

UNIVERSIDADE FEDERAL DE ALFENAS

ANA DE SOUZA SANTOS

**AVALIAÇÃO DA IMUNOGENICIDADE DE EPÍTOPOS DA PROTEÍNA
ESTRUTURAL *SPIKE* DO SARS-CoV-2 RACIONALMENTE SELECIONADOS POR
VACINOLOGIA REVERSA E ASSOCIADOS AO BCG VACINAL COMO
CARREADOR E INDUTOR DA IMUNIDADE INATA TREINADA**

Alfenas/MG

2022

ANA DE SOUZA SANTOS

AVALIAÇÃO DA IMUNOGENICIDADE DE EPÍTOPOS DA PROTEÍNA ESTRUTURAL
SPIKE DO SARS-CoV-2 RACIONALMENTE SELECIONADOS POR VACINOLOGIA
REVERSA E ASSOCIADOS AO BCG VACINAL COMO CARREADOR E INDUTOR DA
IMUNIDADE INATA TREINADA

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre no Programa de Pós-Graduação em Ciências Biológicas pela Universidade Federal de Alfenas.

Área de concentração: Interação Patógeno-Hospedeiro

Orientador: Prof. Dr. Leonardo Augusto de Almeida

Alfenas/MG

2022

Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central

Santos, Ana de Souza.

Avaliação da imunogenicidade de epítomos da proteína estrutural *spike* do SARS-CoV-2 racionalmente selecionados por vacinologia reversa e associados ao BCG vacinal como carreador e indutor da imunidade inata treinada / Ana de Souza Santos. - Alfenas, MG, 2022.

96 f. : il. -

Orientador(a): Leonardo Augusto de Almeida.

Dissertação (Mestrado em Ciências Biológicas) - Universidade Federal de Alfenas, Alfenas, MG, 2022.

Bibliografia.

1. COVID-19. 2. Vacinologia reversa. 3. Epítomos. 4. BCG. 5. Imunidade inata treinada. I. Almeida, Leonardo Augusto de, orient. II. Título.

ANA DE SOUZA SANTOS

**AVALIAÇÃO DA IMUNOGENICIDADE DE EPÍTOPOS DA PROTEÍNA
ESTRUTURAL SPIKE DO SARS-CoV-2 RACIONALMENTE SELECIONADOS
POR VACINOLOGIA REVERSA E ASSOCIADOS AO BCG VACINAL
COMO CARREADOR E INDUTOR DA IMUNIDADE INATA TREINADA**

A Banca examinadora abaixo-assinada aprova a Dissertação/Tese apresentada como parte dos requisitos para a obtenção do título de Mestra em Ciências Biológicas pela Universidade Federal de Alfenas. Área de concentração: Interação Patógeno-Hospedeiro.

Aprovada em: 6 de outubro de 2022.

Prof. Dr. Leonardo Augusto de Almeida
Instituição: Universidade Federal de Alfenas - UNIFAL-MG

Profa. Dra. Marina Quádrio Raposo Branco
Rodrigues Instituição: Universidade Federal de São
João del-Rei - UFSJ

Prof. Rômulo Dias Novaes
Instituição: Universidade Federal de Alfenas - UNIFAL-MG



Documento assinado eletronicamente por **Leonardo Augusto de Almeida, Professor do Magistério Superior**, em 06/10/2022, às 12:07, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Rômulo Dias Novaes, Professor do Magistério Superior**, em 06/10/2022, às 12:08, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Marina Quádrio Raposo Branco Rodrigues, Usuário Externo**, em 07/10/2022, às 07:17, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://sei.unifal-mg.edu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0832964** e o código CRC **03D32657**.

AGRADECIMENTOS

À Deus, por me inspirar e conceder o dom da vida com sabedoria em equilíbrio com a ciência.

Agradeço ao meu pai (in memoriam) e a minha mãe, pela minha família e por todos aqueles que entraram na minha história de vida me ensinando a crescer.

À minha mãe, por representar o motivo pelo qual concluí esta etapa com sucesso. Gratidão pela inspiração, presença e amor de sempre. Nenhum trabalho acadêmico ou profissional com êxito seria formado se não fosse a certeza de que a tenho em minha vida.

As minhas irmãs, pelo apoio, incentivo e por vibrarem comigo cada degrau da vida conquistado.

Ao meu esposo, pelo amor, carinho, cuidado e compreensão em todos os momentos.

Ao meu orientador, por ter me acolhido, por todos os ensinamentos e por contribuir na minha formação.

E por fim, aos grandes amigos do grupo caça ao linfonodo, que colaboraram para o desenvolvimento deste trabalho, através de apoio, amizade e lealdade.

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

A COVID-19 está relacionada a uma síndrome respiratória aguda grave no trato respiratório inferior causada pelo SARS-CoV-2. Análises observacionais mostraram que pessoas vacinadas com o Bacilo Calmette-Guérin (BCG) têm menor probabilidade de desenvolver formas graves da COVID-19. Assim, foi hipotetizado que a associação desse bacilo com epítomos imunogênicos selecionados por técnicas de vacinologia reversa das proteínas estruturais do SARS-CoV-2 pode acarretar o desenvolvimento de uma resposta imune inata treinada capaz de aumentar a resposta imune protetora específica contra esse vírus. Assim, esse trabalho teve como objetivo identificar epítomos imunoestimulantes da proteína *spike* do SARS-CoV-2 e associá-los ao BCG vacinal para a estimulação de uma resposta imune inata treinada. Para isso, 53.838 genomas depositados em bancos de dados do SARS-CoV-2 foram utilizados e as putativas sequências proteicas preditas da *spike* do vírus foram alinhadas para a determinação de uma proteína consenso mundial. A estrutura dessa proteína foi predita por modelagem molecular e foi possível identificar dois peptídeos de maior promiscuidade de se ligar aos BCR, MCH-I e MHC-II, assim como as características quanto a estabilidade, imunogenicidade, não alergenicidade, e com propriedades físico-químicas compatíveis para o desenvolvimento de peptídeos imunogênicos. Por ancoragem molecular foi possível identificar os pontos e as forças de interação com as moléculas de MHC. O epítomo 1 apresenta pontos de alterações já identificadas nas variantes do novo coronavírus, assim como um ponto de *n*-glicosilação. Por outro lado, o epítomo 2 não se encontra em posições de alta variabilidade nas variantes em circulação no momento, incluindo a nova variante ômicron. Duas formulações contendo o BCG e cada peptídeo foram utilizadas para estimular macrófagos derivados da medula óssea murina demonstrando um aumento expressivo na secreção de IL-6, IL-1 β e TNF- α quando do contato prévio com essa formulação. Ademais, a via de MAPK se mostrou mais ativada nessas células com destaque para maior ativação de p38. Análises *in vivo*, utilizando camundongos C57BL/6 vacinados intramuscularmente com as formulações, mostraram um aumento expressivo na produção de IgG total, IgG1 e IgG2c específicos para os peptídeos da proteína *spike* no soro, assim como a secreção aumentada de IFN- γ , IL-6, TNF- α e IL-17. Em suma, a abordagem racional de predição de epítomos imunogênicos ao serem associados ao BCG vacinal foi capaz de induzir a imunidade efetiva tanto inata quanto adaptativa em modelos *in vitro* e *in vivo*.

Palavras-chave: COVID-19; Vacinologia reversa; Epítomos; BCG; Imunidade inata treinada.

ABSTRACT

COVID-19 is related to a severe acute respiratory syndrome in the lower respiratory tract caused by SARS-CoV-2. Observational analyzes have shown that people vaccinated with Bacillus Calmette-Guérin (BCG) are less likely to develop severe forms of COVID-19. Thus, it was hypothesized that the association of this bacillus with immunogenic epitopes selected by reverse vaccinology techniques of the structural proteins of SARS-CoV-2 may lead to the development of a trained innate immune response capable of increasing the protective immune response against this virus. Thus, this work aimed to identify immunostimulatory epitopes of the SARS-CoV-2 spike protein and associate them with vaccine BCG for the specific stimulation of a trained innate immune response. For this, 53.838 genomes deposited in SARS-CoV-2 databases were used and the putative predicted protein sequences of the virus *spike* were aligned to determine a worldwide consensus protein. The structure of this protein was predicted by molecular modeling and it was possible to identify two peptides with greater promiscuity to bind to BCR, MCH-I and MHC-II, as well as the characteristics regarding stability, immunogenicity, non-allergenicity, and physical-chemical properties compatible chemicals for the development of immunogenic peptides. By molecular anchoring it was possible to identify the points and the forces of interaction with the MHC molecules. Epitope 1 presents points of alterations already identified in the novel coronavirus variants, as well as an n-glycosylation point. On the other hand, epitope 2 is not found in positions of high variability in the variants currently in circulation, including the new variant omicron. Two formulations containing BCG and each peptide were used to stimulate macrophages derived from murine bone marrow, demonstrating a significant increase in the secretion of IL-6, IL-1 β and TNF- α upon previous contact with this formulation. In addition, the MAPK pathway was more activated in these cells, highlighting the greater activation of p38. *In vivo* analyzes using C57BL/6 mice vaccinated intramuscularly with the formulations showed a significant increase in the production of total IgG, IgG1 and IgG2c specific for the spike protein peptides in serum, as well as increased secretion of IFN- γ , IL- 6, TNF- α and IL-17. In summary, the rational approach of predicting immunogenic epitopes when associated with vaccine BCG was able to induce effective immunity both innate and adaptive *in vitro* and *in vivo* models.

Keywords: COVID-19; Reverse vaccinology; Epitopes; BCG; Trained innate immunity.

LISTA DE FIGURAS

Figura 1 –	Nomenclatura do Coronavírus em relação a taxonomia de surtos anteriores.....	16
Figura 2 –	Linha do tempo da pandemia de COVID-19.....	17
Figura 3 –	Visão geral da estrutura do SARS-CoV-2 e sua organização genômica.....	18
Figura 4 –	Cronograma de desenvolvimento de putativos candidatos vacinais.....	22
Figura 5 –	Alinhamento das sequências obtidas por região que caracterizam a proteína <i>spike</i> do SARS-CoV-2 em cada continente a partir dos resultados obtidos da sequência consenso por continente.....	41
Figura 6 –	Representação esquemática de regiões da glicoproteína <i>spike</i>	44
Figura 7 –	Alinhamento dos epítomos selecionados da sequência consenso da proteína <i>spike</i> em comparação com as sequências das variantes do vírus SARS-CoV-2.....	45
Figura 8 –	Epítomos da glicoproteína <i>spike</i> localizados na base da proteína.....	46
Figura 9 –	Ilustração da comparação de estruturas entre proteínas <i>spike</i> obtidas por cristalografia e modelagem molecular por homologia.....	47
Figura 10 –	Representação do diagrama 3D e 2D da interação do Epítopo 1 com o receptor MHC de classe I, alelo HLA-A*0203.....	49
Figura 11 –	Representação do diagrama 3D e 2D da interação do Epítopo 1 com o receptor MHC de classe II, alelo HLA-DRBI*04:01.....	51
Figura 12 –	Representação do diagrama 3D e 2D da interação do Epítopo 2 com o receptor MHC de classe I, alelo HLA-B*3501.....	53

Figura 13 –	Representação do diagrama 3D e 2D da interação do Epítopo 2 com o receptor MHC de classe II, alelo HLA-BI*15:01.....	55
Figura 14 –	Representação do diagrama 3D e 2D da interação do Epítopo 2 com o receptor de histocompatibilidade H-2 classe I.....	57
Figura 15 –	A reestimulação em células BMDMs aumentou a expressão de citocinas pró-inflamatórias IL-6 (A-B), IL-1 β (C-D) e TNF- α (E-F) em contato com BCG e as formulações desenvolvidas.....	58
Figura 16 –	As vias de sinalização de MAPK foram ativadas pelas formulações desenvolvidas do vírus SARS-CoV-2 em células BMDMs.....	59
Figura 17 –	Análise de expressão de IL-6 (A), TNF- α (B), INF- γ (C), IL-17 (D) e IL-10 (E) em cultura de esplenócitos de grupos experimentais de camundongos C57/BL6 imunizados com BCG e os respectivos peptídeos da proteína <i>spike</i> e estimulados com meio, peptídeo, BCG e BCG mais peptídeo.....	60
Figura 18 –	Análise de expressão de IL-6 (A), TNF- α (B), INF- γ (C), IL-17 (D) e IL-10 (E) em soro de grupos experimentais de camundongos C57/BL6 imunizados com BCG e BCG com os respectivos peptídeos da proteína <i>spike</i>	63
Figura 19 –	Anticorpos específicos IgG, IgG1, IgG2c no soro de animais não vacinados e animais imunizados.....	64

LISTA DE TABELAS

Tabela 1 –	Avaliações físico-químicas dos epítomos finais mais promíscuos da proteína <i>spike</i>	43
Tabela 2 –	Interações dos alelos de MHC-I e MHC-II dos epítomos selecionados.....	48
Tabela 3 –	Interações dos alelos de H-2 classe I e H-2 classe II dos epítomos selecionados.....	56

LISTA DE ABREVIATURAS E SIGLAS

ACE 2	Enzima Conversora de Angiotensina 2
APCs	Células Apresentadoras de Antígenos
BCG	Bacille Calmette-Guérin
BMDM	Macrófagos Derivados da Medula Óssea
COVID-19	Doença de Coronavírus 2019
MERS-CoV	Síndrome Respiratória do Oriente Médio
OMS	Organização Mundial da Saúde
PAMP	Padrão Molecular Associado ao Patógeno
PRRs	Receptores de reconhecimento de padrões
SARS-CoV	Coronavírus 1 Relacionado à Síndrome Respiratória Aguda e Grave
SARS-CoV-2	Coronavírus 2 Relacionado à Síndrome Respiratória Aguda e Grave
SFB	Soro Fetal Bovino
TMPRSS2	Enzima Serino Protease Transmembrana do Tipo 2

SUMÁRIO

1	INTRODUÇÃO.....	14
2	REFERENCIAL TEÓRICO.....	16
2.1	SARS-COV-2 E A COVID-19.....	16
2.2	REPLICAÇÃO DO VÍRUS E IMUNOPATOGÊNESE DA COVID-19.....	18
2.3	VACINAS CONTRA COVID-19.....	21
2.4	VACINOLOGIA REVERSA.....	23
2.5	BCG E IMUNIDADE INATA TREINADA.....	24
3	JUSTIFICATIVA.....	27
4	OBJETIVOS.....	28
4.1	OBJETIVO GERAL.....	28
4.2	OBJETIVOS ESPECÍFICOS.....	28
5	MATERIAL E MÉTODOS.....	29
5.1	<i>IN SILICO</i>	29
5.1.1	Obtenção de sequência consenso da glicoproteína <i>spike</i> do SARS-CoV-2.....	29
5.1.2	Obtenção e predição de epítopos de célula B, MHC-I e MHC-II.....	29
5.1.3	Alinhamento e identificação de regiões transmembrana.....	30
5.1.4	Teste de antigenicidade, alergenicidade e propriedades físico-químicas.....	30
5.1.5	Identificação de regiões de N-glicosilação.....	31
5.1.6	Modelagem molecular por homologia e <i>molecular docking</i>	31
5.1.7	Validação da estrutura obtida por modelagem molecular por homologia.....	32
5.1.8	Síntese dos epítopos.....	33
5.2	<i>IN VITRO</i>	33
5.2.1	Animais experimentais.....	33
5.2.2	Cultura de macrófagos derivados da medula óssea.....	33
5.2.3	Estimulação de BMDMS com BCG e peptídeos da proteína <i>spike</i>	34
5.2.4	Extração de proteínas e <i>Western Blotting</i>	34
5.2.5	Análise por ELISA em células BMDMS.....	35
5.3	<i>IN VIVO</i>	36
5.3.1	Animais experimentais.....	36
5.3.2	Imunização de animais.....	36
5.3.3	Cultura de esplenócitos.....	36

5.3.4	Análise por CBA.....	37
5.3.5	Ensaio sorológico: quantificação de IGG total e subclasses.....	37
6	ANÁLISES ESTATÍSTICAS.....	39
7	RESULTADOS.....	40
7.1	IN SILICO.....	40
7.1.1	Uma sequência global com 1.276 aminoácidos foi identificada como a sequência consenso da glicoproteína <i>spike</i>	40
7.1.2	Foram identificados 833 epítomos da glicoproteína <i>spike</i> a partir da sequência global.....	42
7.1.3	Os 833 epítomos foram alinhados e 812 epítomos foram removidos por estar na região transmembrana.....	42
7.1.4	Seis epítomos foram considerados antigênicos, não-alérgico e estáveis.....	43
7.1.5	Dois epítomos apresentaram maior probabilidade de serem potencialmente selecionados para análises <i>in vitro</i> e <i>in vivo</i>	44
7.1.6	A modelagem molecular por homologia da proteína consenso mundial da <i>spike</i> mostrou que os epítomos se apresentam na base da proteína e distantes de regiões de maior variabilidade.....	45
7.1.7	Avaliação da ancoragem molecular dos epítomos selecionados apresentam forte ligação às moléculas de MHC de <i>Homo sapiens</i>	47
7.1.8	Ancoragem molecular dos epítomos selecionados apresentam ligação as moléculas de H-2 de <i>Mus musculus</i> similares a interações de MHC de <i>Homo sapiens</i>	56
7.2	IN VITRO.....	58
7.2.1	BMDMs estimuladas com BCG em combinação com os peptídeos da proteína <i>spike</i> apresentaram maior aumento na secreção de IL-6, IL-1 β , TNF- α quando reestimuladas.....	58
7.2.2	A via de sinalização MAPK foi ativada e intensificada em amostras reestimuladas com a formulação contendo os peptídeos associados ao BCG vacinal.....	59
7.3	IN VIVO.....	60
7.3.1	Animais vacinados com os peptídeos associados ao BCG vacinal apresentaram aumento da secreção de IL-6, TNF- α , INF- γ em cultura de	

	esplenócitos e em circulação.....	60
7.3.2	Aumento de anticorpos específicos às proteínas estruturais do SARS-COV-2 foram detectados no soro de animais imunizados.....	63
8	DISCUSSÃO.....	65
9	CONCLUSÃO.....	70
10	REFERÊNCIAS.....	71
	ANEXOS.....	78

1 INTRODUÇÃO

A COVID-19 consiste em uma doença respiratória aguda que é causada pelo coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2). Foi identificada pela primeira vez em Wuhan na província de Hubei na república popular da China. Os primeiros casos em humanos do vírus SARS-CoV-2, causador da COVID-19, ocorreram em meados de dezembro de 2019. A maioria dos casos tinham vínculos epidemiológicos muito fortes da presença de pessoas infectadas no mercado de frutos do mar na cidade de Huanan (HOLSHUE *et al.*, 2020). Zhou e colaboradores 2020, foram os primeiros pesquisadores a publicarem o isolamento do sequenciamento do material do genoma desse agente etiológico até então desconhecido e fizeram a caracterização molecular, análise filogenética e a classificação taxonômica desse novo vírus e identificaram que se tratava de um *Betacoronavírus* e não tinha ainda sido identificado como infectante em humanos. A partir desse momento surgiram diversas hipóteses acerca da origem do SARS-CoV-2.

Visto o cenário de estágio crítico do alastramento desse vírus pelo mundo, uma grande e forte atividade se concentrou no desenvolvimento de pesquisas a fim de se ter uma vacina efetiva contra esse novo coronavírus. As vacinas foram desenvolvidas em tempo recorde e sua implantação teve redução no número de casos, mortes e transmissão de SARS-CoV-2 (DAGAN *et al.*, 2021; PRITCHARD *et al.*, 2021). No entanto, esse vírus sofreu mutações e com o surgimento de novas variantes os casos se ascenderam por serem mais transmissíveis (DAVIES *et al.*, 2021; TEGALLY *et al.*, 2021) e parcialmente escaparem da defesa imune (WANG *et al.*, 2021; WIBMER *et al.*, 2021).

As abordagens de bioinformática se evidenciaram no desenvolvimento de vacinas para SARS-CoV-2, pois as previsões *in silico* auxiliaram no direcionamento de abordagens voltadas em tecnologia vacinal promissoras a fim de garantir mecanismos de resposta imune melhorada e segura (OLI *et al.*, 2020). Vários estudos desenvolveram vacinas baseadas em epítomos do SARS-CoV-2. As ferramentas de imunoinformática usadas por Abraham *et al.* 2020, possibilitaram no desenvolvimento de uma vacina de multiepítopo que desencadearam respostas imunes potencialmente protetoras de células T CD4⁺ e CD8⁺. Além disso, essa abordagem de desenvolvimento de vacinas está concentrada na proteína *spike* do SARS-CoV-2.

Com isso, análises observacionais também se concentraram no Bacilo Calmette-Guerin (BCG). A vacina contra tuberculose tem um impacto não específico na imunidade adaptativa e proteção contra outros patógenos e sobre infecções respiratórias (NETEA *et al.*,

2020b). Essa proteção considerada como inespecífica é descrita pela capacidade do BCG induzir uma imunidade treinada, isto é, uma reprogramação epigenética de resposta imune inata e especificidade para proteger contra diversas doenças. Assim, vários ensaios se concentraram na vacinação e/ou revacinação com BCG e se essa poderia diminuir a gravidade e casos globais da COVID-19 (NETEA, *et al.*, 2020b; TSILIKA *et al.*, 2021)

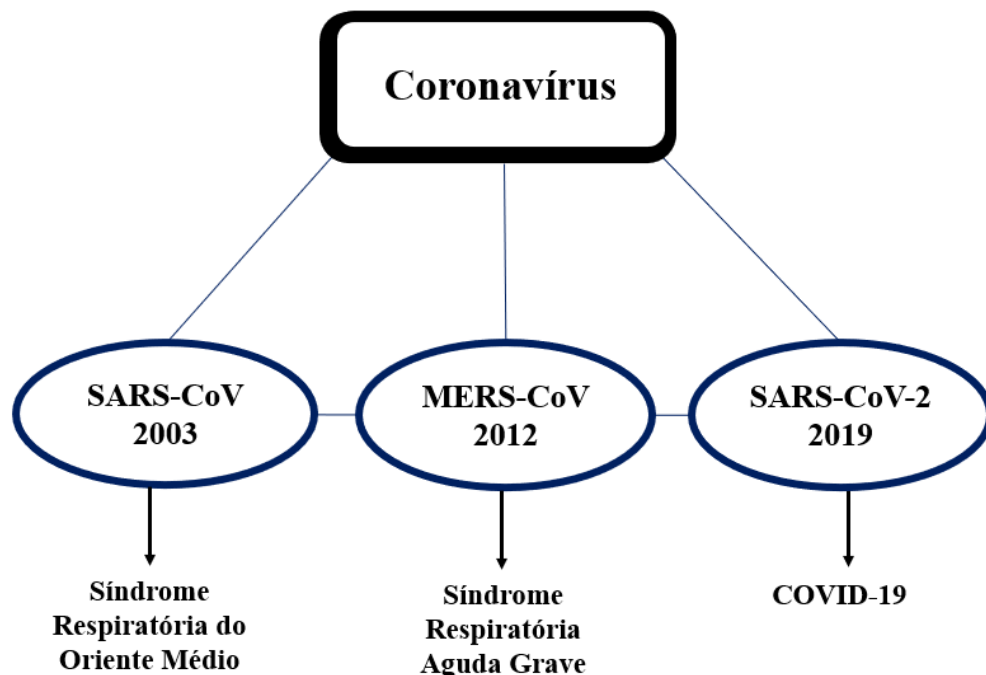
Assim, nesse estudo, o objetivo foi identificar epítomos imunoestimulantes da proteína *spike* do SARS-CoV-2 e associá-los ao BCG vacinal para a estimulação de uma resposta imune inata treinada. Metodologias *in silico* foram usadas para selecionar epítomos imunogênicos como alvos putativos de um candidato vacinal para avaliações *in vitro* e *in vivo* demonstrando resultados promissores em associação com BCG vacinal.

2 REFERENCIAL TEÓRICO

2.1 SARS-COV-2 E A COVID-19

O vírus SARS-CoV-2 faz parte de uma grande família de coronavírus (figura 1). Esse novo coronavírus pertence ao gênero *Betacoronavirus* semelhante ao vírus SARS-CoV (síndrome respiratória aguda e grave), que foi o causador da epidemia acometida em alguns países asiáticos e o Canadá em 2003. Em relação filogenética, o SARS-CoV-2 é distante a MERS-CoV (síndrome respiratória do Oriente Médio), causador de uma outra síndrome respiratória grave que iniciou em 2012, a qual acometeu o Oriente Médio e alguns países da Ásia. O SARS-CoV-2 foi então categorizado dentro da família *Coronaviridae* e subgênero *Sarbecovirus* (CUI; LI; SHI, 2019; CHEN; LIU; GUO, 2020; MALIK, 2020; HU *et al.*, 2021).

Figura 1 – Nomenclatura dos Coronavírus em relação a taxonomia de surtos anteriores.

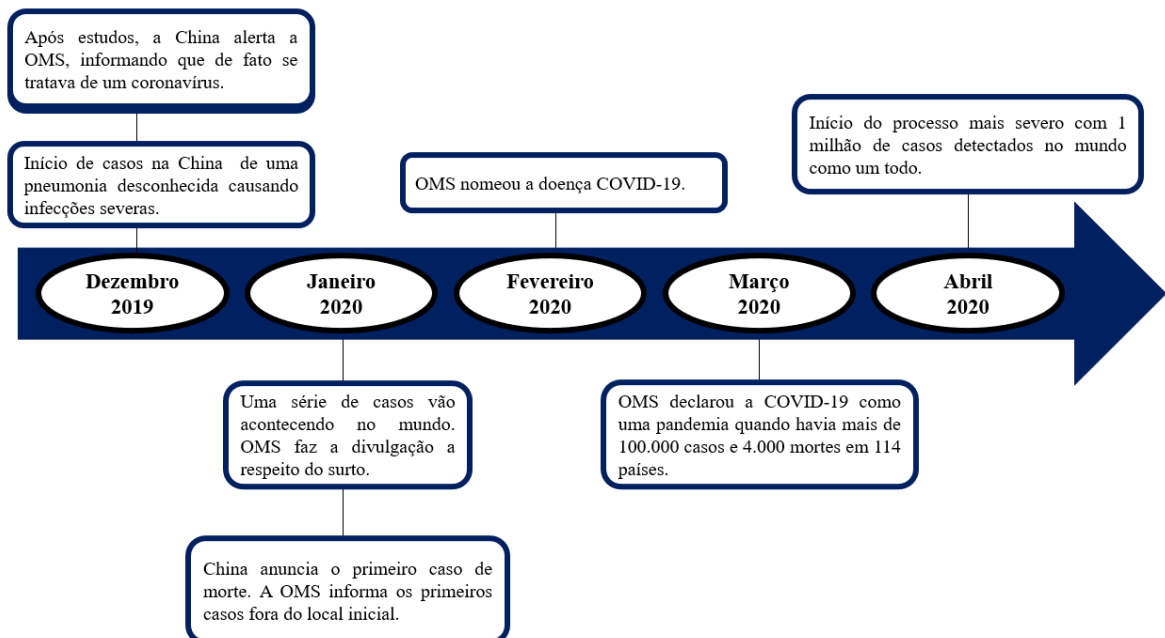


Fonte: Adaptado (GORBALENYA *et al.*, 2020).

Assim como os demais vírus da família *Coronaviridae*, o SARS-CoV-2 também tem o potencial de causar doenças relacionadas ao trato respiratório. A doença de coronavírus 2019 (COVID-19) é uma infecção respiratória, causada pelo novo coronavírus. Foi identificada inicialmente em Dezembro de 2019 como sendo uma pneumonia de origem

desconhecida a qual, acometia pessoas que frequentavam o mesmo mercado de peixes na cidade de Wuhan (China) (MALLAH *et al.*, 2021). Inicialmente, por meio de ensaios e análises de metagenoma, essa doença foi associada a um novo coronavírus, semelhante ao SARS-CoV, denominado inicialmente nCoV-19 e posteriormente como SARS-CoV-2. Em pouco tempo a COVID-19 deixou de ter incidência local e se espalhou pelo mundo (ZHOU *et al.*, 2020). Em fevereiro de 2020 houve o primeiro caso identificado no Brasil e em março do mesmo ano a Organização Mundial da Saúde (OMS) declarou a COVID-19 como sendo uma pandemia (RODRIGUEZ-MORALES *et al.*, 2020; WORLD HEALTH ORGANIZATION, 2020) (figura 2).

Figura 2 – Linha do tempo da pandemia de COVID-19.



Fonte: Adaptado (HU *et al.*, 2021).

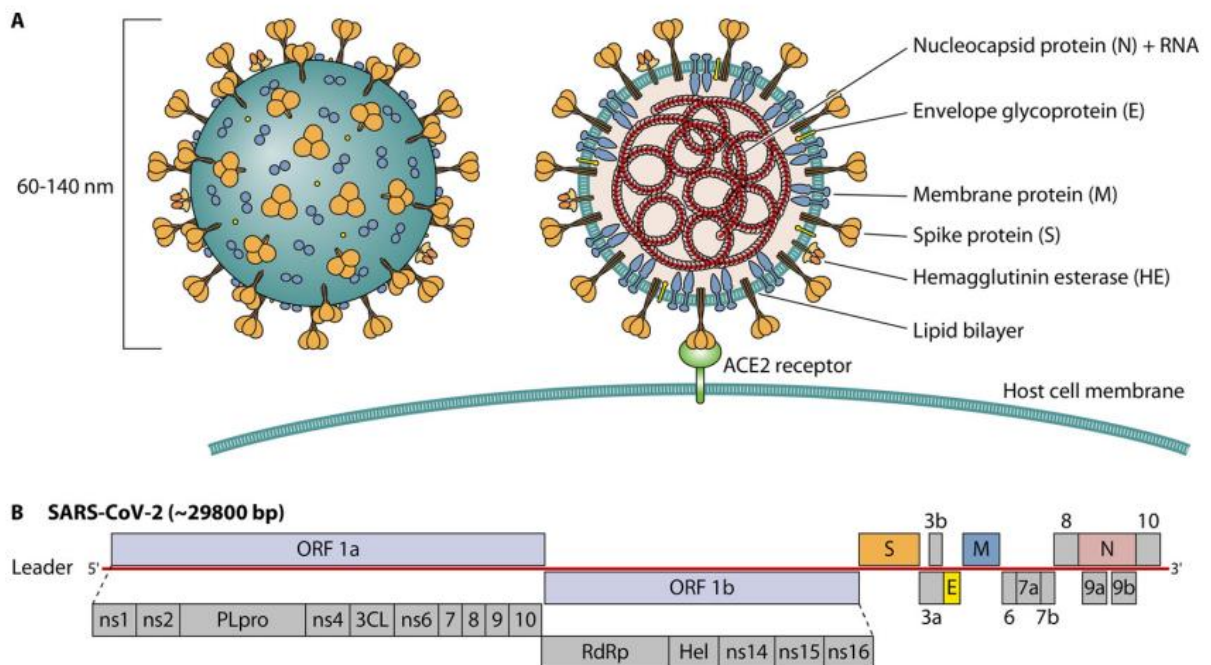
O SARS-CoV-2 é uma abreviação que significa coronavírus 2 relacionado a síndrome respiratória aguda e grave. Como todo coronavírus, o SARS-CoV-2 é um vírus encapsulado, com material genômico composto de RNA de sentido positivo e fita única, genoma em torno de 30 quilobases, vírions que variam em torno de 60 a 140 nanômetros de diâmetro e morfologia de coroa solar (SAFIABADI TALI *et al.*, 2021; SAHU *et al.*, 2021).

A morfologia é dada pela protusão da proteína *spike* e seu envelope. Estruturalmente, os vírions do SARS-CoV-2 são compostos por quatro proteínas estruturais: *spike* (responsável pelo reconhecimento do receptor na superfície da célula a ser infectada), nucleocapsídeo (envolve o material genético e é responsável pela sua condensação), membrana e envelope

que são proteínas transmembranas e estão envolvidas na estabilização de sua estrutura (figura 3) (PENG *et al.*, 2021).

O genoma do SARS-CoV-2 na extremidade 5', possui duas janelas abertas de leitura: a ORF 1a e ORF 1b que codificam as proteínas não estruturais que formam o complexo de transcrição e replicação viral. Na extremidade 3', codificam as proteínas estruturais *spike* (S), envelope (E), membrana (M) e nucleocapsídeo (N) e outras proteínas não estruturais responsáveis pela interação celular (LI *et al.*, 2020; SAFIABADI TALI *et al.*, 2021; SAHU *et al.*, 2021).

Figura 3 – Visão geral da estrutura do SARS-CoV-2 e sua organização genômica.



Fonte: (SAFIABADI TALI *et al.*, 2021).

Legenda: **A:** Estrutura do SARS-CoV-2. **B:** Organização do genoma e suas proteínas estruturais e não-estruturais. O SARS-CoV-2 é um vírus de RNA e quando é colocado dentro da célula do hospedeiro e sua maquinaria é para traduzir o RNA. O RNA é traduzido em 16 proteínas não-estruturais, 9 proteínas acessórias e são proteínas que estão envolvidas no processo de evasão da resposta imune do hospedeiro e também no processo de replicação do vírus dentro da célula.

