécies de morcegos do gênero *Rhinolophus*, e que as amostras obtidas dos primeiros pacientes afetados apontam como epicentro e marco zero o mercado *Huanan Seafood Wholesale Market*, em Wuhan, onde havia a venda e contato direto com animais silvestres passíveis de carregarem variantes de coronavírus (LYTRAS et al., 2022; PEKAR et al., 2022; WOROBEY et al., 2022).

SARS-CoV-2 é um vírus de transmissão respiratória que causa a doença infectocontagiosa COVID-19 (do inglês *Coronavirus Disease 2019*). Pacientes portadores tipicamente

apresentam um período de incubação que pode durar até 14 dias, com média de 5 dias para o aparecimento dos primeiros sintomas, que incluem, mas não se limitam a febre, tosse seca, fadiga, dispneia, mialgia, desordens gastrointestinais (majoritariamente diarreia e náusea), perda de paladar e olfato e falta de ar (JIANG et al., 2020; SUN et al., 2020). É esperada para a maioria dos pacientes positivos para a COVID-19 uma evolução de leve a moderada, e os sintomas podem persistir de uma a duas semanas. Uma parcela relevante dos infectados pode ser assintomática, o que facilita o espalhamento do vírus e dificulta as medidas de contenção. Pacientes que evoluem para a forma grave da COVID-19 necessitam de tratamento em unidade intensiva (UTI) com o uso de respiradores (ZHOU et al., 2020). Com o aparecimento das variantes, houveram mudanças significativa no período de incubação, na sintomatologia, nos marcadores clínicos e na gravidade da doença (ONG et al., 2021; HUANG et al., 2021; CHRISTENSEN et al., 2022; JASSAT et al., 2022).

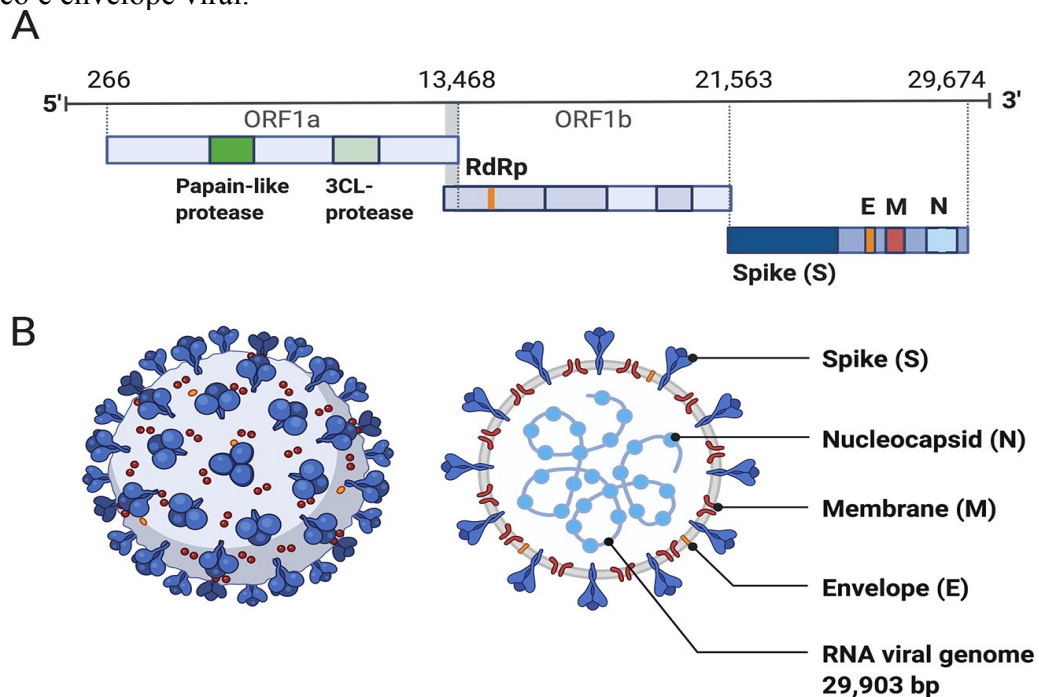
Em termos de achados radiológicos, segundo observações de Guan e colaboradores (GUAN et al., 2020), mais de 86% das Tomografias Computadorizadas (CT scans) realizadas em pacientes chineses no momento de sua admissão nos hospitais se apresentaram alteradas, sendo a Opacidade em Vidro-Fosco e a Opacidade Irregular Bilateral as anomalias mais comuns. No que se diz respeito a achados laboratoriais, foram observados níveis elevados de proteína C-reativa na maioria dos pacientes, bem como níveis elevados de lactato desidrogenase, além de linfocitopenia, trombocitopenia, leucopenia e tempo de protrombina prolongado (GUAN et al., 2020; WANG et al., 2020; ZHOU et al., 2020).

1.2.2 SARS-CoV-2

Em relação aos aspectos de sua virologia, SARS-CoV-2 é um vírus envelopado, esférico e pertencente ao gênero dos *Betacoronavirus* (β CoVs), cujo virion mede entre 100 e 160 nm de diâmetro. Seu genoma possui aproximadamente 30 kb de extensão e é formado por moléculas de RNA fita simples de senso positivo (+ssRNA) com duas regiões bem definidas, a extremidade 5'-replicase e a extremidade 3'-estrutural (Figura 2A) (CHAN et al., 2020). Na extremidade 5' do genoma viral, que cobre aproximadamente dois terços do total, encontram-se as ORFs 1a e 1b, cada qual codifica uma poliproteína (pp1a e pp1b) passível de ser clivada em ao menos 16 proteínas não-estruturais (no total, nsp1 – 16), enquanto na extremidade 3', as ORFs restantes são traduzidas a partir de 9 RNAs mensageiros subgenômicos em cerca de

13 proteínas. É nessa região que se encontram codificadas as 4 proteínas estruturais do vírus: a glicoproteína de superfície (ou spike, S), a glicoproteína de membrana (M), a fosfoproteína do nucleocapsídeo (N) e a proteína do envelope (E). Além das proteínas estruturais, na extremidade 3' do genoma viral também estão presentes 9 fatores acessórios. Essa espécie não possui a hemaglutinina esterase (HE) (CHAN et al., 2020). A Figura 2B esquematiza a estrutura básica da partícula viral de SARS-CoV-2.

Figura 2. Estrutura geral do genoma e do virion de SARS-CoV-2. **a)** Estrutura do genoma, com indicação da localização dos genes estruturais, acessórios e não-estruturais. **b)** Estrutura da partícula viral madura, mostrando a localização relativa das proteínas estruturais, material genético e envelope viral.



Fonte: RANDO et al., 2020.

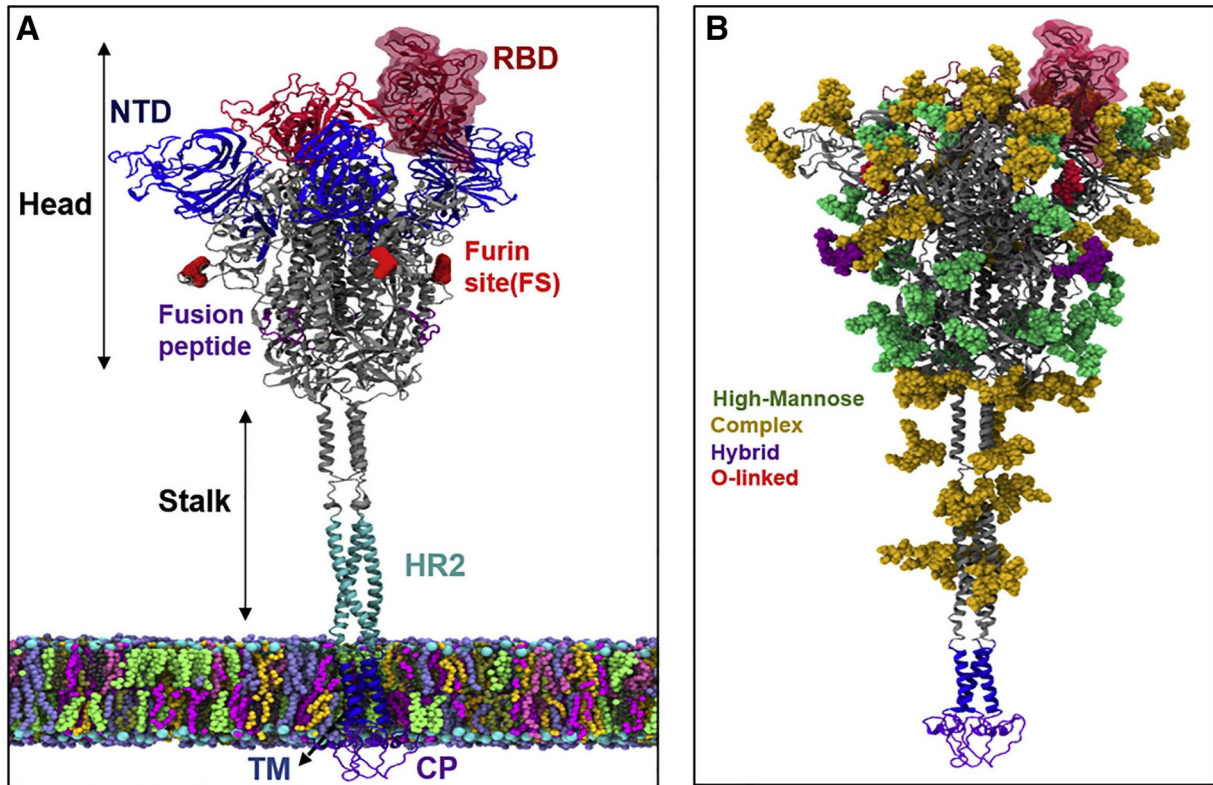
Muito da biologia que envolve o ciclo viral de SARS-CoV-2 foi inicialmente elucidada baseado na literatura sobre SARS-CoV-1, com quem compartilha ~80% de identidade estrutural (LU et al., 2020). SARS-CoV-1 utiliza a glicoproteína S sob a forma de um trímero para o reconhecimento e a fusão junto membrana celular durante o processo de entrada (SIMMONS et al., 2013). O receptor utilizado por SARS-CoV-1 é a enzima ACE2 (*Angiotensin-Converting Enzyme type II*) (XU et al., 2020). Durante o reconhecimento, o trímero S é clivado pela ação de proteases da superfície celular em duas subunidades, S1 e S2, estando localizado na subunidade S1 o domínio de reconhecimento do receptor (RBD), que se liga diretamente ao domínio peptidase de ACE2, enquanto o domínio S2 é responsável pela fusão à membrana (LI et al., 2005; BELOUZARD; CHU; WHITTAKER, 2009; SIMMONS et al., 2013). A estrutura da glicoproteína S de SARS-CoV-2 foi satisfatoriamente determinada por Microscopia Crio-

Eletrônica no início de março de 2020 revelando que a estrutura é otimizada para a interação com o receptor ACE2, com uma ligação cerca de 10 vezes mais forte se comparada a SARS-CoV e uma constante de dissociação de (K_d) de aproximadamente 15 nM (WRAPP et al., 2020; YAN et al., 2020; ANDERSEN et al., 2020).

1.2.3 A glicoproteína *Spike*

Como já comentado, SARS-CoV-2 utiliza a glicoproteína *Spike*, uma proteína de fusão de classe I, como porta de entrada para a infecção de suas células hospedeiras. A proteína S é uma proteína transmembranar encontrada no envelope viral na forma de homotrímeros. Há duas porções distintas na proteína S: a cabeça (ou *head*), que tende a morfologia globular, e o pedestal (ou *stalk*), mais helicoidal/fibroso (Figura 3A). Cada protômero individual que compõe o trímero S maduro possui 1273 aminoácidos de extensão e é fisicamente dividido em duas subunidades funcionais: S1 e S2 (WRAPP et al., 2020). A divisão da proteína S em suas subunidades ocorre durante a passagem da mesma pelo retículo endoplasmático da célula hospedeira, durante a fase final da infecção celular, e é ocasionada pela ação da protease furina, que reconhece uma sequência específica de aminoácidos, a sequência 681-PRRAR↓S-686, e cliva cada uma das três unidades *Spike* entre os aminoácidos R685 e S686 (CHENG et al., 2020).

Figura 3. Aspectos estruturais da glicoproteína *Spike* de SARS-CoV-2. **a)** Modelo tridimensional completo (*head* + *stalk*) da proteína ancorada em uma bicamada lipídica com destaque para os domínios N-terminal (NTD, azul), de ligação ao receptor (RBD, vermelho), Heptad Repeat 2 (HR2, turquesa), transmembranar (TM, azul marinho) e citoplásmico (CP, roxo), além do sítio de clivagem da furina. **b)** Localização das glicosilações conhecidas na proteína S, sendo as complexas marcadas em amarelo-ouro, as híbridas em roxo, as de alto teor de manose em verde-musgo e as O-glicosilações em vermelho.



Fonte: adaptada de GHORBANI; BROOKS; KLAUDA, 2021.

As subunidades da proteína S possuem funcionalidades distintas e essa primeira clivagem é tida como uma das vantagens evolutivas que tornam SARS-CoV-2 mais infeccioso que seu antecessor, o vírus SARS-CoV-1 (ANDERSEN et al., 2020). A subunidade S1 compreende os aminoácidos 14 a 685 (considerando que de 1 – 13 nós temos o peptídeo sinal que direciona a mesma ao retículo endoplasmático) e é composta por dois domínios: o Domínio N-terminal (NTD, aa 14 – 305), e o Domínio de Ligação ao Receptor (RBD, aa 319 - 541), sendo sua função, consequentemente, o reconhecimento do receptor e a ligação do vírus na superfície da célula em que este está prestes a infectar (LAN et al., 2020). A subunidade S2, por sua vez, compreende os aminoácidos 686 a 1273 e é formada pelo Peptídeo de Fusão (FP), pelos Heptad Repeats 1 e 2 (HR1 e HR2), pelo Domínio Transmembranar (TMD) e pela porção citoplásmica (CP). Os HRs e a porção TMD são responsáveis pela sustentação da estrutura da proteína e pela ancoragem dessa no envelope viral, enquanto o FP media o processo de fusão

do envelope viral na membrana da célula hospedeira ou na membrana do endossomo, a depender do processo de invasão disparado (WRAPP et al., 2020; CAI et al., 2020).

Para infectar uma célula, SARS-CoV-2 utiliza como receptor celular a também glicoproteína ACE2, enzima transmembranar componente do ciclo de regulação da pressão sanguínea Renina-Angiotensina-Aldosterona e responsável por converter angiotensina II em angiotensina 1-9 (DONOGHUE et al., 2000; YAN et al., 2020). ACE2 é uma proteína produzida por diversos órgãos e tecidos e encontra-se majoritariamente nos pulmões, córnea, rins e coração (SUN et al., 2020), fato que, em partes, explica o tropismo de SARS-CoV-2 para estes órgãos.

Além da clivagem em S1/S2, a proteína S passa por mais dois outros tipos de modificação pós-traducional extremamente importantes para sua estrutura/função. Uma dessas modificações ocorre durante o processo de fusão do envelope viral com a membrana celular e corresponde a mais uma clivagem, desta vez na região conhecida como S2' (BESTLE et al., 2020). Essa segunda clivagem libera o peptídeo de fusão e é mediada pela protease membrana TMPRSS2, ativando os eventos de fusão entre o envelope viral e a membrana plasmática e garantindo a invasão do vírus (HOFFMANN et al., 2020). O terceiro tipo de modificação corresponde a suas glicosilações.

A extensa presença de carboidratos em proteínas de superfície de vírus patógenos de mamíferos é um fato bem conhecido (VIGERUST; SHEPHERD, 2007). A glicosilação de múltiplos sítios na proteína *Spike* de SARS-CoV-2 já era esperada antes mesmo da publicação das primeiras estruturas experimentais, uma vez que tanto SARS-CoV-1 quanto MERS-CoV, bem como diversos outros coronavírus, possuem uma elevada carga de glicídios em suas proteínas S (CHO et al., 2021). Conhecer a localização e a composição destes carboidratos é essencial se desejamos entender aspectos biológicos do vírus, como sua estrutura, sua capacidade infectiva e a evasão do sistema imune proporcionada pelos açúcares (BAGDONAITIS; WANDALL, 2018).

Quando as primeiras sequências genômicas de SARS-CoV-2 foram publicadas, 22 sítios de N-glicosilação foram preditos na proteína S, e todos eles foram experimentalmente confirmados (WATANABE et al., 2020). Além das N-glicosilações, também foram identificados alguns O-glicopeptídeos em menor extensão (SHAJAHAN et al., 2020; SANDA; MORRISON; GOLDMAN, 2021). Os principais açúcares detectados como componentes dos glicanos da *Spike* são a hexose, a fucose, a galactose, o ácido siálico, a N-acetilglicosamina, a manose e a N-acetilgalactosamina, que formam estruturas com alto teor de manose, híbridas, complexas, dentre outras. A distribuição das glicosilações ao longo da proteína S não é homo-

gênea, estando elas concentradas na subunidade S1 (exceto o RBD), e no entorno do HR2, como pode ser observado na Figura 3B (WATANABE et al., 2020; SHAJAHAN et al., 2020; ZHANG et al., 2021; SANDA, MORRISON; GOLDMAN, 2021).

É interessante ressaltar que, assim como a composição de aminoácidos da glicoproteína *Spike* não é fixa e se altera conforme o vírus adquire mutações em seu genoma, a composição de carboidratos também não é fixa. Esse fato se deve a basicamente três mecanismos: (1) a própria variação na sequência de aminoácidos pode alterar a posição de sítios favoráveis a ação de transferases; (2) estes sítios não necessariamente serão ocupados em 100% das vezes; e (3) a ocupação deles pode variar entre diversos carboidratos diferentes, o que já foi observado (XU et al., 2020; ZHANG et al., 2021; SANDA, MORRISON; GOLDMAN, 2021). Essa variação também impacta diretamente na imunidade gerada contra a proteína S e deve ser levada em consideração se desejamos aplicar essa estrutura, bem como sítios específicos da mesma, em formulações vacinais e plataformas para avaliar a resposta imunológica gerada. Como discutido acima, essas glicosilações possuem diversos propósitos que ainda se encontram sob investigação.

1.3 Epítomos de Célula B: estrutura, predição e aplicações

Por ser a porta de entrada do vírus na célula hospedeira, a proteína *Spike* tem recebido atenção especial quanto ao desenvolvimento de vacinas, tratamentos e métodos diagnósticos que visam tentar mitigar a pandemia e seus efeitos (KRAMMER, 2020; BROWN; MCCULLOUGH, 2020). Diversos estudos que acompanharam pacientes recuperados da doença identificaram vários anticorpos neutralizantes direcionados contra vários epítomos na proteína S (LIU et al., 2020; WAJNBERG et al., 2020; BARNES et al., 2020). Quando um anticorpo reconhece e se liga a um antígeno, a região específica a partir da qual são formadas as interações entre o paratopo do anticorpo e o epítopo do antígeno é chamada de determinante antigênico ou epítopo de célula B (POTOCNAKOVA; BHIDE; PULZOVA, 2016).

Epítomos de célula B são estruturas com potencial aplicação no desenvolvimento de novas vacinas, na produção de anticorpos recombinantes, no desenvolvimento de novos métodos diagnósticos ou mesmo na pesquisa científica para caracterizar a resposta imune contra determinada estrutura imunogênica (SANCHEZ-TRINCADO; GOMES-PEROSANZ; RECHE, 2017; LIU; SHI; LI, 2020). Existem dois tipos de epítomos de célula B: os epítomos line-

ares, formados por resíduos de aminoácidos que se apresentam de forma sequencial na estrutura primária da proteína, e os epítomos conformacionais, formados por aminoácidos que se encontram distantes na estrutura primária da proteína, mas são trazidos para perto uns dos outros pelo *folding* da mesma e permanecem neste estado pelas interações intracadeia polipeptídica (SANCHEZ-TRINCADO; GOMES-PEROSANZ; RECHE, 2017).

Dada sua importância e aplicabilidade, um dos grandes desafios da imunologia foi tentar prever, de forma acurada, quais resíduos de aminoácidos ou regiões de uma proteína fazem parte de um epítomo de célula B, e diversas ferramentas computacionais foram desenvolvidas para tal (POTOCNAKOVA; BHIDE; PULZOVA, 2016). A predição de epítomos lineares é a que concentra a maior parte dos métodos, mesmo que estes epítomos representem apenas cerca de 10% do total de epítomos de célula B, entretanto, algumas vantagens de se trabalhar com epítomos lineares podem ser citadas, como a relativa fácil obtenção dos mesmos através de estratégias de síntese de peptídeos e de biologia molecular, o que otimiza sua aplicação no desenvolvimento de imunizantes e testes diagnósticos, além da possibilidade de reconhecimento por anticorpos mesmo com a desnaturação do epítomo, uma vez que estes estarão direcionados para a sequência e não para a distribuição espacial dos mesmos, o que não é verdadeiro para epítomos conformacionais (VAN REGENMORTEL, 2009; SANCHEZ-TRINCADO; GOMES-PEROSANZ; RECHE, 2017).

Os primeiros métodos para predição de epítomos lineares se baseavam em escalas de propensão e probabilidade de determinado aminoácido ser ou não parte de um epítomo de célula B de acordo com suas características físico-químicas como flexibilidade, hidrofobicidade, acessibilidade superficial, e/ou de acordo com a frequência estatística de aparecimento do resíduo em epítomos experimentalmente validados, como é o caso do método de Kolaskar & Tongaonkar e do método PREDITOP (PELLEQUER; WESTHOF, 1993; KOLASKAR; TONGAONKAR, 1990; SANCHEZ-TRINCADO; GOMES-PEROSANZ; RECHE, 2017). Alguns trabalhos foram desenvolvidos para validar a acurácia de predição destes métodos e todos concluem que se basear em escalas de propensão apenas traz resultados pobres e ligeiramente superiores à escolha randômica entre epítomos e não epítomos (BLYTHE; FLOWER, 2005; ANSARI; RAGHAVA, 2013; SANCHEZ-TRINCADO; GOMES-PEROSANZ; RECHE, 2017).

No intuito de melhorar a capacidade de predição, novos métodos baseados em inteligência artificial e aprendizado de máquina foram introduzidos. A ideia geral por trás destes algoritmos é, em primeiro lugar, obter bancos de dados contendo epítomos e não epítomos validados experimentalmente, seja por técnicas bioquímicas, imunológicas ou estruturais, e então

parametrizar as características básicas destes epítomos para que as mesmas possam ser utilizadas no treinamento de algoritmos de aprendizado de máquina (ANSARI; RAGHAHA, 2013). Aplicações como BepiPred, BepiPred 2.0, ABCpred, Bcepred, iBCE-EL e DLBEpitope foram desenvolvidas, validadas e gratuitamente disponibilizadas através de servidores *web* para serem utilizadas pela comunidade científica (SAHA; RAGHAHA, 2004; LARSEN; LUND; NIELSEN, 2006; SAHA; RAGHAHA, 2006; JESPERSEN et al., 2017; MANAVALAN et al., 2018; LIU; SHI; LI, 2020).

Mesmo com a melhora na acurácia de predição obtida pela implementação da inteligência artificial, o uso dessas ferramentas de forma isolada, apesar de válido, deve ser considerado com cautela. Epítomos são estruturas tão ou mais complexas quanto seus antígenos originários e carregam características únicas em termos de composição, estrutura e propriedades. Ter como alvo apenas a estrutura primária do antígeno significa deixar de lado uma porção grande e importante das características inerentes à esse alvo, como é o caso da glicoproteína *Spike*. Informações relacionadas à estrutura, interações intra e intermoleculares, modificações pós-traducionais e ao comportamento dinâmico podem ser perdidas, mesmo sendo determinantes em processos bioquímicos como o reconhecimento antígeno-anticorpo (MARIUZZA; PHILLIPS; POLJAK, 1987; ABBAS; LICHTMAN; PILLAI, 2022).

1.4 Estudos atomísticos de Dinâmica Molecular

Sistemas biomoleculares não são estáticos, eles se movimentam ao longo do tempo de acordo com as forças que são exercidas sobre os mesmos. Determinar e analisar o movimento de uma biomolécula na escala atômica é de extrema importância se desejamos entender mais detalhes sobre sua estrutura e função, pois esse movimento molecular é inerente a todos os processos físicos e químicos em que o sistema está envolvido (KLEPEIS et al., 2009; OLIVEIRA, 2013). O movimento dinâmico de um sistema biomolecular pode ser estudado através de experimentos de Dinâmica Molecular (MD). As simulações de MD empregam algoritmos desenvolvidos para calcular a evolução dinâmica de cada átomo do sistema ao longo do tempo, ou seja, mapear sua trajetória baseada nas forças aplicadas e nas interações permitidas entre os átomos, descritas matematicamente pelos chamados Campos de Força (GONZÁLEZ, 2011).

O primeiro requisito e muitas vezes o mais difícil de se obter para um estudo atomístico de MD é a descrição detalhada e de qualidade das posições iniciais de cada átomo componente da biomolécula (ou sistema) que se deseja estudar. Essas coordenadas podem ser determinadas por técnicas experimentais de biologia estrutural como Cristalografia por Difração de Raios-X, Ressonância Magnética Nuclear (NMR), Microscopia Crioeletrônica (Cryo-EM), dentre outras, ou por estratégias de bioinformática, como Modelagem por Homologia, e etc. O segundo requisito são as velocidades iniciais, atribuídas a partir da temperatura que selecionamos para nossos experimentos e distribuídas aleatoriamente entre os átomos do sistema via aproximação de Maxwell-Boltzmann (MOTTIN, 2015).

Tendo em mãos as coordenadas iniciais $\vec{r}_i(t=0)$ e as velocidades iniciais $\vec{v}_i(t=0)$ de cada átomo do sistema, e estabelecendo um Campo de Força que descreve como estes irão interagir entre si e como a energia potencial será distribuída dentro da caixa de simulação, calculam-se a força e a aceleração individual dos átomos seguindo a segunda Lei de Newton:

$$\vec{F}_i = m_i \vec{a}_i \quad (1)$$

que também podem ser escrita como:

$$\vec{a}_i = \vec{F}_i / m_i \quad (1.1)$$

Uma vez que a aceleração instantânea do átomo i $\vec{a}_i$ nada mais é que a derivada de sua velocidade instantânea $\vec{v}_i$, e que esta é a derivada da posição $\vec{r}_i$ do átomo i no tempo t , a equação (1) pode ser reescrita como:

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} \quad (2)$$

Tendo em vista que a equação (2) é uma equação diferencial, sua solução pode ser obtida através da integração da mesma, entretanto, como não é possível determinar estes parâmetros para todos os possíveis intervalos de tempo, é necessário integrar através de diferenças finitas, uma vez que os experimentos de dinâmica molecular acontecem em pequenos passos de tempo Δt , na ordem dos femtosegundos, logo, uma integração numérica deve ser aplicada em cada um deles (OLIVEIRA, 2013). O algoritmo de Verlet é um dos métodos mais amplamente utilizados na integração numérica por diferenças finitas das equações do movimento em simulações de MD (VERLET, 1967). Com o método de Verlet, chegamos às seguintes aproximações para o cálculo das velocidades e acelerações (equação 3), e das posições (equação 4):

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \vec{a}_i(t) \cdot \Delta t \quad (3)$$

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t) \cdot \Delta t + \frac{\vec{a}(t) \cdot \Delta t^2}{2} \quad (4)$$

Uma vez estabelecidas essas relações, a cada novo passo tempo Δt é possível atualizar as posições ($\vec{r}_i$), velocidades ($\vec{v}_i$) e acelerações ($\vec{a}_i$) dependentes do tempo para o átomo i com base no passo anterior $t + \Delta t$. O calculo é repetido durante o decorrer da simulação, e com o conjunto destes dados salvos em intervalos regulares temos o resultado de todo experimento de MD: a trajetória dos átomos do sistema em relação ao tempo.

