UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS DA SAÚDE
DEPARTAMENTO DE CIÊNCIAS FARMACÊUTICAS

EMANUELLE DE OLIVEIRA SOUZA

ABORDAGENS *IN SILICO* EM IMUNOINFORMÁTICA: UMA REVISÃO SISTEMÁTICA

Recife
2022

EMANUELLE DE OLIVEIRA SOUZA

ABORDAGENS *IN SILICO* EM IMUNOINFORMÁTICA: UMA REVISÃO SISTEMÁTICA

Trabalho de Conclusão de Curso apresentado à Disciplina de TCC 2 como parte dos requisitos para conclusão do Curso de Graduação em Farmácia do Centro de Ciências da Saúde da Universidade Federal de Pernambuco.

Orientador: Dr. Marcelo Zaldini Hernandes

Coorientador: Dr. Luiz Felipe Gomes Rebello Ferreira

Recife

2022

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

Souza, Emanuelle de Oliveira.
Abordagens in silico em imunoinformática: uma revisão sistemática / Emanuelle de Oliveira Souza. - Recife, 2022.
94 : il., tab.

Orientador(a): Marcelo Zaldini Hernandes
Cooorientador(a): Luiz Felipe Gomes Rebello Ferreira
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Ciências da Saúde, Farmácia - Bacharelado, 2022.
Inclui referências, apêndices.

1. Anticorpo. 2. Biologia Computacional. 3. Epítopo. 4. Imunoinformática. 5. Modelagem. I. Hernandes, Marcelo Zaldini. (Orientação). II. Ferreira, Luiz Felipe Gomes Rebello. (Coorientação). III. Título.

540 CDD (22.ed.)

UNIVERSIDADE FEDERAL DE PERNAMBUCO

CENTRO DE CIÊNCIAS DA SAÚDE

DEPARTAMENTO DE CIÊNCIAS FARMACÊUTICAS

CURSO DE BACHARELADO EM FARMÁCIA

Aprovada em: 27/10/2022.

BANCA EXAMINADORA

_____

Prof. Dr. Marcelo Zaldini Hernandes
(Presidente e Orientador)
Universidade Federal de Pernambuco

_____

Dr. Policarpo Ademar Sales Junior
(Examinador)
CPqAM/FIOCRUZ-PE

_____

Prof. Dr. Thiago André Ramos dos Santos
(Examinador)
Centro Universitário dos Guararapes

_____

Profa. Dra. Elis Dionísio da Silva
(Suplente)
Universidade Federal do Amazonas

Dedico este trabalho ao meu avô Damião que faleceu este ano e não pôde participar desta etapa da minha vida.

AGRADECIMENTOS

Ao corpo docente do Departamento de Ciências Farmacêuticas da Universidade Federal de Pernambuco por todos os ensinamentos durante estes cinco anos de graduação.

Ao professor Dr. Marcelo Zaldini Hernandes e ao Dr. Luiz Felipe Gomes Rebello Ferreira por me orientarem durante minha Iniciação Científica e agora no Trabalho de Conclusão de Curso.

A banca por ter aceitado o convite.

A minha mãe, por todo suporte emocional mesmo durante este tempo longe de casa, por sempre me apoiar em tudo o que faço, por confiar sempre em mim e em minhas decisões, te amo.

A minha tia, Denise, por me ajudar na revisão dos textos de trabalhos durante todo o tempo da graduação e também ao seu esposo Kléber por me acolherem em sua casa no Recife.

A minha avó Edite, por apoiar minha decisão de fazer faculdade fora e sempre apoiar minhas decisões.

As minhas amigas, Beatriz e Thammyris, por todo o companheirismo durante as atividades acadêmicas e fora delas, tornando estes anos difíceis de graduação mais leves e divertidos.

A Matheus pelo apoio demonstrado ao longo de todo o período de tempo em que me dediquei a este trabalho.

RESUMO

A imunoinformática é uma área de pesquisa que integra conhecimentos de imunogenética e biologia computacional e possui amplos campos de aplicação como predição de alérgenos, identificação de genes de virulência, vacinologia reversa, dentre outros. Esta diversidade de aplicações pode auxiliar na triagem e pesquisa de experimentos *in vitro* e *in vivo*, como por exemplo no desenvolvimento de vacinas e testes de diagnóstico. Diante da relevância do tema, o objetivo geral deste presente estudo é realizar uma revisão sistemática qualitativa sobre as ferramentas *in silico* aplicadas a imunoinformática, mais especificamente em duas temáticas intrinsicamente relacionadas: predição de epítopos de células B e modelagem de anticorpos. A revisão foi conduzida de acordo com a seguinte pergunta norteadora: "Quais ferramentas computacionais (*in silico*) são utilizadas para predição de epítopos de células B e modelagem de anticorpos?". Diante disto, o trabalho foi realizado em 4 etapas: i) identificação: onde foram selecionados os artigos escritos na língua inglesa, publicados nos últimos 10 anos, possuindo título e/ou palavras-chaves relacionados aos temas do presente estudo; ii) seleção: onde foram removidos os artigos duplicados e os que não estavam relacionados aos temas após uma leitura do resumo, além de ter sido realizada uma busca pela referência que descrevesse a versão mais recente da ferramenta computacional; iii) disponibilidade: onde foram verificados se as ferramentas possuíam acesso livre e/ou acadêmico, além de verificar se os servidores ainda estavam *on-line* (até setembro de 2022), para os casos onde as ferramentas foram disponibilizadas desta forma; e iv) revisão: onde foi realizada a sumarização e a leitura crítica das referências selecionadas. A partir disso, foram selecionados 26 artigos, das bases de dados eletrônicas *Pubmed/MEDLINE* e *Google* Acadêmico, para a revisão sistemática qualitativa, distribuídos da seguinte forma: 17 artigos foram direcionados para o tópico "predição *in silico* de epítopos de células B" e 9 artigos para "modelagem de anticorpos". Observa-se que dos artigos referentes ao primeiro tema, 65% estão relacionados a preditores de epítopos lineares, seguido de preditores de epítopos conformacionais (29%) e epítopos não peptídicos (6%). Além disso, a maior parte destes algoritmos se encontra disponíveis *on-line* (65%). Em relação ao segundo tema, as proporções entre a disponibilidade através de servidores *on-line* e programas *stand-alone* foram semelhantes. Dentre os artigos voltados para modelagem de anticorpos, 75% relatam ferramentas de modelagem por homologia e *ab initio*, enquanto apenas 25% descrevem protocolos de *redesign*. Ao final da revisão, observou-se a diversidade de aplicações das ferramentas computacionais da imunoinformática voltadas aos campos da predição de epítopos de células B e modelagem de anticorpos. Além disso, ao apresentar as vantagens e desvantagens de cada ferramenta, espera-se auxiliar os pesquisadores na escolha da(s) ferramenta(s) computacional(is) mais adequada(s) para solucionar seus respectivos problemas de pesquisa.

Palavras-chave: Anticorpo; Biologia Computacional; Epítopo; Imunoinformática; Modelagem; Predição.

ABSTRACT

Immunoinformatics is a research field that integrates knowledge of immunogenetics and computational biology and has broad area of application such as allergen prediction, identification of virulence genes, reverse vaccinology, among others. This diversity of applications may help the screening and research activities when performing in vitro and in vivo experiments, such as in the development of vaccines and diagnostic tests. Due to this topic relevance, this present study presents a systematic and qualitative review on the two in silico modeling tools applied to immunoinformatics, more specifically applied on two intrinsically related topics: epitope prediction of B cells and antibody modelling. The review has been conducted based on the following research question: “Which computational tools can be used for B prediction and antibody modeling. From this perspective, the work has been carried out in 4 stages: i) identification: where articles written in English, published in the last 10 years having the title and/or keywords related to the theme of this present study has been selected; ii) selection: where duplicate articles and the ones that were not related to the theme has been removed after reading the abstract, in addition to having carried out a search for the reference that describes the most recent version of the computational tool; iii) availability: where the tools that had free access and/or servers were selected, in addition to checking if they were still online (up to September 2nd, 2022) and iv) revision: where a summary and an judicious reading of the selected references was performed. In this way, a total of 26 articles were selected from the Pubmed/MEDLINE and Google Scholar electronic databases for the systematic review, distributed as follows: 17 articles were directed to the topic “in silico prediction of B cell epitopes” and 9 articles to “antibody modeling”. One may note that out of the articles referring to the first topic, 65% of them are related to linear epitopes predictors, followed by related predictors of conformational epitopes (29%) and non-peptide epitopes (6%). Additionally, most of them are available as an online service (65%). Regarding the second theme, it has been noticed that the way these prediction tools are made available has shown similar proportions for both online servers and stand-alone. Among the essential redesign tools for antibody modeling, 25% of them describes homology modeling protocols, while only 25% describes redesigns. It has been observed at the end of the review the availability of a diversity of computational tools on the field of immunoinformatics, particularly in the areas of B-cell epitope prediction and antibody modeling. Additionally, the summary presenting the advantages and disadvantages of each tool may help the researchers to carried out their studies by finding the most suitable computational tool to address their respective scientific problem.

Keywords: Antibody; Computational Biology; Epitope; Immunoinformatics; Modeling; Prediction.

LISTA DE QUADROS E TABELAS

Quadro 1 –	Critérios de inclusão e exclusão	20
Quadro 2 –	Sumarização dos artigos sobre predição *in silico* de candidatos à epítopos de células B em ordem de ano de publicação	24
Quadro 3 –	Sumarização dos artigos sobre modelagem de anticorpos em ordem de ano de publicação	30
Tabela 1 –	Desempenho da composição de dipeptídeos nos cinco conjuntos de dados do LBtope	37
Quadro 4 –	Características básicas das imunoglobulinas	44
Quadro 5 –	Vantagens e desvantagens dos algoritmos de predição de epítopos lineares em ordem alfabética	54
Quadro 6 –	Vantagens e desvantagens das ferramentas para predição de epítopos conformacionais em ordem alfabética	62
Quadro 7 –	Vantagens e desvantagens das ferramentas para modelagem de anticorpos por homologia e *ab initio* em ordem alfabética	74
Quadro 8 –	Vantagens e desvantagens das ferramentas para *redesign* de anticorpos em ordem alfabética	80

LISTA DE ILUSTRAÇÕES

Figura 1 –	Áreas de pesquisa relacionadas à imunoinformática	13
Figura 2 –	Arquitetura molecular de uma imunoglobulina IgG1	17
Fluxograma 1 –	Etapas da seleção de referências para a revisão sistemática	22
Gráfico 1 –	Motivo da exclusão dos artigos da revisão	23
Gráfico 2 –	Proporção entre algoritmos revisados para predição de epítopos lineares, conformacionais e não peptídicos	28
Gráfico 3 –	Distribuição dos algoritmos de predição de epítopos de células B	29
Gráfico 4 –	Metodologia das ferramentas computacionais relacionadas a modelagem de anticorpo	32
Gráfico 5 –	Distribuição dos algoritmos de modelagem de anticorpos	33
Gráfico 6 –	Distribuição temporal dos artigos revisados	33
Figura 3 –	Etapas de predição do DRREP	49
Figura 4 –	Arquitetura da rede neural profunda *feedforward* do DLBEpitope (A) e formação dos conjuntos de dados de treinamento e teste para IEDB38 (B)	52
Fluxograma 2 –	Procedimento do IsAb	77

LISTA DE ABREVIATURAS E SIGLAS

AA	Aminoácido
ACC	Acurácia
ADFS	*Allergen for Food Safety*
Ag-Ac	Antígeno-Anticorpo
AIP	*Amino-Acid Index Property*
AUC-ROC	Área Sob a Curva da Característica Operacional do Receptor
BCEPS	*B Cell Epitope Prediction Software*
BCR	Receptores de Células B
bNAb	Anticorpo neutralizante viral
CDR	Região Determinante de Complementariedade
CD-HIT	*Cluster Database at High Identity with Tolerance*
ChEBI	*Chemical Entities of Biological Interest*
DRREP	*Deep Ridge Regressed Epitope Predictor*
ECB	Epítopo de Célula B
ECBC	Epítopo de Célula B Conformacional
ECBL	Epítopo de Célula B Linear
ECT	Epítopo de Célula T
ELISA	*Enzyme-Linked Immunosorbent Assay*
ERT	*Extremely Randomized Tree*
DDR	*Design Risk Ratio* (Razão de Risco do *Design*)
Fab	Fragmento de ligação ao antígeno
FFT	*Fast Fourier Transform* (Transformação Rápida de Fourier)
GB	*Gradient Boosting*
HLA	Sistema de antígenos leucocitários humanos
IEDB	*Immune Epitope Database*
Ig	Imunoglobulina
iLBE	*Identification of Linear B-cell Epitope*
IMGT	Sistema internacional de informação em *ImMunoGeneTics*
IPRO	*IterativeProtein Redesign and Optimization*
IRMSD	Desvio Quadrático Médio da Interface
K-NN	*K-Nearest Neighbor*

MAPs	*Modular Antibody Parts*
MCC	Coeficiente de Correlação de Matthew
MHC	Complexo Principal de Histocompatibilidade
MILP	*Mixed-Integer Linear Programming*
NE	Não Epítopo
N-PEP	*Non-peptidic epitope prediction*
NPV	Valor Preditivo Negativo
PDB	*Protein Data Bank*
PKAF	C*omposition of profile-based amino acids frequency*
PPC	Cobertura de Proteção da População
PPV	Valor Preditivo Positivo
PSSM	*Position-Specific Scoring Matrix*
RF	*Random Forest*
RMN	Ressonância Magnética Nuclear
RMSD	Desvio Quadrático Médio
SAbDab	*Structural Antibody Database*
SMO	Otimização Sequencial Mínima
Sn	Sensibilidade
Sp	Especificidade
V_H	Domínio variável da cadeia pesada
V_L	Domínio variável da cadeia leve
WEKA	*Waikato Environment for Knowledge Analysis*