2.2 REPLICAÇÃO DO VÍRUS E IMUNOPATOGÊNESE DA COVID-19

A proteína *spike* possui duas subunidades S1 e S2, sendo a subunidade S1 que se liga ao receptor ACE 2. Após a associação da proteína *spike* com o receptor ACE 2, forma-se um complexo que é clivado por uma enzima serino protease transmembrana do tipo 2 (TMPRSS2) o que faz a proteína *spike* ser ativada e esse processo permite que ocorra a fusão

das membranas do vírus com a membrana celular. Dessa forma, ocorre a entrada do vírus dentro da célula (HARTENIAN *et al.*, 2020; HOFFMANN, M. *et al.*, 2020).

A replicação do vírus inicia-se pela ligação do vírion à membrana da célula através do reconhecimento da *spike* à enzima conversora da angiotensina 2 (ACE2), presente na membrana celular. Nesse contexto duas coisas podem acontecer: o vírus sofrer endocitose e o material genético atinge o interior da célula ou, então, na própria membrana da célula ocorre a fusão com a membrana viral também acarretando a entrada no RNA viral no interior da célula hospedeira. Inicialmente, são traduzidas as proteínas não estruturais que formam um complexo de transcrição e replicação, onde há a transcrição de um novo material genético completo e também de mRNAs subgenômicos que codificam as proteínas estruturais. Essas são traduzidas no retículo endoplasmático e novas vesículas com essas proteínas, associadas aos novos genomas de RNA viral, os novos vírions são formados, excitados e o ciclo replicativo se repete (V'KOVSKI *et al.*, 2021).

O novo coronavírus adentra na mucosa do trato respiratório e entra em contato com células da imunidade inata residentes como macrófagos e células dendríticas, as quais são importantes células apresentadoras de antígenos (APCs) para linfócitos T CD4⁺. O SARS-CoV-2 propriamente ou proteínas virais entrando em contato com moléculas de superfície de macrófagos e células dendríticas, irão ativar essas células levando a indução de vias inflamatórias (DESCHLER *et al.*, 2021; ZHUANG *et al.*, 2021).

De modo geral, a resposta imune contra o SARS-CoV-2, acontece de maneira similar em que acontece a uma resposta imune antiviral e recruta vários componentes da imunidade inata. A partir do momento que o vírus entra em contato com o organismo, ocorrerá a ativação de fatores humorais como citocinas ou irá induzir a ativação de fatores celulares como neutrófilos, macrófagos e células dendríticas (KIM *et al.*, 2021). A resposta imune inata é essencial para indução da resposta imune adaptativa e essa resposta irá construir uma memória imunológica sendo esta específica, ou seja, dará origem aos linfócitos T auxiliares e citotóxicos, linfócitos B e também anticorpos que serão específicos contra o SARS-CoV-2 (FAJGENBAUM; JUNE, 2020).

Quando o vírus entra em contato com a célula, embora ele seja capaz de fazer a replicação viral, existe também uma parte do vírus que é detectada pelo sistema imune e induz a ativação de vários genes como por exemplo, NF-kB ou os genes relacionados com síntese de interferons e, faz com que essas células mandem sinais para o sistema imune. Com isso, irá recrutar células principalmente da imunidade inata a princípio para que essas células sejam

em alguns casos fagocitadas e em outros simplesmente lisadas (LI *et al.*, 2020; PROMPETCHARA; KETLOY; PALAGA, 2020; KING; SPRENT, 2021).

A produção de citocinas a partir da ativação dos genes recrutam células como as células dendríticas que vão ser capazes de capturar os vírus que estão naquele microambiente e apresentar esses antígenos via MHC para linfócitos T CD4⁺ e CD8⁺. Quando já se tem alguns linfócitos específicos do vírus como é o caso dos linfócitos T CD8⁺, ele via moléculas de MHC irá encontrar células que estão infectadas pelo vírus e vai conseguir lisar essas células, eliminado assim, se for possível, o patógeno que está causando aquela patologia (FAJGENBAUM; JUNE, 2020; GUSTINE; JONES, 2021; ZHUANG *et al.*, 2021).

Em geral, o curso da infecção pelo SARS-CoV-2 acontece em semanas. A princípio, quando há o aumento da carga viral, se inicia a ativação da resposta imune inata e quando ela está de fato bem construída começa a indução da resposta imune adaptativa, que acontece depois de mais ou menos duas semanas. Existe também, período de tempo de latência, em que há o crescimento exponencial da carga viral, é nesse intervalo que surge os primeiros sintomas. À medida que se tem o desenvolvimento de uma resposta imune adaptativa, ou seja, o surgimento de linfócitos específicos e a produção de anticorpos, há também o decaimento da carga viral (CHENG *et al.*, 2020; LI *et al.*, 2020; PROMPETCHARA; KETLOY; PALAGA, 2020; KING; SPRENT, 2021).

Em relação aos estágios clínicos da COVID-19, é possível identificar três estados principais: leve, moderado e grave. O que diferencia os estágios clínicos basicamente é o nível do dano tecidual causado pela infecção do próprio vírus SARS-CoV-2 ou pelo nível da inflamação montada pelo sistema imune em resposta a própria infecção. Casos mais leves apresentam danos teciduais e níveis inflamatórios menores em que a pessoa pode ser assintomática ou com sintomas leves que não se diferenciam de outras infecções respiratórias como febre, tosse e diarreia (BUSZKO *et al.*, 2020; RAMAN *et al.*, 2021).

A infecção pelo SARS-CoV-2 pode ser classificada em estágios. Estágio 1, que se refere a infecção. Estágio 2, que é denominado fase pulmonar. Estágio 3, que é denominado hiperinflamação. A maioria das pessoas infectadas pela COVID-19 interrompem o ciclo da doença provavelmente no estágio 2. Dessa forma, já é sabido que alguns pacientes apresentam sinais e sintomas clínicos muito leves, como por exemplo, mal estar, diarreia, tosse seca e febre. No entanto, à medida que evoluir o estágio de severidade da doença, pode ocorrer falta de ar, hipóxia e no estágio 3, por exemplo, pode caracterizar a síndrome do desconforto respiratório e a síndrome respiratória inflamatória sistêmica a qual, os principais sinais clínicos são marcadores pró-inflamatórios elevados (SIDDIQI; MEHRA, 2020).

A barreira entre os estágios clínicos leves das formas clínicas mais graves é justamente a perda da capacidade de troca gasosa pulmonar que caracteriza o sintoma de dificuldade respiratória e podendo levar até mesmo à intubação de pacientes. Há um aumento da inflamação sistêmica, podendo afetar outros órgãos como coração, fígado e o rim, levando o paciente a óbito (MEHMOOD *et al.*, 2021).

Diversos estudos apontam que o sistema imunológico apresenta um papel primordial no direcionamento dos destinos de perfis clínicos. Dessa forma, essas análises mostram que a COVID-19 pode resultar em um perfil de linfopenia que pode ser até mais acentuada em pacientes graves e moderado, podendo se referir para as células T CD4⁺ pois, indivíduos mais graves podem apresentar uma deficiência na montagem expulsiva de perfil Th1. Assim, pacientes graves da doença podem montar de fato uma montagem imunológica, mas esta montagem pode não ser adequada em um processo de infecção viral (LI *et al.*, 2020; PROMPETCHARA; KETLOY; PALAGA, 2020; KIM *et al.*; 2021).

Casos graves da COVID-19 causam hiperinflamações e a tempestade de citocina. As citocinas são moléculas sinalizadoras que recrutam células do sistema imunológico a qual, medeiam uma resposta imunológica balanceada. A resposta pró-inflamatória que é uma resposta de defesa do nosso organismo é balanceada com uma resposta anti-inflamatória e esse balanço é que leva ter uma melhora (COSTELA-RUIZ *et al.*, 2020; CHENG *et al.*, 2020; WAN *et al.*, 2020; WU; TANG, 2020; KHOSROSHAHI *et al.*, 2021).

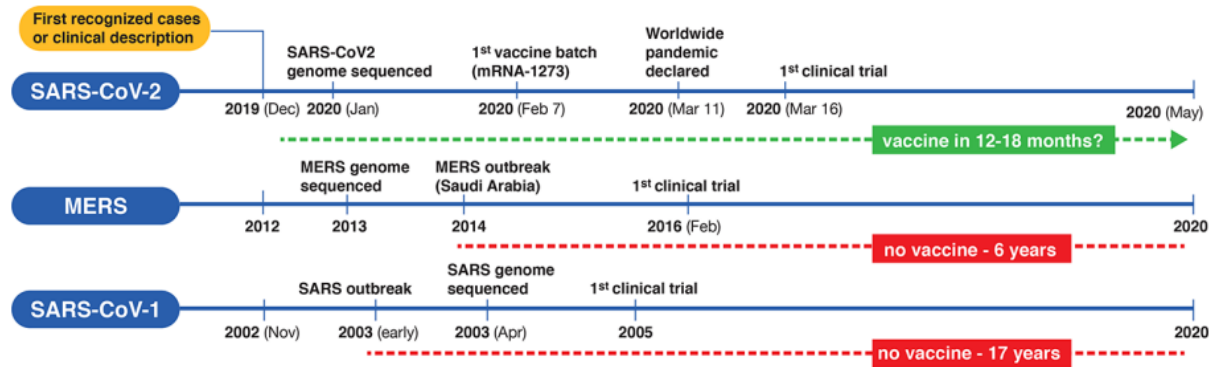
O indivíduo infectado por SARS-CoV-2 pode levar a dois caminhos: ao balanço da resposta pró e anti-inflamatória que leva a uma melhora do indivíduo sem muita gravidade. O outro caminho é onde há o desbalanço, ou seja, uma resposta maior pró-inflamatória do que anti-inflamatória, o que leva a uma tempestade de citocinas e, conseqüentemente ao aumento da gravidade da doença (WU; TANG, 2020; STRICH *et al.*, 2021).

2.3 VACINAS CONTRA COVID-19

A vacinação é o método mais eficaz e seguro de se criar uma barreira imunológica, capaz de romper o ciclo de transmissão do vírus e prevenir da patologia mais grave da doença. Nesse contexto, iniciou-se uma forte atividade de pesquisa e desenvolvimento pelo mundo de uma vacina em busca de uma resposta protetora contra o vírus SARS-CoV-2 (TREGONING *et al.*, 2020). No caso dos coronavírus, a produção de vacinas para o vírus SARS-CoV chegou apenas no seu primeiro ensaio clínico e não foram necessários mais avanços, pois a epidemia foi controlada por medidas não farmacológicas e, apesar de haver eventuais casos no Oriente,

do vírus MERS também não há uma perspectiva de vacina (figura 4). Baseando nas informações já obtidas com os ensaios com outros grupos virais, foi possível avançar rapidamente para a vacina contra a COVID-19 (FUNK; LAFERRIÈRE; ARDAKANI, 2020).

Figura 4 – Cronograma de desenvolvimento de putativos candidatos vacinais.



Fonte: (FUNK; LAFERRIÈRE; ARDAKANI, 2020).

Legenda: Linhas vermelhas pontilhadas representam que não há disponibilidade de vacinas em relação a surtos anteriores (SARS-CoV, MERS). Linha verde pontilhada representa o desenvolvimento emergencial de uma vacina no surto do vírus SARS-CoV-2.

Com o avanço do número de casos e óbitos em decorrência da COVID-19, as vacinas tiveram que ser desenvolvidas no ritmo da pandemia. Assim, em menos de um ano já estava licenciada a primeira vacina (LAMB, 2021). O relatório da Organização Mundial da Saúde mostra que depois de quase dois anos de pandemia, há 334 vacinas em desenvolvimento, sendo que 140 estão em estudos clínicos em diferentes fases e algumas já licenciadas e aprovadas para uso, utilizando diferentes ferramentas de desenvolvimento como de tecnologia de RNA mensageiro (Pfizer, Moderna), vetores virais (AstraZeneca, Gamaleya, Janssen, CanSino), proteína recombinante (Novavax) e vírus inativado (Sinopharm) (HAN; XU; YE, 2021; WHO, 2022).

Todas as plataformas vacinais tem em comum a produção de uma proteção contra o novo coronavírus, porém, a instabilidade genômica do SARS-CoV-2 com altas taxas de mutações, na maior parte das vezes concentradas na região codificadora da sequência da proteína *spike*, podem demandar atualizações constantes das vacinas que tem como alvo proteínas com grandes variações (PANDEY *et al.*, 2020). Com altas taxas de transmissão do vírus, novas variantes têm surgido e têm renovado a preocupação das perspectivas da pandemia (MENGIST *et al.*, 2021). Atualmente existem cinco variantes de ação do vírus SARS-CoV-2: Alpha (B.1.1.7), Beta (B.1.351), Gama (P.1), Delta (B.1.617.2) e Ômicron (B.1.1.529) (ALEEM; AKBAR SAMAD; SLENKER, 2022).

As vacinas foram produzidas para a cepa original do vírus SARS-CoV-2, porém, surgiu a preocupação de sua efetividade frente ao surgimento das novas variantes. Entretanto, no Reino Unido aonde circulava a variante B.1.1.7 teve efetividade já na primeira dose utilizando dois diferentes imunizantes (Pfizer e AstraZeneca), correlato a isso sua eficácia frente a segunda dose (PRITCHARD *et al.*, 2021).

Estudos mostraram a performance da CoronaVac em que a efetividade estimada depois da segunda dose foi significativa no ambiente aonde circulava principalmente a variante Gama (P.1) (FARIA *et al.*, 2021). Avaliando as vacinas Pfizer e AstraZeneca frente a variante B.1.617.2, há um declínio na eficácia, porém com considerável proteção (LOPEZ BERNAL *et al.*, 2021). Mesmo com a diminuição da eficácia das vacinas, a efetividade das mesmas se mostrou importante na queda no número de casos e hospitalizações, exemplificando Israel que foi um dos primeiros países a reportar a redução de risco de hospitalização (HAAS *et al.*, 2021).

2.4 VACINOLOGIA REVERSA

A vacinologia reserva foi descrita por Rino Rappuoli que se dá na obtenção de dados genômicos a fim de identificar possíveis moléculas que tenham uma aplicação como alvo vacinal. Com o advento da era dos genomas e esses depositados em banco de dados é possível predizer de forma *in silico* candidatos vacinais que sejam potencialmente antigênicos, imunogênicos e que não apresentam toxicidade. A vacinologia reversa vem na atualidade para poder encurtar cada vez mais o processo de desenvolvimento de uma vacina convencional até chegar em um modelo pré-clínico (DALSASS *et al.*, 2019; MOXON; RECHE; RAPPUOLI, 2019).

Rino Rappuoli aplicou com sucesso essa metodologia o que resultou em uma vacina comercial contra meningite do tipo B. Nesse caso, haviam 2158 genes que codificavam o genoma e aplicando diferentes estratégias de bioinformática selecionaram proteínas que eram mais promissoras e com isso, três antígenos foram selecionados para a produção de uma vacina (SERRUTO *et al.*, 2012).

A proteína *spike* é um dos principais alvos para o desenvolvimento de vacinas contra a COVID-19. Essa proteína e seus fragmentos podem ser utilizados como antígeno usando trechos específicos denominados epítomos. Os epítomos são pequenos trechos dos antígenos que podem ser usados para desencadear uma resposta imunológica, ou seja, são regiões que

são reconhecidos pelos linfócitos (células de defesa) e também pelos anticorpos (WALLS, 2020).

A base da vacina de epítomos é dar foco na resposta imune para determinadas regiões que são mais importantes para a resposta protetora. A bioinformática tem atuado como um importante funil que aponta epítomos mais promissores para o desenvolvimento de vacinas e isso se faz com um vasto volume de dados disponíveis sobre genes, proteínas, sequências e estruturas (RAPPUOLI *et al.*, 2016). Assim, os algoritmos desenvolvidos permitem identificar características diversas desses epítomos como por exemplo propriedades físico-químicas dos aminoácidos como a carga, o volume e a polaridade. Esses dados são usados como entrada em modelos de aprendizado de máquina e o resultado desses modelos são epítomos candidatos promissores para o desenvolvimento de vacinas. É possível prever epítomos que sejam bons candidatos que tenham um alto potencial antigênico (NOORIMOTLAGH *et al.*, 2020). Ademais, utilizando técnicas de modelagem molecular para construir o modelo teórico dessa estrutura e métodos de ancoragem molecular é possível, inclusive, identificar os melhores alvos de interação com proteínas do ser humano para o desencadeamento efetivo da resposta imune (DONG *et al.*, 2020).

2.5 BCG E IMUNIDADE INATA TREINADA

O BCG (Bacille Calmette-Guérin) é uma vacina que foi desenvolvida contra a tuberculose. Além da proteção contra a tuberculose foi observada a proteção contra outras infecções, especialmente contra infecções no trato respiratório, incluindo infecções virais. Após a introdução da vacina BCG, observou-se que a mortalidade em crianças estava diminuindo após a vacinação e isso não se deu apenas à tuberculose, pois houve a redução do número de óbito também devido à proteção contra outras vias respiratórias, infecções e sepse neonatais (ASPATWAR *et al.*, 2021).

Os mecanismos pelos quais o BCG protege contra outros tipos de infecções é dotado por uma reprogramação das células imunes inatas. Há uma modificação epigenética das células tronco na medula óssea para as células imunes inatas e estas transmitem diferenças de longo prazo em suas funções para que possam reagir melhor as infecções (FOSTER *et al.*, 2021).

A resposta imune é classicamente dividida entre a resposta imune inata e a resposta imune adaptativa. A forma ideal para se ter uma resposta efetiva antiviral é necessário que o sistema imune inato seja capaz de conseguir identificar, por meio de receptores de

reconhecimento de padrões (PRRs), os padrões moleculares associados aos patógenos (PAMPs) sinalizando para a produção de citocinas da imunidade inata capazes de estimular uma resposta imune adaptativa de uma forma mais direta e mais específica mediada pelos linfócitos T CD4⁺ e CD8⁺ (MOLAEI *et al.*, 2021).

Há alguns anos, a resposta imune inata era considerada inespecífica e incapaz de ter uma memória, enquanto a adaptativa já era dotada dessas características pelo rearranjo somático dos seus receptores com capacidade de gerar memória imunológica ao primeiro contato como agente estranho. A quebra desse dogma veio com a constatação de que a resposta imune inata apresenta uma especificidade pelos PRRs e pela existência de uma certa memória imune em organismos que não possuem os mecanismos de defesa adaptativos (NETEA; QUINTIN; VAN DER MEER, 2011; COVIÁN *et al.*, 2019).

Estudos observaram a existência de efeitos não específicos, quando comparados à especificidade da resposta imune adaptativa, associados tanto a vacina BCG bem como a infecções com cepas menos virulentas ou a β -glicanos extraídas da parede celular de *Candida albicans* e esses efeitos promoviam uma resposta protetora em caso de infecções secundárias (VAN DER MEER *et al.*, 2015). Essa memória da imunidade inata é conceituada imunidade inata treinada como uma resposta imune aumentada de células da imunidade inata a estímulos secundários, homólogos ou heterólogos que fornecem uma imunidade cruzada em mamíferos (NETEA; QUINTIN; VAN DER MEER, 2011; DIVANGAHI *et al.*, 2021). A imunidade inata treinada apresenta um caráter independente da resposta imune adaptativa e está associado a plasticidade das células mieloides como os macrófagos, promovendo um aumento de suas funções efetoras (NETEA *et al.*, 2020a).

O BCG gera alterações epigenéticas em algumas células como os monócitos que vão dar origem aos macrófagos. A alteração da estrutura de cromatina, deixando expostas regiões que seriam extremamente interessantes para combater microrganismos que não são aqueles se teve um contato anterior (O'NEILL; NETEA, 2020). Assim, treinar a resposta imune inata com BCG poderia resultar em redução da carga microbiana e, conseqüentemente, numa doença mais branda (KOEKEN *et al.*, 2019; NETEA *et al.*, 2020b).

O BCG induz a produção de citocinas importantes associadas a resposta imune treinada, visto em relação a vacina da febre amarela mostrando que a viremia em pacientes que tiveram vacinação por BCG foi menor para o vírus da febre amarela (ARTS *et al.*, 2018). Por conta disso, existe perspectiva do BCG em outras doenças, principalmente patógenos intracelulares (ARTS *et al.*, 2018). A vacinação por BCG induz a expressão de genes que são importantes para a defesa do hospedeiro, e em idosos induz proteção para infecções

respiratórias virais (GIAMARELLOS-BOURBOULIS *et al.*, 2020). Nesse contexto, o uso de BCG vacinal poderia proteger contra vários tipos de infecções, incluindo ao SARS-CoV-2. Estudos sobre o efeito de BCG durante a pandemia da COVID-19 em idosos realizado na Grécia, demonstraram que a revacinação com esse bacilo, diminuíram o número de casos da doença. Foi então indicado que potencialmente as vacinas atenuadas contribuem na diminuição e o impacto do novo patógeno (NETEA *et al.*, 2020b; TSILIKA *et al.*, 2021).

3 JUSTIFICATIVA

No início deste trabalho, havia alguns estudos clínicos em andamento tanto para o desenvolvimento de vacinas quanto para o desenvolvimento de medicamentos. A vacina é uma estratégia eficaz para o controle e prevenção da disseminação do SARS-CoV-2. Atualmente, existem diversas abordagens de vacinas para o controle da COVID-19. Vale atentar para as proteínas estruturais do vírus, pois, essas proteínas estão intimamente associadas na interação com o hospedeiro e podem ser fontes de componentes imunogênicos específicos contra o vírus.

Até o momento há quatro vacinas aprovadas pela ANVISA (Coronavac, Pfizer, AstraZeneca, Janssen). Uma vez que se sabe que é preciso ter uma resposta imune ideal contra o novo coronavírus e que envolva nesse processo uma imunidade inata eficiente e uma imunidade adaptativa celular e humoral robusta e específica, é necessário identificar de epítomos imunogênicos por meio de banco de dados de DNA. A identificação desses epítomos, possibilita encontrar a sequência de todo o genoma do SARS-CoV-2 onde é possível fazer a predição e a interação desses epítomos com receptores de células B e com alelos de MHC-I e MHC-II. Assim, é possível fazer ensaios imunológicos e até mesmo o designer de uma vacina contra o SARS-CoV-2.

Dados observacionais da vacinação prévia com BCG induzir uma resposta imune inata treinada contra uma série de infecções heterólogas indicam que essa plataforma vacinal pode ser atrelada a outros componentes microbianos a fim de melhorar a especificidade de uma vacina de subunidade. Assim, esse estudo é baseado na associação do BCG vacinal com epítomos imunogênicos provenientes da proteína estrutural *spike* do SARS-CoV-2 selecionados por imunoinformática. Logo, é importante investigar a possibilidade de atrelar esse peptídeo a uma vacina comercial como indutora de uma resposta imune inata treinada e que a mesma colabore em uma resposta ativa contra esse agente infeccioso.

4 OBJETIVOS

4.1 OBJETIVO GERAL

Identificar epítomos imunoestimulantes da proteína *spike* do SARS-CoV-2 e associá-los ao BCG vacinal para a estimulação de uma resposta imune inata treinada.

4.2 OBJETIVOS ESPECÍFICOS

- a) Determinar a sequência proteica da proteína *spike* do novo coronavírus depositadas em bancos de dados genômicos e definir uma sequência consenso mundial;
- b) Definir um modelo estrutural da proteína consenso e identificar epítomos imunodominantes na sequência consenso mundial da proteína *spike* do novo coronavírus;
- c) Avaliar o potencial de ancoragem dos epítomos a receptores do complexo de histocompatibilidade humano;
- d) Estimular BMDMs com epítomos selecionados e sinteticamente produzidos da proteína *spike* do novo coronavírus associados ao BCG vacinal;
- e) Avaliar o potencial imunoestimulador das formulações vacinais *in vivo*.

5 MATERIAL E MÉTODOS

A metodologia do desenvolvimento desse trabalho foi dividida em três partes: *in silico*, *in vitro* e *in vivo*.

5.1 IN SILICO

5.1.1 Obtenção de sequência consenso da glicoproteína *spike* do SARS-CoV-2

A primeira etapa deste trabalho consistiu na obtenção de todas as sequências da glicoproteína *spike* do SARS-CoV-2. Para isso, utilizou-se o servidor e banco de dados NCBI (National Center of Biotechnology Information), em que a aba “NCBI Vírus”, aliado aos descritores “SARS-CoV-2” e “taxid:2697049” foram definidos, no qual coletou-se todas as sequências das respectivas proteínas presentes, que foram depositadas até o dia 06/08/2020 (<https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/>). Em seguida, as sequências foram expostas verticalmente e com o auxílio do software Jalview (<https://www.jalview.org/>) foram alinhadas primeiramente por continente, e depois de maneira global, de forma a se obter uma sequência consenso mundial.

5.1.2 Obtenção e predição de epítomos de célula B, MHC-I e MHC-II

O próximo passo foi obter-se os epítomos a partir da sequência consenso mundial. Assim utilizou-se o servidor IEDB (Immune Epitope Database and Analysis Resource) para as análises necessárias, concentrando-se na identificação de epítomos de células B, MHC-I e MHC-II (<https://www.iedb.org/>).

Para predição de epítomos de célula B utilizou-se a metodologia Bepipred Linear Epitope Prediction 2.0 (<http://tools.iedb.org/bcell/>), que consiste na utilização de um algoritmo chamado “*Random Forest*”, que é treinado para a identificação de epítomos em sequências lineares, definindo-se um “*score*” para cada aminoácido. Dessa forma, para a escolha dos epítomos um *cutoff* de 0,50 é estabelecido, sendo selecionados, aqueles que atingirem pontuação superior.

Para a predição de epítomos de MHC-I utilizou-se a metodologia NetMHCpan EI 4.1 (<http://tools.iedb.org/mhci/>), que consiste em um algoritmo treinado para a análise de dados de ligação de ligante eluído, ou seja, realiza uma quantificação da possível afinidade de

ligação da região ao MHC. Para a obtenção dos epítomos mais promíscuos, definiu-se assim o critério de $IC_{50} < 50Nm$ que é recomendado pelo IEDB, em que 1% a partir da quantidade total de epítomos foi selecionada. Vale ressaltar que o IEDB realiza as predições de epítomos de MHC de classe I e II a partir de 27 alelos de referência que representam uma cobertura populacional mundial de 97 a 99%.

Para a predição de epítomos de MCH-II utilizou-se a metodologia “IEDB recommended 2.22” (<http://tools.iedb.org/mhcii/>), que representa a utilização de todas as ferramentas de maneira consenso presentes no servidor como NN-align, SMM-align, CombLib e Sturniolo, além da adaptação do melhor algoritmo para determinada sequência. Assim, para a escolha dos epítomos, definiu-se que os epítomos com limiar percentil de *consensus rank* menor que 20.0, além da interação com mais ou igual a 50% dos alelos referência (13/14) seriam selecionados.

Vale ressaltar que todas as metodologias, bem como definições de limiar de corte, seguiram os padrões recomendados pelo servidor IEDB, visando manter-se uma coerência nos resultados.

5.1.3 Alinhamento e identificação de regiões transmembrana

A partir de todos os epítomos encontrados, realizou-se o alinhamento deles no servidor MultAlin (<http://multalin.toulouse.inra.fr/multalin/>) e de forma concomitantemente o alinhamento manual, visto que os epítomos apresentavam quantidades diferentes de aminoácidos. Assim, considerou-se um tamanho de 15 aminoácidos, visto que para caber dentro da fenda de MHC-I é necessária uma quantidade de 8 a 11 aminoácidos e no MCH-II pode-se ter até 30 aminoácidos. Com isso, a escolha de 15 aminoácidos atua de forma satisfatória para ambos os tipos de MHC. Em seguida, foram selecionados os epítomos com maior taxa de repetição, que estivessem presentes em regiões fora das regiões transmembrana da proteína, visto a dificuldade de se conectar a estas regiões. Assim, realizou-se a análise das regiões transmembrana da proteína, a partir do servidor SOSUI (<https://harrier.nagahama-i-bio.ac.jp/sosui/>), ou seja, epítomos que estivessem nessa região seriam removidos.

5.1.4 Teste de antigenicidade, alergenicidade e propriedades físico-químicas

Após a obtenção dos epítomos estabelecidos com tamanhos de 15 aminoácidos, deu-se início ao processo de análise dos mesmos. Primeiramente realizou-se o teste de

antigenicidade no servidor *Vaxijen* v2.0 (<http://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>), em que todos os epítomos foram inseridos no servidor que calcula um *score* para as regiões do epítomo, além de um *cutoff* de 0,50, além da etiqueta “provável antígeno” foram definidos. Vale ressaltar, que para este caso, a metodologia “vírus” foi aplicada.

Em seguida, realizou-se o teste de alergenicidade no servidor *AlgPredf: Prediction of Allergenic Proteins and Mapping of IgE Epitopes* (<http://crdd.osdd.net/raghava/algpred/>), no qual os epítomos aprovados no teste de antigenicidade foram inseridos no servidor e a partir do método híbrido (*SVMc*, *IgE*, *epitope*, *ARPs BLAST* e *MAST*) que consiste na utilização consenso de todas as ferramentas para a predição de alergenicidade.

Por fim, utilizou-se a ferramenta *ProtParam* do servidor *Expasy* (<https://web.expasy.org/protparam/>) para os cálculos das características físico-químicas dos epítomos, de modo a encontrar dados como peso molecular, ponto isoelétrico, fórmula estrutural, número de átomos, número de aminoácidos, tempo de meia vida e estabilidade estrutural do epítomo, sendo este último apresentando grande impacto. Desta forma, epítomos que apresentassem um valor de cálculo de estabilidade superiores a 40, foram considerados instáveis e retirados.

5.1.5 Identificação de regiões de N-glicosilação

Para a identificação de regiões de N-glicosilação presentes na glicoproteína *spike* utilizou-se o servidor *NetNglyc* 1.0 *Sever* (<https://services.healthtech.dtu.dk/service.php?NetNGlyc-1.0>). O servidor a partir de algoritmos treinados consegue encontrar sequências de aminoácidos que provocam regiões de dobramento que facilitam e criam um mecanismo de escape para o vírus e conseqüentemente para o epítomo. É importante notar que esta etapa, atua de forma a elucidar as regiões de N-glicosilação, mas não atua como critério de exclusão dos epítomos.

5.1.6 Modelagem molecular por homologia e *molecular docking*

Visto que na época, não se tinha estruturas tridimensionais da glicoproteína *spike* depositada em bancos de dados, fez se necessário a utilização da modelagem molecular por homologia para a análise da proteína e conseqüentemente obtenção do epítomo para análise de

molecular docking. Para isso, utilizou-se os algoritmos do *AlphaFold* (<https://github.com/deepmind/alphafold>).

Já para o ancoramento molecular, usou-se o software AutoDock Vina (<https://vina.scripps.edu/>), no qual para o ligante, recortou-se a região do epítipo da glicoproteína *spike* e para o receptor as estruturas tridimensionais dos alelos referências utilizadas pelo servidor IEDB, além de estruturas de MHC provenientes do organismo *Mus musculus*, visto que os testes *in vivo* seriam realizados em tal animal. Para isso, uma pesquisa foi realizada no banco de dados Protein Data Bank, coletando-se as estruturas tridimensionais dos alelos de *Homo sapiens* e *Mus musculus* respectivamente. Em seguida realizou-se a definição da *GridBox* nestes alelos, que representa um campo de força para que o software AutoDock Vina realiza as análises em que, as regiões de ranhura do MHC foram definidas como sítio ativo e posteriormente o processo de *molecular docking* desses epítipos nas proteínas de MHC, ou seja, definiu-se a estrutura do alelo de maneira rígida e a estrutura do epítipo de maneira flexível. Para a análise de interações como pontes de hidrogênio, pontes de sal, interações hidrofóbicas, entre outras, utilizou-se o software LigPlot +.

Vale ressaltar que além dos ancoramentos moleculares utilizando-se o software AutoDock Vina que realiza um processo de *molecular docking* rígido flexível, realizou-se os processos de ancoramento moleculares nos servidores ClusPro e MdockPep, no qual estes servidores são direcionados para o ancoramento proteína-peptídeo com características de *molecular docking* flexível-flexível.

5.1.7 Validação da estrutura obtida por modelagem molecular por homologia

Foi realizado um teste comparativo para validação do resultado obtido pela metodologia de modelagem molecular por homologia, em que se comparou o resultado com uma estrutura da proteína *spike* obtida pelo método de cristalografia de difração de raio-x e depositada no banco de dados PDB. Utilizou-se o software Pymol para a obtenção do valor de Root Mean Square Deviation (RMSD), este valor representa a medida da distância média entre os átomos de proteínas sobrepostas.

5.1.8 Síntese dos epítomos

Os epítomos selecionados foram sintetizados e sua pureza foi identificada por cromatografia líquida de alta pressão (HPLC) e a sequência foi confirmada por espectrometria de massa (ANEXO II).