É imprescindível ressaltar que a trajetória gerada é nossa amostragem do sistema, portanto, esta deve possuir uma extensão temporal adequada para as análises propostas, para que a evolução dinâmica de nosso sistema calculada permita que este visite um conjunto representativo das configurações/conformações possíveis dentro dos parâmetros especificados (Campo de Força, sistema solvente, íons, presença de outras moléculas, etc) (OLIVEIRA, 2013). Por fim, com a trajetória em mãos, é possível calcular algumas propriedades macroscópicas observáveis do sistema através de aproximações da mecânica estatística, uma vez que, como já comentado, as forças resultantes agindo sobre o sistema podem ser descritas como uma função da energia potencial do sistema, de acordo com a equação:

$$\vec{F}_i = -\nabla_i U(\vec{r}_1(t), \vec{r}_2(t), \dots, \vec{r}_n(t)) \quad (5)$$

onde $\vec{F}_i$ representa a força resultante agindo sobre o átomo i , e $-\nabla_i U(\vec{r}_1(t), \vec{r}_2(t), \dots, \vec{r}_n(t))$ representa o gradiente da energia potencial U do sistema em estudo, em relação as coordenadas do átomo i , que é função das posições de todos os átomos que o compõe. Nota-se que esse gradiente é negativo pois a energia potencial está sendo aplicada através das forças agindo sobre os átomos do sistema para desencadear seu movimento.

Em nosso estudo, foi empregado o *software* Amber 2018, que reúne um conjunto de diferentes programas e ferramentas voltados para a simulações atomísticas via mecânica molecular e mecânica quântica. No Amber é possível escolher entre uma miríade de diferentes campos de força para experimentos com moléculas orgânicas, inorgânicas e biomoléculas (DNA, RNA, carboidratos, lipídeos e proteínas) em vários sistemas solventes. Para distribuir a energia potencial através das diferentes interações permitidas entre os átomos do sistema, um campo de força interpretado pelo Amber deve, ao menos, especificar parâmetros ligados da mecânica molecular e não ligados, sendo estes últimos baseadas nos potenciais Coulumbicos e de Lennard-Jones, normalmente otimizadas a partir de dados experimentais de química

quântica (SALOMON-FERRER; CASE; WALKER, 2012). Neste sentido, a energia potencial $E(\vec{r}_i)$ relacionada ao átomo i em determinada coordenada $\vec{r}_i$ é definida em termos de:

$$E(\vec{r}_i) = E_{\text{Ligação}} + E_{\text{Torção}} + E_{\text{Angular}} + E_{\text{Não-ligada}} \quad (6)$$

ou seja, ela é dependente das energias de ligação, de torção (diédrica), angular e não-ligada entre o átomo i e os outros átomos do sistema. Em azul nós temos os termos ligados e em vermelho o termo não-ligado (que engloba o potencial Coulumbico e o de Lennard-Jones). Cada uma das energias acima pode ser decomposta matematicamente de acordo com as equações:

$$E_{\text{Ligação}} = \sum_{\text{Ligações}} \frac{k_b}{2} (r - r_0)^2 \quad (6.1)$$

$$E_{\text{Torção}} = \sum_{\text{Torções}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma_1)] \quad (6.2)$$

$$E_{\text{Angular}} = \sum_{\text{Ângulos}} k_\theta (\theta - \theta_0)^2 \quad (6.3)$$

$$E_{\text{Não ligada}} = \sum_{i>j} f_{ij} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} + \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} \right) \quad (6.4)$$

Um protocolo típico de simulação por Dinâmica Molecular consiste em preparar a(s) biomolécula(s) alvo e a caixa de simulação, minimizar a energia do sistema, equilibrar a temperatura, e por fim calcular a trajetória (GELPI et al, 2015). A primeira etapa muito dependerá da origem da estrutura alvo. Aqui, átomos/*loops* ausentes devem ser posicionados/modelados, as cargas devem ser precisamente distribuídas entre os átomos, a caixa periódica deve ser criada com uma dimensão adequada ao sistema em estudo, o solvente, íons e demais moléculas devem ser adicionados. No Amber, esta etapa pode ser feita empregando o módulo *tleap* (SALOMON-FERRER; CASE; WALKER, 2012). A minimização de energia é uma etapa crucial para remover as tensões e restrições conformacionais do alvo que se originam de acordo com o método aplicado na obtenção das coordenadas atômicas, sendo características destes, uma vez que a determinação da estrutura ocorre, em grande parte das vezes, em condições não fisiológicas (LINDAHL, 2015). Para minimizar a energia, o algoritmo de MD irá calcular o gradiente da energia potencial do sistema até que seja atingido um mínimo local (JAABEN; MOHAMEDALI; RANGANATHAN, 2019). Por fim, logo antes de iniciarmos nossa simulação, a equilibração da temperatura (ou termalização do sistema) visa levarmos nosso sistema até um ponto próximo ao estado de equilíbrio termodinâmico, ou seja, conduzi-lo gradativamente a uma conformação favorável em determinada temperatura e pressão (OLIVEIRA, 2013). Isso é feito aplicando um sistema barostato/termostato que irá aumentar gradativamen-

te a temperatura dentro da caixa de simulação (até a temperatura desejada), em condições NVT (com volume constante), permitindo que o alvo alcance o equilíbrio sem fortes distorções conformacionais ocasionadas por má distribuição de energia em pontos específicos do sistema.

Tendo em mãos um sistema corretamente preparado, com energia minimizada e temperatura equilibrada, inicia-se a produção da dinâmica, e para isto basta aplicar o método escolhido para a integração numérica das equações do movimento e salvar periodicamente as posições ($\vec{r}_i$), velocidades ($\vec{v}_i$) e acelerações ($\vec{a}_i$) dependentes do tempo para construir a trajetória. Esta etapa normalmente é feita em condições NTP (com pressão e temperatura constantes). Finalizada a etapa de produção da MD, a trajetória deve ser analisada e os parâmetros relevantes ao trabalho calculados. No presente estudo, buscamos avaliar propriedades bioquímicas/biofísicas dos epítomos que são relevantes ao processo de reconhecimento molecular entre o paratopo de um anticorpo e o epítomo do antígeno e que podem ter um papel crucial na capacidade dos mesmos de induzir uma resposta de célula B e serem bons alvos para a detecção de anticorpos anti-S.

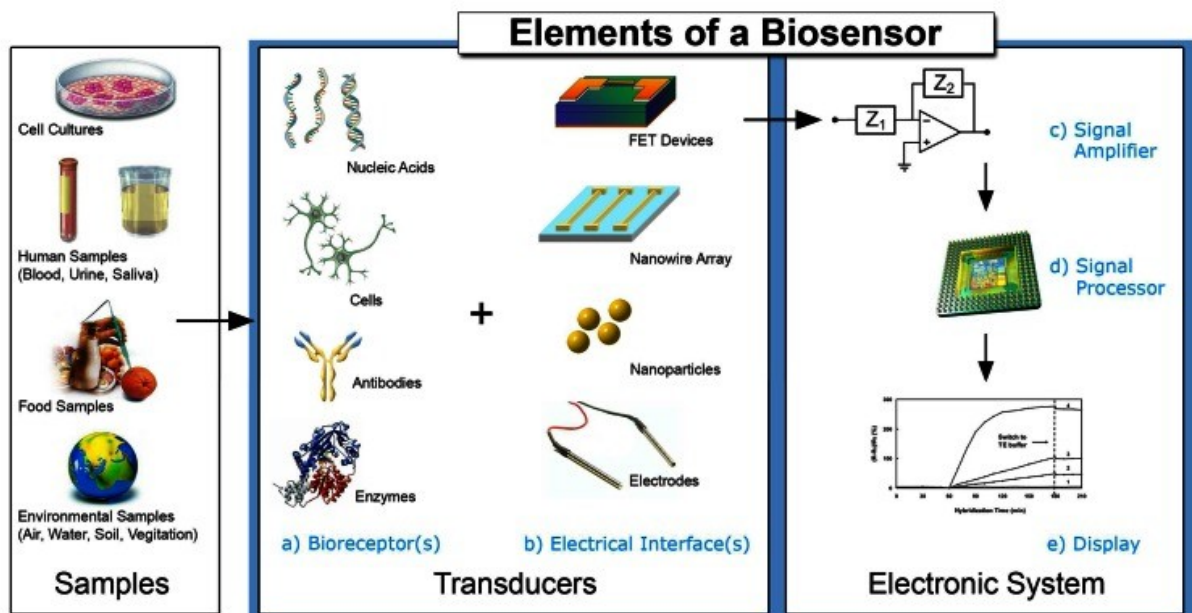
1.5 Biossensores e suas aplicações

Detectar e quantificar componentes e processos biológicos é uma tarefa de extrema importância para as ciências biomédica e biotecnológica, entretanto, converter uma informação biológica em um sinal observável é considerada uma atividade desafiadora, especialmente devido à complexidade em se interfacear um componente eletrônico/óptico a um ambiente biológico (GRIESHABER et al., 2008). Um biossensor é justamente o dispositivo analítico capaz de fazer esta ponte, responsável por converter determinada informação biológica em sinal detectável (LOWE, 1984). Para funcionar, um biossensor depende da conexão de diversos componentes, ilustrados na Figura 4 abaixo. Os principais são:

1. a unidade receptora, capaz de interagir com o analito de interesse presente na amostra biológica;

2. a interface elétrica, composta de materiais como polímeros, nanopartículas e nanoestruturas metálicas, e com arquitetura específica onde ocorrerá um evento biológico capaz de emitir um sinal mensurável;
3. o elemento transdutor, onde o sinal gerado na etapa anterior será amplificado e convertido em sinal eletrônico por um circuito presente no equipamento de leitura;
4. o software que irá compilar, processar e plotar os dados obtidos (GRIESHABER et al., 2008).

Figura 4. Elementos básicos de um Biossensor. No primeiro quadro, há destaque para alguns exemplos de amostras para as quais os biossensores podem ser aplicados. No segundo e terceiro quadros, temos exemplos de unidades receptoras (a), de interfaces elétricas (b), amplificadores (c), processadores (d) e interfaces gráficas para exibição do(s) sinal(is) gerado(s) (e).



Fonte: GRIESHABER et al., 2008.

Um exemplo de grande sucesso dos biossensores é o sensor destinado à aferição da glicemia capilar. Medir com precisão os níveis sanguíneos de glicose é um procedimento extremamente importante para prevenção, diagnóstico e tratamento da Diabetes Mellitus, doença crônica caracterizada pelos níveis elevados e sustentados de açúcar no sangue devido a falta de produção do hormônio insulina (Tipo I) ou a resistência dos tecidos periféricos em responder à seus estímulos (Tipo II) (COLE; FLOREZ, 2020). A grande maioria dos biossensores destinados a esse fim utilizam o método eletroquímico baseado na atividade de enzimas como a glicose desidrogenase (GDH) ou a glicose oxidase (GOx), imobilizadas na superfície de um eletrodo. Quando a glicose está presente na amostra, sua oxidação gera um fluxo de elétrons que é detectado e amplificado pelo sistema. Como esse fluxo é diretamente proporci-

onal à concentração de glicose, é possível inferir sobre os níveis presentes na amostra (YOO; LEE, 2010). Gigantes farmacêuticas como a Abbott Laboratories e a divisão diagnóstica da Hoffmann–La Roche têm participação importante neste mercado (HWANG et al., 2018).

Dentro do universo dos biossensores existem os imunossensores, onde a unidade receptora imobilizada na superfície do transdutor é um antígeno ou um anticorpo, e a detecção do analito se baseia na interação dos mesmos. Antígenos e anticorpos são biomoléculas que possuem altíssima afinidade entre si, na faixa de $K_d = 10^{-12} \text{ mol.L}^{-1}$, e se aproveitar deste fato permite o desenvolvimento de sensores para uma ampla gama de aplicações que vão desde o diagnóstico clínico à pesquisa científica e aplicações industriais (vide abaixo). A imobilização de antígenos e anticorpos é possível em muitos materiais transdutores, o que garante diferentes formas de detecção do sinal gerado, com destaque para os imunossensores eletroquímicos (AIZAWA, 1991; KARUNAKARAN; PANDIARAJ; SANTHARAMAN, 2015).

Estes sensores possuem grande notoriedade pois somam a elevada sensibilidade das análises eletroquímicas com a elevada seletividade das interações antígeno-anticorpo. Os imunossensores eletroquímicos apresentam diversas vantagens como sua versatilidade e simplicidade de uso, uma vez que eles dispensam a necessidade de estruturas laboratoriais complexas e o operador pode fazer medições em praticamente qualquer lugar, os dados podem ser processados diretamente por meio de conexões com um telefone celular ou com um dispositivo simples, os resultados tendem a ser emitidos rapidamente usando pequenos volumes de amostras (na faixa dos microlitros), com elevadas sensibilidade e especificidade, e custo acessível (RICCARDI; COSTA; YAMANAKA, 2002; AYDIN; AYDIN; SEZGINTÜRK, 2019; CASTILHO, 2022).

Algumas aplicações de destaque para os imunossensores eletroquímicos:

- Detecção de Aflatoxina B1 em amostras de azeite de oliva (YU et al., 2015);
- Diagnóstico de Câncer de Ovário pela detecção antígeno CA-125 (GASPAROTTO et al., 2017);
- Diagnóstico simultâneo da Doença de Chagas e da Leishmaniose Visceral pela detecção de anticorpos anti-*Trypanosoma cruzi* e anti-*Leishmania infantum* (CORDEIRO et al., 2020);
- Diagnóstico de Hepatite C pela detecção direta do Vírus da Hepatite C (HCV) em amostra de soro (COSTA et al., 2019);
- Detecção da bactéria patogênica *Salmonella typhimurium* (SALAM; TOTHILL, 2008);

- Detecção de atrazina, um herbicida tóxico para seres humanos e animais em amostras ambientais (IONESCU et al., 2010);

Exemplificar tais aplicações é, na melhor das hipóteses, arranhar a superfície do vasto universo de aplicações e variações dos imunossensores eletroquímicos. Uma vez mapeados e caracterizados os epítomos antigênicos da proteína S, o último passo do trabalho foi testar uma plataforma baseada em sensores imunoeletroquímicos para a detecção de anticorpos anti-S.

1.6 Justificativa

Identificar com acurácia epítomos individuais em estruturas antigênicas é um processo exaustivo porém essencial para caracterizarmos novas doenças e a resposta imune associada, sendo o ponto de partida para o desenvolvimento de novas formas de prevenção, tratamentos e condutas clínicas. Dado que a pandemia de COVID-19, causada pelo vírus respiratório SARS-CoV-2, atingiu violentamente o Brasil e ocasionou irreparáveis perdas humanas, sociais e econômicas, uma resposta eficaz quanto ao manejo dos pacientes e ao desenvolvimento e distribuição de novas vacinas, bem como a avaliação da cobertura vacinal através da caracterização laboratorial da resposta anti-SARS-CoV-2 foi, permanece e será essencial.

Dado o *knowhow* de nosso grupo de pesquisa no desenvolvimento de biossensores para o diagnóstico sorológico e molecular de doenças virais, como a Hepatite C (KENFE et al., 2013; BIASOTO et al., 2019; COSTA et al., 2019), e considerando o exposto nesta breve introdução, o presente trabalho buscou testar um método combinatório utilizando algoritmos de predição de epítomos lineares de célula B que utilizem inteligência artificial para selecionar os provável epítomos presentes na proteína *Spike*; caracterizar estes sítios através de experimentos de Dinâmica Molecular; obtê-los via síntese orgânica de peptídeos e por fim aplicá-los no desenvolvimento de novas plataformas inovadoras baseadas em sensores eletroquímicos baratos e de alta sensibilidade tanto para a triagem de pacientes recém ou previamente infectados, quanto para a avaliação da resposta imunológica produzida por vacinas experimentais.

2 HIPÓTESE

Dada toda problemática da pandemia de COVID-19 apresentada, a hipótese central que guiou o desenvolvimento deste trabalho diz que é possível encontrar, em proteínas grandes e complexas como a glicoproteína *Spike* do vírus SARS-CoV-2, regiões peptídicas altamente específicas e com significativa antigenicidade para serem base do desenvolvimento de plataformas sorológicas de detecção de anticorpos, através de uma análise de bioinformática robusta, e utilizando-se de múltiplos métodos e estratégias imunológicas, bioquímicas e biofísicas computacionais.

3 OBJETIVOS

3.1 Geral

O objetivo principal do presente trabalho foi testar um método computacional para mapear e caracterizar sítios imunodominantes da glicoproteína estrutural *Spike* do vírus SARS-CoV-2 e aplicá-los no desenvolvimento de uma nova plataforma inovadora de detecção de anticorpos anti-S.

3.2 Específicos

- Determinar, através de estratégias *in silico* baseadas em inteligência artificial, quais são as regiões imunogênicas dominantes da proteína S do vírus SARS-CoV-2;
- Caracterizar propriedades bioquímicas e biofísicas inerentes aos sítios preditos que sejam relevantes para o processo de interação antígeno-anticorpo através de Dinâmica Molecular;
- Produzir, via Síntese de Peptídeos em Fase Sólida, alguns epítomos para os testes dos imunossensores;
- Avaliar a identidade estrutural e a pureza dos produtos de síntese através de Espectrometria de Massas e Cromatografia Líquida de Alta Eficiência (CLAE/HPLC);
- Desenvolver e testar uma plataforma com chips específicos funcionalizados com os peptídeos sintetizados de leitura baseada em Voltametria Cíclica com potencial aplicação para detecção de anticorpos anti-*Spike*;

4 MATERIAIS E MÉTODOS

4.1 Análises de bioinformática e imunoinformática da proteína *Spike*

4.1.1 Predição de epítomos lineares de célula B

Para determinar as prováveis regiões antigênicas dominantes a partir da sequência linear da proteína *Spike* de SARS-CoV-2 (*B-Cell Linear Epitope Prediction*), foram selecionados e aplicados três métodos computacionais baseados em inteligência artificial cujos resultados sejam sustentados por dados experimentais. Em um primeiro momento, a sequência de referência da proteína *Spike* (YP_009724390.1) foi obtida na plataforma GenBank no formato FASTA para ser submetida às análises.

O primeiro método utilizado foi o servidor DLBEpitope (<http://ccb1.bmi.ac.cn:81/dlbepitope/index.php?r=site%2Findex>), sendo preditos epítomos sem sobreposição de sequências com 22 aminoácidos de extensão (o que garante uma $AUC > 0.9$) (LIU; SHI; LI, 2020), e escolhidos todos aqueles que obtiveram um mínimo de 11 pontos (*score threshold* padrão = 11). O segundo método aplicado foi o servidor BepiPred 2.0 (<http://www.cbs.dtu.dk/services/bepipred/>), do qual foram selecionados todos os epítomos que superaram o *threshold* padrão de 0.5 (JESPERSEN et al., 2017). Como essa ferramenta atribui um *score* para cada aminoácido individual, não há um padrão de tamanho obtido para os epítomos. Como filtro, selecionamos todos aqueles com um mínimo de 12 aminoácidos de extensão. Caso um *gap* com no máximo três resíduos separasse uma região de outra, estas foram consideradas como um epítomo único. Por fim, o terceiro método utilizado foi o servidor iBCE-EL (<http://www.thegleelab.org/iBCE-EL/ProteinSearch.html>) (MANAVALAN et al., 2018), onde foram preditos epítomos com 22 aminoácidos de extensão e, uma vez que a aplicação não possui filtro de sobreposição de sequências, foram selecionadas todas aquelas com $Prob \geq 0.9$ como forma de filtragem, essas foram alinhadas e toda sua extensão foi considerada como um epítomo único.

4.1.2 Análise comparativa e identificação dos epítomos dominantes

Para identificar as regiões antigênicas imunodominantes da glicoproteína *Spike*, foi realizada uma análise comparativa entre todos os potenciais epítomos previstos individualmente nos três servidores. Para tal, suas sequências foram alinhadas à proteína S referência. Os epítomos identificados em pelo menos duas aplicações individuais com ao menos 12 aminoácidos de extensão foram considerados os epítomos lineares de células B dominantes na sequência da glicoproteína *Spike*. Um resumo da estratégia geral empregada em nosso estudo é apresentado em forma de fluxograma na Figura 5A.

4.1.3 Alinhamento, modelagem tridimensional e análises complementares

Todos os alinhamentos apresentados no trabalho foram realizados através do *software* T-COFFEE (<https://tcoffee.crg.eu/>), exportados no formato MAFFT e editados no servidor ESPript 3.0 (<https://esprpt.ibcp.fr/ESPript/ESPript/>) (NOTREDAME; HIGGINS; HERINGA, 2000; ROBERT; GOUET, 2014). O *software* UCSF ChimeraX (<https://www.cgl.ucsf.edu/chimera/>) foi utilizado para visualizarmos o modelo tridimensional da proteína S, identificarmos a localização dos sítios preditos e obtermos algumas de suas propriedades, além da captura das figuras (PETTERSEN et al., 2020). A conservação estrutural dos epítomos imunodominantes frente a proteína S do vírus SARS-CoV-1 (YP_009825051.1) foi obtida através do servidor IEDB Epitope Conservancy Analysis (<http://tools.iedb.org/conservancy/>) utilizando as configurações padrão (BUI et al., 2007). A massa molecular e outras informações a respeito da química dos epítomos peptídicos foi obtida na ferramenta CEM Peptide Calculator (<https://cem.com/en/peptide-calculator>).

4.2 Simulações atomísticas via Dinâmica Molecular

4.2.1 Preparo do modelo da proteína *Spike*

Para o estudo atomístico, um modelo da proteína *Spike* de SARS-CoV-2 foi gerado baseado na estrutura 6VXX_1_1_1 publicada por WOO e colaboradores (2020). Esta estrutura cobre toda a extensão da proteína S (aa 1 – 1173) e apresenta as três unidades RDB no estado fechado (RBD DOWN), além de possuir 23 pontos de glicosilação em cada protômero que compõe o trîmero *Spike* maduro e de estar ancorada em um envelope viral. 6VXX_1_1_1 foi obtida diretamente no repositório COVID-19 da plataforma CHARMM-GUI (<https://charmm-gui.org/?doc=archive&lib=covid19>). Como o objetivo é estudar a porção onde os epítomos aqui mapeados estão localizados, foi necessário alterar modelo original. O primeiro tratamento feito na estrutura foi remover o peptídeo sinal (aa 1 – 13), e em seguida, cortar na posição 1178 para removermos as porções transmembrana (aa 1213 – 1236) e intracelular (aa 1237 – 1273) da proteína S e a membrana. Por fim, um último corte foi feito entre as posições R685 e S686 para simular a clivagem pela protease furina. Todos os cortes foram igualmente aplicados aos três protômeros. Também foi utilizado o modelo 6VSB_1_1_1 para os cálculos de SASA. As modificações foram igualmente aplicadas.

Para gerar os arquivos de entrada do Amber foi utilizado o servidor online CHARMM-GUI Solution Builder (JO et al., 2008), que já possuía todos os grupos químicos parametrizados nos campos de força Amber, incluindo as glicosilações. A topologia foi construída em uma caixa de simulação ortogonal com 10 Å de espaço da superfície da proteína até a borda da caixa. O modelo original continha moléculas de água TIP3P solvatando a proteína juntamente com uma solução de KCl 0,15 M, que foram mantidas. A Tabela 1 apresenta as informações gerais dos sistemas parametrizados para nossos experimentos de MD.

Tabela 1. Composição dos sistemas parametrizados para os experimentos de Dinâmica Molecular.

Componente*	<i>Spike</i> (6VSB e 6VXX)	E5	E6	E7	E8	E9
Proteína	53.811	200	340	391	173	236
Carboidratos	13.068	-	-	-	-	-
Íons	2.296	1	1	4	2	1
Água	1.205.220	15.049	48.177	35.490	12.465	15.795
Total	1.274.395	15.250	48.518	35.885	12.640	16.032

* Valores apresentados em n° de átomos.

Fonte: o autor.

4.2.2 Preparo dos modelos dos epítomos E5 – E9

Para gerar os arquivos de entrada de cada epítipo individual para a execução dos experimentos de MD, suas coordenadas atômicas foram salvas diretamente da trajetória do sistema 6VXX em um novo arquivo .pdb utilizando o *software* VMD 1.9.3 (HUMPHREY; DALKE; SCHULTEN, 1996). O ponto de tempo selecionado para este processo foi o último *frame* da trajetória. Apenas resíduos de aminoácido foram salvos, simulando um epítipo sintetizado quimicamente. Uma caixa cúbica foi construída em torno do peptídeo, solvatada com água TIP3P e as cargas foram balanceadas com íons de sódio (Na⁺) e cloro (Cl⁻) quando necessário (JORGENSEN et al., 1983). O preparo do sistema foi feito via módulo *tleap*. Uma vez configurado, a topologia foi gerada no Amber pelo módulo *pmemd* (SALOMON-FERRER; CASE; WALKER, 2012). As características dos sistemas também estão dispostas na Tabela 1.

4.2.3 Minimização da energia, termalização do sistema e produção das dinâmicas

No presente estudo, todos os experimentos de MD foram realizados no software Amber 18 usando campos de força ff14SB e GLYCAN_06j para proteínas e carboidratos, respec-

tivamente (KIRSCHNER et al., 2007; SALOMON-FERRER; CASE; WALKER, 2012; MAIER et al., 2015). O sistema *Spike* foi submetido à minimização de energia usando um algoritmo de descida íngreme (*steepest descent algorithm*), aquecido a 310,15 K (37 °C) usando um termostato Langevin, e a pressão foi balanceada para 1 atm usando o baróstato Berendsen. Finalmente, um MD de 100 ns foi gerada com um passo de tempo de 2 fs. As posições e velocidades atômicas foram salvas em arquivos de trajetória a cada 10.000 passos (o que equivale a 1 *frame*). O mesmo protocolo foi usado para realizar as execuções de MD para os epítomos individuais durante 1 μ s a 310,15 K. Os parâmetros de entrada para a execução da minimização, da termalização e da produção podem ser encontrados no Apêndice B.

4.2.4 Análise das trajetórias

Para analisar os dados das trajetórias, usamos o módulo *cpptraj* do Amber e o *software* VMD 1.9.3 juntamente com *scripts* tcl próprios (ROE; CHEATHAM, 2013). As figuras foram criadas usando ambos UCSF ChimeraX 1.3 e VMD 1.9.3. Gráficos e análise estatísticas básicas foram feitos usando o SciDAVis 2.4.1 (<http://scidavis.sourceforge.net/>).