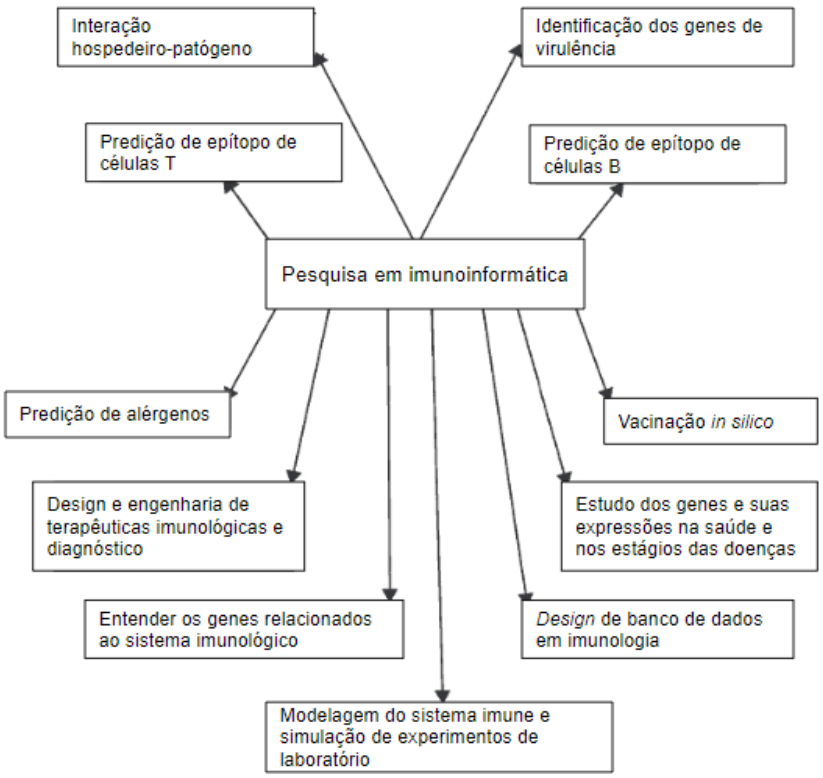
SUMÁRIO

1	**INTRODUÇÃO**	**13**
1.1	OBJETIVOS	15
1.1.1	**Objetivo Geral**	**15**
1.1.2	**Objetivos Específicos**	**15**
2	**REFERENCIAL TEÓRICO**	**16**
2.1	PREDIÇÃO *IN SILICO* DE EPÍTOPOS DE CÉLULAS B	16
2.2	MODELAGEM DE ANTICORPO	17
3	**METODOLOGIA**	**20**
4	**RESULTADOS**	**22**
5	**DISCUSSÃO**	**34**
5.1	PREDIÇÃO *IN SILICO* DE EPÍTOPOS DE CÉLULAS B	34
5.1.1	**Algoritmos para Predição de Epítopos Lineares**	**35**
5.1.2	**Algoritmos para Predição de Epítopos Conformacionais**	**57**
5.1.3	**Algoritmos para Predição de Epítopos não Peptídicos**	**64**
5.2	ALGORTIMOS PARA MODELAGEM DE ANTICORPOS	66
5.2.1	**Algoritmos de modelagem de anticorpos por homologia e *ab initio***	**66**
5.2.2	**Algoritmos de modelagem de anticorpos por *redesign***	**76**
6	**CONCLUSÃO**	**81**
	REFERÊNCIAS	**85**
	APÊNDICE A - ARTIGOS EXCLUÍDOS DA REVISÃO EM ORDEM DE ANO DE PUBLICAÇÃO	**92**

1 INTRODUÇÃO

Em 1989, foi criado o sistema internacional de informação em *ImMunoGeneTics* (IMGT), considerado o marco da criação da imunoinformática, possibilitando a integração da imunogenética e biologia computacional, ao desenvolver e utilizar ferramentas computacionais voltadas para o processamento e análise de dados biológicos (LEFRANC, 2014). A imunoinformática foi aprimorada ao longo dos anos como estratégia para análise e refinamento do grande volume de informações advindas do desenvolvimento nos campos da proteômica e genômica. Em decorrência disso, a imunoinformática tem sido empregada em diversas linhas de pesquisa (Figura 1), incluindo predição de alérgenos e epítopos de células T e B, *design* e engenharia de terapias imunológicas e diagnóstico, dentre outros (TOMAR; DE, 2010).

Figura 1 - Áreas de pesquisa relacionadas à imunoinformática.



Fonte: Traduzido de TOMAR; DE (2010).

Os resultados provenientes dos algoritmos têm auxiliado nas pesquisas *in vitro* e *in vivo*, como no desenvolvimento de novos alvos para diagnóstico (CARVALHO et al., 2017) e vacinologia reversa (TOMAR; DE, 2010). Atualmente, na pandemia da COVID-19, a biologia computacional foi aplicada para desenvolvimento de vacinas, alinhamento de sequenciamento e análise filogenética do vírus SARS-CoV2, além dos métodos de *docking* e dinâmica molecular que foram utilizados para realizar triagens de candidatos à fármacos (MA et al., 2021). Em outro estudo, foram aplicadas diferentes abordagens de dinâmica molecular para o desenvolvimento de dois anticorpos no intuito de serem utilizados na detecção do vírus SARS-CoV2, onde ao final da pesquisa um dos anticorpos se mostrou um candidato adequado para o desenvolvimento de um biossensor (MARTÍ et al., 2021).

As imunoglobulinas tem sido, ao longo dos anos, cada vez mais alvo de pesquisa devido à alta capacidade de reconhecimento de patógenos em decorrência da recombinação V(D)J das cadeias do Fab (fragmento de ligação ao antígeno), ocorrendo no início da maturação dos linfócitos B, permitindo uma variabilidade de reconhecimento de aproximadamente 10^{11} (cem bilhões) de estruturas (MESQUITA JÚNIOR et al., 2010).

O anticorpo se correlaciona com seu antígeno a partir de interações intermoleculares não covalentes como, por exemplo, ligações de hidrogênio, interações hidrofóbicas, interações eletrostáticas (iônicas). Ensaios *in vitro,* como o radioimunoensaio e o ELISA (*Enzyme-Linked Immunosorbent Assay)*, são usados para avaliar a afinidade e especificidade dos anticorpos frente aos antígenos (YANG et al., 2014). Entretanto, essas técnicas experimentais possuem limitações como o alto custo operacional, requerem grandes quantidades de amostra, necessitam de marcadores para visualização do resultado, e há a possibilidade de modificação química do antígeno ou anticorpo durante a imobilização (YANG et al., 2014). Neste contexto, a área de pesquisa relacionada à imunoinformática tem sido desenvolvida e aprimorada no âmbito de auxiliar os métodos experimentais *in vitro*, melhorando a triagem e análise dos dados obtidos, além de fornecer um conjunto de dados teóricos que possibilitam o desenvolvimento de novas hipóteses.

Diante da importância desta temática, o presente estudo se propõe a realizar uma revisão sistemática qualitativa sobre as principais ferramentas *in silico* aplicadas

em imunoinformática. Assim como apresentado anteriormente na Figura 1, a imunoinformática é uma extensa área de aplicação e pesquisa e, em decorrência disso, o presente estudo se concentrou em duas áreas relacionadas: predição de epítopos de células B e modelagem de anticorpos. Desta forma, espera-se contribuir ao apresentar um levantamento sobre o atual estado da arte no uso de ferramentas aplicadas a estas áreas de pesquisa.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

O presente estudo se propõe a realizar uma revisão sistemática qualitativa sobre as ferramentas *in silico* aplicadas à imunoinformática, nas áreas de predição de epítopos de células B e modelagem de anticorpos no intuito de auxiliar a comunidade acadêmica a selecionar as ferramentas mais adequadas às suas metas de pesquisa.

1.1.2 Objetivos Específicos

- Sumarizar os artigos selecionados, com informações sobre autores, ano de publicação, título do artigo, nome do respectivo algoritmo e classificação do algoritmo em *stand-alone*, servidor ou protocolo.
- Realizar um levantamento bibliográfico das principais metodologias *in silico* empregadas na modelagem de anticorpos e na predição de epítopos de células B.
- Apresentar as vantagens e desvantagens das diversas ferramentas computacionais encontradas na literatura.

2 REFERENCIAL TEÓRICO

2.1 PREDIÇÃO *IN SILICO* DE EPÍTOPOS DE CÉLULAS B

Os epítopos são porções do antígeno que desencadeiam a resposta imune de um organismo. Quando os epítopos são reconhecidos pelos receptores de células B (BCR), são denominados epítopos de células B (ECB), e podem ser peptídeos, carboidratos, lipídeos ou ainda pertencer a outras classes químicas (RAS-CARMON, et al., 2021; ZIEREP et al., 2022).

Os epítopos podem ser identificados através de abordagens experimentais, como cristalografia de raios-X, RMN (Ressonância Magnética Nuclear), troca hidrogênio-deutério acoplada à espectrometria de massa, etc. (ABBOTT; DAMSCHRODER, LOWE, 2014). Entretanto, por apresentarem altos custos e dificuldades técnicas inerentes, têm sido desenvolvidos algoritmos (implementados em ferramentas computacionais) que predizem quais porções do antígeno podem ser potenciais epítopos.

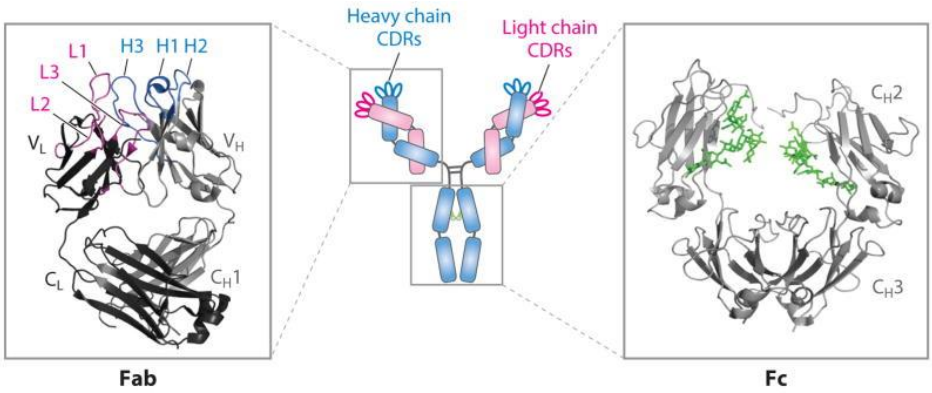
Grande parte das ferramentas computacionais foram desenvolvidas para a identificação dos dois tipos de ECB peptídicos: contínuos (lineares) e descontínuos (conformacionais). Os epítopos lineares (ECBL) são baseados na sequência de aminoácidos (estrutura primária) da proteína, enquanto os descontínuos são baseados na estrutura tridimensional do antígeno. Consequentemente, os epítopos previstos como mais antigênicos são os conformacionais (ECBC), pois levam em consideração os aminoácidos mais próximos à superfície do antígeno, que são os aminoácidos mais disponíveis para interação com receptores celulares (RAOUFI et al., 2019). Entretanto, os algoritmos que predizem ECBC apresentam desafios em seu desenvolvimento devido à baixa quantidade de estruturas tridimensionais determinadas experimentalmente, que servem como estruturas de referência. Por outro lado, as sequências de resíduos de aminoácidos são mais fáceis de serem determinadas e, por isso, estão disponíveis em maior quantidade, facilitando o desenvolvimento de algoritmos de predição de ECBL (RAS-CARMON et al., 2021).

A identificação de epítopos antigênicos pode auxiliar no desenvolvimento de novas vacinas, como é o caso das metodologias de vacinologia reversa, que por meio de métodos computacionais (*in silico*) analisam o proteoma (que pode, inclusive, ser predito a partir do genoma) de vírus, bactérias ou parasitas em busca de candidatos à epítopos com o potencial de provocar uma resposta imune, gerando uma memória imunológica e proteção a longo prazo (CHIANG et al., 2015).

2.2 MODELAGEM DE ANTICORPO

A estrutura monomérica dos anticorpos tem forma de “Y”, e é formada por duas cadeias pesadas (H, *Heavy*) e duas cadeias leves (L, *Light*), como pode ser observado na parte central da Figura 2. Existem dois tipos de cadeias leves (κ, λ) e cinco tipos de cadeias pesadas (α, δ, ε, γ, μ) que classificam as imunoglobulinas em cinco classes (IgA, IgD, IgE, IgG, IgM), cada classe possuindo estruturas e propriedades distintas entre si (MESQUITA JÚNIOR et al., 2010).

Figura 2 - Arquitetura molecular de uma imunoglobulina IgG1.



Fonte: TILLER; TESSIER (2015).

A parte inferior do anticorpo é denominada de Fragmento Cristalizável (Fc), sendo formado apenas pela cadeia pesada (domínios C_{H2} e C_{H3}) (Figura 2), responsável pela interação do anticorpo com os receptores das células efetoras (FcγR), e também pela interação com o sistema complemento (CHIU et al., 2019).

As duas partes superiores do anticorpo constituem o Fab, onde cada fragmento é formado por um conjunto de quatro domínios (dois pertencentes às cadeias leves e dois às pesadas), dividindo-se em domínio variável (V_H e V_L), localizado mais próximo ao local de interação com o antígeno, e domínio constante (C_{H1} e C_L), localizado mais próximo ao Fc (Figura 2) (TILLER; TESSIER, 2015).

O Fab possui, em sua extremidade, a Região Determinante de Complementariedade (CDR), formada por um conjunto de seis alças (*loops*) de aminoácidos que interagem com o antígeno, onde três delas pertencem à cadeia leve (L1, L2, L3), enquanto as outras pertencem à cadeia pesada (H1, H2, H3) (Figura 2). Durante as pesquisas realizadas nos anos 70, foram analisadas estruturas cristalográficas de IgG, onde os pesquisadores notaram que existiam conformações conservadas entre cinco das seis alças, e as denominaram de estruturas canônicas, sendo definidas de acordo com o seu comprimento e pelos aminoácidos constituintes (CHIU et al., 2019; PEDOTTI et al., 2011). Assim, atribuiu-se à alça H3 a imensa e impressionante variabilidade estrutural das IgGs, decorrente da recombinação gênica V(D)J e da hipermutação somática, resultando na diversificação da sequência de aminoácidos e do comprimento da alça (TILLER, TESSIER, 2015; VIDARSSON et al., 2014; WEITZNER et al., 2017).

A modelagem da alça H3 tem se mostrado um grande desafio na predição *in silico* de estruturas de anticorpos, devido à sua hipervariabilidade estrutural, além de sofrer mais alterações em sua conformação durante a interação com o antígeno, além de interagir com resíduos das duas cadeias por situar-se na interface dos dois domínios do fragmento variável (F_V). Ademais, esta alça está localizada no centro da interação com o antígeno, tornando-se um ponto chave para se alcançar bons resultados de estabilidade e afinidade nos complexos Antígeno-Anticorpo (WEITZNER et al., 2017; CHIU et al., 2019; SIRCAR; GRAY, 2010).

Em relação à modelagem molecular de anticorpos, existem duas metodologias mais usadas: modelagem por homologia e modelagem *ab initio*. A modelagem por homologia utiliza fragmentos de estruturas previamente conhecidas para a modelagem de novas estruturas (LEEM et al., 2016; SCHOEDER et al., 2021). Esses fragmentos modelos (“*template*”) são selecionados de acordo com a semelhança do tamanho da alça do CDR, semelhança de sequência primária e

semelhança de geometria. A principal desvantagem desse método é a necessidade de grande diversidade de estruturas já conhecidas, o que limita a predição de novas conformações para partes variáveis estruturalmente, como o CDRH3 (MARKS; DEANE, 2017). Enquanto a exploração conformacional é um fator limitante na modelagem por homologia, os métodos *ab initio* tem esta característica como seu ponto forte na modelagem, explorando as conformações através da variação dos graus de liberdade encontrados nos ângulos diédricos ϕ e ψ das alças do CDR, contudo com o ônus da diminuição na acurácia da modelagem de alças muito longas e um maior esforço computacional (MARKS; DEANE, 2017). Em decorrência disso, frequentemente os algoritmos de modelagem unem as duas metodologias para mitigar as limitações de ambas e tentar obter uma estrutura modelada que esteja o mais próximo possível da estrutura nativa do anticorpo, ou seja, uma precisão a mais na predição da sua estrutura tridimensional.