5.2 IN VITRO

5.2.1 Animais experimentais

Foram utilizados camundongos isogênicos C57BL/6 com seis a oito semanas de idade, provenientes do Biotério Central da Universidade Federal de Alfenas (UNIFAL-MG). Esses animais foram mantidos em sala climatizada, em caixas de polipropileno, com livre acesso à água e alimento e em obediência ao ciclo circadiano de claro/escuro. Os protocolos do experimento foram submetidos e aprovados pela Comissão de Ética no Uso de Animais da UNIFAL-MG, sob registro 0008-2021 (ANEXO I).

5.2.2 Cultura de macrófagos derivados da medula óssea

Camundongos C57BL/6 foram eutanasiados para obtenção de células da medula óssea e foram extraídos o fêmur e a tíbia desses animais. Posteriormente, esses ossos foram cortados em suas epífises e injetado cerca de 5 mL de solução salina estéril por osso. Após isso, as células obtidas foram centrifugadas a 1200 rpm por 10 minutos e ressuspensas em 1 mL de DMEM (Dulbecco's Modified Eagle Medium) suplementado com 10% de soro fetal bovino, 1% de Penicilina/Estreptomicina e 1% de Heps.

Após a centrifugação, foi feita a filtração em separadores de nylon com a malha de 70 µm para a remoção de detritos e ruptura de ligações célula-célula. Em seguida, as células foram plaqueadas em Placas de Petri e incubadas a 37°C, em atmosfera de 5% de CO₂, de maneira a obter somente células não diferenciadas por aproximadamente 2 horas. Foram coletadas as células não aderentes e posteriormente centrifugadas a 1200 rpm por 10 minutos e ressuspensas em 1 mL de DMEM suplementado. As células viáveis foram quantificadas por meio da coloração com o corante Azul de Tripán na câmara de Neubauer. Estas foram plaqueadas para obtenção de macrófagos derivadas da medula óssea (BMDMs).

Foi utilizado o meio DMEM suplementado e sob sua composição foi adicionado o fator de diferenciação de colônia de monócitos (25 ng/mL de M-CSF) na concentração de 5×10^5 células por poço (placa de 24 poços) e na concentração 1×10^6 por poço (placa de 6 poços). As placas foram incubadas a 37°C e sob atmosfera de 5% de CO₂. Cinco dias após o plaqueamento das células, para a diferenciação em macrófagos, foram adicionados 500 µL por poço de meio DMEM suplementado e sob sua composição foi adicionado M-CSF (50 ng/mL). No décimo dia a célula foi analisada quanto sua morfologia e foi realizada a estimulação dos BMDMs.

5.2.3 Estimulação de BMDMS com BCG e peptídeos da proteína *spike*

Após a diferenciação das células BMDMs, estas foram estimuladas com BCG (MOI: 0,1), peptídeo 1 (1 mL/poço), peptídeo 2 (1 mL/poço) e, com a combinação de BCG mais peptídeo. As células foram estimuladas em momentos distintos. Após 24 horas do primeiro estímulo, os sobrenadantes foram coletados, aliquotados e refrigerados. Em seguida, foram adicionados em todos os poços 500 µL do meio DMEM suplementado sem M-CSF e as células foram mantidas em repouso. No segundo momento, as células foram submetidas a um novo estímulo com BCG, peptídeo e BCG mais peptídeo. Após 48 horas da segunda estimulação, os sobrenadantes foram coletados, aliquotados e refrigerados para análise.

5.2.4 Extração de proteínas e *Western Blotting*

BMDMs previamente estimuladas ou não estimuladas foram submetidas ao processo de extração de proteínas totais. Os sobrenadantes das células foram descartados e posteriormente sobre os poços foram adicionados 100 µL do tampão de lise celular e extração RIPA (Sigma, R0278). Além disso, as amostras foram suplementadas com um coquetel com 10% de inibidor de protease (Sigma, P2714) e um coquetel com 1% de inibidor de fosfatase (Sigma, P0044). Em seguida os poços foram raspados e a solução suplementada foi perpassada poço a poço. Subsequente a isso, o material de proteínas foi centrifugado em 12000 rpm por 5 minutos para remoção dos detritos celulares.

Para execução do protocolo de *Western Blotting* foi preparado o gel de poli-acrilamida (água, 30% Acrilamida Mix, 1,5 M TRIS (pH 8,8), 10% SDS, 10% Persulfato de Amônio e Temed) e o gel de entrada (água, 30% Acrilamida Mix, 1,0 M Tris (pH 6,8), 10% SDS, 10% Persulfato de Amônio e Temed). Para o preparo das amostras, elas foram

diluídas em água e receberam a adição de tampão de amostra Leammli 4x (BIO-RAD, 1610747) + 10% de β -mercaptoetanol. Em seguida, essas foram submetidas em fervura a aproximadamente 98 °C por 5 minutos.

Na cuba de eletroforese foi colocado o tampão de corrida (água, glicina, Tris). No primeiro poço do gel, foi colocado o padrão de peso molecular e subsequente as amostras previamente preparadas, com 10 μ g de proteína total por poço. A corrida deu-se início em 80 mV e posteriormente 120 mV, com duração de aproximadamente 120 minutos. Após a corrida o gel foi transferido para uma membrana de nitrocelulose e bloqueada com leite a 5% por 2 horas para diminuir as ligações inespecíficas da membrana, que podem interferir com a ligação dos anticorpos específicos.

Em seguida, para a remoção do bloqueio a membrana foi lavada por 5 vezes em um intervalo de 5 minutos com tampão de lavagem TBS-T. Logo após, a membrana foi incubada com o anticorpo primário anti-rabbit p38 1:1000 (Cell Signaling/8690), anti-rabbit p44/42 1:1000 (Cell Signaling/4695), anti-rabbit SAPK/JNK 1:1000 (Cell Signaling/9252), no qual ficou *overnight* sob agitação e refrigeração (4° C). No dia seguinte, o anticorpo primário foi retirado e novamente a membrana foi lavada por 5 vezes em um intervalo de 5 minutos com TBS-T. Posteriormente, foi inserido o anticorpo secundário anti-rabbit IgG 1:2000 (Cell Signaling/7074) correspondente para marcar especificamente o anticorpo primário. A membrana foi incubada com este anticorpo com agitação por 2 horas em temperatura ambiente. Ao término do período estipulado, lavou-se a membrana novamente com TBS-T, por 5 vezes de 5 minutos. Por fim, a revelação deu-se por quimioluminescência, pela adição do reagente Clarity ECL (BIO-RAD, 1705060) e substrato à membrana por aproximadamente 1 minuto, e depois as bandas foram detectadas pelo dispositivo ChemiDoc XRS + System (BIO-RAD), com o auxílio do *software* Image Lab (BIO-RAD). Foi analisado a via da MAPK a fim de analisar o comportamento visto o primeiro e o segundo contato da formulação desenvolvida.

5.2.5 Análise por ELISA em células BMDMS

Foram utilizados os sobrenadantes das células BMDMs para a análise quantitativa de secreção das citocinas IL-6, IL-1 β e TNF- α através do kit Mirune IL-6/IL-1 β /TNF- α Standard ABTS ELISA Development kit (PreproTech, Cranbury, NJ, USA). Em suma, uma placa Nunc MaxiShopTM foi sensibilizada com o anticorpo de captura na concentração de 1 μ g/mL *overnight*, seguido de 4 lavagens com solução 0,05% Tween-20 em PBS. Em seguida, foi

adicionado o tampão de bloqueio 1% de BSA em PBS, incubado por 1 hora, seguido de 4 lavagens. Após, foram adicionados o padrão (IL-6/2000 pg/mL, IL-1 β /4000 pg/mL e TNF- α /2000 pg/mL e diluído serialmente), as amostras e incubados por 2 horas, seguido de lavagem, adição do anticorpo de detecção (1 μ g/mL) por duas horas, lavagem e adição do conjugado Avidin-HRP na diluição 1:200 e incubado por 30 minutos, com posterior lavagem. Por fim, foi adicionado o substrato ABTS, incubados a temperatura ambiente até o desenvolvimento de cor e leitura em espectrofotômetro com densidade óptica de 405 nm com correção de 650 nm.

5.3 *IN VIVO*

5.3.1 Animais experimentais

Foram utilizados camundongos isogênicos C57BL/6 com seis a oito semanas de idade, provenientes do Biotério Central da Universidade Federal de Alfenas (UNIFAL-MG). Esses animais foram mantidos em sala climatizada, em caixas de polipropileno, com livre acesso à água e alimento e em obediência ao ciclo circadiano de claro/escuro. Os protocolos do experimento foram submetidos e aprovados pela Comissão de Ética no Uso de Animais da UNIFAL-MG, sob registro 0008-2021 (ANEXO I).

5.3.2 Imunização de animais

Camundongos C57/BL6 (fêmeas) foram divididos em grupos distintos (n=6/grupo) e estes foram imunizados com a combinação de BCG (5×10^4 UFC) com o peptídeo 1 (1 μ g/ μ L) e peptídeo 2 (1 μ g/ μ L) da proteína *spike*. Foram realizadas duas imunizações por via intramuscular com intervalo de 15 dias entre elas e 15 dias após a segunda imunização os animais foram eutanasiados. Foram coletados soro e baço para análises. Por animal foi administrado 50 μ L da solução vacinal composta por 5 μ L de BCG, 10 μ L de peptídeo e 35 μ L de solução salina.

5.3.3 Cultura de esplenócitos

Camundongos C57BL/6 foram eutanasiados por dose letal de anestésico (260 μ L de Kitamina, 130 μ L de Xilasina, 1110 μ L de solução salina), sendo administrado 200 μ L/animal

de anestésico. Foram coletados o baço desses animais. A solução contendo os esplenócitos foram obtidas da maceração do baço. Em seguida, essa solução foi centrifugada em 1200 rpm por 7 minutos em temperatura controlada (4°C). O sobrenadante foi descartado e após foi adicionado 1 mL de solução ACK (0,15 M NH₄Cl, 1,0 mM KHCO₃, 0,1 mM Na₂EDTA, pH 7,2) e ressuspenso. Após agir por 5 minutos o volume foi completado com salina estéril e novamente foi submetido a centrifugação em 1200 rpm por 7 minutos. Em seguida, o sobrenadante foi descartado e as células viáveis foram ressuspenso com 1 mL de meio de cultura RPMI 1640 (Gibco, Carlsbad, Calif) suplementado com 10% de soro fetal bovino inativado, 1% de Penicilina/Estreptomicina. As células viáveis foram quantificadas por meio da coloração Azul de Tripán a fim de obter a concentração de 1x10⁶ cel/poço.

Posteriormente as células foram plaqueadas e os respectivos grupos de imunização foram estimulados com meio de cultura suplementado (10% de soro fetal bovino, 1% de Penicilina/Estreptomicina, 1% de HEPES), peptídeo 1 (1 mL/poço), peptídeo 2 (1 mL/poço), BCG (MOI: 0,1) e BCG em combinação com o peptídeo 1 e BCG em combinação com o peptídeo 2 da proteína *spike*. Após a estimulação das células de esplenócitos, o sobrenadante foi aliquoteado para análise.

5.3.4 Análise por CBA

Foi feito o ensaio CBA (Cytometric Bead Array) com a colaboração da Universidade de São Paulo (USP), campus de Ribeirão Preto, para análise da secreção de citocinas que caracterizam os perfis Th1, Th2 e Th17 murino através de citometria de fluxo. A análise foi feita pelo software FCAP array.

5.3.5 Ensaio sorológico: quantificação de IGG total e subclasses

Placas de ensaio imunoenzimático foram revestidas com pool de antígeno (peptídeos estruturais do vírus SARS-CoV-2) sendo administrado 1µg/µL por poço suplementado com o tampão de ligação (0,075g Na₂CO₃ / 0,1465g NAHCO₃ / 50mL H₂O destilada). Os subtipos foram medidos em amostra de soro por anticorpos anti-imunoglobulina G, IgG Total (1:2500), IgG1 (1:5000) e IgG2c (1:20000) conjugados com peroxidase (Sigma Chemical Co., EUA).

As placas foram bloqueadas com tampão de bloqueio (PBS contendo 5% de leite desnatado) e incubadas por 2 horas a 37°C, em atmosfera de 5% de CO₂. As amostras foram preparadas e diluídas em tampão de incubação (PBS-T, 3% de soro fetal bovino).

Subsequente, foi adicionado 100 μ l/poço dos antissoros (não vacinado, vacinado: BCG + Peptídeo 1 e vacinado: BCG + Peptídeo 2) e incubadas por 1 hora a 37°C, em atmosfera de 5% de CO₂.

Após descartado o conteúdo anterior, foi adicionado 100 μ l/poço da solução de anticorpo secundário (IgG Total, IgG1, IgG2c) e incubadas por 1 hora a 37°C, em atmosfera de 5% de CO₂. Após lavagem com PBS-T foi adicionado a cada poço a solução de substrato (1 mg/mL TMB, 4 μ L H₂O₂, pH 5) e as placas foram incubadas no escuro à temperatura ambiente. Para pausar a reação foi adicionado 100 μ L de H₂SO₄2N, após 10 minutos. A absorbância foi medida em espectrofotômetro no comprimento de onda de 450nm.

6 ANÁLISES ESTATÍSTICAS

Os dados obtidos foram analisados usando o programa GraphPadPrism 8 (GraphPad Software, Inc), através de teste one way ANOVA, com pós teste de Tukey. Utilizando nível de significância $p < 0,05$.

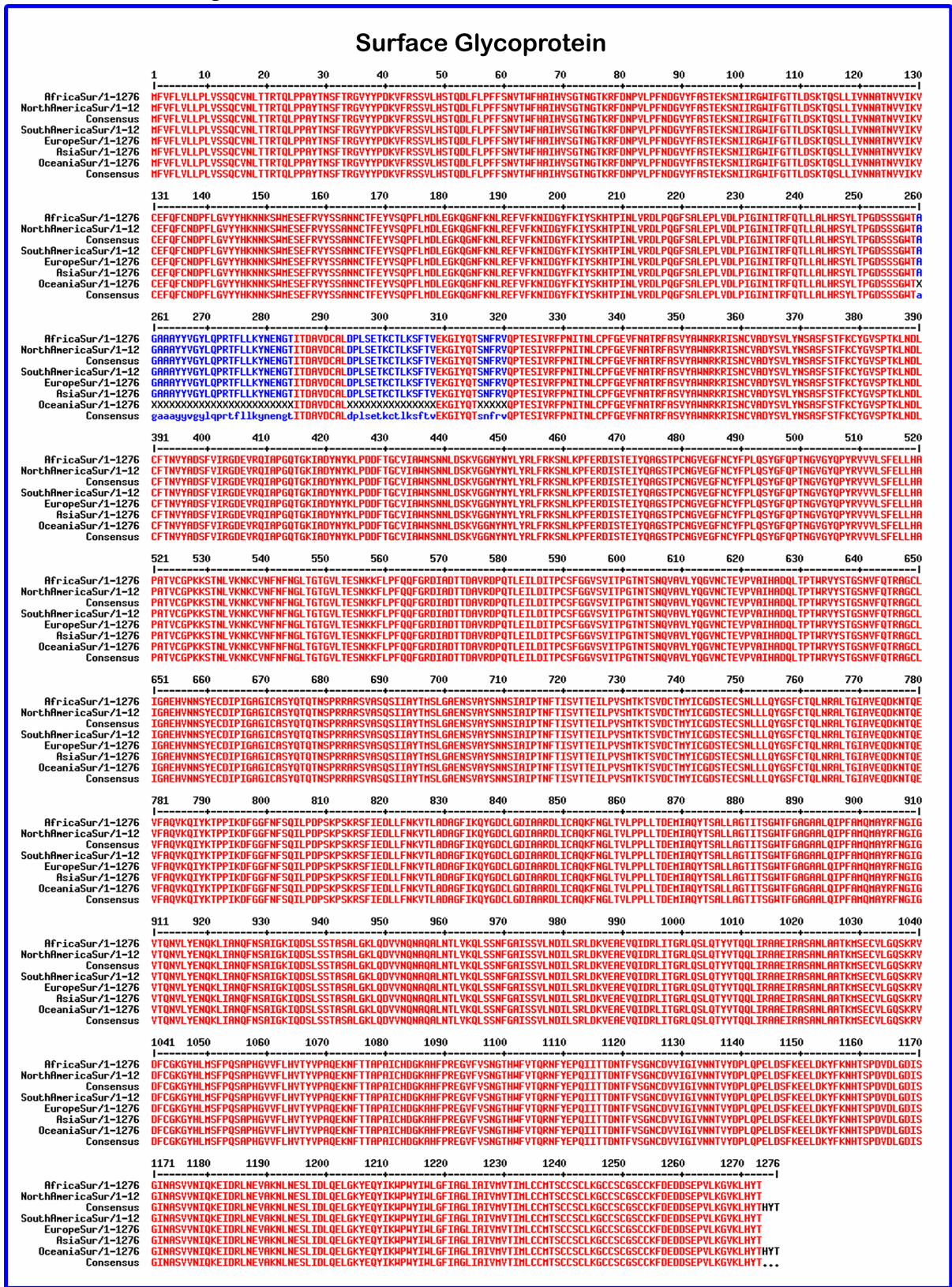
7 RESULTADOS

7.1 *IN SILICO*

7.1.1 Uma sequência global com 1.276 aminoácidos foi identificada como a sequência consenso da glicoproteína *spike*

Foram encontrados 53.838 genomas que haviam sido depositados até em Agosto de 2020 no banco de dados NCBI. Desses genomas foi identificada as putativas sequências codificadores da proteína *spike*, obtendo assim uma sequência consenso por região do globo terrestre, sendo na África (109 sequências), América do Norte (10.331 sequências), América do Sul (31 sequências), Ásia (1.219 sequências), Europa (414 sequências) e Oceania (1.418 sequências), totalizando 13.522 sequências passíveis de identificação da proteína *spike*. Posteriormente aos dados encontrados os 1.276 aminoácidos da glicoproteína *spike* foram alinhados como ilustrado na figura 5. No entanto, a sequência obtida no continente Oceania (figura 5) apresentou diferenças, em que das 1.418 sequências encontradas havia uma região inexistente na sequência consenso da glicoproteína *spike* o que não é o caso das demais regiões. Além disso, nessa mesma região foram encontrados três aminoácidos que as outras regiões não tinham. Tais diferenças podem ser justificadas por possíveis mutações que ocorrem entre as regiões de cada continente.

Figura 5 – Alinhamento das seqüências obtidas por região que caracteriza a proteína *spike* do SARS-CoV-2 em cada continente a partir dos resultados obtidos da seqüência consenso por continente.



Fonte: Da autora.

7.1.2 Foram identificados 833 epítomos da glicoproteína *spike* a partir da sequência global

Utilizou-se o IEDB para realizar a essa análise de predição de epítomos de célula B, usando o método Bepipred Linear Eptpe Prediction 2.0, esse algoritmo treinado com outras sequências já conhecidas verificou os aminoácidos e deu uma pontuação de predição afim de averiguar se é ou não um candidato a ser um epítomo de célula B. Um limiar de corte de 0,50 foi definido, em que os aminoácidos com pontuação acima de 0,50 foram considerados. Assim da sequência consenso mundial foi encontrado para a glicoproteína *spike* um total de 33 epítomos de célula B.

Para predição de epítomos de MHC-I, utilizou-se o método NetMHCpan EL 4.1 sendo esse recomendado pelo próprio IEDB, no qual aliou-se a utilização de 27 alelos de referência, que são os alelos que englobam toda a população mundial. Assim, 68500 epítomos foram obtidos, classificados de mais promíscuo para menos promíscuo. Entretanto, um limiar de corte de 1% sugerido pelo IEDB foi definido e 685 epítomos de MHC classe I foram considerados.

Para a predição de epítomos de MHC-II, utilizou-se o método IEDB recommended 2.2, este comando se refere ao método NetMHCIIpan EL 4.0, no qual utilizou-se cinco ferramentas de forma híbrida disponibilizadas pelo servidor e usando novamente os 27 alelos de referência obteve-se 34.800 epítomos. Em seguida um limiar percentil consenso rank < 20.0 e porcentagem da quantidade de alelos que o epítomo interage > 50% foram definidos, ou seja, era necessário que pelo menos 13 a 14 alelos fizessem uma interação com o epítomo e assim 115 epítomos foram considerados. A glicoproteína *spike* obteve um total de 833 epítomos, sendo 33 epítomos de célula B, 685 de MHC de classe I e 115 de MHC de classe II respectivamente.

7.1.3 Os 833 epítomos foram alinhados e 812 epítomos foram removidos por estar na região transmembrana

Utilizou-se o servidor MutlAlin para o alinhamento dos 833 epítomos obtidos (célula B, MHC-I e MHC-II). Foi necessário o alinhamento manual, pois a ferramenta MultAlin não alinhava os epítomos com a sequência global obtida e sim com os epítomos entre si. Através desse alinhamento, foram selecionados epítomos com 15 aminoácidos. Assim, após o alinhamento completo dos 833 epítomos, retirou-se todos os epítomos presentes nas regiões

transmembrana encontrados nos resultados do SOSUI, visto que um epítopo presente na região transmembrana dificulta a conexão com as proteínas de MHC classe I e II para serem apresentados e servirem de alvos moleculares aos linfócitos T. O total de 21 epítopos foram obtidos ressaltando que epítopos que apresentavam mais de uma repetição eram desconsiderados.

7.1.4 Seis epítopos foram considerados antigênicos, não-alergênico e estáveis

Iniciaram-se os testes de propriedades dos epítopos em que os 21 epítopos foram testados primeiramente para antigenicidade, no qual para ser antigênico o epítopo precisava atingir um limiar de corte maior que 0,50 e apresentar a etiqueta “provável antígeno” dentro do servidor Vaxijen. Dos 21 epítopos iniciais, apenas 9 epítopos passaram no primeiro teste.

O segundo teste aplicado, foi o de alergenicidade usando o servidor AlgPredf em que a metodologia híbrida sugerida pelo servidor foi utilizada. Os resultados obtidos apresentaram que os 9 epítopos eram considerados não alergênicos.

Em seguida, os 9 epítopos foram submetidos a testes de propriedades físico-químicas em que o fator de grau de estabilidade foi definido como limiar de corte. Assim epítopos que apresentavam valor > 40 foram removidos, visto que ao não apresentarem estabilidade, sua produção para testes *in vivo* e *in vitro* não seria possível. Este grau de estabilidade é calculado pela ferramenta ProtParam em que ele visualiza a sequência de aminoácido e a partir das interações entre aminoácidos consegue calcular se esse epítopo é estável físico-quimicamente. Assim, dentre os 9 epítopos, 3 foram excluídos pelo critério de estabilidade. Vale ressaltar que outras informações foram obtidas como ponto isoelétrico, peso molecular, fórmula molecular, predições de meia vida em organismos como *Mammalian reticulocytes* e *Escherichia coli*, conforme descritos na tabela 1.

Tabela 1 – Avaliações físico-químicas dos epítopos finais mais promíscuos da proteína *spike*.

Epítopos	Overall Prediction for the Protective Antigen	Molecular weight	Theoretical pI	Instability index	This classifies the protein as	Mammalian reticulocytes <i>in vitro</i>	Yeast, <i>in vivo</i>	<i>Escherichia coli</i> , <i>in vivo</i>
ALQIPFAMQMAYRFN	1,0112	1801.15	8.79	33.91	stable	4.4 hours	>20 hours	>10 hours
IAIPTNFITISVTTEI	0,7719	1619.88	4.00	7.39	stable	20 hours	30 min	>10 hours
RVVLSFELLHAPAT	0,7485	1651.97	6.75	35.01	stable	1 hours	2 min	2 min
LLFNKVTLADAGFIK	0,6737	1649.99	8.59	2.48	stable	5.5 hours	3 min	2 min
WTAGAAAYVGYLQP	0,6533	1630.82	5.52	25.13	stable	2.8 hours	3 min	2 min
LTPTWRVYSTGSNVF	0,5831	1727.94	8.75	-11.37	stable	5.5 hours	3 min	2 min

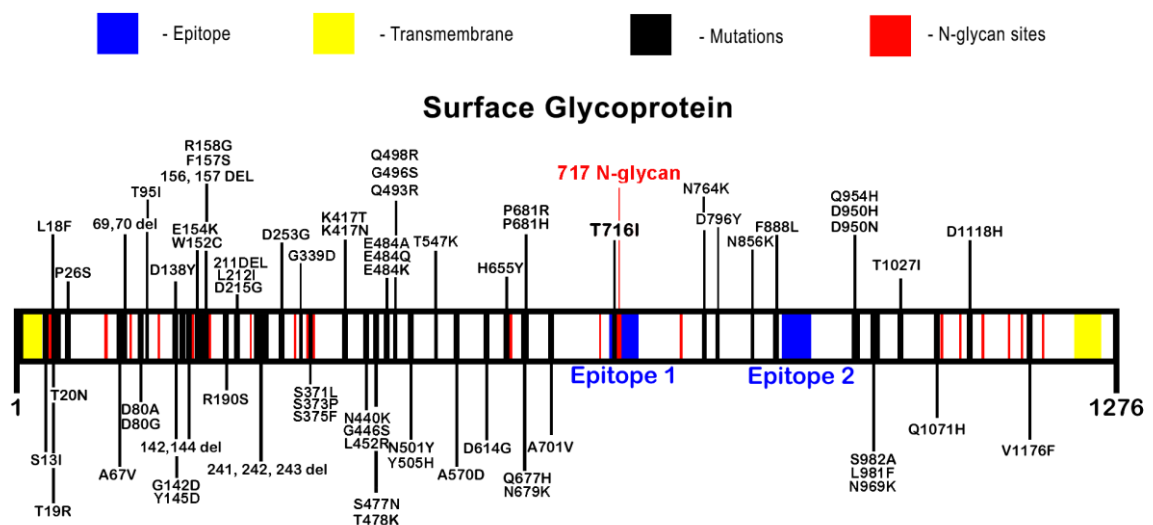
Fonte: Da autora.

7.1.5 Dois epítomos apresentaram maior probabilidade de serem potencialmente selecionados para análises *in vitro* e *in vivo*.

A partir dos dados presentes na tabela 1, dois peptídeos com os melhores resultados apresentados foram selecionados. O epítomo ALQIPFAMQMAYRFN apresentou melhor resultado em antigenicidade que todos os outros epítomos além de em grau de estabilidade inferior a 40 e o IAIPNFTISVTTEI que apesar de seu grau de estabilidade ser menor, no teste de antigenicidade obteve um bom resultado, ou seja, mais provável de ser antigênico quando comparado aos epítomos 3 e 4.

Vale ressaltar que o servidor NetNglyc 1.0 Sever detectou sítio de região de N-glicosilação no epítomo 1, além da presença de uma mutação (T716I) presente na variante Alpha, sendo estes resultados representadas na figura 6.

Figura 6 – Representação esquemática de regiões da glicoproteína *spike*.



Fonte: Da autora.

Legenda: Epítomos (azul). Regiões transmembrana (amarelo). Mutações (preto). Sítio de N-glicosilação (vermelho).

Dessa forma, as variantes do vírus SARS-CoV-2 já descritas, foram alinhadas tanto para o epítomo 1 quanto para o epítomo 2 (figura 7). No epítomo 1, no aminoácido 716, ocorre uma mutação que altera o aminoácido treonina (T) para isoleucina (I) na variante Alpha. Para o epítomo 2 não foram identificadas mutações capazes de alterar qualquer aminoácido na sequência.

Figura 7 – Alinhamento dos epítomos selecionados da sequência consenso da proteína *spike* em comparação com as sequências das variantes do vírus SARS-CoV-2.

	710	720	730	740		900	910	
Alpha	YSNN	IAIP	INFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC
Eta	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Beta	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Gama	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Delta	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Epsilon	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Kappa	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Iota	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Lambda	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
OmicronBA.1	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
OmicronBA.2	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
OmicronBA.2.12.1	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
OmicronBA.4	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
OmicronBA.5	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Zeta	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	
Consensus	YSNN	IAIPTNFTISV	TEII	PV	SMTKTSV	DCTHYICG	STEC	

Alpha	AG	ALQIPFAMQ	HAYRFN
Eta	AG	ALQIPFAMQ	HAYRFN
Beta	AG	ALQIPFAMQ	HAYRFN
Gama	AG	ALQIPFAMQ	HAYRFN
Delta	AG	ALQIPFAMQ	HAYRFN
Epsilon	AG	ALQIPFAMQ	HAYRFN
Kappa	AG	ALQIPFAMQ	HAYRFN
Iota	AG	ALQIPFAMQ	HAYRFN
Lambda	AG	ALQIPFAMQ	HAYRFN
OmicronBA.1	AG	ALQIPFAMQ	HAYRFN
OmicronBA.2	AG	ALQIPFAMQ	HAYRFN
OmicronBA.2.12.1	AG	ALQIPFAMQ	HAYRFN
OmicronBA.4	AG	ALQIPFAMQ	HAYRFN
OmicronBA.5	AG	ALQIPFAMQ	HAYRFN
Zeta	AG	ALQIPFAMQ	HAYRFN
Consensus	AG	ALQIPFAMQ	HAYRFN

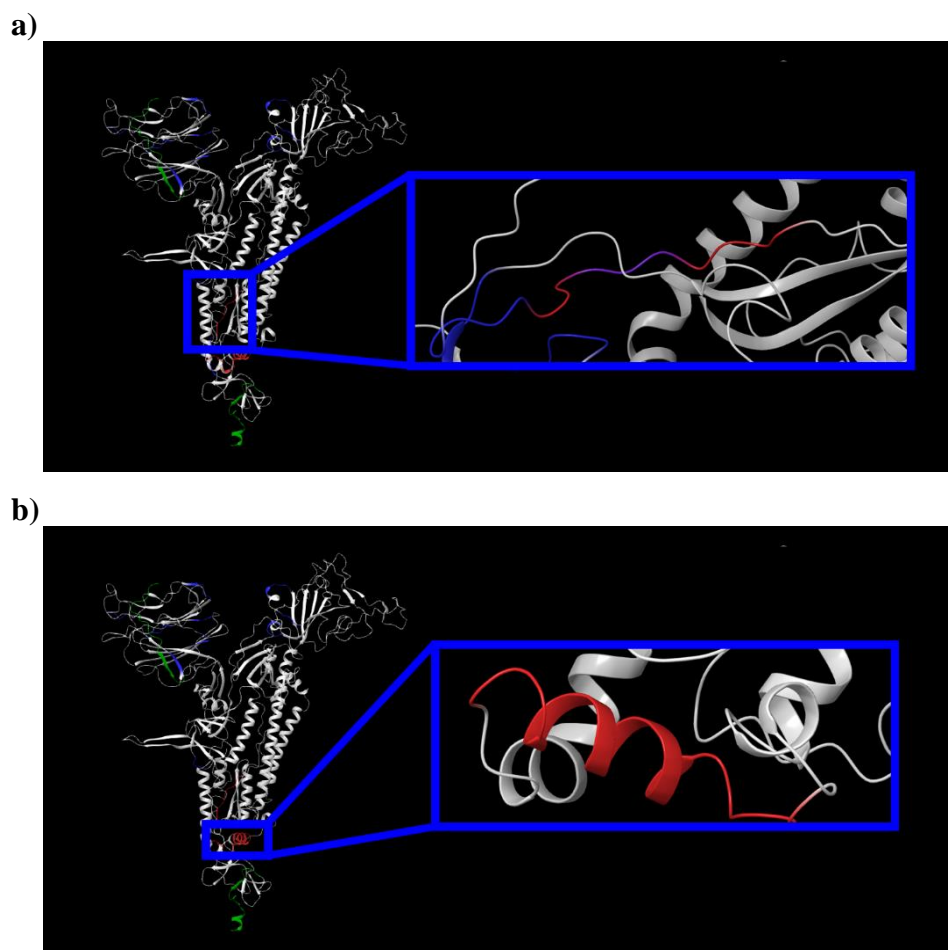
Fonte: Da autora.

Legenda: A sequência do epítopo 1, à esquerda, demonstra diferença no aminoácido da posição 716 na variante alpha em relação as demais variantes e a sequência consenso identificada nesse trabalho. O epítopo 2, à direita, não apresenta variações em relação a qualquer sequência alinhada.

7.1.6 A modelagem molecular por homologia da proteína consenso mundial da *spike* mostrou que os epítomos se apresentam na base da proteína e distantes de regiões de maior variabilidade

A modelagem molecular por homologia consiste no processo realizado quando não se tem a estrutura cristalografada da proteína, em que a confecção da estrutura 3D é realizada a partir de outras estruturas similares presentes em bancos de dados como *Protein Data Bank*. O objetivo da aplicação dessa ferramenta para a glicoproteína *spike*, a priori, é se ter uma estrutura da sequência global dessa glicoproteína, interesse do presente trabalho. A partir da obtenção da modelagem da proteína pode-se observar que os epítomos selecionados se apresentam na base da proteína e de tal forma, distantes de regiões de maior mutação, no qual observa-se que a região superior da cadeia, responsável por se conectar o ACE2 apresenta um maior número de mutações além da presença de inúmeros sítios de N-glicosilação, representados pela cor azul na figura 8.

Figura 8 – Epítomos da glicoproteína *spike* localizados na base da proteína.