4.3 Produção e caracterização dos epítomos

4.3.1 Reagentes

Todos os reagentes e solventes empregados nas etapas de síntese, clivagem e análise dos peptídeos foram obtidos da Merck/Sigma-Aldrich nos graus ACS *Reagent* e HPLC.

4.3.2 Síntese química de Peptídeos em Fase Sólida (SPPS)

As regiões peptídicas que correspondem aos sítios E5 – E9, encontradas no RBD da glicoproteína *Spike*, bem como dois controles (Ct_01 e Ct_2), foram sintetizadas por meio da estratégia Fmoc/tBu em fase sólida, de maneira manual em reatores e com modificações. Os detalhes seguem:

Uma resina Rink Amida (Merck) com grau de substituição de 0,4 – 0,6 (meq/g) foi pesada e acondicionada em reator de polipropileno (12,5 mL) com membrana de filtro de poliestireno. A resina foi então lavada três vezes com dimetilformamida (DMF), três vezes diclorometano (DCM), e expandida em DCM. A remoção do protetor Fmoc dos aminoácidos foi realizada com solução de 4-metilpiperidina a 20% em DMF e 1% Triton-X100 durante 15 minutos. Após cada desproteção e antes de cada teste de ninidrina, o mesmo processo de lavagem se repetiu (3x1' DMF e 3x1' DCM). Para o acoplamento de cada aminoácido-Fmoc foi empregada uma estratégia básica que consiste em solução contendo o aminoácido-Fmoc e os ativadores DIC (N,N'-Diisopropilcarbodiimida) e HOBt (1-Hidroxibenzotriazol), dissolvidos em DCM/DMF. A reação ocorreu de 1 – 2h em estufa a 40 °C. Todos os reagentes foram pesados em excesso (3x), sendo a capacidade máxima de carregamento da resina a condição limitante. A conclusão do acoplamento foi atestada pelo teste da ninidrina (teste de Kaiser) com uma pequena quantidade de resina extraída do reator. Em caso de reação incompleta, também estavam disponíveis os ativadores HBTU [2-(1-H-benzotriazol-1-il)-1,1,3,3-tetrametiluronio hexafluorofosfato], HCTU [*O*-(1-H-6-clorobenzotriazol-1-il)-1,1,3,3-tetrametiluronio hexafluorofosfato, OxymaPure® e DIPEA (*N,N*-diisopropiletilamina). Ao fim do processo de síntese a resina com o peptídeo foi desprotegida, lavada com DMF/DCM, seca e armazenada em tubos Falcons fechados em geladeira até o momento da clivagem.

4.3.3 Clivagem

Para a clivagem dos peptídeos da resina, uma massa dessa foi pesada e suspendida em solução de ácido trifluoroacético (TFA)/triisopropilsilano (TIS)/3,6-dioxa-1,8-octanaditiol (DOT)/água ultrapura na proporção 92,5:2,5:2,5:2,5, respectivamente, em tubo Falcon. Os tubos foram centrifugados em velocidade máxima para forçar toda resina a se depositar junto a

solução e a reação ocorreu, sob agitação dos reatores, durante 3 horas. Após esse período, a solução que contém os peptídeos foi separada da resina por filtração em seringa, os peptídeos clivados foram precipitados com 10 mL éter dietílico (Et₂O) frio durante 30 minutos em freezer, a massa peptídica crua foi lavada 5 vezes com 5 mL de éter frio, o éter restante foi evaporado. Por fim, os peptídeos foram dissolvidos em água ultrapura e finalmente liofilizados.

4.3.4 Análise da identidade molecular por Espectrometria de Massas (CID-MS2)

Uma pequena porção das amostras liofilizadas foi solubilizada em solução de água deionizada e ácido trifluoroacético (TFA) a 0,05% (v/v) e aplicada com auxílio do injetor automático (SIL- 20AXR, Shimadzu) em um sistema de *Ultra Fast Liquid Chromatography* (UFLC) (Nexera-Shimadzu Corporation, Tokyo, Japan) com fluxo de 0,2 mL/min. As análises por espectrometria de massas foram realizadas com a utilização de um sistema Electrospray - Ion trap (ESI/IT), modelo Amazon SL (Bruker Daltonics, Bremen, Germany). Para o controle de aquisição dos dados, foram utilizados os *softwares* HyStar e Trap Control (ambos BRUKER Daltonics), e LC Solution (SHIMADZU). O espectrômetro de massas foi operado no modo íon positivo, com *scan* em um intervalo de 50 a 2200 m/z. As análises de espectrometria de massas sequenciais, que visam obter os espectros de fragmentação em condições de decomposição induzida por colisão (CID-MS2) foram realizadas utilizando Hélio (He) como gás de colisão, a uma pressão de 100 kPa. O *software* Data Analysis (BRUKER Daltonics) foi utilizado para análise de dados. A predição das massas teóricas e da fragmentação dos peptídeos foi feita no software CEM Peptide Calculator (<https://cem.com/en/peptide-calculator>).

4.3.5 Determinação do perfil cromatográfico produtos de síntese por HPLC

Para caracterizarmos a pureza dos peptídeos sintetizados e estabelecermos a necessidade de posterior purificação dos mesmos, seu perfil cromatográfico foi obtido por Cromatografia Líquida de Alta Eficiência (CLAE/HPLC) em cromatógrafo Perkin Elmer empregando uma coluna Supelco Discovery® BIO Wide Pore C18 com dimensões de 250 x 4,6 mm, partí-

cula de 5 μm de diâmetro e poro de 300 Å. Foram utilizados como fases móveis água ultrapura (MiliQ) contendo 0,04% de TFA (solvente A) e acetonitrila (ACN) contendo 0,03% de TFA (solvente B). O experimento consistiu em uma corrida de gradiente exploratório de 5 a 95% de solvente B durante 30 min, seguido de 15 min a 100% de solvente B, com fluxo de 1,0 mL/min e detecção no ultravioleta em $\lambda = 220 \text{ nm}$.

4.4 Desenvolvimento e teste de imunossensor eletroquímico para detecção de anticorpos anti-S

4.4.1 Sensores

Para os testes envolvendo o desenvolvimento dos imunossensores foram adquiridos os Sensores de Ouro Serigrafado modelo 220AT da empresa Metrohm DropSens (Asturias, Espanha). Cada sensor é formado por um transdutor eletroquímico com três eletrodos, sendo o eletrodo de trabalho e o contra-eletrodo serigrafados em ouro e o eletrodo de referência serigrafado em prata.

4.4.2 Lavagem do sensor e imobilização dos epítomos

Antes de utilizar os sensores é necessário limpar a superfície dos eletrodos para garantir que os processos de funcionalização e leitura não serão afetados por quaisquer sujidades decorrentes da produção e/ou transporte. Todos os eletrodos foram devidamente lavados antes de serem utilizados nos testes. A primeira etapa de lavagem consistiu na limpeza manual da superfície metálica dos três eletrodos com cotonete estéril do tipo *swab* embebido em acetona pura seguido de isopropanol. Em seguida, os sensores foram mergulhados em isopropanol puro e levados ao banho de ultrassom por 30 minutos. Após o banho, eles foram secos com gás nitrogênio e levados a um novo banho de ultrassom, desta vez mergulhados em água ultrapura, por mais 30 minutos. Por fim, eles foram secos com gás nitrogênio e armazenados em dessecador até o momento do uso.

Para os testes, apenas o eletrodo de trabalho de ouro foi revestido com os peptídeos de interesse produzidos na etapa de síntese através processo de Adsorção por Monocamadas Automontadas. Em um primeiro momento, foram adicionados 10 µL de solução 20 mM de cistamina, e os eletrodos foram incubados à 37 °C em câmara úmida. Em seguida, o excesso de cistamina foi lavado com água ultrapura e os eletrodos secos com gás nitrogênio. Foram adicionados então 10 µL de solução de glutaraldeído a 5%, passo seguido de nova incubação a 37 °C em câmara úmida. Uma nova lavagem com água ultrapura foi feita após incubação e os eletrodos foram secos com nitrogênio. Por fim, foram adicionados 10 µL de solução contendo 10 µg dos peptídeos em água ultrapura e o sistema foi incubado em geladeira (4 °C) *overnight*. Uma última etapa de lavagem com água ultrapura e secagem com nitrogênio se seguiu.

4.4.3 Incubação com os soros e leitura imunossensores

Após funcionalização, 10 µL dos soros controles comerciais disponíveis foram aplicados sobre o eletrodo de trabalho e estes foram incubados em câmara úmida à 37 °C por 30 minutos. Nesta validação inicial, como controle negativo foi utilizada uma solução de PBS pH 7.4. O excesso de soro/PBS foi lavado com água ultrapura e os eletrodos novamente secos com gás nitrogênio. Os Voltagramas Cíclicos (CV) foram obtidos no aparelho Potenciostato Sensit BT (PalmSens, Houten, Países Baixos). Os eletrodos foram mergulhados em solução de PBS contendo 5 mM de ferri/ferrocianeto de potássio $[\text{Fe}(\text{CN})_6]^{3-/4-}$ como mediador redox e a tensão variou de -0,5 a +0,5 V para medição da corrente, a uma taxa de varredura de 1000 mV/s.

5 RESULTADOS E DISCUSSÃO

5.1 Mapeando os determinantes antigênicos da glicoproteína *Spike*

Tomando como ponto de partida a crise causada pela pandemia de COVID-19, o presente trabalho objetivou a identificação e seleção racional de potenciais epítomos lineares de célula B na proteína estrutural *Spike* do vírus SARS-CoV-2. Para tal, alguns métodos computacionais de predição destes epítomos foram considerados. Os dois principais critérios utilizados na seleção dos métodos que iriam compor a análise foram: (1) ser baseado em algoritmos de inteligência artificial/aprendizado de máquina e (2) ter sido treinado com dados experimentalmente validados. Três aplicações foram selecionadas.

O primeiro método utilizado foi o servidor DLBEpitope, que emprega um algoritmo recentemente publicado baseado em redes neurais profundas *feedforward* (*feedforward deep neural networks*) composto de 11 classificadores independentes para pontuar a sequência alvo e determinar a probabilidade da mesma pertencer à um epítomo. Os classificadores foram treinados a partir de um *set* de composto por cerca de 25 mil epítomos e 215 mil não-epítomos (LIU; SHI; LI, 2020). O segundo método aplicado foi o servidor BepiPred 2.0, algoritmo baseado em regressão de floresta randômica (*random forest regression*) treinado com um conjunto de 649 cristais de complexos antígeno-anticorpo, totalizando 3.542 resíduos de aminoácidos compondo epítomos e 36.785 resíduos não-epítomos (JESPERSEN et al., 2017). Por fim, o terceiro método utilizado foi o servidor iBCE-EL, que é baseado em um modelo de dois algoritmos independentes para formar seus classificadores: incremento de gradiente (*gradient boosting*) e árvore altamente randomizada (*extremely randomized tree*), e ainda conta com um dataset de mais de 12 mil epítomos validados experimentalmente que fundamentam a predição (MANAVALAN et al., 2018). Os dados brutos originados durante o processo de predição podem ser consultados no Apêndice A.

Tanto para o servidor DLBEpitope quanto para o servidor iBCE-EL foram preditos epítomos com 22 aminoácidos de extensão, um tamanho que considera a capacidade de nosso sistema de síntese de peptídeos e a análise feita pelos desenvolvedores do método DLBEpitope, que determinaram a extensão ótima (e padrão de seu método) para predição como 38 aminoácidos (AUC ~0.95), sendo que a extensão de 22 aminoácidos mantém uma AUC > 0.9 (LIU; SHI; LI, 2020).

A primeira análise, utilizando a ferramenta DLBEpitope nas condições acima detalhadas retornou um total de 25 regiões únicas, não sobrepostas (DL_p1 – DL_p25) e potencialmente imunogênicas espalhadas por toda proteína S, estando 14 delas localizadas na subunidade S1, e 11 na subunidade S2. Cinco potenciais epítomos foram identificados no domínio RBD (DL_p6 – DL_p11), um dos mais interessantes para a ocorrência de interações antígeno-anticorpo devido a sua localização e principalmente a sua função. Todos os sítios contém 22 aminoácidos de extensão.

Na análise com o servidor BepiPred 2.0, por sua vez, buscamos selecionar regiões não sobrepostas com mais de 12 aminoácidos de extensão. Aqui, a marca mínima de 12 aminoácidos foi considerada pois nesta aplicação não é possível escolher o tamanho do epítopo predito. Este mínimo também foi escolhido a partir das observações de LIU e colaboradores (2020), que demonstraram que epítomos com 12 aminoácidos de extensão mantêm uma AUC de 0.8 em seus testes de predição. Outra consequência da obtenção de regiões com tamanhos diversos foi a separação de alguns sítios por poucos aminoácidos tidos como não suficientemente imunogênicos para ultrapassarem o *score threshold* determinado. Para não isolar potenciais epítomos muito próximos uns dos outros e que sozinhos poderiam não alcançar a marca mínima de 12 aminoácidos, decidimos levar em consideração o tamanho do *gap* entre eles.

As regiões BP_p7*, BP_p8*, BP_p11*, BP_p12* e BP_p15** foram obtidas a partir da união de duas regiões independentes de tamanhos variados, separadas por não mais que três aminoácidos com *scores* inferiores ao mínimo, e que formaram, juntas, um potencial epítopo único com mais de 12 aminoácidos de extensão. O epítopo BP_p15**, em específico, compreende em sua sequência o sítio de clivagem de furina, zona de divisão entre as subunidades S1 e S2 (HOFFMANN; KLEINE-WEBER; POLHMANN, 2020). No total, 19 regiões de tamanhos variados (entre 12 e 42 aminoácidos de extensão) foram selecionadas desse servidor, estando 14 delas na região S1, uma no *loop* de clivagem por furina e 5 na região S2.

Por fim, na análise através da aplicação iBCE-EL, uma questão parecida foi constatada. Apesar de sermos capazes de escolher o tamanho do epítopo predito, a ferramenta não possui filtro de sobreposição de sequências, logo, foram obtidas 1191 sequências com 22 aminoácidos de extensão, cada qual com seu *score*, e diferentes entre si por apenas um aminoácido. Como primeiro filtro para estes dados, foram selecionadas as 22 regiões que ultrapassaram a marca de 0.9 de probabilidade de representarem um epítopo de célula B (sequências 19 – 20, 69, 136, 302, 436, 438, 441, 457, 459, 1114 – 1120, 1123 – 1127, Apêndice A, Planilha iBCE-EL). Uma vez que o iBCE-EL não possui filtro de sobreposição, os sítios foram alinhados à sequência de referência da proteína *Spike* na ferramenta T-COFFEE. As sequências re-

sultantes das sobreposições que ocorreram entre os 22 sítios iniciais foram consideradas como epítomos únicos, uma vez que toda sua extensão alcança a marca de 0.9 de probabilidade, e, ao todo, 6 prováveis epítomos únicos (não sobrepostos entre si) resultaram desta análise, estando 5 deles na subunidade S1 e apenas um na subunidade S2.

Todas essas considerações e filtragem dos dados foram imprescindíveis para chegarmos à etapa da análise comparativa proposta. No Apêndice A podem ser encontrados todos os dados brutos resultantes das predições iniciais (planilhas DLBEpitope, BepiPred 2.0 e iBCE-EL). A Tabela 2 apresenta os potenciais epítomos obtidos após as filtrações e considerações aplicadas. Na Figura 5A pode ser encontrado um fluxograma que detalha o raciocínio da análise. O Apêndice B apresenta o alinhamento feito para os epítomos identificados no *software* iBCE-EL.

Tendo em mãos as 50 sequências obtidas nas ferramentas de predição, uma análise comparativa entre elas foi realizada. Essa análise consistiu no alinhamento das mesmas à sequência de referência da proteína S, como ilustrado na Figura 5C. Dos 50 prováveis epítomos únicos de partida foi possível identificar 13 regiões altamente imunogênicas (E1 – E13, Tabela 3). Tais regiões aparecem em ao menos duas aplicações individuais e possuem um comprimento de 12 ou mais resíduos de aminoácido e foram consideradas como sendo os sítios imunodominantes da proteína *Spike*. Cinco regiões foram identificadas individualmente nos aplicativos iBCE-EL e BepiPred 2.0, cinco nos aplicativos DLBEpitope e BepiPred 2.0 e três regiões foram identificadas nas três aplicações (Figura 5B). O Alinhamento completo da análise comparativa pode ser encontrado no Apêndice C.

Tabela 2. Peptídeos possivelmente imunogênicos obtidos nas plataformas DBLEpitope, BepiPred 2.0 e iBCE-EL.

Application	Peptide ID	Spike Subunit	Mapped		Peptide Sequence	Epitope Lenght ¹
			Start	End		
<u>iBCE-EL</u>	EL_p1	S1	20	42	TRTQLPPAYTNSFTRGVYYPDKV	23
	EL_p2	S1	70	91	VSGTNGTKRFDNPVLPFNDGVY	22
	EL_p3	S1	137	158	NDPFLGVYYHKNNKSWMESEFR	22
	EL_p4	S1	303	324	LKSFTVEKGIYQTSNFRVQPT	22
	EL_p5	S1	437	481	NSNNLDSKVGGNYNYLYRLFRKSNLKPFERDISTEIQAGSTPCN	45
	EL_p6	S2	1129	1163	VIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPD	35
<u>BepiPred 2.0</u>	BP_p1	S1	14	36	QCVNLTTTRTQLPPAYTNSFTRGV	23
	BP_p2	S1	59	81	FSNVTWFHAIHVSGTNGTKRFDN	23
	BP_p3	S1	141	163	LGVYYHKNNKSWMESEFRVYSSA	23
	BP_p4	S1	178	191	DLEGKQGNFKNLRE	14
	BP_p5	S1	207	222	HTPINLVRLDLPQGFS	16
	BP_p6	S1	248	260	YLTPGDSSSGWTA	13
	BP_p7*	S1	306	322	FTVE KGI YQTSNFRVQP	17
	BP_p8*	S1	331	356	NITNLC P FGEVFNATRFASVYAWNRRK	26
	BP_p9	S1	370	395	NSASFSTFKCYGVSPTKLNDLCFTNV	26
	BP_p10	S1	404	424	GDEVQRQIAPGQTGKIADYNYK	21
	BP_p11*	S1	439	478	NNLDSKVGGNYNY LYR LFRKSNLKPFERDISTEIQAGST	40
	BP_p12*	S1	483	502	VEGFNCYFPLQ SYGF QPTNG	20
	BP_p13	S1	516	535	ELLHAPATVCGPKKSTNLVK	20
	BP_p14	S1	626	645	ADQLTPTWRVYSTGSNVFQT	20

Application	Peptide ID	Spike Subunit	Mapped		Peptide Sequence	Epitope Lenght ¹
			Start	End		
<u>BepiPred 2.0</u>	BP_p15**	S1/S2	673	709	SYQTQTNSPRRAR↓SVASQSIIAYTMSLGAENSVAYSN	37
	BP_p16	S2	786	800	KQIYKTPPIKDFGGF	15
	BP_p17	S2	828	843	LADAGFIKQYGDCLGD	16
	BP_p18*	S2	1133	1174	VNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGI NA	42
	BP_p19	S2	1252	1270	SCCKFDEDDSEPVLKGVKL	19
<u>DLBEpitope</u>	DL_p1	S1	21	42	RTQLPPAYTNSFTRGVYYPDKV	22
	DL_p2	S1	117	138	LLIVNNATNVVIKVCEFQFCND	22
	DL_p3	S1	151	172	SWMESEFRVYSSANNCTFEYVS	22
	DL_p4	S1	186	207	FKNLREFVFKNIDGYFKIYSKH	22
	DL_p5	S1	213	234	VRDLPQGFSALEPLVDLPIGIN	22
	DL_p6	S1	322	343	PTESIVRFPNITNLCPFGEVFN	22
	DL_p7	S1	373	394	SFSTFKCYGVSPTKLNDLCFTN	22
	DL_p8	S1	416	437	GKIADYNYKLPDDFTGCVIAWN	22
	DL_p9	S1	445	466	VGGNYNYLYRLFRKSNLKPFER	22
	DL_p10	S1	476	497	GSTPCNGVEGFNCYFPLQSYGF	22
	DL_p11	S1	527	548	PKKSTNLVKNKCVNFNFNGLTG	22
	DL_p12	S1	597	618	VITPGTNTSNQVAVLYQDVNCT	22
	DL_p13	S1	623	644	AIHADQLTPTWRVYSTGSNVFQ	22
	DL_p14	S1	654	675	EHVNNSYECDIPIGAGICASYQ	22
	DL_p15	S2	747	768	TECSNLLLQYGSFCTQLNRALT	22

Application	Peptide ID	Spike Subunit	Mapped		Peptide Sequence	Epitope Length ¹
			Start	End		
DLBEpitope	DL_p16	S2	823	844	FNKVTLADAGFIKQYGDCLGDI	22
	DL_p17	S2	894	915	LQIPFAMQMAYRFNGIGVTQNV	22
	DL_p18	S2	920	941	QKLIANQFNSAIGKIQDSLST	22
	DL_p19	S2	955	976	NAQALNTLVKQLSSNFGAIVS	22
	DL_p20	S2	988	1009	EAEVQIDRLITGRLQSLQTYVT	22
	DL_p21	S2	1011	1032	QLIRAAEIRASANLAATKMSEC	22
	DL_p22	S2	1057	1078	PHGVVFLHVTYVPAQEKNFTTA	22
	DL_p23	S2	1150	1171	EELDKYFKNHTSPDVLGDISG	22
	DL_p24	S2	1199	1220	DLQELGKYEQYIKWPWYIWLGF	22
	DL_p25	S2	1241	1262	CSCLKGCCSCGSCCKFDEDDSE	22

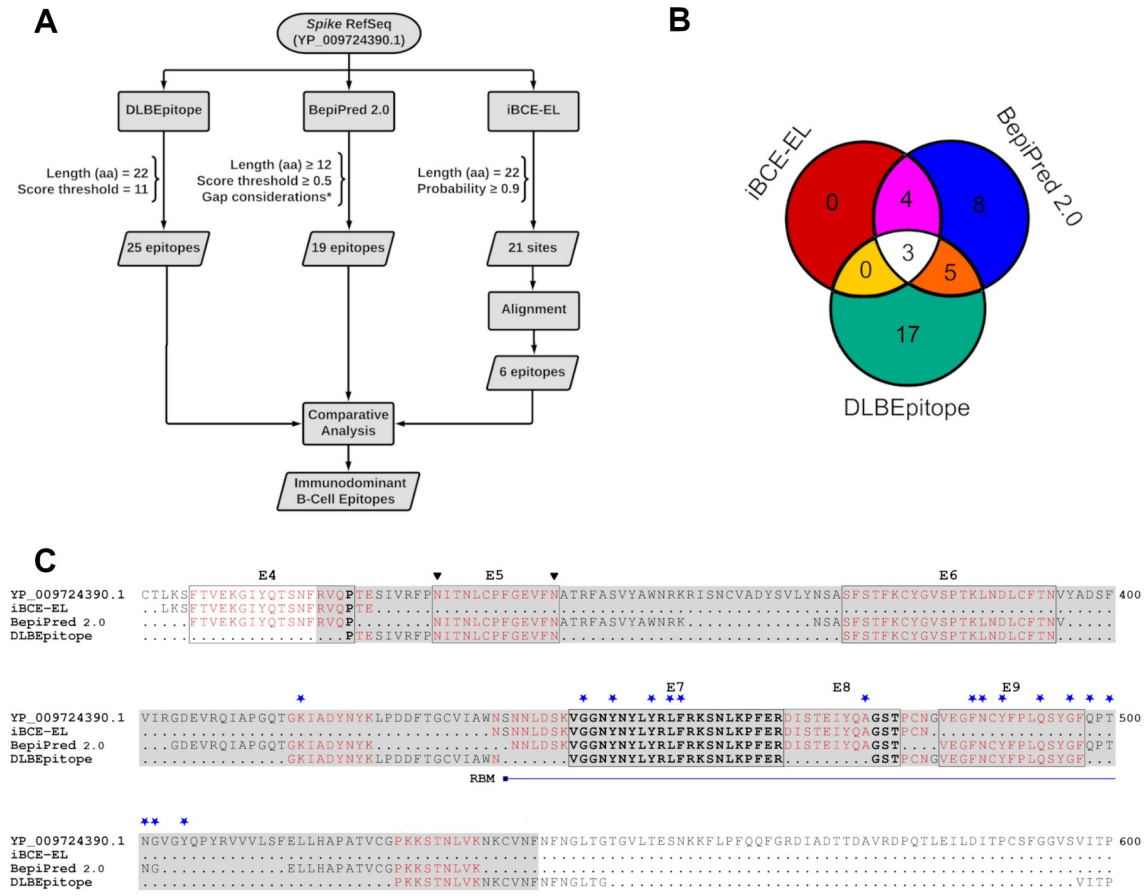
* Sítios obtidos a partir da união de duas ou mais regiões independentes, separadas por não mais que três aminoácidos de extensão, que juntas formam um epítipo com mais de 12 aminoácidos de extensão. Os aminoácidos marcados em **vermelho** são os aminoácidos considerados para a união das regiões independentes.

** Provável epítipo em que se encontra o sítio de clivagem de furina onde ocorre a separação da região S1 e da região S2 da proteína (o local da clivagem está marcado com o símbolo ↓).

¹ Em n° de aminoácidos.

Fonte: o autor.

Figura 5. Estratégia de mapeamento dos potenciais epítomos lineares de célula B. **A)** Fluxograma que condensa as principais etapas do método proposto. **B)** Diagrama de Venn indicando a origem e a quantidade de potenciais epítomos preditos individualmente e entre aplicações. **C)** Secção ilustrativa do alinhamento comparativo aplicado com destaque para a região RBD da proteína S (área de fundo cinza). Os epítomos são destacados nas caixas retangulares. Resíduos identificados em duas aplicações diferentes são mostrados em vermelho, e os identificados nas três, em negrito. A linha roxa indica a região RBM, interna ao RBD, e as estrelas azuis os resíduos da proteína S que fazem interação com resíduos do receptor ACE2, de acordo com observações de LAN e colaboradores (2020).



Fonte: o autor.

Tabela 3. Epítomos de Célula B imunodominantes da proteína *Spike* encontrados na análise combinatória.

