

UNIVERSIDADE FEDERAL DO TRIÂNGULO MINEIRO
PROGRAMA DE PÓS-GRADUAÇÃO EM MEDICINA TROPICAL E
INFECTOLOGIA

ANDREI GIACCHETTO FELICE

Construção *in silico* de uma vacina multi-epítopo contra o gênero
***Rickettsia* através de imunoinformática**

Uberaba

2023

ANDREI GIACCHETTO FELICE

Construção *in silico* de uma vacina multi-epítopo contra o gênero
Rickettsia através de imunoinformática

Dissertação apresentada ao Programa de Pós-Graduação em Medicina Tropical e Infectologia, ênfase em Imunologia e Parasitologia Aplicadas, da Universidade Federal do Triângulo Mineiro, como requisito para obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Siomar de Castro
Soares

Uberaba

2023

**Catálogo na fonte: Biblioteca da Universidade Federal do
Triângulo Mineiro**

F347c Felice, Andrei Giacchetto
Construção *in silico* de uma vacina multi-epítopo contra o gênero
Rickettsia através de imunoinformática / Andrei Giacchetto Felice. -- 2023.
71 p. : il., graf., tab.

Dissertação (Mestrado em Medicina Tropical e Infectologia) – Univer-
sidade Federal do Triângulo Mineiro, Uberaba, MG, 2023
Orientador: Prof. Dr. Siomar de Castro Soares

1. Vacinas. 2. *Rickettsia*. 3. Proteínas. 4. Biologia computacional.
I. Soares, Siomar de Castro. II. Universidade Federal do Triângulo Mineiro.
III. Título.

CDU 615.371

ANDREI GIACCHETTO FELICE

Construção *in silico* de uma vacina multi-epítopo contra o gênero
Rickettsia através de imunoinformática

Dissertação apresentada ao Programa de Pós-Graduação em Medicina Tropical e Infectologia, ênfase em Imunologia e Parasitologia Aplicadas, da Universidade Federal do Triângulo Mineiro, como requisito para obtenção do título de Mestre em Ciências.

24 de janeiro de 2023.

Banca examinadora:

Prof. Dr. Siomar de Castro Soares – Orientador
Universidade Federal do Triângulo Mineiro

Prof. Dr. André Luiz Pedrosa
Universidade Federal do Triângulo Mineiro

Prof. Dr. Sandeep Tiwari
Universidade Federal da Bahia

AGRADECIMENTOS

Agradeço primeiramente à Deus, pela saúde e por sempre me guiar e iluminar todos os meus passos até hoje.

À minha família, meu pai Renato, minha mãe Selma, minha irmã Jéssica, minha avó Elza e meu cunhado Renan, por estarem ao meu lado em todos os momentos, me apoiarem sempre em todas as minhas decisões, sempre acreditando no meu potencial. Aos demais familiares, obrigado por tudo.

À minha namorada Victoria, por também estar ao meu lado, todos os dias, me dando sempre suporte emocional e deixar os meus dias mais felizes.

Aos meus amigos de Colina/SP, de Uberaba/MG e da República LisosSomos por sempre estarem comigo me apoiando.

Agradeço ao meu orientador, Prof. Dr. Siomar de Castro Soares por estar nessa jornada comigo desde a minha primeira iniciação científica, sempre me encorajando a aprender novas coisas. Agradeço por toda a paciência, responsabilidade e por deixar esse caminho mais leve e divertido.

Aos meus colegas do Laboratório de Bioinformática da UFTM e do Laboratório de Genética Celular e Molecular da UFMG, por sempre me ajudarem em minhas análises e sempre apoiar possibilitando sempre expandir novos horizontes.

Ao Programa de Pós-graduação em Medicina Tropical e Infectologia e à Universidade Federal do Triângulo Mineiro por realizar mais esse passo na minha carreira. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), a Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo todo o apoio financeiro.

À todas as pessoas que ajudaram direta ou indiretamente na contribuição para a realização deste trabalho.

RESUMO

O gênero *Rickettsia*, pertencente ao filo das Proteobacterias, é composto principalmente por quatro grupos de microrganismos: o grupo tifo, o grupo *Spotted fever*, grupo ancestral e o grupo transicional. Essas espécies são causadoras de doenças de importância para a saúde pública como por exemplo tifo epidêmico, tifo murino, febre maculosa das Montanhas Rochosas e febre maculosa do Mediterrâneo. Apesar do aumento no número de casos, ainda não existe uma vacina contra as doenças causadas por esse gênero. Este trabalho visa o desenvolvimento *in silico* de uma vacina multi-epítopo com potencial imune contra o gênero *Rickettsia*. Então, oito proteínas previamente identificadas como potenciais candidatas a vacinais contra essas bactérias, passaram por um processo de predição de epítopos ligantes às moléculas do complexo de histocompatibilidade, do tipo I e II, e depois por análises de qualidade quanto à sua antigenicidade, alergenicidade e toxicidade. Aqueles que passaram por essa filtragem foram unidos com *linkers* e um adjuvante para formar uma única proteína quimérica multi-epítopo, a qual passou pela predição da estrutura tridimensional e posterior análise de *docking* molecular contra o receptor do tipo *Toll Like-4*, presente nas células do hospedeiro. Por fim, a proteína foi testada quanto a simulação da sua resposta imune através de *machine learning*. Um total de vinte e seis epítopos imunogênicos, sendo 13 ligantes para cada tipo das moléculas do complexo de histocompatibilidade, foram ligados para a construção da proteína multi-epítopo. A proteína apresentou ótimos parâmetros imunogênicos e não foi considerada tóxica e nem alérgica ao hospedeiro. Além disso, apresentou uma boa estimulação de células imunes, com potencial geração de células de memória. As validações *in silico* sugerem uma vacina com eficácia contra a ação dessas bactérias e devem ser validadas *in vitro*.

Palavras-chave: *Rickettsia*; proteínas; bioinformática; vacinas.

ABSTRACT

The genus *Rickettsia*, belonging to the phylum Proteobacteria, is mainly composed of four groups of microorganisms: the typhus group, the spotted fever group, the ancestral group and the transitional group. These species cause diseases of public health importance, for example epidemic typhus, murine typhus, Rocky Mountain spotted fever, and Mediterranean spotted fever. Despite the increasing number of cases, there is still no vaccine against the diseases caused by this genus. This work aims at the *in silico* development of a multi-epitope vaccine with immune potential against the genus *Rickettsia*. Then eight proteins previously identified as potential vaccine candidates against these bacteria, went through a process of prediction of epitopes binding to histocompatibility complex molecules, type I and II, and then by quality analysis as to their antigenicity, allergenicity, toxicity. Those that passed this screening were joined with linkers and an adjuvant to form a chimeric multi-epitope protein, which underwent three-dimensional structure prediction and subsequent molecular docking analysis against the Toll Like-4 receptor present on host cells. Finally, the protein was tested for an immune response simulation using machine learning. A total of twenty-six immunogenic epitopes with 13 ligands for each type of histocompatibility complex molecule, were linked to construct the multi-epitope protein. The protein showed excellent immunogenic parameters and was not considered toxic or allergic to the host. In addition, it showed good stimulation of immune cells, with potential generation of memory cells. The *in silico* validations suggest a vaccine with efficacy against the action of these bacteria and must be validated *in vitro*.

Keywords: *Rickettsia*; proteins; bioinformatic; vaccines.

LISTA DE ILUSTRAÇÕES

Figura 1 - Gráfico representativo da sequência da proteína multi-epítopo.	32
Figura 2 - Representação das estruturas secundária e terciária da proteína multi-epítopo.....	33
Figura 3 - Plot da interação da proteína multi-epítopo com o receptor TLR4 e suas interações.	35
Figura 4 - Gráficos da simulação dinâmica para o complexo proteína multi-epítopo-ligante.	36
Figura 5 - Simulação da resposta imune <i>in silico</i> gerada pela proteína multi-epítopo em relação às células T, <i>natural killer</i> , dendríticas, macrófagos e células epiteliais.....	38
Figura 6 - Simulação da resposta imune <i>in silico</i> gerada pela proteína multi-epítopo em relação às células B, citocinas e imunoglobulinas.	39
Figura 7 - Simulação da resposta imune <i>in silico</i> gerada somente pelo adjuvante em relação às células T, <i>natural killer</i> , dendríticas, macrófagos e células epiteliais.	41
Figura 8 - Simulação da resposta imune <i>in silico</i> gerada somente pelo adjuvante em relação às células B, citocinas e imunoglobulinas.	42

LISTA DE TABELAS

Tabela 1 - Informações das oito proteínas potencialmente candidatas à vacina contra o gênero <i>Rickettsia</i>	24
Tabela 2 - Epítopos selecionados para a construção da proteína multi-epítipo.....	30
Tabela 3 - Etapa de filtragem de epítopos para o complexo MHC I/CTL de cada uma das 8 proteínas utilizadas nos programas IEDB e NetCTL realizada pelo programa Microsoft Excel.	57
Tabela 4 - Etapa de filtragem de epítopos para o complexo MHC II/HTL de cada uma das 8 proteínas utilizadas nos programas IEDB e NetCTL realizada pelo programa Microsoft Excel.	57
Tabela 5 - Epítopos finais preditos pelo programa ABCpred para cada uma das 8 proteínas.	58
Tabela 6 - Técnica de overlap entre os complexos MHC I/CTL e MHC II/HTL com células B para cada uma das proteínas utilizadas.....	58
Tabela 7 - Predição de epítopos lineares de células B.....	60
Tabela 8 - Predição de epítopos descontínuos de células B.	60
Tabela 9 - Predição de estimulação de Interferon dos epítopos da proteína multi-epítipo.....	62

LISTA DE SIGLAS

- AG: Grupo Ancestral, do inglês, *Ancestral Group*.
- ANN: Rede Neural Artificial, do inglês, *Artificial Neural Network*.
- BZD: Doença de Brill-Zinsser.
- CTL: Linfócitos T Ritotóxicos, do inglês, *Cytotoxic T lymphocytes*.
- DNA: Ácido Desoxirribonucleico
- EUA: Estados Unidos da América.
- HTL: Linfócitos T Auxiliares, do inglês, *Helper T lymphocytes*.
- IC50: Metade da concentração inibitória mínima, do inglês, *Half maximal inhibitory concentration*,
- IEDB: Banco de Dados de Epítomos Imunes, do inglês, *Immune Epitope Database*.
- IFA: Ensaio de Imunofluorescência Indireta.
- IFN: Interferon.
- IgG: Imunoglobulina G.
- IgM: Imunoglobulina M.
- IL-17: Interleucina 17.
- MHC: Moléculas do Complexo de Histocompatibilidade.
- MSF: Febre Maculosa do Mediterrâneo, do inglês, *Mediterranean Spotted Fever*.
- NCBI: Centro Nacional de Informação em Biotecnologia, do inglês, *National Center for Biotechnology Information*.
- OMPs: Proteínas de membrana externa, do inglês, *Outer Membrane Protein*.
- OMS: Organização Mundial da Saúde.
- PCR: Reação em Cadeia da Polimerase.
- PDB: Banco de Dados de Proteína, do inglês, *Protein Data Bank*.
- pI: Ponto isoeletrico.
- PSE: Possivelmente Expostas à Superfície.
- RMSD: Raiz Quadrada do Desvio Quadrático Médio, do inglês, *Root-Mean-Square Deviation*.
- RMSF: Febre das Montanhas Rochosas, do inglês, *Rocky Mountain Spotted Fever*.
- RNA: Ácido Ribonucleico.
- SFG: Grupo Febre Maculosa, do inglês, *Spotted Fever Group*.
- SVM: Máquina de Vetores de Suporte, do inglês, *Support Vector Machine*.
- TG: Grupo Tifo, do inglês, *Typhus Group*.
- TLR: Receptor do Tipo Toll, do inglês, *Toll Like Receptor*.
- VHB: Vacina de Hepatite B.

SUMÁRIO

1	INTRODUÇÃO	13
1.1	GÊNERO <i>Rickettsia</i>	13
1.1.1	Classificação	13
1.1.2	Fisiopatologia	13
1.1.3	Epidemiologia	14
1.1.4	Diagnóstico	16
1.1.5	Tratamento	17
1.1.6	Prevenção	17
1.2	VACINAS	18
1.3	BIOINFORMÁTICA	18
1.4	IMUNOINFORMÁTICA.....	19
1.4.1	Trabalhos com abordagens multi-epítipo	20
2	JUSTIFICATIVA	22
3	OBJETIVOS	23
3.1	OBJETIVOS GERAIS	23
3.2	OBJETIVOS ESPECÍFICOS	23
4	MATERIAIS E MÉTODOS	24
4.1	SELEÇÃO DOS DADOS	24
4.2	PREDIÇÃO DE EPÍTOPOS QUE SE LIGAM A ALELOS MHC-I.....	24
4.3	PREDIÇÃO DE EPÍTOPOS QUE SE LIGAM A ALELOS MHC-II.....	25
4.4	PREDIÇÃO DE EPÍTOPOS QUE SE LIGAM A ALELOS DE CÉLULAS B.....	25
4.5	SELEÇÃO DOS MELHORES EPÍTOPOS	26
4.6	CONSTRUÇÃO DA PROTEÍNA MULTI-EPÍTOPO.....	27
4.7	PREDIÇÃO DAS ESTRUTURAS SECUNDÁRIA, TERCIÁRIA E DE EPÍTOPOS CONFORMACIONAIS DE CÉLULAS B	27
4.8	AVALIAÇÃO DA PROTEÍNA MULTI-EPÍTOPO E DA INTERAÇÃO COM O HOSPEDEIRO	28

4.9	<i>DOCKING</i> MOLECULAR E DINÂMICA MOLECULAR NA INTERAÇÃO PROTEÍNA-RECEPTOR	28
4.10	AVALIAÇÃO DA POSSÍVEL RESPOSTA IMUNE GERADA NO HOSPEDEIRO.....	28
5	RESULTADOS	30
5.1	OS MELHORES EPÍTOPOS PREDITOS PARA MHCI/CTL, MHCII/HTL E CÉLULAS B	30
5.2	NA CONSTRUÇÃO DA PROTEÍNA MULTI-EPÍTOPO FORAM UTILIZADOS VINTE E SEIS EPÍTOPOS.....	31
5.3	A ESTRUTURA 3D DA PROTEÍNA MULTI-EPÍTOPO APRESENTOU MAIS DE 90% DE AMINOÁCIDOS EM REGIÕES MAIS FAVORECIDAS.....	32
5.4	A PROTEÍNA MULTI-EPÍTOPO APRESENTOU ÓTIMOS PARÂMETROS IMUNOGÊNICOS	33
5.5	A PROTEÍNA APRESENTOU ALTA AFINIDADE E BOA DINÂMICA MOLECULAR QUANDO SE LIGA AO RECEPTOR TLR4.....	34
5.6	UMA VACINA CANDIDATA COM A PROTEÍNA MULTI-EPÍTOPO APRESENTARIA UMA BOA RESPOSTA IMUNE	36
6	DISCUSSÃO	43
7	CONCLUSÃO	47
	REFERÊNCIAS	48
	ANEXO A – Filtragem e <i>overlap</i> das proteínas	57
	ANEXO B – Epítomos conformacionais de células B	60
	ANEXO C – Predição do IFNepitope	62

1 INTRODUÇÃO

1.1 GÊNERO *Rickettsia*

O gênero *Rickettsia* pertence à classe das alfa-Proteobacterias e os microrganismos representativos desse gênero são caracterizados como seres *Gram*-negativos intracelulares obrigatórios, ou seja, necessitam ocupar uma célula viva para continuar seu processo de replicação, e têm como alvo principal as células endoteliais (LIMA et al., 2020; SCHROEDER et al., 2015). As rickettsias são patógenos considerados emergentes e residem em uma ampla gama de vetores artrópodes, como pulgas, carrapatos, ácaros e piolhos (PAROLA et al., 2013).

1.1.1 Classificação

Tradicionalmente, as rickettsias são classificadas em dois principais grupos que causam doenças nocivas aos seres humanos. O primeiro agrupa espécies causadoras do tifo, o *Typhus Group* (TG), doença letal sem tratamento, representadas por espécies como *Rickettsia prowazekii*, causadora do tifo exantemático epidêmico, a qual tem como seu vetor o piolho do corpo humano, chamado de *Pediculus humanus corporis*; e *Rickettsia typhi*, causadora do tifo murino, que tem como vetor principalmente a pulga dos ratos, *Xenopsylla cheopis*. O segundo grupo é chamado de *Spotted Fever group* (SFG), que agrupa principalmente espécies como *Rickettsia rickettsii*, causadora da febre das Montanhas Rochosas (*RMSF, do inglês, Rocky Mountain Spotted Fever*) e da febre maculosa brasileira, que tem como seu vetor carrapatos do gênero *Amblyomma*; e *Rickettsia conorii*, agente da febre maculosa do Mediterrâneo (*MSF, do inglês, Mediterranean Spotted Fever*) ou febre botonosa, que tem como vetor carrapatos do gênero *Rhipicephalus* (BLANTON, 2019; GALVÃO et al., 2005; GILLESPIE et al., 2007; LIMA et al., 2020).

Há ainda alguns organismos desse gênero que foram classificados como pertencentes a um grupo chamado de ancestral (AG), como *Rickettsia bellii* e *Rickettsia canadensis*, que apesar de já terem sido classificados como pertencentes a um determinado grupo, filogenias demonstram serem espécies ancestrais para os grupos TG e SFG (PHILIP et al., 1983; STOTHARD; CLARK; FUERST, 1994; STOTHARD; FUERST, 1995). Por fim, quando o grupo AG é adicionado para ser raiz nas análises filogenéticas desse gênero, espécies como *Rickettsia akari*, *Rickettsia felis* e *Rickettsia australis*, as quais possuem características dos grupos TG e SFG, costumam se aproximar e formar o quarto grupo, chamado de grupo transicional (BLANTON, 2019; GILLESPIE et al., 2007).

1.1.2 Fisiopatologia

1
2 Dados encontrados na literatura trazem que essas bactérias realizam um processo de
3 transmissão vertical nos vetores, o que ajuda a manter a infecção na natureza, porém algumas
4 delas precisam concluir seu ciclo de vida em mais hospedeiros para garantir sua sobrevivência
5 (DAVOUST B et al., 2010).

6 Baseado nisso, um segundo hospedeiro, no caso os seres humanos, pode se infectar com
7 a picada do artrópode. E, a transmissão das rickettsias, de uma forma geral, vai ocorrer por meio
8 de produtos salivares ou por inoculação das fezes do vetor infectado na ferida. Uma vez dentro
9 deste hospedeiro, a bactéria é fagocitada por células dendríticas e transportada via vasos
10 linfáticos para os linfonodos locais, onde irão se replicar. Após a replicação, esses organismos
11 entram na corrente sanguínea e se disseminam para a infecção de células endoteliais dos vasos
12 sanguíneos (COLONNE; EREMEEVA; SAHNI, 2011; RYDKINA; TURPIN; SAHNI, 2010).
13 Dados sugerem que as espécies do grupo SFG podem espalhar sua infecção entre células através
14 da motilidade de actinas que compõem as células sem destruí-las, ao contrário do grupo TG, que
15 se mantém na célula até causar a lise (OSTERLOH, 2020). À medida que a infecção dos
16 endotélios ocorre, o dano causado pela inflamação vai levar a um aumento da permeabilidade
17 vascular, que conseqüentemente, vai se manifestar através das erupções cutâneas, e em casos
18 mais graves, podem causar pneumonia intersticial, lesão renal e falência múltipla de órgãos
19 (BLANTON, 2019; MANSUETO et al., 2012). A infecção vai ser seguida por um período de
20 incubação que varia de sete à quinze dias antes do início dos sintomas e os infectados
21 normalmente apresentam sintomas como náusea, vômito, escaras no local da picada do
22 carrapato, erupção maculopapular e febre (GAYWEE et al., 2007; PAROLA et al., 2013).

23 Ainda dentro da fisiopatologia, *R. prowazekii* é o único conhecido desse gênero que
24 consegue sobreviver fora do seu vetor por algum tempo, já que o seu processo de replicação
25 rompe o trato gastrointestinal do vetor, levando-o a morte. Além de também ser o único membro
26 do gênero que pode causar uma infecção latente, chamada de doença de Brill-Zinsser (BZD),
27 que pode se manifestar anos depois, devido a uma baixa no sistema imunológico do hospedeiro,
28 levando a uma nova ascensão da bactéria (AKRAM; LADD; KING, 2022; PIERACCI et al.,
29 2017).

30 31 **1.1.3 Epidemiologia**

32 Os organismos do gênero *Rickettsia* têm sua distribuição ao redor do mundo de acordo
33 com a prevalência dos seus vetores em cada uma das regiões. Apesar de serem organismos
34 conhecidos, as rickettsioses são negligenciadas por serem mais presentes em países em

1 desenvolvimento, o que dificulta ainda mais o processo de avaliação da quantidade de casos
2 dessas doenças (BERMÚDEZ; TROYO, 2018).

3 O grupo TG tem dois principais representantes, sendo o primeiro *R. prowazekii*, causador
4 do tifo epidêmico transmitido por piolhos. A infecção ocorre quando as condições não
5 higiênicas favorecem a infestação dos piolhos no corpo humano, como acontece em guerras,
6 desastres naturais, e em roupas não lavadas (ANGELAKIS; BECHAH; RAOULT, 2016;
7 BURNS; ACUNA-SOTO; STAHLER, 2014). Essa espécie foi a primeira a ter uma vacina
8 produzida e utilizada durante as guerras (AKRAM; LADD; KING, 2022). O piolho não é
9 acostumado com temperaturas elevadas, então quando a pessoa já está infectada e febril, eles
10 migram para um próximo hospedeiro, espalhando a infecção. Segundo a história, ele é
11 responsável por mais de 20 epidemias, sendo as últimas em Burundi e em Ruanda, com taxa de
12 mortalidade chegando a 10% de todos os infectados (BECHAH et al., 2008; NYATANYI et
13 al., 2016). Já o segundo, o tifo murino, também chamado de tifo endêmico, causado por *R.*
14 *typhi*, apresenta um tropismo por regiões costeiras tropicais e subtropicais, que têm um tráfego
15 intenso de navios repletos do seu reservatório primário, o *Rattus spp.* (KUO et al., 2017). Foi
16 responsável por muitos casos no Vietnã e nos EUA durante os anos de 1900, porém o uso de
17 diclorodifeniltricloroetano (DDT), um pesticida, em abrigos de ratos resultou em um forte
18 declínio nos casos (BLANTON; WALKER, 2017; HAMAGUCHI et al., 2015). Por ser
19 indistinguível de algumas outras doenças febris presentes nos trópicos, provavelmente o tifo
20 murino têm sua prevalência subnotificada (BLANTON, 2019).

21 As rickettsioses do grupo SFG são transmitidas por diversas espécies de carrapatos, que
22 vão funcionar, não só como vetores, mas como hospedeiros reservatórios, que mantém a espécie
23 na natureza através das transmissões transestadial e transovariana (BLANTON, 2019). A
24 RMSF, causada por *R. rickettsii* foi descoberta em Idaho, Estados Unidos da América (EUA),
25 porém hoje se sabe que é uma doença transmitida em toda a América. Espécies do carrapato
26 *Dermacentor* são comumente responsáveis pela transmissão no sudeste e centro-sul e nos
27 estados montanhosos do oeste dos EUA (PAROLA et al., 2013). Em algumas terras tribais do
28 Arizona, EUA e em Sonora, México, diversas infestações peridomésticas são relacionadas com
29 *Rhipicephalus sanguineus* (STRAILY et al., 2016). Já em terras da América Central e do Sul,
30 a RMSF é transmitida por espécies de *Amblyomma* (TARRAGONA et al., 2016). Na última
31 década houve um aumento de notificação de casos de rickettsioses SFG nos EUA
32 (DAHLGREN et al., 2016), porém a taxa de mortalidade ainda fica em torno de 0,3%
33 nacionalmente. Muitos desses casos se devem à exposição a outras espécies além de *R.*
34 *rickettsii*, muito prevalente no estado do Arizona, com taxa de mortalidade em

1 aproximadamente 10% (WALKER, 2016). Esses dados podem tentar ser explicados, pois pode
2 ocorrer um fenômeno de interferência na transmissão dessas espécies quando há mais de uma
3 *Rickettsia* presente dentro de um mesmo carrapato (SAKAI et al., 2014).