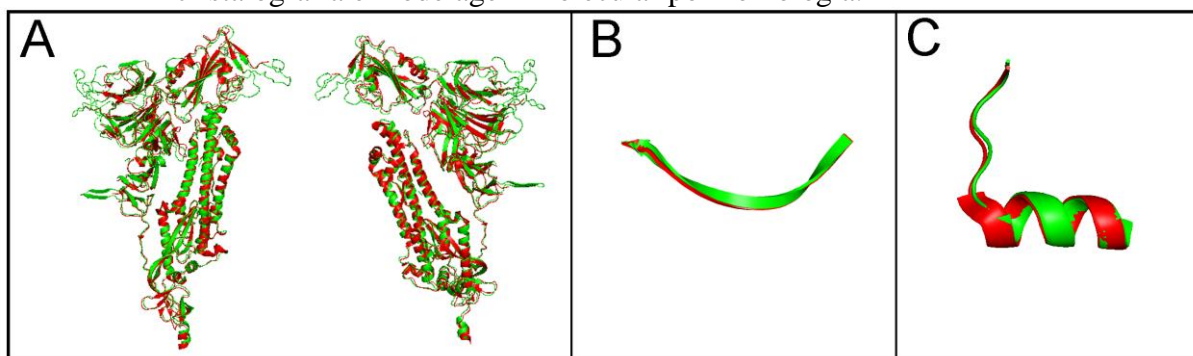


Fonte: Da autora.

Legenda: Epítopo 1 selecionado e destacado em vermelho **(a)**, regiões transmembrana (verde), regiões de N-glicosilação (azul). Epítopo 2 selecionado e destacado em vermelho **(b)**, regiões transmembrana (verde), regiões de N-glicosilação (azul).

Conforme mencionado anteriormente, a época de obtenção dos epítomos em Agosto de 2020, ainda não se tinha estruturas cristalografadas da proteína *spike* e de tal forma para se solucionar tal problema, realizou-se o processo de modelagem molecular por homologia. Entretanto, atualmente já se tem estruturas cristalografadas da estrutura. Um teste comparativo para se avaliar o resultado de modelagem molecular por homologia foi realizado, no qual para a comparação utilizou-se a estrutura de cristalografia da *spike* depositada no banco de dados PDB sob código 6VXX e a estrutura obtida por AlphaFold. Para este teste utilizou-se o software Pymol com o intuito de se obter o valor de Root Mean Square Deviation (RMSD), este valor representa a medida da distância média entre os átomos de proteínas sobrepostas (figura 9).

Figura 9 – Ilustração da comparação de estruturas entre proteínas *spike* obtidas por cristalografia e modelagem molecular por homologia.



Fonte: Da autora.

Legenda: Estruturas obtidas por cristalografia (PDB: 6VXX; verde) e modelagem molecular por homologia (AlphaFold; vermelho), no qual em (A) tem-se a visualização da sobreposição toda, em (B) tem-se o epítipo 1 e em (C) tem-se o epítipo 2.

O valor de RMSD obtido durante a avaliação foi de 1,08 e de tal forma, observando-se que para se ter um valor satisfatório é necessário que os resultados estejam abaixo de 2. Pode-se concluir que a proteína obtida e utilizada durante o processo de ancoragem molecular condiz com dados baseados em estruturas cristalografadas. Vale ressaltar que o valor de 1,08 ainda pode ser justificado a aminoácidos diferentes devido a mutações ou átomos ausentes durante a cristalografia, entretanto, apesar destes pontos, o resultado ainda é considerado satisfatório.

7.1.7 Avaliação da ancoragem molecular dos epítipos selecionados apresentam forte ligação às moléculas de MHC de *Homo sapiens*

Para as análises de ancoragem molecular utilizou-se como alvos os receptores de MHC-I e MHC-II em que os 27 alelos de referência utilizados na identificação de epítipos foram identificados no banco de dados *Protein Data Bank*. Entretanto, nem todos os alelos referências com suas estruturas cristalografadas foram encontradas. Onze alelos de MCH-I e 14 alelos de MHC-II foram identificados. Vale ressaltar que os alelos, além de estarem presentes nos 27 alelos referência determinado pelo IEDB, também deveriam apresentar interação com o epítipo conforme seleção de limiar de corte. Segundo essas informações obtidas, é provável que o epítipo 1 interaja mais com MHC-I e o epítipo 2 interaja mais com MHC-II. Os resultados obtidos estão demonstrados na tabela 2.

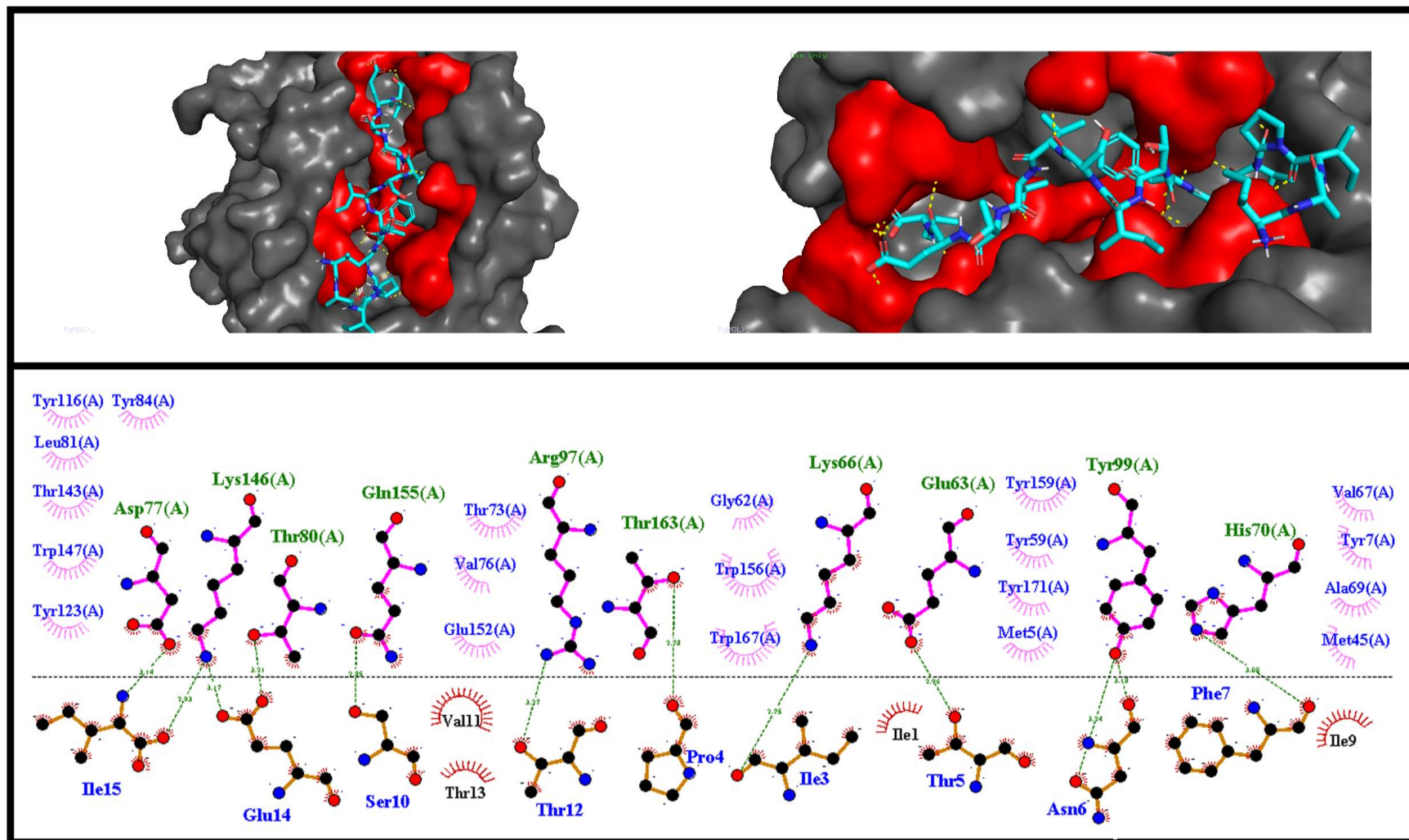
Tabela 2 – Interações dos alelos de MHC-I e MHC-II dos epítomos selecionados.

		ALELO	ClusPro		MdockPep	AutoDock Vina
			Center	Lowest Energy		
EPÍTOPO 1	MHC-I	HLA-A*0203	-792,0	-931,7	-187,7	-8,3
	MHC-II	HLA- DRB1*04:01	-722,0	-844,1	-200,8	-7,2
EPÍTOPO 2	MHC-I	HLA-B*3501	-794,6	-1044,9	-234,9	-7,9
	MHCII	HLA- DRB1*15:01	-797,3	-996,6	-263,6	-7,7

Fonte: Da autora.

Além disso, essas interações foram quantificadas no qual foram visualizadas 12 pontes de hidrogênio, 24 interações hidrofóbicas e nenhuma interação com as pontes de sal para o epítomo 1 no alelo HLA-A*0203 de MHC-I (figura 10).

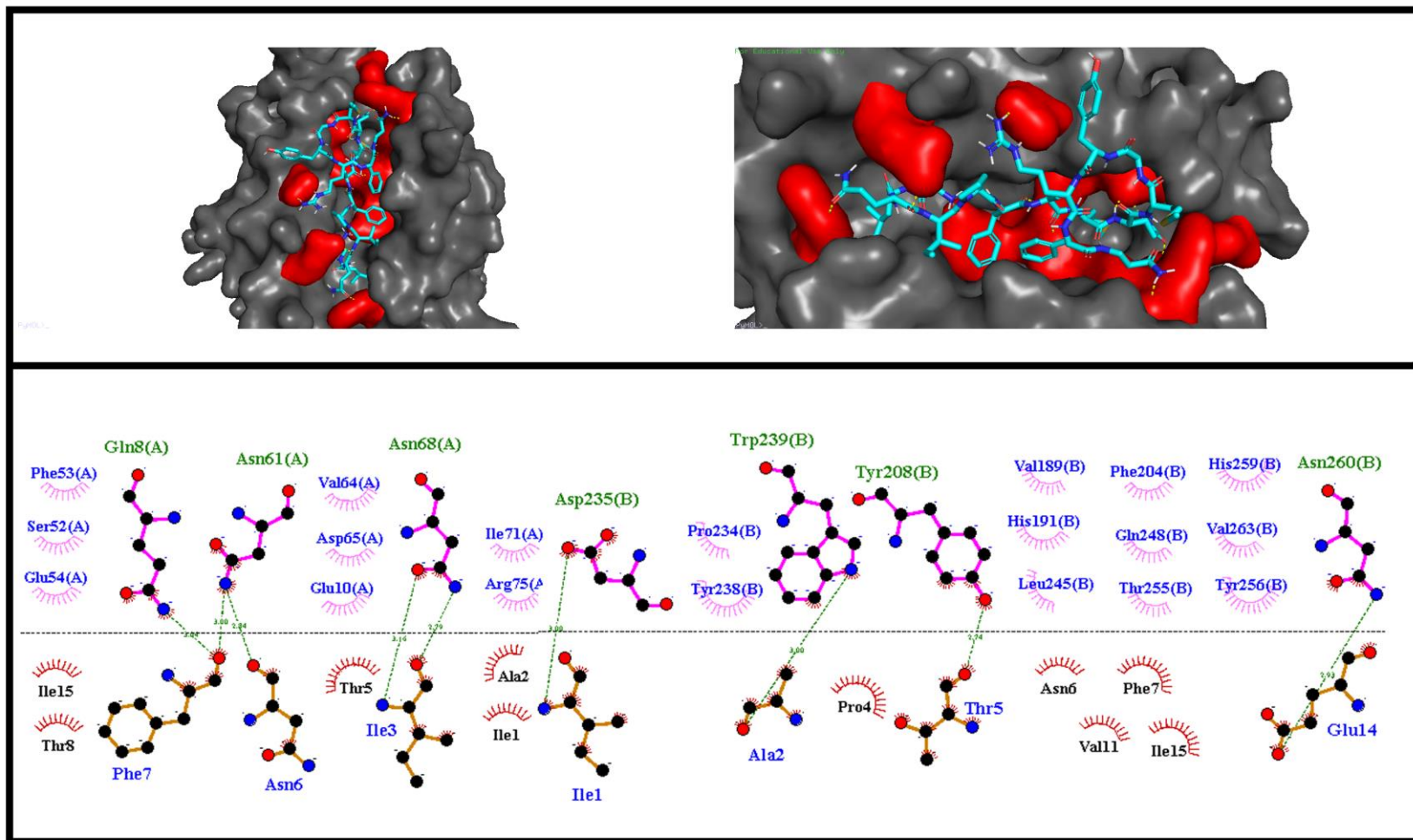
Figura 10 – Representação do diagrama 3D e 2D da interação do Epítipo 1 com o receptor MHC de classe I, alelo HLA-A*0203.



Fonte: Da autora.

No alelo HLA-DRBI*04:01 de MHC-II, para o mesmo epítipo, foram visualizadas 9 pontes de hidrogênio, 29 interações hidrofóbicas e também nenhuma interação com as pontes de sal (figura 11). Vale ressaltar que as análises de interações, foram realizadas nestes respectivos alelos, pois os mesmos apresentaram os melhores resultados de ancoramento molecular no software AutoDock Vina.

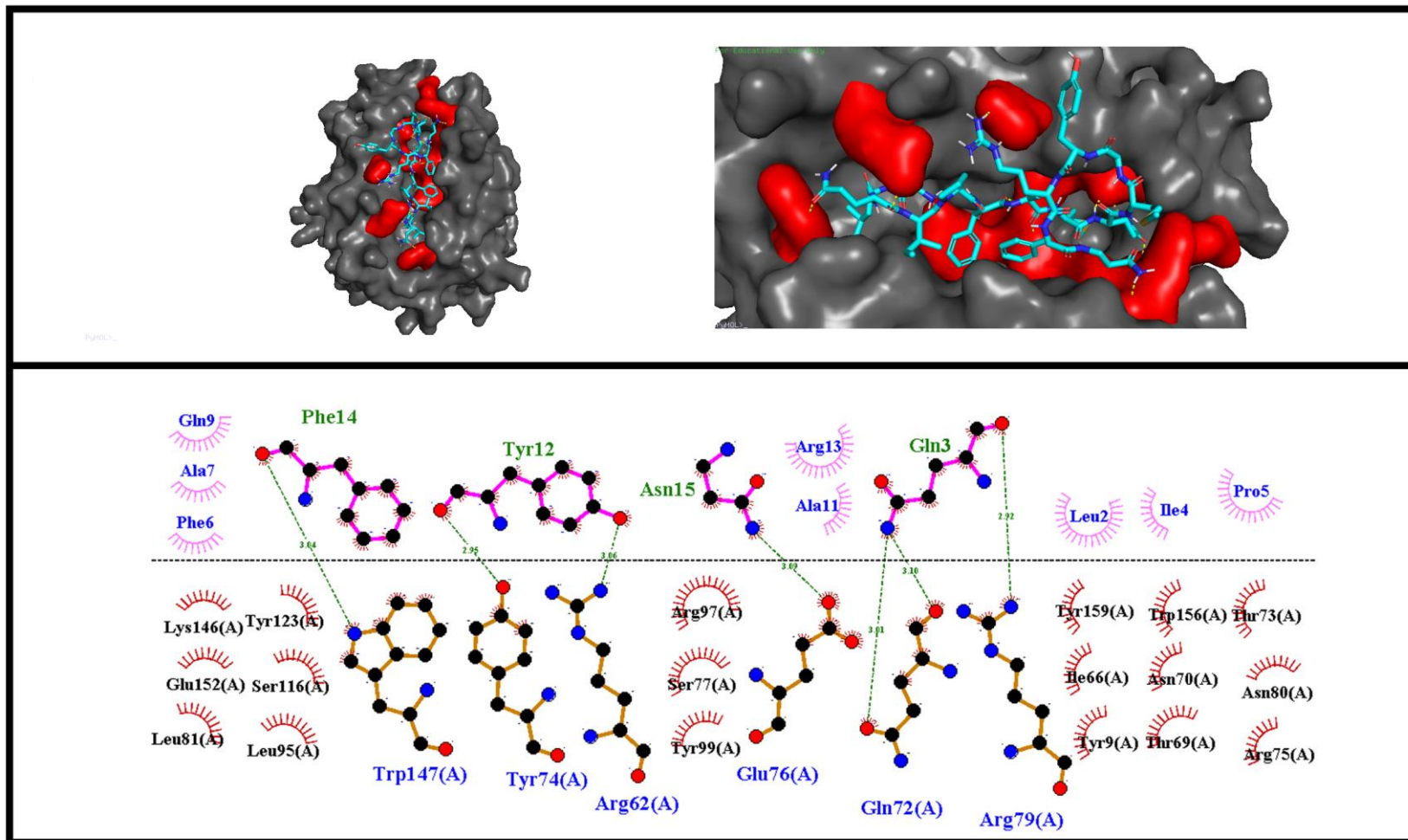
Figura 11 – Representação do diagrama 3D e 2D da interação do Epítipo 1 com o receptor MHC de classe II, alelo HLA-DRBI*04:01.



Fonte: Da autora.

Para o epítipo 2 foram visualizadas interações com 9 pontes de hidrogênio, 29 interações hidrofóbicas e nenhuma interação com as pontes de sal para o epítipo 2 no alelo HLA-B*3501 de MCH-I (figura 12).

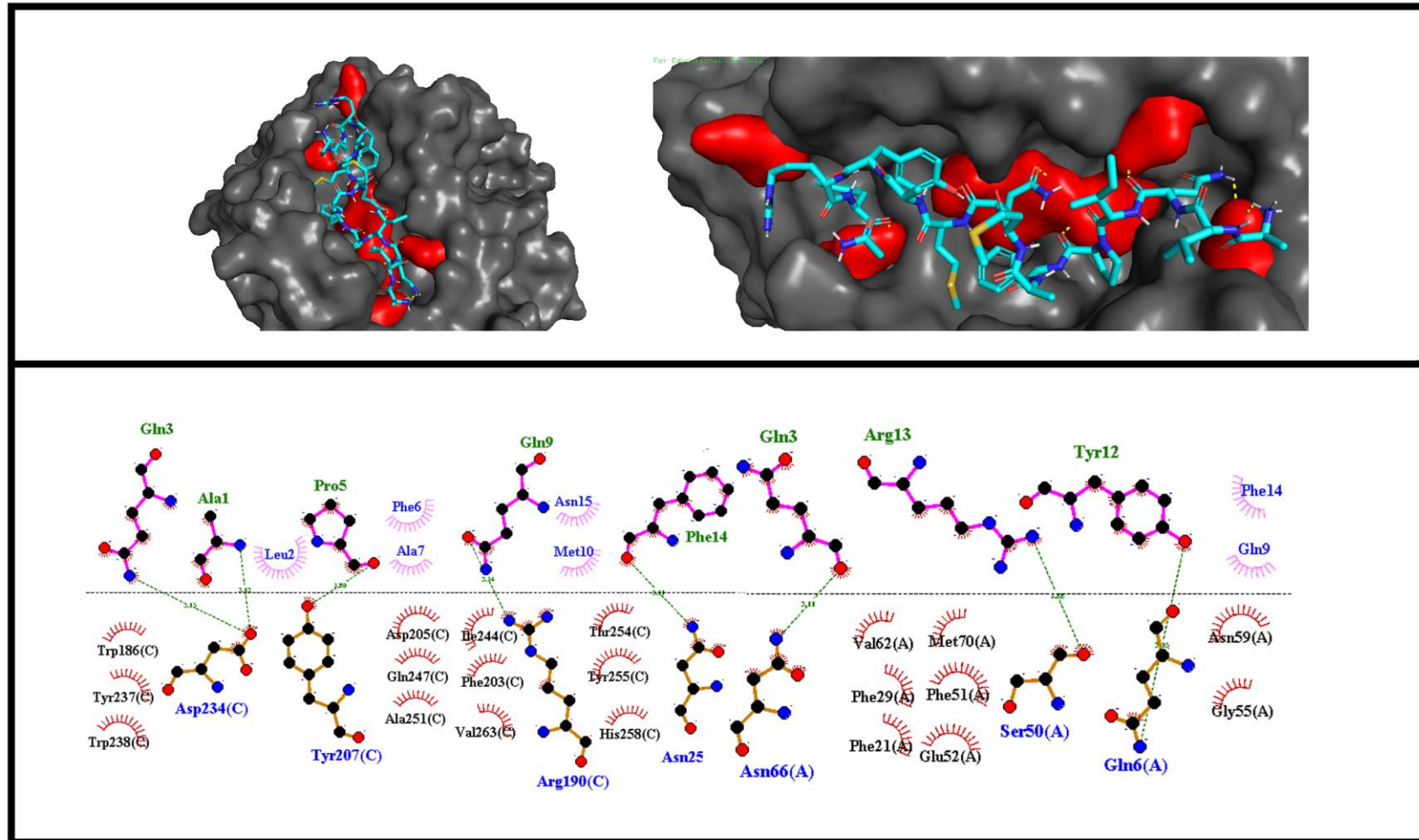
Figura 12 – Representação do diagrama 3D e 2D da interação do Epítipo 2 com o receptor MHC de classe I, alelo HLA-B*3501.



Fonte: Da autora.

No alelo HLA-DRBI*15:01 de MHC-II (figura 13) para interação com o mesmo epítipo, foram visualizadas 8 pontes de hidrogênio, 27 interações hidrofóbicas e também nenhuma interação com as pontes de sal.

Figura 13 – Representação do diagrama 3D e 2D da interação do Epítipo 2 com o receptor MHC de classe II, alelo HLA-BI*15:01.



Fonte: Da autora.

7.1.8 Ancoragem molecular dos epítomos selecionados apresentam ligação as moléculas de H-2 de *Mus musculus* similares a interações de MHC de *Homo sapiens*

Após a realização da ancoragem molecular com os alelos de *Homo sapiens*, realizou-se o mesmo procedimento para as moléculas de MHC do organismo *Mus musculus*. Foi realizado uma busca no banco de dados PDB pelas proteínas cristalografadas de H-2, no qual, encontrou-se a proteína 1NEZ para H-2 de classe I com resolução de 2,10 Å e 1FNG para H-2 de classe II com resolução de 1,90Å, respectivamente. As proteínas foram preparadas e colocadas em execução contra os epítomos, no qual os resultados estão descritos na tabela abaixo.

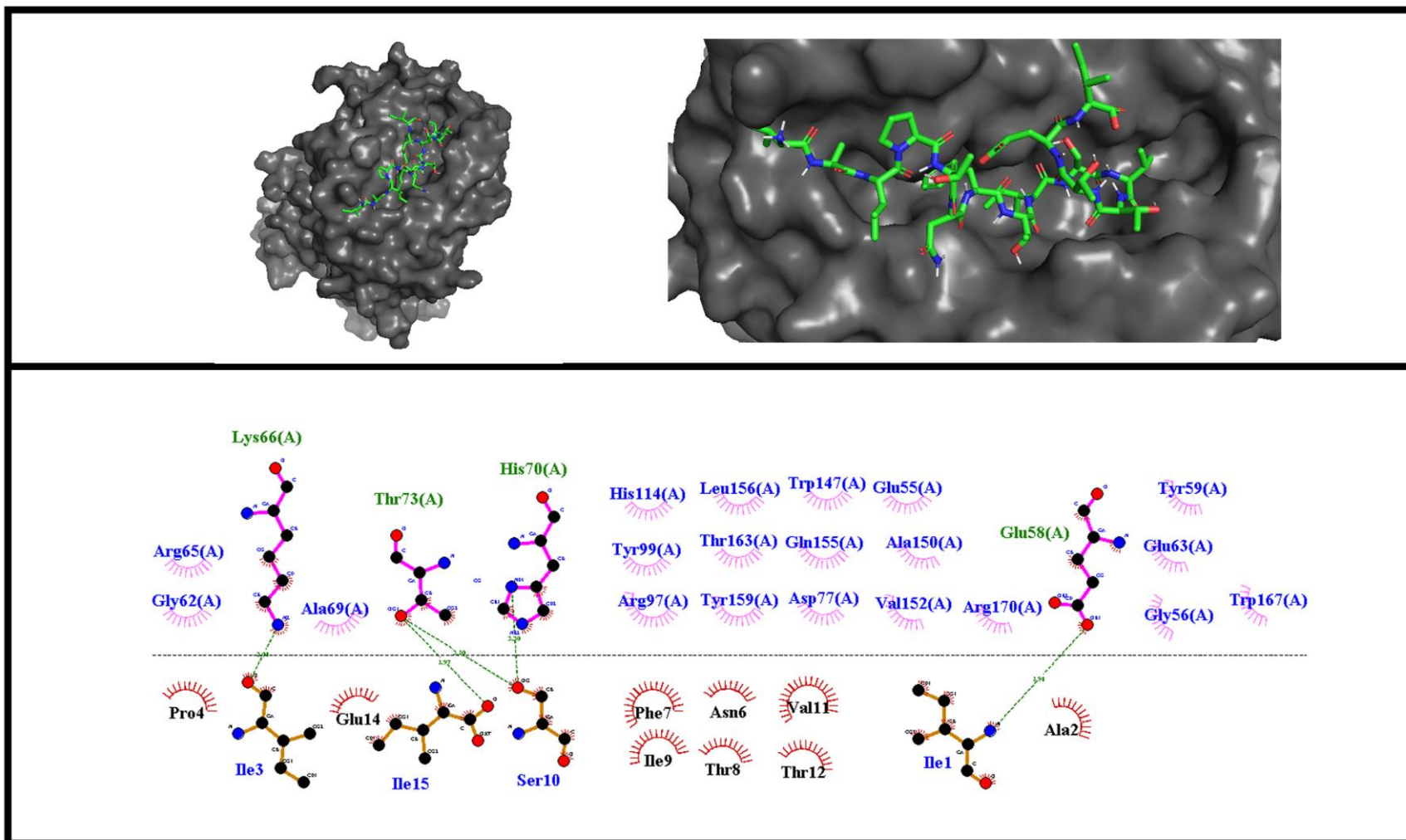
Tabela 3 – Interações dos alelos de H-2 classe I e H-2 classe II dos epítomos selecionados.

Epítomo	Proteína	AutoDock Vina
1	H-2 classe I	-7,5
	H-2 classe II	-6,9
2	H-2 classe I	-8,0
	H-2 classe II	7,6

Fonte: Da autora.

A interação entre o epítomo 2 e o complexo de histocompatibilidade H-2 de classe I, apresentou 5 pontes de hidrogênio e 29 interações hidrofóbicas (figura 14).

Figura 14 – Representação do diagrama 3D e 2D da interação do Epítipo 2 com o receptor de histocompatibilidade H-2 classe I.



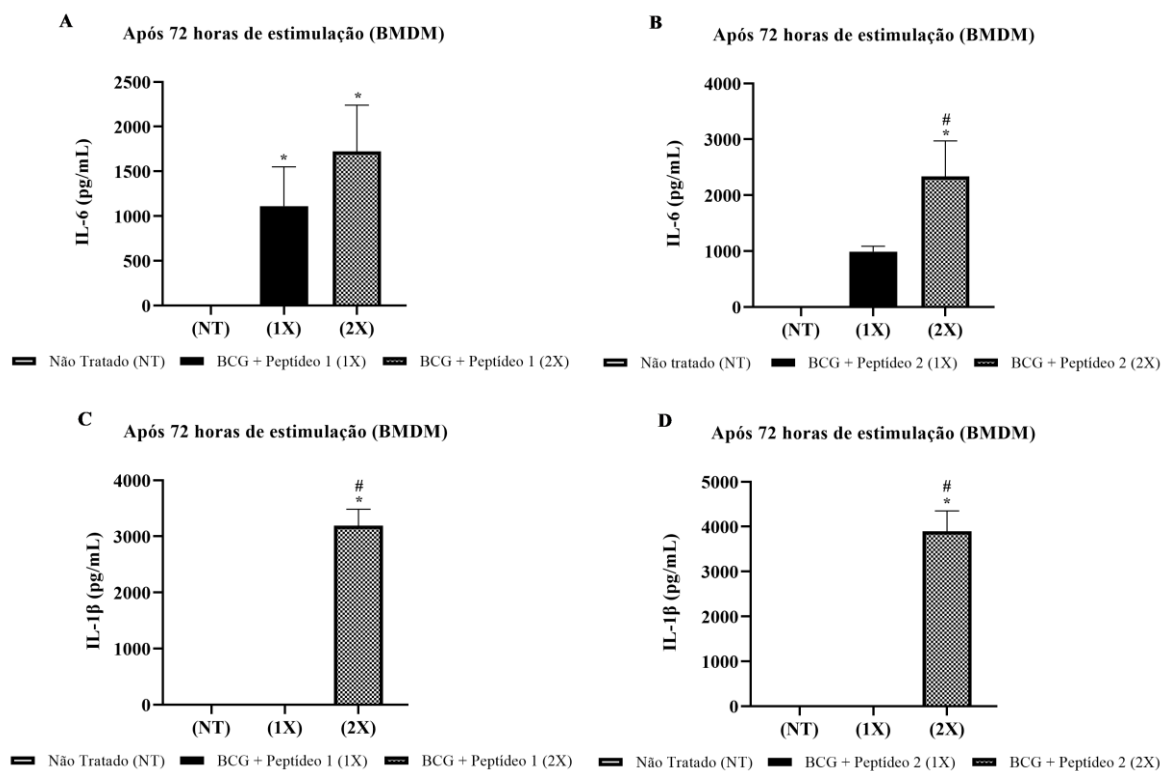
Fonte: Da autora.

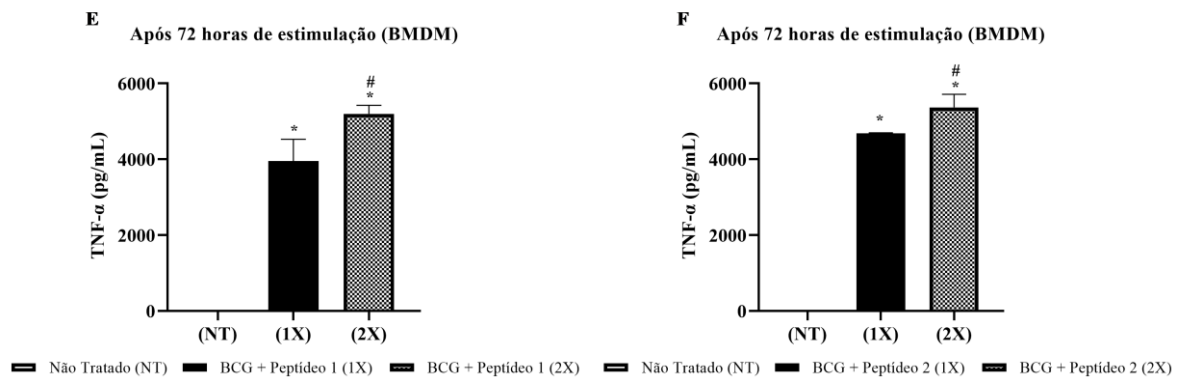
7.2 IN VITRO

7.2.1 BMDMs estimuladas com BCG em combinação com os peptídeos da proteína *spike* apresentaram maior aumento na secreção de IL-6, IL-1 β , TNF- α quando reestimuladas

As células BMDMs quando estimulados com BCG em combinação com peptídeo 1 e BCG em combinação com peptídeo 2 respectivamente, em um primeiro momento, aumentaram consideravelmente a resposta dos alvos analisados de citocinas da imunidade inata IL-6 (figura 15A) e TNF- α (figura 15E) em exceção da citocina IL-1 β (figura 15C) em relação ao grupo não tratado. Quando reestimuladas, houve uma maior concentração e ativação expressiva na secreção dessas citocinas pró-inflamatórias IL-6, IL-1 β e TNF- α (figura 15B, 15D e 15F) detectadas por ELISA, em contato com BCG e as formulações desenvolvidas.

Figura 15 – A reestimulação em células BMDMs aumentou a expressão de citocinas pró-inflamatórias IL-6 (A-B), IL-1 β (C-D) e TNF- α (E-F) em contato com BCG e as formulações desenvolvidas.





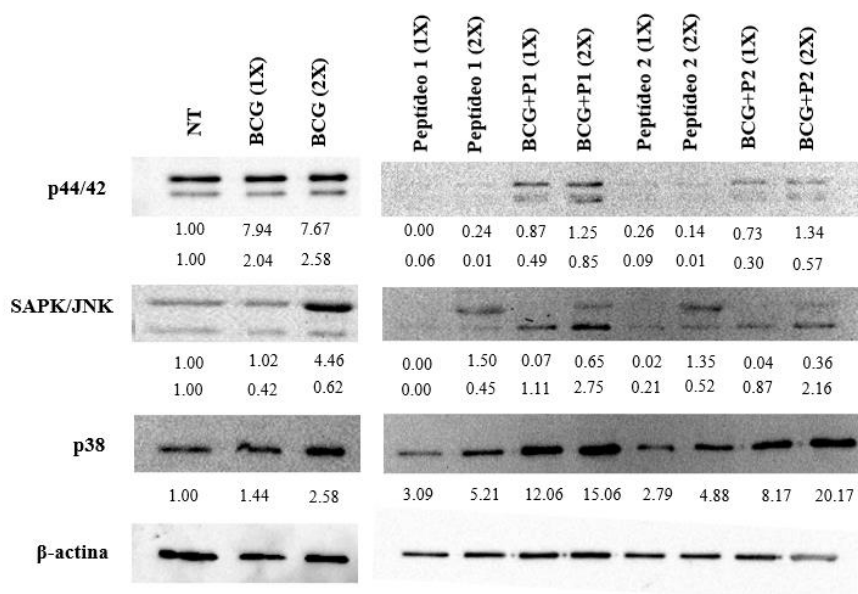
Fonte: Da autora.