Neste contexto, uma nova abordagem de modelagem de anticorpos é o *redesign*, tendo como objetivo a modificação da estrutura de um anticorpo no intuito de melhorar sua afinidade com o antígeno alvo, podendo até mesmo ser aplicada com o objetivo de tornar o anticorpo específico para um antígeno diferente do seu alvo inicial (LIANG et al., 2021; NIMROD et al., 2018).

3 METODOLOGIA

Neste trabalho foi realizada uma revisão bibliográfica do tema imunoinformática, com base na seguinte pergunta norteadora: "***Quais ferramentas computacionais (in silico) são utilizadas para predição de epítopos de células B e modelagem de anticorpos?***". A intenção foi responder a esta pergunta a partir de um levantamento bibliográfico do tipo revisão sistemática qualitativa. Neste tipo de metodologia são utilizados métodos sistematizados e replicáveis de busca e seleção, permitindo ao leitor uma apresentação abrangente dos principais pontos abordados sobre o tema em pesquisa (SAMPAIO; MANCINI, 2007). Esta revisão sistemática qualitativa foi realizada em 4 etapas bem definidas: identificação, seleção, disponibilidade e revisão (BRASIL, 2012). Os critérios de inclusão e exclusão foram distribuídos nessas etapas da revisão e estão listados no Quadro 1.

Quadro 1 - Critérios de inclusão e exclusão.

Critérios de inclusão	Critérios de exclusão
● Referências que descrevem algoritmos ou protocolos computacionais com acesso livre e/ou acadêmico e, em caso de ferramentas *on-line*, o servidor deve estar em funcionamento. ● Referências escritas na língua inglesa. ● Referências publicadas nos últimos 10 anos. ● Referências com título e/ou palavras-chaves e resumo relacionados à temática.	● Referências que utilizam algoritmos em estudos de caso. ● Referências que descrevem algoritmos com uso exclusivamente comercial ou servidores *on-line* que não estão mais disponíveis. ● Referências escritas em outras línguas. ● Referências publicadas correspondentes a mais de 10 anos atrás. ● Referências não relacionadas a nenhuma das duas temáticas: predição *in silico* de epítopos de células B e modelagem de anticorpos.

Fonte: A autora (2022).

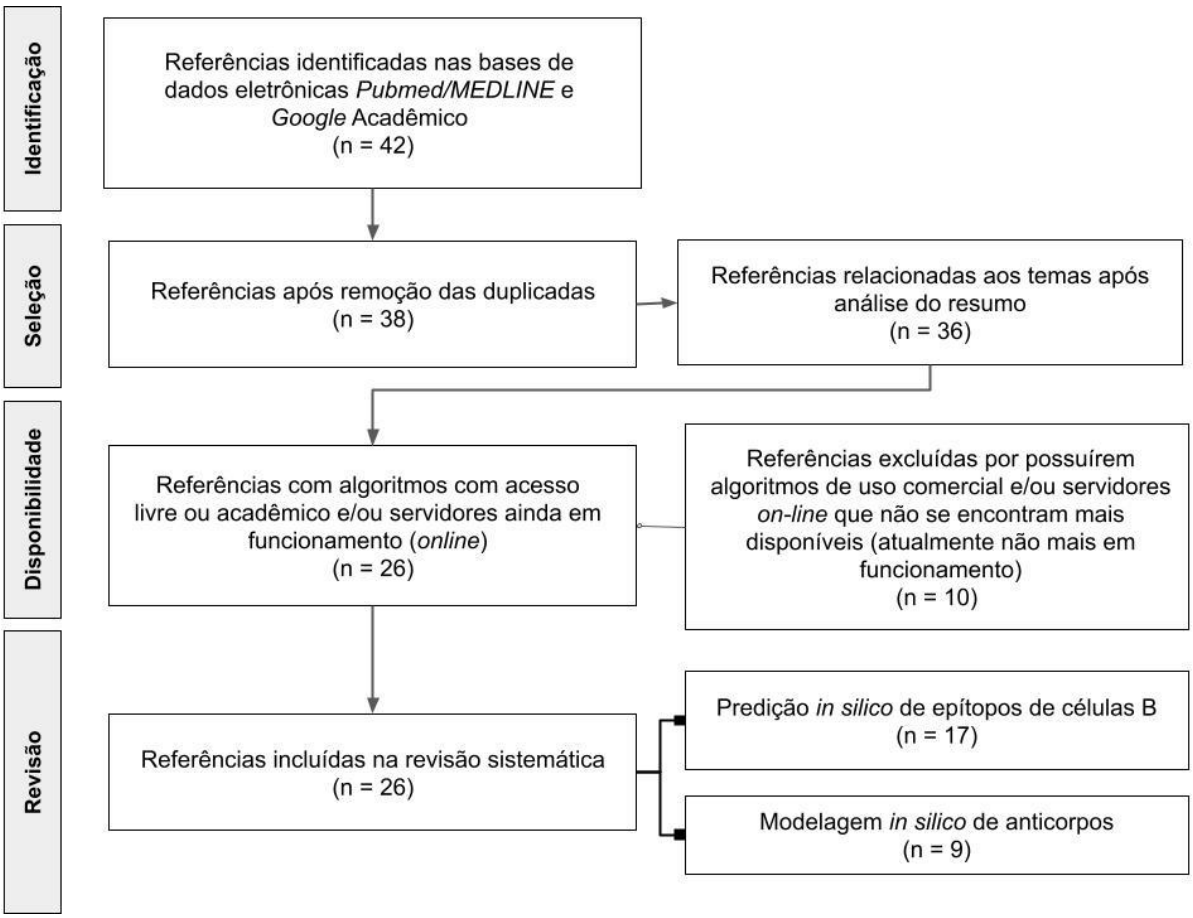
A primeira etapa da pesquisa (identificação) foi composta pela seleção de artigos científicos que descrevessem algoritmos ou protocolos computacionais relacionados aos temas de predição *in silico* de epítopos de células B ou modelagem de anticorpos. A busca de referências foi realizada utilizando os seguintes descritores nas bases de dados eletrônicas *Pubmed/MEDLINE* e *Google* Acadêmico: "*modeling epitope*", "*B-cell epitope prediction*", "*B-cell epitope conformational prediction*", "*B-cell epitope linear prediction*", "*modeling antibody in silico*", "*modeling antibody ab initio*", "*modeling antibody homology*", "*antibody computer simulation*", "*antibody design in silico*" e "*antibody design de novo*". Diante disto, foram selecionados artigos escritos na língua inglesa, publicados nos últimos 10 anos, possuindo título e/ou palavras-chaves relacionados aos temas do presente estudo.

A etapa da seleção foi direcionada para a remoção dos artigos que haviam sido selecionados em duplicidade, ou seja, foram encontrados e selecionados em ambas as bases de dados eletrônicas. Além disso, foi realizada uma seleção mais robusta, através da análise dos resumos das publicações, removendo as que não estavam relacionadas a nenhum dos dois temas, e buscando referências mais recentes em relação àquela respectiva ferramenta computacional. Posteriormente, na etapa de disponibilidade, foram selecionadas as referências que apresentavam algoritmos disponíveis em servidores *on-line* (até setembro de 2022) e com acesso livre e/ou acadêmico, excluindo os artigos que apresentavam algoritmos *on-line* que não se encontravam mais disponíveis (atualmente não mais em funcionamento) e/ou possuíam distribuição do algoritmo apenas comercial. Na última etapa (revisão), foi realizada a sumarização e leitura crítica das referências selecionadas.

4 RESULTADOS

As quatro etapas da revisão sistemática (identificação, seleção, disponibilidade e revisão) estão esquematizadas no Fluxograma 1.

Fluxograma 1 - Etapas da seleção de referências para a revisão sistemática.



Fonte: A autora (2022) com base em BRASIL (2012).

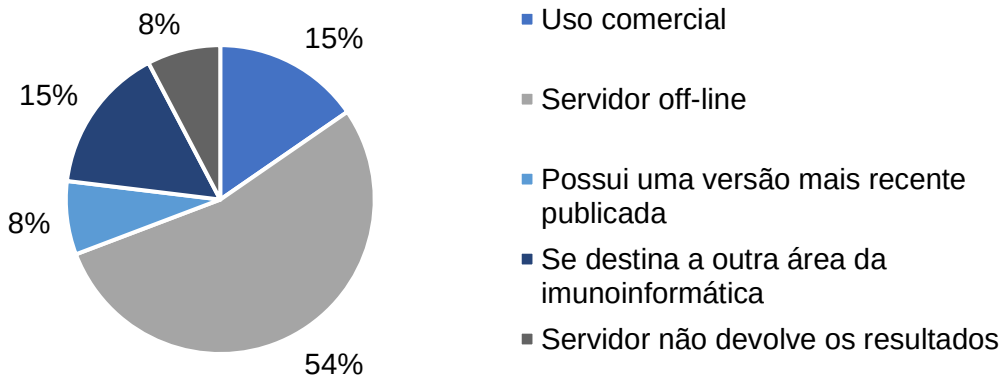
Na etapa de identificação foram encontrados 42 artigos em inglês. Entretanto, na etapa de seleção, observou-se que 4 artigos eram duplicados e, portanto, suas cópias foram excluídas, enquanto que uma referência foi substituída pela versão mais recente do algoritmo, alterando o número de artigos para 38. Além disso, após a leitura e análise dos resumos, o número de artigos passou para 36,

pois 2 artigos estavam relacionados a outros tópicos da imunoinformática como, por exemplo, humanização de anticorpos, que não é foco desta revisão.

Alguns dos artigos selecionados disponibilizavam seus algoritmos em servidores, portanto analisou-se na etapa de disponibilidade se os servidores estavam atualmente disponíveis *on-line* (até setembro de 2022), pois é comum na área da biologia computacional ocorrer descontinuidade de serviços/servidores *on-line*, pelo menos depois de algum tempo transcorrido no projeto. Além disso, foi analisado se o licenciamento do algoritmo estava disponível apenas para o uso comercial, dificultando a utilização por parte da comunidade acadêmica. No intuito de realizar uma revisão sobre as ferramentas que possam ser usadas em projetos de pesquisa, decidiu-se pela exclusão dos artigos que se classificavam nas duas situações apresentadas acima.

Após estas quatro etapas, foram incluídos 26 artigos para a revisão sistemática qualitativa, distribuindo-os da seguinte forma: 17 artigos foram direcionados para o tópico “predição *in silico* de epítopos de células B” (Quadro 2) e 9 artigos para “modelagem de anticorpos” (Quadro 3). Os artigos excluídos durante a metodologia estão sumarizados no apêndice A, onde observa-se que o principal motivo para exclusão está relacionado à falta de disponibilidade atual dos algoritmos em servidores que foram desativados (*off-line*) até setembro de 2022 (54%), seguido de algoritmos com uso exclusivamente comercial e algoritmos destinados a outra área da imunoinformática com 15%, cada (Gráfico 1).

Gráfico 1 – Motivo da exclusão dos artigos da revisão.



Fonte: A autora (2022).

Quadro 2 - Sumarização dos artigos sobre predição *in silico* de candidatos à epítopos de células B em ordem de ano de publicação (continua).

Autores e ano de publicação	**Título do artigo e DOI**	**Nome do algoritmo**	**Tipo de epítopo para predição**	**Tipo de distribuição do algoritmo**
YAO et al., 2012	SVMTriP: a method to predict antigenic epitopes using support vector machine to integrate tri-peptide similarity and propensity (DOI: 10.1371/journal.pone.0045152)	SVMTriP	Linear	*Stand-alone* e Servidor (http://sysbio.unl.edu/SVMTriP/)
SINGH; ANSARI; RAGHAVA, 2013	Improved Method for Linear B-Cell Epitope Prediction Using Antigen's Primary Sequence (DOI: 10.1371/journal.pone.0062216)	LBtope	Linear	Servidor (http://crdd.osdd.net/raghava/lbtope/)
KRAWCZYK et al., 2014	Improving B-cell epitope prediction and its application to global antibody-antigen docking (DOI:10.1093/bioinformatics/btu190)	EpiPred	Conformacional	*Stand-alone*
SELA-CULANG et al., 2014	Using a Combined Computational-Experimental Approach to Predict Antibody-Specific B Cell Epitopes (DOI: 10.1016/j.str.2014.02.003)	Predição Específica de Anticorpo (PEAc)	Conformacional	*Stand-alone*
DALKAS; ROOMAN, 2017	SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence (DOI:10.1186/s12859-017-1528-9)	SEPIa	Conformacional	*Stand-alone*

Quadro 2 - Sumarização dos artigos sobre predição *in silico* de candidatos à epítopos de células B em ordem de ano de publicação (continua).

Autores e ano de publicação	**Título do artigo e DOI**	**Nome do algoritmo**	**Tipo de epítopo para predição**	**Tipo de distribuição do algoritmo**
JESPERSEN et al., 2017	BepiPred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes (DOI:10.1093/nar/gkx346)	BepiPred-2.0	Linear	Servidor (https://services.healthtech.dtu.dk/service.php?BepiPred-2.0)
REN et al., 2017	Staged heterogeneity learning to identify conformational B-cell epitopes from antigen sequences (DOI:10.1186/s12864-017-3493-0)	SePre	Conformacional	*Stand-alone*
SHER; ZHI; ZHANG, 2017	DRREP: deep ridge regressed epitope predictor (DOI: 10.1186/s12864-017-4024-8)	DRREP	Linear	*Stand-alone*
MANAVALAN et al., 2018	iBCE-EL: a new ensemble learning framework for improved linear b-cell epitope prediction (DOI:10.3389/fimmu.2018.01695)	iBCE-EL	Linear	Servidor (http://thegleelab.org/iBCE-EL/)
HASAN; KHATUN; KURATA, 2020	iLBE for Computational Identification of Linear B-cell Epitopes by Integrating Sequence and Evolutionary Features (DOI:10.1016/j.gpb.2019.04.004)	iLBE	Linear	Servidor (http://kurata14.bio.kyutech.ac.jp/iLBE/)

Quadro 2 - Sumarização dos artigos sobre predição *in silico* de candidatos à epítopos de células B em ordem de ano de publicação (continua).

Autores e ano de publicação	**Título do artigo e DOI**	**Nome do algoritmo**	**Tipo de epítopo para predição**	**Tipo de distribuição do algoritmo**
LIU; SHI; LI, 2020	Deep learning methods improve linear B-cell epitope prediction (DOI:10.1186/s13040-020-00211-0)	DLBEpitope	Linear	Servidor (http://ccb1.bmi.ac.cn:81/dlbepitope/)
ASHFORD et al., 2021	Organism-specific training improves performance of linear B-cell epitope prediction (DOI:10.1093/bioinformatics/btab536)	RF-OrgSpec	Linear	*Stand-alone*
RAS-CARMON et al., 2021	BCEPS: a web server to predict linear b cell epitopes with enhanced immunogenicity and cross-reactivity (DOI:10.3390/cells10102744)	BCEPS	Linear	Servidor (http://imbio.med.ucm.es/bceps/)
TUNG et al., 2021	NIgPred: class-specific antibody prediction for linear b-cell epitopes based on heterogeneous features and machine-learning approaches (DOI: 10.3390/v13081531)	NIgPred	Linear	Servidor (Link não disponibilizado no artigo)
MONTIEL--GARCIA et al., 2022	Epitope-Analyzer: a structure-based webtool to analyze broadly neutralizing epitopes (DOI:10.1016/j.jsb.2022.107839)	Epitope--Analyzer (EA)	Conformacional	Servidor (http://viperdb.scripps.edu/ea.php)

Quadro 2 - Sumarização dos artigos sobre predição *in silico* de candidatos à epítopos de células B em ordem de ano de publicação (conclusão).