4 Já a MSF, causada por *R. conorii*, também é transmitida por *Rhipicephalus sanguineus*,
5 porém com foco na Europa, África e Ásia (PAROLA et al., 2013) e principalmente nos meses
6 do verão, quando os carrapatos são mais ativos, e com uma prevalência maior em pessoas que
7 viajam para áreas endêmicas (DELORD; SOCOLOVSKI; PAROLA, 2014).

8 9 **1.1.4 Diagnóstico**

10 O diagnóstico das rickettsioses é complexo e difícil, sendo o reconhecimento dos
11 sintomas clínicos e conhecimento da epidemiologia local cruciais para a realização de um
12 diagnóstico diferencial de outras doenças febris. Importante destacar também a investigação de
13 viagens a locais endêmicos e de possíveis mordidas desses artrópodes, informações as quais
14 podem ajudar no diagnóstico diferencial (BLANTON, 2019).

15 O primeiro meio de diagnóstico para a identificação dessas bactérias é a cultura da
16 bactéria. Esta pode ser realizada com amostras de sangue e biópsia de tecido, porém requer
17 técnicas usando meio livre de antibiótico e um laboratório de nível três de biossegurança; já
18 que essas bactérias podem ser aerossolizadas e representar um risco laboral a equipe de trabalho
19 do laboratório. Outra dificuldade desse método é o crescimento muito lento e o difícil cultivo
20 por se tratar de um microrganismo intracelular obrigatório. Durante a fase aguda, apesar de não
21 estar disponível na maioria dos laboratórios clínicos, a detecção imuno-histoquímica pode ser
22 realizada através da biópsia de pele das lesões como uma alternativa (FANG; BLANTON;
23 WALKER, 2017).

24 A amplificação de DNA da bactéria por reação em cadeia da polimerase (PCR) também
25 pode ser um segundo método diagnóstico, sendo usada em amostras de sangue periférico,
26 plasma e *swabs* coletados nas escaras. Quando realizado por meio de primers específicos,
27 oferece um diagnóstico espécie-específico; todavia, é mais útil quando aplicada em amostras
28 de biópsia das escaras ou erupções cutâneas (BLANTON, 2019; PENICHE-LARA et al., 2013).

29 Por fim, mesmo que não seja espécie-específico, a sorologia, como por exemplo o ensaio
30 de imunofluorescência indireta (IFA), é o mais indicado, utilizando anticorpos conjugados
31 marcados com fluoresceína para detectar os anticorpos séricos ligados à rickettsias fixadas na
32 lâmina. Os ensaios de IFA se tornam mais sensíveis na busca de anticorpos de duas a três
33 semanas após o início da doença, com melhores interpretações se comparadas fases agudas e
34 de convalesça da doença (BIGGS et al., 2016). Ensaios de imunoabsorção enzimática estão

1 disponíveis para comercialização para oferecer um diagnóstico mais qualitativo dos grupos
2 SFG e TG do que os títulos com IFA (DUNN et al., 2014).

3 4 **1.1.5 Tratamento**

5 Alguns estudos trazem a comparação de alguns medicamentos para o tratamento das
6 rickettsioses com certas variações na eficácia entre os grupos, o que torna muito importante o
7 reconhecimento e identificação de qual microrganismo é o causador daquela rickettsiose, além
8 de melhorar o prognóstico com a utilização correta dos medicamentos.

9 Embora ainda existam poucos estudos que avaliem de forma prospectiva os resultados, a
10 classe de drogas com melhor eficácia *in vitro* para o tratamento de uma forma geral são as
11 tetraciclina, principalmente a doxiciclina. Além da doxiciclina, o cloranfenicol e as
12 fluoroquinolonas apresentam bons resultados na ação contra os organismos, ao contrário da
13 amoxicilina (beta-lactâmico) e da gentamicina (aminoglicosídeo), que tiveram um valor de
14 MIC alto, sendo menos efetivos. Há ainda medicamentos que variaram com os grupos, como a
15 eritromicina que foi efetiva contra o grupo TG, mas não efetiva contra o grupo SFG. De forma
16 similar, a rifampicina tem certa eficácia contra o grupo TG e algumas SFG, mas não efetiva
17 contra outras (ROLAIN, 2005; VANROMPAY et al., 2018).

18 Outros estudos ainda descrevem que a doxiciclina deve ser administrado em doses de
19 acordo com o quadro e recomendação médica, e diferente dos seus parceiros de classe, é
20 biodisponível na presença de alimentos, melhorando o desconforto gastrointestinal
21 (BLANTON, 2019; DUMLER, 2014). O tratamento deve ser realizado via oral até que o
22 paciente melhore e esteja afebril de 24 a 48 horas, tratamento que dura por volta de uma semana
23 (AKRAM; LADD; KING, 2022).

24 25 **1.1.6 Prevenção**

26 Não existem vacinas disponíveis como meios de prevenção contra as rickettsioses. E os
27 métodos de prevenção disponíveis atualmente são evitar os vetores através de repelentes, uso
28 de roupas de proteção, como por exemplo, tratadas com permetrina que dura cerca de um ano
29 e protege de picadas de carrapatos (VAUGHN et al., 2014). No caso do tifo murino, o uso de
30 venenos de ratos é importante para diminuir, de forma indireta, a população dos artrópodes que
31 os infectam. Em algumas situações, a Organização Mundial da Saúde (OMS) recomenda o
32 tratamento em massa com a pulverização de ar comprimido misturado com permetrina no
33 ambiente e nas roupas (WORLD HEALTH ORGANIZATION, 1994).

1.2 VACINAS

Estudos relacionados a criação das vacinas vindo sendo realizados desde meados do século XVIII e segundo a OMS, os atuais programas de vacinação, como por exemplo, vacinas contra difteria, poliomielite, sarampo, coqueluche, etc., contribuem para salvar mais de dois milhões de pessoas a cada ano, principalmente de crianças menores que cinco anos de idade (WORLD HEALTH ORGANIZATION, 2020).

As vacinas contêm antígenos derivados de patógenos ou antígenos que são sintetizados para representar os componentes de um patógeno, e por isso são consideradas um produto biológico. Elas são usadas principalmente para induzir, de forma segura, uma resposta imune que vai conferir ao organismo uma proteção contra uma futura exposição a esse patógeno (POLLARD; BIJKER, 2021).

Geralmente, as vacinas podem ser divididas em dois tipos: atenuadas, inativadas ou de subunidades. As vacinas atenuadas representam aquelas que possuem cepas dos patógenos na sua forma atenuada, ou seja, em indivíduos imunocompetentes, possuem uma capacidade de se replicar até conseguirem estimular uma resposta imune, porém que não causem manifestações clínicas da doença, por exemplo, vacinas do sarampo, caxumba, rubéola. As vacinas de subunidades ou inativadas são aquelas que possuem apenas uma parte de um patógeno ou o patógeno inteiro morto, respectivamente, por exemplo, vacina de coqueluche de célula inteira e vacina de hepatite B (VHB) com proteína recombinante. Recentemente, vêm surgindo outros tipos de vacinas, como vacinas de vetores virais, vacinas baseadas em ácidos nucleicos de DNA e RNA, vacinas de subunidades, por exemplo (POLLARD; BIJKER, 2021).

1.3 BIOINFORMÁTICA

A bioinformática é uma área da ciência que mescla a tecnologia da computação com os conhecimentos básicos de matemática, física, química e biologia molecular, como forma de tentar ajudar a entender e responder questões relacionadas às ciências da vida (HAGEN, 2000). Seu surgimento esteve intimamente relacionado aos primeiros computadores por volta dos anos de 1960, sendo importante até hoje quando se trata do estudo das ciências ômicas, área que têm foco no estudo nas funções, diferenças e nos tipos de interações entre as moléculas de um organismo (GAUTHIER et al., 2019).

Dentre as áreas de pesquisa da bioinformática dentro das ciências ômicas, podemos citar os estudos relacionados ao DNA como um todo (genômica), relacionados a transcrição do DNA para RNA (transcriptômica), ao processo de tradução de RNA para proteínas (proteômica) e por fim, os metabólitos relacionados a essas proteínas (metabolômica) (AIZAT; ISMAIL;

1 NOOR, 2018). Dentro da bioinformática, todas essas áreas são estudadas no formato *in silico*,
2 ou seja, no contexto da pesquisa, termo utilizado para se referir a métodos experimentais
3 realizados em computadores (PAPPALARDO et al., 2019).
4

5 1.4 IMUNOINFORMÁTICA

6 Dentro das vacinas de subunidades citadas acima, por exemplo, as tecnologias de
7 sequenciamento de nova geração e os avanços da bioinformática, permitem que a
8 imunoinformática utilize diversas ferramentas *in silico* para otimizar a triagem laboratorial da
9 construção dessas vacinas, em alvos que foram previamente selecionados devido à sua alta
10 afinidade imunogênica (BAH et al., 2018).

11 A imunoinformática é uma vertente dentro da área da genômica da Bioinformática, que
12 se baseia em ferramentas e algoritmos que buscam, principalmente, estudar as correlações de
13 genes e proteínas do patógeno com células do sistema imune do hospedeiro, como por exemplo,
14 linfócitos T e B, o que ajuda a potencializar o processo de criação de testes diagnósticos e
15 vacinas contra diversas doenças (BACKERT; KOHLBACHER, 2015; TOMAR; DE, 2010).

16 Nesse contexto, algumas dessas ferramentas da imunoinformática foram criadas para
17 ajudar no processo de desenvolvimento de vacinas a partir de epítomos das proteínas do
18 patógeno, as quais são chamadas de vacinas multi-epítomo. Essas vacinas são compostas por
19 uma proteína quimérica e um adjuvante, que são unidos por peptídeos de ligação, chamados de
20 *linkers* (NEGAHDARIPOUR et al., 2018). Os adjuvantes desempenham um papel fundamental
21 nessa vacina porque funcionam como imunoestimulantes, já que os peptídeos sintéticos
22 normalmente utilizados nessas vacinas, apresentam poucos determinantes imunogênicos
23 (AZMI et al., 2014; LEI et al., 2019). Já os *linkers*, além da função de conexão dos epítomos,
24 auxiliam na atividade biológica, na estabilidade da proteína e por vezes, como estimulantes do
25 processo de expressão (CHEN; ZARO; SHEN, 2013a; HU et al., 2004).

26 As vacinas multi-epítomo possuem diversas vantagens em comparação com outras
27 metodologias de criação de vacina, como por exemplo, a fusão epítomos específicos dos
28 patógenos de interesse. Essa fusão visa uma maior modulação e aumento da imunogenicidade,
29 além de remover epítomos que possam causar efeitos colaterais. Também podem desencadear
30 uma resposta imune mais direcionada, mediada por anticorpos e células, à medida que varia a
31 sobreposição dos epítomos de células T auxiliares, citotóxicas e células B (OLI et al., 2020).

32 Por serem desenvolvida a partir de vários epítomos, essas vacinas normalmente não têm
33 possibilidade de integração genética como em vacinas de DNA e nem de reversão de virulência
34 como nas vacinas atenuadas (MOYLE; TOTH, 2008). Além disso, possuem maior facilidade

1 no manejo e apresentam uma diminuição de custos de síntese de peptídeos, já que são criadas
2 de forma *in silico* e também produzidas na forma liofilizada (PURCELL; MCCLUSKEY;
3 ROSSJOHN, 2007).

4 5 **1.4.1 Trabalhos com abordagens multi-epítopo**

6 Dados recentes mostraram resultados promissores utilizando abordagens multi-epítopo
7 em modelos animais, como no caso de Cao e colaboradores (2013). Este estudo selecionou
8 epítomos de proteínas de três sorotipos do vírus causador da febre aftosa em porcos e os
9 conectou a um adjuvante poli I:C para a criação de uma vacina multi-epítopo. Uma dessas
10 vacinas testadas com esse adjuvante conferiu proteção completa após realizarem um ensaio de
11 infecção do vírus da febre aftosa em camundongos e porcos. Outro estudo, esse realizado com
12 *Leishmania infantum*, utilizou 4 proteínas antigênicas acopladas a dois agonistas do receptor
13 Toll 4 para a criação de uma vacina. Eles utilizaram camundongos do tipo Balb/c e fizeram um
14 desafio com o parasito, observando um aumento de resposta Th1 e uma diminuição de resposta
15 Th2 acoplada com redução parasitária significativa (VAKILI et al., 2020).

16 Na literatura, já foram realizados testes com uma potencial vacina multi-epítopo contra
17 carrapatos, como no caso de Tshilwane e colaboradores (2019), em que formularam dois
18 constructos baseado em multi-epítopo de DNA, pSignal plus e pLamp, de *Ehrlichia*
19 *ruminantium* (também conhecida como *Rickettsia ruminantium*). Ambos foram aplicados
20 topicamente, mas sem conferir proteção as ovelhas; entretanto, um dos constructos, o pLamp,
21 foi aplicado junto a um adjuvante monofosforil lipídio A via intramuscular em ovelhas, das
22 quais 60% das ovelhas tiveram proteção contra o ensaio de carrapatos em campo. Alguns
23 estudos disponíveis em formato *pré-print*, tratam de abordagens multi-epítopo em outros
24 carrapatos também, como das espécies *Rhipicephalus microplus*, um carrapato bovino (NANDI
25 et al., 2022), e de carrapatos como *Rhipicephalus turanicus* e *Hyalomma maginatum*, que
26 transmitem bactérias do gênero *Anaplasma*, pertencentes a ordem das *Rickettsiales*
27 (KÖSEOĞLU et al., 2022).

28 Lemes e colaboradores (2022) selecionaram epítomos ligantes às moléculas de MH C I e
29 II para a construção de uma proteína multi-epítopo com atuação em *Mycobacterium leprae* e
30 *Mycobacterium lepromatosis*. Segundo suas análises *in silico*, a proteína não apresentou
31 reações adversas ao hospedeiro e tinha potencial para elicitar resposta imune, sendo
32 considerado como um potencial imunógeno para a utilização em kits diagnóstico para a
33 hanseníase. Seguindo uma metodologia parecida, Vilela-Rodrigues e colaboradores (2022) se
34 basearam em um trabalho de vacinologia reversa para identificar proteínas como potencial ação

1 contra *Mycoplasma pneumoniae*. Dessas proteínas, selecionaram epítomos potenciais e
2 acoplaram com um adjuvante, a enterotoxina termolábil de *E. coli*, desenvolvendo uma
3 proteína estável, não alérgica e com potencial de geração de resposta imune inata e adaptativa
4 contra a pneumonia. Apesar de todos esses estudos citados, não existem dados na literatura com
5 pesquisas vacinais utilizando abordagens de imunoinformática para as bactérias do gênero
6 *Rickettsia*.

2 JUSTIFICATIVA

As rickettsioses são doenças consideradas negligenciadas, visto que costumam estar mais presentes dentro dos trópicos, principalmente nos países em desenvolvimento. Esses países normalmente apresentam condições mais vulneráveis, o que favorece a proliferação dos vetores dessas doenças. Além disso, possivelmente os números relacionados a epidemiologia das rickettsioses são subnotificados, pela falta de controle e de gestão dos sistemas de saúde desses países.

Até o momento não existem meios de prevenção, como vacinas contra essas doenças. Isso pode ser explicado por serem organismos intracelulares obrigatórios, o que dificulta todo o processo de cultivo e isolamento do patógeno, além da necessidade de utilização de laboratórios de alta segurança para o manejo dos mesmos. Dessa maneira, as abordagens *in silico* surgem com uma alternativa para a busca de novos alvos vacinais contra as rickettsioses. Essa abordagem além de agilizar o processo de criação da vacina, diminui os custos para o seu desenvolvimento, além de diminuir também a utilização de animais submetidos aos experimentos.

3 OBJETIVOS

3.1 OBJETIVOS GERAIS

Construir uma vacina multi-epítopo contra o gênero *Rickettsia* por meio de ferramentas de imunoinformática disponíveis online e gratuitamente, baseada nos dados de proteínas já encontrados em trabalho de vacinologia reversa.

3.2 OBJETIVOS ESPECÍFICOS

- a) Identificar epítomos específicos de células B, células T citotóxicas e células T auxiliares.
- b) Avaliar cada epítopo quanto a antigenicidade, potencial citotóxico e alérgico, cobertura populacional.
- c) Determinar a sequência de aminoácidos da proteína quimérica da vacina, prever e avaliar a estrutura tridimensional.
- d) Avaliar a potencial interação da vacina com o sistema imune do hospedeiro através de ferramentas de bioinformática.

4 MATERIAIS E MÉTODOS

4.1 SELEÇÃO DOS DADOS

Para a análise e construção da proteína quimérica multi-epítopo foram utilizadas oito proteínas identificadas do *core* genoma do gênero *Rickettsia* como possíveis alvos vacinais em um trabalho realizado anteriormente (FELICE et al., 2021). (TABELA 1).

Tabela 1 - Informações das oito proteínas potencialmente candidatas à vacina contra o gênero *Rickettsia*.

<i>Locus ID</i>	Nome (NCBI)	Localização Subcelular	Probabilidade de Adesão
WP_012148219.1	Porina	Secretada	0,620
WP_012148227.1	OmpW Família de Proteína	Secretada	0,632
WP_012148552.1	Chaperone de Cobre PCu(A)C	Secretada	0,620
WP_012148811.1	Dioxigenase	Secretada	0,598
WP_012148819.1	D-alanil-D-alanina carboxipeptidase	Secretada	0,510
WP_012149174.1	Família de Proteínas - Porina	Secretada	0,533
WP_012149175.1	Família de Proteínas - Porina	Secretada	0,510
WP_012149185.1	Alpha/Beta hidrolase	PSE	0,564

Fonte: Adaptado de Felice AG e colaboradores, 2021.

Nota: PSE – possivelmente exposta a superfície (do inglês, *putative exposed surface*)

4.2 PREDIÇÃO DE EPÍTOPOS QUE SE LIGAM A ALELOS MHC-I

Epítopos ligantes a Moléculas do Complexo de Histocompatibilidade (MHC) ou *Human Leucocyte Antigen* (HLA), além de serem reconhecidos por essas moléculas, são apresentados aos linfócitos T citotóxicos (CTL - *cytotoxic T lymphocyte*) e induzem uma resposta imune (PARVIZPOUR et al., 2020). Para identificar esses epítopos, foram utilizadas duas ferramentas para dar maior confiabilidade no resultado: MHC-I *binding prediction tool* do banco de dados IEDB (Immune Epitope Database)(VITA et al., 2019), e o servidor NetCTL 1.2 (LARSEN et al., 2007).

O IEDB é um banco de dados com vários epítopos já classificados em alta ou baixa imunogenicidade, os quais são ranqueados de acordo com o seu valor de IC50 (metade da concentração inibitória máxima): epítopos considerados de alta imunogenicidade possuem IC50 < 50nM, e os considerados com baixa imunogenicidade, possuem IC50 > 50nM (SOUTHWOOD et al., 1998). Essa plataforma também proporciona uma predição da ligação

1 dos epítomos com o MHC I, resultados os quais sugerem que epítomos com nove aminoácidos
2 sejam mais antigênicos e com alta afinidade de ligação com MHC (VITA et al., 2019). Como
3 nosso trabalho tem o intuito de atingir a maior parte da população mundial, um conjunto de 27
4 alelos que são considerados mais frequentes na população foram utilizados (VILELA
5 RODRIGUES et al., 2022). Então, com os parâmetros padrão, foram selecionados os epítomos
6 que possuíam uma classificação percentual $> 1\%$, os quais têm maior probabilidade de serem
7 imunogênicos, para as próximas análises.

8 Já o servidor NetCTL, foram caracterizados os epítomos seguindo três critérios
9 relacionados aos estágios do processo de apresentação de antígeno aos MHC I: clivagem no
10 terminal C, transporte eficiente e boa afinidade de ligação com MHC I. Para o peso na clivagem
11 no terminal C foi utilizado o *cut-off* de 0,15 e para o peso no transporte eficiente foi utilizado
12 um *cut-off* de 0,05. Por fim, os epítomos que obtiveram um resultado de pontuação $\geq 0,75$
13 (classificação “<E” pelo próprio programa) foram considerados bons epítomos para as próximas
14 análises (ALI et al., 2017).

15 16 4.3 PREDIÇÃO DE EPÍTOPOS QUE SE LIGAM A ALELOS MHC-II

17 Seguindo o mesmo raciocínio, os epítomos ligantes à MHC II são definidos como aqueles
18 capazes de serem reconhecidos por moléculas MHC II e apresentados aos linfócitos T auxiliares
19 (HTL – *helper T lymphocyte*), elicitando uma resposta imune (SOUTHWOOD et al., 1998).
20 Para realizar tal identificação, a ferramenta MHC-II *binding prediction tool* do banco de dados
21 IEDB (VITA et al., 2019) e o servidor NetMHCII 2.3 (RANA; AKHTER, 2016) foram
22 utilizados.

23 Diferentemente dos encontrados com CTL, os epítomos de HTL são ligeiramente maiores
24 (13 a 25 aminoácidos); então um padrão de 15 aminoácidos por epítopo foi estabelecido, devido
25 ao tamanho da fenda existente neste MHC formado pelas cadeias alfa e beta. Nesta etapa, a
26 métrica utilizada também foi a relacionada com o IC50, então para a ferramenta do IEDB,
27 seguiram para a filtragem os epítomos com IC50 > 500 nM e com *Percentil Rank* < 10 .

28 O servidor NetMHCII utiliza uma Rede Neural Artificial (ANN) acoplada a um banco de
29 dados de epítomos; então, foram selecionados todos aqueles que tiveram classificação “SB” pelo
30 próprio programa.

31 32 4.4 PREDIÇÃO DE EPÍTOPOS QUE SE LIGAM A ALELOS DE CÉLULAS B

33 A ferramenta ABCpred foi utilizada para a predição dos epítomos lineares de linfócitos
34 B. Essa ferramenta trabalha com abordagens de redes neurais artificiais (ANN, Artificial Neural

1 Network) e Máquina de Vetores de Suporte (SVM, Support Vector Machine). Foram
2 selecionados epítomos utilizando parâmetros padrão, ou seja, um tamanho de 16 resíduos para
3 cada epítomo e uma boa acurácia com *threshold* de 0,51. Então, esses epítomos seguiram para
4 as próximas análises (CHEN et al., 2007; EL-MANZALAWY; DOBBS; HONAVAR, 2008;
5 SAHA; RAGHAVA, 2006).

6 7 4.5 SELEÇÃO DOS MELHORES EPÍTOPOS

8 Como o foco da criação de uma vacina multi-epítomo é utilizar somente aqueles que fazem
9 parte de todas as proteínas utilizadas, primeiro utilizamos de filtragens com o intuito de reduzir
10 a quantidade de epítomos, selecionando apenas os melhores. Para os epítomos que se ligam ao
11 MHC I preditos pela ferramenta do IEDB, os que tiveram o *percentil rank* < 1 passaram pela
12 ferramenta de “remoção de duplicatas” do programa Microsoft Excel (2019) para remover os
13 epítomos duplicados. Para os epítomos de MHC I preditos pelo NetCTL, além da classificação
14 de qualidade “<- E” pelo próprio programa, também foi utilizada a ferramenta de remoção de
15 duplicatas.

16 Já para os epítomos que se ligam à MHC II preditos pelos IEDB, após selecionar os que
17 tinham classificação de *percentil rank* < 10 e IC50 < 500 nM, eles também passaram pela
18 ferramenta de remoção de duplicatas. Porém, ainda continuava um número alto de epítomos,
19 então, passaram pelo *software* ToxinPred (GUPTA et al., 2013) que removeu os epítomos que
20 eram considerados tóxicos ao hospedeiro. Por fim, os epítomos preditos pelo programa
21 NetMHCII, após o filtro de qualidade do programa, também passaram pela ferramenta de
22 remoção de duplicatas. Todos os epítomos preditos pelo ABCpred foram utilizados.