Legenda: Amostras testadas em triplicata (n=3). * P<0,05 em relação ao grupo não tratado. # P<0,05 em relação ao grupo tratado BCG + Peptídeo (1X).

7.2.2 A via de sinalização MAPK foi ativada e intensificada em amostras reestimuladas com a formulação contendo os peptídeos associados ao BCG vacinal

A via de sinalização MAPK foi analisada pelo processo de fosforilação das proteínas p44/42, SAPK/JNK e p38. Os resultados apontaram que quando reestimuladas (figura 16), estas intensificaram a via da MAPK, dando ênfase ao p38 que frente a esse reestímulo demonstrou a via mais ativada.

Figura 16 – As vias de sinalização de MAPK foram ativadas pelas formulações desenvolvidas do vírus SARS-CoV-2 em células BMDMs.



Fonte: Da autora.

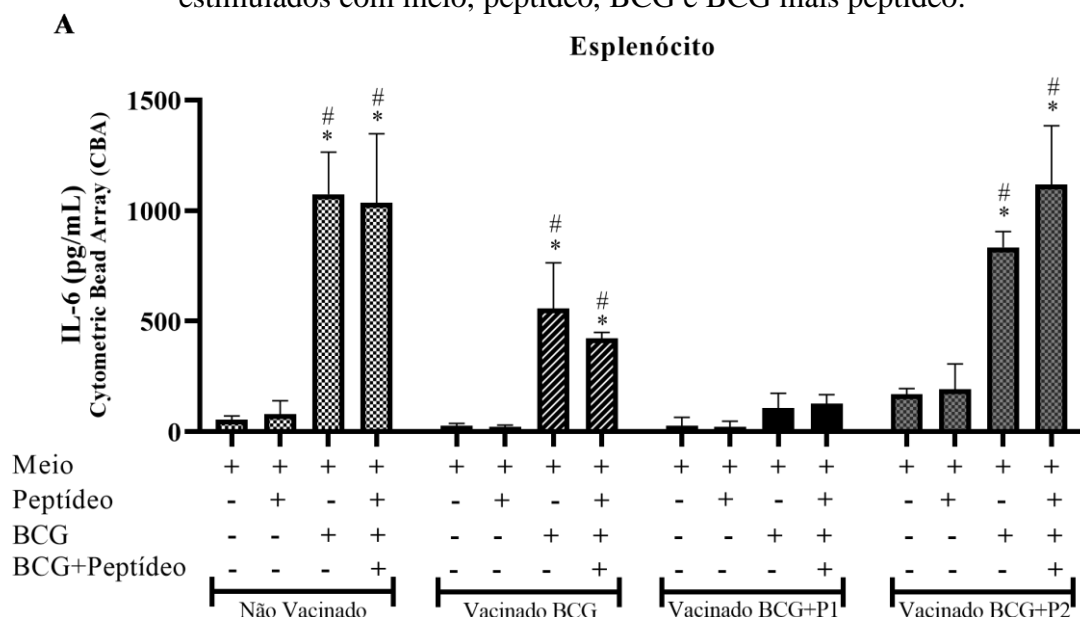
Legenda: β-actina foi usada como controle. Valores abaixo das bandas indicam a quantificação da intensidade em relação ao carregamento.

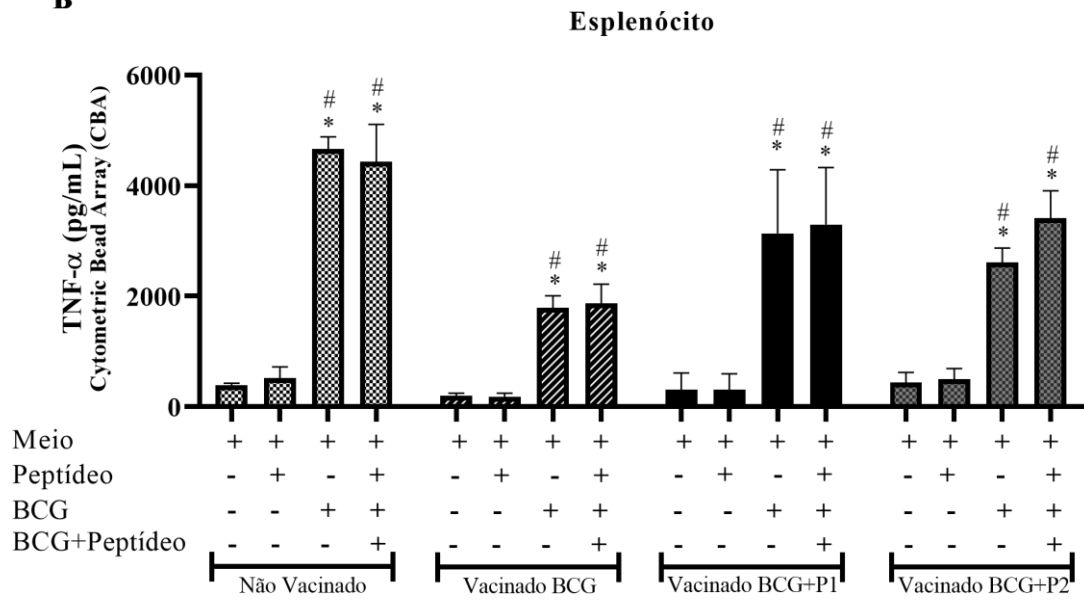
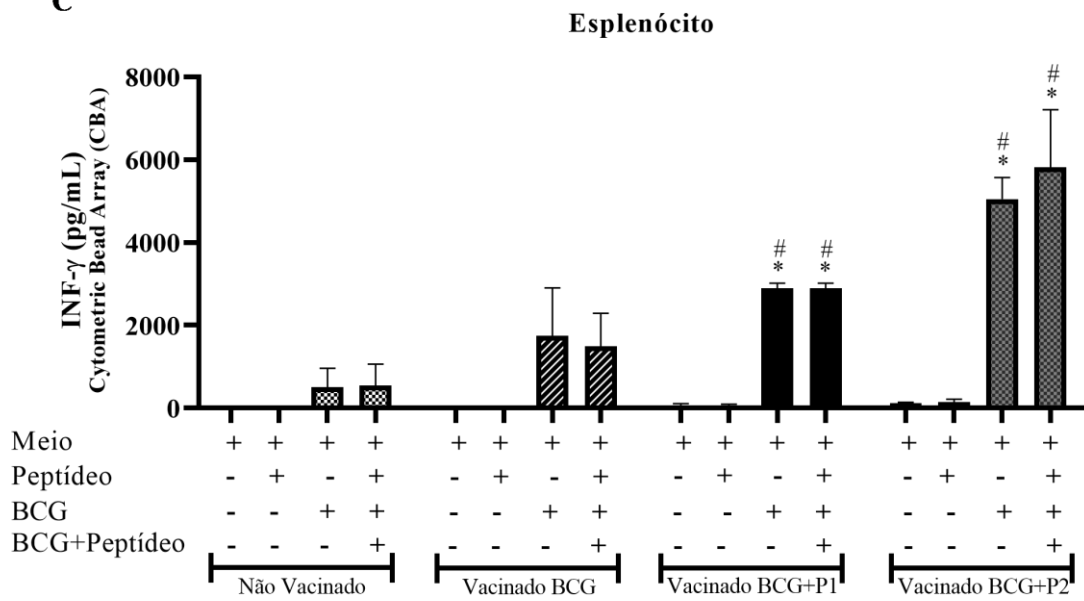
7.3 *IN VIVO*

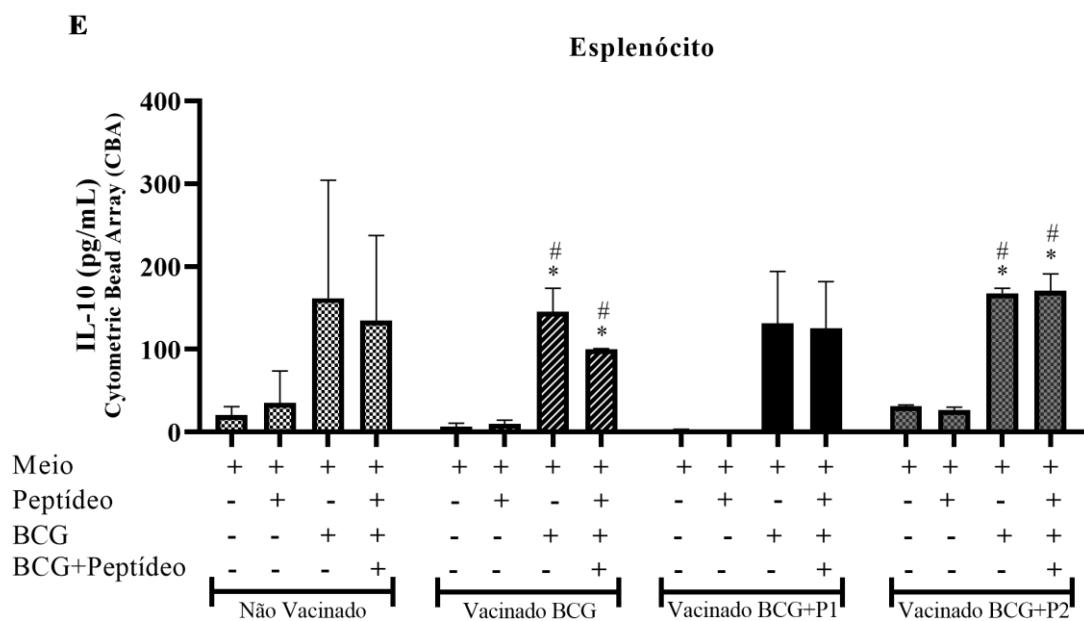
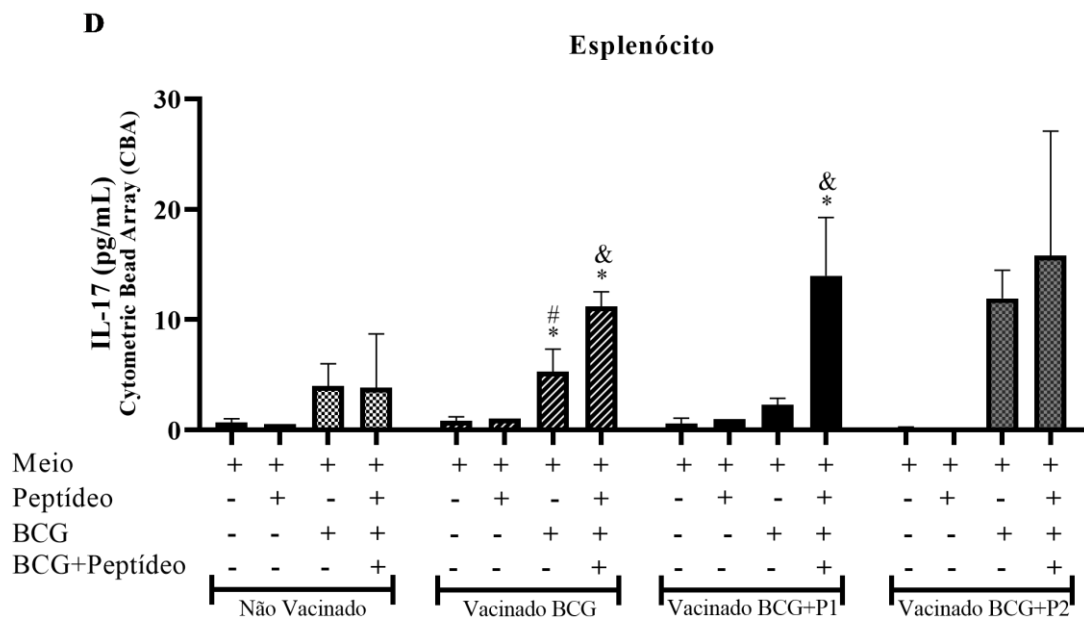
7.3.1 Animais vacinados com os peptídeos associados ao BCG vacinal apresentaram aumento da secreção de IL-6, TNF- α , INF- γ em cultura de esplenócitos e em circulação

Nos tecidos esplênicos de animais vacinados com os peptídeos em associação com o BCG vacinal e reestimulados, observou-se o aumento da secreção de citocinas pró-inflamatórias como IL-6 (figura 17A), TNF- α (figura 17B), INF- γ (figura 17C). Não foram detectados níveis de produção de IL-4 nas amostras (dado não apresentado). De maneira inesperada, foi observado uma diminuição da secreção de IL-6 nos esplenócitos reestimulados. Contudo, os esplenócitos reestimulados se destacaram por apresentarem um aumento significativo da secreção de INF- γ , sugestivo que há uma indução de resposta imune do perfil Th1. Há um aumento de IL-17 da associação principalmente de BCG com os peptídeos, mas, a quantidade de secreção dessa citocina se apresenta muito baixa. Uma vez observado que há uma intensa secreção de citocinas pró-inflamatórias, foi avaliada a secreção de IL-10. É possível observar que esplenócitos reestimulados apresentam certa secreção dessa citocina anti-inflamatória sugestivo de indução da tentativa de um controle da resposta pró-inflamatória mediada pela IL-6, TNF- α e INF- γ .

Figura 17 – Análise de expressão de IL-6 (A), TNF- α (B), INF- γ (C), IL-17 (D) e IL-10 (E) em cultura de esplenócitos de grupos experimentais de camundongos C57/BL6 imunizados com BCG e os respectivos peptídeos da proteína *spike* e estimulados com meio, peptídeo, BCG e BCG mais peptídeo.



B**C**



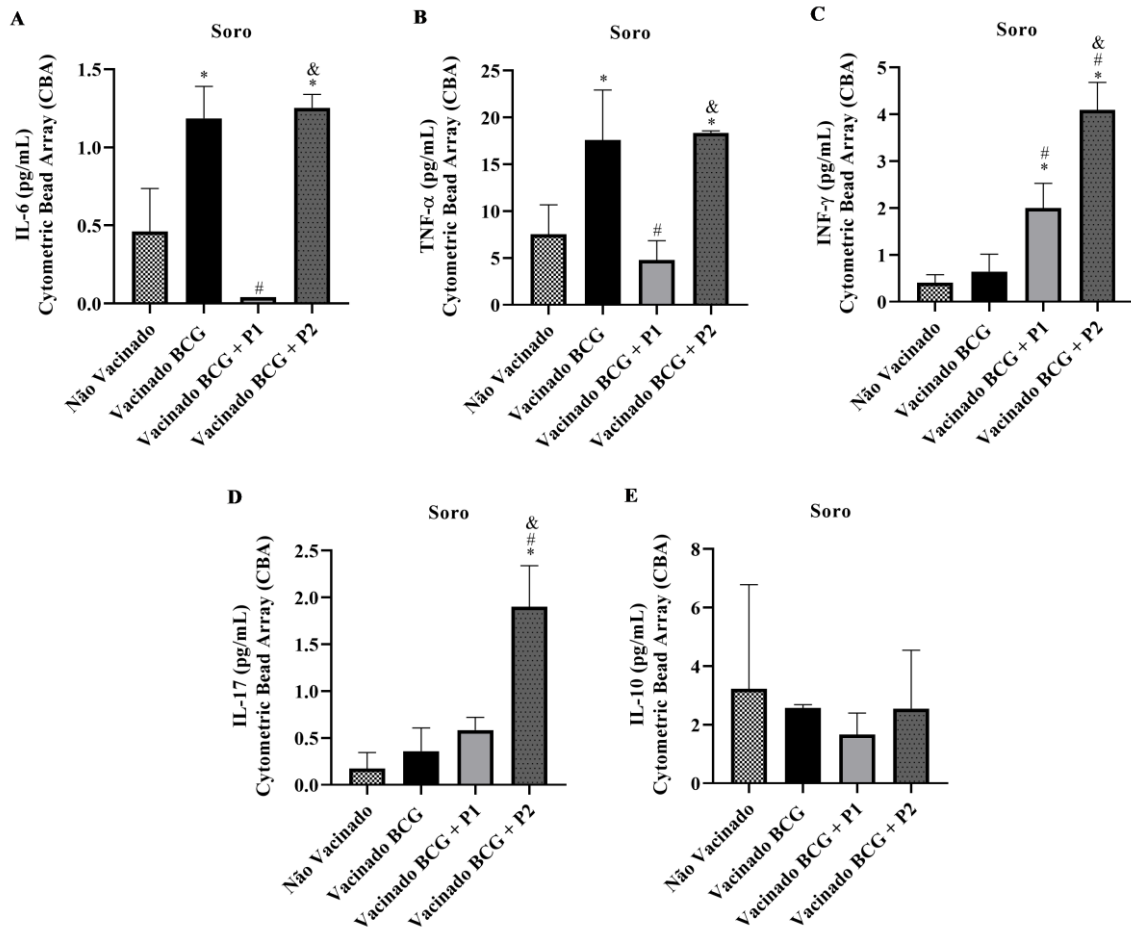
Fonte: Da autora.

Legenda: Amostras testadas em triplicata (n=3). * P<0,05 em relação ao estímulo meio. # P<0,05 em relação ao estímulo peptídeo. & P<0,05 em relação ao estímulo BCG.

Para identificar se as citocinas secretadas pelo tecido esplênico também estavam em circulação, os animais vacinados tiveram o soro coletado e avaliados quanto a secreção das mesmas citocinas. É possível identificar o aumento da secreção de citocinas pró-inflamatórias como IL-6 (figura 18A), no soro de animais vacinados com BCG e vacinados com BCG+P2 e uma diminuição dessa citocina no grupo vacinado com BCG+P1, aumento significativo de TNF- α (figura 18B) no grupo vacinado com BCG e vacinado com BCG+P2 e aumento de

INF- γ (figura 18C) e IL-17, sendo mais proeminente no grupo vacinado com BCG+P2. Esses resultados corroboram com os achados nos tecidos esplênicos.

Figura 18 – Análise de expressão de IL-6 (A), TNF- α (B), INF- γ (C), IL-17 (D) e IL-10 (E) em soro de grupos experimentais de camundongos C57/BL6 imunizados com BCG e BCG com os respectivos peptídeos da proteína *spike*.



Fonte: Da autora.

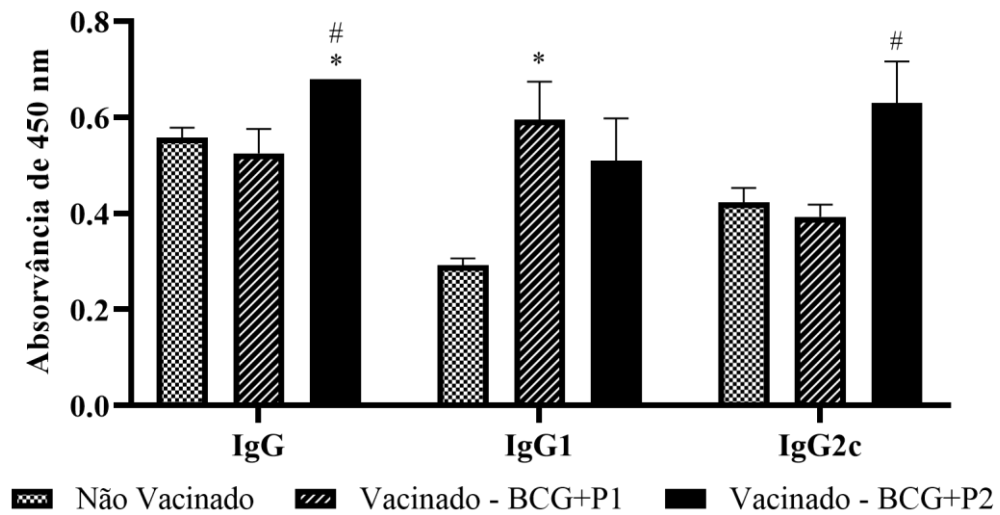
Legenda: Amostras testadas em triplicata (n=3). * P<0,05 em relação ao grupo não vacinado. # P<0,05 em relação ao grupo vacinado com BCG. & P<0,05 em relação ao grupo vacinado com BCG + Peptídeo 1.

7.3.2 Aumento de anticorpos específicos às proteínas estruturais do SARS-COV-2 foram detectados no soro de animais imunizados

Foi quantificado por ELISA o potencial dos peptídeos em relação de induzir uma resposta imune humoral. O peptídeo 1 não apresentou uma boa resposta imune humoral, mas em contrapartida o peptídeo 2 apresentou uma boa produção de IgG e suas subclasses (figura 19) concluindo que IgG, IgG1 e IgG2c estão acima do limite em relação dos animais não

vacinados.

Figura 19 – Anticorpos específicos IgG, IgG1, IgG2c no soro de animais não vacinados e animais imunizados.



Fonte: Da autora.

Legenda: * $P < 0,05$ em relação ao grupo não vacinado. # $P < 0,05$ em relação ao grupo vacinado com BCG+P1. Amostras testadas em duplicata (n=2).

8 DISCUSSÃO

O SARS-CoV-2 é caracterizado por ser um vírus altamente transmissível e sua rápida disseminação fez com que medidas protetivas fossem necessárias bem como o desenvolvimento e implantação a nível global de uma vacina (LU, 2020). As tentativas atuais da pesquisa têm se concentrado na busca de candidatos vacinais que sejam eficazes utilizando diferentes ferramentas de bioinformática. Nesse contexto, a vacinologia reversa passou a ser uma metodologia de desenvolvimento de vacinas que são potencialmente imunogênicas e seguras.

Várias linhas de pesquisa foram desenvolvidas em prol de se obter a vacina do SARS-CoV-2 por métodos de bioinformáticas. Baruah e Bose, 2020 por exemplo, desenvolveram a pesquisa sobre a proteína *spike* para descobrir epítomos de células B e linfócitos T citotóxicos por meio de ferramentas de imunoinformática. Além disso, Dong *et al.* 2020, também projetaram uma vacina de multiepítomo para prevenção do COVID-19. PORAN *et al.* 2020, com base em multi-proteína localizaram epítomos para células T CD4⁺ e CD8⁺.

Nesse estudo, foram utilizadas abordagens *in silico* para obter um epítomo da proteína estrutural *spike* do SARS-CoV-2 para que fosse identificado um peptídeo imunogênico e assim, fosse desenvolvido em uma vacina segura a base de peptídeos. As técnicas tradicionais de desenvolvimento de vacinas, apresentam algumas desvantagens que podem ser superadas pelas técnicas *in silico*. A técnica *in silico* pode diminuir o tempo e o custo para o desenvolvimento de uma vacina.

Foi previsto epítomos da proteína *spike* com afinidade de ligação com células B, MHC-I e MHC-II. Em seguida, foram alinhados com intuito de obter epítomos imunogênicos que desenvolvessem a ação em ambos, não necessariamente significa que o epítomo final é de todas os resultados encontrados nessas células, ou seja, os epítomos finais significam que apresentaram maior taxa de repetição de todos os epítomos de células B, MHC-I e MHC-II.

A estratégia de se obter uma sequência consenso, auxilia na observação e diminuição de variantes e aumenta a possibilidade de definir sequências de aminoácidos mais conservados. Além disso, as técnicas de modelagem molecular e inteligência artificial de dados possibilitaram na definição de estrutura 3D da proteína consenso e identificação de pontos dos epítomos selecionados dentro da proteína estrutural.

As técnicas racionais *in silico* podem prever o melhor epítomo para interagir com receptores de células B gerando uma resposta adaptativa e humoral específica, além de prever

o melhor epítopo que possa interagir com MHC-I ativando linfócitos T CD8⁺ e prever o melhor epítopo que interaja com alelos de MHC-II ativando linfócitos T CD4⁺, promovendo assim uma resposta adaptativa celular específica.

As vacinas peptídicas são vacinas composta por pequenas regiões de proteínas com componentes imunogênicos e podem ser manejadas de duas formas: através do pool de peptídeos e através de substituição da proteína. Esse pool de peptídeos não é encontrado com frequência em regiões de mutações, enquanto que a substituição da proteína pode ser feita em regiões de mutação. Os resultados obtidos *in silico* nesse trabalho, foram obtidos em um curto período de tempo, com baixo custo onde foi definido epítomos prováveis de serem imunogênicos.

A escolha de epítomos de tamanho 15 foi para que eles encaixassem tanto na fenda de MHC-I e MHC-II que ao ser mostrado, seja mostrado não somente aquilo que está interagindo com a fenda, mas também a fenda que não está interagindo. Regiões de N-glicosilação foram detectadas no epítopo 1 e essas reações estão em sua superfície que é aonde se conectam com o ACE 2. O fato de ter N-glicosilações nessa região indicam que o mecanismo de escape pode mudar, pois qualquer variação e inserção de região de N-glicosilação fazem com que o vírus escape do sistema imune.

Foi detectada a mutação T716I no epítopo 1 que fica dentro da subunidade S2 da linhagem B.1.1.7 da proteína *spike*. As evidências mostraram que essa variante B.1.1.7 é mais virulenta (COLLIER *et al.*, 2021; LEUNG *et al.*, 2021). Na literatura não foi encontrado ferramentas que façam a predição ou a previsão de qual é a N-glicosilação que ocorre no aminoácido 717 que é o que está presente dentro do epítopo 1. Porém, Watanabe *et al.* 2020, explicam *in vitro* a identificação de qual era a N-glicosilação no mesmo aminoácido 717. Como resultado, identificaram que a N-glicosilação presente nesse aminoácido se tratava de 75% aproximadamente de glicanos (oligomanose) e uma mistura de híbrido com complexo. Dessa forma, é sugestivo que a glicosilação encontrada no aminoácido 717 do epítopo 1 se trata de uma oglicosilação de várias manoses presente no vírus Alpha.

Nas análises *in vitro* os macrófagos derivados da medula óssea têm perfis de ação e em paralelo aos resultados obtidos, mostraram significativamente o treinamento de células imunes (VIOLA *et al.*, 2019). Estudos epidemiológicos mostram que a vacina de BCG tem

uma proteção por outras infecções virais e que o rearranjo da cromatina induz o treinamento da célula quando exposta a um patógeno inespecífico e com isso altera o acesso para expressão gênica de citocinas pró-inflamatórias que são aumentadas quando reestimuladas (COVIÁN *et al.*, 2019).

O'Neill e Netea 2020, descreveram que a vacinação de BCG em humanos saudáveis induz no aumento de produção de citocinas como IL-6, IL-1 β e TNF- α quando em ensaios *in vitro* os monócitos são estimulados com patógenos inespecíficos. Com o conhecimento de mecanismos envolvidos na ativação da imunidade inata é possível compreender o que a vacina BCG é capaz de induzir afim de gerar uma resposta protetora.

Curiosamente, Covián *et al.* 2020, também descreveram que os monócitos treinados com a vacinação de BCG têm uma rápida resposta e com isso secretam IL-6, IL-1 β e TNF- α após a estimulação com *S. aureus* e *C. albicans* e isso pode estar relacionado com as alterações epigenéticas e com os genes reguladores dessas citocinas pró-inflamatórias.

Nesse contexto, há uma correlação de citocinas inatas e estas medeiam um papel crucial de respostas imunes treinadas como por exemplo IL-1 β , visto que esta citocina desempenha uma imunidade treinada devido as modificações epigenéticas de metilação de histonas e pode ser fundamentada por reguladores intrínsecos nas células imunes progenitores da medula óssea estimuladas por BCG (ARTS *et al.*, 2018; COVIÁN *et al.*, 2020).

As vias de sinalização das MAPK são fundamentais e tem um papel sobre as células imunes e regulam processos celulares (GOEL *et al.*, 2021). Nesse contexto, Kumar *et al.*, descreveram que a MAPK é essencial via de sinalização celular por ser conhecida e ativada por uma diversidade de vírus. Assim, a ativação dessa via intracelular é relevante para a produção de citocinas frente a uma infecção por SARS-CoV-2 (BATTAGELLO *et al.*, 2020). Concomitantemente, essa via propicia que haja separação celular e que esses sejam alterados em estímulos extracelulares que geram respostas celulares (CARGNELLO; ROUX, 2011; KUMAR *et al.*, 2018).

Dessa forma, as citocinas pró-inflamatórias como IL-6, TNF- α e IL-1 β também são ativadas uma vez que se tem a expressão de p38 frente a uma regulação positiva da via MAPK (ZARUBIN; HAN, 2005; GRIMES; GRIMES, 2020). Já foi descrito que células epiteliais brônquicas humanas primárias infectadas com SARS-CoV-2 e analisadas a partir de fatores de transcrição e expressão gênica que são derivadas dessas células, p38 que regulam esses fatores de transcrição é intensamente mais ativado após a infecção (BLANCO-MELO *et al.*, 2020; GOEL *et al.*, 2021). De fato, os resultados obtidos desse estudo, reforçam que p38 foi altamente ativado mostrando uma mediação de resposta celular em relação aos estímulos

propostos em sobrenadantes de células BMDMs culminando no treinamento de resposta.

Níveis de IL-6, TNF- α , INF- γ e IL-17 foram produzidos pelas células do baço de camundongos imunizados e posteriormente estimulados, sugerindo uma indução de resposta imune do perfil Th1 pelos peptídeos vacinais. Além disso, a vacina com BCG e os peptídeos aumentaram a resistência de resposta de perfil de linfócitos T CD4⁺, sugerindo uma resposta imune efetora após estimulação. A especificidade das células T sobre a proteína não só desse estudo, mas qualquer proteína viral é proeminente para uma imunidade protetora (GRIFONI *et al.*, 2020; SEKINE *et al.*, 2020; SETTE; CROTTY, 2021).

Os peptídeos vacinais da proteína *spike* estimulam o perfil de resposta e faz com que apresentem uma atividade antiviral sobre essas células Th1 mediada pela produção de INF- γ e demais citocinas (SETTE; CROTTY, 2021). Vale ressaltar que esse INF- γ observado nas análises tem um papel muito interessante para atuar sobre os linfócitos B, pois quando são ativados se transformam em plasmócitos e secretam anticorpos. Níveis de IL-10 também foram encontrados nas células do baço de camundongos imunizados. Esses resultados colaboram com esse cenário aonde se tem citocinas pró-inflamatórias e anti-inflamatórias, sugestivo que já tenha ocorrido uma resposta inflamatória a ponto de gerar uma resposta anti-inflamatória com característica de controle, ou seja, com características de resolução do problema. Já foi demonstrado que nos ensaios clínicos com as vacinas para o SARS-CoV-2 tem como alvo esse perfil encontrado nesse estudo, sendo esse muito proeminente e não outro perfil no qual não apresentou aumento da citocina por exemplo IL-4 de perfil Th2 (KRAMMER, 2020; SAHIN *et al.*, 2020).

Nesse contexto, Xavier e colaboradores 2013, pontuam que esse equilíbrio entre produção de citocinas pró e anti-inflamatórias parece ser crucial para a capacidade do hospedeiro de erradicar a infecção que é exatamente o que se viu nesse estudo quando avaliamos esses níveis de citocinas no tecido esplênico, tendo esse equilíbrio entre as citocinas pró-inflamatórias e a IL-10 estando regulada para equilibrar essa expressão das citocinas pró-inflamatórias.

Fernandez *et al.* 2022, identificaram um conjunto de peptídeos de células T CD4⁺ e CD8⁺ promíscuos que foram reconhecidos por indivíduos convalescentes. Essa resposta de célula T, foi analisada *in vitro* em células produtoras de INF- γ a peptídeos específicos do SARS-CoV-2. De fato, a eficácia vacinal depende da ativação dos tipos celulares responsáveis pelo controle da infecção. As células T CD4⁺ tem como função regular o sistema imune. A busca para o desenvolvimento de vacinas para SARS-CoV-2 é que as células T produzem INF- γ , pois essa citocina auxilia no controle de patógenos intracelulares. Corroborando aos

resultados encontrados nesse estudo, os epítomos selecionados podem ser alvos potenciais para avaliação da imunidade uma vez que a imunização em conjunto com BCG reforça a ideia de proteção associado as células T de memória.

Níveis de anticorpos IgG, IgG1 e IgG2c foram vistos no soro de camundongos imunizados. Já foi demonstrado que no sangue de camundongos vacinados com BCG e a combinação de uma forma trimérica do antígeno de pico do SARS-CoV-2 potencializou no aumento de anticorpos IgG específicos (COUNOUPAS *et al.*, 2021) e isso pode estar relacionado a uma resposta humoral específica do tipo Th1 em particular tendo papel fundamental a infecções por vírus (VISCIANO *et al.*, 2012; DEL FRESNO *et al.*, 2021)

9 CONCLUSÃO

A abordagem *in silico* foi promissora no desenvolvimento de candidatos vacinais potencialmente imunogênicos contra o vírus SARS-CoV-2.

Células da imunidade inata foram capazes de ser treinadas com os peptídeos em combinação com BCG desencadeando em uma resposta eficiente e segura.