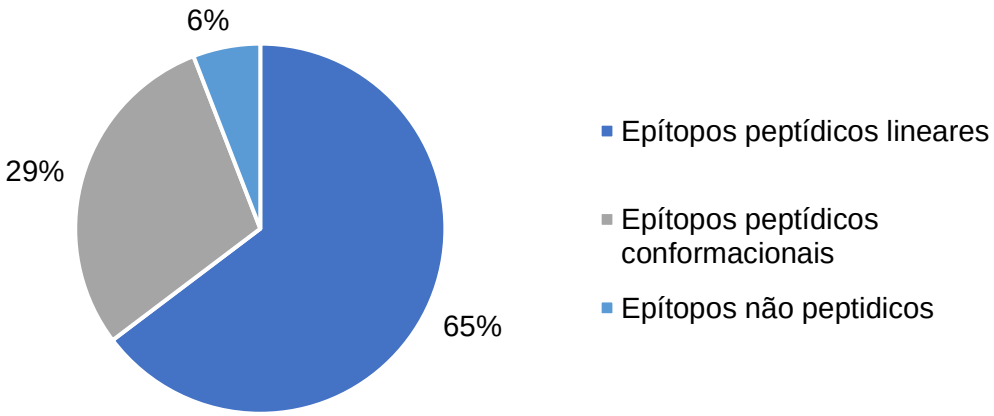
Autores e ano de publicação	**Título do artigo e DOI**	**Nome do algoritmo**	**Tipo de epítopo para predição**	**Tipo de distribuição do algoritmo**
ZHU et al., 2022	BepiTBR: T-B reciprocity enhances b cell epitope prediction (DOI: 10.1016/j.isci.2022.103764)	BepiTBR	Linear	Servidor (https://dbai.biohpc.swmed.edu/bepitbr/)
ZIEREP et al., 2022	Towards the prediction of non-peptidic epitopes (DOI: 10.1371/journal.pcbi.1009151)	Non-peptidic epitope prediction (N-PEP)	Não peptídico	Servidor (http://tools-staging.iedb.org/np_epitope_predictor)

Fonte: A autora (2022).

.

A proporção entre algoritmos para predição de epítopos lineares, conformacionais e não peptídicos está apresentado no Gráfico 2. Observa-se que houve uma maior proporção de artigos referentes a algoritmos para predição de epítopos lineares (65%), o que pode estar relacionado a menor disponibilidade de estruturas tridimensionais de epítopos em bancos de dados, o que interfere diretamente na quantidade de dados de treinamento para o algoritmo de predição de epítopos conformacionais (29%). Dentre as referências selecionadas para esta revisão, apenas 1 (6%) descreve um algoritmo para predição de epítopos não peptídicos, pois ainda é uma área pouco explorada.

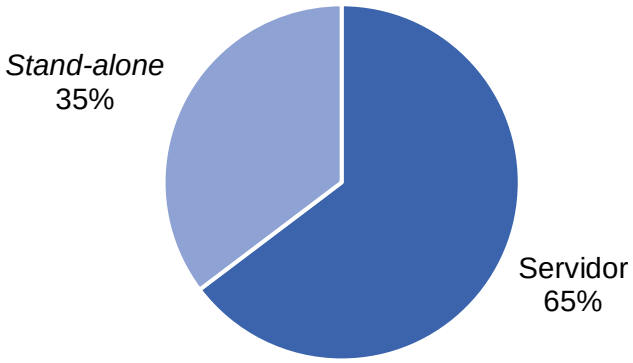
Gráfico 2 – Proporção entre algoritmos revisados para predição de epítopos lineares, conformacionais e não peptídicos.



Fonte: A autora (2022).

Além disso, nenhum dos artigos revisados descreve um protocolo computacional. Entretanto, observa-se que 65% dos algoritmos revisados estão disponíveis em servidores *on-line* (Gráfico 3) o que facilita o acesso e utilização por parte dos pesquisadores, especialmente para os inexperientes na área de imunoinformática ou que não possuem recursos computacionais suficientes para executar estas metodologias em seus próprios computadores.

Gráfico 3 – Distribuição dos algoritmos de predição de epítopos de células B



Fonte: A autora (2022).

Quadro 3 - Sumarização dos artigos sobre modelagem de anticorpos em ordem de ano de publicação (continua).

Autores e ano de publicação	**Título do artigo e DOI**	**Nome do algoritmo**	**Metodologia de modelagem**	**Tipo de distribuição do algoritmo**
LAPIDOTH et al., 2015	AbDesign: an algorithm for combinatorial backbone design guided by natural conformations and sequences (DOI: 10.1002/prot.24779)	AbDesign	Homologia e *ab initio*	*Stand-alone*
LEEM et al., 2016	ABodyBuilder: Automated antibody structure prediction with data–driven accuracy estimation (DOI: 10.1080/19420862.2016.1205773)	AbodyBuilder	Homologia e *ab initio*	Servidor (http://opig.stats.ox.ac.uk/webapps/newsabdab/sabpred/abodybuilder/)
MARKS et al., 2017	Sphinx: merging knowledge-based and ab initio approaches to improve protein loop prediction (DOI: 10.1093/bioinformatics/btw823)	Sphinx	Homologia e *ab initio*	*Stand-alone*
WEITZNER et al., 2017	Modeling and docking antibody structures with Rosetta (DOI: 10.1038/nprot.2016.180)	RosettaAntibody	Homologia e *ab initio*	*Stand-alone* e Servidor (https://rosie.graylab.jhu.edu/antibody)
ADOLF-BRYFOGLE et al., 2018	RosettaAntibodyDesign (RAbD): A general framework for computational antibody design (DOI: 10.1371/journal.pcbi.1006112)	RosettaAntibody Design (RAbD)	Homologia e *ab initio*	*Stand-alone*

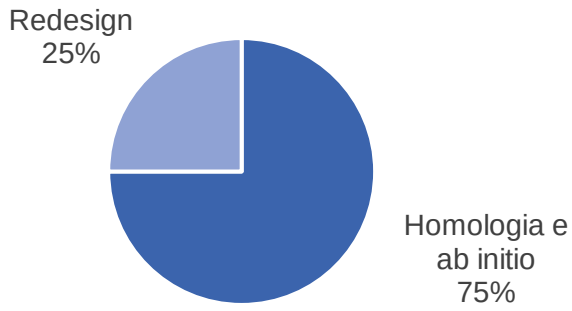
Quadro 3 - Sumarização dos artigos sobre modelagem de anticorpos em ordem de ano de publicação (conclusão)

Autores e ano de publicação	**Título do artigo e DOI**	**Nome do algoritmo**	**Metodologia de modelagem**	**Tipo de distribuição do algoritmo**
CHOWDHURY; ALLAN; MARANAS, 2018	OptMAVEn-2.0: De novo Design of Variable Antibody Regions against Targeted Antigen Epitopes (DOI: 10.3390/antib7030023)	OptMAVEn-2.0	Homologia e *ab initio*	*Stand-alone*
NIMROD et al., 2018	Computational Design of Epitope-Specific Functional Antibodies (DOI: 10.1016/j.celrep.2018.10.081)	Re-epitoping	*Redesign*	Protocolo
LIANG et al., 2021	IsAb: a computational protocol for antibody design (DOI: 10.1093/bib/bbab143)	IsAb	*Redesign*	Protocolo
SCHOEDER et al., 2021	Modeling Immunity with Rosetta: methods for antibody and antigen design (DOI: 10.1021/acs.biochem.0c00912)	AbDesign	Homologia e *ab initio*	*Stand-alone*
		RosettaAntibody	Homologia e *ab initio*	*Stand-alone* e Servidor (https://rosie.graylab.jhu.edu/antibody)
		RosettaAntibody Design (RAbD)	Homologia e *ab initio*	*Stand-alone*

Fonte: A autora (2022).

Nove referências foram selecionadas para o tema de modelagem de anticorpos e oito ferramentas foram revisadas. Destas, 75% utilizam a metodologia de modelagem por homologia em conjunto com *ab initio*, enquanto apenas 25% utilizam o *redesign* (Gráfico 4).

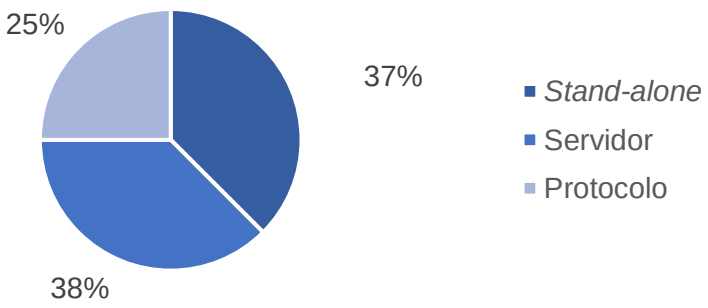
Gráfico 4 – Metodologia das ferramentas computacionais relacionadas a modelagem de anticorpo



Fonte: A autora (2022).

Além disso, estas metodologias apresentam uma proporção semelhante na distribuição dos algoritmos, onde 37% estão disponíveis como programas instaláveis (*stand-alone*), enquanto que 38% estão disponíveis para uso através de servidores *on-line*. Ao contrário do tópico de predição *in silico* de epítopos, a modelagem de anticorpos apresentou dois artigos sobre “protocolos computacionais” que representam 25% de todas as ferramentas revisadas para este tema.

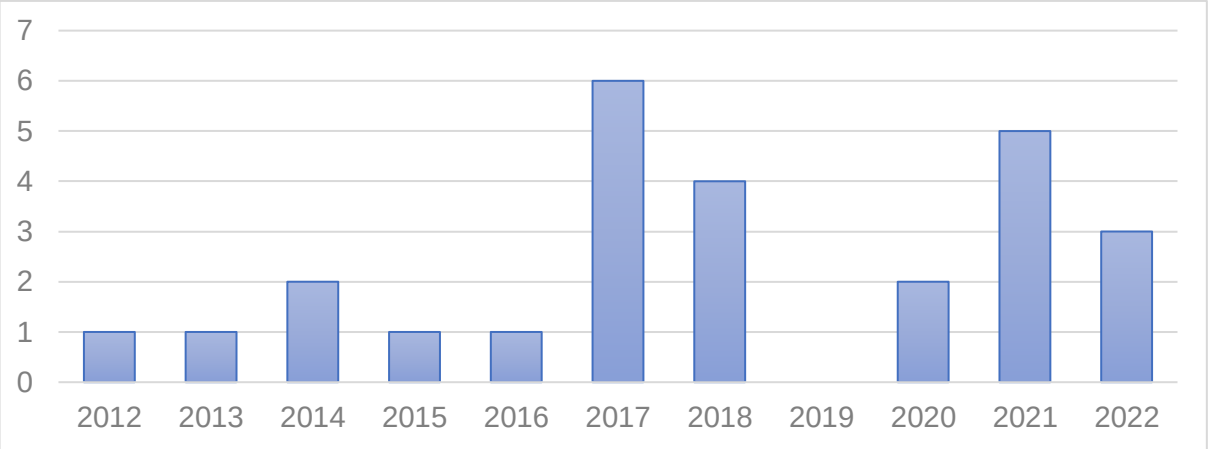
Gráfico 5 - Distribuição dos algoritmos de modelagem de anticorpos.



Fonte: A autora (2022).

O Gráfico 6 apresenta a distribuição temporal das 26 referências revisadas, onde percebe-se que o ano de 2017 apresentou o maior quantitativo de artigos, seguido do ano de 2021, 2018 e 2022. Apesar disso, observa-se que este presente estudo conseguiu incluir ao menos um artigo de cada ano dentro do período designado de 10 anos, mostrando uma boa homogeneidade na distribuição temporal, excetuando-se o ano de 2019.

Gráfico 6 - Distribuição temporal dos artigos revisados.



Fonte: A autora (2022).

.

5 DISCUSSÃO

5.1 PREDIÇÃO *IN SILICO* DE CANDIDATOS À EPÍTOPOS DE CÉLULAS B

Os epítopos de células B (ECB) podem ser provenientes de diferentes naturezas químicas, como peptídeos, carboidratos, lipídeos, dentre outros (RAS-CARMON et al., 2021; ZIEREP et al., 2022). Os ECB peptídicos são os mais estudados e por isso possuem mais algoritmos de predição desenvolvidos. Estes epítopos são divididos em lineares (ECBL) e conformacionais (ECBC). Os epítopos lineares, também denominados contínuos, tomam como base a sequência primária do epítopo, ou seja, levam em consideração a sequência de aminoácidos do peptídeo (JESPERSEN et al., 2017; RAS-CARMON et al., 2021). Por outro lado, os epítopos conformacionais ou descontínuos são assim denominados por levarem em consideração a estrutura tridimensional do peptídeo, e desta forma identificar os aminoácidos do antígeno que estão em contato direto com o parátopo do anticorpo (local de interação com o antígeno, mais especificamente com o epítopo), o que nem sempre segue a sequência primária do peptídeo (ASHFORD et al., 2021; JESPERSEN et al., 2017).

Algoritmos que predizem ECBC podem indicar epítopos mais imunogênicos, pois levam em consideração a informação da interação com o anticorpo, entretanto apresentam restrição de dados de referência, devido a poucas estruturas resolvidas experimentalmente para complexos antígeno-anticorpo, enquanto que a sequência de aminoácidos está disponível para um conjunto maior de estruturas (RAS-CARMON, et al., 2021), e isto se traduz em mais dados para o treinamento dos algoritmos de predição de ECBL. Outro fator que proporciona uma desvantagem dos algoritmos de predição de ECBC em relação ao ECBL é a dificuldade de recriar a conformação do peptídeo, pois mudanças no pH, temperatura e outros fatores influenciam na conformação da proteína e exposição dos aminoácidos, potencialmente fazendo com que os epítopos preditos possam estar inacessíveis ao anticorpo durante os testes *in vitro* ou *in vivo* (ASHFORD et al., 2021). Além disto, vale ressaltar que ECBL podem ser facilmente sintetizados e utilizados *in vitro* para

desenvolvimento de vacinas ou ferramentas diagnósticas (RAS-CARMONA et al., 2021).