23 Após as filtragens, foi utilizado um *script in house* para sobrepor os resultados dos dois
24 programas: primeiro para os epítomos filtrados de MHC I: IEDB vs NetCTL e, depois o
25 resultado dessa sobreposição vs epítomos de células B. Após esse resultado, foi utilizado a
26 ferramenta *Class I Immunogenicity* (CALIS et al., 2013) do IEDB para prever a
27 imunogenicidade dos epítomos restantes de MHC I. Os epítomos restantes ainda passaram por
28 mais uma curadoria, dessa vez manual, para remoção de até dois aminoácidos (AA), com o
29 intuito de reduzir ainda mais a quantidade. Por fim, estes foram submetidos a ferramenta
30 Vaxijen (DOYTCHINOVA; FLOWER, 2010), sendo utilizado um *threshold* de 0.7, para
31 selecionar os epítomos mais imunogênicos.

32 Para os epítomos filtrados de MHC II, o mesmo procedimento foi realizado: utilizado um
33 *script in house* para sobrepor os resultados do IEDB vs NetMHCII e depois, o resultado vs

1 epítomos de células B. Os epítomos restantes passaram pela remoção de até dois AA e
2 posteriormente, foram submetidos a ferramenta Vaxijen, com o mesmo *threshold* de 0.7.

4 4.6 CONSTRUÇÃO DA PROTEÍNA MULTI-EPÍTOPO

5 Os epítomos finais, que passaram por todos os critérios de filtragem, foram agrupados
6 para a criação da sequência da proteína multi-epítopo. Para esse processo foi utilizado outro
7 *script in house*, o qual gerou 1000 modelos randômicos, e que posteriormente foram submetidos
8 à *web service* Vaxijen, para selecionar somente o modelo mais antigênico. A sequência desse
9 modelo então foi inserida novamente nesse mesmo *script*, o qual adicionou os *linkers* peptídicos
10 AAY e GPGPG entre os epítomos CTL/B e somente GPGPG entre os epítomos HTL/B
11 (CHAUHAN et al., 2019). Esses ligantes são importantes para diversas funções como ajudar
12 na estabilização da proteína e no processo de apresentação de antígeno (CHEN; ZARO; SHEN,
13 2013b). Além disso, também foi adicionada a sequência precursora da cadeia B da enterotoxina
14 termolábil de *Escherichia coli* (Acesso Genbank: [ALO79813.1](#)) como adjuvante (MA, 2016),
15 seguida da sequência peptídica EAAAK (LI et al., 2016), com intuito de melhorar a atividade
16 imunogênica da sequência multi-epítopo.

18 4.7 PREDIÇÃO DAS ESTRUTURAS SECUNDÁRIA, TERCIÁRIA E DE EPÍTOPOS 19 CONFORMACIONAIS DE CÉLULAS B

20 O *WebService* PSIPRED (WARD et al., 2003) foi utilizado para predizer a estrutura
21 secundária da proteína com base em redes neurais e em *Position-Specific Scoring Matrix*
22 (PSSM). Para a estrutura terciária foi utilizado o ColabFold (MIRDITA et al., 2022), preditor
23 online que se baseia em uma busca rápida de homologia com outros programas como MMseqs2
24 com AlphaFold2 ou RoseTTAFold, hospedado no site
25 <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>.
26 A melhor estrutura gerada pelo ColabFold foi submetida a ferramenta GalaxyRefine (HEO;
27 PARK; SEOK, 2013), que refina a estrutura da proteína mudando as regiões de cadeias laterais
28 em regiões permitidas. Por fim, a ferramenta PROCHECK (LASKOWSKI et al., 1993) através
29 do *WebService* SAVES v6.0 (<https://saves.mbi.ucla.edu/>) gerou um gráfico de Ramachandran
30 para verificar a qualidade da estrutura. Quanto mais próximo de regiões permitidas os
31 aminoácidos estiverem, melhor é o benefício para a qualidade da estrutura.

32 Epítomos conformacionais de células B são regiões de aminoácidos importantes por
33 influenciarem na resposta imune do hospedeiro e, para prever tais epítomos, será utilizado o
34 servidor ElliPro (PONOMARENKO et al., 2008), em seus parâmetros padrão.

4.8 AVALIAÇÃO DA PROTEÍNA MULTI-EPÍTOPO E DA INTERAÇÃO COM O HOSPEDEIRO

Como forma de validação da sequência multi-epítipo, ela foi submetida a diversos programas como o Vaxijen para determinar sua potencial antigenicidade (HE; XIANG; MOBLEY, 2010), com intuito de confirmar sua capacidade de ligação à MHC/II. Para avaliação da alergenicidade foi utilizada a ferramenta Allertop v2.0 (DIMITROV et al., 2014) e para a toxicidade foi utilizada a ferramenta ToxinPred.

A ferramenta ProtParam (GASTEIGER E., 2005) foi utilizada para analisar a composição dos aminoácidos, o ponto isoelétrico (pI) da proteína, o índice alifático, o peso molecular, o coeficiente de meia vida e a média da hidropaticidade da sequência (GRAVY). O Protein-Sol (HEBDITCH et al., 2017) utilizou dados de *E. coli* para avaliar o índice de solubilidade. Por fim, a sequência foi submetida a ferramenta BLAST para a análise de homologia com as proteínas do hospedeiro (ALTSCHUL et al., 1990). Todos os programas foram utilizados em seus parâmetros *default*.

4.9 DOCKING MOLECULAR E DINÂMICA MOLECULAR NA INTERAÇÃO PROTEÍNA-RECEPTOR

O receptor do tipo Toll (TLR) escolhido para esta análise foi o tipo 4, devido a sua importância para o gênero *Rickettsia* (JORDAN et al., 2008a; RUMFIELD et al., 2020a), estrutura a qual foi recuperada do banco de dados do PDB (Protein Data Bank) (BERMAN, 2000). O *software* Chimera (PETTERSEN et al., 2004) foi utilizado para remover as moléculas de água, ligantes e outras cadeiras laterais. Então, a técnica de *docking* molecular (análise de acoplagem) foi realizada entre a proteína e a estrutura do TLR-4 usando o programa Cluspro v2.0. Esse programa divide a análise em três passos, sendo uma ancoragem rígida em bilhões de conformações, um desvio RMSD em clusters de 1000 estruturas que representam as mais prováveis conformações e por fim o refinamento das estruturas baseado na minimização de energia do complexo (KOZAKOV et al., 2017). O programa LigPlot+ foi utilizado para visualizar os resíduos de aminoácidos que participavam na interação proteína-receptor.

A análise de dinâmica molecular foi realizada com o programa IMODs (LÓPEZ-BLANCO et al., 2014). Esse programa analisa as vibrações das macromoléculas e projeta a dinâmica da interação entre as estruturas da proteína e do receptor TLR-4.

4.10 AVALIAÇÃO DA POSSÍVEL RESPOSTA IMUNE GERADA NO HOSPEDEIRO

1 A proteína final foi testada quanto à indução de IFN-gama seguida de ativação de
2 linfócitos TCD4+ (HTL) e TCD8+ (CTL) com o servidor IFNepitope, o qual sobrepõe os
3 epítomos contra seu banco de dados e os classifica em indutores ou não indutores, baseado na
4 pontuação > 1 (DHANDA; VIR; RAGHAVA, 2013). Ela também foi testada quanto a sua
5 capacidade de indução de IL-17 com o programa IL17eScan, através de *machine learning*
6 (GUPTA et al., 2017).

7 Por fim, para tentar descrever a resposta imune gerada pela vacina multi-epítomo criada,
8 utilizamos o servidor C-ImmSimm para realizar estimulações imune. Dados gerados com esse
9 programa representam fortemente o desenvolvimento real e robusto da imunidade no humano
10 (RAPIN et al., 2010). Foram simuladas três injeções cada com intervalo de quatro semanas,
11 sendo realizadas duas simulações: uma para a proteína inteira e outra somente da sequência do
12 adjuvante, como forma de comparar a eficácia dos compostos na administração da vacina. O
13 programa apresenta a contagem de células em *cells per mm³*, a contagem de imunoglobulinas
14 em *antigen count per mL* e a dosagem de citocinas em ng/ml.

5 RESULTADOS

5.1 OS MELHORES EPÍTOPOS PREDITOS PARA MHCI/CTL, MHCII/HTL E CÉLULAS B

Para a ligação MHCI/CTL foram preditos 55728 epítomos pela ferramenta do IEDB e 2064 epítomos pela ferramenta NetCTL quando somados os valores para todas as 8 proteínas. Seguindo o mesmo raciocínio, para a ligação MHCII/HTL foram preditos para as 8 proteínas um total de 53116 epítomos pela ferramenta do IEDB e 3358 epítomos pela ferramenta NetMHCII. Por fim, foram preditos um total de 194 epítomos ligantes à Células B pela ferramenta ABCpred.

Após a realização das filtragens com os parâmetros citados previamente, foi feita a sobreposição dos epítomos, totalizando 13 epítomos para o complexo MHCI/B e 13 epítomos para o complexo MHCII/B, os quais foram utilizados para as próximas etapas (TABELA 2). Uma descrição mais detalhada da quantidade de epítomos filtrados em cada uma das etapas pode ser encontrada no (ANEXO A – TABELAS 3, 4, 5 e 6). Nenhum epítomo da proteína WP_012149185.1 (Alpha/Beta hidrolase) obteve o limiar mínimo de 0,7 do Vaxijen, então não foram selecionados para as próximas análises.

Tabela 2 - Epítomos selecionados para a construção da proteína multi-epítomo.

Proteínas (ID)	Epítomos CTL/B	Imunogenicidade		
		Vaxijen (<i>threshold</i> > 0,7)	Epítomos HTL/B Vaxijen (<i>threshold</i> > 0,7)	
Porina /	KTTRGTVGL	1,21	IHTKMDRRTKGVEVLN	0,8
WP_012148219.1	QVTAGTGGW	1	EYKIMPGLLPYAEIS	0,76
	RVNFFTPKM	0,8	NSDTSIKLEGFYLL	0,82
	TANTGGNRY	2,1	-	-
OmpW Família de Proteína /	TGQFHIAPY	0,9	PKLEYKPNLVGNKTV	1,2
	GQIGVDFYA	0,85	-	-
WP_012148227.1				
Chaperone de	SSFKITFFF	2,1	GMHIMLYDLKGPLNV	0,97
Cobre PCu(A)C /	-	-	SKMVKVDYPFLIADN	0,98
WP_012148552.1				

Dioxigenase / WP_012148811.1	NECGEVNPI	1,2	-	-
D-alanil-D- alanina carboxipeptidase/ WP_012148819.1	PYIFIATFF	1,7	MALKRDYPKYYPLFS	0,7
	FLLPYIFIA	2,7	-	-
Porina / WP_012149174.1	KSTRGNIVK	1,1	MENQWYLKLNAGTMI	0,9
	GTVINVHLK	1,6	NANGITLTVNNPSLT	1,3
Porina / WP_012149175.1	SSMENQWYL	0,8	KNKINFAYQLSLGTS	0,8
	-	-	YQLSLGTSFEVAQGV	0,8
	-	-	RTNISYKLTLG TSAQ	1
	-	-	ATSVK LKSNMTVSVD	1,3

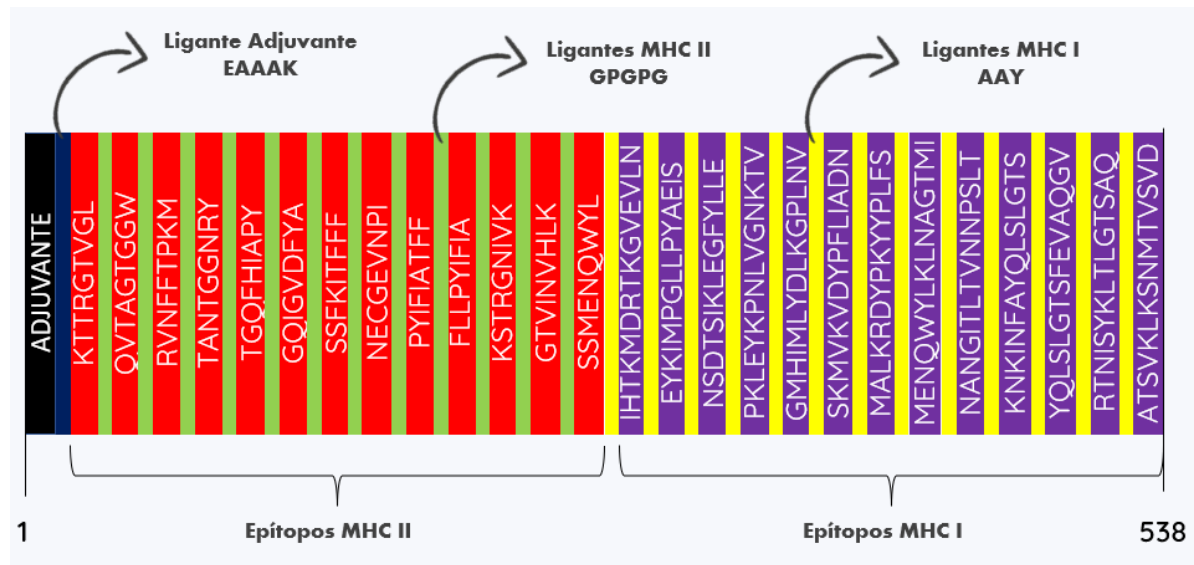
Fonte: Do Autor, 2022.

Nota: ID: identidade; CTL: Linfócitos T citotóxicos; HTL: Linfócitos T auxiliares.

5.2 NA CONSTRUÇÃO DA PROTEÍNA MULTI-EPÍTOPO FORAM UTILIZADOS VINTE E SEIS EPÍTOPOS

A proteína multi-epítipo foi construída com um total de 13 epítipos do complexo HTL/B intercalados com ligantes peptídicos GPGPG e 13 epítipos do complexo CTL/B intercalados com ligantes peptídicos AAY. Foi adicionado no início da sequência o adjuvante enterotoxina termo-lábil de *Escherichia coli* seguida por um ligante peptídico EAAAK. Assim, somando o adjuvante, os 3 tipos de ligantes adicionados e os 26 epítipos selecionados dos complexos HTL/B e CTL/B, o comprimento total da proteína multi-epítipo foi de 538 aminoácidos. A sequência final da proteína está plotada na Figura 1.

1 Figura 1 - Gráfico representativo da sequência da proteína multi-epítopo.



2 Fonte: Do Autor, 2022.

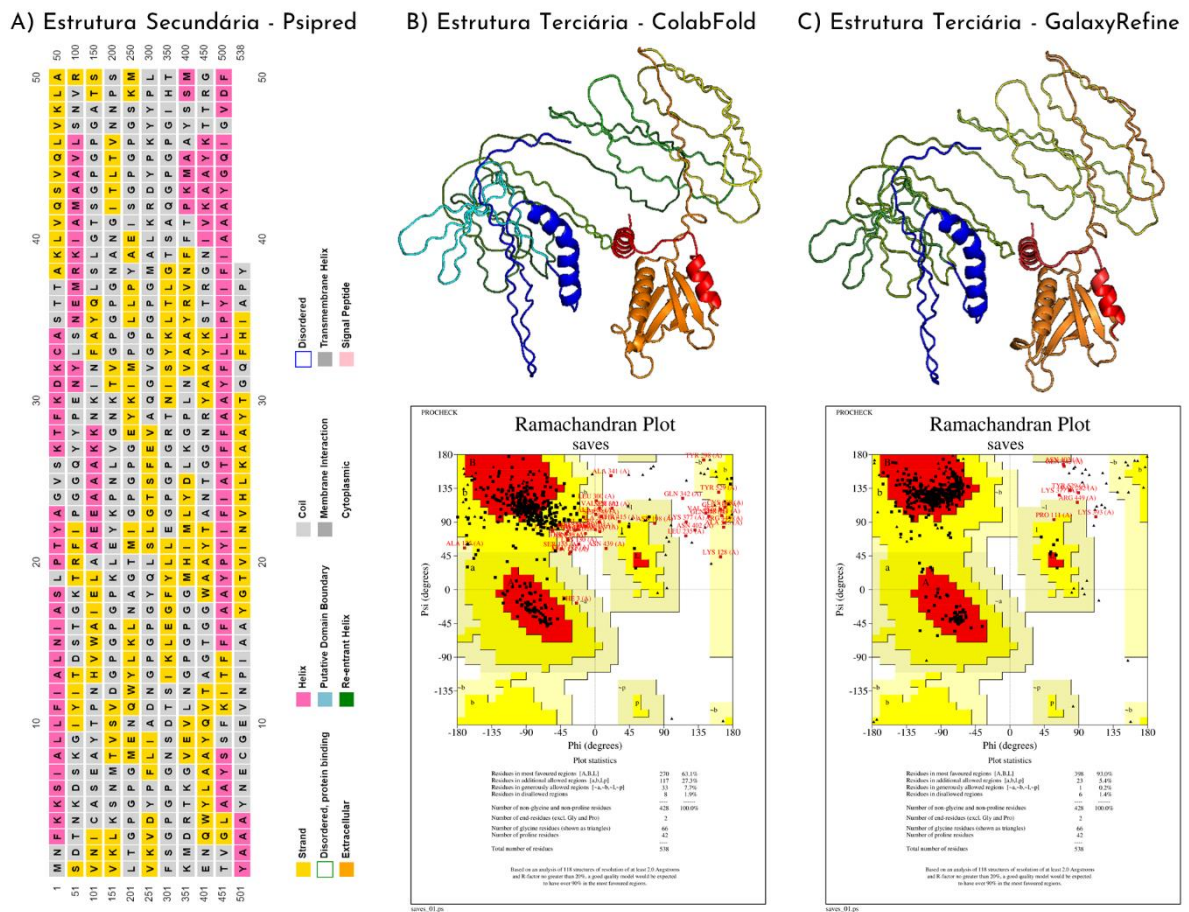
3 Nota: A imagem representa todos os epítopos selecionados intercalados com seus respectivos ligantes. Após o
 4 adjuvante (preto), na sequência são representados os epítopos do complexo HTL/B (vermelhos) intercalados com
 5 seus ligantes GPGPG (verde) e depois os epítopos do complexo CTL/B (roxos) intercalados com seus ligantes AAY
 6 (amarelo). A estrutura final possui 538 aminoácidos.
 7
 8

9 5.3 A ESTRUTURA 3D DA PROTEÍNA MULTI-EPÍTOPO APRESENTOU MAIS DE 10 90% DE AMINOÁCIDOS EM REGIÕES MAIS FAVORECIDAS

11 A estrutura inteira da proteína (538 AA) foi submetida ao programa Psipred. Segundo o
 12 programa, a estrutura da proteína apresentou 49,64% de alça, 30,29% de folhas β e 20,07% de
 13 hélices (FIGURA 2A). A melhor estrutura 3D predita pelo ColabFold apresentou 63,1% de
 14 aminoácidos em regiões mais favorecidas, 27,3% em regiões adicionais permitidas, 7,7% em
 15 regiões generosamente permitidas e 1,9% em regiões não permitidas, segundo *Ramachandran*
 16 *plot* na Figura 2B. A estrutura 3D refinada pelo GalaxyRefine aumentou para 93% de
 17 aminoácidos em regiões mais favorecidas, 5,4% em regiões adicionais permitidas, 0,2% em
 18 regiões generosamente permitidas e 1,4% em regiões não permitidas, conforme o
 19 *Ramachandran plot* na Figura 2, C. O programa Ellipro predisse sete epítopos lineares de
 20 células B e 10 epítopos descontínuados de células B para a proteína multi-epítopo (ANEXO B,
 21 TABELAS 7 e 8).
 22

1

Figura 2 - Representação das estruturas secundária e terciária da proteína multi-epítipo.



2

Fonte: Do Autor, 2022.

Nota: Resultados dos programas Psipred, ColabFold, GalaxyRefine e SAVES para as estruturas da proteína. Em A) está mostrado a estrutura secundária predita pelo Psipred com uma legenda explicativa das colorações de cada um dos aminoácidos da sequência. Em B) está mostrado na parte de cima a estrutura 3D predita pelo programa ColabFold e em baixo o *Ramachandran Plot* gerado na *webservice* SAVES com 63,1% dos aminoácidos em regiões favoráveis. Em C) está mostrado na parte superior a estrutura 3D da proteína refinada pelo GalaxyRefine e em baixo o *Ramachandran Plot* gerado pela *webservice* SAVES com 93% dos aminoácidos em regiões favoráveis.

10

11

12

13

14

5.4 A PROTEÍNA MULTI-EPÍTOPO APRESENTOU ÓTIMOS PARÂMETROS IMUNOGÊNICOS

15

O *software* Vaxijen considerou nossa proteína como antigênica com um *score* de 0,9788.

16

O Allertop não encontrou nenhum epítipo da proteína que fosse considerado alérgico e o Toxinpred também não encontrou epítipos tóxicos para humanos. Indicando um importante potencial imunogênico da proteína multi-epítipo.

17

18

19

20

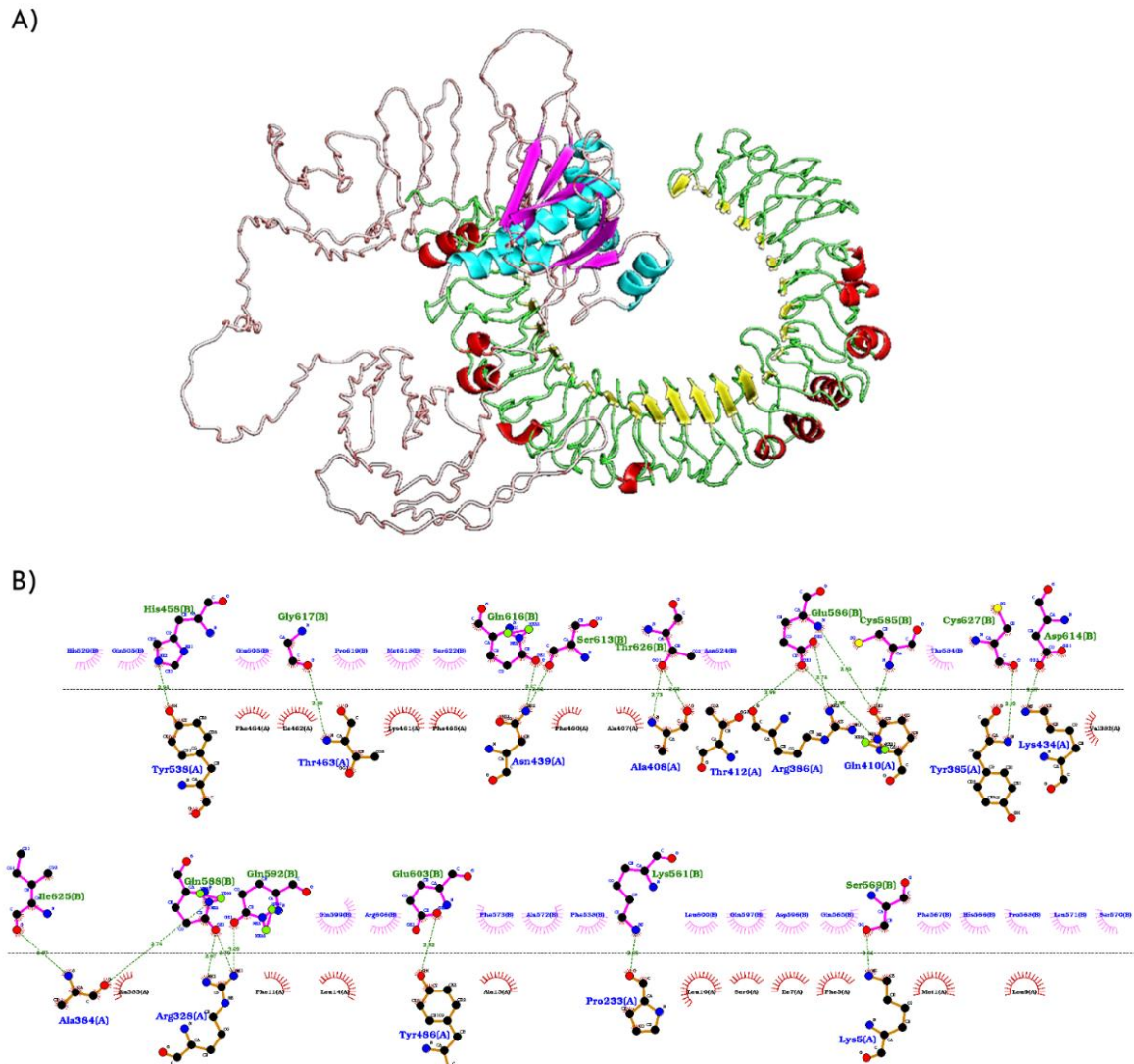
A massa molecular da proteína foi calculada em 57135,45 (57,14 KD) com um ponto isoeletrico (pI) teórico de 9.36. Ela apresentou ainda um índice de instabilidade de 18,89,

1 associado com o índice alifático de estabilidade medido em 76,62, o que segundo os parâmetros
2 do programa Protparam, ela é considerada estável. Este programa ainda calculou a meia-vida
3 da proteína em: 30h para reticulócitos de mamíferos *in vitro*, > 20h para leveduras *in vivo*, e >
4 10h em *E. coli in vivo*. A proteína apresentou uma média da hidropaticidade da sequência
5 (GRAVY) em -0,089. Por fim, de acordo com a ferramenta Protein-Sol, a solubilidade da
6 proteína foi de 0.427 e na análise de BLAST contra o proteoma humano (*Taxid: 9606*) não foi
7 encontrada homologia significativa.