Epitope	Subunit (Domain)	Start	End	Sequence	Lenght	Glycosylated?	Conservation (%)	
							SARS-CoV-1	MERS-CoV
E1	S1 (NTD)	21	36	RTQLPPAYTNSFTRGV	16	-	31,2	31,2
E2	S1 (NTD)	70	81	VSGTNGTKRFDN	12	N74	41,7	33,3
E3	S1 (NTD)	141	158	LGVYYHKNNKSWMESEFR	18	N149	22,2	27,8
E4	S1	306	322	FTVEKGIYQTSNFRVQP	17	-	76,5	47,1
E5	S1 (RBD)	331	343	NITNLCPFGEVFN	13	N331, N343	100	38,5
E6	S1 (RBD)	373	394	SFSTFKCYGVSPTKLNDLCFTN	22	-	86,4	27,3
E7	S1 (RBD/M)	445	466	VGGNYNYLYRLFRKSNLKPFR	22	-	59,1	22,7
E8	S1 (RBD/M)	467	478	DISTEIQAGST	12	-	41,7	41,7
E9	S1 (RBD/M)	483	497	VEGFNCYFPLQSYGF	15	-	53,3	33,3
E10	S1	626	644	ADQLTPTWRVYSTGSNVFQ	19	-	84,2	31,6
E11	S2	828	843	LADAGFIKQYGDCLGD	16	-	87,5	37,5
E12	S2	1133	1149	VNNTVYDPLQPELDSFK	17	N1134	94,1	35,3
E13	S2	1150	1163	EELDKYFKNHTSPD	14	N1158	100	42,9

Fonte: o autor.

5.2 Características gerais dos epítomos preditos

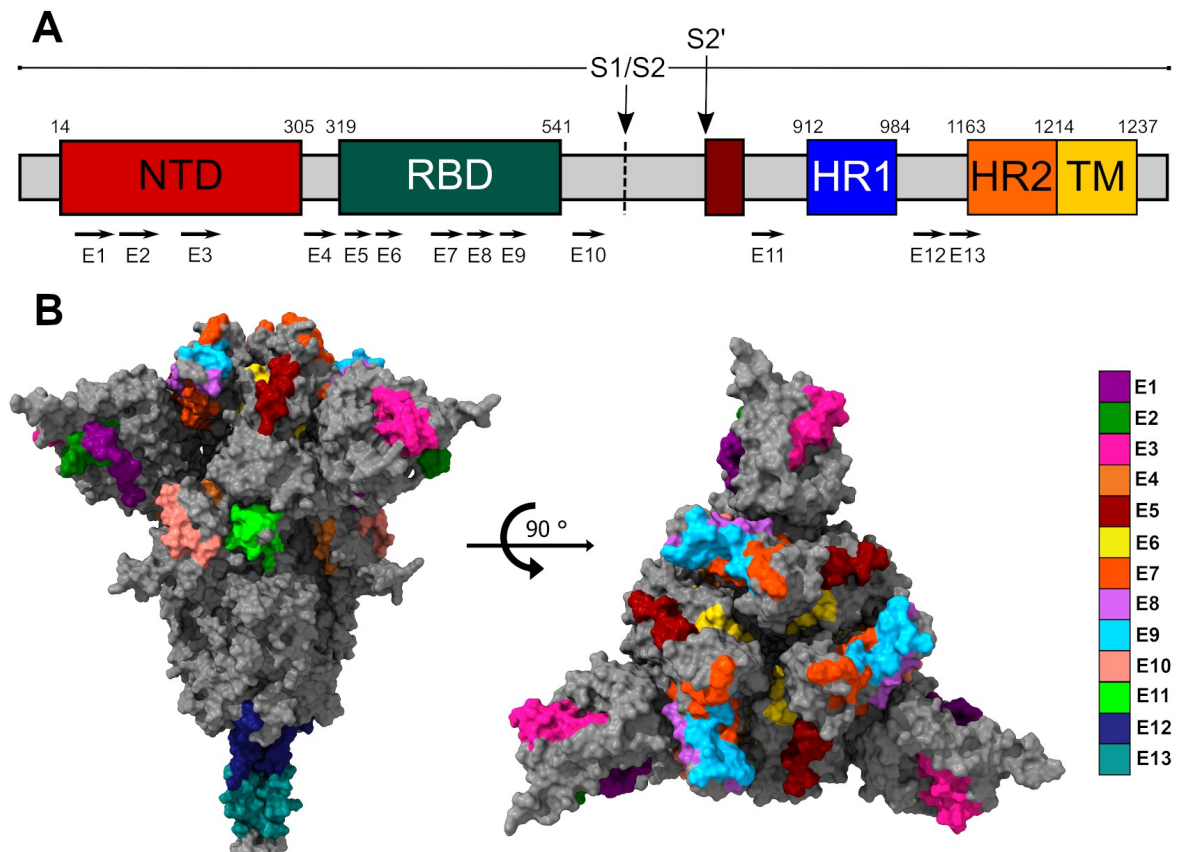
Dos 13 epítomos imunodominantes identificados em nossa análise combinatória, 10 (E1 – E10) encontram-se na subunidade S1 da proteína, estando três deles (E1 – E3) no Domínio N-Terminal (NTD, 14 – 305), cinco (E5 – E9) na Região de Ligação ao Receptor (RBD, 319 – 541), um entre o NTD e o RBD (E4) e um entre o RBD e o sítio de clivagem de furina (E10). Os três sítios subsequentes (E11 – E13) encontram-se na subunidade S2 próximos aos *Heptad Repeats* (HR) 1 (E11) e 2 (E12 e E13) (Figura 6A). É interessante observar que a maior população de determinantes imunogênicos estar presente na subunidade S1 é um fato que vai de encontro com a literatura, uma vez que esta porção da proteína S é conhecida-mmente mais imunogênica e estruturalmente acessível aos anticorpos (WRAPP et al., 2020; JU et al., 2020; CHI et al., 2020).

Também é na subunidade S1, mais precisamente no RBD, onde o primeiro evento necessário à infecção ocorre: o reconhecimento do receptor ACE2. Anticorpos direcionados à essa região possuem o potencial de neutralizar o vírus e são extremamente importantes para o desenvolvimento da imunidade natural, aquela obtida via infecção, e da imunidade vacinal, sendo avaliados durante os testes dos candidatos à antígeno vacinal e nos estudos da resposta à infecção (BARNES et al., 2020; LI et al., 2022). Além disso, tais anticorpos possuem o potencial de serem aplicados no desenvolvimento de soros hiperimunes e de biofármacos voltados ao combate da infecção, estando alguns deles, como o casirivimab e o imdevimab, já disponíveis para uso emergencial (FDA, 2020). Com isso, os epítomos encontrados no domínio RBD, especialmente aqueles encontrados no RBM (*Receptor Binding Motif*) que compreendem resíduos envolvidos na interação RBD-ACE2 (E7 – E9), receberão atenção especial e serão direcionados para análises mais detalhadas (Figura 5C).

Outro dado interessante é a identidade estrutural dos epítomos comparada com os dois principais coronavírus de interesse médico: seu antecessor SARS-CoV-1 e MERS-CoV. Para a proteína S de SARS-1 (YP_009825051.1), a identidade variou de 22 à 100%, com média próxima a 60%, inferior a conservação geral de ~80% observada entre as duas proteínas (LAN et al., 2020). Para a proteína S de MERS-CoV, a identidade variou de 22,7 a 47,1%, com média foi de 34,6%, uma vez que esse vírus é evolutivamente mais distante tanto de SARS-1 quanto de SARS-2 (GORBALENYA et al., 2020). Epítomos idênticos aos encontrados em outros vírus, como é o caso do E5 e do E13, ou mesmo muito parecidos, como o E12 (ambos casos em relação à SARS-1), podem resultar em reações cruzadas com anticorpos for-

mados contra os vírus em questão, fato extremamente relevante para, por exemplo, flavivírus como o vírus da Dengue e o da ZIKA (ROCKSTROH et al., 2017). Mesmo que SARS-CoV-1 e MERS-CoV não sejam realidade em território Brasileiro, em nosso caso, buscamos epítomos com o menor nível de reações cruzadas possível, o que torna tais epítomos menos atrativos. Tais resultados podem ser encontrados na Tabela 2.

Figura 6. Localização dos epítomos preditos na glicoproteína *Spike*. **A)** Destaque para a estrutura primária, com as setas indicando a posição dos epítomos. Também é indicada o sítio de clivagem de furina (S1/S2), o sítio de clivagem do Peptídeo Sinal (S2') e os domínios NTD = Domínio N-Terminal (vermelho), RBD = Domínio de Ligação ao Receptor (verde musgo), FP = Peptídeo de Fusão (vinho, sigla não mostrada), HR1/2 = *Heptad Repeat* 1 (azul) e 2 (laranja), e TM = Domínio Transmembranar. **B)** Destaque para a estrutura quaternária em visão lateral e superior.



Fonte: o autor.

A Figura 6B, obtida no UCSF ChimeraX, mostra a localização exata de cada um dos sítios identificados no modelo tridimensional da proteína S com destaque para a superfície da mesma. Com base no campo de observação delimitado, fica claro que os epítomos 1, 3, 5, 10, 11, 12 e 13 aparentam possuir boa acessibilidade superficial a partir da perspectiva lateral da proteína, e os epítomos 3, 5, 7 e 9 aparentam possuir boa acessibilidade superficial a partir da visão superior da proteína. Ainda, da perspectiva superior, podemos visualizar nitidamente os

domínios NTD e RBD, com três unidades de cada epítopo, uma em cada protômero. A presença de múltiplas variações do mesmo epítopo próximas uma das outras é um dos fatos que pode tornar a região um excelente alvo para a ação de anticorpos, em especial anticorpos neutralizadores direcionados ao RBD. Os sítios 2, 4, 6 e 8 se apresentam mais inacessíveis em ambos planos de observação, estando estes enterrados em depressões e cavidades internas à estrutura do trímero.

Avaliar a acessibilidade superficial de potenciais epítopos de célula B é uma etapa importante quando desejamos compreender a resposta imune contra os mesmos, uma vez que, para estes serem capazes de interagir com BCRs e ativar linfócitos B, eles precisam, necessariamente, estarem expostos (SANCHEZ-TRINCADO; GOMES-PEROSANZ; RECHE, 2017). Estas primeiras observações nos oferecem uma boa ideia de quais sítios têm mais potencial de serem bons epítopos de célula B, entretanto, ainda é uma avaliação rasa e, ironicamente, superficial, pois este modelo em questão não está levando em consideração outra característica muito importante da proteína *Spike*: suas glicosilações.

Já não é novidade que proteínas superficiais de vírus tendem a possuir glicosilações, uma vez que estas podem desempenhar diversas funções, como mediar o reconhecimento de seus receptores celulares e a ancoragem do vírus na membrana da célula, direcionar o *folding* ou mesmo blindar a proteína contra a ação de anticorpos, como é o caso dos coronavírus (ZHANG et al., 2021). A proteína *Spike* de SARS-CoV-2 possui cerca de 22 glicosilações conhecidas, espalhadas por toda sua superfície (WATANABE et al., 2020) e dos nossos 13 epítopos, 5 deles (a saber E2, E3, E5, E12 e E13) possuem ao menos um sítio de N-glicosilação conhecido em sua sequência, o que os torna pouco atrativos para nossa aplicação devido ao provável impedimento estérico que os carboidratos ocasionam e pela dificuldade em obtê-los artificialmente.

Por mais que mimetizar tais glicosilações via sintética em nosso sistema síntese de peptídeos não seja possível, o impacto do escudo de carboidratos na acessibilidade superficial da proteína como um todo e de nossos epítopos é um dado interessante que pretendemos acessar com mais detalhes. Para tal, um modelo da proteína *Spike* completamente glicosilado foi gerado e submetido à um estudo de Dinâmica Molecular (MD) no *software* Amber, como especificado no item 4.2.

5.3 Dinâmica Molecular com a proteína S e seus epítomos

5.3.1 O trímero S é estável e seu comportamento dinâmico é “sincronizado”

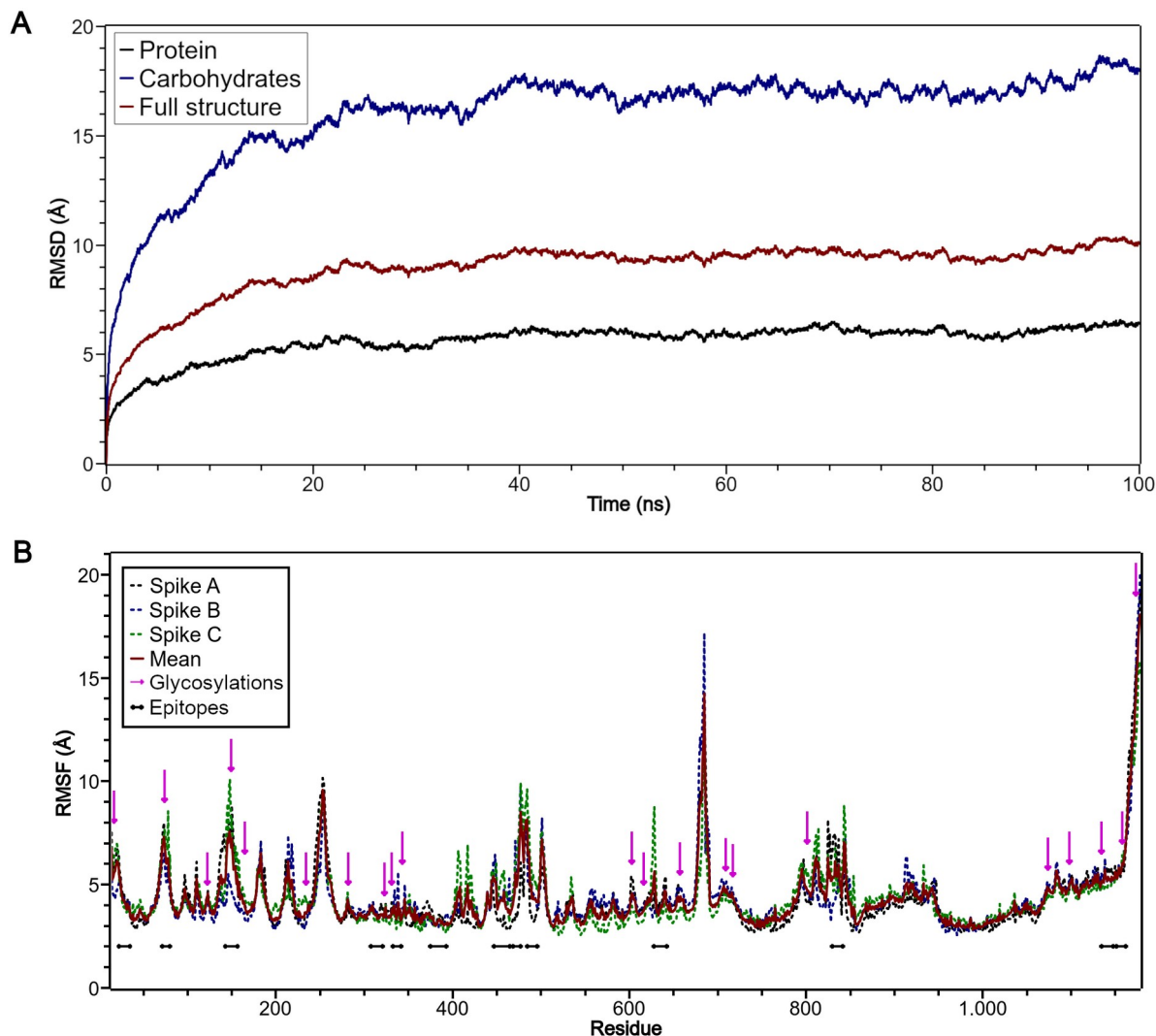
Uma das primeiras análises normalmente aplicadas após a obtenção da trajetória em experimentos de MD e que irá não apenas atestar a qualidade da simulação, mas também revelar importantes detalhes a respeito do comportamento dinâmico do sistema, é avaliar seu deslocamento espacial através do cálculo do RMSD. O RMSD (do inglês *Root Mean Square Deviation*) nos mostra o quanto os átomos de uma seleção específica do sistema se movimentaram ao longo da trajetória em relação a mesma estrutura no *frame* inicial (ou seja, em sua posição inicial), sendo calculado a partir da raiz quadrada dos desvios médios seguindo a equação:

$$\text{RMSD}(t_1, t_2) = \sqrt{\frac{1}{M} \cdot \sum_{i=1}^N m_i \|r_i(t_1) - r_i(t_2)\|^2} \quad (7)$$

onde $r_i(t_1)$ e $r_i(t_2)$ representam, respectivamente, as posições (compostas pelas coordenadas x , y e z) do átomo i nos tempos t_1 e t_2 ; e M representa a somatória das massas de todos os átomos da seleção (OLIVEIRA, 2013).

A Figura 7A mostra o RMSD para a proteínas *Spike* ao longo dos 100 ns de trajetória no sistema baseado em 6VXX. Os valores foram obtidos através do próprio Amber utilizando o módulo *cpptraj*, e selecionando dois diferentes componentes do sistema: resíduos de aminoácido e os carboidratos. Durante os primeiros 20 ns de simulação, todas estruturas tendem a se deslocar de sua posição inicial, com maior variação para os resíduos de açúcar (~17 angstroms) e a menor para os resíduos de aminoácido (~5 angstroms). Isso mostra que, mesmo com a energia potencial minimizada e com a temperatura equilibrada, as estruturas ainda carregavam tensões estruturais a serem liberadas. Ao longo dos 80 ns finais, o RMSD de todos os componentes convergiu e se estabilizou, com pequena variação próxima a média. Outro dado interessante que o gráfico mostra é a correta distribuição da energia potencial nos sistemas, uma vez que não há grandes picos ou saltos no RMSD. Levando tais resultados em consideração, todas as análises posteriores dependentes do tempo serão feitas nos 50 ns finais da trajetória, onde o sistema entra em equilíbrio dinâmico.

Figura 7. Aspectos dinâmicos da proteína S durante 100 ns de simulação. **A)** RMSD dos componentes principais. **B)** RMSF dos resíduos de aminoácido.



Fonte: o autor.

Ainda a respeito o gráfico de RMSD, podemos observar que as estruturas se movimentam de maneira muito similar no tempo observado, sendo as diferenças calculadas próximas a ~ 2 Å nos pontos onde essas atingiram sua maior disparidade, estando a maioria dentro de 1 Å. Como esperado, os carboidratos possuem uma flexibilidade muito superior à estrutura proteica, e essa flexibilidade interfere globalmente no deslocamento da proteína.

Para avaliarmos os efeitos locais da presença dos carboidratos no movimento da proteína S foi calculado o RMSF (do inglês *Root Mean Square Fluctuation*) dos 1163 resíduos de aminoácido que compõe cada um dos três protômeros. O significado do RMSF é muito próximo ao do RMSD, entretanto, ele nos fornece informações sobre partículas (átomos, resíduos, etc) individuais, e não da estrutura como um todo observado no RMSD. Ele pode ser calculado segundo a equação:

$$\text{RMSF}(t_1, t_2) = \sqrt{\sum_{i=1}^N \langle r_i(t_1) - r_i(t_2) \rangle^2} \quad (8)$$

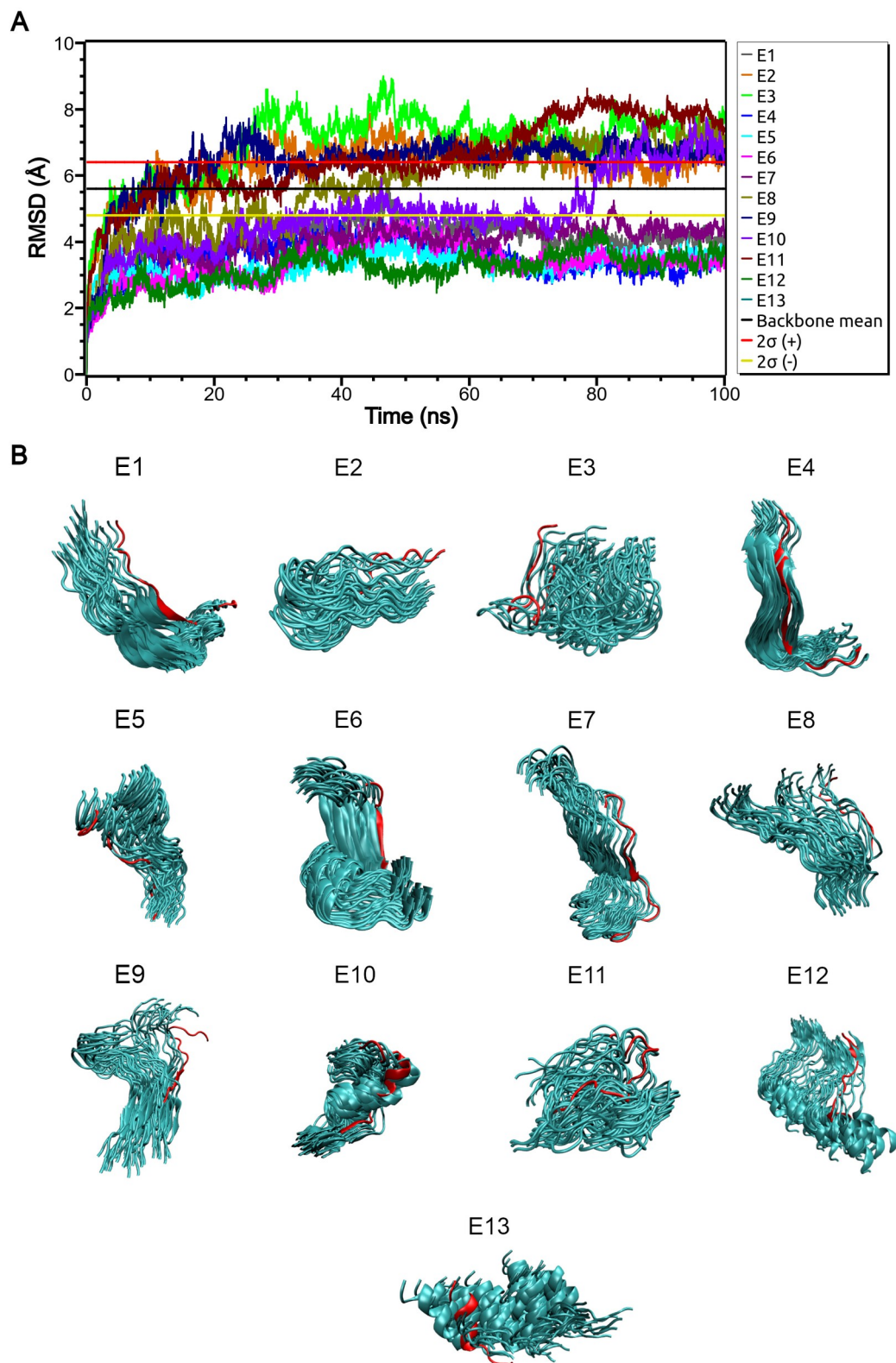
onde $r_i(t_1)$ e $r_i(t_2)$ representam, respectivamente, as posições (compostas pelas coordenadas x , y e z) do átomo i nos tempos t_1 e t_2 ; e os parênteses angulares $\langle r_i \rangle$ indicam que estamos trabalhando com as médias das posições para r_i , ou seja, com o *ensemble* das mesmas. O RMSF também foi obtido através do módulo *cpptraj*. A Figura 7B apresenta os resultados para este parâmetro.

O movimento individual dos protômeros que compõe o trímero *Spike* nos mostra o quão estável e sincronizada é a proteína, que apesar de seu tamanho e estrutura quaternária complexa, se comporta como uma estrutura única. Como era esperado, diferentes domínios possuem diferentes liberdades conformacionais, sendo o deslocamento mínimo observado por resíduo próximo a 3 Å. Se considerarmos a média entre os protômeros, o local de maior flexibilidade é justamente a alça onde ocorre a clivagem pela protease furina e a separação física entre as subunidade S1 e S2, entre os resíduos de número 680 e 690, com deslocamento de cerca de 15 Å. Dentre os nossos epítomos preditos, os que se encontram em regiões de maior flexibilidade são o E2, E3, E8 e E11, com um RMSF menor que 8 Å em média.

Por fim, é interessante ressaltar que o grande pico no RMSF da porção terminal da *Spike*, em ambos os sistemas, muito provavelmente é um artefato de nossa simulação, uma vez que, na proteína completa, existe continuidade na cadeia após o corte realizado na posição 1173, e a região adjacente não incluída em nossas simulações está ancorada no envelope viral, portanto, deve possuir uma mobilidade muito próxima à observada para a maior parte da subunidade S2, isto é, entre 3 e 4 Å.

Se selecionarmos apenas os resíduos que compõem nossos epítomos e calcularmos seu RMSD, podemos ter uma ideia da sua estabilidade conformacional ao longo da simulação. Na Figura 8A é possível encontrar tais valores. O comportamento observado nos 100 ns de trajetória é equiparável ao da proteína como um todo, visto que estamos apenas focalizando a análise em pontos específicos da estrutura, demonstrado pelo suave aumento nos valores dentro dos 20 ns iniciais, e uma estabilização nos 80 ns finais. Ao compará-los com o RMSD médio para a proteína S inteira, que é de ~5,2 Å (desconsiderando os carboidratos), são facilmente identificados dois grupos distintos. O primeiro é o dos epítomos cujo RMSD encontra-se próximo ou acima de dois desvios padrões da média da proteína ($2\sigma +$), que compreende o E2, E3, E9 e E11, e o segundo contém aqueles cujo RMSD encontra-se próximo ou abaixo de dois desvios padrões da média da proteína ($2\sigma -$), como o E1, E4, E5, E6, E7, E12 e E13.

Figura 8. Evolução dinâmica dos epítopos preditos quando internos a estrutura da proteína S no sistema 6VXX. **A)** Valores de RMSD dos epítopos durante 100 ns de simulação. **B)** Sobreposição das estruturas adotadas durante a dinâmica. A estrutura inicial está marcada em vermelho.



Fonte: o autor.

Na Figura 8B são expostos alguns estados visitados pelos epítomos na *Spike*. Foram sobrepostas 50 estruturas ao longo dos 100 ns de trajetória. Apesar de parecer um tanto quanto congestionada, se associarmos as estruturas aos valores de RMSD, é possível constatar o quão estáveis os epítomos são em relação a sua conformação, pois apesar de transladarem no espaço durante a produção da MD, isto é, se movimentarem no espaço, a estrutura global é mantida. Em termos de resposta imune, a estabilidade conformacional é muito importante para o processo de reconhecimento entre um anticorpo e seu alvo, mostrando a potencialidade de nossos epítomos.