5.1.1 Algoritmos para Predição de Epítopos Lineares

Embora a maioria dos algoritmos de predição de ECBL se baseiem em algoritmos de aprendizado de máquina (*machine learning*), onde cada ferramenta apresenta suas particularidades, o desenvolvimento destes algoritmos pode ser resumido da seguinte forma:

1. Organização de dois conjuntos de dados de treinamento, o conjunto de teste e o conjunto independente, onde as mesmas estruturas ou sequências não são utilizadas simultaneamente para os conjuntos de treinamento ou teste, com o objetivo de não criar vieses/tendências durante a comparação com outros algoritmos. Estes conjuntos são subdivididos em ECB e não epítopos (NE), geralmente extraídos do IEDB (*Immune Epitope Database*) (https://www.iedb.org/), um banco público de epítopos que também disponibiliza informações de interações destas estruturas com anticorpos, células T e MHC (Complexo Principal de Histocompatibilidade) (VITA et al., 2018);
2. Exclusão de estruturas com semelhança de sequência entre 70 a 80% do conjunto de dados, geralmente utilizando o CD-HIT (*Cluster Database at High Identity with Tolerance*), gerando um conjunto de estruturas não redundantes, ou seja, não semelhantes;
3. Treinamento do algoritmo com o conjunto de dados de treinamento, onde podem ser utilizadas propriedades físico-químicas (hidrofobicidade, carga, acessibilidade a superfície, etc.), composição dipeptídica, índice de aminoácidos, dentre outras abordagens para permitir que o algoritmo identifique os aminoácidos pertencentes ao epítopo e distinga-os do restante do antígeno. Posteriormente, são avaliados em experimentos de validação cruzada de 5 ou 10 vezes;

4. Avaliação do desempenho do algoritmo em um grupo teste com epítopos e não epítopos;
5. Uso do conjunto de dados independente para comparação com outros algoritmos já consolidados na literatura.

A avaliação da qualidade da predição dos algoritmos é realizada através dos índices: i) sensibilidade (Sn) ou *recall* que verifica a capacidade do algoritmo em detectar os resultados positivos (MARIANO, 2018); ii) acurácia (ACC) que calcula o percentual de acerto em relação ao número total de itens de entrada (MARIANO, 2018); iii) precisão que verifica a quantidade de verdadeiros positivos encontrados em relação a totalidade de valores positivos (MARIANO, 2018); iv) especificidade (Sp) que verifica a capacidade do algoritmo em detectar os resultados negativos (MARIANO, 2018); v) AUC-ROC (Área Sob a Curva da Característica de Operação do Receptor), a curva ROC é um gráfico cartesiano com valores de sensibilidade e especificidade, permitindo avaliar a capacidade de discriminação do classificador. A partir da área curva do gráfico ROC, se determina a AUC por Soma de Riemann (SOUSA, 2016); e vi) MCC (Coeficiente de Correlação de Matthews) que mede a qualidade de uma predição ou classificação com valores entre -1 e +1, onde +1 indica predição perfeita, 0 indica predição aleatória e -1 indica predição inversa (SOUSA, 2016). Estes índices utilizam valores de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos obtidos pelos algoritmos para os cálculos (MARIANO, 2018).

Os primeiros preditores *in silico* de ECB utilizavam poucas estruturas para treinamento, além de utilizar peptídeos aleatórios como NE (SINGH; ANSARI; RAGHAVA, 2013). Diante disto, o LBtope foi o primeiro algoritmo de predição a utilizar estruturas confirmadas experimentalmente. Para seu desenvolvimento, foram obtidas do IEDB 49.694 estruturas de ECB e 50.324 estruturas de NE, dividindo-os em cinco conjunto de dados: i) Lbtope_Fixed contendo 19.803 ECB e 28.329 NE de comprimento definido em 20 aminoácidos (AAs) a partir de estratégias de extensão e truncamento; ii) Lbtope_Fixed_non_redundant contendo 7.824 ECB e 7.853 NE obtidos do Lbtope_Fixed após utilizar o CD-HIT para retirar estruturas com identidade de sequência superiores a 80%; iii) Lbtope_Variable contendo 14.876 ECB e 23.321 NE, após retirar estruturas com menos de 5 AAs e com mais de 50

AAs e estruturas comuns nos dois grupos; iv) Lbtope_Variable_non_redundant contendo 8.011 de ECB e 10.868 de NE advindos do conjunto de dados Lbtope_Variable após retirada de estruturas com semelhanças superiores a 80% de sequência; v) Lbtope_Confirm contendo 1.042 de ECB e 1.795 de NE confirmados experimentalmente em pelo menos 2 estudos distintos (SINGH; ANSARI; RAGHAVA, 2013).

O LBtope implementou dois algoritmos de aprendizado de máquina: SVM_light (*Support Vector Machine*) e k-NN (*K-Nearest Neighbor*). Estes foram treinados com os recursos de perfil binário, propriedades físico-químicas, AAP (par de aminoácidos de Chen), composição dipeptídica, CTD (Composição-Transição-Distribuição) e AAP modificado (AAP*) onde cada valor de dipeptídeo foi atribuído a um respectivo dipeptídeo de uma matriz (SINGH; ANSARI; RAGHAVA, 2013). O desempenho do SVM e k-NN foram avaliados nos cinco bancos de dados com cada um dos seis recursos (SINGH; ANSARI; RAGHAVA, 2013). A composição de aminoácidos apresentou os maiores valores de ACC, MCC e AUC-ROC em todos os conjuntos de dados tanto para SVM quanto para k-NN dentre todos os recursos. Diante disto, a composição de dipetídeos foi aqui utilizada para comparação do desempenho entre os conjuntos de dados (Tabela 1).

Tabela 1 – Desempenho da composição de dipeptídeos nos cinco conjuntos de dados do LBtope.

Conjunto de dados		**Sn**	**Sp**	**ACC**	**MCC**	**AUC-ROC**
Lbtope_Fixed	**SVM**	80,5	81,67	81,24	0,61	0,88
	k-NN	80,33	81,67	81,18	0,61	0,86
Lbtope_Variable	**SVM**	75,18	76,34	75,89	0,51	0,83
	k-NN	80,90	77,50	78,82	0,57	0,84
Lbtope_Confirm	**SVM**	84,62	81,01	81,33	0,64	0,91
	k-NN	82,69	87,71	85,87	0,70	0,92
Lbtope_Fixed_non_redundant	**SVM**	65,75	63,97	64,86	0,30	0,69
	k-NN	-	-	-	-	-
Lbtope_Variable_non_redundant	**SVM**	66	67,24	66,7	0,33	0,73
	k-NN	-	-	-	-	-

Fonte: A autora com dados de SINGH; ANSARI; RAGHAVA (2013).

Observa-se na Tabela 1 que SVM e k-NN obtiveram melhores desempenhos em dados de estruturas confirmadas experimentalmente (Lbtope_Confirm). Entretanto, quando treinados em dados não redundantes, ou seja, com identidade de sequência inferior a 80%, observa-se redução do desempenho de SVM de 16,38% de ACC, 0,31 de MCC e 0,19 de AUC-ROC para Lbtope_Fixed_non_redundant quando comparado ao conjunto de dados redundante (Lbtope_Fixed) e uma redução de 11,03% de ACC, 0,18 de MCC e 0,10 de AUC-ROC para Lbtope_Variable_non_redundant em comparação com Lbtope_Variable.

LBtope foi comparado com outras ferramentas de predição, como ABCpred, BCPred e AAP de Chen. Para uma comparação justa com ABCpred, foi utilizado o conjunto de dados de comprimento fixo (Lbtope_Fixed) e Lbtope treinado por ele, mas testado no conjunto de dados ABCpred. Neste teste, ABCpred apresentou maior Sp e AUC-ROC igual a LBtope, entretanto LBtope supera em Sn e ACC com diferença de 15,45% e 6,15%, respectivamente (SINGH; ANSARI; RAGHAVA, 2013).

O conjunto Lbtope_Fixed também foi utilizado para os testes com BCPred e AAP de Chen, para a comparação com BCPred foi utilizado o LBtope treinado com Lbtope_Fixed, Lbtope_Variable e Lbtope_Confirm no conjunto de dados do BCPred. Os resultados mostram que LBtope treinado com Lbtope_Fixed e Lbtope_Confirm apresentaram maior Sn ao custo de uma redução na Sp, ACC e MCC em comparação ao BCPred, o que não ocorre para LBtope treinado em Lbtope_Variable que apresenta maior Sn, ACC e MCC comparado ao BCPred (SINGH; ANSARI; RAGHAVA, 2013). Enquanto que para a comparação com o AAP de Chen foi utilizado apenas o LBtope treinado em Lbtope_Fixed no conjunto de dados de Chen, neste caso LBtope também apresentou aumento significativo de Sn, mas com redução de Sp e pouca diferença nos valores de ACC e MCC quando comparado ao AAP de Chen no conjunto de dados Lbtope_Fixed (SINGH; ANSARI; RAGHAVA, 2013).

Neste contexto, os desenvolvedores do LBtope chegaram à conclusão que o desempenho apresentado pode ser reflexo do fato de que os conjuntos de dados utilizados nos testes (ABCpred, BCPred e AAP de Chen) designam NE como estruturas aleatórias, potencialmente incluindo epítopos ainda não validados,

enquanto que LBtope é treinado com estruturas experimentalmente confirmadas tanto para ECB quanto para NE (SINGH; ANSARI; RAGHAVA, 2013). Entretanto, quando testado em um conjunto de dados com peptídeos aleatórios para NE, observou-se que LBtope obteve desempenho comparável com Lbtope_Confirm (SINGH; ANSARI; RAGHAVA, 2013).

Pode-se observar que LBtope apresenta desempenhos diferentes de acordo com o conjunto de dados analisados. Durante as rodadas de testes e com dados independentes apresentam valores para MCC, AUC-ROC, ACC, Sp e Sn diferentes de acordo com o conjunto de dados utilizado no treinamento (SINGH; ANSARI; RAGHAVA, 2013; SHER; ZHI, ZHANG, 2017). Esta falta de consistência dificulta a comparação entre os algoritmos e na própria avaliação do desempenho da ferramenta.

Apesar disso, o LBtope se apresenta como uma boa opção para predição de epítopos lineares, apresentando uma acurácia acima de 70% quando treinado com estruturas de comprimento variável (Lbtope_Variable) e acima de 80% com estruturas de comprimento fixo (Lbtope_Fixed) e estruturas confirmadas experimentalmente (Lbtope_Confirm) (SINGH; ANSARI; RAGHAVA, 2013). LBtope foi implementado em um servidor que disponibiliza todas as opções de treinamento, dando liberdade ao usuário escolher qual se adequa mais ao seu problema de pesquisa. Além da predição de epítopos, o servidor do LBtope também disponibiliza uma ferramenta para mutações, permitindo realizar mudanças nas estruturas dos epítopos (SINGH; ANSARI; RAGHAVA, 2013).

Assim como o LBtope, o iBCE-EL é um modelo *ensemble* (EM) que une diferentes algoritmos de predição por aprendizado de máquina (MANAVALAN et al., 2018). Para a formação do iBCE-EL foram testadas seis combinações (EM_{1-6}) entre os algoritmos SVM, k-NN, RF (*Random Forest*), ERT (*Extremely Randomized Tree*), AB (*AdaBoost*) e GB (*Gradient Boosting*). Todos os algoritmos foram treinados utilizando combinações de informações extraídas de 4.440 estruturas de ECB e 5.485 de não epítopos (NE) validados experimentalmente. Os principais parâmetros utilizados no treinamento foram: i) composição de aminoácidos (AAC) que representa a porcentagem de aminoácidos padrões nas sequências; ii) índice de aminoácidos (AAI) que apresenta as propriedades físico-químicas e bioquímicas dos

aminoácidos advindas do banco de dados AAIndex; iii) distribuição de transição de cadeia (CTD) que analisa a distribuição dos aminoácidos polares, neutros e hidrofóbicos para predizer os dobramentos das proteínas; iv) composição dipeptídica (DPC) que combina todos os 20 aminoácidos em dipeptídeos, gerando 400 variáveis; v) propriedades físico-químicas, como hidrofobicidade, carga, peso molecular e alifaticidade (MANAVALAN et al., 2018).

Dentre os EM formados, a combinação entre GB e ERT foi a selecionada para formar o iBCE-EL por apresentar valores mais elevados de MCC nos dados de treinamento e independentes (MANAVALAN et al., 2018). O iBCE-EL também apresentou melhor desempenho quando comparado aos algoritmos isolados e LBtope (Lbtope_Variable_non_redundant) no conjunto de dados independente. Neste experimento, o iBCE-EL superou o LBtope com uma diferença de 13,5% de MCC, 8% de ACC e 15,7% de especificidade, enquanto que os valores de AUC-ROC foram semelhantes com uma diferença de apenas 0,8% (MANAVALAN et al., 2018). Entretanto, após análise de informações posicionais e termos de características do algoritmo, não foram observadas diferenças entre iBCE-EL e LBtope (Lbtope_Variable_non_redundant) (MANAVALAN et al., 2018).

Dois conjuntos de dados do iBCE-EL foram utilizados para treinamento e teste do iLBE (*Identification of Linear B-cell Epitope*) que apresenta uma abordagem que integra as informações de características evolutivas e sequenciais dos candidatos a epítopos (HASAN; KHATUN; KURATA, 2020). As características evolutivas obtidas pelo iLBE advém da utilização dos valores de PSSM (*position-specific scoring matrix*), através da utilização do PSI-BLAST, e PKAF (*composition of profile-based amino acids frequency*) que é obtido após os valores de PSSM. Enquanto as características de sequência são obtidas de acordo com os valores de AFC (*amino acid frequency composition*) e 8 características físico-químicas e bioquímicas de aminoácidos obtidas do AIP (*amino-acid index property*). Por fim, os dados de cada um dos 4 recursos são otimizados por WR (*non-parametric Wilcoxon-rank sum*) de acordo com o grau de importância para predição do epítopo e a sequência do epítopo; os recursos são escolhidos por RF e combinados com LR (HASAN; KHATUN; KURATA, 2020).

O iLBE foi comparado com seus quatro recursos no conjunto de dados de treinamento, e demonstrou um desempenho superior em todos os parâmetros de análise. O iLBE apresentou uma diferença de 4,4% de Sp, 2% de Sn, 3,3% de ACC, 6,4% de MCC e 3,4% de AUC-ROC em relação ao AFC, considerado o que teve melhor desempenho dentre os comparados (HASAN; KHATUN; KURATA, 2020).

Neste sentido, outro teste também foi realizado no conjunto de dados independente e, novamente, o iLBE mostrou um desempenho superior. Resultado similar ao obtido na comparação com as ferramentas LBtope e iBCE-EL, tanto para o conjunto de dados de treinamento, quanto para o conjunto de dados independentes, onde o iLBE utilizando limite moderado supera esses dois algoritmos em Sp, ACC, MCC e AUC-ROC nos dois conjuntos de dados. Entretanto, para o conjunto de dados independente, o iLBE apresenta valores semelhantes com LBtope em relação a sensibilidade, mesmo LBtope apresentando desempenho inferior em todos os índices analisados. Além disso, quando a especificidade é aumentada, o desempenho do iLBE sofre uma redução drástica de sensibilidade frente ao iBCE-EL, LBtope e iLBE com limite moderado nos dois conjuntos de dados e os valores de ACC e MCC obtidos são inferiores ao iBCE-EL no conjunto de dados independentes (HASAN; KHATUN; KURATA, 2020).