8 9 5.5 A PROTEÍNA APRESENTOU ALTA AFINIDADE E BOA DINÂMICA 10 MOLECULAR QUANDO SE LIGA AO RECEPTOR TLR4

11 A análise de *docking molecular* mostrou que a nossa proteína multi-epítipo apresenta boa
12 afinidade de ligação com o receptor TLR4, apresentando 21 ligações de hidrogênio e 22
13 interações hidrofóbicas entre os resíduos com 17 átomos correspondentes (FIGURA 3B).
14 Segundo o programa ClusPro, o modelo 0 foi o melhor ranqueado para a ligação entre a proteína
15 e o receptor, apresentando um cluster com 53 membros, e um *score* ponderado de menor energia
16 da molécula em -1132,7 (FIGURA 3A).

1 Figura 3 - Plot da interação da proteína multi-epítipo com o receptor TLR4 e suas interações.



2

3 Fonte: Do Autor, 2022.

4 Nota: Técnica de *docking molecular* aplicada para a proteína multi-epítipo e o receptor do tipo Toll 4. Em A),
 5 observa-se a interação gerada pelo programa ClusPro entre a proteína (à esquerda, cor predominante cinza claro)
 6 com o receptor do tipo Toll 4 (à direita, cor predominante verde). Em B) é a visualização da interação dos resíduos
 7 que participam da interação entre as moléculas, gerado com o programa LigPlot+. Neste caso, em verde estão
 8 representadas as ligações de hidrogênio, os semicírculos mostrados a baixo da linha tracejada na coloração
 9 vermelha representam as interações hidrofóbicas, e os semicírculos mostrados acima da linha tracejada na
 10 coloração rosa representam seus átomos correspondentes.

11

12

13

14

15

16

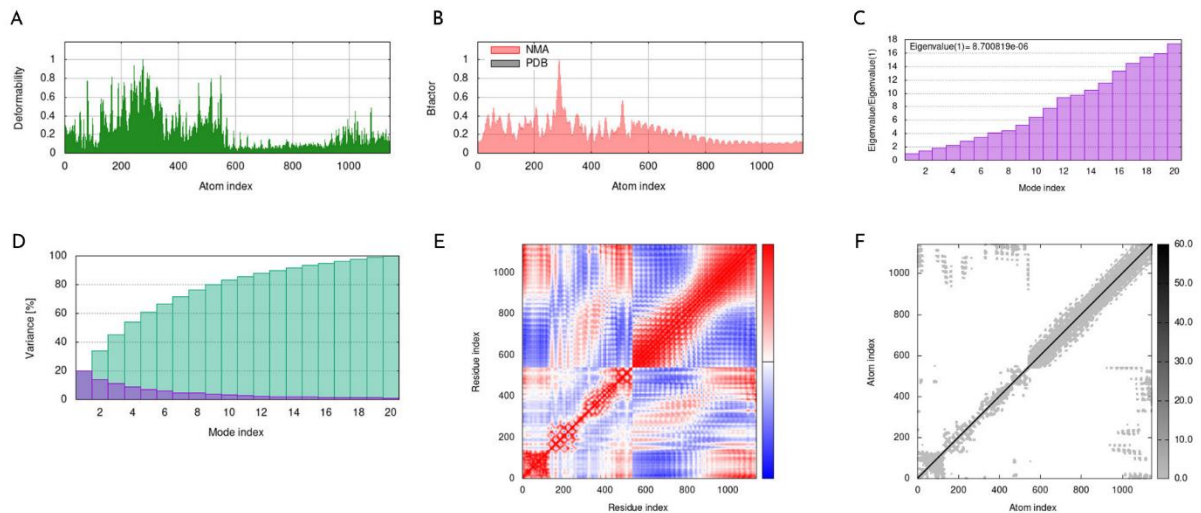
17

O complexo proteína-TLR4 foi utilizado para a simulação de dinâmica molecular da interação pelo programa IMods. Em A e B, na Figura 4, podemos observar valores mais altos entre 200 e 400 átomos, sendo mais baixos no restante dos gráficos, indicando uma flexibilidade da dinâmica desse complexo nos valores mais altos, e uma rigidez do complexo nos valores mais baixos. Em C, pode ser observados os valores próprios da dinâmica, no qual seu valor está

1 relacionado a quantidade de energia necessária para deformar a estrutura. No caso, observamos
 2 um valor de $8.700819e-06$, o que é considerado baixo, indicando que com pouca energia pode
 3 se deformar a estrutura.

4 Em D, a análise de variância apresenta o entendimento dos valores sendo inversamente
 5 proporcionais ao valor próprio. No caso, uma menor coloração roxa no gráfico indica maior
 6 estabilidade do complexo, o que condiz com o resultado encontrado nos gráficos A e B. O mapa
 7 de covariância em E mostra os movimentos relacionados (vermelho), não correlacionados
 8 (branco) e anticorrelacionados (azul) para avaliação da dinâmica do complexo. Este resultado
 9 mostra uma grande região apresentando movimentos relacionados no 2º e 3º quadrantes, e no
 10 1º e 4º algo mais homogêneo entre os três tipos de movimento. Por fim, em F é observado as
 11 redes elásticas no qual os pares de átomos são correlacionados, em qual quanto mais cinza
 12 escuro, mais rígidas são as molas. O gráfico apresenta na sua totalidade molas coradas em cinza
 13 claro, indicando também a flexibilidade na dinâmica desse complexo.

15 Figura 4 - Gráficos da simulação dinâmica para o complexo proteína multi-epítopo-ligante.



16 Fonte: Do Autor, 2022.

17 Nota: O gráfico A representa a deformidade, o gráfico B, o fator B, em C, os valores próprios, em D, a variância
 18 para o complexo. Em E, podemos ver um mapa de calor representando a covariância e o modelo elástico usados
 19 para o cálculo da dinâmica em F.

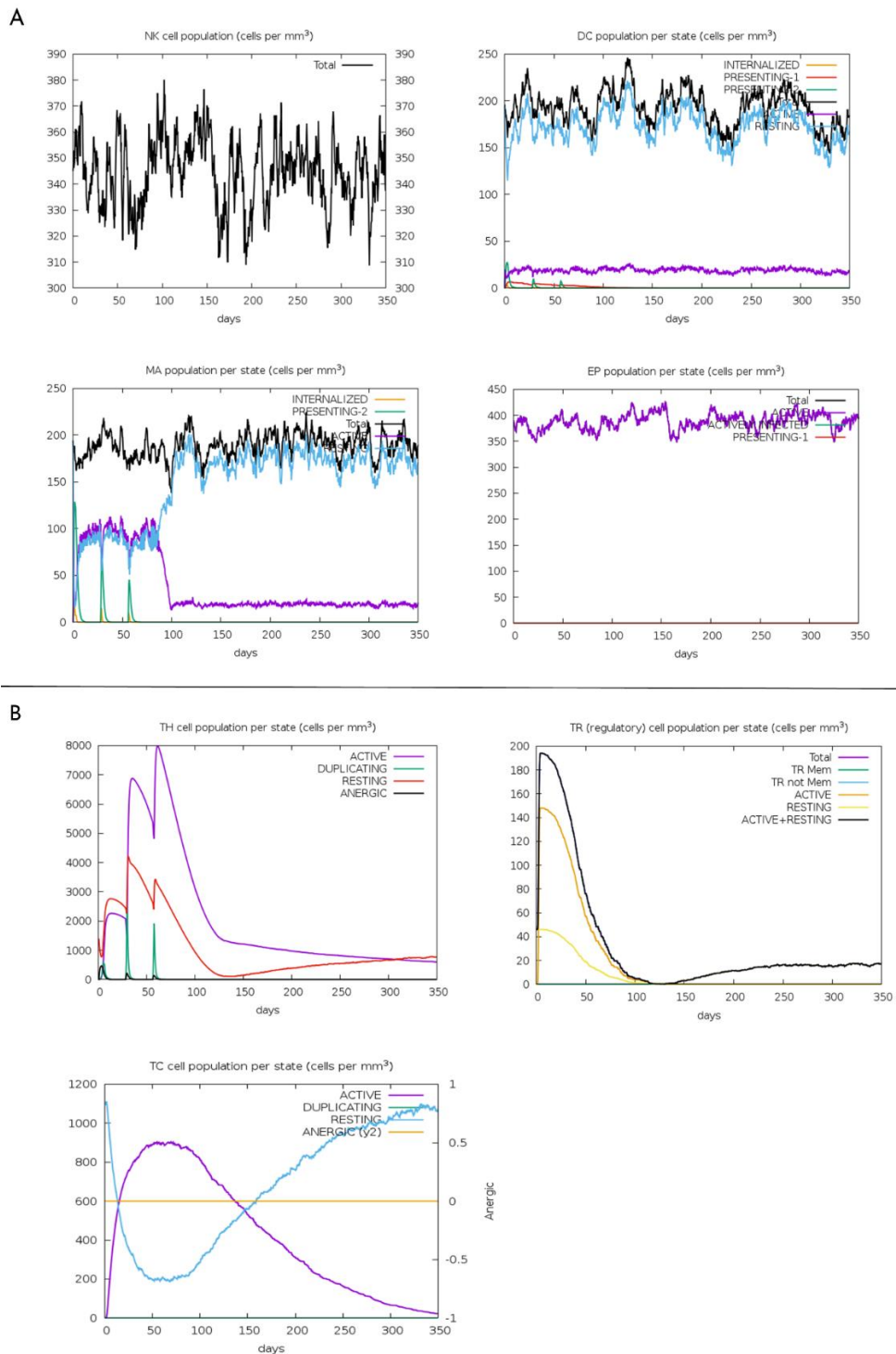
22 5.6 UMA VACINA CANDIDATA COM A PROTEÍNA MULTI-EPÍTOPO 23 APRESENTARIA UMA BOA RESPOSTA IMUNE

24 O programa IFNepitope identificou 559 epítomos, dos quais 153 epítomos foram
 25 considerados positivos, ou seja, com capacidade de induzir a produção de Interferon- γ e 406

1 foram considerados com pontuação negativa, ou seja, não tinham capacidade de induzir a
2 produção desta citocina (ANEXO C – TABELA 9). O *webservice* IL17eScan não encontrou
3 epítomos com capacidades indutoras da Interleucina 17.

4 Os resultados da simulação imune mostram que uma vacina baseada em nossa proteína
5 possui grande potencial de estimular a ativação de células natural killer por todo o tempo após
6 a injeção e também macrófagos após a primeira e segunda dose. As células dendríticas
7 aparecem na maior parte em estado de repouso, porém uma parcela mostra-se ativa após a
8 injeção; as células epiteliais aparecem ativas ao longo de todo o tempo (FIGURA 5A). A
9 previsão apontou um aumento das populações de células T auxiliares, T reguladoras e T
10 citotóxicas durante a administração das doses da vacina, com maior ativação de células T
11 auxiliares, seguida pelas células T citotóxicas e por último as T reguladoras (FIGURA 5B).

1 Figura 5 - Simulação da resposta imune *in silico* gerada pela proteína multi-epítipo em relação
 2 às células T, *natural killer*, dendríticas, macrófagos e células epiteliais.



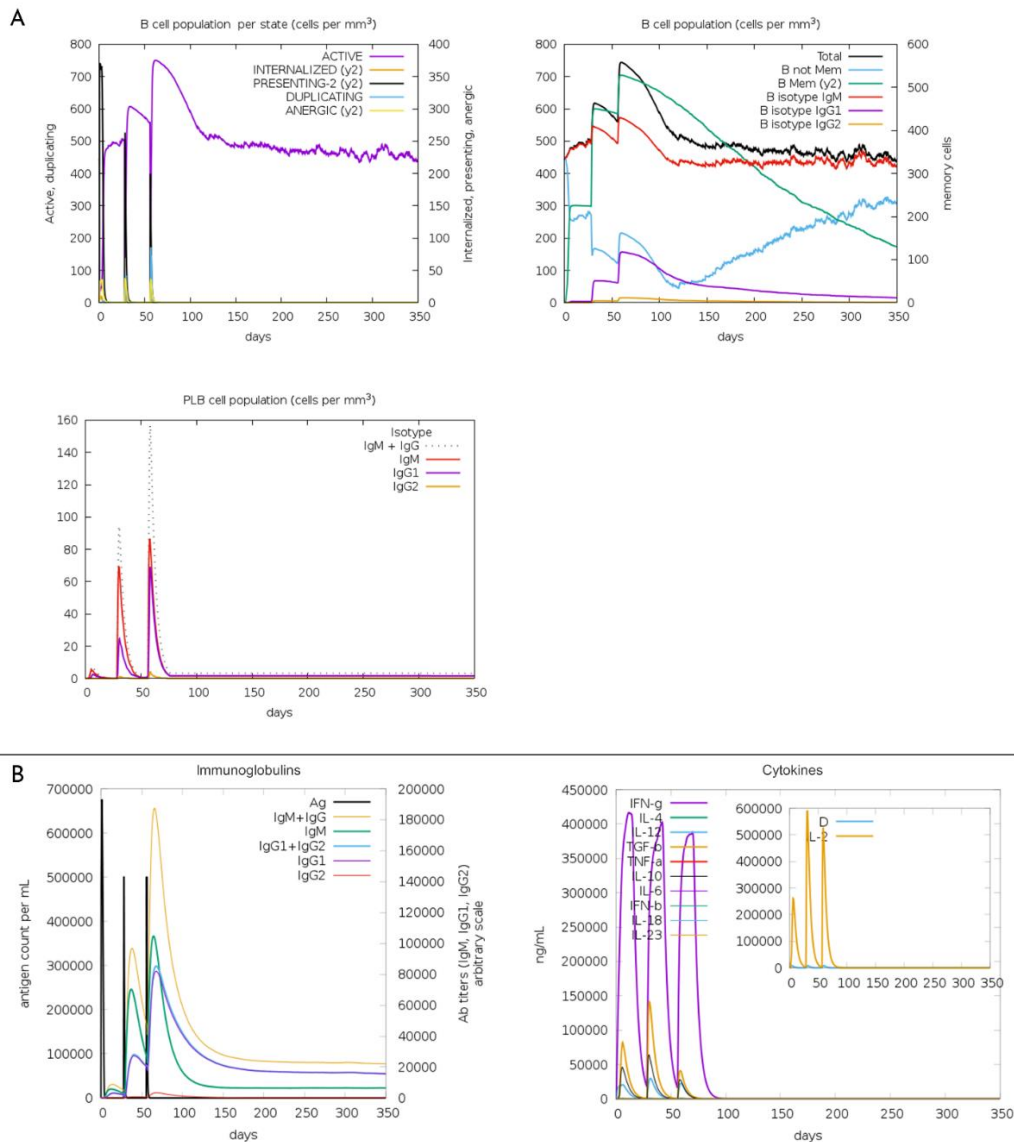
3
 4
 5
 6
 7
 8

Fonte: Do Autor, 2022.

Nota: Em A, estão representados os dados da simulação para as populações de células *natural killers* (NK), células dendríticas (DC), macrófagos (MA) e células epiteliais (EP). Em B, estão representados os dados da simulação para as populações de células T auxiliares (TH), células T regulatórias (Treg) e células T citotóxicas (TC).

1 A administração dessa vacina ainda mostrou fortemente um aumento de populações
 2 ativas de células B ao longo do período de injeções *in silico*. Nessa população ativa, pode-se
 3 observar um aumento de células B de memória com posterior queda e aumento e manutenção
 4 das células B do isotipo IgM. Durante esse período há uma queda na quantidade de células B
 5 sem memória, com posterior aumento (FIGURA 6A). Por fim, ainda foi observado um aumento
 6 da síntese de imunoglobulinas, principalmente da concentração de IgM + IgG; além de aumento
 7 significativo de IFN- γ , seguido de aumento discreto de TGF- β , com diminuição de todas as
 8 citocinas após a administração das doses da vacina (FIGURA 6B).

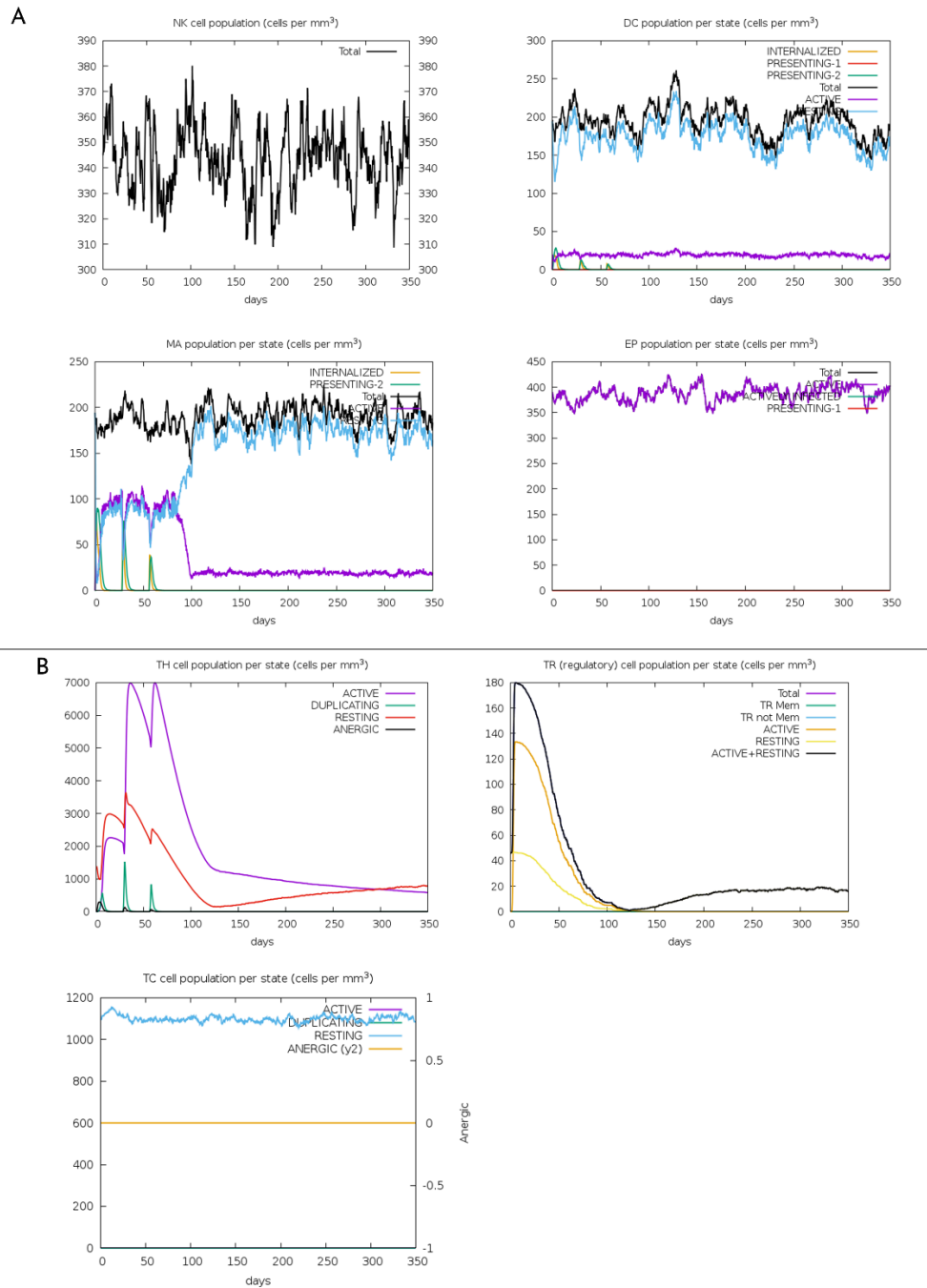
9 Figura 6 - Simulação da resposta imune *in silico* gerada pela proteína multi-epítipo em relação
 10 às células B, citocinas e imunoglobulinas.



1 Nota: Em A, estão representados os dados da simulação para as populações de células B e células B no plasma.
2 Em B, estão representados os dados da simulação relacionados as imunoglobulinas (linhas coloridas) durante as
3 injeções (linhas pretas verticais - Ag) no gráfico à esquerda; e, citocinas (linhas coloridas) e o valor de diversidade
4 (letra D, representado em azul) de clones de células T, no gráfico à direita.
5

6
7 Para finalizar, ainda realizamos a simulação *in silico* da administração da vacina se fosse
8 composta somente pelo adjuvante, como forma de avaliar a nossa proteína multi-epítipo
9 (FIGURAS 7 e 8). A simulação mostrou uma manutenção dos níveis de macrófagos, *natural*
10 *killers* e células epiteliais e um leve aumento das células da resposta inata como as células
11 dendríticas (FIGURA 7A). Pode-se observar uma ativação das células T auxiliares e T
12 regulatórias durante as aplicações da vacina, e uma manutenção das células T citotóxicas em
13 estado de repouso em comparação com a proteína multi-epítipo (FIGURA 7B). Para as células
14 B, foi observado um padrão parecido de estimulação, porém com menor produção de células B
15 e menos células B relacionadas ao estado de apresentação (FIGURA 8A). Por fim, a produção
16 de citocinas também apresenta o mesmo padrão, ao contrário da produção de imunoglobulinas,
17 que parecem ter mecanismos diferentes, com uma estimulação mais dependente do adjuvante
18 (FIGURA 8B).

1 Figura 7 - Simulação da resposta imune *in silico* gerada somente pelo adjuvante em relação às
 2 células T, *natural killer*, dendríticas, macrófagos e células epiteliais.