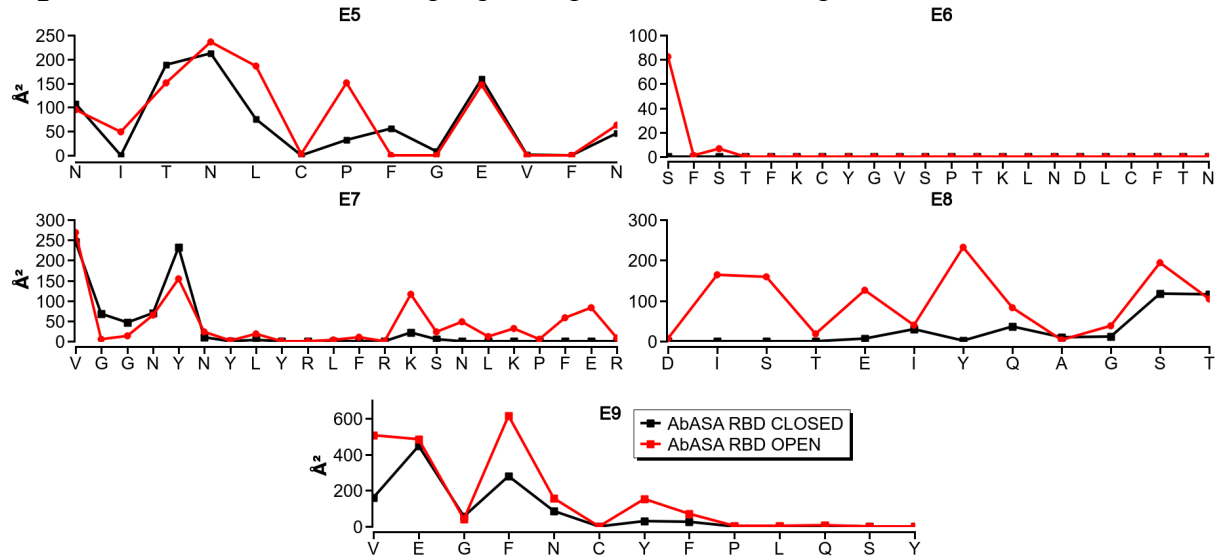
5.3.2 Acessibilidade superficial dos epítomos

Existem diversas abordagens para avaliarmos, em uma biomolécula, o grau de acessibilidade de determinado resíduo ao solvente ou à outra estrutura através de MD. Uma das mais populares é o cálculo da Área Superficial Acessível ao Solvente, ou simplesmente SASA (*Solvent Accessible Surface Area*). Em termos técnicos, o SASA mede a área acessível de uma biomolécula à um corpo de prova esférico com raio definido, considerando o raio de van der Waals da estrutura em alvo como ponto de partida para o cálculo (SHRAKE; RUPLEY, 1973). Uma esfera de raio igual a 1.4 Å é normalmente a escolhida por representar, aproximadamente, uma molécula de água, entretanto, pode-se variar seu tamanho de acordo com a análise pretendida. Ao calcular este parâmetro é possível determinar resíduos/átomos que encontram-se tanto acessíveis quanto inacessíveis ao corpo de prova, além da extensão dessa acessibilidade. Para direcionar nossas análises, estaremos apenas avaliando o SASA dos epítomos encontrados na região RBD uma vez que estes são os mais promissores, como mostra a Figura 9.

Nesta análise, utilizamos tanto a MD gerada para o sistema 6VXX (RBD DOWN) quando a gerada para o sistema 6VSB (RBD UP). O corpo de prova escolhido possui raio de 10 Å, que se aproxima ao tamanho médio de uma porção Fab de um anticorpo (WINTJENS; BIFANI; BIFANI, 2020). Isso nos permitiu avaliar a acessibilidade relativa dos resíduos aminoácidos que compõe estes sítios frente a uma estrutura de tamanho comparável a um anticorpo e inferir se estes epítomos podem despertar uma resposta de célula B, e como consequência, sua aplicabilidade no desenvolvimento de uma plataforma que visa detectar anticorpos no soro humano. Foram amostrados 50 *frames* igualmente espaçados de cada trajetória entre os

50 – 100 ns finais. Os valores plotados para o sistema 6VXX representam as médias entre os três protômeros da proteína S, enquanto o dados para o sistema 6VSB representam apenas os valores obtidos no protômero A, onde o RBD se encontra no estado aberto (UP).

Figura 9. Valores de SASA dos epítomos mapeados no RBD da proteína S.



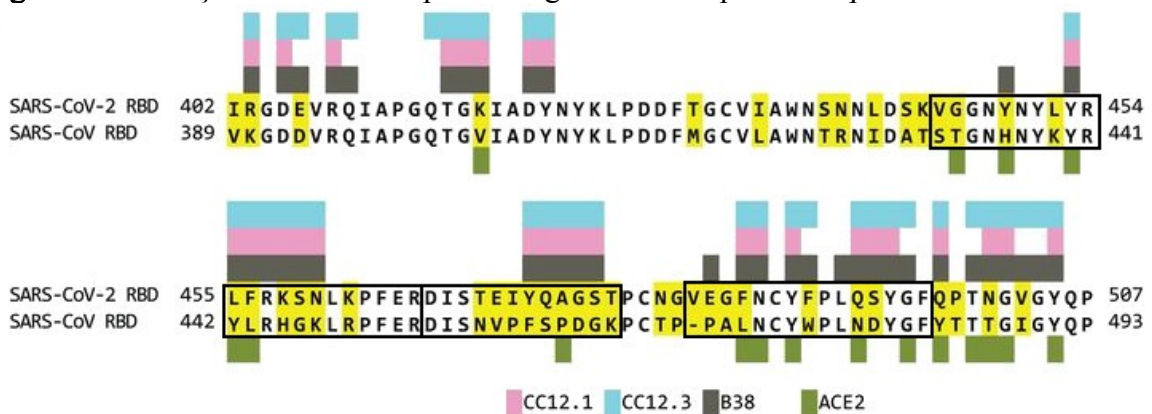
Fonte: o autor.

A primeira constatação obtida a partir dos dados indica que a acessibilidade não é homogênea entre os resíduos de aminoácido que compõe o epítipo e pode variar com grande amplitude mesmo entre resíduos vizinhos, como é o caso da leucina 335 (L335), encontrada em E5, que possui SASA $\approx 200 \text{ \AA}^2$ no sistema 6VSB, ou seja, se apresenta bem acessível, e sua vizinha cisteína 336 (C336), que possui SASA = 0 em ambos sistemas. Tal observação pode ser explicada pela orientação individual que cada resíduo de aminoácido assume na estrutura da proteína, de acordo com sua posição, vizinhança e estrutura secundária adotada (NELSON; COX, 2014).

Outro dado interessante observado é o fato de a acessibilidade dos resíduos variar de acordo com o estado conformacional adotado pelo RBD. De forma geral, quando o RBD se encontra no estado aberto (RBD UP), os resíduos tendem a apresentar um SASA maior e mais resíduos tendem a estar acessíveis e disponíveis para interagir com estruturas grandes, como é o caso da porção Fab dos anticorpos. Dentre os epítomos analisados, apenas o E6 se mostrou completamente inacessível em ambos sistemas, dada sua localização mais interna em relação à estrutura do RBD. Se considerarmos apenas o sistema 6VXX, onde o RBD encontra-se fechado (RBD DOWN) em todos os protômeros, os dados podem nos levar a crer que apenas o epítipo E5 seria um bom alvo na região RBD, entretanto, é importante lembrar que essa região deve possuir acessibilidade superficial pela sua função: se ligar ao receptor ACE2 (WRAPP et al., 2020).

Tanto o epítipo E7 quanto o E9 possuem 5 resíduos cada que fazem contato com resíduos de aminoácido da enzima ACE2 (Figura 5A). Segundo mostrado por YUAN e colaboradores (2020), ao menos três anticorpos monoclonais diferentes se ligam na região RBD e fazem contato com resíduos presentes nos epítipos E7, E8 e E9, como mostrado na Figura 10. O domínio proteico que compreende o RBD possui uma liberdade conformacional muito elevada e durante o processo de ligação do vírus na superfície da célula hospedeira sua conformação pode ser alterada da forma fechada para a forma aberta, o que eleva a área disponível para contato com o parceiro de interação (HENDERSON et al., 2020).

Figura 10. Interações entre anticorpos e a região RBD da proteína *Spike*.



Fonte: adaptado de YUAN et al., 2020.

A hipótese que explica tais valores de SASA é justamente a conformação adotada pelo RBD no modelo 6VXX. Ao repetirmos o cálculo com o modelo 6VSB, onde o RBD encontra-se na conformação aberta (RBD UP), os valores de SASA obtidos para os epítipos são consideravelmente mais elevados, confirmando que a acessibilidade da região depende não só do tamanho do corpo de prova parceiro de interação, mas também da conformação adotada pelo RBD. Tais dados também estão de acordo com outros estudos computacionais publicados na literatura (WINTJENS; BIFANI; BIFANI, 2020; CASALINO et al., 2020; GRANT et al., 2020; SIKORA et al., 2021).

5.3.3 Evolução dinâmica dos epítipos encontrados na região RBD

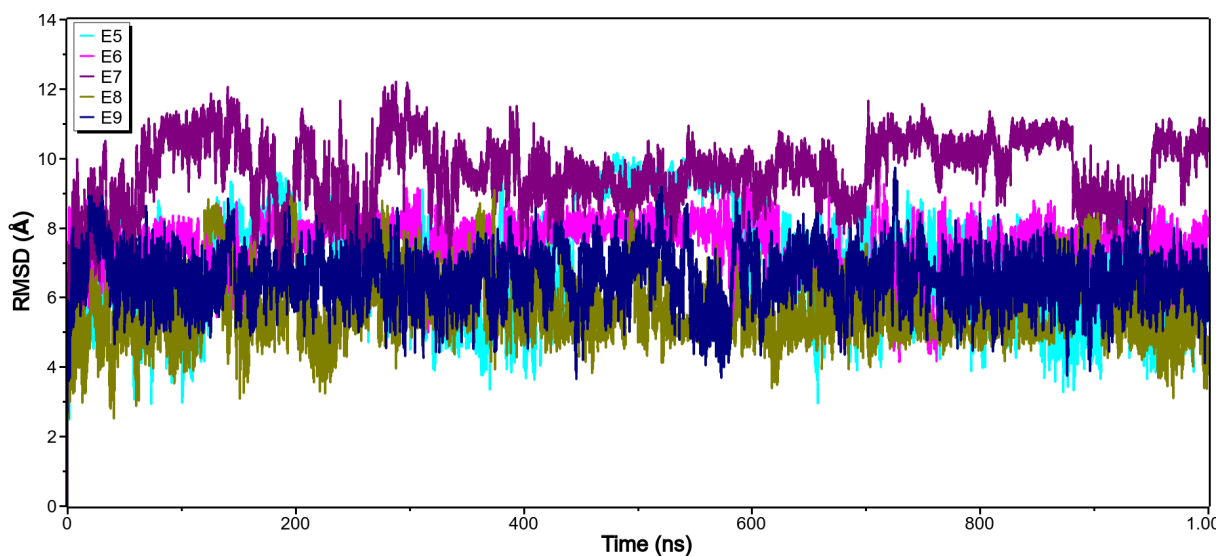
Por fim, uma última análise proposta foi avaliar o possível comportamento dos epítipos mapeados no RBD da proteína S em sua forma livre, como se os mesmos fossem obtidos por síntese química e solubilizados em água. Como já discutido, a estabilidade conformacio-

nais de todos os nossos epítomos preditos quando esses se encontram associados a um dos prótomeros é muito elevada, sem grandes mudanças observadas durante os 100 ns de estudo. Nos resta caracterizar as possíveis mudanças em sua conformação quando estes estão isolados.

Para tal, os epítomos E5 – E9 foram parametrizados individualmente em um sistema próprio e uma trajetória de 1 μ s foi obtida. A Figura 11 mostra os valores de RMSD obtidos durante as dinâmicas. O epítomo E7 foi o que alcançou o maior valor, de cerca de 10 Å em média. Os outros se mantiveram abaixo de 8 Å, valor não muito distante do observado para eles quando associados à proteína S, o que reforça a estabilidade conformacional dos mesmos.

Além do RMSD, também avaliamos as estruturas secundárias adotadas pelos epítomos ao longo da simulação. Essa análise é feita mapeando as ligações de hidrogênio intra-cadeia e calculando o potencial dos ângulos diedros através do algoritmo STRIDE (FRISHMAN; ARGOS, 1995). Na Figura 12 podemos encontrar os resultados. Juntamente com o DSSP também são apresentados três estados dos peptídeos ao longo da simulação: sua estrutura inicial, no *frame* 1, a sobreposição de 100 estruturas visitadas durante toda a simulação, e a estrutura final, no *frame* 50000.

Figura 11. RMSD dos epítomos E5 - E9 ao longo de 1 μ s de simulação.



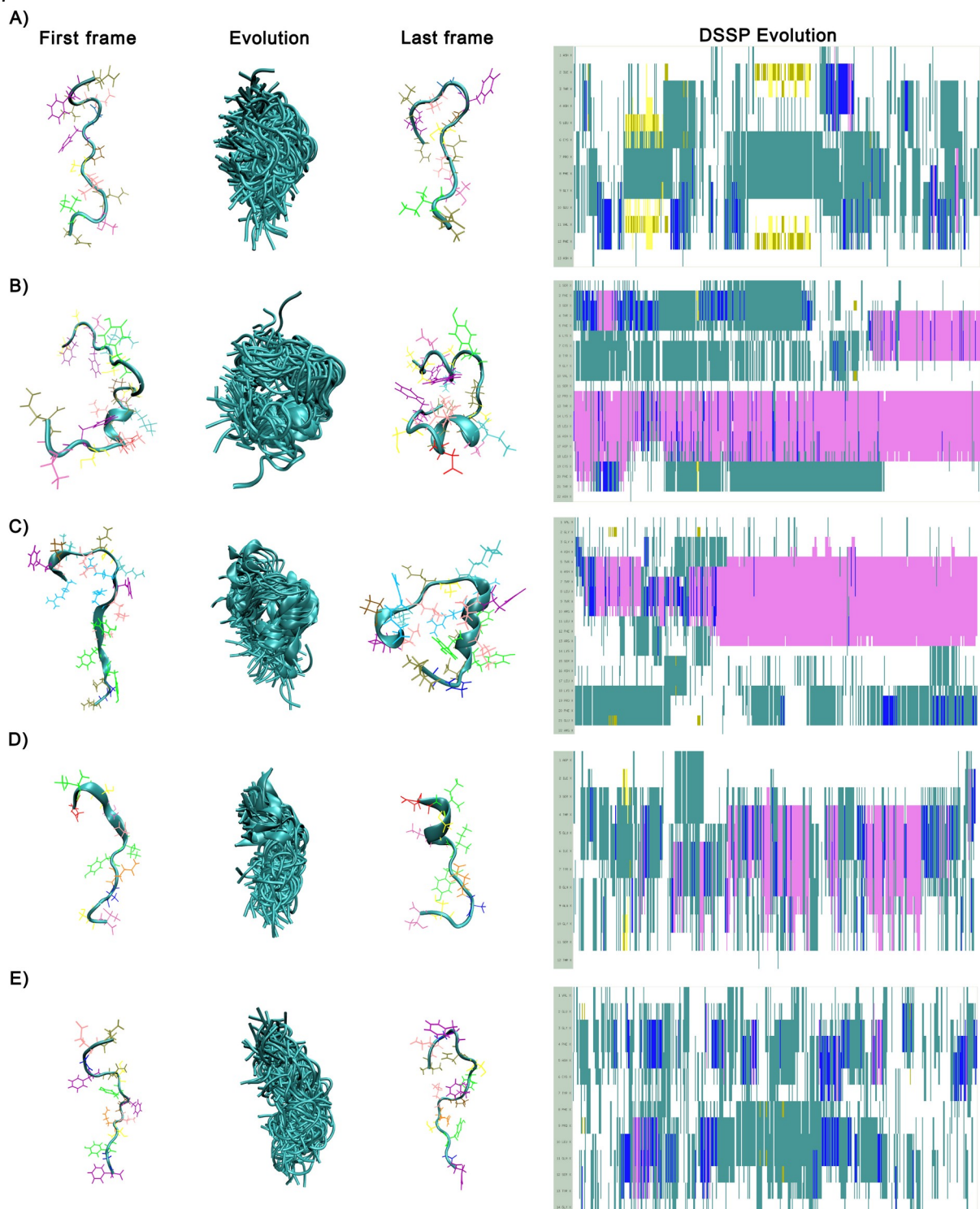
Fonte: o autor.

Dentre os epítomos avaliados, o E5, o E8 e o E9 não adotaram estruturas secundárias estáveis ao longo dos 1000 ns de trajetória e se mantiveram como *random coils*. Nos últimos nanossegundos do experimento, parte do epítomo E8 parece ter começado a se estruturar em uma hélice, o que é ilustrado tanto pela captura do *frame* final na Figura 12D, quanto pela sobreposição dos estados, entretanto, como a análise foi interrompida na marca de 1 microsegundo, não é possível afirmar se esta estrutura se manterá estável. Para os epítomos E6 e E7, houve a estruturação de um par de hélices para cada epítomo, partindo das extremidades. Essa

evolução fica mais evidente no caso do epítopo E6, que já traz em sua estrutura uma pequena hélice formada próxima a sua extremidade C-terminal, que passa a interagir com uma segunda hélice na extremidade N-terminal no quarto final da simulação. Ambas hélices do epítopo E7 se estruturam ao longo da simulação, sendo que a formação da estrutura mais próxima a extremidade C-terminal só ocorre nos momentos finais da trajetória, assim como para o E8.

Entender as possíveis conformações adotadas pelos epítopos que iremos aplicar no imunossensor é extremamente importante pela natureza dos anticorpos que desejamos detectar. Como já discutido, foram preditos epítopos lineares de célula B, ou seja, que devem interagir com anticorpos capazes de identificar a sequência de aminoácidos da cadeia, e não necessariamente sua conformação espacial. Neste caso, a problemática maior seria a adoção de uma estrutura muito estável pelos epítopos, o que impediria as pequenas mudanças conformacionais induzidas pela porção Fab do anticorpo e reduziria consideravelmente a capacidade de ligação entre as estruturas. De todo modo, apenas os epítopos E6 e E7 adotaram estruturas secundárias estáveis por um tempo considerável durante o experimento. No primeiro caso, uma das hélices já era parte de sua estrutura inicial, já no segundo caso, a hélice se formou ao longo da trajetória.

Figura 12. Evolução dinâmica e *folding* dos epítomos E5 - E9 quando isolados. **A)** E5. **B)** E6. **C)** E7. **D)** E8. **E)** E9. As cores no gráfico de DSSP representam as estruturas secundárias adotadas, sendo ciano = *turns*, azul = hélice 3-10, rosa = α -hélice, branco = *coils*, amarela = folha β .



Fonte: o autor.

5.4 Obtenção e caracterização química dos epítomos

Peptídeos são biomoléculas extremamente versáteis em termos de composição, estrutura e funções. Assim como as proteínas, os peptídeos são polímeros de L- α -aminoácidos com uma extensão que varia de 2 (dipeptídeos) a ~100 aminoácidos, sendo sua estrutura tridimensional e função determinadas diretamente pela sua estrutura primária (NELSON; COX, 2014). Nas células, peptídeos são produzidos pelos ribossomos, pela ação de proteases em proteínas maiores ou por uma família especial de enzimas chamadas Sintetases Peptídicas não Ribossomais (NRPSs) (YAZAWA; NUMATA, 2014; NELSON; COX, 2014; SUSSMUTH; MAINZ, 2017). Peptídeos possuem uma miríade de aplicações na ciência, na medicina e na indústria biotecnológica, podendo ser utilizados como agentes antimicrobianos (HUAN et al., 2020), agentes anti-hipertensivos (LEE; HUR, 2017), inibidores enzimáticos (VLIEGHE et al., 2010), adoçantes (RAINES; WENNEMERS, 2017) e epítomos para reações imunoquímicas e biossensores (LIU; WANG; BOYD, 2015).

Uma das formas de obtermos peptídeos específicos com alta pureza é a síntese química de peptídeos em fase sólida (SPPS, do inglês *Solid Phase Peptide Synthesis*), que só é possível pela aplicação de uma resina plástica insolúvel em água e em solventes orgânicos que está funcionalizada com grupos químicos especiais, grupos estes que darão suporte a construção do peptídeo através da adição sequencial de aminoácido por aminoácido, em um sistema solvente adequado e na presença de moléculas ativadoras que irão direcionar a formação das ligações peptídicas (BODANSZKY, 2012; BEHRENDT; WHITE; OFFER, 2016). Em nosso laboratório, recentemente implementamos a estratégia de SPPS para a síntese de possíveis novos alvos com aplicação diagnóstica e antiviral.

Uma vez que o protocolo de SPPS foi recém implementado em nosso laboratório, a partir de uma parceria com o Dr. Deivys Leandro Portuondo, faz-se necessário otimizar o procedimento a fim de estabelecermos um método viável, econômico e rápido para as sínteses de nosso grupo. Para validarmos a nossa estratégia de síntese, foi escolhido um peptídeo controle recentemente publicado como epítomo da proteína *Spike* capaz de elicitar uma resposta de anticorpos neutralizantes (POH et al., 2020), que também servirá de controle nas reações imunoquímicas futuras. Este epítomo está localizado na região de clivagem S2', que possui papel crucial na invasão das células hospedeiras pelo vírus. Os dados sobre este epítomo seguem:

Peptídeo: Controle

Sequência: H₂N-PSKPSKRSFIEDLLFNKV-CONH₂

n° de Aminoácidos: 18

Fórmula Química: $C_{97}H_{158}N_{26}O_{26}$

Massa Teórica: 2104.46

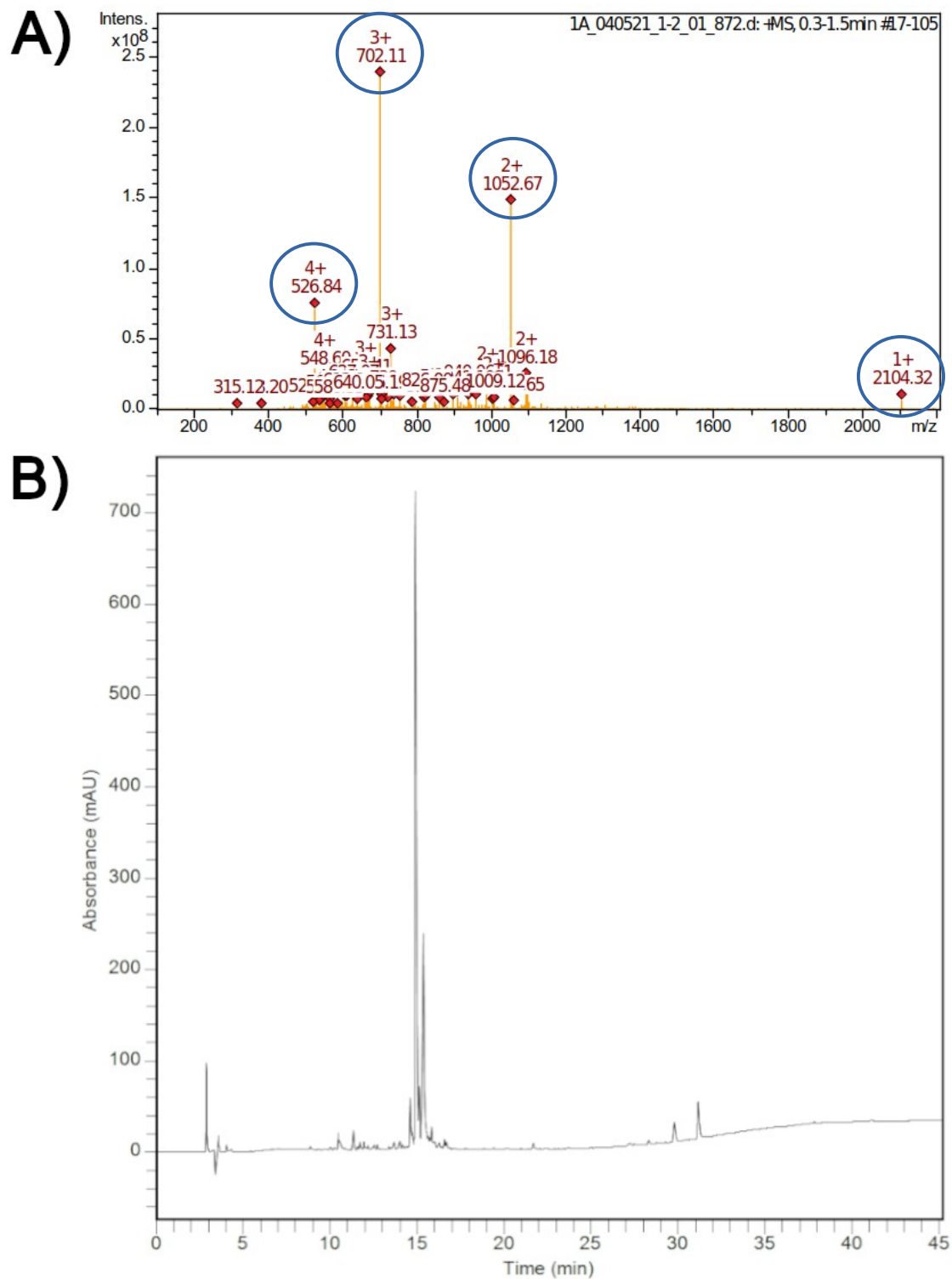
Fragmentação Teórica:

Fragmento	Massa	Detectado?
$[M + H]^+$	2104.19	Sim
$[M + 2H]^{2+}$	1052.60	Sim
$[M + 3H]^{3+}$	702.07	Sim
$[M + 4H]^{4+}$	526.80	Sim
$[M + 5H]^{5+}$	421.64	Não

A síntese durou ao todo 8 dias corridos, com uma média de 2 a 3 aminoácidos acoplados por dia, e contou com a participação da Mestranda Anna Carolina Toledo Borges e da estudante de Iniciação Científica Nicoly Simões de Melo. Para confirmar a identidade molecular do produto de síntese, é necessário analisar o espectro de massas e a fragmentação do mesmo, comparando as massas obtidas às massas teóricas calculadas. Para peptídeos lineares, a fragmentação pode ser obtida de forma manual ou através de alguma ferramenta de bioinformática (PAIZS; SUHAI, 2005), sendo, em nosso caso, utilizada a CEM Peptide Calculator. Na Figura 13A abaixo é apresentado o espectro de massas do produto obtido e com ele podemos concluir que o epítipo de interesse fora sintetizado com sucesso, uma vez que todos os principais fragmentos (íons filhos $M2^+$ a $M4^+$) e o íon integro (íon pai M^+) foram detectados.

Para caracterizarmos a pureza do peptídeo produzido, uma análise via Cromatografia Líquida de Alta Eficiência foi conduzida através de um gradiente exploratório de solvente B, partindo de 5% e atingindo 95% em 30 minutos, seguidos de 15 minutos à 100% de solvente B. O cromatograma é apresentado na Figura 13B. Nele, podemos observar um pico acentuado com tempo de retenção igual a 14,9 minutos, que deve corresponder ao produto de interesse, e vários outros picos menores, com intensidades e tempos de retenção variados, que devem representar as impurezas geradas durante a síntese e a clivagem do mesmo.

Figura 13. Identificação do peptídeo Controle obtido através de SPPS. **A)** Espectro de Massas. Os círculos azuis indicam os fragmentos teóricos calculados que foram observados na análise. **B)** Perfil cromatográfico da amostra em gradiente exploratório de 5-100% de solvente B.



Fonte: o autor.

Estando estabelecida e validada a estratégia, foram realizadas as sínteses de alguns dos sítios imunodominantes identificados e caracterizados através das análises de bioinformática. Os epítomos escolhidos foram o E7, E8 e E9. Tal preferência se dá pois estes encontram-se no domínio RBD, mais precisamente no domínio RBM, um dos mais importantes da proteína *Spike* por mediar diretamente o contato dessa com o receptor celular ACE2 e ser o alvo da grande maioria dos anticorpos neutralizadores (BARNES et al., 2020). Além disso, nossas análises computacionais de Dinâmica Molecular também demonstraram que esses epítomos possuem propriedades bioquímicas e biofísicas propícias para a eliciação de uma resposta de células B. Todos os epítomos foram satisfatoriamente sintetizados.