Portanto, a imunização demonstrou uma resposta sobre esses antígenos, sendo esses epítomos imunoestimulatórios plausíveis de serem candidatos vacinais no desenvolvimento de vacinas utilizando a vacinologia reversa a fim de candidatos que sejam seguros e eficazes na proteção contra o vírus SARS-CoV-2.

10 REFERÊNCIAS

- ABRAHAM PEELE, K. *et al.* Design of multi-epitope vaccine candidate against SARS-CoV-2: a in-silico study. **Journal of Biomolecular Structure and Dynamics**, v. 39, n. 10, p. 3793-3801, 2020.
- ABU-RADDAD, L. J.; CHEMAITELLY, H.; BUTT, A. A. Effectiveness of the BNT162b2 Covid-19 Vaccine against the B. 1.1. 7 and B. 1.351 Variants. **New England Journal of Medicine**, v. 385, n. 2, p. 187-189, 2021.
- ALEEM, A., AKBAR SAMAD, A. B., SLENKER, A. K. Emerging Variants of SARS-CoV-2 And Novel Therapeutics Against Coronavirus (COVID-19). In StatPearls. StatPearls Publishing. 2022.
- ARTS, R. J. W. *et al.* BCG Vaccination Protects against Experimental Viral Infection in Humans through the Induction of Cytokines Associated with Trained Immunity. **Cell Host & Microbe**, v. 23, n. 1, p. 89- 100. e5, 2018.
- ASPATWAR, A. *et al.* Tuberculosis vaccine BCG: the magical effect of the old vaccine in the fight against the COVID-19 pandemic. **International Reviews of Immunology**, v. 41, n. 2, p. 283-296, 2021.
- BARUAH, V.; BOSE, S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. **Journal of medical virology**, v. 92, n. 5, p. 495-500, 2020.
- BATTAGELLO, D. S. *et al.* Unpuzzling COVID-19: Tissue-related signaling pathways associated with SARS-CoV-2 infection and transmission. **Clinical Science**, v. 134, n. 16, p. 2137–2160, 2020.
- BLANCO-MELO, D. *et al.* Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. **Cell**, v. 181, n. 5, p. 1036- 1045. e9, 2020.
- BUSZKO, M. *et al.* The dynamic changes in cytokine responses in COVID-19: a snapshot of the current state of knowledge. **Nature Immunology**, v. 21, n. 10, p. 1146–1151, 2020.
- CARGNELLO, M.; ROUX, P. P. Activation and Function of the MAPKs and Their Substrates, the MAPK-Activated Protein Kinases. **Microbiology and Molecular Biology Reviews**, v. 75, n. 1, p. 50–83, 2011.
- CHEN, Y.; LIU, Q.; GUO, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. **Journal of Medical Virology**, v. 92, n. 4, p. 418–423, 2020.
- CHENG, Mary Hongying *et al.* An insertion unique to SARS-CoV-2 exhibits superantigenic character strengthened by recent mutations. **Biorxiv**, 2020.
- COLLIER, D. A. *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. **Nature**, v. 593, n. 7857, p. 136-141, 2021.

COSTELA-RUIZ, V. J. *et al.* SARS-CoV-2 infection: The role of cytokines in COVID-19 disease. **Cytokine & growth factor reviews**, v. 54, p. 62-75, 2020.

COUNOUPAS, C. *et al.* A single dose, BCG-adjuvanted COVID-19 vaccine provides sterilising immunity against SARS-CoV-2 infection. **NPJ Vaccines**, v. 6, n. 1, p.1-10, 2021.

COVIÁN, C. *et al.* BCG-Induced Cross-Protection and Development of Trained Immunity: Implication for Vaccine Design. **Frontiers in Immunology**, v. 10, p. 2806, 2019.

COVIÁN, C. *et al.* Could BCG Vaccination Induce Protective Trained Immunity for SARS-CoV-2? **Frontiers in Immunology**, v. 11, n. May, p. 970, 2020.

CUI, J.; LI, F.; SHI, Z. L. Origin and evolution of pathogenic coronaviruses. **Nature Reviews Microbiology**, v. 17, n. 3, p. 181–192, 2019.

DAGAN, N. *et al.* BNT162b2 Mrna Covid-19 Vaccine in a Nationwide Mass Vaccination Setting. **New England Journal of Medicine**, v. 384, n. 15, p. 1412–1423, 2021.

DALSASS, M. *et al.* Comparison of open-source reverse vaccinology programs for bacterial vaccine antigen discovery. **Frontiers in Immunology**, v. 10, p. 113, 2019.

DAVIES, N. G. *et al.* Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. **Nature**, v. 593, n. 7858, p. 270–274, 2021.

DEL FRESNO, C. *et al.* The Bacterial Mucosal Immunotherapy MV130 Protects Against SARS-CoV-2 Infection and Improves COVID-19 Vaccines Immunogenicity. **Frontiers in Immunology**, v. 12, 2021.

DESCHLER, S. *et al.* Mucosal-associated invariant T (MAIT) cells are highly activated and functionally impaired in COVID-19 patients. **Viruses**, v. 13, n. 2, p. 241, 2021.

DIVANGAHI, M. *et al.* Trained immunity, tolerance, priming and differentiation: distinct immunological processes. **Nature Immunology**, v. 22, n. 1, p. 2–6, 2021.

DONG, R. *et al.* Contriving Multi-Epitope Subunit of Vaccine for COVID-19: Immunoinformatics Approaches. **Frontiers in Immunology**, p. 1784, 2020.

FAJGENBAUM, D. C., JUNE, C. H. Cytokine storm. **New England Journal of Medicine**, v. 383, n. 23, p. 2255-2273, 2020.

FARIA, E. *et al.* Performance of vaccination with CoronaVac in a cohort of healthcare workers (HCW) - preliminary report. **MedRxivp**, 2021.

FOSTER, M. *et al.* BCG-induced protection against Mycobacterium tuberculosis infection: Evidence, mechanisms, and implications for next-generation vaccines. **Immunological Reviews**, v. 301, n. 1, p. 122–144, 2021.

FUNK, C. D.; LAFERRIÈRE, C.; ARDAKANI, A. A Snapshot of the Global Race for Vaccines Targeting SARS-CoV-2 and the COVID-19 Pandemic. **Frontiers in Pharmacology**, v. 11, p. 937, 2020.

GIAMARELLOS-BOURBOULIS, E. J. *et al.* Activate: Randomized Clinical Trial of BCG Vaccination against Infection in the Elderly. **Cell**, v. 183, n. 2, p. 315- 323. e9, 2020.

GOEL, S. *et al.* SARS-CoV-2 Switches ‘on’ MAPK and NFκB Signaling via the Reduction of Nuclear DUSP1 and DUSP5 Expression. **Frontiers in Pharmacology**, v. 12, 2021.

GORBALENYA, A. E. *et al.* The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. **Nature Microbiology**, v. 5, n. 4, p. 536–544, 2020.

GRIMES, J. M.; GRIMES, K. V. p38 MAPK inhibition: A promising therapeutic approach for COVID-19. **Journal of Molecular and Cellular Cardiology**, v. 144, p. 63-65, 2020.

GUSTINE, J. N.; JONES, D. Immunopathology of Hyperinflammation in COVID-19. **The American journal of pathology**, v. 191, n. 1, p. 4-17, 2021.

HAAS, E. J. *et al.* Impact and effectiveness of mRNA BNT162b2 vaccine against SARS-CoV-2 infections and COVID-19 cases, hospitalisations, and deaths following a nationwide vaccination campaign in Israel: an observational study using national surveillance data. **The Lancet**, v. 397, n. 10287, p. 1819–1829, 2021.

HAN, X.; XU, P.; YE, Q. Analysis of COVID-19 vaccines : Types , thoughts , and application. **Journal of clinical laboratory analysis**, v. 35, n. 9, p. e23937, 2021.

HARTENIAN, E. *et al.* The molecular virology of coronaviruses. **Journal of Biological Chemistry**, v. 295, n. 37, p. 12910-12934, 2020.

HOFFMANN, M. *et al.* SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. **cell**, v. 181, n. 2, p. 271-280. e8, 2020.

HOLSHUE, M. L. *et al.* First Case of 2019 Novel Coronavirus in the United States. **New England Journal of Medicine**, 2020.

HU, B. *et al.* Characteristics of SARS-CoV-2 and COVID-19. **Nature Reviews Microbiology**, v. 19, n. 3, p. 141–154, 2021.

KIM, Jae Seok *et al.* Immunopathogenesis and treatment of cytokine storm in COVID-19. **Theranostics**, v. 11, n. 1, p. 316, 2021.

KING, C.; SPRENT, J. Dual nature of type I interferons in SARS-CoV-2-induced inflammation. **Trends in immunology**, v. 42, n. 4, p. 312-322, 2021.

KOEKEN, V. A. C. M. *et al.* Trained innate immunity and resistance to Mycobacterium tuberculosis infection. **Clinical Microbiology and Infection**, v. 25, n. 12, p. 1468–1472, 2019.

KRAMMER, F. SARS-CoV-2 vaccines in development. **Nature**, v. 586, n. 7830, p. 516–527, 2020.

KHOSROSHAHI, L. M. *et al.* Immunology, immunopathogenesis and immunotherapeutics of COVID-19; an overview. **International immunopharmacology**, v. 93, p. 107364, 2021.

KUMAR, R. *et al.* Role of MAPK/MNK1 signaling in virus replication. **Virus Research**, v. 253, p. 48–61, 2018.

LAMB, Y. N. BNT162b2 mRNA COVID-19 Vaccine: First Approval. **Drugs**, v. 81, n. 4, p. 495–501, 2021.

LEUNG, K. *et al.* Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. **Eurosurveillance**, v. 26, n. 1, p. 2002106, 2021.

LI *et al.* Molecular immune pathogenesis and diagnosis of COVID-19. **Journal of pharmaceutical analysis**, v. 10, n. 2, p. 102-108, 2020.

LI, X. *et al.* Molecular immune pathogenesis and diagnosis of COVID-19. **Journal of Pharmaceutical Analysis**, v. 10, n. 2, p. 102–108, 2020.

LOPEZ BERNAL, J. *et al.* Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. **New England Journal of Medicine**, 2021.

LU, S. Timely development of vaccines against SARS-CoV-2. **Emerging Microbes and Infections**, v. 9, n. 1, p. 542–544, 2020.

MALIK, Y. A. Properties of Coronavirus and SARS-CoV-2. **Malays J Pathol**, v. 42, n. 1, p. 3–11, 2020.

MALLAH, S. I. *et al.* COVID-19: breaking down a global health crisis. **Annals of Clinical Microbiology and Antimicrobials**, v. 20, n. 1, p. 1–36, 2021.

MEHMOOD, I. *et al.* SARS-CoV-2: An Update on Genomics, Risk Assessment, Potential Therapeutics and Vaccine Development. **International Journal of Environmental Research and Public Health**, v. 18, n. 4, p. 1626, 2021.

MENGIST, H. M. *et al.* Mutations of SARS-CoV-2 spike protein: Implications on immune evasion and vaccine-induced immunity. **Seminars in Immunology**, v. 55, p. 101533, 2021.

MOLAEI, S. *et al.* The immune response and immune evasion characteristics in SARS-CoV, MERS-CoV, and SARS-CoV-2: Vaccine design strategies. **International Immunopharmacology**, v. 92, p. 107051, 2021.

MOXON, R.; RECHE, P. A.; RAPPUOLI, R. Editorial: Reverse Vaccinology. **Frontiers in Immunology**, p. 2776, 2019.

NETEA, M. G. *et al.* Defining trained immunity and its role in health and disease. **Nature Reviews Immunology**, v. 20, n. 6, p. 375–388, 2020a.

NETEA, M. G. *et al.* Trained Immunity: a Tool for Reducing Susceptibility to and the Severity of SARS-CoV-2 Infection. **Cell**, v. 181, n. 5, p. 969–977, 2020b.

NETEA, M. G.; QUINTIN, J.; VAN DER MEER, J. W. M. Trained immunity: A Memory for Innate Host Defense. **Cell Host and Microbe**, v. 9, n. 5, p. 355–361, 2011.

NOORIMOTLAGH, Z. *et al.* Immune and bioinformatics identification of T cell and B cell epitopes in the protein structure of SARS-CoV-2: A systematic review. **International Immunopharmacology**, v. 86, p.106738, 2020.

O'NEILL, L. A. J.; NETEA, M. G. BCG-induced trained immunity: can it offer protection against COVID-19? **Nature Reviews Immunology**, v. 20, n. 6, p. 335–337, 2020.

OLI, D. C. *et al.* Immunoinformatics and Vaccine Development: An Overview. v. 9, p. 13, 2020.

PANDEY, S. C. *et al.* Vaccination strategies to combat novel corona virus SARS-CoV-2. **Life Sciences**, v. 256, p. 117956, 2020.

PENG, X.-L. *et al.* Advances in the design and development of SARS-CoV-2 vaccines. **Military Medical Research**, v. 8, n. 1, p. 1–31, 2021.

PRITCHARD, E. *et al.* Impact of vaccination on new SARS-CoV-2 infections in the United Kingdom. **Nature Medicine**, v. 27, n. 8, p. 1370–1378, 2021.

PORAN, A. *et al.* Sequence-based prediction of vaccine targets for inducing T cell responses to SARS-CoV-2 utilizing the bioinformatics predictor RECON. **bioRxiv**, 2020.

PROMPETCHARA, E., KETLOY, C., PALAGA, T. Immune responses in COVID-19 and potential vaccines: Lessons learned from SARS and MERS epidemic. **Asian Pacific journal of allergy and immunology**, v. 38, n. 1, p. 1-9, 2020.

RAMAN, B. *et al.* Medium-term effects of SARS-CoV-2 infection on multiple vital organs, exercise capacity, cognition, quality of life and mental health, post-hospital discharge. **EClinicalMedicine**, v. 31, p. 100683, 2021.

RAPPUOLI, R. *et al.* Reverse vaccinology 2.0: Human immunology instructs vaccine antigen design. **Journal of Experimental Medicine**, v. 213, n. 4, p. 469–481, 2016.

RODRIGUEZ-MORALES, A. J. *et al.* COVID-19 in Latin America: The implications of the first confirmed case in Brazil. **Travel Medicine and Infectious Disease**, v. 35, p. 101613, 2020.

SAFIABADI TALI, S. H. *et al.* Tools and techniques for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)/COVID-19 detection. **Clinical Microbiology Reviews**, v. 34, n. 3, p. e00228-20, 2021.

SAHU, A. K. *et al.* SARS-CoV-2: phylogenetic origins, pathogenesis, modes of transmission, and the potential role of nanotechnology. **VirusDisease**, v. 32, n. 1, p. 1–12, 2021.

SERRUTO, D. *et al.* The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: Immunological, functional and structural characterization of the antigens. **Vaccine**, v. 30, p. B87-B97, 2012.

SETTE; CROTTY. Adaptive immunity to SARS-CoV-2 and COVID-19. **Cell**, v. 184, n. 4, p. 861-880, 2021.

SIDDIQI, H. K., MEHRA, M. R. COVID-19 illness in native and immunosuppressed states: A clinical–therapeutic staging proposal. **The journal of heart and lung transplantation**, v. 39, n. 5, p. 405-407, 2020.

STRICH, J. R. *et al.* Fostamatinib inhibits neutrophils extracellular traps induced by COVID-19 patient plasma: a potential therapeutic. **The Journal of infectious diseases**, v. 223, n. 6, p. 981-984, 2021.

TEGALLY, H. *et al.* Detection of a SARS-CoV-2 variant of concern in South Africa. **Nature**, v. 592, n. 7854, p. 438–443, 2021.

TREGONING, J. S. *et al.* Vaccines for COVID-19. **Clinical and Experimental Immunology**, v. 202, n. 2, p. 162–192, 2020.

TSILIKA, M. *et al.* Activate-2: a Double-Blind Randomized Trial of BCG Vaccination Against COVID-19 in Individuals at Risk. **MedRxiv**, 2021.

V'KOVSKI, P. *et al.* Coronavirus biology and replication: implications for SARS-CoV-2. **Nature Reviews Microbiology**, v. 19, n. 3, p. 155–170, 2021.

VAN DER MEER, J. W. M. *et al.* Trained immunity: A smart way to enhance innate immune defence. **Molecular Immunology**, v. 68, n. 1, p. 40–44, 2015.

VIOLA, A. *et al.* The metabolic signature of macrophage responses. **Frontiers in Immunology**, p. 1462, 2019.

VISCIANO, M. L. *et al.* Effects of adjuvants on IgG subclasses elicited by virus-like Particles. **Journal of Translational Medicine**, v. 10, n. 1, p. 1-9, 2012.

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

WAN *et al.* Characteristics of lymphocyte subsets and cytokines in peripheral blood of 123 hospitalized patients with 2019 novel coronavirus pneumonia (NCP). **MedRxiv**, 2020.

WANG, P. *et al.* Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. **Nature**, v. 593, n. 7857, p. 130–135, 2021.

WATANABE, Y. *et al.* Site-specific glycan analysis of the SARS-CoV-2 spike. **Science**, v. 369, n. 6501, p. 330-333, 2020.

WIBMER, C. K. *et al.* SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. **Nature Medicine**, v. 27, n. 4, p. 622–625, 2021.

WORLD HEALTH ORGANIZATION. Coronavirus Disease 2019 (COVID-19) Situation Report 51. 2020. World Health Organization. Disponível em: <<https://apps.who.int/iris/handle/10665/331475>>. Acesso em: 25 jan. 2022.

WORLD HEALTH ORGANIZATION (WHO). COVID-19 vaccine tracker and landscape. World Health Organization, 2022. Disponível em: <<https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines>>. Acesso em: 25 jan. 2022.

WU, J., TANG, Y. Revisiting the immune balance theory: a neurological insight into the epidemic of COVID-19 and its alike. **Frontiers in Neurology**, v. 11, p. 566680, 2020.

XAVIER, M. N. *et al.* CD4+ T cell-derived IL-10 promotes *Brucella abortus* persistence via modulation of macrophage function. **PLoS pathogens**, v. 9, n. 6, p. e1003454, 2013.

ZARUBIN, T.; HAN, J. Activation and signaling of the p38 MAP kinase pathway. **Cell Research**, v. 15, n. 1, p. 11–18, 2005.

ZHOU, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. **Nature**, v. 579, n. 7798, p. 270–273, 2020.

ZHUANG, Z. *et al.* Mapping and role of T cell response in SARS-CoV-2–infected mice. **Journal of Experimental Medicine**, v. 218, n. 4, 2021.

ANEXO I

MINISTÉRIO DA EDUCAÇÃO
 Universidade Federal de Alfenas - UNIFAL-MG
 Lei nº 11.154, de 29 de julho de 2005

Comissão de Ética no Uso de Animais - CEUA/UNIFAL-MG

Certificado

Certificamos que a proposta intitulada **Avaliação da imunogenicidade de epítomos racionalmente selecionados por vacinologia reversa de proteínas estruturais do SARS-CoV-2 e associados ao BCG vacinal como carreador e indutor da imunidade inata treinada**, registrada com o nº **0008/2021**, sob a responsabilidade de **Leonardo Augusto de Almeida**, que envolve a produção, manutenção ou utilização de animais pertencentes ao filo Chordata, subfilo Vertebrata (exceto humanos), para fins de **pesquisa científica**, com vigência de **01/03/2021 a 30/12/2022**, encontra-se de acordo com os preceitos da Lei nº 11.794, de 8 de outubro de 2008, do Decreto nº 6.899, de 15 de julho de 2009, e com as normas editadas pelo Conselho Nacional de Controle de Experimentação Animal (CONCEA), e foi aprovado pela COMISSÃO DE ÉTICA NO USO DE ANIMAIS (CEUA-UNIFAL) DA UNIVERSIDADE FEDERAL DE ALFENAS.

Espécie/linhagem/raça	Total de animais	Total de machos	Total de fêmeas	Origem
Rodent / C57bl/6	120	60	60	Biotério central

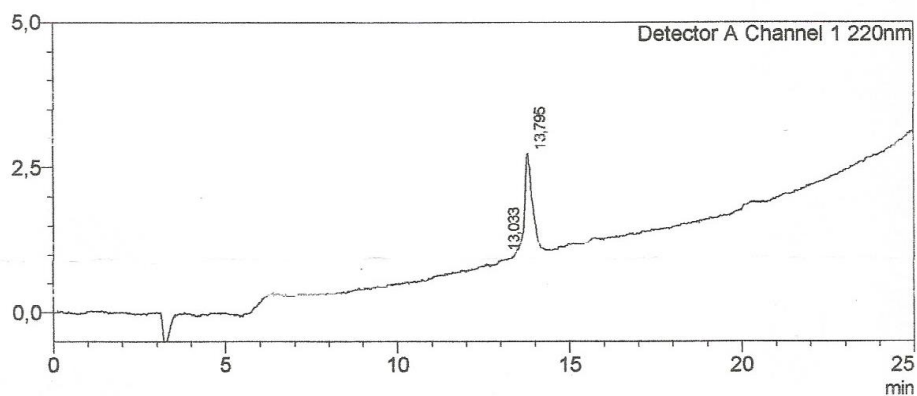
Alfenas, 26 de Abril de 2021

Prof(a). Dr(a). Pollyanna Francielli de Oliveira
 Coordenador(a) do CEUA/UNIFAL - MG

Para verificar autenticidade acesse: <http://sistemas.unifal-mg.edu.br/app/cvua/autenticidadecertificada/> e digite a chave: 934228c55916306be0481d6ae854f028

ANEXO II

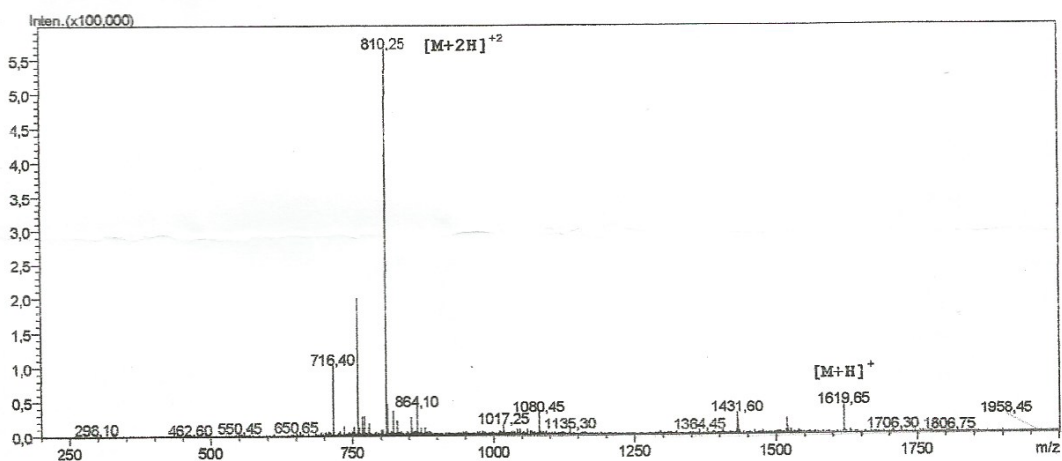
Sintetização do Epítopo 1



Peak Table

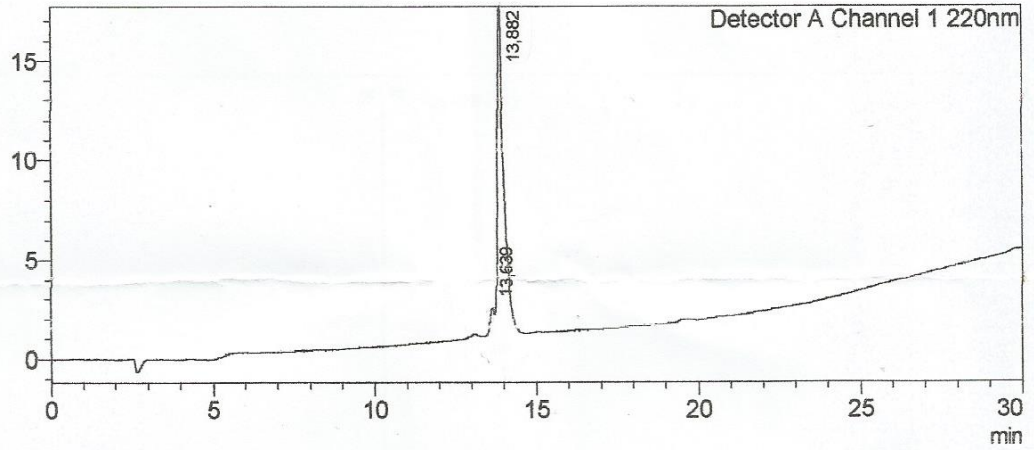
Peak#	Ret. Time	Area	Area%	Height	Height%	Conc.	Mark
1	13,033	285	0,923	27	1,499	0,923	M
2	13,795	30574	99,077	1769	98,501	99,077	V
Total		30859	100,000	1795	100,000		

Condições de Análise
 Coluna RP-C18, 4,6x250mm - Fluxo: 1mL/min
 Solventes: A=TFA 0,1%/H₂O, B=CH₃CN/H₂O 9:1 com TFA 0,1%
 Detector: 220nm - Gradiente: 5 a 100% de B em 25 min.



ANEXO II

Sintetização do Epítipo 2



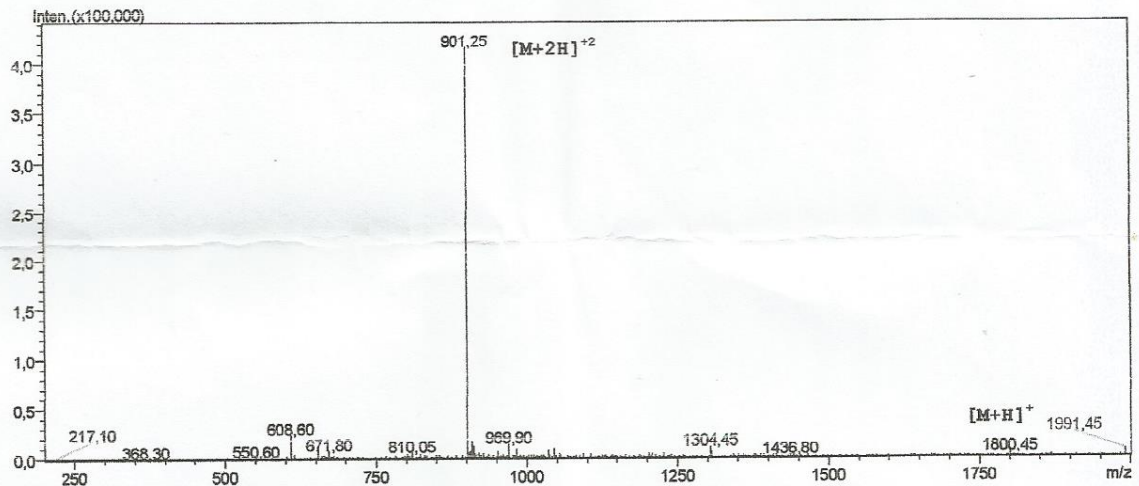
Peak Table

Detector A Channel 1 220nm

Peak#	Ret. Time	Area	Area%	Height	Height%	Conc.	Mark
1	13,639	7570	3,242	1200	6,809	3,242	M
2	13,882	225921	96,758	16417	93,191	96,758	M
Total		233491	100,000	17617	100,000		

Condições de Análise

Coluna RP-C18, 4,6x250mm - Fluxo: 1mL/min
 Solventes: A=TFA 0,1%/H₂O, B=CH₃CN/H₂O 9:1 com TFA 0,1%
 Detector: 220nm - Gradiente: 5 a 70% de B em 20 min.



ANEXO III

Immunoinformatics 7 (2022) 100015



Contents lists available at ScienceDirect

Immunoinformatics



Epitope identification of SARS-CoV-2 structural proteins using in silico approaches to obtain a conserved rational immunogenic peptide

Leonardo Pereira de Araújo ^{a, b}, Maria Eduarda Carvalho Dias ^a, Gislaine Cristina Scodeler ^a, Ana de Souza Santos ^a, Letícia Martins Soares ^a, Patrícia Paiva Corsetti ^a, Ana Carolina Barbosa Padovan ^a, Nelson José de Freitas Silveira ^b, Leonardo Augusto de Almeida ^{a,*}

^a Laboratory of Molecular Biology of Microorganisms, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil

^b Laboratory of Molecular Modeling and Computer Simulation, Federal University of Alfenas, Alfenas, Minas Gerais, Brazil

ARTICLE INFO

Keywords:

COVID-19

New coronavirus

Reverse vaccinology

SARS-CoV-2

ABSTRACT

The short time between the first cases of COVID-19 and the declaration of a pandemic initiated the search for ways to stop the spread of SARS-CoV-2. There are great expectations regarding the development of effective vaccines that protect against all variants, and in the search for it, we hypothesized the obtention of a predicted rational immunogenic peptide from structural components of SARS-CoV-2 might help the vaccine research direction. In the search for a candidate of an immunogenic peptide of the SARS-CoV-2 envelope (E), membrane (M), nucleocapsid (N), or spike (S) proteins, we access the predicted sequences of each protein after the genome sequenced worldwide. We obtained the consensus amino acid sequences of about 14,441 sequences of each protein of each continent and the worldwide consensus sequence. For epitope identification and characterization from each consensus structural protein related to MHC-I or MHC-II interaction and B-cell receptor recognition, we used the IEDB reaching 68 epitopes to E, 174 to M, 245 to N, and 833 to S proteins. To select an epitope with the highest probability of binding to the MHC or BCR, all epitopes of each consensus sequence were aligned. The curation indicated 1, 4, 8, and 21 selected epitopes for E, M, N, and S proteins, respectively. Those epitopes were tested in silico for antigenicity obtaining 16 antigenic epitopes. Physicochemical properties and allergenicity evaluation of the obtained epitopes were done. Ranking the results, we obtained one epitope of each protein except for the S protein that presented two epitopes after the selection. To check the 3D position of each selected epitope in the protein structure, we used molecular homology modeling. Afterward, each selected epitope was evaluated by molecular docking to reference MHC-I or MHC-II allelic protein sequences. Taken together, the results obtained in this study showed a rational search for a putative immunogenic peptide of SARS-CoV-2 structural proteins that can improve vaccine development using in silico approaches. The epitopes selected represent the most conserved sequence of new coronavirus and may be used in a variety of vaccine development strategies since they are also presented in the described variants of SARS-CoV-2.

1. Introduction

1.1. SARS-CoV-2 and COVID-19

Coming from a large family of single-stranded RNA viruses, SARS-CoV-2 is a polyadenylated, positive-sense virus that is responsible for the worsening of the severe acute respiratory syndrome, elevating the outbreak that began in December 2019 in Wuhan, China, to the pandemic status declared on March 11, 2020, by the World Health

Organization [1,2]. Seven coronaviruses capable of infecting humans are known, three of which are potential causes of serious diseases: MERS-CoV, SARS-CoV, and SARS-CoV-2; meanwhile, 229E, NL63, HKU1, and OC43 are linked to mild symptoms not related to complications [3]. Due to the higher transmissibility and faster spread of SARS-CoV-2, the WHO announced the pandemic status in February 2020 [1]. The coronavirus disease 2019 (COVID-19) causes a variety of symptoms from mild to severe cases characterized by pneumonia with a pro-inflammatory cytokine storm or extrapulmonary responses to virus

* Corresponding author.

E-mail address: leonardo.almeida@unifal-mg.edu.br (L.A. de Almeida).

<https://doi.org/10.1016/j.immuno.2022.100015>

Received 23 July 2021; Received in revised form 8 April 2022; Accepted 10 June 2022

Available online 11 June 2022

2667-1190/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

leading to systemic effects as vasculitis and subsequently inducing thrombosis [4].

1.2. Search for COVID-19 treatment and vaccines

Until now, about 8,000 clinical trials are in progress for COVID-19, including the development of new drugs or drug repositioning and vaccine development. However, despite the great efforts of scientists around the world, there is still no fully efficient treatment for COVID-19 [5]. An effective strategy for the control and prevent the new corona-virus spread is the use of vaccines as perceived with the decrease in the number of cases in countries with high rates of vaccinated people [6]. Although there are different approaches to obtain a vaccine to SARS-CoV-2, it is important to explore the structural proteins of the virus since they are fundamental to directly interacting with the host and may be the source of specific immunogenic components against it [7].