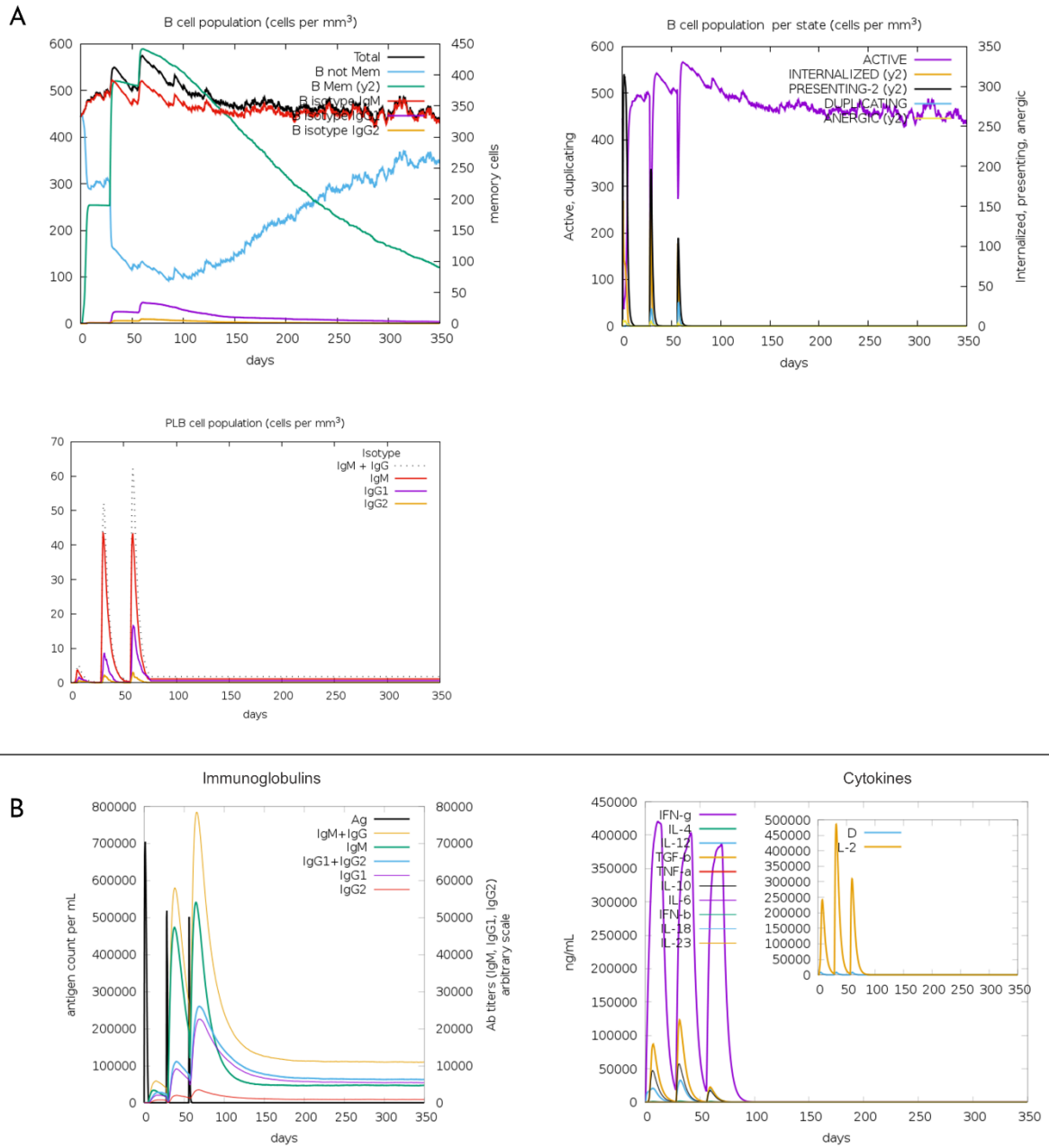


3
 4
 5
 6
 7
 8

Fonte: Do Autor, 2022.

Nota: Em A, estão representados os dados de simulação para as populações de células *natural killer* (NK), células dendríticas (DC), macrófagos (MA), e células epiteliais (EP). Em B, são representados os dados de simulação para as populações de células T auxiliares (TH), células T reguladoras (Treg) e células T citotóxicas (TC).

1 Figura 8 - Simulação da resposta imune *in silico* gerada somente pelo adjuvante em relação às
 2 células B, citocinas e imunoglobulinas.



3
 4
 5
 6
 7

Fonte: Do Autor, 2022.
 Nota: Em A, são representados os dados de simulação para as populações de células B e de células B no plasma. Em B, os dados de simulação relacionados às imunoglobulinas e citocinas, respectivamente, são representados.

6 DISCUSSÃO

A vacinação é considerada como o melhor meio de prevenção contra as mais diversas doenças existentes. As abordagens mais recentes de desenvolvimento de vacinas, como as vacinas de subunidade, partem de estudos relacionados a imunoinformática. Essas abordagens vêm crescendo cada vez mais devido a vantagens como a aceleração da produção das vacinas e detecção de imunógenos mais eficazes. Dentro dessas vacinas de subunidade, existe a chamada vacina multi-epítipo. Esse tipo de vacina permite a escolha de imunógenos que possuem capacidade de estimular uma resposta imune tanto inata quanto adaptativa, induzindo ainda mais o desenvolvimento de uma memória mais adaptativa.

Com isso, utilizamos como base metodológica estudos anteriores, como de Vilela-Rodrigues e colaboradores (2022) e Lemes e colaboradores (2022), para selecionar os melhores epítipos das proteínas centrais classificadas como potenciais candidatas a alvo de vacina contra as bactérias do gênero *Rickettsia* por Felice e colaboradores (2021). As proteínas centrais, chamadas de *core* genoma, implicam um papel muito importante em representar todo o conjunto de organismos que está sendo estudado, logo permite adquirir um conhecimento sobre diversidade das espécies desse conjunto e conseqüentemente, como elas se comportam (JUNCKER et al., 2009).

No processo de desenvolvimento de vacinas, dos diversos tipos existentes, a principal preocupação dos pesquisadores é em relação à eficácia da vacina para a população global como um todo. Este trabalho utilizou-se dessa premissa e selecionou aqueles alelos de MHC que são considerados mais frequentes na população global, para depois selecionar os epítipos das proteínas que se ligavam aos complexos MHCI/CTL, MHCII/HTL e também com células B, como forma de estimular a imunidade celular e humoral. Dessa forma, 26 epítipos foram selecionados de 7 proteínas para a construção de uma proteína multi-epítipo a ser utilizada como uma vacina peptídica quimérica contra as bactérias do gênero *Rickettsia*.

Esses 26 epítipos foram retirados das proteínas WP_012148219.1, WP_012149174.1 e WP_012149175.1, pertencentes a família das Porinas, WP_012148227.1, pertencente à família OmpW, WP_012148552.1, conhecida como uma chaperona de cobre, WP_012148811.1, uma dioxigenase, e por fim, WP_012148819.1, uma D-alanil-D-alanina carboxipeptidase. As quatro primeiras proteínas pertencem a uma grande família chamada proteínas de membrana externa (OMPs) que são encontrados dentro dos folhetos da membrana externa de bactérias *Gram*-negativas. Essas OMPs formam barris/poros dentro dessa membrana que podem ter diversas funções como auxiliar na estrutura dessas membranas, auxiliar na adesão a células, no metabolismo lipídico e também no transporte de substâncias hidrofílicas, dificultado pela

1 presença de lipopolissacarídeos nesse folheto de membrana externa (DE LA CRUZ; CALVA,
2 2010; HONG et al., 2006). Essas funções se tornaram muito exploradas visto que dados
3 demonstraram que esses canais auxiliavam na patogênese de alguns organismos, conferindo
4 certa resistência relacionada a sobrevivência dos organismos ao inibir ou facilitar o transporte
5 de substâncias através da membrana (SAIER JR., 2007).

6 As chaperonas possuem funções relacionadas a degradação de componentes
7 intracelulares indesejados, importante para ajudar na adaptação celular a mudanças ambientais
8 e na resposta contra diversos tipos de estresses vindos do meio intra ou extracelular
9 (BEJARANO E, 2010). As dioxigenases e as carboxipeptidases são moléculas bioquímicas que
10 iniciam a degradação e clivagem de compostos aromáticos e aminoácidos básicos que são
11 metabolicamente ativos no processo de sobrevivência celular bacteriana (DOBLE; KUMAR,
12 2005; MATTHEWS; MUELLER-ORTIZ; WETSEL, 2004). Todas as funções relacionadas a
13 essas proteínas demonstram interação na relação patógeno/hospedeiro e com as doenças
14 causadas, apoiando a ideia de que a vacina baseada nos epítomos selecionados potencializam a
15 eficácia e a indução de uma boa resposta imune na população global contra as rickettsias. Além
16 disso, na proteína final foram utilizados ligantes entre os epítomos selecionados como forma de
17 melhorar o processamento e apresentação de antígenos dentro do hospedeiro, ajudar na
18 flexibilidade estrutural da proteína e a reduzir a probabilidade de formação de antígenos
19 juncionais (CHEN; ZARO; SHEN, 2013b).

20 Sabe-se que as vacinas clássicas podem utilizar o patógeno inteiro, o qual pode possuir
21 proteínas semelhantes às do hospedeiro, induzindo diversos tipos de reações adversas. Seguindo
22 a mesma linha das vacinas anteriores citadas neste trabalho acima, nossa proteína foi
23 cuidadosamente construída para que não houvesse nenhuma homologia com o hospedeiro, não
24 induzindo tais reações adversas. A proteína também foi avaliada quanto à sua toxicidade e sua
25 alergenicidade, sendo considerada não tóxica e não alérgica para o humano. Outra forma de
26 avaliar a interação com o hospedeiro, se deve a questões da meia-vida, dinâmica molecular e
27 solubilidade da vacina. Uma meia vida baixa desse método de prevenção ou sua pouca
28 solubilidade podem pôr em risco a sua aplicação prática (MATHUR et al., 2016;
29 SMIALOWSKI et al., 2007). A meia vida do nosso constructo foi de pelo menos 10h em células
30 de *E. coli in vivo*, com uma solubilidade de 0,427, abaixo do limiar de 0,45; ela foi considerada
31 estável e com um índice de hidropaticidade de -0.089. Apesar de nossa vacina multi-epítomo ir
32 em contrapartida de trabalhos como o de Vilela-Rodrigues (VILELA RODRIGUES et al.,
33 2022); Nandi (2022); e apresentar uma hidropaticidade com tendência a interação com meio

1 aquoso, os outros resultados parecem seguir seus padrões, sugerindo que pode ser considerada
2 como uma vacina promissora para testes *in vitro* e *in vivo*.

3 A rickettsiologia é uma área que têm o foco no estudo e análise crítica da patogênese das
4 bactérias pertencentes ao gênero *Rickettsia*. Dentro dessa área de estudo, a etapa de interação
5 com a células hospedeira e seus mecanismos subjacentes são pontos chaves para entender o
6 processo inflamatório causado pela infecção dessas bactérias. Mecanismos como a adesão às
7 células hospedeiras, indução de fagocitose, aquisição ou degradação de nutrientes, por exemplo,
8 resultante da utilização de funções das proteínas utilizadas em nossa vacina, ajudam a entender
9 o processo inflamatório gerado como resposta imune (WALKER; VALBUENA; OLANO,
10 2003). O processo de danos biológicos e produção de espécies reativas de oxigênio são
11 iniciados sempre que o patógeno é reconhecido por receptores como os do tipo Toll (JORDAN
12 et al., 2009; QUEVEDO-DIAZ et al., 2010; RUMFIELD et al., 2020b) e que gera citocinas
13 inflamatórias, como IL-17 e IFN- γ (JORDAN et al., 2008b; TURCO; WINKLER, [s.d.]). As
14 interações entre essas proteínas e os receptores do tipo Toll também podem ajudar na indução
15 de respostas imunes inatas e adaptativas (CLEM, 2011). Nesse sentido, uma gama de epítomos
16 (153) presentes em nossa proteína foram considerados como bons indutores de IFN- γ ; apesar
17 disso, nenhum deles foi considerado como indutor pela análise que fizemos para a IL-17. A
18 presença de epítomos indutores de IFN- γ em nossa vacina apoia seu potencial imune, o que
19 também foi encontrado nos trabalhos de testes *in vivo* de Cao (2021); Vakili (2020); Thsilwane
20 (2019); teste *in vitro* de KÖSEOĞLU (2022) e testes *in silico* de Vilela-Rodrigues (VILELA
21 RODRIGUES et al., 2022) Nandi (2022); para a criação de vacinas multi-epítomos.

22 Dentro da imunoinformática e das análises *in silico*, diversos preditores seguem sendo
23 desenvolvidos com resultados que demonstram boas precisões de eficácia dessas vacinas,
24 porém o campo que estuda a estimulação de respostas imunes do hospedeiro, parece ainda ser
25 bastante complexo (SIX et al., 2012). Apesar disso, tentamos avaliar a resposta tanto inata
26 quanto adaptativa para nossa proteína. Durante os tempos que injeção da vacina, foi possível
27 perceber que há uma estimulação de células que têm esse primeiro contato com o hospedeiro,
28 como macrófagos e células NK, com ativação dos linfócitos T e linfócitos B, o que sugere que
29 nossa vacina consegue ativar as respostas imune inata e adaptativa e ainda estimular células de
30 memória contra um segundo contato contra esse patógeno. Apesar da ferramenta não permitir
31 a identificação de quais células são as principais responsáveis por desencadear essa
32 estimulação, verifica-se também uma produção de certa forma equilibrada entre todas as
33 citocinas inflamatórias, com aumento de TGF- β e IFN- γ , sugerindo que nossa vacina projetada
34 possa gerar efeitos colaterais, porém considerados mínimos.

1 Apesar de técnicas diferentes, os ensaios de febre aftosa em animais vacinados com a
2 vacina multi-epítopo B4 do trabalho de TSHILWANE (2019) também apresentaram um
3 aumento de citocinas e células com um padrão de resposta Th1. Com a abordagem *in silico*,
4 Vilela-Rodrigues (VILELA RODRIGUES et al., 2022) também encontrou aumento de
5 macrófagos e células NK e ativação de linfócitos T e B, apoiando o potencial imune de nossos
6 constructos.

7 CONCLUSÃO

As bactérias do gênero *Rickettsia* são as responsáveis por diversas doenças como as febres maculosas e os tifos em seres humanos. Essas doenças são consideradas de alto contágio e até o momento não existem meios preventivos que sejam realmente eficazes contra elas. Desse modo, a criação de uma vacina se torna o principal objetivo dos pesquisadores para diminuir a prevalência delas ao redor do mundo. Neste trabalho, utilizamos uma abordagem de imunoinformática *in silico* para construir uma vacina multi-epítipo capaz de ser eficaz contra todas as bactérias do gênero, gerando tanto uma resposta imune inata quanto adaptativa no seu hospedeiro.

De uma forma geral, todas as validações *in silico* aqui feitas suportam a ideia de que nossa vacina é uma forte candidata para atuar contra a infecção causada por essas bactérias. No entanto, são necessárias outras validações, dessa vez, *in vitro* e *in vivo* para confirmar a sua eficácia e segurança em seres humanos.

REFERÊNCIAS

- 1
2 AIZAT, W. M.; ISMAIL, I.; NOOR, N. M. Recent Development in Omics Studies. Em: [s.l:
3 s.n.]. p. 1–9.
- 4 AKRAM, S. M.; LADD, M.; KING, K. C. **Rickettsia Prowazekii**. [s.l: s.n.].
- 5 ALI, M. et al. Exploring dengue genome to construct a multi-epitope based subunit vaccine by
6 utilizing immunoinformatics approach to battle against dengue infection. **Scientific Reports**,
7 v. 7, n. 1, p. 9232, 23 dez. 2017.
- 8 ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v.
9 215, n. 3, p. 403–410, out. 1990.
- 10 ANGELAKIS, E.; BECHAH, Y.; RAOULT, D. The History of Epidemic Typhus.
11 **Microbiology Spectrum**, v. 4, n. 4, 12 ago. 2016.
- 12 AZMI, F. et al. Recent progress in adjuvant discovery for peptide-based subunit vaccines.
13 **Human Vaccines & Immunotherapeutics**, v. 10, n. 3, p. 778–796, 3 mar. 2014.
- 14 BACKERT, L.; KOHLBACHER, O. Immunoinformatics and epitope prediction in the age of
15 genomic medicine. **Genome Medicine**, v. 7, n. 1, p. 119, 20 dez. 2015.
- 16 BAH, S. Y. et al. Highlights on the Application of Genomics and Bioinformatics in the Fight
17 Against Infectious Diseases: Challenges and Opportunities in Africa. **Frontiers in Genetics**, v.
18 9, 27 nov. 2018.
- 19 BECHAH, Y. et al. Epidemic typhus. **The Lancet Infectious Diseases**, v. 8, n. 7, p. 417–426,
20 jul. 2008.
- 21 BEJARANO E, C. A. Chaperone-mediated autophagy. **Proc Am Thorac Soc**, v. 7, 2010.
- 22 BERMAN, H. M. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 1
23 jan. 2000.
- 24 BERMÚDEZ, S.; TROYO, A. A review of the genus *Rickettsia* in Central
25 America. **Research and Reports in Tropical Medicine**, v. Volume 9, p. 103–112, jun. 2018.
- 26 BIGGS, H. M. et al. Diagnosis and Management of Tickborne Rickettsial Diseases: Rocky
27 Mountain Spotted Fever and Other Spotted Fever Group Rickettsioses, Ehrlichioses, and
28 Anaplasmosis — United States. **MMWR. Recommendations and Reports**, v. 65, n. 2, p. 1–
29 44, 13 maio 2016.
- 30 BLANTON, L. S. The Rickettsioses. **Infectious Disease Clinics of North America**, v. 33, n.
31 1, p. 213–229, mar. 2019.
- 32 BLANTON, L. S.; WALKER, D. H. Flea-Borne Rickettsioses and Rickettsiae. **The American**
33 **Journal of Tropical Medicine and Hygiene**, v. 96, n. 1, p. 53–56, 11 jan. 2017.

- 1 BURNS, J. N.; ACUNA-SOTO, R.; STAHL, D. W. Drought and Epidemic Typhus, Central
2 Mexico, 1655–1918. **Emerging Infectious Diseases**, v. 20, n. 3, p. 442–447, mar. 2014.
- 3 CALIS, J. J. A. et al. Properties of MHC Class I Presented Peptides That Enhance
4 Immunogenicity. **PLoS Computational Biology**, v. 9, n. 10, p. e1003266, 24 out. 2013.
- 5 CAO, H. et al. Multi-Omics Approach Reveals the Potential Core Vaccine Targets for the
6 Emerging Foodborne Pathogen *Campylobacter jejuni*. **Frontiers in Microbiology**, v. 12, 2021.
- 7 CAO, Y. et al. Poly(I:C) combined with multi-epitope protein vaccine completely protects
8 against virulent foot-and-mouth disease virus challenge in pigs. **Antiviral Research**, v. 97, n.
9 2, p. 145–153, fev. 2013.
- 10 CHAUHAN, V. et al. Designing a multi-epitope based vaccine to combat Kaposi Sarcoma
11 utilizing immunoinformatics approach. **Scientific Reports**, v. 9, n. 1, p. 2517, 21 dez. 2019.
- 12 CHEN, J. et al. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale.
13 **Amino Acids**, v. 33, n. 3, p. 423–428, 26 set. 2007.
- 14 CHEN, X.; ZARO, J. L.; SHEN, W.-C. Fusion protein linkers: Property, design and
15 functionality. **Advanced Drug Delivery Reviews**, v. 65, n. 10, p. 1357–1369, out. 2013a.
- 16 CHEN, X.; ZARO, J. L.; SHEN, W.-C. Fusion protein linkers: Property, design and
17 functionality. **Advanced Drug Delivery Reviews**, v. 65, n. 10, p. 1357–1369, out. 2013b.
- 18 CLEM, A. Fundamentals of vaccine immunology. **Journal of Global Infectious Diseases**, v.
19 3, n. 1, p. 73, 2011.
- 20 COLONNE, P. M.; EREMEEVA, M. E.; SAHNI, S. K. Beta Interferon-Mediated Activation
21 of Signal Transducer and Activator of Transcription Protein 1 Interferes with *Rickettsia conorii*
22 Replication in Human Endothelial Cells. **Infection and Immunity**, v. 79, n. 9, p. 3733–3743,
23 set. 2011.
- 24 DAHLGREN, F. S. et al. National Surveillance of Spotted Fever Group Rickettsioses in the
25 United States, 2008–2012. **The American Journal of Tropical Medicine and Hygiene**, v. 94,
26 n. 1, p. 26–34, 6 jan. 2016.
- 27 DAVOUST B et al. Are vertebrates reservoir hosts for *Rickettsia*? **Acad. Vet. France**, v. 163,
28 p. 291–302, 2010.
- 29 DE LA CRUZ, M. Á.; CALVA, E. The Complexities of Porin Genetic Regulation. **Microbial**
30 **Physiology**, v. 18, n. 1, p. 24–36, 2010.
- 31 DELORD, M.; SOCOLOVSKI, C.; PAROLA, P. Rickettsioses and Q fever in travelers
32 (2004–2013). **Travel Medicine and Infectious Disease**, v. 12, n. 5, p. 443–458, set. 2014.
- 33 DHANDA, S. K.; VIR, P.; RAGHAVA, G. P. Designing of interferon-gamma inducing MHC
34 class-II binders. **Biology Direct**, v. 8, n. 1, p. 30, 5 dez. 2013.

- 1 DIMITROV, I. et al. AllerTOP v.2—a server for in silico prediction of allergens. **Journal of**
2 **Molecular Modeling**, v. 20, n. 6, p. 2278, 31 jun. 2014.
- 3 DOBLE, M.; KUMAR, A. Chlorinated Hydrocarbons and Aromatics, and Dioxins. Em:
4 **Biotreatment of Industrial Effluents**. [s.l.] Elsevier, 2005. p. 65–82.
- 5 DOYTCHINOVA, I. A.; FLOWER, D. R. Bioinformatic approach for identifying parasite and
6 fungal candidate subunit vaccines. **Open Vaccine Journal**, v. 3, n. 1, p. 22–26, 2010.
- 7 DUMLER, J. S. Clinical Disease: Current Treatment and New Challenges. Em: **Intracellular**
8 **Pathogens II**. Washington, DC: ASM Press, 2014. p. 1–39.
- 9 DUNN, J. et al. Inadequacy of IgM Antibody Tests for Diagnosis of Rocky Mountain Spotted
10 Fever. **The American Journal of Tropical Medicine and Hygiene**, v. 91, n. 4, p. 767–770, 1
11 out. 2014.
- 12 EL-MANZALAWY, Y.; DOBBS, D.; HONAVAR, V. Predicting flexible length linear B-cell
13 epitopes. **Computational systems bioinformatics. Computational Systems Bioinformatics**
14 **Conference**, v. 7, p. 121–32, 2008.
- 15 FANG, R.; BLANTON, L. S.; WALKER, D. H. Rickettsiae as Emerging Infectious Agents.
16 **Clinics in Laboratory Medicine**, v. 37, n. 2, p. 383–400, jun. 2017.
- 17 FELICE, A. G. et al. Pan-genomic analyses of 47 complete genomes of the Rickettsia genus
18 and prediction of new vaccine targets and virulence factors of the species. **Journal of**
19 **Biomolecular Structure and Dynamics**, 2021.
- 20 GALVÃO, M. A. M. et al. Riquetsioses no Brasil e Portugal: ocorrência, distribuição e
21 diagnóstico. **Revista de Saúde Pública**, v. 39, n. 5, p. 850–856, out. 2005.
- 22 GASTEIGER E., H. C. , G. A. , D. S. , W. M. R. , A. R. D. , B. A. Protein Identification and
23 Analysis Tools on the ExPASy Server. Em: **The Proteomics Protocols Handbook, Humana**
24 **Press**. [s.l: s.n.]. p. 571–607.
- 25 GAUTHIER, J. et al. A brief history of bioinformatics. **Briefings in Bioinformatics**, v. 20, n.
26 6, p. 1981–1996, 27 nov. 2019.
- 27 GAYWEE, J. et al. Human Infection with *Rickettsia* sp. related to *R. japonica* , Thailand.
28 **Emerging Infectious Diseases**, v. 13, n. 4, p. 657–659, abr. 2007.
- 29 GILLESPIE, J. J. et al. Plasmids and Rickettsial evolution: Insight from *Rickettsia felis*. **PLoS**
30 **ONE**, v. 2, n. 3, 2007.
- 31 GUPTA, S. et al. In Silico Approach for Predicting Toxicity of Peptides and Proteins. **PLoS**
32 **ONE**, v. 8, n. 9, p. e73957, 13 set. 2013.
- 33 GUPTA, S. et al. IL17eScan: A Tool for the Identification of Peptides Inducing IL-17
34 Response. **Frontiers in Immunology**, v. 8, 31 out. 2017.