Peptídeo: E7

Sequência: H₂N-VGGNYNYLYRLFRKSNLKPFER-CONH₂

Fórmula Química: C₁₂₇H₁₉₃N₃₇O₃₁

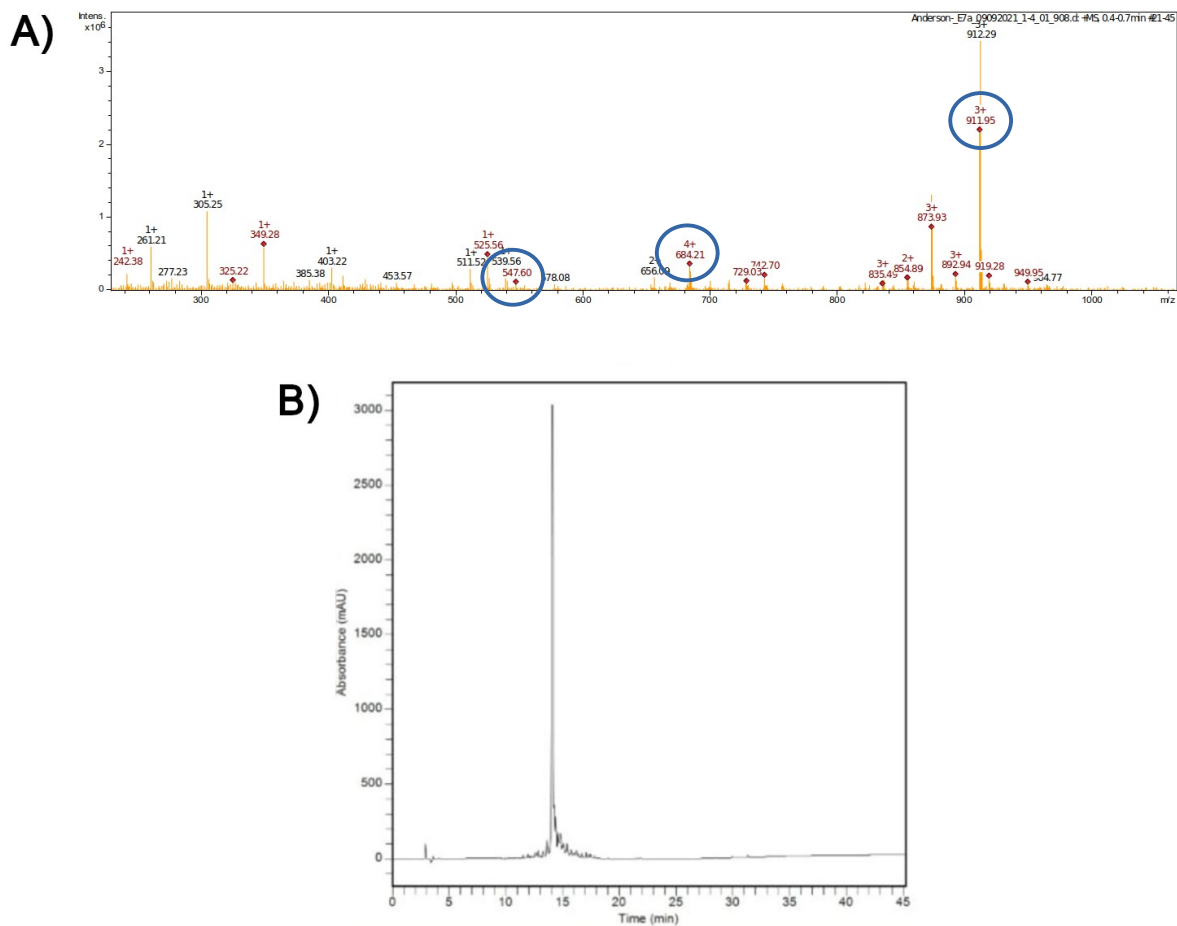
Massa Teórica: 2734.13

Fragmentação Teórica:

Fragmento	Massa	Detectado?
[M + H] ⁺	2733.47	Não
[M + 2H] ²⁺	1367.24	Não
[M + 3H] ³⁺	911.83	Sim
[M + 4H] ⁴⁺	684.12	Sim
[M + 5H] ⁵⁺	547.50	Sim
[M + 6H] ⁶⁺	456.42	Não

Fonte: o autor.

Figura 14. Identificação do peptídeo E7 obtido através de SPPS. **A)** Espectro de Massas. Os círculos azuis indicam os fragmentos teóricos calculados que foram observados na análise. **B)** Perfil cromatográfico da amostra em gradiente exploratório de 5-100% de solvente B.



Fonte: o autor.

Peptídeo: E8

Sequência: H₂N-DISTEIQAGST-CONH₂

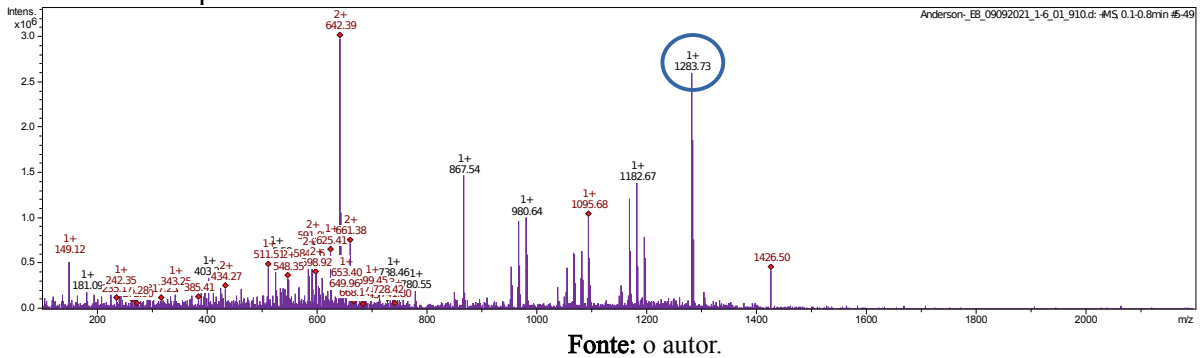
Fórmula Química: C₅₄H₈₆N₁₄O₂₂

Massa Teórica: 1283.35

Fragmentação Teórica:

Fragmento	Massa	Detectado?
[M + H] ⁺	1283.61	Sim

Figura 15. Espectro de massas do peptídeo E8. Os círculos azuis indicam os fragmentos teóricos calculados que foram observados na análise.



Peptídeo: E9

Sequência: H₂N-VEGFNCYFPLQSYGF-CONH₂

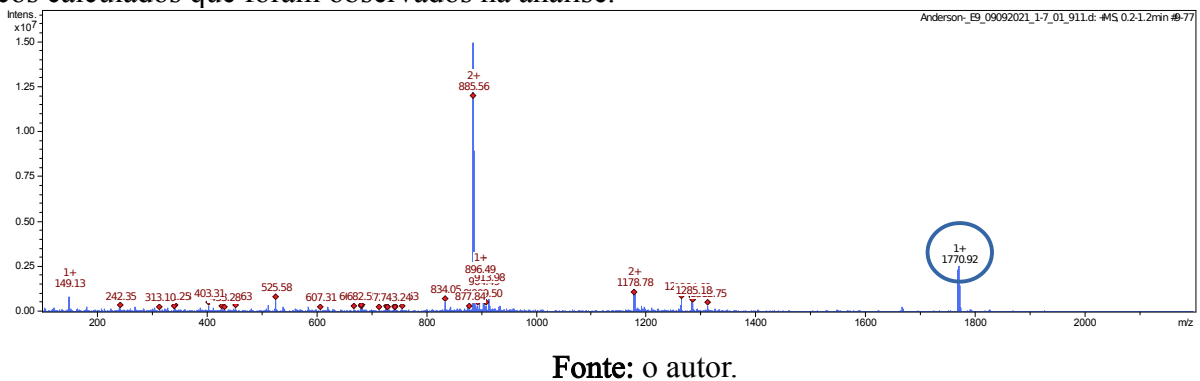
Fórmula Química: C₈₅H₁₁₂N₁₈O₂₂S

Massa Teórica: 1769.97

Fragmentação Teórica:

Fragmento	Massa	Detectado?
[M + H] ⁺	1769.80	Sim

Figura 16. Espectro de massas do peptídeo E9. Os círculos azuis indicam os fragmentos teóricos calculados que foram observados na análise.



Para o epítipo E7, o íon molecular com a massa total do peptídeo não pode ser observado pois seu valor se encontra fora do espectro de análise de 250 a 2200 m/z, entretanto, alguns de seus fragmentos filhos foram detectados, o que indica sua presença na amostra. A análise por HPLC em modo gradiente de solvente orgânico mostra um grande pico com tempo de retenção igual a 14,1 minutos, com a presença de algumas poucas impurezas. Para todos os peptídeos sintetizados, diversos outros picos com massas variadas podem ser observados no espectro, com acúmulo destes na faixa de 250 a 1200 m/z. Isso é um indicativo de que há

impurezas presentes na amostra e que essas impurezas possuem massa menor que os peptídeos de interesse.

Para qualquer síntese, seja de peptídeos ou de outras moléculas orgânicas, é esperado que haja a produção de impurezas na forma de peptídeos geralmente menores, originados em pontos onde a reação não foi completa ou onde ocorreu alguma reação paralela/degradação. Mesmo que atestemos a cada ciclo de acoplamento a ausência de grupos amina livres, o Teste de Kaiser é um teste visual e qualitativo, que nem sempre irá indicar com 100% de precisão a ausência total de grupos amino onde a reação de acoplamento não foi completa e que servem de suporte para a produção das impurezas. Há também a possibilidade de reações entre cadeias laterais dos aminoácido e destas com grupos carboxílicos ativados, em pontos onde, por alguma razão, o grupo protetor lateral fora removido.

5.5 Teste dos imunossensores funcionalizados com E7 para detecção de anticorpos anti-S

Para avaliar a ocorrência de interações antígeno-anticorpo entre os epítopos mapeados e sintetizados e os soros controles disponíveis no laboratório, foram preparados Sensores de Ouro Serigrafado visando experimentos de Voltametria Cíclica (CV). Por questões de disponibilidade dos sensores no laboratório, e dado que esta é apenas uma avaliação inicial do sistema, só foi possível realizar apenas uma duplicata de cada soro disponível. Os valores apresentados são a média aritmética dos experimentos.

A Voltametria Cíclica é uma técnica amplamente utilizada em biossensores, onde a informação eletroquímica sobre o analito/reação de interesse é obtida variando-se o potencial elétrico e medindo-se a corrente elétrica resultante. Para tal, são necessários três eletrodos: um eletrodo de trabalho, um eletrodo de referência e um contra-eletrodo. O potencial é medido entre o eletrodo de trabalho e o eletrodo de referência, enquanto a corrente é medida entre o eletrodo de trabalho e o contra-eletrodo (GRIESHABER et al., 2008). Como apenas o eletrodo de trabalho encontra-se funcionalizado com os antígenos de interesse, as diferenças observadas nas medidas são resultado dos eventos ocorridos na superfície do mesmo. A Figura 17 abaixo ilustra os eletrodos disponíveis para os experimentos.

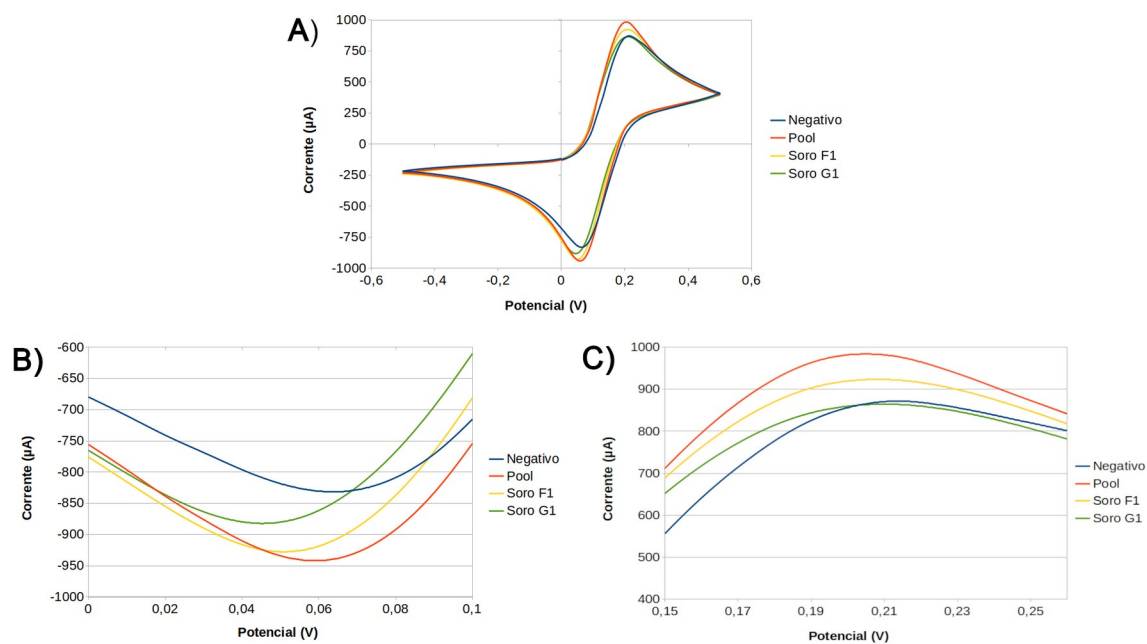
Figura 17. Modelo do Sensor de Ouro Serigrafado Metrohm DropSens 220AT utilizado nos experimentos. São indicadas as localizações dos eletrodos de trabalho (ouro), referência (prata) e contra-eletrodo (ouro), bem como suas respectivas conexões com o Potenciostato Palm-Sens Sensit BT.



Fonte: CASTILHO, 2022.

Foram utilizados três soros controles positivos para anticorpos anti-SARS-CoV-2 obtidos diretamente do Centro de Referência Diagnóstica do Núcleo de Atendimento à Comunidade (FCFAR -UNESP). Um Pool foi feito com todos os soros, e os aqui identificados como soros F1 e G1 foram também aplicados puros sobre o sensor. Os voltagramas cíclicos obtidos para cada soro podem ser consultados na Figura 18A. É interessante observar que ambas regiões onde os potenciais de redução (pico de corrente negativa) e oxidação (pico de corrente positiva) atingem seus valores máximos são muito bem definidas, entretanto, o potencial de redução (aproximado pela Figura 18B) parece oferecer as melhores medidas, uma vez que o controle negativo se mantém distante de todos os soros, diferente do que ocorre para o potencial de redução (aproximado pela Figura 18C). No potencial de redução, o controle negativo atinge o valor máximo de corrente igual a $-831 \mu\text{A}$, enquanto o Pool de amostras atinge $-941 \mu\text{A}$, o soro F1 atinge $-927 \mu\text{A}$ e o soro G1 – $880 \mu\text{A}$. A menor diferença entre o negativo e um dos soros positivos é de aproximadamente $50 \mu\text{A}$, e a maior de $110 \mu\text{A}$. Esses resultados, apesar de qualitativos e iniciais, nos indicam a ligação de anticorpos nos antígenos adsorvidos na superfície do sensor.

Figura 18. Teste dos sensores. A) Voltagramas Cíclicos completos. B) Potencial de redução. C) Potencial de oxidação.



6 CONCLUSÕES

- A partir da estratégia combinatória para predição de epítomos lineares de célula B proposta e testada no trabalho, 13 potenciais sítios imunodominantes foram encontrados na glicoproteína *Spike*;
- Com análises de bioinformática complementares, foi possível avaliar a identidade estrutural destes sítios frente a proteína S de SARS-CoV-1 e de MERS-CoV, além do acúmulo de mutações em decorrência do surgimento das variantes do vírus (Apêndice E);
- Com os experimentos atomísticos de Dinâmica Molecular, dentro das limitações do método e de nosso poder computacional, foi possível caracterizar propriedades bioquímicas e biofísicas importantes ao processo de interação antígeno-anticorpo que são inerentes aos epítomos preditos, e com isso avaliar quais seriam os melhores alvos para a validação experimental;
- Os epítomos E7, E8, E9 e Controle foram satisfatoriamente obtidos pela estratégia de SPPS implementada e validada no laboratório, sendo a identidade molecular dos produtos de síntese confirmada por Espectrometria de Massas e o perfil cromatográfico dos epítomos Controle e E7 obtido via HPLC;
- O epítomo E7 foi satisfatoriamente aplicado na validação inicial da plataforma de detecção de anticorpos anti-S em biossensor imunoeletroquímico com resultados preliminares promissores;

7 PERSPECTIVAS

- Obter os perfis cromatográficos dos epítopos E8 e E9;
- Purificar, quando necessário, os epítopos sintetizados através de Cromatografia Líquida de Alta Eficiência (CLAE/HPLC) semipreparativa;
- Avaliar a estrutura global em solução dos peptídeos sintetizados através de Dicroísmo Circular;
- Funcionalizar mais sensores e realizar as medições com mais replicadas e com significância estatística;

8 REFERÊNCIAS

- ABBAS, A. K.; LICHTMAN, A. H. H.; PILLAI, S. Cellular and Molecular Immunology. 10th ed. Amsterdam: Elsevier Health Sciences Division, 2022.
- AIZAWA, M. Principles and applications of electrochemical and optical biosensors. **Analytica Chimica Acta**, vol. 250, p. 249–256, Oct. 1991. [https://doi.org/10.1016/0003-2670\(91\)85073-2](https://doi.org/10.1016/0003-2670(91)85073-2).
- ANDERSEN, K. G.; RAMBAUT, A.; LIPKIN, W. I.; HOLMES, E. C.; GARRY, R. F. The Proximal Origin of SARS-CoV-2. **Nature Medicine**, vol. 26, p. 450–452, Mar. 2020. <https://doi.org/10.1038/s41591-020-0820-9>.
- ANSARI, H. R.; RAGHAVA, G. P. S. In Silico Models for B-Cell Epitope Recognition and Signaling. **Methods in Molecular Biology**, vol. 993, p. 129–138, 2013. https://doi.org/10.1007/978-1-62703-342-8_9.
- AYDIN, E. B.; AYDIN, M.; SEZGINTÜRK, M. K. Advances in electrochemical immunosensors. **Advances in Clinical Chemistry**, p. 1–57, 2019. <https://doi.org/10.1016/bs.acc.2019.04.006>.
- BAGDONAITE, I.; WANDALL, H. H. Global aspects of viral glycosylation. **Glycobiology**, vol. 28, no. 7, p. 443–467, Mar. 2018. <https://doi.org/10.1093/glycob/cwy021>.
- BARNES, C. O.; JETTE, C. A.; ABERNATHY, M. E.; DAM, K.-M. A.; ESSWEIN, S. R.; GRISTICK, H. B.; MALYUTIN, A. G.; SHARAF, N. G.; HUEY-TUBMAN, K. E.; LEE, Y. E.; ROBBIANI, D. F.; NUSSENZWEIG, M. C.; WEST, A. P.; BJORKMAN, P. J. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. **Nature**, vol. 588, p. 682–687, Oct. 2020. <https://doi.org/10.1038/s41586-020-2852-1>.
- BELOUZARD, S.; CHU, V. C.; WHITTAKER, G. R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. **Proceedings of the National Academy of Sciences**, vol. 106, no. 14, p. 5871–5876, Mar. 2009. <https://doi.org/10.1073/pnas.0809524106>.
- BEHRENDT, R.; WHITE, P.; OFFER, J. Advances in Fmoc solid-phase peptide synthesis. **Journal of Peptide Science**, vol. 22, no. 1, p. 4–27, Jan. 2016. <https://doi.org/10.1002/psc.2836>.
- BESTLE, D.; HEINDL, M. R.; LIMBURG, H.; VAN LAM VAN, T.; PILGRAM, O.; MOULTON, H.; STEIN, D. A.; HARDES, K.; EICKMANN, M.; DOLNIK, O.; ROHDE, C.;

KLENK, H.-D.; GARTEN, W.; STEINMETZER, T.; BÖTTCHER-FRIEBERTSHÄUSER, E. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. **Life Science Alliance**, vol. 3, no. 9, p. e202000786, Jul. 2020. <https://doi.org/10.26508/lsa.202000786>.

BIASOTTO, G.; COSTA, J. P. C.; COSTA, P. I.; ZAGHETE, M. A. ZnO nanorods-gold nanoparticle-based biosensor for detecting hepatitis C. **Applied Physics A**, vol. 125, no. 12, p. 1–7, Nov. 2019. <https://doi.org/10.1007/s00339-019-3128-1>.

BLYTHE, M. J.; FLOWER, D. R. Benchmarking B cell epitope prediction: Underperformance of existing methods. **Protein Science**, vol. 14, no. 1, p. 246–248, Jan. 2009. <https://doi.org/10.1110/ps.041059505>.

BODANSZKY, M. Principles of Peptide Synthesis. Berlin: Heidelberg Springer, 1984.

BRASIL. Painel Coronavírus. Mar. 2022. Available at: <https://covid.saude.gov.br/>. Accessed on: 15 Mar. 2022.

BROWN, B. L.; MCCULLOUGH, J. Treatment for emerging viruses: Convalescent plasma and COVID-19. **Transfusion and Apheresis Science**, vol. 59, no. 3, Apr. 2020. <https://doi.org/10.1016/j.transci.2020.102790>.

BUI, H.-H.; SIDNEY, J.; LI, W.; FUSSEDER, N.; SETTE, A. Development of an epitope conservancy analysis tool to facilitate the design of epitope-based diagnostics and vaccines. **BMC Bioinformatics**, vol. 8, no. 1, Sep. 2007. <https://doi.org/10.1186/1471-2105-8-361>.

CAI, Y.; ZHANG, J.; XIAO, T.; PENG, H.; STERLING, S. M.; WALSH, R. M.; RAWSON, S.; RITS-VOLLOCH, S.; CHEN, B. Distinct conformational states of SARS-CoV-2 spike protein. **Science**, vol. 369, no. 6511, p. 1586–1592, Jul. 2020. <https://doi.org/10.1126/science.abd4251>.

CHAN, J. F.-W.; KOK, K.-H.; ZHU, Z.; CHU, H.; TO, K. K.-W.; YUAN, S.; YUEN, K.-Y. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. **Emerging Microbes & Infections**, vol. 9, no. 1, p. 221–236, Jan. 2020. <https://doi.org/10.1080/22221751.2020.1719902>.

CHENG, Y.-W.; CHAO, T.-L.; LI, C.-L.; CHIU, M.-F.; KAO, H.-C.; WANG, S.-H.; PANG, Y.-H.; LIN, C.-H.; TSAI, Y.-M.; LEE, W.-H.; TAO, M.-H.; HO, T.-C.; WU, P.-Y.; JANG, L.-T.; CHEN, P.-J.; CHANG, S.-Y.; YEH, S.-H. Furin Inhibitors Block SARS-CoV-2 Spike Pro-

tein Cleavage to Suppress Virus Production and Cytopathic Effects. **Cell Reports**, vol. 33, no. 2, p. 108254, Oct. 2020. <https://doi.org/10.1016/j.celrep.2020.108254>.

CHI, X.; YAN, R.; ZHANG, J.; ZHANG, G.; ZHANG, Y.; HAO, M.; ZHANG, Z.; FAN, P.; DONG, Y.; YANG, Y.; CHEN, Z.; GUO, Y.; ZHANG, J.; LI, Y.; SONG, X.; CHEN, Y.; XIA, L.; FU, L.; HOU, L.; XU, J. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. **Science**, vol. 369, no. 6504, p. 650–655, Jun. 2020. <https://doi.org/10.1126/science.abc6952>.

BIASOTTO, G.; COSTA, J. P. C.; COSTA, P. I.; ZAGHETE, M. A. ZnO nanorods-gold nanoparticle-based biosensor for detecting hepatitis C. **Applied Physics A**, vol. 125, no. 12, Nov. 2019. <https://doi.org/10.1007/s00339-019-3128-1>.

CASTILHO, B. M. Identificação, caracterização e síntese peptídica associada à enolase de *Sporothrix schenckii*: aplicação em plataforma imunossensora eletroquímica portátil para o diagnóstico da Esporotricose Felina. 2022. Thesis – Faculdade de Ciências Farmacêuticas, Universidade Estadual Paulista, Campus Araraquara, 2022.

CHO, B. G.; GAUTAM, S.; PENG, W.; HUANG, Y.; GOLI, M.; MECHREF, Y. Direct Comparison of N-Glycans and Their Isomers Derived from Spike Glycoprotein 1 of MERS-CoV, SARS-CoV-1, and SARS-CoV-2. **Journal of Proteome Research**, vol. 20, no. 9, p. 4357–4365, Aug. 2021. <https://doi.org/10.1021/acs.jproteome.1c00323>.

CHRISTENSEN, P. A.; OLSEN, R. J.; LONG, S. W.; SNEHAL, R.; DAVIS, J. J.; SAAVEDRA, M. O.; REPPOND, K.; SHYER, M. N.; CAMBRIC, J.; GADD, R.; THAKUR, R. M.; BATAJOO, A.; MANGHAM, R.; PENA, S.; TRINH, T.; KINSKEY, J. C.; WILLIAMS, G.; OLSON, R.; GOLLIHAR, J.; MUSSER, J. M. Signals of Significantly Increased Vaccine Breakthrough, Decreased Hospitalization Rates, and Less Severe Disease in Patients with Coronavirus Disease 2019 Caused by the Omicron Variant of Severe Acute Respiratory Syndrome Coronavirus 2 in Houston, Texas. **The American Journal of Pathology (Preprint)**, Feb. 2022. <https://doi.org/10.1016/j.ajpath.2022.01.007>.

COLE, J. B.; FLOREZ, J. C. Genetics of diabetes mellitus and diabetes complications. **Nature Reviews. Nephrology**, vol. 16, no. 7, p. 377–390, 12 May 2020. <https://doi.org/10.1038/s41581-020-0278-5>.

CORDEIRO, T. A. R.; MARTINS, H. R.; FRANCO, D. L.; SANTOS, F. L. N.; CELEDON, P. A. F.; CANTUÁRIA, V. L.; DE LANA, M.; REIS, A. B.; FERREIRA, L. F. Impedimetric

immunosensor for rapid and simultaneous detection of chagas and visceral leishmaniasis for point of care diagnosis. **Biosensors and Bioelectronics**, vol. 169, p. 112573, Dec. 2020. <https://doi.org/10.1016/j.bios.2020.112573>.

CORMAN, V. M.; LANDT, O.; KAISER, M.; MOLENKAMP, R.; MEIJER, A.; CHU, D. K.; BLEICKER, T.; BRÜNINK, S.; SCHNEIDER, J.; SCHMIDT, M. L.; MULDER, D. G.; HAAGMANS, B. L.; VAN DER VEER, B.; VAN DEN BRINK, S.; WIJSMAN, L.; GORDERSKI, G.; ROMETTE, J.-L.; ELLIS, J.; ZAMBON, M.; PEIRIS, M. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. **Eurosurveillance**, vol. 25, no. 3, Jan. 2020. <https://doi.org/10.2807/1560-7917.es.2020.25.3.2000045>.