Outro algoritmo que utiliza RF é o BepiPred 2.0 que utiliza 46 variáveis que se distribuem em características como hidrofobicidade, volume calculado, polaridade, acessibilidade relativa da superfície, estrutura secundária e volume total do antígeno (JESPERSEN et al., 2017). Foram utilizadas 155 estruturas para o treinamento e 5 para os testes, todas extraídas de complexos Ag-Ac disponíveis no PDB (*Protein Data Bank*). Os epítopos foram identificados como aqueles aminoácidos não pertencentes ao anticorpo e que estavam a uma distância de 4Å de raio das CDR (Regiões Determinantes de Complementariedade) do anticorpo (JESPERSEN et al., 2017), resultando na linearização de epítopos conformacionais, ou seja, são identificados os AAs dos epítopos conformacionais e estes são unidos em sequência (RAS-CARMONA et al., 2021).

Seu desempenho de predição foi comparado ao LBtope, BCPred, CBTope e BepiPred 1.0 em cinco estruturas extraídas do PDB, os resultados das médias de AUC-ROC mostram que mesmo com o CBTope apresentando resultado próximo, o

BepiPred 2.0 teve melhor desempenho que todas as quatro ferramentas (JESPERSEN et al., 2017). Não obstante, outro teste foi realizado com o BepiPred 1.0, desta vez utilizando epítopos lineares extraídos do IEDB, sendo 11.839 validados como ECBL e 18.722 validados como não epítopos (NE). Enquanto o BepiPred 2,0 apresentou AUC-ROC de 0,574, sua versão anterior apresentou 0,548 (JESPERSEN et al., 2017), reforçando a ideia que o treinamento a partir de estruturas cristalográficas melhorou o desempenho do algoritmo de predição desta ferramenta.

Como relatado anteriormente, de forma geral, o desempenho do BepiPred 2.0 foi superior que o desempenho obtido pelas demais ferramentas preditoras de epítopos que foram utilizadas na comparação. Entretanto, analisando dados individuais, percebe-se que o CBTope apresentou maiores valores de AUC-ROC para três das cinco estruturas testadas, sendo elas: 0,920 para 4WFF, 0,600 para 5BVP e 0,560 para 5C0N, em comparação a BepiPred 2.0 que alcançou 0,738, 0,569 e 0,473, respectivamente (JESPERSEN et al., 2017). Entretanto, CBTope apresentou valores muito reduzidos de AUC-ROC para as outras duas estruturas, o que resultou em redução de desempenho na média final (JESPERSEN et al., 2017). Além disso, poucas estruturas foram utilizadas no conjunto de testes para o comparativo com outras ferramentas e, diante disso, é válido ressaltar a necessidade da realização de testes adicionais utilizando uma quantidade maior de estruturas (amostras).

Pode-se observar que a maioria das ferramentas que foram revisadas neste trabalho utilizam o banco de dados IEDB como fonte para a obtenção de estruturas para a formação dos conjuntos de dados, inclusive para o conjunto de treinamento do algoritmo, onde são incluídas diversas espécies e epítopos de diversos tipos de anticorpos, ampliando suas estruturas de treinamento e sua aplicabilidade. Entretanto, o algoritmo RF-OrgSpec propõe que treinar o algoritmo com estruturas de determinadas espécies pode ser mais vantajoso, pois aumenta a qualidade dos dados de amostras e, consequentemente, a predição de epítopos (ASHFORD et al., 2021).

As espécies alvo para o treinamento do RF-OrgSpec foram *Onchocerca volvulus*, vírus Epstein-Barr e o vírus da hepatite C, pois possuem grande volume de

dados validados no IEDB (ASHFORD et al., 2021). Uma das limitações desta estratégia de predição direcionada está diretamente relacionada ao fato de que, em casos de espécies filogeneticamente distintas das espécies previamente treinadas, será necessário um grande volume de dados referentes a espécie em estudo para treinar o algoritmo, o que nem sempre é possível de ser feito para todos os organismos, resultando em uma possível redução do desempenho da predição. Nestes casos, ferramentas mais generalistas como, por exemplo, LBtope e BepiPred 2.0, podem potencialmente predizer epítopos de forma mais acurada.

Os conjuntos de dados específicos dos organismos (OrgSpec) foram representados em uma janela deslizante de 15 AA, onde foram calculadas 845 características para cada AA que incluíam combinação de dipeptídeos, propriedades físico-químicas dos AA, peso molecular, entre outros. Estes dados foram utilizados no treinamento de predição de ECB em RF (ASHFORD et al., 2021).

O desempenho foi comparado em um conjunto de dados independente, onde o RF que foi treinado com dados específicos dos organismos (RF-OrgSpec) apresentou melhor desempenho para os três patógenos quando comparado ao RF-Heterogêneo (treinado com dados sem informações do patógeno relacionado) e o RF-Híbrido (treinado com os mesmos dados do grupo Heterogêneo e OrgSpec), além de ferramentas de predição amplas como SVMtrip, LBtope, iBCE-EL, BepiPred 2.0 e ABCpred (ASHFORD et al., 2021).

Para o conjunto de dados do vírus da hepatite C, o LBtope apresentou desempenho semelhante ao algoritmo RF-OrgSpec nos seis parâmetros utilizados, sendo eles: ACC, AUC-ROC, MCC, Sn, PPV (Valor Preditivo Positivo) e NPV (Valor Preditivo Negativo). Este resultado pode estar relacionado ao fato das estruturas deste vírus estarem presentes no conjunto de dados de treinamento do LBtope (ASHFORD et al., 2021). Além disso, o LBtope apresentou valores superiores de PPV para os epítopos do vírus Epstein-Barr. As ferramentas SVMtrip, BepiPred 2.0, iBCE-EL e ABCpred apresentaram desempenho inferior frente ao RF-OrgSpec. Entretanto, no conjunto de dados de *O. volvulus*, o BepiPred 2.0 apresentou um desempenho melhor no quesito sensibilidade (ASHFORD et al., 2021).

Portanto, o RF-OrgSpec demonstra que a utilização de um conjunto de dados mais direcionado para o patógeno em estudo pode melhorar a predição dos epítopos para este organismo. No entanto, como afirmado anteriormente, esta predição está interligada a um maior número de estruturas disponíveis em banco de dados, o que não se aplica a todos os organismos.

Epítopos de IgG correspondem a grande maioria de epítopos em bancos de dados (TUNG et al., 2021). Entretanto, existem outros 4 tipos de imunoglobulinas (IgA, IgD, IgE e IgM) que diferem por suas estruturas e propriedades (MESQUITA JÚNIOR et al., 2010) (Quadro 4), mas que também apresentam papéis importantes na defesa imunológica, como combate a bactérias, parasitas e vírus (TUNG et al., 2021).

Quadro 4: Características básicas das imunoglobulinas.

Classe	**Estrutura**	**Propriedades**
IgA	Dimérica ou monomérica	Encontrada em mucosas do trato gastrointestinal, respiratório e urogenital. Previne a colonização por patógenos. Presente também na saliva, lágrimas e leite.
IgD	Monomérica	Imunoglobulina de membrana. Faz parte do receptor de membrana dos linfócitos B virgens (BCR).
IgE	Monomérica	Envolvida em processos alérgicos e parasitários. Sua interação com basófilos e mastócitos causa liberação de histamina.
IgG	Monomérica	Principal imunoglobulina da imunidade adquirida. Tem capacidade de atravessar a barreira placentária.
IgM	Monomérica ou pentamérica	Faz parte do receptor de membrana dos linfócitos B virgens (BCR). Forma encontrada no soro, secretada precocemente na resposta imune adquirida.

Fonte: MESQUITA JÚNIOR et al. (2010).

Considerando a importância destas outras classes de imunoglobulinas (Ig), desenvolveu-se a ferramenta NIgPred, um preditor de ECBL para IgA, IgE e IgG (TUNG et al., 2021). O NIgPred utiliza o algoritmo de Otimização Sequencial Mínima

(SMO) e o algoritmo *hill-climbing* para a predição dos epítopos de IgA. O algoritmo *hill-climbing* está associado ao RF para predição dos epítopos de IgG, enquanto que para IgE foram utilizados o FLDA e o algoritmo genético (TUNG et al., 2021). O conjunto de amostras para as 3 classes de Igs foram selecionados do IEDB, dividindo-as em 4 grupos: amostras de IgA (631), amostras de IgE (2.435), amostras de IgG (49.093) e amostras de não epítopos (127.010) para controle negativo, onde destes quantitativos 80% foi utilizado para treinamento e o restante para a rodada de teste (TUNG et al., 2021).

A ferramenta NIgPred difere de outras abordagens semelhantes como, por exemplo, IgPred e BCIgEPred, pois utiliza duas características adicionais na classificação de epítopos: autocorrelação e recursos de correlação de características de peptídeo, além das características comumente utilizadas de composição e características físico-químicas (TUNG et al., 2021). Após a comparação dos resultados utilizando estas características tanto de forma separada quanto em conjunto, observou-se que os autores do NIgPred acertaram em suas escolhas, pois o conjunto dos 4 recursos utilizados tiveram os melhores valores de MCC para os 3 tipos de Ig: 70,1% para IgA, 67,9% para IgE e 91,1% para IgG (TUNG et al., 2021).

Os resultados obtidos para o conjunto de testes apresentaram valores de precisão e ACC superiores a 80% para os 3 tipos de Igs. Além disso, o conjunto de testes de IgE foi submetido ao BCIgEPred para comparação, onde este apresentou desempenho inferior comparado ao NIgPred com valores de precisão de 68,6%, ACC de 56,9% e MCC de 18,7% frente aos valores do NIgPred de 82%, 81,9% e 64%, respectivamente (TUNG et al., 2021). Em outra análise com 429 estruturas de epítopos de IgE extraídas dos bancos de dados *Allergen for Food Safety* (ADFS) e *Allerbase*, o NIgPred apresentou valores mais elevados de precisão quando comparado com as ferramentas IgPred e BCIgEPred (TUNG et al., 2021). No conjunto de 410 amostras extraídas do ADFS, o NIgPred apresentou melhor desempenho com diferença de 32% de precisão para o BCIgEPred e 74% para o IgPred. Adicionalmente, para as amostras do Allerbase, a diferença de precisão passou para 49% comparado ao BCIgEPred e 28% para o IgPred (TUNG et al., 2021).

O NIgPred traz uma nova abordagem para o problema da predição de epítopos, além de mostrar que é possível utilizar ferramentas *in silico* para a predição de epítopos de outras classes de Igs. Apesar desta abordagem já ser utilizada em outros preditores, o NIgPred apresentou um desempenho melhor na predição de epítopos voltados aos IgEs, como abordado anteriormente. No entanto, ainda é necessário um maior desenvolvimento na área de predição de epítopos voltados aos IgAs para uma melhor análise do desempenho desta ferramenta. Além disso, o trabalho deixa a desejar quando não compara sua ferramenta com outros preditores de epítopos de IgG que, ao contrário das outras classes de Igs, é bem explorada.

Como pode ser observado, as ferramentas anteriores trabalham com diferentes abordagens, principalmente, voltadas às informações referentes apenas a interação do epítopo com o anticorpo. Entretanto, deixam de lado informações de um outro aliado da resposta imune: os linfócitos T.

Para produzir uma resposta imune, o antígeno deve primeiro ser apresentado para as células efetoras da imunidade adaptativa (linfócitos T e B) pelo HLA (Antígenos Leucocitários Humanos), o qual se apresenta em duas classes: I e II. O HLA de classe I está presente em todas as células nucleadas do organismo e tem a finalidade de apresentar antígenos endógenos para os linfócitos T $CD8^+$, com função citotóxica. Enquanto isto, o HLA de classe II está presente apenas nas APCs (Células Apresentadoras de Antígenos), como os linfócitos B, macrófagos e células dendríticas, e tem como função a apresentação de antígenos exógenos para os linfócitos T $CD4^+$, que também são denominados linfócitos T auxiliares (CRUVINEL et al., 2010; RAOUFI et al., 2019). Os linfócitos T auxiliares ativados liberam citocinas que ativam linfócitos B para liberação de anticorpos, em um processo conhecido como reciprocidade T-B (TOMAR; DE, 2010; ZHU et al., 2022). Portanto, os ECBs que também possuem afinidade ao HLA II têm maiores chances de serem imunogênicos e, consequentemente, podem ser usados para desenvolvimento de vacinas (ZHU et al., 2022).

Diante da importância da reciprocidade T-B, ainda pouco compreendida, foi desenvolvida a ferramenta BepiTBR, *software* de predição de ECB que analisa a afinidade do epítopo para células B e HLA classe II (ZHU et al., 2022). Em resumo, o

BepiTBR prediz possíveis epítopos imunogênicos a partir das pontuações obtidas dos três preditores de ECB (BepiPred 1.0, BepiPred 2.0 e LBEEP) e associa com os preditores de epítopos de HLA (netMHCIIpan e mixMHC2pred). Além disto, utiliza-se o *Elastic-Net framework* para designar duas penalidades aos dados (L1 e L2): L1 que declara pontos de corte entre 0,01 a 2 de ligantes de HLA, separando os mais importantes para a predição, e L2 para redução da correlação dos ligantes dos alelos DRB, DP e DQ do HLA de classe II (ZHU et al., 2022).

Os autores relatam que uma das limitações do método é não ter levado em consideração a diversidade alélica de HLA entre os indivíduos, por conta das limitações dos próprios bancos de dados que ainda não conseguem abarcar toda esta diversidade (ZHU et al., 2022). Mesmo com esta limitação, quando comparado às suas ferramentas isoladas e estas associadas em pares, o BepiTBR se mostrou superior, demonstrando que a reciprocidade T-B é uma informação importante para desenvolvimento de epítopos potencialmente imunogênicos (ZHU et al., 2022).

A ferramenta BCEPS (*B Cell Epitope Prediction Software*) conseguiu incluir em sua metodologia parte desta diversidade alélica do HLA no seu desenvolvimento, no intuito de identificar a cobertura de proteção da população (PPC), valor que indica a imunogenicidade do epítopo frente a população (RAS-CARMONA et al., 2021). O PPC é calculado a partir dos índices de ligação ao alelo HLA-DRB1 para caucasianos, afro-americanos, asiáticos e nativos norte-americanos, obtidos no *Allele Frequency Net Database* (RAS-CARMONA et al., 2021). As ligações com o HLA de classe II foram previstas pelo RANKPEP e netMHCIIpan, este último também foi utilizado pelo BepiTBR (RAS-CARMONA et al., 2021).

Assim como o BepiPred 2.0, o BCEPS utiliza estruturas cristalográficas de complexos Ag-Ac, considerando os aminoácidos que se encontravam em um raio de 4Å do anticorpo como a estrutura dos epítopos (RAS-CARMONA et al., 2021). Diferente das demais ferramentas, os autores do BCEPS utilizaram o CD-HIT para que a identidade da sequência de aminoácidos fosse inferior a 80% (RAS-CARMONA et al., 2021), enquanto em sua maioria, as ferramentas utilizam identidade de sequência inferior a 70%. Além da predição de epítopos e imunogenicidade, o BCEPs também disponibiliza em seu servidor informações

acerca da hidrofilicidade, flexibilidade, acessibilidade, localização do ectodomínio e sítios de glicosilação do epítopo (RAS-CARMONA et al., 2021).