- 1 HAGEN, J. B. The origins of bioinformatics. **Nature Reviews Genetics**, v. 1, n. 3, p. 231–236,
2 dez. 2000.
- 3 HAMAGUCHI, S. et al. Clinical and Epidemiological Characteristics of Scrub Typhus and
4 Murine Typhus among Hospitalized Patients with Acute Undifferentiated Fever in Northern
5 Vietnam. **The American Journal of Tropical Medicine and Hygiene**, v. 92, n. 5, p. 972–978,
6 6 maio 2015.
- 7 HE, Y.; XIANG, Z.; MOBLEY, H. L. T. Vaxign: The first web-based vaccine design program
8 for reverse vaccinology and applications for vaccine development. **Journal of Biomedicine
9 and Biotechnology**, v. 2010, 2010.
- 10 HEBDITCH, M. et al. Protein–Sol: a web tool for predicting protein solubility from sequence.
11 **Bioinformatics**, v. 33, n. 19, p. 3098–3100, 1 out. 2017.
- 12 HEO, L.; PARK, H.; SEOK, C. GalaxyRefine: protein structure refinement driven by side-
13 chain repacking. **Nucleic Acids Research**, v. 41, n. W1, p. W384–W388, 1 jul. 2013.
- 14 HONG, H. et al. The Outer Membrane Protein OmpW Forms an Eight-stranded β -Barrel with
15 a Hydrophobic Channel. **Journal of Biological Chemistry**, v. 281, n. 11, p. 7568–7577, mar.
16 2006.
- 17 HU, W. et al. A flexible peptide linker enhances the immunoreactivity of two copies HBsAg
18 preS1 (21–47) fusion protein. **Journal of Biotechnology**, v. 107, n. 1, p. 83–90, jan. 2004.
- 19 JORDAN, J. M. et al. The Absence of Toll-Like Receptor 4 Signaling in C3H/HeJ Mice
20 Predisposes Them to Overwhelming Rickettsial Infection and Decreased Protective Th1
21 Responses. **Infection and Immunity**, v. 76, n. 8, p. 3717–3724, ago. 2008a.
- 22 JORDAN, J. M. et al. The Absence of Toll-Like Receptor 4 Signaling in C3H/HeJ Mice
23 Predisposes Them to Overwhelming Rickettsial Infection and Decreased Protective Th1
24 Responses. **Infection and Immunity**, v. 76, n. 8, p. 3717–3724, ago. 2008b.
- 25 JORDAN, J. M. et al. Rickettsiae Stimulate Dendritic Cells through Toll-Like Receptor 4,
26 Leading to Enhanced NK Cell Activation In Vivo. **The Journal of Infectious Diseases**, v. 199,
27 n. 2, p. 236–242, 15 jan. 2009.
- 28 JUNCKER, A. S. et al. Systematic Characterisation of Cellular Localisation and Expression
29 Profiles of Proteins Containing MHC Ligands. **PLoS ONE**, v. 4, n. 10, p. e7448, 14 out. 2009.
- 30 KÖSEOĞLU, A. E. et al. Molecular characterization of Anaplasma ovis Msp4 protein in strains
31 isolated from ticks in Turkey: A multi-epitope synthetic vaccine antigen design against
32 Anaplasma ovis using immunoinformatic tools. 2022.
- 33 KOZAKOV, D. et al. The ClusPro web server for protein–protein docking. **Nature Protocols**,
34 v. 12, n. 2, p. 255–278, 12 fev. 2017.

- 1 KUO, C.-C. et al. Significance of major international seaports in the distribution of murine
2 typhus in Taiwan. **PLOS Neglected Tropical Diseases**, v. 11, n. 3, p. e0005430, 6 mar. 2017.
- 3 LARSEN, M. V et al. Large-scale validation of methods for cytotoxic T-lymphocyte epitope
4 prediction. **BMC Bioinformatics**, v. 8, n. 1, p. 424, 31 dez. 2007.
- 5 LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of
6 protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, 1 abr. 1993.
- 7 LEI, Y. et al. Application of built-in adjuvants for epitope-based vaccines. **PeerJ**, v. 6, p. e6185,
8 14 jan. 2019.
- 9 LEMES, M. R. et al. In silico designing of a recombinant multi-epitope antigen for leprosy
10 diagnosis. **Journal of Genetic Engineering and Biotechnology**, v. 20, n. 1, p. 128, 2 set. 2022.
- 11 LI, G. et al. Construction of a linker library with widely controllable flexibility for fusion
12 protein design. **Applied Microbiology and Biotechnology**, v. 100, n. 1, p. 215–225, 22 jan.
13 2016.
- 14 LIMA, D. S. et al. ASPECTOS DO GÊNERO RICKETTSIA: UMA REVISÃO
15 SISTEMÁTICA. **Educação, Ciência e Saúde**, v. 7, n. 1, 28 jun. 2020.
- 16 LÓPEZ-BLANCO, J. R. et al. iMODS: internal coordinates normal mode analysis server.
17 **Nucleic Acids Research**, v. 42, n. W1, p. W271–W276, 1 jul. 2014.
- 18 MA, Y. Recent advances in nontoxic *Escherichia coli* heat-labile toxin and its derivative
19 adjuvants. **Expert Review of Vaccines**, v. 15, n. 11, p. 1361–1371, 12 nov. 2016.
- 20 MANSUETO, P. et al. New Insight into Immunity and Immunopathology of Rickettsial
21 Diseases. **Clinical and Developmental Immunology**, v. 2012, p. 1–26, 2012.
- 22 MATHUR, D. et al. PEPlife: A Repository of the Half-life of Peptides. **Scientific Reports**, v.
23 6, n. 1, p. 36617, 7 dez. 2016.
- 24 MATTHEWS, K. W.; MUELLER-ORTIZ, S. L.; WETSEL, R. A. Carboxypeptidase N: a
25 pleiotropic regulator of inflammation. **Molecular Immunology**, v. 40, n. 11, p. 785–793, jan.
26 2004.
- 27 MIRDITA, M. et al. ColabFold: making protein folding accessible to all. **Nature Methods**, v.
28 19, n. 6, p. 679–682, 30 jun. 2022.
- 29 MOYLE, P.; TOTH, I. Self-Adjuvanting Lipopeptide Vaccines. **Current Medicinal**
30 **Chemistry**, v. 15, n. 5, p. 506–516, 1 fev. 2008.
- 31 NANDI, A. et al. **Designing of multi-epitope vaccine construct employing immuno-**
32 **informatics approach to combat multi-tick species infestations.** [s.l: s.n.]. Disponível em:
33 <<https://ssrn.com/abstract=4046359>>.

- 1 NEGAHDARIPOUR, M. et al. Structural vaccinology considerations for in silico designing of
2 a multi-epitope vaccine. **Infection, Genetics and Evolution**, v. 58, p. 96–109, mar. 2018.
- 3 NYATANYI, T. et al. A Mixed Outbreak of Epidemic Typhus Fever and Trench Fever in a
4 Youth Rehabilitation Center: Risk Factors for Illness from a Case-Control Study, Rwanda,
5 2012. **The American Journal of Tropical Medicine and Hygiene**, v. 95, n. 2, p. 452–456, 3
6 ago. 2016.
- 7 OLI, A. N. et al. <p>Immunoinformatics and Vaccine Development: An Overview</p>.
8 **ImmunoTargets and Therapy**, v. Volume 9, p. 13–30, fev. 2020.
- 9 OSTERLOH, A. The neglected challenge: Vaccination against rickettsiae. **PLOS Neglected**
10 **Tropical Diseases**, v. 14, n. 10, p. e0008704, 22 out. 2020.
- 11 PAPPALARDO, F. et al. In silico clinical trials: concepts and early adoptions. **Briefings in**
12 **Bioinformatics**, v. 20, n. 5, p. 1699–1708, 27 set. 2019.
- 13 PAROLA, P. et al. Update on Tick-Borne Rickettsioses around the World: a Geographic
14 Approach. **Clinical Microbiology Reviews**, v. 26, n. 4, p. 657–702, out. 2013.
- 15 PARVIZPOUR, S. et al. Epitope-based vaccine design: a comprehensive overview of
16 bioinformatics approaches. **Drug Discovery Today**, v. 25, n. 6, p. 1034–1042, jun. 2020.
- 17 PENICHE-LARA, G. et al. Simple Method to Differentiate among Rickettsia Species. **Journal**
18 **of Molecular Microbiology and Biotechnology**, v. 23, n. 3, p. 203–208, 2013.
- 19 PETTERSEN, E. F. et al. UCSF Chimera - A visualization system for exploratory research and
20 analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, 2004.
- 21 PHILIP, R. N. et al. Rickettsia bellii sp. nov.: a Tick-Borne Rickettsia, Widely Distributed in
22 the United States, That Is Distinct from the Spotted Fever and Typhus Biogroups. **International**
23 **Journal of Systematic Bacteriology**, v. 33, n. 1, p. 94–106, 1 jan. 1983.
- 24 PIERACCI, E. G. et al. Fatal Flea-Borne Typhus in Texas: A Retrospective Case Series, 1985–
25 2015. **The American Journal of Tropical Medicine and Hygiene**, v. 96, n. 5, p. 1088–1093,
26 3 maio 2017.
- 27 POLLARD, A. J.; BIJKER, E. M. A guide to vaccinology: from basic principles to new
28 developments. **Nature Reviews Immunology**, v. 21, n. 2, p. 83–100, 22 fev. 2021.
- 29 PONOMARENKO, J. et al. ElliPro: a new structure-based tool for the prediction of antibody
30 epitopes. **BMC Bioinformatics**, v. 9, n. 1, p. 514, 2 dez. 2008.
- 31 PURCELL, A. W.; MCCLUSKEY, J.; ROSSJOHN, J. More than one reason to rethink the use
32 of peptides in vaccine design. **Nature Reviews Drug Discovery**, v. 6, n. 5, p. 404–414, maio
33 2007.

- 1 QUEVEDO-DIAZ, M. A. et al. Involvement of TLR2 and TLR4 in cell responses to *Rickettsia*
2 *akari*. **Journal of Leukocyte Biology**, v. 88, n. 4, p. 675–685, out. 2010.
- 3 RANA, A.; AKHTER, Y. A multi-subunit based, thermodynamically stable model vaccine
4 using combined immunoinformatics and protein structure based approach. **Immunobiology**, v.
5 221, n. 4, p. 544–557, abr. 2016.
- 6 RAPIN, N. et al. Computational Immunology Meets Bioinformatics: The Use of Prediction
7 Tools for Molecular Binding in the Simulation of the Immune System. **PLoS ONE**, v. 5, n. 4,
8 p. e9862, 16 abr. 2010.
- 9 ROLAIN, J. M. Genome Comparison Analysis of Molecular Mechanisms of Resistance to
10 Antibiotics in the Rickettsia Genus. **Annals of the New York Academy of Sciences**, v. 1063,
11 n. 1, p. 222–230, 1 dez. 2005.
- 12 RUMFIELD, C. et al. Activation of ASC Inflammasome Driven by Toll-Like Receptor 4
13 Contributes to Host Immunity against Rickettsial Infection. **Infection and Immunity**, v. 88, n.
14 4, 23 mar. 2020a.
- 15 RUMFIELD, C. et al. Activation of ASC Inflammasome Driven by Toll-Like Receptor 4
16 Contributes to Host Immunity against Rickettsial Infection. **Infection and Immunity**, v. 88, n.
17 4, 23 mar. 2020b.
- 18 RYDKINA, E.; TURPIN, L. C.; SAHNI, S. K. *Rickettsia rickettsii* Infection of Human
19 Macrovascular and Microvascular Endothelial Cells Reveals Activation of Both Common and
20 Cell Type-Specific Host Response Mechanisms. **Infection and Immunity**, v. 78, n. 6, p. 2599–
21 2606, jun. 2010.
- 22 SAHA, S.; RAGHAVA, G. P. S. Prediction of continuous B-cell epitopes in an antigen using
23 recurrent neural network. **Proteins: Structure, Function, and Bioinformatics**, v. 65, n. 1, p.
24 40–48, 7 ago. 2006.
- 25 SAIER JR., M. H. Active Transport in Communication, Protection and Nutrition. **Journal of**
26 **Molecular Microbiology and Biotechnology**, v. 12, n. 3–4, p. 161–164, 2007.
- 27 SAKAI, R. K. et al. Experimental infection with *Rickettsia rickettsii* in an *Amblyomma*
28 *dubitatum* tick colony, naturally infected by *Rickettsia bellii*. **Ticks and Tick-borne Diseases**,
29 v. 5, n. 6, p. 917–923, out. 2014.
- 30 SCHROEDER, C. L. C. et al. Bacterial small RNAs in the Genus *Rickettsia*. **BMC Genomics**,
31 v. 16, n. 1, p. 1075, 18 dez. 2015.
- 32 SIX, A. et al. Systems biology in vaccine design. **Microbial Biotechnology**, v. 5, n. 2, p. 295–
33 304, mar. 2012.

- 1 SMIALOWSKI, P. et al. Protein solubility: sequence based prediction and experimental
2 verification. **Bioinformatics**, v. 23, n. 19, p. 2536–2542, 1 out. 2007.
- 3 SOUTHWOOD, S. et al. Several common HLA-DR types share largely overlapping peptide
4 binding repertoires. **Journal of immunology (Baltimore, Md. : 1950)**, v. 160, n. 7, p. 3363–
5 73, 1 abr. 1998.
- 6 STOTHARD, D. R.; CLARK, J. B.; FUERST, P. A. Ancestral Divergence of *Rickettsia bellii*
7 from the Spotted Fever and Typhus Groups of *Rickettsia* and Antiquity of the Genus *Rickettsia*.
8 **International Journal of Systematic Bacteriology**, v. 44, n. 4, p. 798–804, 1 out. 1994.
- 9 STOTHARD, D. R.; FUERST, P. A. Evolutionary Analysis of the Spotted Fever and Typhus
10 Groups of *Rickettsia* Using 16S rRNA Gene Sequences. **Systematic and Applied**
11 **Microbiology**, v. 18, n. 1, p. 52–61, jan. 1995.
- 12 STRAILY, A. et al. *Notes from the Field*: Community-Based Prevention of Rocky Mountain
13 Spotted Fever — Sonora, Mexico, 2016. **MMWR. Morbidity and Mortality Weekly Report**,
14 v. 65, n. 46, p. 1302–1303, 25 nov. 2016.
- 15 TARRAGONA, E. L. et al. Vectorial competence of *Amblyomma tonelliae* to transmit
16 *Rickettsia rickettsii*. **Medical and Veterinary Entomology**, v. 30, n. 4, p. 410–415, dez. 2016.
- 17 TOMAR, N.; DE, R. K. Immunoinformatics: an integrated scenario. **Immunology**, v. 131, n.
18 2, p. 153–168, out. 2010.
- 19 TSHILWANE, S. I. et al. A multi-epitope DNA vaccine co-administered with monophosphoryl
20 lipid A adjuvant provides protection against tick transmitted *Ehrlichia ruminantium* in sheep.
21 **Vaccine**, v. 37, n. 31, p. 4354–4363, jul. 2019.
- 22 TURCO, J.; WINKLER, H. H. Cytokines Influencing Infections by *Rickettsia* Species. Em:
23 **Rickettsial Infection and Immunity**. Boston: Kluwer Academic Publishers, [s.d.]. p. 29–52.
- 24 VAKILI, B. et al. A new multi-epitope peptide vaccine induces immune responses and
25 protection against *Leishmania infantum* in BALB/c mice. **Medical Microbiology and**
26 **Immunology**, v. 209, n. 1, p. 69–79, 6 fev. 2020.
- 27 VANROMPAY, D. et al. Antimicrobial Resistance in *Chlamydiales*, *Rickettsia*, *Coxiella*, and
28 Other Intracellular Pathogens. **Microbiology Spectrum**, v. 6, n. 2, 6 abr. 2018.
- 29 VAUGHN, M. F. et al. Long-lasting Permethrin Impregnated Uniforms. **American Journal of**
30 **Preventive Medicine**, v. 46, n. 5, p. 473–480, maio 2014.
- 31 VILELA RODRIGUES, T. C. et al. An immunoinformatics-based designed multi-epitope
32 candidate vaccine (mpme-VAC/STV-1) against *Mycoplasma pneumoniae*. **Computers in**
33 **Biology and Medicine**, v. 142, p. 105194, mar. 2022.

- 1 VITA, R. et al. The Immune Epitope Database (IEDB): 2018 update. **Nucleic Acids Research**,
2 v. 47, n. D1, p. D339–D343, 8 jan. 2019.
- 3 WALKER, D. H. Changing Dynamics of Human–Rickettsial Interactions. **The American**
4 **Journal of Tropical Medicine and Hygiene**, v. 94, n. 1, p. 3–4, 6 jan. 2016.
- 5 WALKER, D. H.; VALBUENA, G. A.; OLANO, J. P. Pathogenic Mechanisms of Diseases
6 Caused by *Rickettsia*. **Annals of the New York Academy of Sciences**, v. 990, n. 1, p. 1–11,
7 jun. 2003.
- 8 WARD, J. J. et al. Secondary structure prediction with support vector machines.
9 **Bioinformatics**, v. 19, n. 13, p. 1650–1655, 1 set. 2003.
- 10 WORLD HEALTH ORGANIZATION. Epidemic typhus risk in Rwandan refugee camps.
11 **Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire**, p. 259–259, 26
12 ago. 1994.
- 13 WORLD HEALTH ORGANIZATION. **Child mortality and causes of death**.
- 14

ANEXO A – Filtragem e *overlap* das proteínas

Tabela 3 - Etapa de filtragem de epítomos para o complexo MHCII/CTL de cada uma das 8 proteínas utilizadas nos programas IEDB e NetCTL realizada pelo programa Microsoft Excel.

ID Proteína	MHCII/CTL IEDB			MHCII/CTL NetCTL		
	Total inicial	Percentil rank < 1	Remoção de duplicatas	Total inicial	<- E	Remoção de duplicatas
WP_012148219.1	11664	402	140	432	117	94
WP_012148227.1	6345	220	78	235	67	52
WP_012148552.1	4563	127	44	169	50	35
WP_012148811.1	5400	174	55	200	61	44
WP_012148819.1	8046	261	105	298	87	63
WP_012149174.1	6453	221	72	239	66	50
WP_012149175.1	5832	186	64	216	61	46
WP_012149185.1	7425	210	76	275	53	46
Total	55728	1801	634	2064	562	430

Fonte: Do Autor, 2022.

Tabela 4 - Etapa de filtragem de epítomos para o complexo MHCII/HTL de cada uma das 8 proteínas utilizadas nos programas IEDB e NetCTL realizada pelo programa Microsoft Excel.

ID Proteína	MHCII/HTL IEDB				MHCII/HTL NetMHCII			
	Total Inicial	Percentil rank < 10	Toxinpred	Remoção de Duplicatas	IC50 < 500	Total Inicial	SB	Remoção de Duplicatas
WP_012148219.1	11319	668	662	265	206	674	110	77
WP_012148227.1	5988	521	510	157	126	491	103	64
WP_012148552.1	4401	404	394	100	82	264	92	50
WP_012148811.1	5052	304	256	100	72	306	82	33
WP_012148819.1	7695	744	716	202	131	561	151	89
WP_012149174.1	6101	408	402	148	106	399	121	59
WP_012149175.1	5481	316	316	116	87	328	57	44
WP_012149185.1	7079	333	317	146	105	335	49	36
Total	53116	3698	3573	1234	915	3358	765	452

Fonte: Do Autor, 2022.

Tabela 5 - Epítomos finais preditos pelo programa ABCpred para cada uma das 8 proteínas.

Células B	
ID Proteína	Total Inicial
WP_012148219.1	44
WP_012148227.1	25
WP_012148552.1	12
WP_012148811.1	19
WP_012148819.1	30
WP_012149174.1	25
WP_012149175.1	19
WP_012149185.1	20
Total	194

Fonte: Do Autor, 2022.

Tabela 6 - Técnica de overlap entre os complexos MHCI/CTL e MHCII/HTL com células B para cada uma das proteínas utilizadas.

MHC I					
	IEDB vs NetCTL	MHC I vs Cel B	<i>Class I Immunogenicity para MHC I > 0.1</i>	Remoção de até 2 AA	Vaxijen (cut off 0.7)
WP_012148219.1	86	70	17	15	4
WP_012148227.1	48	36	11	7	2
WP_012148552.1	30	16	1	1	1
WP_012148811.1	35	21	3	3	1
WP_012148819.1	57	50	16	10	2
WP_012149174.1	45	39	7	6	3
WP_012149175.1	37	23	6	5	0
WP_012149185.1	42	21	5	5	0
Total	380	276	66	52	13

MHC II				
	IEDB vs NetMHCII	MHC II vs Cel B	Remoção de até 2 AA	Vaxijen (cut off 0.7)
WP_012148219.1	58	15	10	3

WP_012148227.1	51	9	5	1
WP_012148552.1	42	4	3	2
WP_012148811.1	17	1	1	0
WP_012148819.1	54	6	5	1
WP_012149174.1	50	15	10	4
WP_012149175.1	33	7	5	2
WP_012149185.1	23	2	2	0
Total	328	59	41	13

Fonte: Do Autor, 2022.

ANEXO B – Epítomos conformacionais de células B

Tabela 7 - Predição de epítomos lineares de células B.

Epítomos	Peptídeos	Número de Resíduos	Scores
1	LGTSFEVAQGVGPGPGMALKRDYPKYYPLFSGP GPGNSDTSIKLEG	46	0.764
2	VKAAYKTTRGTVGLAAYS	18	0.755
3	CASEAYTPNHVWA	13	0.752
4	GTSAQGPGGIHTKMDRTKGVEVLNGPGPGG	31	0.747
5	SLGTSGPGPGATSVKLLSNMTVSVDGPGPGPKLE YKPNLVGNKTVGPGPGNANGITLTVNNPSLTGPG PGM	71	0.745
6	HLKAAAYTGQFHIAPY	15	0.743
7	DTNKDSKGIY	10	0.716

Fonte: Do Autor, 2022.

Tabela 8 - Predição de epítomos descontinuados de células B.

Epítomos	Resíduos	Número de Resíduos	Scores
1	A:H349, A:T350, A:K351	3	0.929
2	A:R449, A:G450, A:T451, A:V452, A:G453, A:L454, A:A455, A:A456, A:Y457	9	0.849
3	A:L525, A:K526, A:A527, A:A528, A:Y529, A:T530, A:G531, A:Q532, A:F533, A:H534	10	0.828
4	A:L272, A:G273, A:T274, A:S275, A:F276, A:E277, A:V278, A:A279, A:Q280, A:G281, A:V282, A:G283, A:P284, A:G285, A:P286, A:G287, A:M288, A:A289, A:L290	19	0.824
5	A:S138, A:L139, A:G140, A:T141, A:S142, A:G143, A:P144, A:G145, A:P146, A:G147, A:A148, A:T149, A:S150, A:V151, A:K152	15	0.821

6	A:Y297, A:P299, A:L300, A:F301, A:S302, A:G303, A:P304, A:G305, A:P306, A:G307, A:N308, A:S309, A:D310, A:T311, A:S312, A:I313, A:K314	17	0.808
7	A:L153, A:K154, A:S155, A:N156, A:M157, A:T158, A:V159, A:S160, A:V161, A:D162, A:G163, A:P164, A:G165, A:P166, A:G167, A:P168, A:K169, A:L170, A:E171, A:Y172, A:K173, A:P174, A:N175, A:L176, A:V177, A:G178, A:N179, A:K180	28	0.762
8	A:E507, A:C508, A:G509, A:E510, A:V511, A:N512, A:P513, A:I514	8	0.737
9	A:G338, A:T339, A:S340, A:A341, A:Q342, A:G343, A:P344, A:G345, A:P346, A:G347, A:I348, A:V358, A:E359, A:V360, A:L361, A:N362, A:G363, A:P364, A:G365, A:P366, A:G367, A:G368, A:K393	23	0.725
10	A:M352, A:D353, A:R354, A:T355, A:K356, A:G357	6	0.7

Fonte: Do Autor, 2022.

ANEXO C – Predição do IFNepitope

Tabela 9 - Predição de estimulação de Interferon dos epítomos da proteína multi-epítomo.