CUI, J.; LI, F.; SHI, Z.-L. Origin and evolution of pathogenic coronaviruses. **Nature Reviews Microbiology**, vol. 17, no. 3, p. 181–192, Dec. 2018. <https://doi.org/10.1038/s41579-018-0118-9>.

COSTA, J. P.; BASTOS, W. B.; DA COSTA, P. I.; ZAGHETE, M. A.; LONGO, E.; CARMO, J. P. Portable Laboratory Platform With Electrochemical Biosensors for Immunodiagnostic of Hepatitis C Virus. **IEEE Sensors Journal**, vol. 19, no. 22, p. 10701–10709, Nov. 2019. <https://doi.org/10.1109/jsen.2019.2930957>.

DONOGHUE, M.; HSIEH, F.; BARONAS, E.; GODBOUT, K.; GOSSELIN, M.; STAGLIANO, N.; DONOVAN, M.; WOOLF, B.; ROBISON, K.; JEYASEELAN, R.; BREITBART, R. E.; ACTON, S. A Novel Angiotensin-Converting Enzyme–Related Carboxypeptidase (ACE2) Converts Angiotensin I to Angiotensin 1-9. **Circulation Research**, vol. 87, no. 5, p. e1–e9, Sep. 2000. <https://doi.org/10.1161/01.res.87.5.e1>.

FDA. FACT SHEET FOR HEALTH CARE PROVIDERS EMERGENCY USE AUTHORIZATION (EUA) OF REGEN-COV™ (casirivimab with imdevimab). [S. l.: s. n.], Jul. 2021. Available at: <https://www.fda.gov/media/145611/download>. Accessed on: 16 Mar. 2022.

FORNI, D.; CAGLIANI, R.; CLERICI, M.; SIRONI, M. Molecular Evolution of Human Coronavirus Genomes. **Trends in Microbiology**, vol. 25, no. 1, p. 35–48, Jan. 2017. <https://doi.org/10.1016/j.tim.2016.09.001>.

FRISHMAN, D.; ARGOS, P. Knowledge-based protein secondary structure assignment. **Proteins: Structure, Function, and Genetics**, vol. 23, no. 4, p. 566–579, Dec. 1995. <https://doi.org/10.1002/prot.340230412>.

GASPAROTTO, G.; COSTA, J. P. C.; COSTA, P. I.; ZAGHETE, M. A.; MAZON, T. Electrochemical immunosensor based on ZnO nanorods-Au nanoparticles nanohybrids for ovarian cancer antigen CA-125 detection. **Materials Science and Engineering: C**, vol. 76, p. 1240–1247, Jul. 2017. <https://doi.org/10.1016/j.msec.2017.02.031>.

GELPI, J.; HOSPITAL, A.; GOÑI, R.; OROZCO, M. Molecular dynamics simulations: advances and applications. **Advances and Applications in Bioinformatics and Chemistry**, vol. 8, p. 37–47, Nov. 2015. <https://doi.org/10.2147/aabc.s70333>.

GHORBANI, M.; BROOKS, B. R.; KLAUDA, J. B. Exploring dynamics and network analysis of spike glycoprotein of SARS-COV-2. **Biophysical Journal**, vol. 120, no. 14, p. 2902–2913, Mar. 2021. <https://doi.org/10.1016/j.bpj.2021.02.047>.

GONZÁLEZ, M. A. Force fields and molecular dynamics simulations. **École thématique de la Société Française de la Neutronique**, vol. 12, p. 169–200, Jun. 2011. <https://doi.org/10.1051/sfn/201112009>.

GUAN, W.; NI, Z.; HU, Y.; LIANG, W.; OU, C.; HE, J.; LIU, L.; SHAN, H.; LEI, C.; HUI, D. S. C.; DU, B.; LI, L.; ZENG, G.; YUEN, K.-Y.; CHEN, R.; TANG, C.; WANG, T.; CHEN, P.; XIANG, J.; LI, S. Clinical Characteristics of Coronavirus Disease 2019 in China. **New England Journal of Medicine**, vol. 382, no. 18, Feb. 2020. <https://doi.org/10.1056/nejmoa2002032>.

GRIESHABER, D.; MACKENZIE, R.; VÖRÖS, J.; REIMHULT, E. Electrochemical Biosensors - Sensor Principles and Architectures. **Sensors**, vol. 8, no. 3, p. 1400–1458, Mar. 2008. <https://doi.org/10.3390/s80314000>.

HENDERSON, R.; EDWARDS, R. J.; MANSOURI, K.; JANOWSKA, K.; STALLS, V.; GOBEIL, S. M. C.; KOPP, M.; LI, D.; PARKS, R.; HSU, A. L.; BORGNIA, M. J.; HAYNES, B. F.; ACHARYA, P. Controlling the SARS-CoV-2 spike glycoprotein conformation. **Nature Structural & Molecular Biology**, vol. 27, no. 10, p. 925–933, Oct. 2020. <https://doi.org/10.1038/s41594-020-0479-4>.

HOFFMANN, M.; KLEINE-WEBER, H.; SCHROEDER, S.; KRÜGER, N.; HERRLER, T.; ERICHSEN, S.; SCHIERGENS, T. S.; HERRLER, G.; WU, N.-H.; NITSCHKE, A.; MÜLLER, M. A.; DROSTEN, C.; PÖHLMANN, S. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. **Cell**, vol. 181, no. 2, p. 271–280, Mar. 2020. <https://doi.org/10.1016/j.cell.2020.02.052>.

HUAN, Y.; KONG, Q.; MOU, H.; YI, H. Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. **Frontiers in Microbiology**, vol. 11, Oct. 2020. <https://doi.org/10.3389/fmicb.2020.582779>.

HUANG, C.; HUANG, L.; WANG, Y.; LI, X.; REN, L.; GU, X.; KANG, L.; GUO, L.; LIU, M.; ZHOU, X.; LUO, J.; HUANG, Z.; TU, S.; ZHAO, Y.; CHEN, L.; XU, D.; LI, Y.; LI, C.; PENG, L.; LI, Y. 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. **The Lancet**, vol. 397, no. 10270, p. 220–232, Jan. 2021. [https://doi.org/10.1016/S0140-6736\(20\)32656-8](https://doi.org/10.1016/S0140-6736(20)32656-8).

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. **Journal of Molecular Graphics**, vol. 14, no. 1, p. 33–38, Feb. 1996. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).

HWANG, D.-W.; LEE, S.; SEO, M.; CHUNG, T. D. Recent advances in electrochemical non-enzymatic glucose sensors – A review. **Analytica Chimica Acta**, vol. 1033, p. 1–34, Nov. 2018. <https://doi.org/10.1016/j.aca.2018.05.051>.

IONESCU, R. E.; GONDRAN, C.; BOUFFIER, L.; JAFFREZIC-RENAULT, N.; MARTELLET, C.; COSNIER, S. Label-free impedimetric immunosensor for sensitive detection of atrazine. **Electrochimica Acta**, vol. 55, no. 21, p. 6228–6232, Aug. 2010. <https://doi.org/10.1016/j.electacta.2009.11.029>.

JABEEN, A.; MOHAMEDALI, A.; RANGANATHAN, S. Protocol for protein structure modelling. *In*: RANGANATHAN, S.; NAKAI, K.; SCHÖNBACH, C.; GRIBSKOV, M. (eds.). **Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics**. [S. l.]: Elsevier, 2019. p. 252–272.

JASSAT, W.; ABDOL KARIM, S. S.; MUDARA, C.; WELCH, R.; OZOUGWU, L.; GROOME, M. J.; GOVENDER, N.; VON GOTTFBERG, A.; WOLTER, N.; WOLMARANS, M.; ROUSSEAU, P.; BLUMBERG, L.; COHEN, C. Clinical severity of COVID-19 patients admitted to hospitals during the Omicron wave in South Africa. **medRxiv (Preprint)**, Feb. 2022. <https://doi.org/10.1101/2022.02.22.21268475>.

JESPERSEN, M. C.; PETERS, B.; NIELSEN, M.; MARCATILI, P. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. **Nucleic Acids Research**, vol. 45, no. W1, p. W24–W29, May 2017. <https://doi.org/10.1093/nar/gkx346>.

JI, W.; WANG, W.; ZHAO, X.; ZAI, J.; LI, X. Homologous recombination within the spike glycoprotein of the newly identified coronavirus may boost cross-species transmission from snake to human. **Journal of Medical Virology**, vol. 92, no. 4, p. 433–440, Jan. 2020. <https://doi.org/10.1002/jmv.25682>.

JIANG, F.; DENG, L.; ZHANG, L.; CAI, Y.; CHEUNG, C. W.; XIA, Z. Review of the Clinical Characteristics of Coronavirus Disease 2019 (COVID-19). **Journal of General Internal Medicine**, vol. 35, no. 5, p. 1545–1549, Mar. 2020. <https://doi.org/10.1007/s11606-020-05762-w>.

JO, S.; KIM, T.; IYER, V. G.; IM, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. **Journal of Computational Chemistry**, vol. 29, no. 11, p. 1859–1865, Mar. 2008. <https://doi.org/10.1002/jcc.20945>.

JORGENSEN, W. L.; CHANDRASEKHAR, J.; MADURA, J. D.; IMPEY, R. W.; KLEIN, M. L. Comparison of simple potential functions for simulating liquid water. **The Journal of Chemical Physics**, vol. 79, no. 2, p. 926–935, Jul. 1983. <https://doi.org/10.1063/1.445869>.

KARUNAKARAN, C.; PANDIARAJ, M.; SANTHARAMAN, P. Immunosensors. **Biosensors and Bioelectronics**, p. 205–245, 2015. <https://doi.org/10.1016/b978-0-12-803100-1.00004-9>.

KENFE, F. R.; URBACZEK, A. C.; SILVA, J. C.; NÉO, T. A.; DA SILVA, F. H.; DA COSTA, P. I. Development of diagnostic methods and study of the immunoreactivity of a mixture of recombinant core and E2 proteins fused to GST with control serum positive for hepatitis C. **Talanta**, vol. 110, p. 32–38, Jun. 2013. <https://doi.org/10.1016/j.talanta.2013.02.017>.

JU, B.; ZHANG, Q.; GE, J.; WANG, R.; SUN, J.; GE, X.; YU, J.; SHAN, S.; ZHOU, B.; SONG, S.; TANG, X.; YU, J.; LAN, J.; YUAN, J.; WANG, H.; ZHAO, J.; ZHANG, S.; WANG, Y.; SHI, X.; LIU, L. Human neutralizing antibodies elicited by SARS-CoV-2 infection. **Nature**, vol. 584, p. 115–119, May 2020. <https://doi.org/10.1038/s41586-020-2380-z>.

KIRSCHNER, K. N.; YONGYE, A. B.; TSCHAMPEL, S. M.; GONZÁLEZ-OUTEIRIÑO, J.; DANIELS, C. R.; FOLEY, B. L.; WOODS, R. J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. **Journal of Computational Chemistry**, vol. 29, no. 4, p. 622–655, Sep. 2007. <https://doi.org/10.1002/jcc.20820>.

KLEPEIS, J. L.; LINDORFF-LARSEN, K.; DROR, R. O.; SHAW, D. E. Long-timescale molecular dynamics simulations of protein structure and function. **Current Opinion in Structural Biology**, vol. 19, no. 2, p. 120–127, Apr. 2009. <https://doi.org/10.1016/j.sbi.2009.03.004>.

KOLASKAR, A. S.; TONGAONKAR, P. C. A semi-empirical method for prediction of antigenic determinants on protein antigens. **FEBS Letters**, vol. 276, no. 1, p. 172–174, Dec. 1990. [https://doi.org/10.1016/0014-5793\(90\)80535-q](https://doi.org/10.1016/0014-5793(90)80535-q).

KRAMMER, F. SARS-CoV-2 vaccines in development. **Nature**, vol. 586, p. 516–527, Sep. 2020. <https://doi.org/10.1038/s41586-020-2798-3>.

LAM, T. T.-Y.; SHUM, M. H.-H.; ZHU, H.-C.; TONG, Y.-G.; NI, X.-B.; LIAO, Y.-S.; WEI, W.; CHEUNG, W. Y.-M.; LI, W.-J.; LI, L.-F.; LEUNG, G. M.; HOLMES, E. C.; HU, Y.-L.; GUAN, Y. Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. **Nature**, vol. 583, p. 282–285, Mar. 2020. <https://doi.org/10.1038/s41586-020-2169-0>.

LAN, J.; GE, J.; YU, J.; SHAN, S.; ZHOU, H.; FAN, S.; ZHANG, Q.; SHI, X.; WANG, Q.; ZHANG, L.; WANG, X. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. **Nature**, vol. 581, p. 215–220, Mar. 2020. <https://doi.org/10.1038/s41586-020-2180-5>.

LARSEN, J.; LUND, O.; NIELSEN, M. Improved method for predicting linear B-cell epitopes. **Immunome Research**, vol. 2, no. 1, 2006. <https://doi.org/10.1186/1745-7580-2-2>.

LEE, S. Y.; HUR, S. J. Antihypertensive peptides from animal products, marine organisms, and plants. **Food Chemistry**, vol. 228, p. 506–517, Aug. 2017. <https://doi.org/10.1016/j.foodchem.2017.02.039>.

LI, D.; SEMPOWSKI, G. D.; SAUNDERS, K. O.; ACHARYA, P.; HAYNES, B. F. SARS-CoV-2 Neutralizing Antibodies for COVID-19 Prevention and Treatment. **Annual Review of Medicine**, vol. 73, no. 1, p. 1–16, Aug. 2021. <https://doi.org/10.1146/annurev-med-042420-113838>.

LI, F.; LI, W.; FARZAN, M.; HARRISON, S. C. Structure of SARS Coronavirus Spike Receptor-Binding Domain Complexed with Receptor. **Science**, vol. 309, no. 5742, p. 1864–1868, Sep. 2005. <https://doi.org/10.1126/science.1116480>.

LI, W. Bats Are Natural Reservoirs of SARS-Like Coronaviruses. **Science**, vol. 310, no. 5748, p. 676–679, Oct. 2005. <https://doi.org/10.1126/science.1118391>.

- LINDAHL, E. Molecular Dynamics Simulations. *In: KUKOL, A. (ed.). Molecular Modeling of Proteins. Methods in Molecular Biology (Methods and Protocols)*. New York, NY: Humana Press, 2015. p. 3–26.
- LIU, Q.; WANG, J.; BOYD, B. J. Peptide-based biosensors. *Talanta*, vol. 136, p. 114–127, May 2015. <https://doi.org/10.1016/j.talanta.2014.12.020>.
- LIU, T.; SHI, K.; LI, W. Deep learning methods improve linear B-cell epitope prediction. *Bio-Data Mining*, vol. 13, no. 1, Apr. 2020. <https://doi.org/10.1186/s13040-020-00211-0>.
- LOWE, C. R. Biosensors. *Trends in Biotechnology*, vol. 2, no. 3, p. 59–65, May 1984. [https://doi.org/10.1016/0167-7799\(84\)90011-8](https://doi.org/10.1016/0167-7799(84)90011-8).
- LU, R.; ZHAO, X.; LI, J.; NIU, P.; YANG, B.; WU, H.; WANG, W.; SONG, H.; HUANG, B.; ZHU, N.; BI, Y.; MA, X.; ZHAN, F.; WANG, L.; HU, T.; ZHOU, H.; HU, Z.; ZHOU, W.; ZHAO, L.; CHEN, J. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet*, vol. 395, no. 10224, p. 565–574, Jan. 2020. [https://doi.org/10.1016/s0140-6736\(20\)30251-8](https://doi.org/10.1016/s0140-6736(20)30251-8).
- LYTRAS, S.; HUGHES, J.; MARTIN, D.; SWANEPOEL, P.; DE KLERK, A.; LOURENS, R.; KOSAKOVSKY POND, S. L.; XIA, W.; JIANG, X.; ROBERTSON, D. L. Exploring the Natural Origins of SARS-CoV-2 in the Light of Recombination. *Genome Biology and Evolution*, vol. 14, no. 2, Feb. 2022. <https://doi.org/10.1093/gbe/evac018>.
- MAIER, J. A.; MARTINEZ, C.; KASAVAJHALA, K.; WICKSTROM, L.; HAUSER, K. E.; SIMMERLING, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, vol. 11, no. 8, p. 3696–3713, Jul. 2015. <https://doi.org/10.1021/acs.jctc.5b00255>.
- MANAVALAN, B.; GOVINDARAJ, R. G.; SHIN, T. H.; KIM, M. O.; LEE, G. iBCE-EL: A New Ensemble Learning Framework for Improved Linear B-Cell Epitope Prediction. *Frontiers in Immunology*, vol. 9, Jul. 2018. <https://doi.org/10.3389/fimmu.2018.01695>.
- MARIUZZA, R. A.; PHILLIPS, S. E. V.; POLJAK, R. J. The Structural Basis of Antigen-Antibody Recognition. *Annual Review of Biophysics and Biophysical Chemistry*, vol. 16, no. 1, p. 139–159, Jun. 1987. <https://doi.org/10.1146/annurev.bb.16.060187.001035>.
- MOHD, H. A.; AL-TAWFIQ, J. A.; MEMISH, Z. A. Middle East Respiratory Syndrome Coronavirus (MERS-CoV) origin and animal reservoir. *Virology Journal*, vol. 13, no. 1, Jun. 2016. <https://doi.org/10.1186/s12985-016-0544-0>.

MOTTIN, M. **Dinâmica Molecular do receptor ativador da proliferação de peroxissomos gama: associação com ligantes e proteínas corregulatórias**. 2015. Tese – Universidade Estadual de Campinas (UNICAMP), 2015. Available at: repositorio.unicamp.br/Busca/Download?codigoArquivo=505773.

NELSON, D. L.; COX, M. M. **Princípios de Bioquímica de Lehninger**. 6th ed. Porto Alegre: Artmed, 2014.

NOTREDAME, C.; HIGGINS, D. G.; HERINGA, J. T-coffee: a novel method for fast and accurate multiple sequence alignment 1 Edited by J. Thornton. *Journal of Molecular Biology*, vol. 302, no. 1, p. 205–217, Sep. 2000. <https://doi.org/10.1006/jmbi.2000.4042>.

OLIVEIRA, G. S. **Análise da interação molecular proteína-herbicida através de simulação computacional: aplicação no desenvolvimento de nanobiossensores**. 2013. Tese – Universidade Federal de São Carlos (UFSCar), 2013. Available at: <https://repositorio.ufscar.br/handle/ufscar/6267?show=full>.

ONG, S. W. X.; CHIEW, C. J.; ANG, L. W.; MAK, T.-M.; CUI, L.; TOH, M. P. H. S.; LIM, Y. D.; LEE, P. H.; LEE, T. H.; CHIA, P. Y.; MAURER-STROH, S.; LIN, R. T. P.; LEO, Y.-S.; LEE, V. J.; LYE, D. C.; YOUNG, B. E. Clinical and virological features of SARS-CoV-2 variants of concern: a retrospective cohort study comparing B.1.1.7 (Alpha), B.1.315 (Beta), and B.1.617.2 (Delta). **Clinical Infectious Diseases**, p. 1–9, Aug. 2021. <https://doi.org/10.1093/cid/ciab721>.

PAIZS, B.; SUHAI, S. Fragmentation pathways of protonated peptides. **Mass Spectrometry Reviews**, vol. 24, no. 4, p. 508–548, Aug. 2005. <https://doi.org/10.1002/mas.20024>.

PEIRIS, J. S. M. Coronaviruses. **Clinical Virology**. [S. l.]: ASM Press, 2016. p. 1243–1265. <https://doi.org/10.1128/9781555819439.ch52>.

PEKAR, J. E.; MAGEE, A.; PARKER, E.; MOSHIRI, N.; IZHIKEVICH, K.; HAVENS, J. L.; GANGAVARAPU, K.; MALPICA SERRANO, L. M.; CRITS-CHRISTOPH, A.; MATTESSON, N. L.; ZELLER, M.; LEVY, J. I.; WANG, J. C.; HUGHES, S.; LEE, J.; PARK, H.; PARK, M.-S.; CHING ZI YAN, K.; TZER PIN LIN, R.; MAT ISA, M. N. SARS-CoV-2 emergence very likely resulted from at least two zoonotic events. Feb. 2022. **Zenodo (Pre-print)**. Available at: <https://zenodo.org/record/6291628#.YiJv6xtv9kh>.

PELLEQUER, J. L.; WESTHOF, E. PREDITOP: A program for antigenicity prediction. **Journal of Molecular Graphics**, vol. 11, no. 3, p. 204–210, Sep. 1993. [https://doi.org/10.1016/0263-7855\(93\)80074-2](https://doi.org/10.1016/0263-7855(93)80074-2).

PETTERSEN, E. F.; GODDARD, T. D.; HUANG, C. C.; MENG, E. C.; COUCH, G. S.; CROLL, T. I.; MORRIS, J. H.; FERRIN, T. E. UCSF ChimeraX : Structure visualization for researchers, educators, and developers. **Protein Science**, vol. 30, no. 1, p. 70–82, Oct. 2020. <https://doi.org/10.1002/pro.3943>.

POH, C. M.; CARISSIMO, G.; WANG, B.; AMRUN, S. N.; LEE, C. Y.-P.; CHEE, R. S.-L.; FONG, S.-W.; YEO, N. K.-W.; LEE, W.-H.; TORRES-RUESTA, A.; LEO, Y.-S.; CHEN, M. I.-C.; TAN, S.-Y.; CHAI, L. Y. A.; KALIMUDDIN, S.; KHENG, S. S. G.; THIEN, S.-Y.; YOUNG, B. E.; LYE, D. C.; HANSON, B. J. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. **Nature Communications**, vol. 11, no. 1, Jun. 2020. <https://doi.org/10.1038/s41467-020-16638-2>.

POTOCNAKOVA, L.; Bhide, M.; PULZOVA, L. B. An Introduction to B-Cell Epitope Mapping and In Silico Epitope Prediction. **Journal of Immunology Research**, vol. 2016, p. 1–11, Dec. 2016. <https://doi.org/10.1155/2016/6760830>.

RABI, F. A.; AL ZOUBI, M. S.; KASASBEH, G. A.; SALAMEH, D. M.; AL-NASSER, A. D. SARS-CoV-2 and Coronavirus Disease 2019: What We Know So Far. **Pathogens**, vol. 9, no. 3, Mar. 2020. <https://doi.org/10.3390/pathogens9030231>.

RAINES, R. T.; WENNEMERS, H. Peptides on the Rise. **Accounts of Chemical Research**, vol. 50, no. 10, p. 2419–2419, Oct. 2017. <https://doi.org/10.1021/acs.accounts.7b00471>.

RANDO, H. M.; MACLEAN, A. L.; LEE, A. J.; LORDAN, R.; RAY, S.; BANSAL, V.; SKELLY, A. N.; SELL, E.; DZIAK, J. J.; SHINHOLSTER, L.; D'AGOSTINO MCGOWAN, L.; BEN GUEBILA, M.; WELLHAUSEN, N.; KNYAZEY, S.; BOCA, S. M.; CAPONE, S.; QI, Y.; PARK, Y.; MAI, D.; SUN, Y. Pathogenesis, Symptomatology, and Transmission of SARS-CoV-2 through Analysis of Viral Genomics and Structure. **mSystems**, vol. 6, no. 5, Oct. 2021. <https://doi.org/10.1128/msystems.00095-21>.

ROBERT, X.; GOUET, P. Deciphering key features in protein structures with the new ENDs-script server. **Nucleic Acids Research**, vol. 42, no. W1, p. W320–W324, Apr. 2014. <https://doi.org/10.1093/nar/gku316>.

- ROCKSTROH, A.; MOGES, B.; BARZON, L.; SINIGAGLIA, A.; PALÙ, G.; KUMBUK-GOLLA, W.; SCHMIDT-CHANASIT, J.; SARNO, M.; BRITES, C.; MOREIRA-SOTO, A.; DREXLER, J. F.; FERREIRA, O. C.; ULBERT, S. Specific detection of dengue and Zika virus antibodies using envelope proteins with mutations in the conserved fusion loop. **Emerging Microbes & Infections**, vol. 6, no. 1, p. 1–9, Jan. 2017. <https://doi.org/10.1038/emi.2017.87>.
- ROE, D. R.; CHEATHAM, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. **Journal of Chemical Theory and Computation**, vol. 9, no. 7, p. 3084–3095, Jun. 2013. <https://doi.org/10.1021/ct400341p>.
- SAHA, S.; RAGHAVA, G. P. S. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. **Proteins: Structure, Function, and Bioinformatics**, vol. 65, no. 1, p. 40–48, Aug. 2006. <https://doi.org/10.1002/prot.21078>.
- SALAM, F.; TOTHILL, I. E. Detection of Salmonella typhimurium using an electrochemical immunosensor. **Biosensors and Bioelectronics**, vol. 24, no. 8, p. 2630–2636, Apr. 2009. <https://doi.org/10.1016/j.bios.2009.01.025>.
- SALOMON-FERRER, R.; CASE, D. A.; WALKER, R. C. An overview of the Amber biomolecular simulation package. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, vol. 3, no. 2, p. 198–210, 25 Sep. 2012. <https://doi.org/10.1002/wcms.1121>.
- SANCHEZ-TRINCADO, J. L.; GOMEZ-PEROSANZ, M.; RECHE, P. A. Fundamentals and Methods for T- and B-Cell Epitope Prediction. **Journal of Immunology Research**, vol. 2017, p. 1–14, Dec. 2017. <https://doi.org/10.1155/2017/2680160>.
- SANDA, M.; MORRISON, L.; GOLDMAN, R. N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. **Analytical Chemistry**, vol. 93, no. 4, p. 2003–2009, Feb. 2021. <https://doi.org/10.1021/acs.analchem.0c03173>.
- SELA, M. Antigenicity: Some Molecular Aspects. **Science**, vol. 166, no. 3911, p. 1365–1374, Dec. 1969. <https://doi.org/10.1126/science.166.3911.1365>.
- SHAJAHAN, A.; SUPEKAR, N. T.; GLEINICH, A. S.; AZADI, P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. **Glycobiology**, vol. 30, no. 12, p. 981–988, May 2020. <https://doi.org/10.1093/glycob/cwaa042>.
- SHRAKE, A.; RUPLEY, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. **Journal of Molecular Biology**, vol. 79, no. 2, p. 351–371, Sep. 1973. [https://doi.org/10.1016/0022-2836\(73\)90011-9](https://doi.org/10.1016/0022-2836(73)90011-9).

SIMMONS, G.; ZMORA, P.; GIERER, S.; HEURICH, A.; PÖHLMANN, S. Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. **Antiviral Research**, vol. 100, no. 3, p. 605–614, Dec. 2013. <https://doi.org/10.1016/j.antiviral.2013.09.028>.