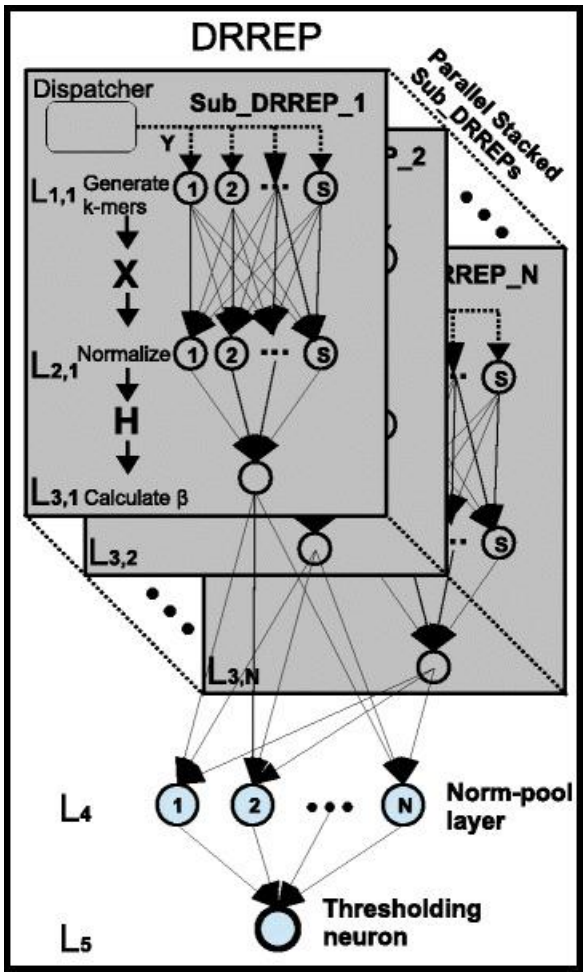
Foram avaliados diversos algoritmos de aprendizado de máquina contidos no pacote WEKA (*Waikato Environment for Knowledge Analysis*), como redes neurais, K-NN, RF e SVM com Kernel de função de base radial gaussiana (RBF) e função polinomial (RBF-SVM). Estes algoritmos foram treinados por vetores de composição de aminoácidos, composição de dipeptídeos e ambos (RAS-CARMONA et al., 2021). Os três algoritmos foram mais precisos utilizando os vetores de composição de aminoácidos, dentre eles, o RBF-SVM obteve um melhor desempenho em relação a ACC, MCC e sensibilidade, sendo este algoritmo escolhido para implementação no BCEPS (RAS-CARMONA et al., 2021).

A partir disso, o BCEPS foi testado com o BepiPred 2.0, iBCE-EL e LBtope em dados independentes obtidos do IEDB contendo ECBL, ECBC linearizados e não epítopos (RAS-CARMONA et al., 2021). O BepiPred 2.0, iBCE-EL e LBtope não conseguiram atingir 50% de ACC para os EBCL, onde o iBCE-EL obteve ACC de 48,15% frente aos 60,38% obtido pelo BCEPS (RAS-CARMONA et al., 2021). Em relação aos ECBCs, o LBtope atingiu apenas 56,8% de ACC frente aos 67,05% do BCEPS (RAS-CARMONA et al., 2021). O aumento de desempenho do BCEPS durante os testes de ECBC comparação aos EBCL pode estar relacionado à semelhança dos dados de treinamento, uma vez que foram utilizados epítopos conformacionais linearizados (RAS-CARMONA et al., 2021).

As ferramentas até aqui descritas utilizam aprendizado de máquina para realizar a predição de epítopos, mas outras abordagens também podem ser utilizadas para esta função como a rede neural profunda (em inglês, DNN) utilizada no DRREP (*Deep Ridge Regressed Epitope Predictor*) e no DLBEpitope.

O DRREP utiliza incompatibilidade de *strings* para identificar os resíduos de aminoácidos pertencentes ao epítopo. Nesta ferramenta, diversas subcamadas (Sub_DRREP) funcionam em paralelo contendo 3 camadas de treinamento independentes e 2 camadas que utilizam os dados obtidos para a predição (SHER; ZHI; ZHANG, 2017). O funcionamento do DRREP é ilustrado na figura 3.

Figura 3 – Etapas de predição do DRREP



Fonte: SHER; ZHI; ZHANG (2017).

Inicialmente, o DRREP utiliza a sequência total do alvo como entrada, esta sequência é dividida em janelas deslizantes de comprimento Y definido pelo usuário durante a fase de treinamento (SHER; ZHI; ZHANG, 2017). Posteriormente, estas subsequências Y são submetidas a primeira camada (L1) que utiliza um número aleatório de funções básicas de ativação de incompatibilidade (S) em um k-mer de tamanho aleatório com base entre 1 o tamanho da janela deslizante, em L1 são calculadas quantas vezes o k-mer ocorreu na sequência (SHER; ZHI; ZHANG, 2017). Esse resultado é passado para a segunda camada (L2) que calcula a proporção de cada k-mer encontrado na subsequência (SHER; ZHI; ZHANG, 2017). A terceira camada (L3) é a camada de aprendizado e utiliza os dados resultantes da L2 para classificar se aquela subsequência é ou não um epítopo, a partir do uso do inverso generalizado de Moore-Penrose (SHER; ZHI; ZHANG, 2017). A quarta

camada (L4) une os resultados de cada Sub_DRREP e envia para a quinta e última camada (L5) que utiliza a pontuação de validação relativa de cada Sub_DRREP para a pontuação final, ou seja, identifica o epítopo linear a partir dos resíduos sequenciais que apresentaram maior pontuação (SHER; ZHI; ZHANG, 2017).

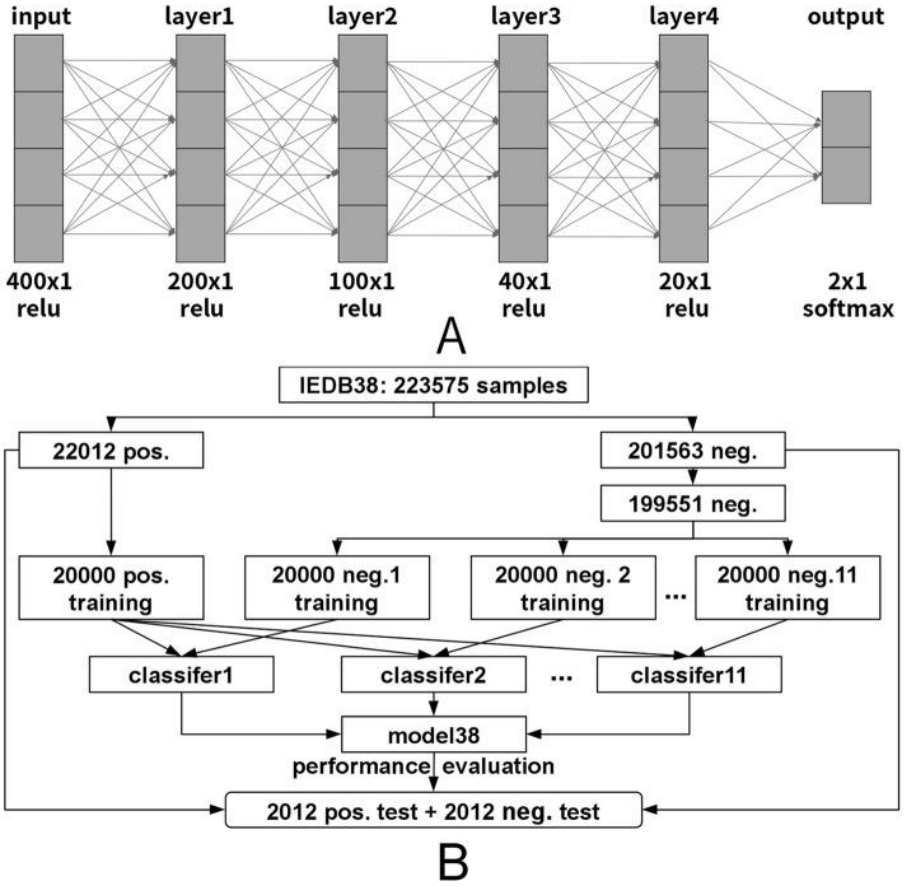
O DRREP foi treinado com o conjunto de dados BciPep, um conjunto cuidadosamente curado contendo 701 estruturas possuindo identidade de sequência inferior a 80% e epítopos contendo 20 AA. Este conjunto também foi utilizado durante os testes realizados em outros preditores como, por exemplo, os preditores BCPred e BEST. O conjunto de dados negativo (não epítopos), foi formado através da seleção de 701 peptídeos aleatórios no SwissProt (SHER; ZHI; ZHANG, 2017). Cinco conjuntos de dados que foram utilizados nos testes de comparação de desempenho de outros preditores foram utilizados, possuindo diferentes densidades de epítopos, são eles: i) SEQ194: é o que contém o maior número de estruturas (128.180) disponíveis quando comparado com os demais conjuntos escolhidos. Este conjunto foi utilizado pelo preditor BEST e a densidade de seus epítopos é de 6,6%; ii) AntiJen: assim como SEQ194, este conjunto também possui epítopos com baixa densidade (1,4%) quando comparado com os demais conjuntos de dados. Este conjunto foi utilizado pelo preditor BepiPred 1.0; iii) Pellequer: também utilizado pelo BepiPred 1.0, este conjunto possui uma densidade de epítopos mais elevada que os conjuntos anteriores (37,6%); iv) HIV: possui 2.706 epítopos e uma densidade de 37,1% e iv) SARS: conjunto com 193 resíduos de alta densidade (63,7%) que foi utilizado para o desenvolvimento do BCPred (SHER; ZHI; ZHANG, 2017).

No conjunto de dados SEQ194, o DRREP foi a ferramenta de predição que apresentou o maior valor de AUC-ROC (0,732) quando comparado com as demais ferramentas analisadas, seguido pelas ferramentas LBtope (0,71) e Epitopia (0,59) (SHER; ZHI; ZHANG, 2017). Observa-se que a AUC-ROC obtido para o DRREP difere apenas em 0,022 do valor de AUC-ROC obtido para o LBtope, mas nota-se que esta diferença aumenta para 0,142 quando comparado a terceira ferramenta com melhor desempenho (Epitopia). A ferramenta LBtope aparece na segunda colocação para os demais conjuntos de dados, atrás somente do DRREP, excetuando-se o teste realizado com o conjunto AntiJen, onde as duas ferramentas apresentam o mesmo valor de AUC-ROC (0,702), com o LBtope superando o

DRREP em 1,2% na ACC para uma especificidade fixa de 75%. Este mesmo valor é encontrado para DMN-LBE no conjunto de dados de HIV, com o DRREP superando o valor AUC-ROC com uma diferença de 0,053. Por fim, o DRREP obteve valores mais elevados de AUC-ROC para os testes realizados no conjunto SARS, onde apresentou ACC de 86% e AUC-ROC de 0,862, enquanto BCPred apresentou segunda melhor ACC com 80,3% e LBtopo segunda melhor AUC-ROC de 0,758. Portanto, observa-se que para os cinco conjuntos de dados analisados, o DRREP apresentou o maior valor de AUC-ROC dentre as ferramentas analisadas, demonstrando, assim, uma consistência de resultados mesmo com estruturas de densidades variadas.

Assim como o DRREP, o preditor DLBEpitope também utiliza algoritmos de rede neural profunda, especificamente uma rede neural profunda *feedforward* construído a partir de pacote R ”keras” (LIU; SHI; LI, 2020). A composição dipeptídica foi utilizada como recurso para predição, estratégia similar a empregada pelo preditor iBCE-EL. Este recurso gera 400 vetores para serem utilizados como dados de entrada (*input*) para a predição do DLBEpitope e, a partir disso, segue mais quatro camadas com 200, 100, 40 e 20 unidades que resultando como saída (*output*) 2 unidades: epítopo ou não epítopo (LIU; SHI; LI, 2020). A arquitetura do DLBEpitope está ilustrada na Figura 4A.

Figura 4 – Arquitetura da rede neural profunda *feedforward* do DLBEpitope (A) e formação dos conjuntos de dados de treinamento e teste para IEDB38 (B).



Fonte: LIU; SHI, LI (2020).

O conjunto de dados de treinamento contém 25.884 estruturas de ECB e 214.679 de NE confirmadas experimentalmente e extraídas do IEDB (LIU; SHI; LI, 2020). A partir de técnicas de truncamento e extensão, estas estruturas foram divididas em diversos grupos de acordo com seu comprimento entre 11 e 50 resíduos (IEDBx; X=11, 12, ..., 50) (LIU; SHI; LI, 2020). Os números de amostras positivas e negativas são desiguais e, portanto, houve a necessidade de balancear as amostras. Neste sentido, foram selecionadas 20.000 amostras aleatórias positivas e negativas para cada IEDBx e, como o número de dados negativos é maior, foram formados 11 subconjuntos com 20.000 amostras negativas que foram utilizadas para treinamento com um conjunto também de 20.000 amostras positivas (Figura 4B) (LIU; SHI; LI, 2020). Estes subconjuntos negativos e positivos foram utilizados para treinamento da rede neural e formaram 11 classificadores que são usados para indicar se o peptídeo em análise é ou não um epítopo, de acordo com cada comprimento de sequência (Figura 4B) (LIU; SHI; LI, 2020).

As estruturas que sobraram após a retirada das amostras de treinamento, foram utilizadas na comparação com outras ferramentas (LIU; SHI; LI, 2020). Para uma comparação justa com as demais ferramentas, foram selecionados os conjuntos de dados IEDB16, IEDB18 e IEDB20 para ABCpred e APCpred, sendo este último também utilizado para comparação com AAP. Os conjuntos IEDB22, IEDB31, IEDB38, além dos anteriormente citados, foram utilizados para comparar o desempenho com a ferramenta BepiPred nas versões 1.0 e 2.0 (LIU; SHI; LI, 2020). O preditor DLBEpitope apresentou valores de AUC-ROC mais elevados para todos os conjuntos de IEDBx. A título de exemplo, o BepiPred 2.0 apresentou valores de AUC-ROC em um intervalo entre 51,26% a 54,02, enquanto o DLBEpitope apresentou valores de AUC-ROC no intervalo entre 88,8% a 95,68%, resultando em uma diferença de 39,93 no conjunto IEDB31, conjunto no qual ambos apresentaram melhor desempenho (LIU; SHI; LI, 2020).