Número	Sequência/Epítomo	Método	Resultado	Score
1	GPGPGEYKIMPGLLP	SVM	POSITIVO	0.00081751901
2	AAAKMAENPNIDDL	SVM	NEGATIVO	-0.0083222059
3	AAKMAENPNIDDLPA	SVM	NEGATIVO	-0.33371237
4	AKMAENPNIDDLPAP	SVM	NEGATIVO	-0.45271403
5	KMAENPNIDDLPAPL	SVM	NEGATIVO	-0.4840845
6	MAENPNIDDLPAPLL	SVM	NEGATIVO	-0.4134494
7	AENPNIDDLPAPLLA	SVM	NEGATIVO	-0.71224553
8	ENPNIDDLPAPLLAA	SVM	NEGATIVO	-0.45431911
9	NPNIIDDLPAPLLAAL	SVM	NEGATIVO	-0.74891022
10	PNIDDLPAPLLAALG	SVM	NEGATIVO	-0.52187068
11	NIDDLPAPLLAALGA	SVM	NEGATIVO	-0.57934186
12	IDDLPAPLLAALGAA	SVM	NEGATIVO	-0.6699177
13	DDLAPLLAALGAAD	SVM	NEGATIVO	-0.72351262
14	DLAPLLAALGAADL	SVM	NEGATIVO	-0.4164655
15	LPAPLLAALGAADLA	SVM	NEGATIVO	-0.29926524
16	PAPLLAALGAADLAL	SVM	NEGATIVO	-0.19584417
17	ELTQEALGTVASQTR	SVM	POSITIVO	0.0053471244
18	GYLEAATNRYNELVE	SVM	POSITIVO	0.0099973701
19	EGYVDQAVELTQEAL	SVM	POSITIVO	0.013816801
20	YNELVERGEAALQRL	SVM	POSITIVO	0.015511647
21	AYNECGEVNPIAAYG	SVM	POSITIVO	0.018395856
22	IELRDKFTTEELRKA	SVM	POSITIVO	0.019814547
23	VDQAVELTQEALGTV	SVM	POSITIVO	0.021024276
24	FIELRDKFTTEELRK	SVM	POSITIVO	0.022702707
25	TRGTVGLAAYSSFKI	SVM	POSITIVO	0.025585894
26	NECGEVNPIAAYGTV	SVM	POSITIVO	0.026568109
27	FIAAAYGQIGVDFYA	SVM	POSITIVO	0.030906254
28	LALATVNDLIANLRE	SVM	NEGATIVO	-0.033176546
29	ALATVNDLIANLRER	SVM	NEGATIVO	-0.22745613
30	LATVNDLIANLRERA	SVM	NEGATIVO	-0.047548487
31	FLLPYIFIAAAYGQI	SVM	POSITIVO	0.032212977
32	TVNDLIANLRERAEE	SVM	NEGATIVO	-0.16534372
33	VNDLIANLRERAET	SVM	NEGATIVO	-0.65707349
34	NDLIANLRERAETR	SVM	NEGATIVO	-0.59219489
35	VAAYRVNFFTPKMAA	SVM	POSITIVO	0.035810386
36	AYGTVINVHLKAAYT	SVM	POSITIVO	0.041909028
37	AAYKTTRGTVGLAAY	SVM	POSITIVO	0.043361796
38	SQTAFEDASARAEGY	SVM	POSITIVO	0.044101115
39	NRYNELVERGEAALQ	SVM	POSITIVO	0.046841932
40	GEAALQRLRSQTAFE	SVM	POSITIVO	0.047961904
41	EDASARAEGYVDQAV	SVM	POSITIVO	0.051516293
42	VGKNTVGPMPGNANG	SVM	POSITIVO	0.055051808

43	LTLGTSAQGGPGPIH	SVM	POSITIVO	0.055145257
44	SIKLEGFYLLEGGP	SVM	POSITIVO	0.060315712
45	RGEAALQRLRSQTAF	SVM	POSITIVO	0.060445107
46	QVTAGTGGWAAYTAN	SVM	POSITIVO	0.062536902
47	TLTVNNPSLTGPGPG	SVM	POSITIVO	0.068458746
48	AAYTANTGGNRYAAY	SVM	POSITIVO	0.071460395
49	AETRTRVEERRARLT	SVM	NEGATIVO	-0.070375625
50	ETRTRVEERRARLTK	SVM	NEGATIVO	-0.26865463
51	TRTRVEERRARLTKF	SVM	NEGATIVO	-0.1707879
52	RTRVEERRARLTKFQ	SVM	NEGATIVO	-0.33710604
53	TRVEERRARLTKFQE	SVM	NEGATIVO	-0.5377821
54	RVEERRARLTKFQED	SVM	NEGATIVO	-0.5641927
55	VEERRARLTKFQEDL	SVM	NEGATIVO	-0.6915817
56	EERRARLTKFQEDLP	SVM	NEGATIVO	-0.86347699
57	ERRARLTKFQEDLPE	SVM	NEGATIVO	-0.81750768
58	RRARLTKFQEDLPEQ	SVM	NEGATIVO	-0.70032566
59	RARLTKFQEDLPEQF	SVM	NEGATIVO	-0.75787356
60	ARLTKFQEDLPEQFI	SVM	NEGATIVO	-0.76045003
61	RLTKFQEDLPEQFIE	SVM	NEGATIVO	-0.79239767
62	LTKFQEDLPEQFIEL	SVM	NEGATIVO	-0.62274853
63	TKFQEDLPEQFIELR	SVM	NEGATIVO	-0.50992006
64	KFQEDLPEQFIELRD	SVM	NEGATIVO	-0.43037405
65	FQEDLPEQFIELRDK	SVM	NEGATIVO	-0.46237585
66	QEDLPEQFIELRDKF	SVM	NEGATIVO	-0.54791675
67	EDLPEQFIELRDKFT	SVM	NEGATIVO	-0.63503871
68	DLPEQFIELRDKFTT	SVM	NEGATIVO	-0.39834432
69	LPEQFIELRDKFTTE	SVM	NEGATIVO	-0.39696285
70	PEQFIELRDKFTTEE	SVM	NEGATIVO	-0.57994486
71	EQFIELRDKFTTEEL	SVM	NEGATIVO	-0.17647669
72	QFIELRDKFTTEELR	SVM	NEGATIVO	-0.030704287
73	AAVKSTRGNIVKAA	SVM	POSITIVO	0.071611421
74	CGEVNPIAAYGTVIN	SVM	POSITIVO	0.072554322
75	TRGNIVKAAKTTTRG	SVM	POSITIVO	0.075834649
76	LRDKFTTEELRKAEE	SVM	NEGATIVO	-0.097349091
77	RDKFTTEELRKAEEG	SVM	NEGATIVO	-0.082332581
78	DKFTTEELRKAEEGY	SVM	NEGATIVO	-0.040332333
79	KFTTEELRKAEEGYL	SVM	NEGATIVO	-0.094277929
80	FTTEELRKAEEGYLE	SVM	NEGATIVO	-0.017476923
81	PSLTGPGPGMENQWY	SVM	POSITIVO	0.075968726
82	ATVNDLIANLRERAE	SVM	POSITIVO	0.076253348
83	AADLALATVNDLIAN	SVM	POSITIVO	0.076778464
84	VNPIAAYGTVINVHL	SVM	POSITIVO	0.078877951
85	LVGNKTVGPGPGNAN	SVM	POSITIVO	0.08247417
86	PGNSDTSIKLEGFYL	SVM	POSITIVO	0.082718777
87	SAQGGPGPIHTKMD	SVM	POSITIVO	0.085461753
88	ARAEQYVDQAVELTQ	SVM	POSITIVO	0.085530564
89	AAVRVNFFTPKMAAY	SVM	POSITIVO	0.089416816
90	IAAYGTVINVHLKAA	SVM	POSITIVO	0.094507614

91	PNLVGNKTVGPGPGN	SVM	POSITIVO	0.095238414
92	YLEAATNRYNELVER	SVM	NEGATIVO	-0.055192107
93	PIAA YGTVINVHLKA	SVM	POSITIVO	0.095878676
94	ELRDKFTTEELRCAA	SVM	POSITIVO	0.096391677
95	AYGQIGVDFYAAAYN	SVM	POSITIVO	0.096642774
96	TNISYKLTGTSAAQG	SVM	POSITIVO	0.097038842
97	AA YFLLPYIFIAAAY	SVM	POSITIVO	0.097397326
98	RYNELVERGEAALQR	SVM	POSITIVO	0.099048937
99	VNNPSLTGPGPGMEN	SVM	POSITIVO	0.10153799
100	TNRYNELVERGEAAL	SVM	POSITIVO	0.10441409
101	NELVERGEAALQRLR	SVM	NEGATIVO	-0.064146658
102	ELVERGEAALQRLRS	SVM	NEGATIVO	-0.31591531
103	LVERGEAALQRLRSQ	SVM	NEGATIVO	-0.20848673
104	VERGEAALQRLRSQT	SVM	NEGATIVO	-0.10304081
105	ERGEAALQRLRSQTA	SVM	NEGATIVO	-0.07642814
106	YAA YKSTRGNIVKAA	SVM	POSITIVO	0.11124026
107	YLLEGPGPGRTNISY	SVM	POSITIVO	0.11252606
108	EAA LQRLRSQTAFED	SVM	NEGATIVO	-0.079015532
109	AALQRLRSQTAFEDA	SVM	NEGATIVO	-0.15268613
110	ALQRLRSQTAFEDAS	SVM	NEGATIVO	-0.34101947
111	LQRLRSQTAFEDASA	SVM	NEGATIVO	-0.36158074
112	QRLRSQTAFEDASAR	SVM	NEGATIVO	-0.21286094
113	RLRSQTAFEDASARA	SVM	NEGATIVO	-0.20318048
114	LR SQTAFEDASARAE	SVM	NEGATIVO	-0.091971595
115	RSQTAFEDASARAEG	SVM	NEGATIVO	-0.048411334
116	EGFYLLEGPGPGRTN	SVM	POSITIVO	0.11729918
117	QTAFEDASARAEGYV	SVM	NEGATIVO	-0.045260359
118	TAFEDASARAEGYVD	SVM	NEGATIVO	-0.054936166
119	AFEDASARAEGYVDQ	SVM	NEGATIVO	-0.154556
120	FEDASARAEGYVDQA	SVM	NEGATIVO	-0.061549755
121	DLIANLRERAETRA	SVM	POSITIVO	0.12148915
122	IKLEGFYLLEGPGPG	SVM	POSITIVO	0.12402411
123	YAAAYNECGEVNPIA	SVM	POSITIVO	0.12723401
124	TTRGTVGLAA YSSFK	SVM	POSITIVO	0.12761045
125	AYKSTRGNIVKAA YK	SVM	POSITIVO	0.13281497
126	TMIGPGPGGEYKIMPG	SVM	POSITIVO	0.13458411
127	AEGYVDQAVELTQEA	SVM	NEGATIVO	-0.039107974
128	AGTGGWAA YTANTGG	SVM	POSITIVO	0.13658353
129	GYVDQAVELTQEALG	SVM	NEGATIVO	-0.11955742
130	YVDQAVELTQEALGT	SVM	NEGATIVO	-0.11440732
131	RGTVGLAA YSSFKIT	SVM	POSITIVO	0.14007839
132	DQAVELTQEALGTVA	SVM	NEGATIVO	-0.084015188
133	QAVELTQEALGTVAS	SVM	NEGATIVO	-0.15625201
134	AVELTQEALGTVASQ	SVM	NEGATIVO	-0.092045326
135	VELTQEALGTVASQT	SVM	NEGATIVO	-0.059555371
136	ADLALATVNDLIANL	SVM	POSITIVO	0.14128778
137	TGQFHIAPY	SVM	POSITIVO	0.14189666
138	MIGPGPGGEYKIMPGL	SVM	POSITIVO	0.14258261

139	LEAATNRYNELVERG	SVM	POSITIVO	0.14280402
140	RYAAYKSTRGNIVKA	SVM	POSITIVO	0.14750068
141	KLEGFYLLEGPGPGR	SVM	POSITIVO	0.14849352
142	GPGPGNSDTSIKLEG	SVM	POSITIVO	0.14960645
143	AAAYNECGEVNPIAA	SVM	POSITIVO	0.14968424
144	GRTNISYKLTLG TSA	SVM	POSITIVO	0.15096884
145	PGPGIHTKMD	SVM	POSITIVO	0.15930109
146	YQV TAGTGGWAAYTA	SVM	POSITIVO	0.16270366
147	AA YQV TAGTGGWAAY	SVM	POSITIVO	0.16511064
148	EAATNRYNELVERGE	SVM	POSITIVO	0.17927253
149	AATNRYNELVERGEA	SVM	POSITIVO	0.17927253
150	ATNRYNELVERGEAA	SVM	POSITIVO	0.17927253
151	TLG TSAQGP GPGIHT	SVM	POSITIVO	0.18080032
152	NTGGNRYAAYKSTRG	SVM	POSITIVO	0.18580132
153	GERAAKLVGIELKNK	SVM	NEGATIVO	-0.12147046
154	ERAAKLVGIELKNKI	SVM	NEGATIVO	-0.41396249
155	RAAKLVGIELKNKIN	SVM	NEGATIVO	-0.19259786
156	AAKLVGIELKNKINF	SVM	NEGATIVO	-0.1915538
157	AKLVGIELKNKINF A	SVM	NEGATIVO	-0.31868965
158	KLVGIELKNKINFAY	SVM	NEGATIVO	-0.34470609
159	LVGIELKNKINFAYQ	SVM	NEGATIVO	-0.12721729
160	VGIELKNKINFAYQL	SVM	NEGATIVO	-0.11427754
161	GIELKNKINFAYQLS	SVM	NEGATIVO	-0.20831596
162	IELKNKINFAYQLSL	SVM	NEGATIVO	-0.3869098
163	ELKNKINFAYQLSLG	SVM	NEGATIVO	-0.31783601
164	LKNKINFAYQLSLGT	SVM	NEGATIVO	-0.52445227
165	KNKINFAYQLSLGTS	SVM	NEGATIVO	-0.55575013
166	NKINFAYQLSLGTSG	SVM	NEGATIVO	-0.63870886
167	KINFAYQLSLGTSGP	SVM	NEGATIVO	-0.95307898
168	INFAYQLSLGTSGPG	SVM	NEGATIVO	-10.184.834
169	NFAYQLSLGTSGPGP	SVM	NEGATIVO	-0.97514565
170	FAYQLSLGTSGPGPG	SVM	NEGATIVO	-0.7746566
171	AYQLSLGTSGPGPGA	SVM	NEGATIVO	-0.76924683
172	YQLSLGTSGPGPGAT	SVM	NEGATIVO	-0.90028228
173	QLSLGTSGPGPGATS	SVM	NEGATIVO	-0.89839153
174	LSLGTSGPGPGATSV	SVM	NEGATIVO	-0.85513829
175	SLGTSGPGPGATSVK	SVM	NEGATIVO	-0.6978686
176	LGTSGPGPGATSVKL	SVM	NEGATIVO	-0.76675732
177	GTSGPGPGATSVK LK	SVM	NEGATIVO	-0.82693225
178	TSGPGPGATSVKLKS	SVM	NEGATIVO	-0.81012001
179	SGPGPGATSVKLKSN	SVM	NEGATIVO	-0.92714853
180	GPGPGATSVKLKSNM	SVM	NEGATIVO	-0.98571887
181	PGPGATSVKLKSNMT	SVM	NEGATIVO	-0.8708089
182	GPGATSVKLKSNMTV	SVM	NEGATIVO	-12.871.509
183	PGATSVKLKSNMTVS	SVM	NEGATIVO	-0.87389666
184	GATSVKLKSNMTVSV	SVM	NEGATIVO	-0.76196648
185	ATSVKLKSNMTVSV D	SVM	NEGATIVO	-0.61868005
186	TSVKLKSNMTVSV D G	SVM	NEGATIVO	-0.79884227

187	SVKLKSNMTVSVDGP	SVM	NEGATIVO	-10.118.588
188	VKLKSNMTVSVDGPG	SVM	NEGATIVO	-11.670.819
189	KLKSNMTVSVDGPGP	SVM	NEGATIVO	-10.669.956
190	LKSNMTVSVDGPGPG	SVM	NEGATIVO	-0.85366904
191	KSNMTVSVDGPGPGP	SVM	NEGATIVO	-0.71678664
192	SNMTVSVDGPGPGPK	SVM	NEGATIVO	-0.79266066
193	NMTVSVDGPGPGPKL	SVM	NEGATIVO	-0.76674984
194	MTVSVDGPGPGPKLE	SVM	NEGATIVO	-0.53222368
195	TVSVDGPGPGPKLEY	SVM	NEGATIVO	-0.43960322
196	VSVDGPGPGPKLEYK	SVM	NEGATIVO	-0.51418992
197	SVDGPGPGPKLEYKP	SVM	NEGATIVO	-0.52287797
198	VDGPGPGPKLEYKPN	SVM	NEGATIVO	-0.5172335
199	DGPGPGPKLEYKPNL	SVM	NEGATIVO	-0.53581897
200	GPGPGPKLEYKPNLV	SVM	NEGATIVO	-0.60970824
201	PGPGPKLEYKPNLVG	SVM	NEGATIVO	-0.60764339
202	GPGPKLEYKPNLVGN	SVM	NEGATIVO	-0.75549418
203	PGPKLEYKPNLVGNK	SVM	NEGATIVO	-0.72348396
204	GPKLEYKPNLVGNKT	SVM	NEGATIVO	-10.337.245
205	PKLEYKPNLVGNKTV	SVM	NEGATIVO	-0.90038399
206	KLEYKPNLVGNKTVG	SVM	NEGATIVO	-0.41959291
207	LEYKPNLVGNKTVGP	SVM	NEGATIVO	-0.62741765
208	EYKPNLVGNKTVGPG	SVM	NEGATIVO	-0.50356374
209	YKPNLVGNKTVGPGP	SVM	NEGATIVO	-0.77810594
210	KPNLVGNKTVGPGPG	SVM	NEGATIVO	-0.32285321
211	EGYLEAATNRYNELV	SVM	POSITIVO	0.18595309
212	NLVGNKTVGPGPGNA	SVM	NEGATIVO	-0.00077354888
213	AALGAADLALATVND	SVM	POSITIVO	0.19038392
214	WAAYTANTGGNRYAA	SVM	POSITIVO	0.19286017
215	GNKTVGPGPGNANGI	SVM	NEGATIVO	-0.11467578
216	NKTVGPGPGNANGIT	SVM	NEGATIVO	-0.46074431
217	KTVGPGPGNANGITL	SVM	NEGATIVO	-0.45846355
218	TVGPGPGNANGITLT	SVM	NEGATIVO	-0.21058253
219	VGPGPGNANGITLTV	SVM	NEGATIVO	-0.21058253
220	GPGPGNANGITLTVN	SVM	NEGATIVO	-0.3519491
221	PGPGNANGITLTVNN	SVM	NEGATIVO	-0.25116477
222	GPGNANGITLTVNNP	SVM	NEGATIVO	-0.44479488
223	PGNANGITLTVNNPS	SVM	NEGATIVO	-0.28321988
224	GNANGITLTVNNPSL	SVM	NEGATIVO	-0.5739404
225	NANGITLTVNNPSLT	SVM	NEGATIVO	-0.48927408
226	ANGITLTVNNPSLTG	SVM	NEGATIVO	-0.57787533
227	NGITLTVNNPSLTGP	SVM	NEGATIVO	-0.61248747
228	GITLTVNNPSLTGPG	SVM	NEGATIVO	-0.25126042
229	ITLTVNNPSLTGPGP	SVM	NEGATIVO	-0.37543856
230	SARAEGYVDQAVELT	SVM	POSITIVO	0.19607458
231	EVNPIAAYGTVINVH	SVM	POSITIVO	0.20688699
232	TVNNPSLTGPGPGME	SVM	NEGATIVO	-0.014119672
233	AAAYGTVINVHLKAAY	SVM	POSITIVO	0.2098964
234	NNPSLTGPGPGMENQ	SVM	NEGATIVO	-0.0047159632

235	NPSLTGPGPGMENQW	SVM	NEGATIVO	-0.010134443
236	IGPGPGGEYKIMPGLL	SVM	POSITIVO	0.21231044
237	SLTGPMPGMENQWYL	SVM	NEGATIVO	-0.12938235
238	LTGPMPGMENQWYLK	SVM	NEGATIVO	-0.18046298
239	TGPMPGMENQWYLKL	SVM	NEGATIVO	-0.38898278
240	GPGMPGMENQWYLKLN	SVM	NEGATIVO	-0.18668857
241	PGMPGMENQWYLKLN	SVM	NEGATIVO	-0.14642392
242	GPGMPGMENQWYLKLNAG	SVM	NEGATIVO	-0.34121278
243	PGMPGMENQWYLKLNAGT	SVM	NEGATIVO	-0.17444882
244	GMENQWYLKLNAGTM	SVM	NEGATIVO	-0.064187547
245	MENQWYLKLNAGTMI	SVM	NEGATIVO	-0.04436069
246	ENQWYLKLNAGTMIG	SVM	NEGATIVO	-0.21880781
247	NQWYLKLNAGTMIGP	SVM	NEGATIVO	-0.63838174
248	QWYLKLNAGTMIGPG	SVM	NEGATIVO	-0.47000225
249	WYLKLNAGTMIGPGP	SVM	NEGATIVO	-0.62041533
250	YLKLNAGTMIGPGPG	SVM	NEGATIVO	-0.40561541
251	LKLNAGTMIGPGPGE	SVM	NEGATIVO	-0.39641993
252	KLNAGTMIGPGPGEY	SVM	NEGATIVO	-0.28128392
253	LNAGTMIGPGPGEYK	SVM	NEGATIVO	-0.35728398
254	NAGTMIGPGPGEYKI	SVM	NEGATIVO	-0.4823715
255	AGTMIGPGPGEYKIM	SVM	NEGATIVO	-0.26023432
256	GTMIGPGPGEYKIMP	SVM	NEGATIVO	-0.46243093
257	LTVNPNPSLTGPGPGM	SVM	POSITIVO	0.21383011
258	SDTSIKLEGFYLLEG	SVM	POSITIVO	0.21573808
259	TGGWAAYTANTGGNR	SVM	POSITIVO	0.21950658
260	ECGEVNPIAAYGTVI	SVM	POSITIVO	0.22297771
261	PGPGEYKIMPGLLPY	SVM	NEGATIVO	-0.081710037
262	GPGGEYKIMPGLLPYA	SVM	NEGATIVO	-0.45888939
263	PGEYKIMPGLLPYAE	SVM	NEGATIVO	-0.47504599
264	GEYKIMPGLLPYAEI	SVM	NEGATIVO	-0.81632467
265	EYKIMPGLLPYAEIS	SVM	NEGATIVO	-0.6842841
266	YKIMPGLLPYAEISG	SVM	NEGATIVO	-0.78555028
267	KIMPGLLPYAEISGP	SVM	NEGATIVO	-10.631.175
268	IMPGLLPYAEISGPG	SVM	NEGATIVO	-0.68605427
269	MPGLLPYAEISGPGP	SVM	NEGATIVO	-0.76998373
270	PGLLPYAEISGPGPG	SVM	NEGATIVO	-0.36433537
271	GLLPYAEISGPGPGS	SVM	NEGATIVO	-0.84522185
272	LLPYAEISGPGPGSK	SVM	NEGATIVO	-10.900.427
273	LPYAEISGPGPGSKM	SVM	NEGATIVO	-11.323.852
274	PYAEISGPGPGSKMV	SVM	NEGATIVO	-12.700.792
275	YAEISGPGPGSKMVK	SVM	NEGATIVO	-10.571.596
276	AEISGPGPGSKMVKV	SVM	NEGATIVO	-10.829.879
277	EISGPGPGSKMVKVD	SVM	NEGATIVO	-10.272.501
278	ISGPGPGSKMVKVDY	SVM	NEGATIVO	-10.183.881
279	SGPGPGSKMVKVDYP	SVM	NEGATIVO	-1.366.205
280	GPGPGSKMVKVDYPF	SVM	NEGATIVO	-10.820.981
281	PGPGSKMVKVDYPFL	SVM	NEGATIVO	-0.8275652
282	GPGSKMVKVDYPFLI	SVM	NEGATIVO	-0.92315445