SU, S.; WONG, G.; SHI, W.; LIU, J.; LAI, A. C. K.; ZHOU, J.; LIU, W.; BI, Y.; GAO, G. F. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. **Trends in Microbiology**, vol. 24, no. 6, p. 490–502, Jun. 2016. <https://doi.org/10.1016/j.tim.2016.03.003>.

SUN, P.; QIE, S.; LIU, Z.; REN, J.; LI, K.; XI, J. Clinical characteristics of 50 466 hospitalized patients with 2019-nCoV infection. **Journal of Medical Virology**, vol. 92, p. 612–617, Mar. 2020. <https://doi.org/10.1002/jmv.25735>.

SÜSSMUTH, R. D.; MAINZ, A. Nonribosomal Peptide Synthesis-Principles and Prospects. **Angewandte Chemie International Edition**, vol. 56, no. 14, p. 3770–3821, Mar. 2017. <https://doi.org/10.1002/anie.201609079>.

VAN REGENMORTEL, M. H. V. What Is a B-Cell Epitope? *In*: SCHUTKOWSKI, M.; REINEKE, U. (eds.). **Epitope Mapping Protocols**. New Jersey: Humana Press, 2009. p. 3–20.

VERLET, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. **Physical Review**, vol. 159, no. 1, p. 98–103, 5 Jul. 1967. <https://doi.org/10.1103/physrev.159.98>.

VIGERUST, D. J.; SHEPHERD, V. L. Virus glycosylation: role in virulence and immune interactions. **Trends in Microbiology**, vol. 15, no. 5, p. 211–218, May 2007. <https://doi.org/10.1016/j.tim.2007.03.003>.

VLIEGHE, P.; LISOWSKI, V.; MARTINEZ, J.; KHRESTCHATISKY, M. Synthetic therapeutic peptides: science and market. **Drug Discovery Today**, vol. 15, no. 1-2, p. 40–56, Jan. 2010. <https://doi.org/10.1016/j.drudis.2009.10.009>.

WAJNBERG, A.; AMANAT, F.; FIRPO, A.; ALTMAN, D. R.; BAILEY, M. J.; MANSOUR, M.; MCMAHON, M.; MEADE, P.; MENDU, D. R.; MUELLERS, K.; STADLBAUER, D.; STONE, K.; STROHMEIER, S.; SIMON, V.; ABERG, J.; REICH, D. L.; KRAMMER, F.; CORDON-CARDO, C. Robust neutralizing antibodies to SARS-CoV-2 infection persist for months. **Science**, vol. 370, no. 6521, p. 1227–1230, Oct. 2020. <https://doi.org/10.1126/science.abd7728>.

WANG, D.; HU, B.; HU, C.; ZHU, F.; LIU, X.; ZHANG, J.; WANG, B.; XIANG, H.; CHENG, Z.; XIONG, Y.; ZHAO, Y.; LI, Y.; WANG, X.; PENG, Z. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. **JAMA**, vol. 323, no. 11, p. 1061–1069, Feb. 2020. <https://doi.org/10.1001/jama.2020.1585>.

WATANABE, Y.; ALLEN, J. D.; WRAPP, D.; MCLELLAN, J. S.; CRISPIN, M. Site-specific glycan analysis of the SARS-CoV-2 spike. **Science**, vol. 369, no. 6501, p. 330–333, May 2020. <https://doi.org/10.1126/science.abb9983>.

WEISS, S. R.; LEIBOWITZ, J. L. Coronavirus Pathogenesis. *In*: MARAMOROSCH, K.; SHATKIN, A. J.; MURPHY, F. A. (eds.). **Advances in Virus Research**. [S. l.]: Academic Press, 2011. p. 85–164. <https://doi.org/10.1016/b978-0-12-385885-6.00009-2>.

WHO. **Situation Report 11**. Geneva: World Health Organization, Jan. 2020a. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4. Accessed on: 15 May 2020.

WHO. **Situation Report 51**. Geneva: World Health Organization, Mar. 2020b. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10. Accessed on: 20 Apr. 2020.

WHO. **Situation Report 158**. Geneva: World Health Organization, Jun. 2020c. Available at: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200626-covid-19-sitrep-158.pdf?sfvrsn=1d1aae8a_2. Accessed on: 26 Jun. 2020.

WINTJENS, R.; BIFANI, A. M.; BIFANI, P. Impact of glycan cloud on the B-cell epitope prediction of SARS-CoV-2 Spike protein. **npj Vaccines**, vol. 5, no. 1, Sep. 2020. <https://doi.org/10.1038/s41541-020-00237-9>.

WOO, H.; PARK, S.-J.; CHOI, Y. K.; PARK, T.; TANVEER, M.; CAO, Y.; KERN, N. R.; LEE, J.; YEOM, M. S.; CROLL, T. I.; SEOK, C.; IM, W. Developing a Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein Model in a Viral Membrane. **The Journal of Physical Chemistry B**, vol. 124, no. 33, p. 7128–7137, Jun. 2020. <https://doi.org/10.1021/acs.jpcc.0c04553>.

WOROBAY, M.; LEVY, J. I.; SERRANO, L. M. M.; CRITS-CHRISTOPH, A.; PEKAR, J. E.; GOLDSTEIN, S. A.; RASMUSSEN, A. L.; KRAEMER, M. U. G.; NEWMAN, C.; KO-OPMANS, M. P. G.; SUCHARD, M. A.; WERTHEIM, J. O.; LEMEY, P.; ROBERTSON, D.

L.; GARRY, R. F.; HOLMES, E. C.; RAMBAUT, A.; ANDERSEN, K. G. The Huanan market was the epicenter of SARS-CoV-2 emergence. Feb. 2022. **Zenodo (Preprint)**. Available at: <https://zenodo.org/record/6299116#.YiJyCBtv9kg>.

WRAPP, D.; WANG, N.; CORBETT, K. S.; GOLDSMITH, J. A.; HSIEH, C.-L.; ABIONA, O.; GRAHAM, B. S.; MCLELLAN, J. S. Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. **Science**, vol. 367, no. 6483, p. 1260–1263, Feb. 2020. <https://doi.org/10.1126/science.abb2507>.

XIAO, K.; ZHAI, J.; FENG, Y.; ZHOU, N.; ZHANG, X.; ZOU, J.-J.; LI, N.; GUO, Y.; LI, X.; SHEN, X.; ZHANG, Z.; SHU, F.; HUANG, W.; LI, Y.; ZHANG, Z.; CHEN, R.-A.; WU, Y.-J.; PENG, S.-M.; HUANG, M.; XIE, W.-J. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. **Nature**, vol. 583, no. 7815, p. 286–289, May 2020. <https://doi.org/10.1038/s41586-020-2313-x>.

XU, H.; ZHONG, L.; DENG, J.; PENG, J.; DAN, H.; ZENG, X.; LI, T.; CHEN, Q. High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. **International Journal of Oral Science**, vol. 12, no. 8, 1 Feb. 2020. <https://doi.org/10.1038/s41368-020-0074-x>.

XU, W.; WANG, M.; YU, D.; ZHANG, X. Variations in SARS-CoV-2 Spike Protein Cell Epitopes and Glycosylation Profiles During Global Transmission Course of COVID-19. **Frontiers in Immunology**, vol. 11, Sep. 2020. <https://doi.org/10.3389/fimmu.2020.565278>.

YAN, R.; ZHANG, Y.; LI, Y.; XIA, L.; GUO, Y.; ZHOU, Q. Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. **Science**, vol. 367, no. 6485, p. 1444–1448, Mar. 2020. <https://doi.org/10.1126/science.abb2762>.

YAZAWA, K.; NUMATA, K. Recent Advances in Chemoenzymatic Peptide Syntheses. **Molecules**, vol. 19, no. 9, p. 13755–13774, Sep. 2014. <https://doi.org/10.3390/molecules190913755>.

YOO, E.-H.; LEE, S.-Y. Glucose Biosensors: An Overview of Use in Clinical Practice. **Sensors**, vol. 10, no. 5, p. 4558–4576, May 2010. <https://doi.org/10.3390/s100504558>.

YU, L.; ZHANG, Y.; HU, C.; WU, H.; YANG, Y.; HUANG, C.; JIA, N. Highly sensitive electrochemical impedance spectroscopy immunosensor for the detection of AFB1 in olive oil. **Food Chemistry**, vol. 176, p. 22–26, Jun. 2015. <https://doi.org/10.1016/j.foodchem.2014.12.030>.

YUAN, M.; LIU, H.; WU, N. C.; LEE, C.-C. D.; ZHU, X.; ZHAO, F.; HUANG, D.; YU, W.; HUA, Y.; TIEN, H.; ROGERS, T. F.; LANDAIS, E.; SOK, D.; JARDINE, J. G.; BURTON, D. R.; WILSON, I. A. Structural basis of a shared antibody response to SARS-CoV-2. **Science**, vol. 369, no. 6507, p. 1119–1123, Aug. 2020. <https://doi.org/10.1126/science.abd2321>.

ZHANG, T.; WU, Q.; ZHANG, Z. Probable Pangolin Origin of SARS-CoV-2 Associated with the COVID-19 Outbreak. **Current Biology**, vol. 30, no. 7, p. 1346–1351, Mar. 2020. <https://doi.org/10.1016/j.cub.2020.03.022>.

ZHANG, Y.; ZHAO, W.; MAO, Y.; CHEN, Y.; WANG, S.; ZHONG, Y.; SU, T.; GONG, M.; DU, D.; LU, X.; CHENG, J.; YANG, H. Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins. **Molecular & Cellular Proteomics**, vol. 20, p. 100058, Feb. 2021. <https://doi.org/10.1074/mcp.ra120.002295>.

ZHOU, F.; YU, T.; DU, R.; FAN, G.; LIU, Y.; LIU, Z.; XIANG, J.; WANG, Y.; SONG, B.; GU, X.; GUAN, L.; WEI, Y.; LI, H.; WU, X.; XU, J.; TU, S.; ZHANG, Y.; CHEN, H.; CAO, B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. **The Lancet**, vol. 395, no. 10229, p. 1054–1062, Mar. 2020. [https://doi.org/10.1016/s0140-6736\(20\)30566-3](https://doi.org/10.1016/s0140-6736(20)30566-3).

ZHOU, P.; YANG, X.-L.; WANG, X.-G.; HU, B.; ZHANG, L.; ZHANG, W.; SI, H.-R.; ZHU, Y.; LI, B.; HUANG, C.-L.; CHEN, H.-D.; CHEN, J.; LUO, Y.; GUO, H.; JIANG, R.-D.; LIU, M.-Q.; CHEN, Y.; SHEN, X.-R.; WANG, X.; ZHENG, X.-S. A pneumonia outbreak associated with a new coronavirus of probable bat origin. **Nature**, vol. 579, no. 7798, p. 270–273, Feb. 2020. <https://doi.org/10.1038/s41586-020-2012-7>.

ZHU, N.; ZHANG, D.; WANG, W.; LI, X.; YANG, B.; SONG, J.; ZHAO, X.; HUANG, B.; SHI, W.; LU, R.; NIU, P.; ZHAN, F.; MA, X.; WANG, D.; XU, W.; WU, G.; GAO, G. F.; TAN, W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. **New England Journal of Medicine**, vol. 382, no. 8, p. 727–733, Jan. 2020. <https://doi.org/10.1056/nejmoa2001017>.

APÊNDICE A

As tabelas que compõe o “Apêndice A” do presente trabalho encontram-se em documento Excel separado nomeado “Apêndice A – Dissertação Anderson Pinheiro.xlsx” devido ao grande volume dos dados compilados. São parte do documento: planilha DLBEpitope, planilha BepiPred 2.0 e planilha iBCE-EL.

APÊNDICE B

Figura 19. Alinhamento aplicado aos dados brutos da análise no servidor iBCE-EL.

```

1      10      20      30      40      50      60      70      80      90      100     110     120     130     140     150
YP_009724390.1 MFVFLVLLPLVSSQCVNLTTTRTQLPPAYTNSFTRGVVYYPDKVFRSSVLHSTQDLFLPFFSNVTWFHAIHVSGTNGTKRFDNPLPFNDGVYFASTEKSNIRGWIFGTTLDSKTQSLIVNNATNVVVKVCEFCQFCNDPFLGVYYHKNNKSWMESEFR
EL19             .TRTQLPPAYTNSFTRGVVYYPDK.
EL20             .RTQLPPAYTNSFTRGVVYYPDKV.
EL69             .VSGTNGTKRFDNPLPFNDGVY.
EL136           .NDPFLGVYYHKNNKSWMESEFR
EL302
EL436
EL438
EL441
EL457
EL459
EL1114
EL1115
EL1116
EL1117
EL1118
EL1119
EL1120
EL1123
EL1124
EL1125
EL1126
EL1127

160     170     180     190     200     210     220     230     240     250     260     270     280     290     300     310
YP_009724390.1 VYSSANNCTFEYVSQPFLMDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQTLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETKCTLKSFVEKGIYQTS
EL19
EL20
EL69
EL136
EL302
EL436
EL438
EL441
EL457
EL459
EL1114
EL1115
EL1116
EL1117
EL1118
EL1119
EL1120
EL1123
EL1124
EL1125
EL1126
EL1127

320     330     340     350     360     370     380     390     400     410     420     430     440     450     460     470
YP_009724390.1 NFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVLVNSASFSTFKCYGVSPTKLNDLCFTNVYADSVIRGDEVQRQIAPGGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNLYRLFRKSNLKPFFERDISTEIIYQ
EL19
EL20
EL69
EL136
EL302 NFRVQPT.
EL436
EL438
EL441
EL457
EL459
EL1114
EL1115
EL1116
EL1117
EL1118
EL1119
EL1120
EL1123
EL1124
EL1125
EL1126
EL1127

```

```

      480      490      500      510      520      530      540      550      560      570      580      590      600      610      620      630
YP_009724390.1 AGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYPYRVVLSFELLHAPATVCGPKKSTNLVKNKCVNFNFNGLTGTGVLTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQDVNCTEVPVAIHADQLTPT
EL19 .....
EL20 .....
EL69 .....
EL136 .....
EL302 .....
EL436 .....
EL438 .....
EL441 .....
EL457 AGSTP .....
EL459 AGSTPCN .....
EL1114 .....
EL1115 .....
EL1116 .....
EL1117 .....
EL1118 .....
EL1119 .....
EL1120 .....
EL1123 .....
EL1124 .....
EL1125 .....
EL1126 .....
EL1127 .....

      640      650      660      670      680      690      700      710      720      730      740      750      760      770      780      790
YP_009724390.1 WRVYSTGSNVFQTRAGCLIGAHEVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNNSIAIPTNFTISVTEILPVSMTKTSVDCTMYICGDSSTECNLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYK
EL19 .....
EL20 .....
EL69 .....
EL136 .....
EL302 .....
EL436 .....
EL438 .....
EL441 .....
EL457 .....
EL459 .....
EL1114 .....
EL1115 .....
EL1116 .....
EL1117 .....
EL1118 .....
EL1119 .....
EL1120 .....
EL1123 .....
EL1124 .....
EL1125 .....
EL1126 .....
EL1127 .....

      800      810      820      830      840      850      860      870      880      890      900      910      920      930      940
YP_009724390.1 TPIKDFGGFNFSQILPDPSKPSKRSFIEDLLFNKVTLDAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPLLLTDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSTASALGKL
EL19 .....
EL20 .....
EL69 .....
EL136 .....
EL302 .....
EL436 .....
EL438 .....
EL441 .....
EL457 .....
EL459 .....
EL1114 .....
EL1115 .....
EL1116 .....
EL1117 .....
EL1118 .....
EL1119 .....
EL1120 .....
EL1123 .....
EL1124 .....
EL1125 .....
EL1126 .....
EL1127 .....
```

```

          950      960      970      980      990      1000      1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
YP_009724390.1 QDVVNQNAQALNTLVKQLSSNFGAISSVLNDILSRDKVEAEVQIDRLITGRQLQTYVTQQILIRAAEIRASANLAATKMSECVLGQSKRVDFCGKGYHLSFPQSAPHGVVFLHVTYVPAQEKNFTTAPAICHGDKAHFPREGVFVSNGTHWFTQ
EL19 .....
EL20 .....
EL69 .....
EL136 .....
EL302 .....
EL436 .....
EL438 .....
EL441 .....
EL457 .....
EL459 .....
EL1114 .....
EL1115 .....
EL1116 .....
EL1117 .....
EL1118 .....
EL1119 .....
EL1120 .....
EL1123 .....
EL1124 .....
EL1125 .....
EL1126 .....
EL1127 .....

          1110      1120      1130      1140      1150      1160      1170      1180      1190      1200      1210      1220      1230      1240      1250      1260
YP_009724390.1 RNFYEPQIITTDNTFVSGNCDVVIGIVNNTVYDPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFIAGLIAIVMVTIMLCCMTSCCSCCLKGCCSCGSCCKFDEDDSEPV
EL19 .....
EL20 .....
EL69 .....
EL136 .....
EL302 .....
EL436 .....
EL438 .....
EL441 .....
EL457 .....
EL459 .....
EL1114 .....VIGIVNNTVYDPLQPELDSFKE.....
EL1115 .....IGIVNNTVYDPLQPELDSFKEE.....
EL1116 .....GIVNNTVYDPLQPELDSFKEEL.....
EL1117 .....IVNNTVYDPLQPELDSFKEELD.....
EL1118 .....VNNTVYDPLQPELDSFKEELDK.....
EL1119 .....NNTVYDPLQPELDSFKEELDKY.....
EL1120 .....NTVYDPLQPELDSFKEELDKYF.....
EL1123 .....YDPLQPELDSFKEELDKYFKNH.....
EL1124 .....DPLQPELDSFKEELDKYFKNHT.....
EL1125 .....PLQPELDSFKEELDKYFKNHTS.....
EL1126 .....LQPELDSFKEELDKYFKNHTSP.....
EL1127 .....QPELDSFKEELDKYFKNHTSPD.....

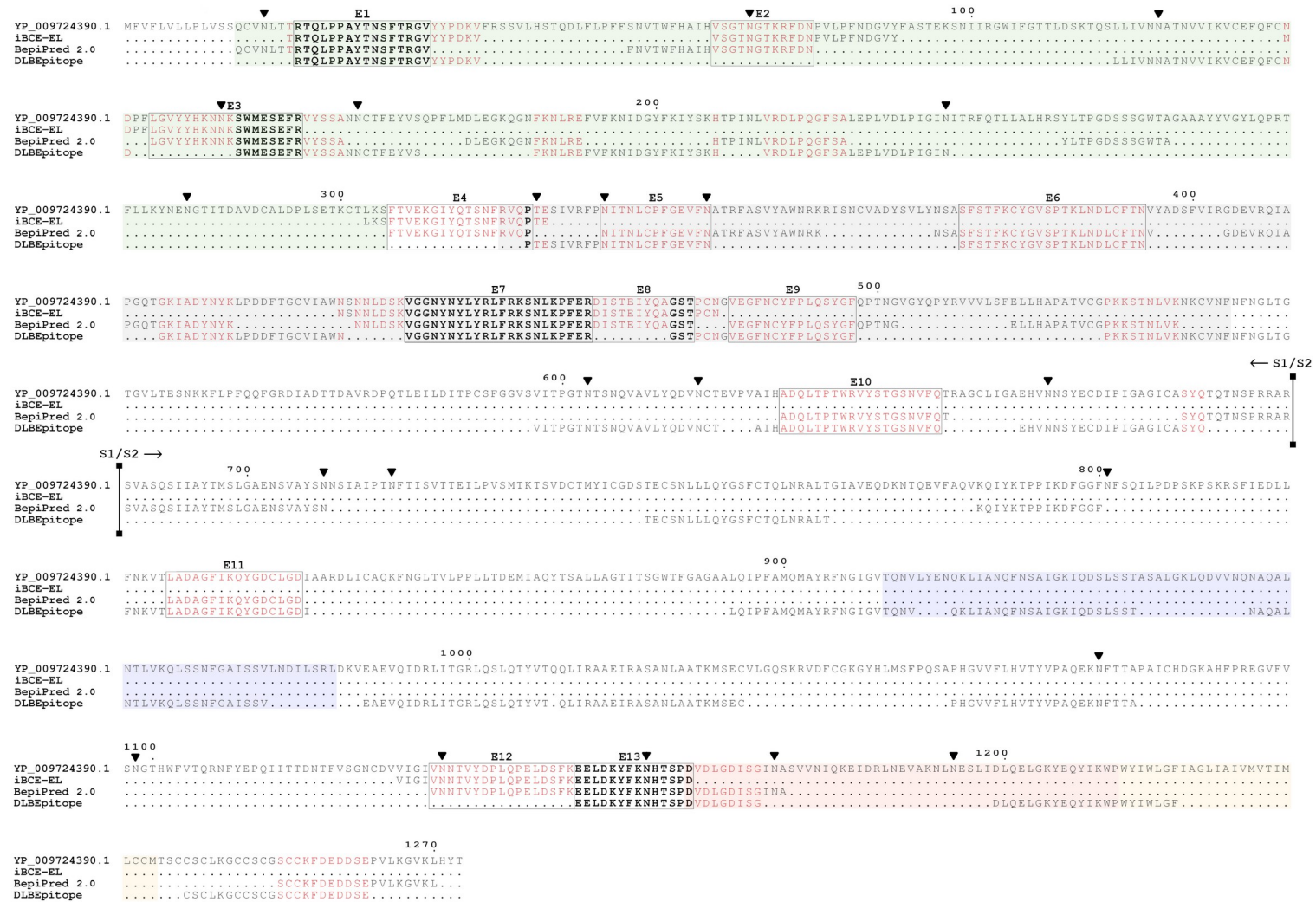
          1270
YP_009724390.1 LKGVKLIHYT
EL19 .....
EL20 .....
EL69 .....
EL136 .....
EL302 .....
EL436 .....
EL438 .....
EL441 .....
EL457 .....
EL459 .....
EL1114 .....
EL1115 .....
EL1116 .....
EL1117 .....
EL1118 .....
EL1119 .....
EL1120 .....
EL1123 .....
EL1124 .....
EL1125 .....
EL1126 .....
EL1127 .....

```

Fonte: o autor.

APÊNDICE C

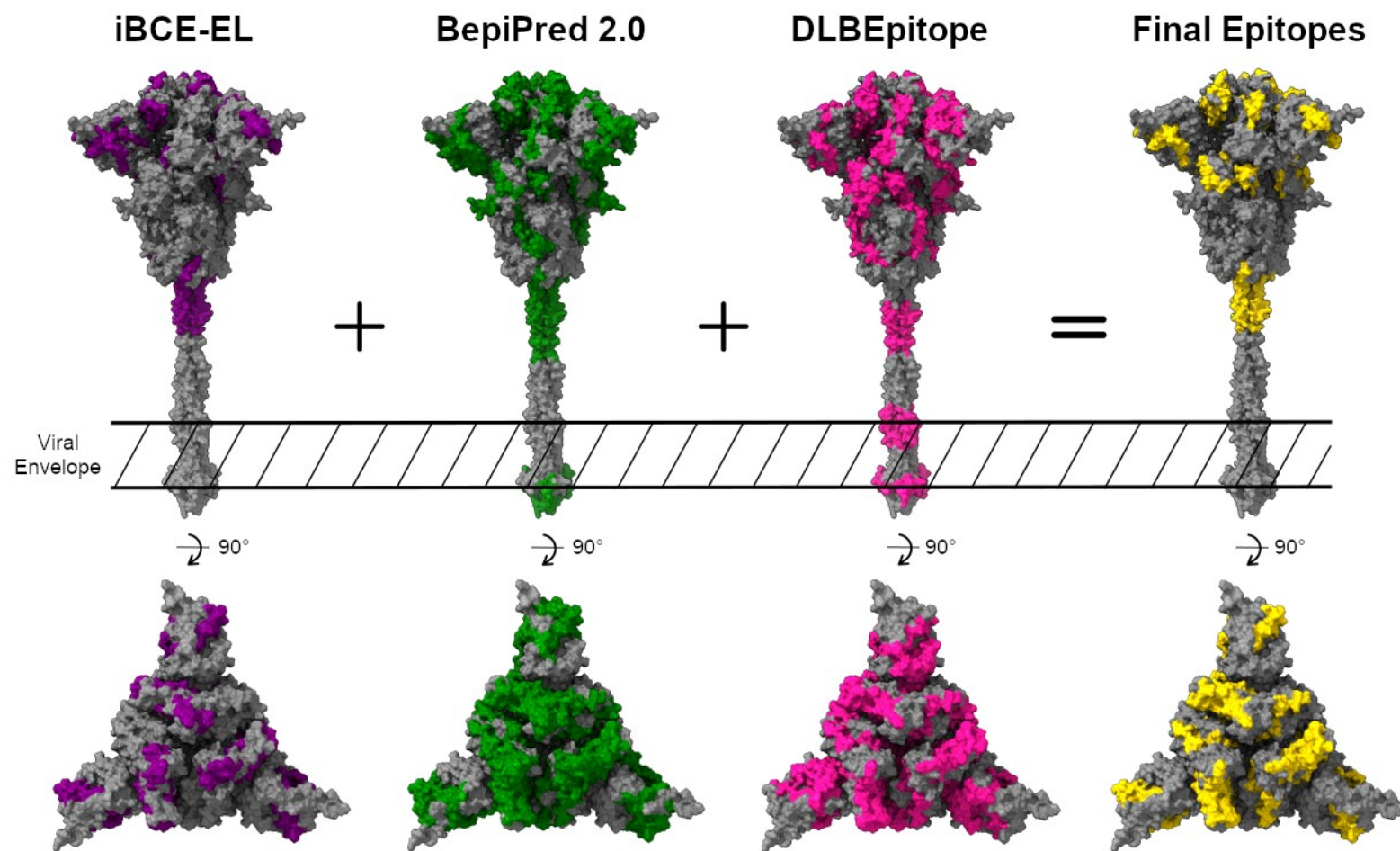
Figura 20. Alinhamento comparativo completo, com destaque para os epítomos imunodominantes identificados.



Fonte: o autor.

APÊNDICE D

Figura 21. Estrutura tridimensional da proteína *Spike* com destaque para a área coberta pelos epítomos identificados por aplicação individual.



Fonte: o autor.

APÊNDICE E

Tabela 4. Evolução genética e a presença de mutações nos epítomos mapeados dentre as principais variantes de SARS-CoV-2.

Epitope	Accumuled mutations					
	Alpha (B.1.1.7)	Beta (B.1.351)	Gamma (P.1)	Delta (B.1.617.2)	Zeta (P.2)	Omicron (B.1.1.529)
E1	-	-	P26S	-	-	-
E2	Δ70	D80A	-	-	-	Δ70
E3	Δ144	-	-	G142D, E156G, Δ157-158	-	G142D, Δ143-145
E4	-	-	-	-	-	-
E5	-	-	-	-	-	G339D
E6	-	-	-	-	-	S373P, S375F
E7	-	-	-	L452R	-	G446S
E8	-	-	-	T478K	-	S477N, T478K
E9	-	E484K	E484K	-	E484K	E484A, Q493R, G496S
E10	-	-	-	-	-	-
E11	-	-	-	-	-	-
E12	-	-	-	-	-	-
E13	-	-	-	-	-	-

Fonte: o autor.