1.3. Immunoinformatics and COVID-19 vaccine development based on SARS-CoV-2 structural proteins

With the advent of immunoinformatics where the bioinformatics have helped a quick way to develop epitope-based vaccines *in silico* as a preliminary study for the *in vivo* validation study [8]. Sohail and colleagues [9] proposed this strategy to identify *in silico* T cell epitope identification for SARS-CoV-2. Since the ideal immune response against the new coronavirus should include an efficient antiviral innate immunity and a robust, specific cellular and humoral adaptative response against this virus [10] it is necessary to identify possible immunogenic epitopes based on the different SARS-CoV-2 genomes present in the DNA repositories. SARS-CoV-2 expresses 29 proteins, four of which are structural proteins: envelope protein (E), membrane glycoprotein (M), nucleocapsid phosphoprotein (N), and surface glycoprotein (S). They are closely associated with host interaction during the infection, which makes them important targets to be recognized by the immune system [11]. Among the structural proteins, the S protein is used as the most important to host recognition due to its presence around the virion, and it might be the best target for neutralizing antibodies. However, the SARS-CoV-2 variants described during the pandemic are mostly based on differences in amino acids in the primary sequences of the S protein, decreasing the capacity of neutralizing antibodies. In addition, researchers have observed interesting mechanisms of the new coronavirus to induce filopodia in the infected cells, which can augment the spread of the virus [12]. Therefore, a robust and efficient cellular immune response should be triggered by epitope-based vaccines that can induce elimination of the focus of infection using T cell activation-dependent mechanisms [13,14]. In order to obtain a rationally selected, specific immunogenic epitope from SARS-CoV-2 structural proteins that surpasses the possible issues associated with the virus variants or only one type of adaptive immune response, we identified a consensus sequence of the four structural proteins with a high possibility to elicit T and B cell responses against this virus. We used different established immunoinformatic approaches and structural analysis to identify the position of each selected epitope in the consensus 3D structure modeled by us for the structural proteins of SARS-CoV-2.

2. Methods

2.1. Obtaining SARS-CoV-2 putative structural protein consensus sequences

The methods followed the pipeline presented in Supplementary Fig. 1, showing the main immunoinformatic and molecular modeling and docking established tools used in this manuscript. The protein sequences for SARS-CoV-2 were obtained from the National Center of Biotechnology Information (NCBI) database. The specific database for Sars-CoV-2, present on the website in the "NCBI Virus" tab, was used by

means of the following identification: "taxid: 2697049" (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/virus?SeqType_s=Protein&VirusLineage_ss=SARS-CoV-2,%20taxid:2697049). All sequences of the envelope protein (E), membrane glycoprotein (M), nucleocapsid phosphoprotein (N), and surface glycoprotein (S) from all continents present on the site were downloaded in FASTA format. The Jalview software (<https://www.jalview.org/>) was used to obtain the consensus sequences for each region and the global consensus sequence for each of the four target proteins [15].

2.2. Epitopes identification

For all epitope predictions of the target proteins, we used the algorithms of the Immune Epitope Database and Analysis Resource (IEDB; <https://www.iedb.org/>), of which only the methods were changed for the selection of each prediction.

2.3. MHC-I or MHC-II and B-cell receptor epitopes prediction

The NetMHCpan EL 4.1 server (<http://tools.iedb.org/mhci/>) was used and the option to obtain epitopes that interact with the 27 reference alleles of MHC-I [16] that simulate a greater coverage of the global population, among 97% and 99% and IC50 < 500nM [17] in sequence the cutoff of 1% of the total amount of each target structure was used; all options were recommended by the IEDB platform itself [18,19]. The "IEDB recommended 2.22" method (<http://tools.iedb.org/mhcii/>) was used, which consists of the consensus use of the NN-align [20], SMM-align [21], CombLib [22], and Sturniolo [23] methods, to obtain the best possible result for a given protein, as recommended by the IEDB. The combination of the 27 MHC-II alleles recommended for the world-wide coverage of the population was also used [24], for the cutoff, the epitopes that fit into two observations, including a consensus percentile threshold rank < 20.0 and interacting with more than 50% of the selected alleles [25], were selected. The BepiPred Linear Epitope Prediction 2.0 server (<http://tools.iedb.org/bcell/>) was used to obtain B cell interaction epitopes, based on a system of machine learning algorithms called Random Forest that classifies the amino acid sequences of the target proteins from simulations in crystallized proteins and amino acids not considered epitopes. A standard cutoff threshold of 0.5 was used to obtain promiscuous epitopes [26].

2.4. Alignment and determination of worldwide consensus epitopes

All epitopes found from each structural protein that interacts with B-cell receptors MHC-I and MHC-II were aligned using the MultiAlign server (<http://multalin.toulouse.inra.fr/multalin/>; [27] in addition to a manual alignment for comparison and greater precision. From the alignment, the epitopes with the highest repetition rate per region with an established size of 15 amino acids were selected. The selection of the most promiscuous 15 amino acids epitope was based on the peptide length able to fit in the MHC groves with best affinity for interaction, stability, and specificity to SARS-CoV-2 proteins.

2.5. Transmembrane epitopes identification

All regions of epitopes that were within the transmembrane regions of the SARS-CoV-2 structural proteins were removed in such a way the SOSUI server was used (<http://harrier.nagahama-i-bio.ac.jp/sosui/sosuisubmit.html>) [28], which identifies the transmembrane regions present in the proteins. It is worth mentioning that the nucleocapsid phosphoprotein is a completely soluble protein with no transmembrane regions.

2.6. Selected epitopes antigenicity, allergenicity, putative N-glycan sites and physical and chemical properties test

For the antigenicity test, the Vaxijen 2.0 server (<http://www.ddg-pharmfac.net/vaxijen/Vaxijen/Vaxijen.html>), with 89% accuracy, was used with the 0.5 and “probable antigen” cutoff [29]. For the allergenicity test, the server AlgPred 2.0, an update version of “AlgPred: Prediction of Allergenic Proteins and Mapping of IgE Epitopes” (<http://crrd.osdd.net/raghava/algpred/>), with sensitivity of 93.1%, of 95.36% specificity and 94.26% accuracy, was used with the hybrid method that consists of using the following five tools available on the server: SVMc, IgE, epitope, ARPs BLAST, and MAST [30]. The Putative N-glycan sites were determined using the NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>), 76% accuracy [31]. For the physical and chemical properties test, the ExPasy server’s ProtParam tool (<http://web.expasy.org/protparam/>) was used with a cutoff of 40 for structure stability values [32].

2.7. SARS-CoV-2 structural proteins modeling

For the molecular modeling of SARS-CoV-2 envelope protein, membrane glycoprotein, nucleocapsid phosphoprotein, and surface glycoprotein membrane glycoprotein, the script AlphaFold v2.0 [33] (<https://github.com/deepmind/alphafold>) was used to predict the conformation of the proteins with an extremely high level of accuracy (>95%). Multiple databases as UniRef90 [34], MGnify [35], BFD [36], Uniclust30 [37], PDB70 and PDB [38] were used to homology modeling the structural proteins.

2.8. Refinement of human leukocyte antigen (HLA) and selected epitopes from structural proteins from SARS-CoV-2 for molecular docking

Receptor proteins for the docking process were selected from the 27 interaction alleles made available by the IEDB, and then a search was carried out on the PDB database for all these alleles to perform molecular docking. All proteins, both receptor (HLA) and ligands (Epitopes), were subjected to a preparation, in which hydrogens absent from the structures were added, the waters present in the .pdb files were deleted, the protein was scanned using the script complete_pdb.py from the MODELLER software in order to find missing atoms, and the grid box in the active site of the structure was defined [39].

2.9. HLA and selected epitopes from structural proteins from SARS-CoV-2 docking

To carry out the dockings, the protein-peptide method was defined, which consists of determining the receptor protein as a rigid structure with some small conformations of the side chains as flexible and the ligand-protein as a flexible structure. For that, we used the MDockPep servers (<https://zougrouptoolkit.missouri.edu/mdockpep/index.html>; [40] XU X et al., 2018) and ClusPro (<https://cluspro.org/help.php>) [41].

In addition, the AutoDock Vina software was used with the intention of obtaining a greater range of results and thus having a greater precision of results [42]. LigPlot [43] and Pymol [44] were used to determine the epitope-MHC interaction in 2D and 3D diagrams, respectively.

3. Results

3.1. SARS-CoV-2 structural proteins consensus sequence from different deposited genomes in a database.

From the NCBI database, we analyzed 53,765 putative SARS-CoV-2 structural proteins from translated, deposited genomes (table 1), representing sequences from six continents. North America is the continent with the highest number of obtained SARS-CoV-2 genome sequences with a total of 41,112 sequences (10,278 per protein). Asia presents 4,784 sequences (1,196 per protein), followed by Oceania with 5,672 sequences (1,418 per protein), Europe with 1,644 sequences (411 per protein), Africa with 428 sequences (107 per protein), and finally South America with 124 sequences (31 per proteins). After obtaining all sequences from all continents, the process of obtaining consensus sequences was carried out. In addition to the respective conservation values, in which a consensus sequence of each protein per continent was obtained (Supp. 1-4), we identified global consensus sequences from each new coronavirus structural protein to be used as targets in obtaining putative epitopes.

3.2. Putative immunogenic epitopes from the consensus sequence of SARS-CoV-2 structural proteins are described to interact with MHC-I, and the S protein contains a higher number of epitopes

According to Table 1, the higher number of epitopes from structural proteins from SARS-CoV-2 interact more with MHC-I and less with B-cell receptors. Since the surface glycoprotein is considered the highest immunogenic protein of SARS-CoV-2 and the bigger structural protein from this virus, the majority of putative epitopes were found in this protein (773 putative epitopes). On the other hand, the envelope protein with 75 amino acids presented only 68 putative epitopes in this sequence. All results are presented in Table 1, and they were obtained using the IEDB software performing a projection of 98,55% of population coverage according to “Population Coverage Calculation Result” tool.

3.3. Rational selection of putative immunogenic epitopes from SARS-CoV-2 structural proteins indicates at least one candidate epitope in each protein

To obtain the best candidates of putative immunogenic epitopes from each SARS-CoV-2 structural protein, the obtained epitopes were aligned, and the linear amino acids sequence performing about 15 amino acids that presenting repeating was selected. Furthermore, epitopes presented in the transmembrane anchoring site were excluded since they present

Table 1

Rational selection of consensus structural proteins epitopes of SARS-CoV-2.

	Africa	Asia	Europe	North America	Oceania	South America	Total
Genomes	428	4,784	1,644	41,112	5,672	124	53,764
Epitopes from consensus structural proteins of SARS-CoV-2							
	Envelope		Membrane		Nucleocapsi		Surface
B-cell receptor	2		6		11		33
MHC-I	36		116		222		625
MHC-II	30		52		12		115
Rational selection of the best epitopes							
	Envelope		Membrane		Nucleocapsi		Surface
Multalign	1		4		8		21
Vaxijen	1		3		3		9
ProtParam	1		1		2		6
AlgPred	1		1		1		2

hydrophobic characteristics that cause difficult interactions with the amino acids present in the MHC-I or MHC-II grooves. Since the nucleocapsid phosphoprotein is a soluble protein, all repeated epitopes identified in this protein were kept for the next evaluations. After those selections, 34 epitopes were obtained, and they were in silico tested for antigenicity using the Vaxijen 2.0 software, obtaining 16 antigenic epitopes (Table 2). Physicochemical properties evaluation of the obtained epitopes was based on molecular weight, isoelectric point, structural formula, number of atoms, number of amino acids, half-life, structural stability, and n-glycan sites. At this point, 10 epitopes were selected. It is important to note that the epitope from envelope protein is the only epitope that did not reach the cutoff value for stability. To overcome it, it was necessary to augment the selected epitope to 17 amino acids to make it stable but with a n-glycan site. The epitope was subjected to the antigenicity evaluation again, yielding a score necessary to be considered a putative immunogenic epitope. To increase the confidence of using the selected epitopes in future in vitro and in vivo evaluation, the allergenicity was tested using the AlgPred software with the following five tools in the server: SVMc, IgE, epitope, ARPs BLAST, and MAST. At this point, only five epitopes were not allergenic. Ranking the results, we obtained one epitope of each protein except for the S protein that presented two epitopes after selection (Table 1).

3.4. Position identification of selected putative immunogenic epitopes in SARS-CoV-2 structural proteins after molecular modeling by homology

To identify the position of selected putative immunogenic epitopes in the structural proteins from SARS-CoV-2, we performed molecular modeling for each global consensus protein. Since there is no information about the 3D consensus protein in the PDB database, we used the opensource script AlphaFold v2.0 for the envelope protein (E; Fig. 1A), membrane glycoprotein (M; Fig. 1B), nucleocapsid phosphoprotein (N; Fig. 1C), and surface glycoprotein (S; Fig. 1D). The selected epitopes inserted in the 3D structure that were identified for each protein are highlighted in red in Fig. 1. The transmembrane helices are indicated in the models in green.

3.5. Molecular docking between the selected putative immunogenic epitopes in SARS-CoV-2 structural proteins and MHC-I and MHC-II showshigh probabilities of interaction between them

Although not all structures of MHC-I and MHC-II were found in the database (IEDB), we used the structure of 14 proteins of MHC-I alleles and three of MHC-II alleles according to the underlined presented in Table 3. The results demonstrated great energy of interaction based on two different online servers and by AutoDock Vina software. To represent the molecular docking between the epitope and the MHC-I and MHC-II alleles, we chose the interaction with the highest power of interaction indicated by the servers and the software. The best interaction

between the envelope protein (E) and MHC-I was with the HLA-B*1501 allele (Figs 2A and 2C) and MHC-II with the HLA-DRB1*04:01 allele (Figs 2B and 2D). The points of contact are demonstrated in the 3D Figs 2A and 2B, and twenty-two hydrophobic interactions between the epitope and MHC-I are 2D represented in Fig. 2C. For epitope and MHC-II interactions it was observed eighteen hydrogen bonds and twenty-eight hydrophobic interactions (Fig. 2D). For the membrane glycoprotein (M), we demonstrated the interaction between the HLA-B*57:01 allele for MHC-I (Figs 3A and 3C) and the HLA-DRB1*15:01 allele for MHC-II (Figs 3B and 3D). The points of contact are demonstrated in the 3D Figs 3A and 3B, and eight hydrogen bonds and thirty-five hydrophobic interactions between the epitope and MHC-I are 2D represented in Fig. 3C. For epitope and MHC-II interactions it was observed six hydrogen bonds and thirty-two hydrophobic interactions (Fig. 3D). The epitope present in the nucleocapsid phosphoprotein (N) was dock-ing with the HLA-B*35:01 allele for MHC-I (Figs 4A and 4C) and HLA-DRB1*04:01 allele for MHC-II (Figs 4B and 4D). The 3D diagram shows the points of contact in Figs 4A and 4B. Seven hydrogen bonds and twenty-six hydrophobic interactions between the selected epitope and MHC-I are represented in the Fig. 4C, while ten hydrogen bonds and thirty-seven hydrophobic interactions between the selected epitope and MHC-II are represented in the Fig. 4D. The surface glycoprotein (S) presented two epitopes, and the results of the interaction are demonstrated in Figs 5 and 6. For epitope 1, we demonstrated the interaction between the MHC-I allele HLA-A*02:03 (Figs 5A and 5C), with twelve hydrogen bonds and twenty-four hydrophobic interactions between the epitope 1 and MHC-I (Fig. 5C). The interaction between the epitope1 and MHC-II allele HLA-DRB1*04:01 showed nine hydrogen bonds and twenty-nine hydrophobic interactions (Fig. 5D). For epitope 2 from the S protein demonstrated the interaction between the allele HLA- B*35:01 (Figs 6A and 6C) and allele HLA-DRB1*15:01 (Figs 6B and 6D) with MHC-I and MHC-II, respectively. The points of contact are demonstrated in the 3D Figs 6A and 6B, and nine hydrogen bonds and twenty-nine hydrophobic interactions between the epitope 2 and MHC-I (Fig. 6C). For epitope 2 and MHC-II interactions it was observed eight hydrogen bonds and twenty-seven hydrophobic interactions (Fig. 6D).

4. Discussion

4.1. Short time to find putative immunogenic epitopes to COVID-19 vaccine development by immunoinformatic

The urgency of an effective vaccine against emerging and rapidly transmitting pathogens, such as the new coronavirus, is a reality. The WHO report, after a year and a half of the COVID-19 pandemic status, that there are more than 284 vaccines in development, 100 of which are in clinical studies in different phases, nine are in phase III already published, two of messenger RNA technology (Pfizer, Moderna), four of viral vectors (AstraZeneca, Gamaleya, Jansen, CanSino), one of

Table 2

Antigenicity, allergenicity, and physical-chemical properties of selected epitopes from consensus structural proteins epitopes of SARS-CoV-2.

	Envelope	Membrane	Nucleocapsid	Surface 1	Surface 2
Sequence	LVKPSFYVYSRVKLNLS	TVATSRTLSEYKLG	MEVTPSGTWLTYTGA	IAIPTNFTISVTTEI	ALQIPFAMQMAYRFN
Antigenicity	0,6582	0,7300	0,7886	0,7719	1,0112
MW	2014,36	1630,86	1613,80	1619,88	1801,15
PI	10,00	9,70	4,00	4,00	8,79
Stability	38,99	19,08	-7,92	7,39	33,91
Half-life	5,5 h	7,2 h	30 h	20 h	4,4 h
Reticulocytes					
Half-life	3 min	>20 h	>20	30 min	>20 h
Yeast					
Half-life	2 min	>10 h	>10 h	>10 h	>10 h
E. coli					
IgE epitope	No	No	No	No	No
Allergenicity	No	No	No	No	No
n-glycan site	Yes	No	No	Yes	No

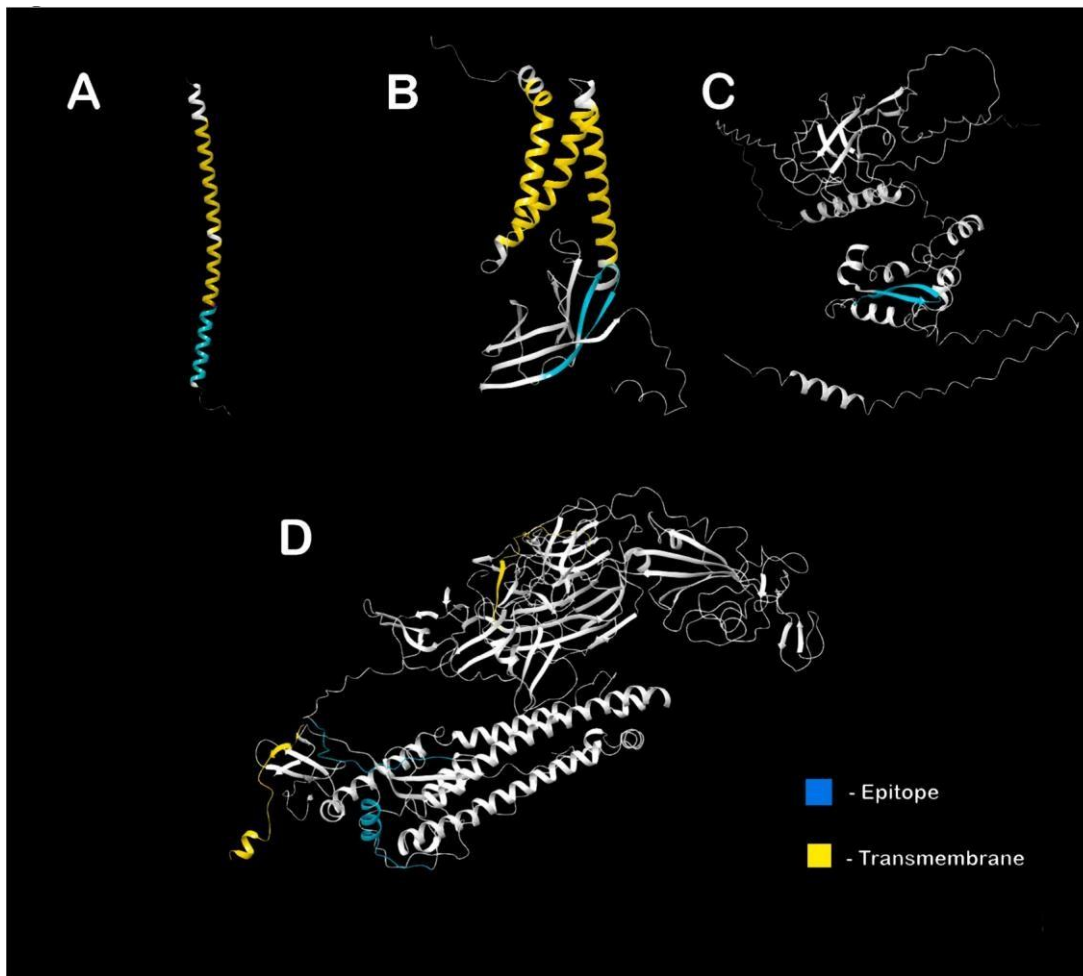


Fig. 1. Position identification of selected epitopes in consensus 3D structural proteins from SARS-CoV-2. Molecular modeling by homology for the envelope protein (A), membrane glycoprotein (B), nucleocapsid phosphoprotein (C), and surface glycoprotein (D). The molecular modeling of the proteins was based on homology using AlphaFold v2.0 scrip. The selected epitopes are highlighted in red, and the transmembrane helices are indicated in the models in green.

recombinant protein (Novavax), and two using an inactivated virus (Sinopharm). Here we used *in silico* approaches to obtain epitopes from structural proteins from SARS-CoV-2 to reach, in a short period of time, a specific immunogenic peptide able to be used as a safe peptide-based vaccine. Currently, numerous projects for the development of effective vaccines have been carried out, and some of these vaccines are already commercially available for humans [45]. However, traditional vaccine production techniques have some disadvantages, which can be overcome by using computational approaches [46]. In addition, recent advances in bioinformatics have provided a variety of tools and servers capable of reducing the cost and time of advancing the traditional vaccine [5]. Immunoinformatics approaches can be used to analyze pathogen antigens, predict their epitopes, and assess their immunogenicity [5]. Furthermore, reverse vaccinology, epitope prediction, structural vaccinology, rational approaches, and molecular docking are of great use in designing a potential vaccine against COVID-19 [46].

4.2. Global consensus amino acid sequence and modeling of SARS-CoV-2 structural proteins

Therefore, in this work, bioinformatics techniques were used to reach a consensus amino acid sequence from the four structural proteins of SARS-CoV-2 using the sequences of genomes deposited in the database from different regions of the world following obtaining a global consensus sequence. The strategy to obtain a consensus sequence of each protein may decrease the possibility of observing variants, but it

increases the possibility to define the conserved amino acid sequences from those proteins on available genomes around the world. Here, we were able to define the three-dimensional structure of the consensus proteins and identify their secondary structures using molecular modeling and artificial intelligence data. Furthermore, the immunoinformatics approaches may direct in a rational way the identification of putative immunogenic epitopes to design an epitope-based vaccine to control SARS-CoV-2 [47].

4.3. Rational selection of putative immunogenic epitopes from global consensus SARS-CoV-2 structural protein

We also demonstrated that sequentially using rational *in silico* techniques, based on well-established bioinformatic tools, made it possible to select the best epitope able to interact with the B-cell receptor, which may stimulate a specific adaptative humoral response, or MHC-I and MHC-II alleles, which may stimulate a cellular adaptative immune response based on TCD8 or TCD4 activation, respectively. The best example is the search for epitopes in the ⁺sequence of the surface glycoprotein (S). There are about 770 epitopes in the protein S, but with our sequence of analysis, only two potential epitopes are selected in this protein. S glycoprotein, also called spike, is one of the main proteins for therapeutic and vaccine targets, in addition to being one of the most important proteins to study, as it is responsible for the SARS-CoV-2 virus infecting host cells. It has a structure with covalently linked carbohydrate molecules that are N-glycans and is extremely glycosylated. This

Table 3

Molecular docking between selected epitopes of SARS-CoV-2 consensus structural proteins and HLAs.

	Epitope	Allele	ClusPro		MdockPep	AutoDock Vina	
			Center	Lowest Energy			
MHC-I	Envelope	HLA-A*0203	-842,2	-842,2	-266,03	-7,1	
		HLA-B*0801	-636,7	-963,7	-239,0	-7,3	
		HLA-B*1501	-635,5	-791,4	-261,9	-7,4	
	Membrane	HLA-B*3501	-679,4	-923,0	-273,9	-6,4	
		HLA-A*0206	-749,8	-749,8	-270,2	-5,7	
		HLA-B*5701	-655,0	-850,3	-230,4	-8	
		HLA-B*3501	-693,9	-833,7	-221,1	-7,3	
		HLA-B*5801	-578,8	-689,7	-220,4	-7,4	
		HLA-A*0301	-731,9	-731,9	-223,6	-7,0	
	Nucleocapsid	HLA-A*0206	-809,0	-809,6	-247,7	-6,5	
		HLA-B*3501	-711,2	-911,1	-203,6	-7,8	
	Surface 1	HLA-B*4403	-681,4	-833,6	-204,3	-7,4	
		HLA-A*0203	-792,0	-931,7	-187,7	-8,3	
		HLA-B*5101	-796,5	-1112,8	-185,4	-7,9	
	Surface 2	HLA-B*3501	-792,4	-843,8	-187,8	-8,3	
		HLA-B*5701	-762	-914,2	-213,7	-7,7	
		HLA-B*5801	-787,9	-905,3	-210,3	-7,2	
		HLA-B*3501	-794,6	-1044,9	-234,9	-7,9	
		HLA-A*2402	-664,7	-789,9	-229,3	-7,8	
		HLA-B*5801	-730,5	-982,8	-264,4	-7,0	
	MHC-II	Envelope	HLA-A*0206	-754,5	-949,9	-255,2	-7,1
			HLA-B*1501	-752,5	-1016,6	-242,4	-6,6
			HLA-DRB1*04:01	-695,2	-806,9	-275,3	-7,2
		Membrane	HLA-DRB1*15:01	-847,4	-911,3	-273,2	-6,9
			HLA-DRB1*01:01	-788,7	-917,0	-276,5	-5,8
			HLA-DRB1*04:01	-743,9	-747,3	-228,1	-6,6
			HLA-DRB1*15:01	-814,2	-814,2	-219,9	-6,3
			HLA-DRB1*01:01	-762,1	-762,1	-230,2	-7,0
			HLA-DRB1*04:01	-697,9	-809,8	-226,4	-8,2
		Nucleocapsid	HLA-DRB1*15:01	-723,0	-858,8	-240,5	-6,5
			HLA-DRB1*01:01	-660,5	-751,0	-214,5	-6,3
			HLA-DRB1*04:01	-722,0	-844,1	-200,8	-7,2
		Surface 1	HLA-DRB1*15:01	-784,1	-857,2	-209,0	-6,7
			HLA-DRB1*01:01	-729,1	-827,8	-201,4	-7,1
			HLA-DRB1*04:01	-770,7	-806,3	-252,1	-7,7
		Surface 2	HLA-DRB1*15:01	-797,3	-996,6	-263,6	-7,7
HLA-DRB1*01:01			-766,7	-888,1	-258,6	-7,4	

glycosylation plays a very important role in the interaction of this protein with the ACE2 receptor because water molecules can disrupt the interaction of the S protein with this receptor. These glycans, like carbohydrates, are quite polar, and in addition to stabilizing this molecule, they work by capturing water molecules on the surface of S glycoprotein as well as capturing water molecules on the surface of ACE2, favoring its interaction with the receptor [48]. However, the S protein is where most variations are found between strains, favoring the evasion of the virus by the host's immune system [49,50]. Based on knowledge of how the infection occurs, which residues are responsible, and which atoms are involved, we can infer the impact of these mutations on the protein's activity.

4.4. The concerns about the SARS-CoV-2 new variants during vaccine development

Viruses mutate constantly, but just a few key mutations can affect in some way, making a variant more virulent or more lethal, and it depends a lot on where this mutation occurs [51]. One of the two selected epitopes in this work is not presented in the described variants, while the other, epitope 1, is related to a variant T7161 and contains a n-glycan site (Fig. 7). If a mutation with changes in the amino acid sequence occurred in the same region of a selected epitope, there would be a loss of efficiency of the same, or even a loss of functionality of the immune system to efficiently recognize the pathogen. The n-glycan site may worsen the antibody interaction to neutralize the virus. However, it is not hampering the cellular response against the infection. Among the variants described the ones that cause the most concern is related to changes in the amino acid sequence of the S glycoprotein. The US Department of

Health and Human Services (HHS) created the Sars-CoV-2 Interagency Group (SIG), formed by the largest agencies for Disease Control and Prevention (CDC). The GIS is responsible for defining the classification of the new coronavirus mutations, observing the characteristics that a given mutation can generate; these three groups are the variant of interest (VOI), variant of concern (VOC), and variant of high consequence (VOHC). For a mutation to qualify for the VOI classification, it must have at least one of the characteristics, such as specific genetic markers, that lead the virus to increase its degree of transmission or change the escape mechanism to the immune system or any evidence to prove that it is responsible by an expansion of contamination. Currently, seven strains fall into this category, and they are Iota (United States), B.1.526.1 (United States), Eta (Nigeria/United Kingdom), B.1.617 (India), Kappa (India), B.1.617.3 (India), and Zeta (Brazil). To define a mutation as VOC, it must have, in addition to the attributes mentioned for VOI, any evidence that proves increased disease severity, impact on vaccines, treatments, diagnoses, increased transmissibility, and reduced vaccine effect among others. Currently, seven strains are present in this category: Alpha (United Kingdom), Gamma (Japan/Brazil), Delta (India), Beta (South Africa), Epsilon (B.1.427 and B.1.429; United States), and Omicron (South Africa). And finally, for a mutation to be characterized as VOHC, in addition to having all the characteristics mentioned for VOC, it must have an impact on medical countermeasures (MCM), evidence of a drastic reduction in the effect of vaccines, and an increase in hospitalizations with worsening clinical conditions; however, to date, no mutations fall into this category [52]. It is noteworthy that the variant VOC B.1.351 from South Africa was the object of a study for researchers at the Faculty of Medicine of São Paulo (FM-USP) in which the 417N mutation responsible for replacing the amino acid lysine with asparagine is

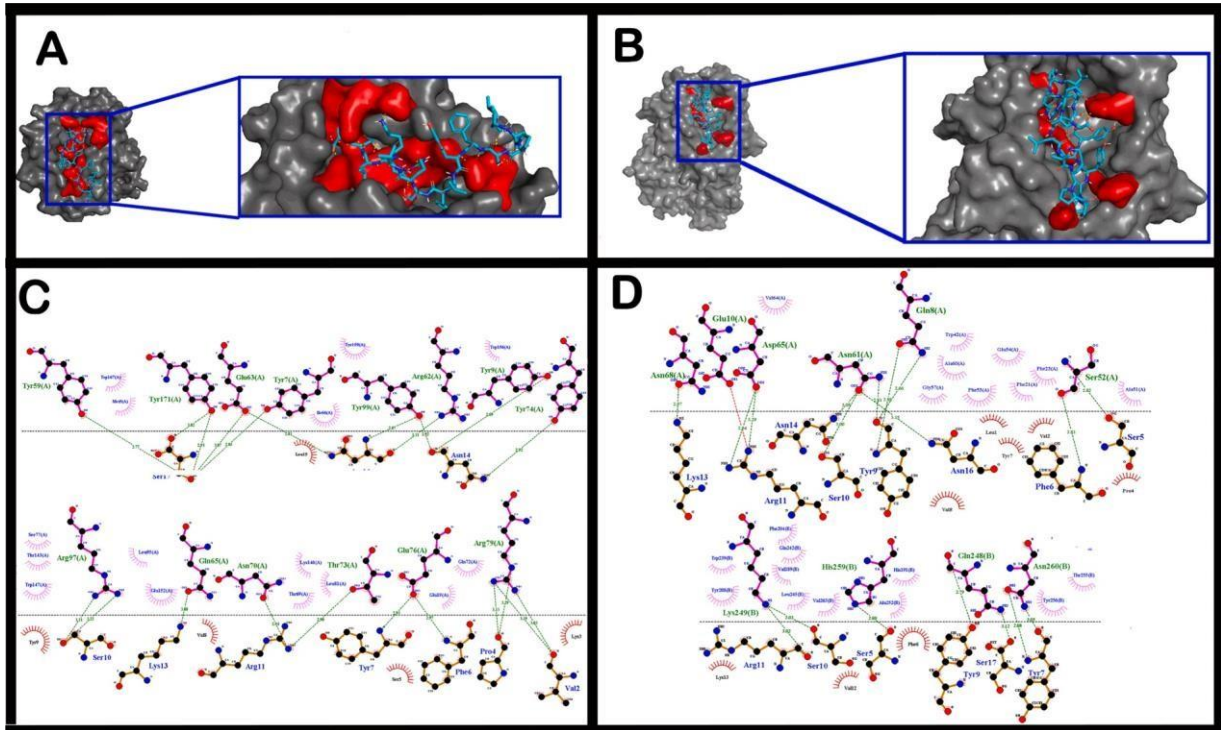


Fig. 2. Interaction between the selected epitope from the consensus sequence of the SARS-CoV-2 envelope protein and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the envelope protein epitope and the MHC-I allele HLA-B*1501(A and C) and MHC-II allele HLA-DRB1*04:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