283	PGSKMVKVDYPFLIA	SVM	NEGATIVO	-0.44153551
284	GSKMVKVDYPFLIAD	SVM	NEGATIVO	-0.47197913
285	SKMVKVDYPFLIADN	SVM	NEGATIVO	-0.37520278
286	KMVKVDYPFLIADNG	SVM	NEGATIVO	-0.36995935
287	MVKVDYPFLIADNGP	SVM	NEGATIVO	-0.58326539
288	VKVDYPFLIADNGPG	SVM	NEGATIVO	-0.37476913
289	KVDYPFLIADNGPGP	SVM	NEGATIVO	-0.53671927
290	VDYPFLIADNGPGPG	SVM	NEGATIVO	-0.20747583
291	DYPFLIADNGPGPGY	SVM	NEGATIVO	-0.11816894
292	YPFLIADNGPGPGYQ	SVM	NEGATIVO	-0.087526193
293	PFLIADNGPGPGYQL	SVM	NEGATIVO	-0.053297345
294	FLIADNGPGPGYQLS	SVM	NEGATIVO	-0.39887687
295	LIADNGPGPGYQLSL	SVM	NEGATIVO	-0.52702483
296	IADNGPGPGYQLSLG	SVM	NEGATIVO	-0.72373305
297	ADNGPGPGYQLSLGT	SVM	NEGATIVO	-0.87996177
298	DNGPGPGYQLSLGTS	SVM	NEGATIVO	-0.91378502
299	NGPGPGYQLSLGTSF	SVM	NEGATIVO	-0.99592658
300	GPGPGYQLSLGTSFE	SVM	NEGATIVO	-0.83712817
301	PGPGYQLSLGTSFEV	SVM	NEGATIVO	-0.72884912
302	GPGYQLSLGTSFEVA	SVM	NEGATIVO	-0.86504631
303	PGYQLSLGTSFEVAQ	SVM	NEGATIVO	-0.43800422
304	GYQLSLGTSFEVAQG	SVM	NEGATIVO	-0.38446651
305	YQLSLGTSFEVAQGV	SVM	NEGATIVO	-0.55028793
306	QLSLGTSFEVAQGVG	SVM	NEGATIVO	-0.6796587
307	LSLGTSFEVAQGVGP	SVM	NEGATIVO	-0.75344882
308	SLGTSFEVAQGVGPG	SVM	NEGATIVO	-0.6253869
309	LGTSFEVAQGVGPGP	SVM	NEGATIVO	-0.70497141
310	GTSFEVAQGVGPGPG	SVM	NEGATIVO	-0.36510859
311	TSFEVAQGVGPGPGM	SVM	NEGATIVO	-0.33047011
312	SFEVAQGVGPGPGMA	SVM	NEGATIVO	-0.098965602
313	FEVAQGVGPGPGMAL	SVM	NEGATIVO	-0.11968773
314	EVAQGVGPGPGMALK	SVM	NEGATIVO	-0.14676456
315	VAQGVGPGPGMALKR	SVM	NEGATIVO	-0.10916247
316	AQGVGPGPGMALKRD	SVM	NEGATIVO	-0.081824753
317	QGVGPGPGMALKRDY	SVM	NEGATIVO	-0.0014119525
318	GVGPGPGMALKRDYP	SVM	NEGATIVO	-0.29558011
319	VGPGPGMALKRDYPK	SVM	NEGATIVO	-0.81385926
320	GPGPGMALKRDYPKY	SVM	NEGATIVO	-0.66087395
321	PGPGMALKRDYPKYY	SVM	NEGATIVO	-0.5418075
322	GPGMALKRDYPKYYP	SVM	NEGATIVO	-0.71435069
323	PGMALKRDYPKYYPL	SVM	NEGATIVO	-0.40607288
324	GMALKRDYPKYYPLF	SVM	NEGATIVO	-0.444461694
325	MALKRDYPKYYPLFS	SVM	NEGATIVO	-0.52293236
326	ALKRDYPKYYPLFSG	SVM	NEGATIVO	-0.60232406
327	LKRDYPKYYPLFSGP	SVM	NEGATIVO	-0.78760055
328	KRDYPKYYPLFSGPG	SVM	NEGATIVO	-0.7023448
329	RDYPKYYPLFSGPGP	SVM	NEGATIVO	-0.78468921
330	DYPKYYPLFSGPGPG	SVM	NEGATIVO	-0.42400323

331	YPKYYPLFSGPGPGN	SVM	NEGATIVO	-0.28829713
332	PKYYPLFSGPGPGNS	SVM	NEGATIVO	-0.37995864
333	KYYPLFSGPGPGNSD	SVM	NEGATIVO	-0.2351915
334	YYPLFSGPGPGNSDT	SVM	NEGATIVO	-0.26901535
335	YPLFSGPGPGNSDTS	SVM	NEGATIVO	-0.60199335
336	PLFSGPGPGNSDTSI	SVM	NEGATIVO	-0.37477089
337	LFSGPGPGNSDTSIK	SVM	NEGATIVO	-0.33132092
338	FSGPGPGNSDTSIKL	SVM	NEGATIVO	-0.21874709
339	SGPGPGNSDTSIKLE	SVM	NEGATIVO	-0.17702502
340	EAAAKMAENPNIDDL	SVM	POSITIVO	0.22408368
341	ALGAADLALATVNDL	SVM	POSITIVO	0.2269467
342	GPGNSDTSIKLEGFY	SVM	NEGATIVO	-0.1103031
343	TTEELRKAEGYLEA	SVM	POSITIVO	0.23080213
344	GNSDTSIKLEGFYLL	SVM	NEGATIVO	-0.065420044
345	NSDTSIKLEGFYLL	SVM	NEGATIVO	-0.023329883
346	DLALATVNDLIANLR	SVM	POSITIVO	0.23304405
347	DTSIKLEGFYLLLEGP	SVM	NEGATIVO	-0.047076169
348	TSIKLEGFYLLLEGG	SVM	NEGATIVO	-0.028646032
349	GTGGWAAYTANTGGN	SVM	POSITIVO	0.23699595
350	VGERAAKLVGIELKN	SVM	POSITIVO	0.24884831
351	APLLAALGAADLALA	SVM	POSITIVO	0.25404264
352	GEVNPAAAYGTVINV	SVM	POSITIVO	0.25441553
353	RAEGYVDQAVELTQE	SVM	POSITIVO	0.25934691
354	GFYLLEGGPGGRTNI	SVM	NEGATIVO	-0.27713361
355	FYLLEGGPGGRTNIS	SVM	NEGATIVO	-0.11217144
356	AEGYLEAATNRYNEL	SVM	POSITIVO	0.26370284
357	LLEGGPGGRTNISYK	SVM	NEGATIVO	-0.046314171
358	LEGGPGGRTNISYKL	SVM	NEGATIVO	-0.1471395
359	EGPGGRTNISYKLT	SVM	NEGATIVO	-0.1755268
360	GPGGRTNISYKLT	SVM	NEGATIVO	-0.20472778
361	PGGRTNISYKLTG	SVM	NEGATIVO	-0.029313913
362	GPGRTNISYKLTGTL	SVM	NEGATIVO	-0.30618056
363	PGRRTNISYKLTGTS	SVM	NEGATIVO	-0.13961339
364	RTNISYKLTGTSAQ	SVM	POSITIVO	0.27056192
365	DASARAEGYVDQAVE	SVM	POSITIVO	0.27419673
366	GGWAAYTANTGGNRY	SVM	POSITIVO	0.27490477
367	NISYKLTGTSAQGP	SVM	NEGATIVO	-0.13711911
368	ISYKLTGTSAQGPG	SVM	NEGATIVO	-0.095125701
369	SYKLTGTSAQGPGP	SVM	NEGATIVO	-0.45694781
370	YKLTGTSAQGPGPG	SVM	NEGATIVO	-0.21092397
371	KLTGTSAQGPGPGI	SVM	NEGATIVO	-0.036915205
372	IAAAYGQIGVDFYAA	SVM	POSITIVO	0.28221447
373	TQEALGTVASQTRAV	SVM	POSITIVO	0.28301464
374	LGTSAQGPGPGIHTK	SVM	NEGATIVO	-0.10368019
375	GTSAQGPGPGIHTKM	SVM	NEGATIVO	-0.090647472
376	TSAQGPGPGIHTKMD	SVM	NEGATIVO	-0.090959896
377	LGAADLALATVNDLI	SVM	POSITIVO	0.29475444
378	AQGPGPGIHTKMD	SVM	NEGATIVO	-0.024570467

379	QGPGGIHTKMD	SVM	NEGATIVO	-0.013180579
380	GPPGGIHTKMD	SVM	NEGATIVO	-0.023156261
381	ASARAEGYVDQAVEL	SVM	POSITIVO	0.30587225
382	GPGIHTKMD	SVM	NEGATIVO	-0.14375157
383	RTKGVEVLNGPGG	SVM	NEGATIVO	-0.10823802
384	TKGVEVLNGPGGGM	SVM	NEGATIVO	-0.29710325
385	KGVEVLNGPGGGMH	SVM	NEGATIVO	-0.41909509
386	GVEVLNGPGGGMHI	SVM	NEGATIVO	-0.26436672
387	VEVLNGPGGGMHIM	SVM	NEGATIVO	-0.61374913
388	EVLNGPGGGMHIML	SVM	NEGATIVO	-0.5410201
389	VLNGPGGGMHIMLY	SVM	NEGATIVO	-0.5225017
390	LNGPGGGMHIMLYD	SVM	NEGATIVO	-0.68046189
391	NGPGGGMHIMLYDL	SVM	NEGATIVO	-0.66511339
392	GPPGGGMHIMLYDLK	SVM	NEGATIVO	-0.67329397
393	PGPGGGMHIMLYDLKG	SVM	NEGATIVO	-0.65010033
394	GPPGGGMHIMLYDLKGP	SVM	NEGATIVO	-10.222.806
395	PGGMHIMLYDLKGPL	SVM	NEGATIVO	-0.87208497
396	GGMHIMLYDLKGPLN	SVM	NEGATIVO	-0.9990277
397	GMHIMLYDLKGPLNV	SVM	NEGATIVO	-10.703.545
398	MHIMLYDLKGPLNVA	SVM	NEGATIVO	-10.634.026
399	HIMLYDLKGPLNVAA	SVM	NEGATIVO	-0.75410297
400	IMLYDLKGPLNVAAAY	SVM	NEGATIVO	-0.78763233
401	MLYDLKGPLNVAAAYR	SVM	NEGATIVO	-0.62631051
402	LYDLKGPLNVAAAYRV	SVM	NEGATIVO	-0.5003443
403	YDLKGPLNVAAAYRVN	SVM	NEGATIVO	-0.52141954
404	DLKGPLNVAAAYRVNF	SVM	NEGATIVO	-0.4777662
405	LKGPLNVAAAYRVNFF	SVM	NEGATIVO	-0.33148922
406	KGPLNVAAAYRVNFFT	SVM	NEGATIVO	-0.37214865
407	GPLNVAAAYRVNFFTP	SVM	NEGATIVO	-0.37969539
408	PLNVAAAYRVNFFTPK	SVM	NEGATIVO	-0.12716089
409	LNVAAYRVNFFTPKM	SVM	NEGATIVO	-0.21902906
410	NVAAYRVNFFTPKMA	SVM	NEGATIVO	-0.079865622
411	LTQEALGTVASQTRA	SVM	POSITIVO	0.30595246
412	IANLRERAEETRAET	SVM	POSITIVO	0.30682484
413	AYRVNFFTPKMAAYS	SVM	NEGATIVO	-0.10422298
414	YRVNFFTPKMAAYSS	SVM	NEGATIVO	-0.3775773
415	RVNFFTPKMAAYSSM	SVM	NEGATIVO	-0.67444568
416	VNFFTPKMAAYSSME	SVM	NEGATIVO	-0.66881444
417	NFFTPKMAAYSSMEN	SVM	NEGATIVO	-0.51183719
418	FFTPKMAAYSSMENQ	SVM	NEGATIVO	-0.66179478
419	FFTPKMAAYSSMENQW	SVM	NEGATIVO	-0.68141827
420	TPKMAAYSSMENQWY	SVM	NEGATIVO	-0.56608184
421	PKMAAYSSMENQWYL	SVM	NEGATIVO	-0.62227546
422	KMAAYSSMENQWYLA	SVM	NEGATIVO	-0.53475378
423	MAAYSSMENQWYLA	SVM	NEGATIVO	-0.34911677
424	AAAYSSMENQWYLAAY	SVM	NEGATIVO	-0.26365417
425	AYSSMENQWYLAAYQ	SVM	NEGATIVO	-0.15746494
426	YSSMENQWYLAAYQV	SVM	NEGATIVO	-0.33498103

427	SSMENQWYLAAYQVT	SVM	NEGATIVO	-0.38314134
428	SMENQWYLAAYQVTA	SVM	NEGATIVO	-0.34487992
429	MENQWYLAAYQVTAG	SVM	NEGATIVO	-0.30993475
430	ENQWYLAAYQVTAGT	SVM	NEGATIVO	-0.48385965
431	NQWYLAAYQVTAGTG	SVM	NEGATIVO	-0.57966716
432	QWYLAAYQVTAGTGG	SVM	NEGATIVO	-0.51852304
433	WYLAAYQVTAGTGGW	SVM	NEGATIVO	-0.11832399
434	YLAAYQVTAGTGGWA	SVM	NEGATIVO	-0.043665071
435	LAAYQVTAGTGGWAA	SVM	NEGATIVO	-0.00016427105
436	ANLRERAEETRAETR	SVM	POSITIVO	0.30828533
437	AYQVTAGTGGWAAYT	SVM	NEGATIVO	-0.011873184
438	PGPGNSDTSIKLEGF	SVM	POSITIVO	0.3122424
439	GAADLALATVNDLIA	SVM	POSITIVO	0.31790981
440	VTAGTGGWAAYTANT	SVM	NEGATIVO	-0.025886878
441	GWAAYTANTGGNRYA	SVM	POSITIVO	0.31875773
442	NRYAAYKSTRGNIVK	SVM	POSITIVO	0.3275812
443	RKAAEGYLEAATNRY	SVM	POSITIVO	0.33922804
444	AAYGQIGVDFYAAAY	SVM	POSITIVO	0.34010472
445	LRKAAEGYLEAATNR	SVM	POSITIVO	0.34433339
446	RAETRTRVEERRARL	SVM	POSITIVO	0.34996506
447	LEGFYLLEGPGPGRT	SVM	POSITIVO	0.36412773
448	KAAEGYLEAATNRYN	SVM	POSITIVO	0.3645018
449	AYTANTGGNRYAAYK	SVM	NEGATIVO	-0.23441511
450	YTANTGGNRYAAYKS	SVM	NEGATIVO	-0.19718565
451	TANTGGNRYAAYKST	SVM	NEGATIVO	-0.13528146
452	ANTGGNRYAAYKSTR	SVM	NEGATIVO	-0.084446565
453	TAGTGGWAAYTANTG	SVM	POSITIVO	0.39957715
454	AAEGYLEAATNRYNE	SVM	POSITIVO	0.4059626
455	QEALGTVASQTRAVG	SVM	POSITIVO	0.40993056
456	NLRERAEETRAETRT	SVM	POSITIVO	0.41405318
457	ELRKAAEGYLEAATN	SVM	POSITIVO	0.42301994
458	EETRAETRTRVEERR	SVM	POSITIVO	0.43695956
459	LIANLRERAEETRAE	SVM	POSITIVO	0.45548391
460	EALGTVASQTRAVGE	SVM	POSITIVO	0.47919824
461	PLLAALGAADLALAT	SVM	POSITIVO	0.49926943
462	YKSTRGNIVKAAAYKT	SVM	NEGATIVO	-0.21814903
463	KSTRGNIVKAAAYKTT	SVM	NEGATIVO	-0.39225937
464	STRGNIVKAAAYKTTR	SVM	NEGATIVO	-0.078681695
465	ALGTVASQTRAVGER	SVM	POSITIVO	0.52937284
466	RGNIVKAAAYKTTRGT	SVM	NEGATIVO	-0.20089573
467	GNIVKAAAYKTTRGTV	SVM	NEGATIVO	-0.56754019
468	NIVKAAAYKTTRGTVG	SVM	NEGATIVO	-0.58871394
469	IVKAAAYKTTRGTVGL	SVM	NEGATIVO	-0.37349172
470	VKAAAYKTTRGTVGLA	SVM	NEGATIVO	-0.25098069
471	KAAAYKTTRGTVGLAA	SVM	NEGATIVO	-0.17520884
472	EELRKAAEGYLEAAT	SVM	POSITIVO	0.55006559
473	AYKTTRGTVGLAAYS	SVM	NEGATIVO	-0.041432703
474	YKTTRGTVGLAAYSS	SVM	NEGATIVO	-0.2013011

475	KTTRGTVGLAAYSSF	SVM	NEGATIVO	-0.074321144
476	VASQTRAVGERAAKL	SVM	POSITIVO	0.55072089
477	TGGNRYAAKSTRGN	SVM	POSITIVO	0.5532313
478	AEETRAETRTRVEER	SVM	POSITIVO	0.55920212
479	GTVGLAAYSSFKITF	SVM	NEGATIVO	-0.15417429
480	TVGLAAYSSFKITFF	SVM	NEGATIVO	-0.11339719
481	VGLAAYSSFKITFFF	SVM	NEGATIVO	-0.17431327
482	GLAAYSSFKITFFFA	SVM	NEGATIVO	-0.26243455
483	LAAAYSSFKITFFFAA	SVM	NEGATIVO	-0.45249276
484	AAAYSSFKITFFFAAY	SVM	NEGATIVO	-0.13233084
485	AYSSFKITFFFAAYP	SVM	NEGATIVO	-0.15977549
486	YSSFKITFFFAAYPY	SVM	NEGATIVO	-0.61950951
487	SSFKITFFFAAYPYI	SVM	NEGATIVO	-0.45366911
488	SFKITFFFAAYPYIF	SVM	NEGATIVO	-0.5448695
489	FKITFFFAAYPYIFI	SVM	NEGATIVO	-0.43000795
490	KITFFFAAYPYIFIA	SVM	NEGATIVO	-0.57581681
491	ITFFFAAYPYIFIAT	SVM	NEGATIVO	-0.63435883
492	TFFFAAYPYIFIATF	SVM	NEGATIVO	-0.65260825
493	FFFAAYPYIFIATFF	SVM	NEGATIVO	-0.50029468
494	FFAAYPYIFIATFFA	SVM	NEGATIVO	-0.46264627
495	FAAYPYIFIATFFAA	SVM	NEGATIVO	-0.47530544
496	AAYPYIFIATFFAAY	SVM	NEGATIVO	-0.51814797
497	AYPYIFIATFFAAYF	SVM	NEGATIVO	-0.68199196
498	YPYIFIATFFAAYFL	SVM	NEGATIVO	-0.89554833
499	PYIFIATFFAAYFLL	SVM	NEGATIVO	-0.63401709
500	YIFIATFFAAYFLLP	SVM	NEGATIVO	-0.38603493
501	IFIATFFAAYFLLPY	SVM	NEGATIVO	-0.59592354
502	FIATFFAAYFLLPYI	SVM	NEGATIVO	-0.42400306
503	IATFFAAYFLLPYIF	SVM	NEGATIVO	-0.68881184
504	ATFFAAYFLLPYIFI	SVM	NEGATIVO	-0.51799692
505	TFFAAYFLLPYIFIA	SVM	NEGATIVO	-0.68573776
506	FFAAYFLLPYIFIAA	SVM	NEGATIVO	-0.40326731
507	FAAYFLLPYIFIAAA	SVM	NEGATIVO	-0.018197573
508	TRAETRTRVEERRAR	SVM	POSITIVO	0.56407767
509	AYFLLPYIFIAAAYG	SVM	NEGATIVO	-0.016929804
510	YFLLPYIFIAAAYGQ	SVM	NEGATIVO	-0.22091168
511	AAAYGQIGVDFYAAA	SVM	POSITIVO	0.57637913
512	LLPYIFIAAAYGQIG	SVM	NEGATIVO	-0.15236759
513	LPYIFIAAAYGQIGV	SVM	NEGATIVO	-0.29150337
514	PYIFIAAAYGQIGVD	SVM	NEGATIVO	-0.31002772
515	YIFIAAAYGQIGVDF	SVM	NEGATIVO	-0.14951665
516	IFIAAAYGQIGVDFY	SVM	NEGATIVO	-0.32751191
517	TEELRKAAGEYLEAA	SVM	POSITIVO	0.58030736
518	TRAVGERAAKLVGIE	SVM	POSITIVO	0.62604005
519	ETRAETRTRVEERRA	SVM	POSITIVO	0.6328511
520	AVGERAAKLVGIELK	SVM	POSITIVO	0.63611803
521	GGNRYAAKSTRGNI	SVM	POSITIVO	0.65128431
522	YGQIGVDFYAAAYNE	SVM	NEGATIVO	-0.062260931

523	GQIGVDFYAAAYNEC	SVM	NEGATIVO	-0.12662259
524	QIGVDFYAAAYNECG	SVM	NEGATIVO	-0.2272139
525	IGVDFYAAAYNECGE	SVM	NEGATIVO	-0.18350616
526	GVDFYAAAYNECGEV	SVM	NEGATIVO	-0.21775756
527	VDFYAAAYNECGEVN	SVM	NEGATIVO	-0.1407766
528	DFYAAAYNECGEVNP	SVM	NEGATIVO	-0.16003992
529	FYAAAYNECGEVNPI	SVM	NEGATIVO	-0.050204864
530	RERAEETRAETRTRV	SVM	POSITIVO	0.66965338
531	GNNRYAAYKSTRGNIV	SVM	POSITIVO	0.67006343
532	AAAYNECGEVNPIAAY	SVM	NEGATIVO	-0.028880976
533	ASQTRAVGERAAKL	SVM	POSITIVO	0.67014771
534	YNECGEVNPIAAYGT	SVM	NEGATIVO	-0.14912259
535	TVASQTRAVGERAAK	SVM	POSITIVO	0.67359072
536	LAALGAADLALATVN	SVM	POSITIVO	0.6812668
537	LRERAEETRAETRTR	SVM	POSITIVO	0.6887185
538	SQTRAVGERAAKLVG	SVM	POSITIVO	0.70298562
539	ERAEETRAETRTRVE	SVM	POSITIVO	0.71554854
540	LLAALGAADLALATV	SVM	POSITIVO	0.75124621
541	NPIAAYGTVINVHLK	SVM	NEGATIVO	-0.005414537
542	GTVASQTRAVGERAA	SVM	POSITIVO	0.77149353
543	QTRAVGERAAKLVG	SVM	POSITIVO	0.7913839
544	RAVGERAAKLVGIEL	SVM	POSITIVO	0.86574214
545	RAEETRAETRTRVEE	SVM	POSITIVO	0.86835363
546	YGTVINVHLKAAAYTG	SVM	NEGATIVO	-0.25823997
547	GTVINVHLKAAAYTGQ	SVM	NEGATIVO	-0.49112193
548	TVINVHLKAAAYTGQF	SVM	NEGATIVO	-0.39419775
549	VINVHLKAAAYTGQFH	SVM	NEGATIVO	-0.19730602
550	INVHLKAAAYTGQFHI	SVM	NEGATIVO	-0.58936486
551	NVHLKAAAYTGQFHIA	SVM	NEGATIVO	-0.55483486
552	VHLKAAAYTGQFHIA	SVM	NEGATIVO	-0.61972469
553	HLKAAAYTGQFHIA	SVM	NEGATIVO	-0.74858747
554	LKAAAYTGQFHIA	SVM	NEGATIVO	-0.6444112
555	KAAAYTGQFHIA	SVM	NEGATIVO	-0.66480221
556	AAAYTGQFHIA	SVM	NEGATIVO	-0.59809789
557	AYTGQFHIA	SVM	NEGATIVO	-0.44073237
558	YTGQFHIA	SVM	NEGATIVO	-0.2689636
559	LGTVASQTRAVGERA	SVM	POSITIVO	0.88559246

Fonte: Do Autor, 2022.