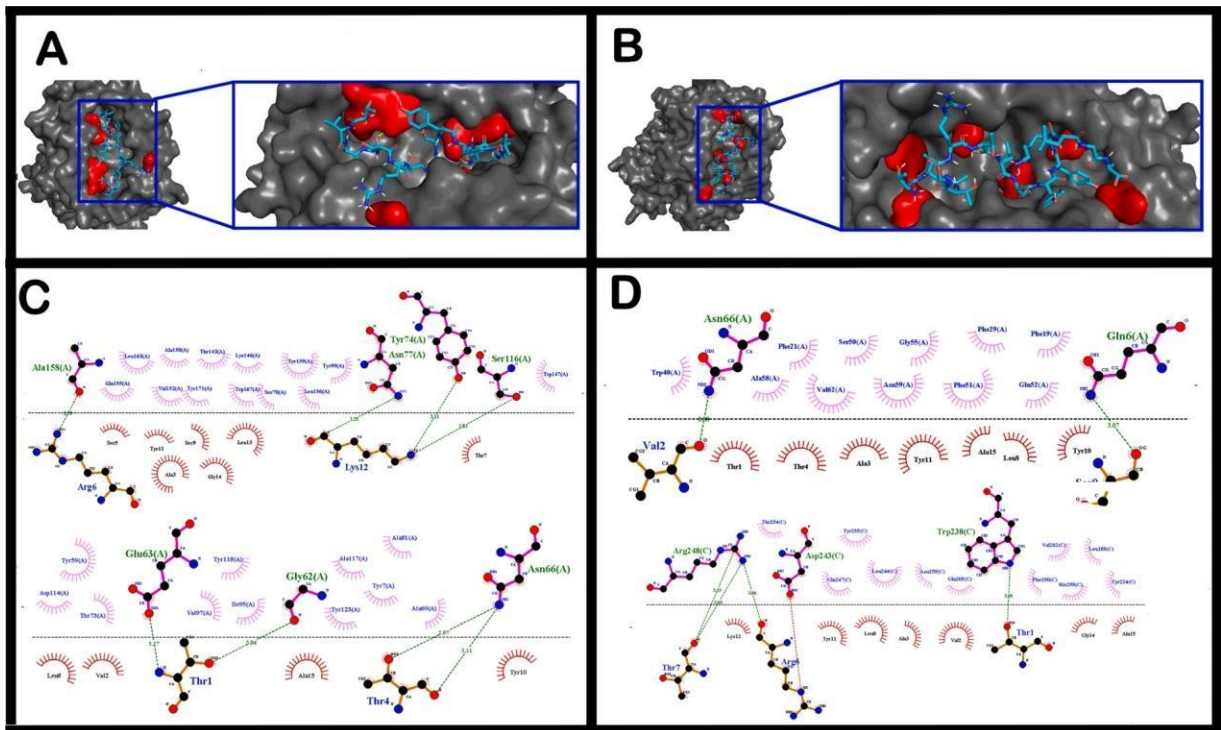


Fig. 3. Interaction between the selected epitope from the consensus sequence of the SARS-CoV-2 membrane glycoprotein and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the membrane glycoprotein epitope and the MHC-I allele HLA-B*57:01 (A and C), and MHC-II allele HLA-DRB1*15:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

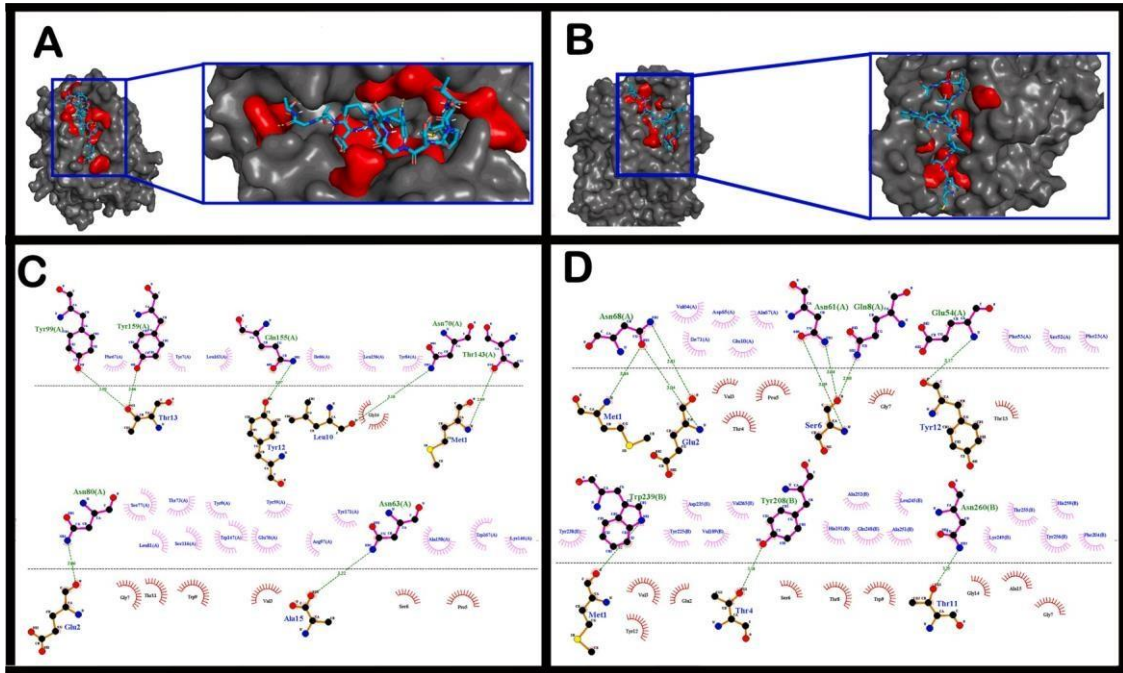


Fig. 4. Interaction between the selected epitope from the consensus sequence of the SARS-CoV-2 nucleocapsid phosphoprotein and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the nucleocapsid phosphoprotein epitope and the MHC-I allele HLA-B*35:01 (A and C), and MHC-II allele HLA-DRB1*04:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

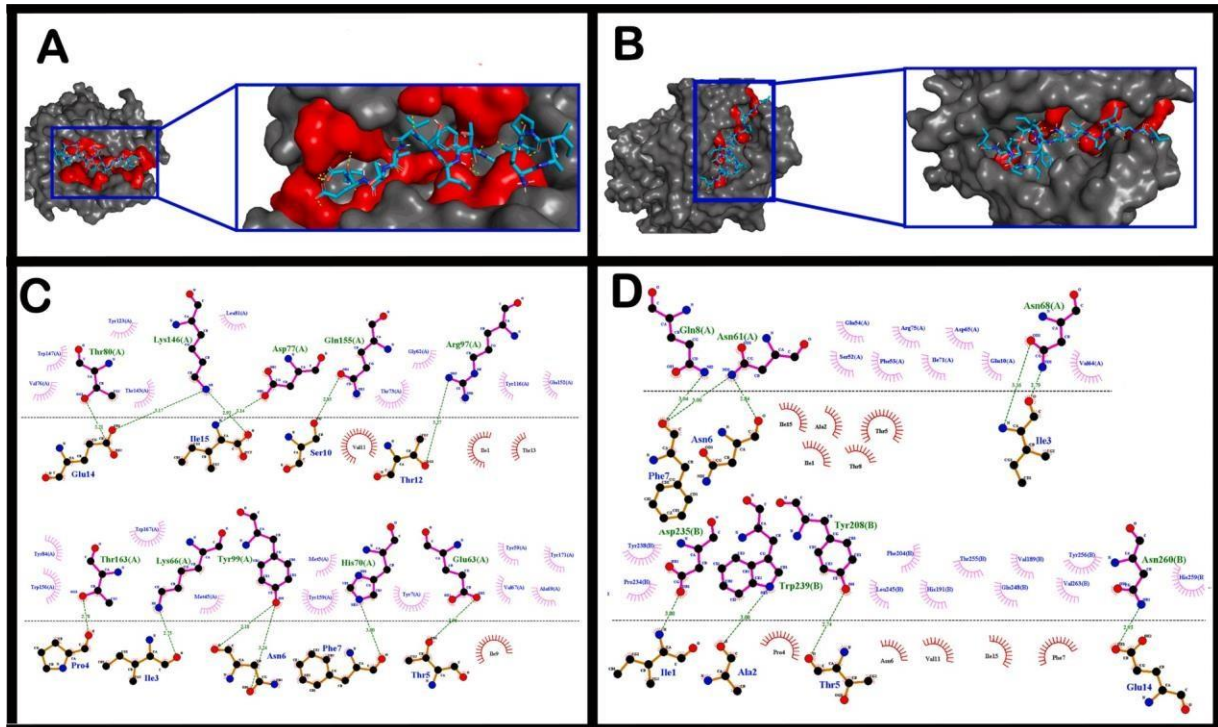


Fig. 5. Interaction between the selected epitope 1 from the SARS-CoV-2 surface glycoprotein consensus sequence and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the surface glycoprotein epitope 1 and the MHC-I allele HLA-A*02:03 (A and C) and MHC-II allele HLA-DRB1*04:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

responsible for glycolysis of proteins making an escape mechanism for the virus, changing its conformation [53]. Currently, the new Indian variant, B.1617, of the new coronavirus raises the concern and warning

signal regarding transmission, lethality, and resistance to available vaccines [54,55]. Using the approaches indicated in this work, peptide-based vaccines facilitate the handling of variants.

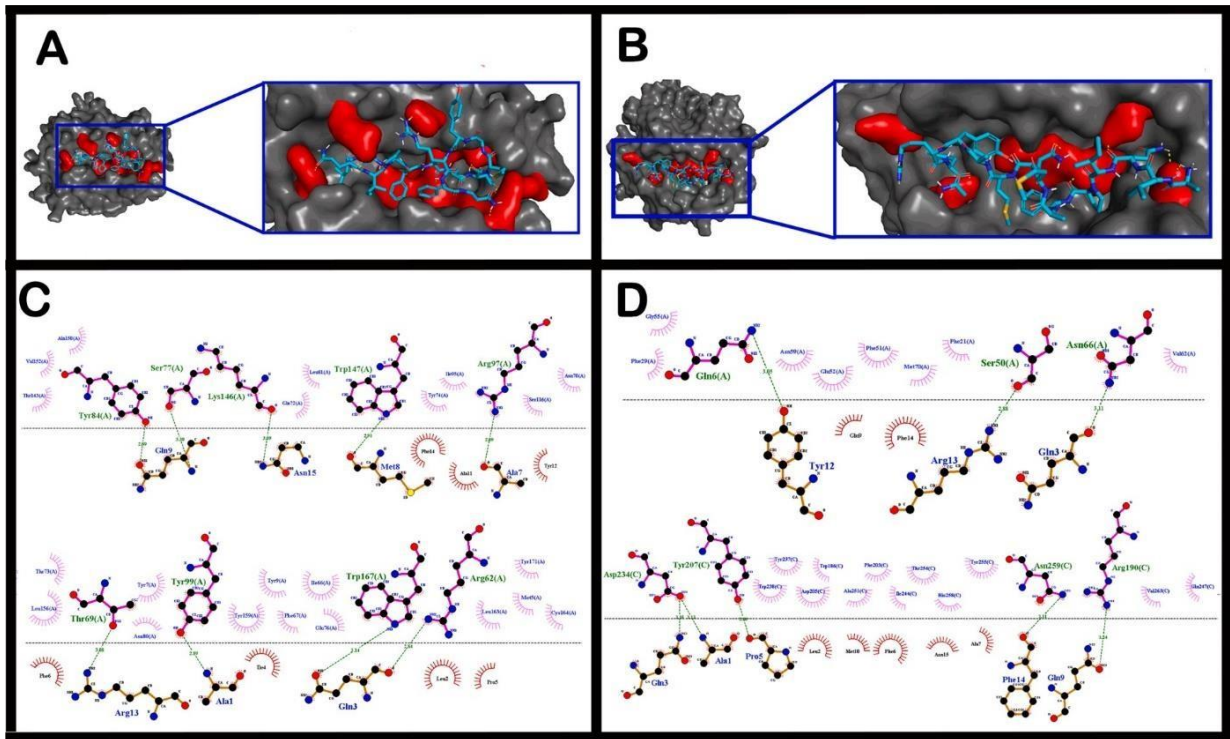


Fig. 6. Interaction between the selected epitope 2 from the SARS-CoV-2 surface glycoprotein consensus sequence and MHC-I and MHC-II alleles. The best interaction evaluated by molecular docking between the surface glycoprotein epitope 2 and the MHC-I allele HLA-B*35:01 (A and C), and MHC-II allele HLA-DRB1*15:01 (B and D). The points of contact are demonstrated in A and B in red, and the points of interaction between the epitope and MHC molecules are 2D represented in C and D, respectively.

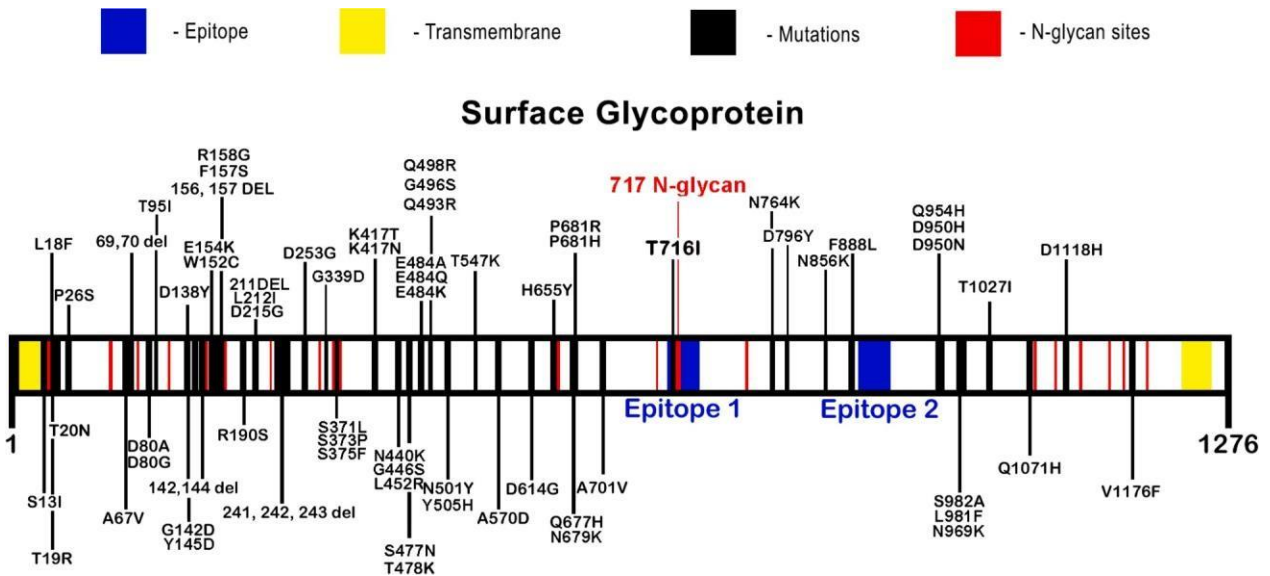


Fig. 7. SARS-CoV-2 surface glycoprotein variants. The described variants of SARS-CoV-2 based on amino acid changes in the surface glycoprotein indicate a variety of points across the protein indicated by a red mark. The selected epitopes are highlighted in blue. N-glycan sites are highlighted in green. The selected epitope 1 lies within a described variant T716I and a putative n-glycan site, while epitope 2 is not included within any yet-described variant.

4.5. Handling of a peptide-based putative vaccine against SARS-CoV-2 rationally selected by immunoinformatic

The only way to perform the management of a vaccine with the complete protein sequence is to change the sequence itself as a whole. On the other hand, peptide vaccines, which are the small regions of the protein that have characteristics of an immune response, can be managed in two ways: using a pool of peptides that escape and are not

within the most frequent areas that present mutation and replacing the protein or in the advent of the emergence of a new variant where that substitution is precisely in the peptide used and can perform only that substitution [56]. Besides, generating multiple target sites as other structural proteins of a specific pathogen may increase the possibilities for the host's immune system to recognize and eliminate the invader. The results presented in this work also identify the most likely epitope, using a variety of established bioinformatic tools with different accuracy

level, to be immunogenic in the envelope protein (E), membrane glycoprotein (M), or nucleocapsid phosphoprotein (N). It is important to note that variants in these proteins are not described as more transmissible, lethal, or resistant to vaccines. All epitopes were evaluated to their potential to interact with immune system molecules and the security to be used in humans since they are not allergenic but antigenic with a good half-life in mammals, yeast, or *E. coli*. Finally, the results obtained by our group are based on a relatively reduced cost and time compared to traditional techniques, but it is necessary to evaluate the action of selected peptides associated with adjuvants to functionally define them as immunogenic and possible candidates to be used as a peptide-based vaccine. Lee and colleagues (2021), remarks the importance of different approaches to correlate the immunogenicity prediction of MHC-bound peptides, including large scale *in vitro* or *in vivo* evaluation of selected epitopes to increase the confidence of *in silico* epitope selection [57]. Vaccination is the most effective and safest method of creating an immune barrier, capable of breaking the SARS-CoV-2 transmission and preventing the most serious pathology of the disease. When the effective vaccination campaign is combined with other preventive methods of social distancing and the use of a mask, over time, there is a significant reduction in the number of cases of the disease, even though the social life of its population returns [58].

CRedit authorship contribution statement

Leonardo Pereira de Araújo: Conceptualization, Data curation, Methodology, Writing – original draft. **Maria Eduarda Carvalho Dias:** Data curation, Methodology, Formal analysis, Writing – original draft. **Gislaine Cristina Scodeler:** Data curation, Methodology, Formal analysis, Writing – original draft. **Ana de Souza Santos:** Data curation, Methodology, Formal analysis, Writing – original draft. **Letícia Martins Soares:** Data curation, Methodology, Formal analysis. **Patrícia Paiva Corsetti:** Data curation, Methodology, Formal analysis, Supervision. **Ana Carolina Barbosa Padovan:** Data curation, Methodology, Formal analysis, Supervision. **Nelson José de Freitas Silveira:** Data curation, Methodology, Formal analysis, Supervision. **Leonardo Augusto de Almeida:** Conceptualization, Data curation, Formal analysis, Writing – original draft, Project administration, Resources, Supervision.

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (Finance Code 001), Universidade Federal de Alfenas and the Brazilian Ministry of Education (MEC) (EDITAL N° 002/2020 - PRPPG/REITORIA), and by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), process 115462/2020-1).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.immuno.2022.100015.

References

- WORLD HEALTH ORGANIZATION. WHO announces COVID-19 outbreak a pandemic. March 12, 2020. <https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic> (accessed 19 July 2021). Reference to a dataset.
- WU Di, et al. The SARS-CoV-2 outbreak: what we know. *Int J Infect Dis* 2020;94: 44–8. <https://doi.org/10.1016/j.ijid.2020.03.004>.
- PETERSEN Eskild, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis* 2020. [https://doi.org/10.1016/S1473-3099\(20\)30484-9](https://doi.org/10.1016/S1473-3099(20)30484-9).
- Yao XH, Luo T, et al. A cohort autopsy study defines COVID-19 systemic pathogenesis. *Cell Res* 2021 Jun 16. <https://doi.org/10.1038/s41422-021-00523-8>.
- UL QAMAR, Tahir Muhammad, et al. Reverse vaccinology assisted designing of multi-epitope-based subunit vaccine against SARS-CoV-2. *Infect Dis Poverty* 2020;9(1):1–14. <https://doi.org/10.1186/s40249-020-00752-w>.
- MOORE Sam, et al. Modelling optimal vaccination strategy for SARS-CoV-2 in the UK. *PLoS Comput Biol* 2021;17(5):e1008849. <https://doi.org/10.1371/journal.pcbi.1008849>.
- ABRAHAM PEELE K, et al. Design of multi-epitope vaccine candidate against SARS-CoV-2: a *in-silico* study. *J Biomol Struct Dyn* 2020:1–9. <https://doi.org/10.1080/07391102.2020.1770127>.
- OLI Angus Nnamdi, et al. Immunoinformatics and vaccine development: an overview. *Immuno Targets Therapy* 2020;9(13). <https://doi.org/10.2147/ITT.S241064>. PMID: 32161726.
- SOHAIL Muhammad Saqib, et al. *In silico* T cell epitope identification for SARS-CoV-2: progress and perspectives. *Adv Drug Deliv Rev* 2021. <https://doi.org/10.1016/j.addr.2021.01.007>.
- SCHIJNS V LAVELLE EC. Prevention and treatment of COVID-19 disease by controlled modulation of innate immunity. *Eur J Immunol* 2020;50(7):932–8. <https://doi.org/10.1002/eji.202048693>. Epub 2020 Jun 15.
- ONG Edison, et al. COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. *Front Immunol* 2020;11:1581. <https://doi.org/10.3389/fimmu.2020.01581>.
- BARRETO-VIEIRA Debora Ferreira, et al. Morphology and morphogenesis of SARS-CoV-2 in Vero-E6 cells. *Memórias do Instituto Oswaldo Cruz* 2021;116. <https://doi.org/10.1590/0074-02760200443>.
- SMITH Christof C, et al. Landscape and selection of vaccine epitopes in SARS-CoV-2. *Genome Med* 2021;13(1):1–23. <https://doi.org/10.1186/s13073-021-00910-1>.
- LEHMANN Alexander A, et al. Deconvoluting the T cell response to SARS-CoV-2: specificity versus chance and cognate cross-reactivity. *Front Immunol* 2021;12. <https://doi.org/10.3389/fimmu.2021.635942>.
- WATERHOUSE Andrew M, et al. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 2009;25(9):1189–91. <https://doi.org/10.1093/bioinformatics/btp033>.
- WEISKOPF Daniela, et al. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. *Proc Natl Acad Sci* 2013;110(22):E2046–53. <https://doi.org/10.1073/pnas.1305271110>.
- SETTE Alessandro, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol* 1994;152(153): 5586–92.
- MOUTAFTSI Magdalini, et al. A consensus epitope prediction approach identifies the breadth of murine T CD8+ cell responses to vaccinia virus. *Nat Biotechnol* 2006;24(7):817–9. <https://doi.org/10.1038/nbt1215>.
- KOTTURI Maya F, et al. The CD8+ T-cell response to lymphocytic choriomeningitis virus involves the L antigen: uncovering new tricks for an old virus. *J Virol* 2007;81(10):4928–40. <https://doi.org/10.1128/JVI.02632-06>.
- JENSEN Kamilla Kjaergaard, et al. Improved methods for predicting peptide binding affinity to MHC class II molecules. *Immunology* 2018;154(3):394–406. <https://doi.org/10.1111/imm.12889>.
- NIELSEN Morten, LUNDEGAARD Claus, LUND Ole. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinf* 2007;8(1):1–12. <https://doi.org/10.1186/1471-2105-8-238>.
- SIDNEY John, et al. Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Res* 2008;4(1):1–14. <https://doi.org/10.1186/1745-7580-4-2>.
- STURNIOLO Tiziana, et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol* 1999;17(6):555–61. <https://doi.org/10.1038/9858>.
- GREENBAUM Jason, et al. Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. *Immunogenetics* 2011;63(6):325–35. <https://doi.org/10.1007/s00251-011-0513-0>.
- OSEROFF Carla, et al. Molecular determinants of T cell epitope recognition to the common Timothy grass allergen. *J Immunol* 2010;185(2):943–55. <https://doi.org/10.4049/jimmunol.1000405>.
- JESPERSEN Martin Closter, et al. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. *Nucleic Acids Res* 2017;45(W1): W24–9. <https://doi.org/10.1093/nar/gkx346>.
- CORPET Florence. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* 1988;16(22):10881–90. <https://doi.org/10.1093/nar/16.22.10881>.
- MITAKU Shigeki, HIROKAWA Takatsugu, TSUJI Toshiyuki. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane–water interfaces. *Bioinformatics* 2002;18(4):608–16. <https://doi.org/10.1093/bioinformatics/18.4.608>.
- DOYTCHEVA Irimi A., FLOWER Darren R. Vaxijen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf* 2007;8(1): 1–7. <https://doi.org/10.1186/1471-2105-8-4>.
- SHARMA N, PATIYAL S, DHALL A, PANDE A, ARORA C, RAGHAVA GPS. AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE

- epitopes. *Brief Bioinform* 2021;22(4):bbaa294. <https://doi.org/10.1093/bib/bbaa294>. Jul 20.
- [31] SAHA Sudipto, RAGHAV A GPS. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res* 2006;34(2):W202-9. <https://doi.org/10.1093/nar/gkl343>. suppl.
- [32] GASTEIGER Elisabeth, et al. EXPASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003;31(13):3784-8. <https://doi.org/10.1093/nar/gkg563>.
- [33] Gupta R, Brunak S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* 2002:310-22.
- [34] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583-9. <https://doi.org/10.1038/s41586-021-03819-2>.
- [35] The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2020;49:D480-9.
- [36] Mitchell AL, et al. MGnify: the microbiome analysis resource in 2020. *Nucleic Acids Res* 2020;48:D570-8.
- [37] Steinegger M, Mirdita M, Söding J. Protein-level assembly increases protein sequence recovery from metagenomic samples manifold. *Nat Methods* 2019;16:603-6. <https://doi.org/10.1038/s41592-019-0437-4>.
- [38] Mirdita M, et al. Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res* 2017;45:D170-6.
- [39] Consortium wwPDB. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res* 2018;47:D520-8.
- [40] WEBB B, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol* 2014;1137:1-15. https://doi.org/10.1007/978-1-4939-0366-5_1.
- [41] XU Xianjin, YAN Chengfei, ZOU Xiaolin. MDockPeP: An ab-initio protein-peptide docking server. *J Comput Chem* 2018;39(28):2409-13. <https://doi.org/10.1002/jcc.25555>.
- [42] PORTER Kathryn A, et al. ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. *Bioinformatics* 2017;33(20):3299-301. <https://doi.org/10.1093/bioinformatics/btx216>.
- [43] TROTT Oleg, OLSON Arthur J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;31(2):455-61. <https://doi.org/10.1002/jcc.21334>.
- [44] Laskowski R A, Swindells M B. LigPlot+ multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model* 2011;51:2778-86.
- [45] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
- [46] HISHAM Yasmin, et al. Identification of highly conserved SARS-CoV-2 antigenic epitopes with wide coverage using reverse vaccinology approach. *Viruses* 2021;13(5):787. <https://doi.org/10.3390/v13050787>.
- [47] SINGH Nidhi, et al. Molecular characterization, pathogen-host interaction pathway and in silico approaches for vaccine design against COVID-19. *J Chem Neuroanat* 2020;101874. <https://doi.org/10.1016/j.jchemneu.2020.101874>.
- [48] NAVEED Muhammad, et al. Design of a novel multiple epitope-based vaccine: an immunoinformatics approach to combat SARS-CoV-2 strains. *J Infect Public Health* 2021. <https://doi.org/10.1016/j.jiph.2021.04.010>.
- [49] WALLS Alexandra C, et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020;181(2):281-92. <https://doi.org/10.1016/j.cell.2020.02.058>. e6.
- [50] HARVEY William T, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat Rev Microbiol* 2021;19(7):409-24. <https://doi.org/10.1038/s41579-021-00573-0>.
- [51] VERKHIVKER Gennady, et al. Dynamic profiling of binding and allosteric propensities of the SARS-CoV-2 spike protein with different classes of antibodies: mutational and perturbation-based scanning reveals the allosteric duality of functionally adaptable hotspots. *J Chem Theory Comput* 2021. <https://doi.org/10.1021/acs.jctc.1c00372>.
- [52] NELSON Gard, et al. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y. V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. *bioRxiv* 2021.
- [53] CDC, Centers For Disease Control And Prevention. SARS-CoV-2 Variant Classifications and Definitions. 2021. <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html#print> (access: 20 april 2021) Reference to a dataset.
- [54] SANTOS Keity Souza, et al. Immunodominant B cell epitope in SARS-CoV-2 RBD comprises a B. 1.351 and P. 1 mutation hotspot: implications for viral spread and antiescape. *medRxiv* 2021.
- [55] QUINONEZ Elena, et al. Structural analysis of the novel variants of SARS-CoV-2 and forecasting in North America. *Viruses* 2021;13(5):930. <https://doi.org/10.3390/v13050930>.
- [56] WALL Emma C, et al. Neutralising antibody activity against SARS-CoV-2 VOCs B. 1.617. 2 and B. 1.351 by BNT162b2 vaccination. *Lancet North Am Ed* 2021;397(10292):2331-3. [https://doi.org/10.1016/S0140-6736\(21\)01290-3](https://doi.org/10.1016/S0140-6736(21)01290-3).
- [57] CHAUHAN Nidhi, et al. Interpretative immune targets and contemporary position for vaccine development against SARS-CoV-2: A systematic review. *J Med Virol* 2021;93(4):1967-82. <https://doi.org/10.1002/jmv.26709>.
- [58] LEE Chloe H, et al. To what extent does MHC binding translate to immunogenicity in humans? *Immuno Informatics* 2022. <https://doi.org/10.1016/j.immuno.2021.100006>.

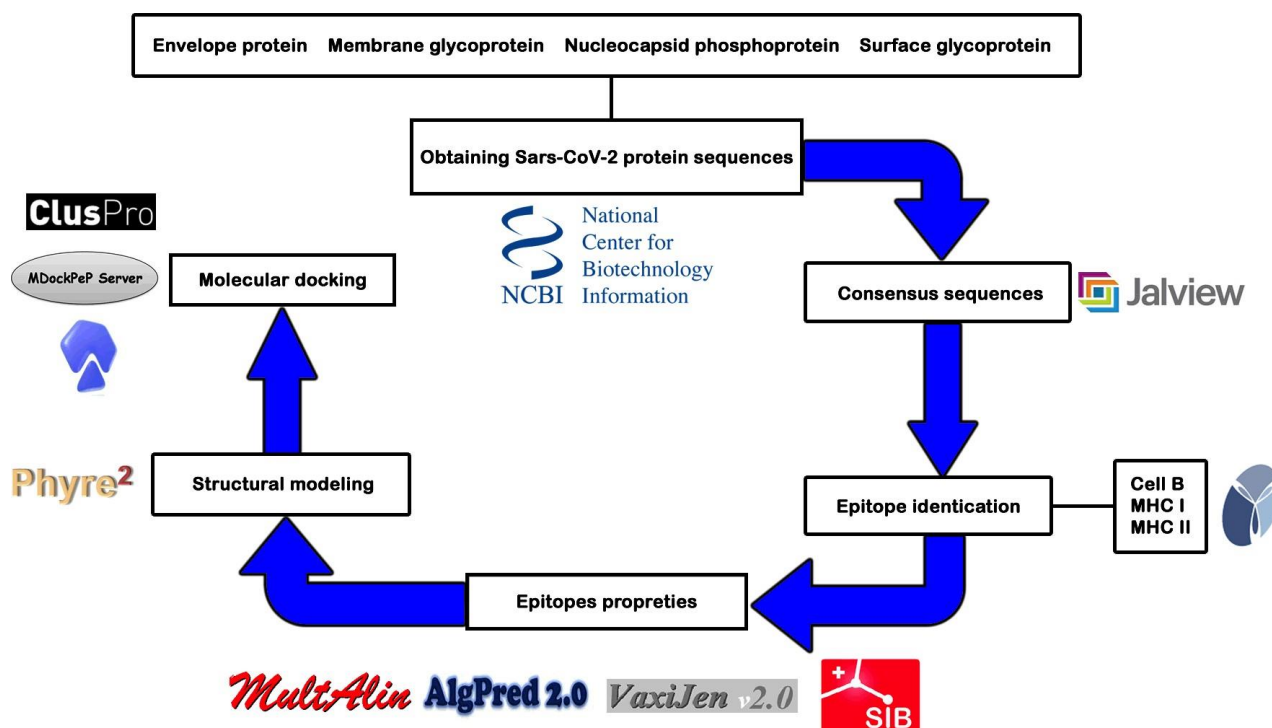
ANEXO IV

Supplementary Figure 1. The pipeline of methods used in the rational selection of SARS-CoV-2 epitopes presented in consensus structural proteins. Flowchart of the steps taken to obtain promiscuous epitopes and their respective software.

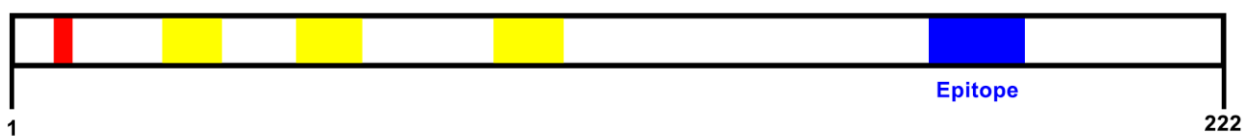
Supplementary Figures 2-5. Alignment of consensus amino acid sequences from structural proteins obtained for SARS-CoV-2 genomes from the six continents deposited in the database to reach a global consensus sequence. The amino acid sequences of the envelope protein (Supp. 2), membrane glycoprotein (Supp. 3), nucleocapsid phosphoprotein (Supp. 4), and surface glycoprotein (Supp. 5) are presented according to the consensus sequence obtained in each continent following the alignment consensus sequence of the global sequence. The conserved amino acids are represented in red and the variations in blue and lowercase.

Supplementary Figure 6. Epitopes, transmembrane and n-glycan sites of SARS-CoV-2 structural proteins. Each consensus structural protein sequence of the new coronavirus was submitted to identification of putative epitopes (blue), and transmembrane (yellow), and n-glycan (green) sites. In the S protein it is indicated the points of mutations described in variants of this virus.

Supplementary Figure 1



Supplementary Figure 6

**Envelope Protein****Membrane Glycoprotein****Nucleocapsid Phosphoprotein**