Outros testes de desempenho foram realizados com o conjunto Lbtope_Fixed, estruturas com comprimento fixo de 20 AAs contendo 8.661 amostras de ECB e 16.492 amostras NE após remoção das estruturas que se encontravam também no conjunto de dados de treinamento (IEDB20). O conjunto de dados do ABCpred, também foi utilizado no teste, contendo amostras de comprimento de 16 AAs, resultando em 107 amostras de ECB e 196 NE após remoção de estruturas que participavam do IEDB16 (LIU; SHI; LI, 2020). Neste cenário, o DLBEpitope apresentou valores de AUC-ROC mais elevados nos dois conjuntos de dados. Levando em consideração o conjunto Lbtope_Fixed modificado, o DLBEpitope obteve 73,83% de AUC-ROC, enquanto que o BepiPred 1.0 apresentou 55,79%. No conjunto de dados ABCpred modificado, observou-se que o DLBEpitope obteve 65,5% de AUC-ROC, resultado similar ao obtido pelo BepiPred 1.0 (61,88%). Estes resultados podem estar relacionados a menor quantidade de estruturas disponíveis neste conjunto (ABCpred modificado).

Por fim, observa-se que as ferramentas que utilizam algoritmos de redes neurais apresentaram desempenho próximos ou, por vezes, superior a muitas das ferramentas que utilizam algoritmos de aprendizado de máquina, indicando que este método também é uma abordagem útil e relevante para predição de epítopos de células B.

Diante das diversas ferramentas apresentadas, levando em consideração suas respectivas abordagens e inovações, observa-se no Quadro 5 um resumo das vantagens e desvantagens encontradas para cada ferramentas revisadas. Este quadro tem como objetivo facilitar a escolha da ferramenta mais apropriada para cada tipo de pesquisa realizada pela comunidade científica.

Quadro 5 - Vantagens e desvantagens dos algoritmos de predição de epítopos lineares em ordem alfabética (continua).

Algoritmos	Vantagens	Desvantagens
BCEPS (RAS--CARMONA et al., 2021)	○ Treinado em estruturas cristalográficas. ○ Predição de epítopos conformacionais linearizados ou lineares. ○ Informa sobre hidrofilicidade, flexibilidade, acessibilidade, localização do ectodomínio e sítios de glicosilação do epítopo. ○ Promove a predição de epítopos mais imunogênicos, com informações sobre a interação destes com o HLA de classe II e sua diversidade alélica. ○ Disponível em servidor *on-line*.	○ Apresentou desempenho um pouco inferior para predição de epítopos lineares frente a epítopos conformacionais linearizados.
BepiPred 2.0 (JESPERSEN et al., 2017)	○ Treinado em estruturas cristalográficas extraídas do banco de dados PDB. ○ Disponível em servidor *on-line*.	○ A comparação com outras ferramentas usa poucas estruturas; ○ Apresenta desempenho semelhante a BepiPred 1.0.
BepiTBR (ZHU et al., 2022)	○ Leva em consideração a reciprocidade T-B; ○ Predição de possíveis epítopos mais imunogênicos. ○ Disponível em servidor *on-line*.	○ Não considerou a diversidade alélica entre os indivíduos. ○ Não é direcionado para a interação Ag-Ac e por isso pode não ser o mais apropriado para engenharia de anticorpos.

Quadro 5 - Vantagens e desvantagens dos algoritmos de predição de epítopos lineares em ordem alfabética (continua).

Algoritmos	Vantagens	Desvantagens
DLBEpitope (LIU; SHI; LI, 2020)	○ Utiliza composição dipeptídica em rede neural profunda *feedforward*. ○ Disponível em servidor *on-line*.	○ Comparado ao DRREP, que também utiliza rede neural profunda, o DLBEpitope não pode ser usado para análise de uma sequência inteira de proteína, mas apenas para peptídeos de comprimento fixo.
DRREP (SHER; ZHI, ZHANG, 2017)	○ Utiliza incompatibilidade de *strings* em uma rede neural profunda (DNN) com aprendizado rápido em uma única etapa. ○ Pode ser aplicado a uma sequência sem comprimento definido.	○ Disponível apenas como *stand-alone*, o que pode dificultar sua utilização para usuários iniciantes, no primeiro contato com a biologia computacional.
iBCE-EL (MANAVALAN et al., 2018)	○ Modelo *ensemble* que une ERT e GB. ○ Disponível em servidor *on-line*.	○ Desempenho semelhante ao LBtope (Lbtope_Variable_non ○ _redundant)
iLBE (HASAN; KHATUN; KURATA, 2020)	○ Utiliza características evolutivas e de sequência para predição. ○ Disponível em servidor *on-line*.	○ Desconsidera epítopos com menos de 24 AA. ○ Além de redução de desempenho, após aumento da especificidade.
LBtope (SINGH; ANSARI; RAGHAVA, 2013)	○ Utiliza SVM e k-NN com diversas opções de treinamento, incluindo estruturas confirmadas experimentalmente. ○ Disponibiliza no seu servidor uma ferramenta para mutações no epítopo.	○ Apresenta desempenho reduzido para não epítopos em alguns conjuntos de dados.

Quadro 5 - Vantagens e desvantagens dos algoritmos de predição de epítopos lineares em ordem alfabética (conclusão).

Algoritmos	Vantagens	Desvantagens
RF-OrgSpec (ASHFORD et al., 2021)	○ Melhor predição de epítopos de organismos específicos, como *Onchocerca volvulus*, vírus Epstein-Barr e o vírus da hepatite C.	○ A aplicabilidade pode ser reduzida para espécies diferentes das utilizadas para treinamento e/ou espécies pouco exploradas na literatura, pois há dificuldade na formação de conjunto de dados de treinamento. Nestes casos, as ferramentas mais generalistas podem apresentar uma melhor performance. ○ Disponível apenas como *stand-alone*, o que pode dificultar sua utilização para usuários iniciantes.
NIgPred (TUNG et al., 2021)	○ Prediz epítopos para IgA, IgE e IgG, através de autocorrelação, características de correlação característica de peptídeo, composição e características físico-químicas. ○ Apresenta melhor desempenho na predição de epítopos de IgE quando comparada a outras ferramentas já apresentadas na literatura. ○ Disponível em servidor *on-line*.	○ Poderia ter apresentado mais comparações com outras ferramentas de predição de epítopos de IgG.

Fonte: A autora (2022).

5.1.2 Algoritmos para Predição de Epítopos Conformacionais

Os vírus possuem regiões conservadas entre diferentes espécies virais, como, por exemplo, as glicoproteínas do envelope para os vírus envelopados. Esta característica permite o desenvolvimento de anticorpos atuantes contra mais de um agente infeccioso, sendo denominados bNAbs (anticorpos neutralizantes virais) (MONTIEL-GARCIA et al., 2022; SILVA, 2019). Tomando por base essa

característica, foi desenvolvido o Epitope-Analyser (EA) que analisa as semelhanças estruturais entre um antígeno com afinidade por um bNAb e um antígeno em estudo, indicando se este último também pode se tornar alvo do mesmo anticorpo.

O preditor EA utiliza duas estruturas como dados de entrada (*input*): i) PDB-1 que corresponde ao complexo Ag-Ac de referência e ii) PDB-2 que corresponde ao antígeno alvo que será comparado ao antígeno do PDB-1 (MONTIEL-GARCIA et al., 2022). A primeira etapa do EA busca identificar os aminoácidos pertencentes ao parátopo e epítopo da estrutura PDB-1. Posteriormente são identificados os resíduos conservados do epítopo de PDB-1, através da relação de semelhança da sequência e conformação estrutural com o epítopo do PDB-2, obtidos a partir de *docking* rígido deste antígeno com o anticorpo de PDB-1 (MONTIEL-GARCIA et al., 2022). Para a etapa de análise de semelhanças, primeiro é realizada a superposição dos dois antígenos (PDB-1 e PDB-2) através do algoritmo Kpax, identificando os resíduos que pertencem aos mesmos grupos em termos de propriedades físico-químicas como, por exemplo, resíduos pouco ou muito hidrofóbicos, resíduos carregados positivamente ou negativamente, resíduos muito ou pouco polares e resíduos aromáticos. Após mapear as propriedades físico-químicas destes resíduos, é feito um cálculo do RMSD entre os Cα dos epítopos, identificando semelhanças conformacionais (MONTIEL-GARCIA et al., 2022). O termo em inglês RMSD (Root Mean Square Deviation) refere-se à raiz do desvio quadrático médio, uma métrica comumente utilizada em modelagem molecular, na avaliação comparativa entre duas conformações diferentes. Quanto menor for o valor obtido no cálculo de RMSD, mais próximas e sobrepostas estarão as estruturas comparadas.

Para analisar o desempenho do EA, foi realizado um estudo de caso com seis bNAbs anti hemaglutinina de diferentes cepas do vírus da gripe do grupo 1 e 2, dentre estes, quatro complexos pertenciam ao mesmo bNAb (CR6261) (MONTIEL-GARCIA et al., 2022). Os resultados obtidos para os diferentes epítopos de CR6261 apresentaram alta semelhança de sequência e baixos valores de RMSD, confirmando que o bNAb possuí afinidade pelo antígeno analisado devido a semelhança estrutural (MONTIEL-GARCIA et al., 2022). Vale ressaltar que quando a estrutura PDB-1 é colocada na posição de PDB-2 e vice-versa, os valores de semelhança de sequência e RMSD podem sofrer certa variação, uma vez que o

epítopo de referência foi modificado. Além disso, como era de se esperar, os antígenos do mesmo grupo do vírus da gripe apresentaram mais semelhanças (RMSD: 0,38 a 1,09; identidade de sequência: 83,4 a 94,7%) entre si, do que com os epítopos de outro grupo (RMSD: 2,12 a 2,54; identidade de sequência: 37 a 63,2%) (MONTIEL-GARCIA et al., 2022).

Neste sentido, os resultados iniciais demostraram o grande potencial de utilização do Epitope-Analyser em indicar se determinado antígeno possui afinidade por determinado alvo bNAb já conhecido e, com isto, aumentar a aplicabilidade deste anticorpo. Diante dos resultados obtidos, estima-se que o preditor EA também possa ser utilizado como uma opção de ferramenta durante o processo de *redesign* de anticorpos, tópico que será discutido na seção 5.2.2 desta monografia, durante a fase identificação de anticorpos que apresentam complementariedade parcial com outros antígenos, selecionando estes Ac para futuras mutações no intuito de aumentar a afinidade com o novo alvo. Por fim, apesar do fato de que a ferramenta EA apresentou um grande potencial de aplicação, comparar apenas uma estrutura por execução, pode se tornar uma tarefa onerosa para o usuário para os casos onde o número de estruturas a serem comparadas é muito grande.

Vários dos algoritmos citados durante este trabalho utilizaram as propriedades físico-químicas e as semelhanças estruturais compartilhadas entre diversos epítopos como guia para realizar a predição. Entretanto, como qualquer sequência do antígeno pode vir a ser um epítopo de um anticorpo (SELA-CULANG et al., 2014), não necessariamente compartilhando das mesmas características comuns dos demais epítopos, a seleção de epítopos que possuam apenas caraterísticas em comum pode acabar excluindo potencias epítopos em detrimento de uma predição mais generalizada. Em decorrência disto, foi desenvolvida uma abordagem para predição específica de anticorpo que utiliza informações de interação entre pares de aminoácidos da interface Ag-Ac e outras características de interface para prever *patches* de epítopos (SELA-CULANG et al., 2014). Vale ressaltar que embora os autores não tenham nomeado esta metodologia, ela será referenciada neste estudo como PEAc (Predição Específica de Anticorpo).

Inicialmente, foram extraídos todos os complexos Ag-Ac não redundantes do banco de dados PDB, obtendo-se pares de resíduos que integram a superfície de

contato entre as estruturas, totalizando 6.605 pares de interação e 1.772.624 pares que não estão em contato. Esse grande conjunto de dados foi dividido em três subgrupos, sendo dois para treinamento e um para teste (SELA-CULANG et al., 2014). Os grupos de treinamento foram utilizados para identificação de características dos resíduos de interface e distingui-los dos que não fazem parte, nesta pesquisa foram obtidas ou calculadas 10 características: resíduos vizinhos ao AA de contato de antígeno ou anticorpo; acessibilidade do antígeno ao solvente; estrutura secundária do antígeno; estado de desordem do antígeno; *hotspots* de interação do antígeno; cadeia pesada ou leve do anticorpo e qual alça do CDR; posição do AA no CDR; e cálculo da energia estatística do par de resíduos Ag-Ac que reflete o potencial de contato (SELA-CULANG et al., 2014).

Posteriormente, estas informações foram utilizadas para treinamento do algoritmo RF, designando uma pontuação de *patch* (PS) entre 0 e 1. Nesta métrica de pontuação, quanto maior for a pontuação direcionada para pares de resíduos, maior será a probabilidade de pertencer a interface Ag-Ac (SELA-CULANG et al., 2014). A PS advém da média da pontuação de resíduo (indica a possibilidade do AA pertencer ao epítopo) por *patch* (conjunto do AA com alta pontuação de resíduo e mais quatro AAs vizinhos) (SELA-CULANG et al., 2014). Essa abordagem é denominada como *patch-per-Ab* e mostrou maior precisão quando comparada a abordagem *per-residue,* abordagem que toma como base apenas a pontuação de resíduo, e EPSVR, um método de predição de epítopos baseado em estrutura (SELA-CULANG et al., 2014).

Um teste também foi realizado para identificar os epítopos do antígeno D8 de anticorpo produzidos pela vacina da varíola. Neste teste, foram utilizados a abordagem *patch-per-Ab* e EPSVR, além da junção de *patch-per-Ab* com dados experimentais de bloqueio cruzado de anticorpos, denominado de abordagem *patch-per-group* (SELA-CULANG et al., 2014). Os resultados da abordagem *patch-per-group* foram os mais precisos, pois fizeram a predição de cerca de 78% dos resíduos pertencentes à epítopos, enquanto que a predição do *patch-per-Ab* foi de 40% e a do EPSVR cerca de 20%. Entretanto, testes adicionais serão necessários para verificar a aplicação e o desempenho da abordagem *patch-per-group* (SELA-CULANG et al., 2014.

O PEAc apesar de apresentar uma abordagem útil para predição, infelizmente não prediz toda a estrutura do epítopo descontínuo, mas apenas fragmentos (*patches*), além de necessitar da sequência de resíduos do anticorpo, conhecimento sobre o padrão de bloqueio cruzado dos anticorpos e estrutura tridimensional do antígeno para identificar os resíduos em posições mais próximas (SELA-CULANG et al., 2014).

Seguindo esta mesma proposta de predição de epítopos direcionados a anticorpos, foi desenvolvido o EpiPred que considera como resíduos de epítopos aqueles que se apresentam na superfície dentro de uma distância de até 4,5 Å da estrutura do anticorpo. Estes resíduos são unidos em *patches* e então pontuados de acordo com as interações entre os aminoácidos por ajuste geométrico e pontuação do complexo Ag-Ac (KRAWCZYK et al., 2014).

O desempenho de predição do EpiPred foi comparado ao DiscoTope 2.0 e o método aleatório. No conjunto de 30 estruturas cristalográficas de complexos Ag-Ac, o preditor EpiPred não conseguiu prever oito epítopos e seu melhor desempenho foi para um antígeno de 381 resíduos (3r1g), apresentando 37% de precisão, 100% de sensibilidade e 0,57 de MCC. Apesar disso, a ferramenta EpiPred apresentou, em média, melhores resultados durante as predições realizadas em estruturas de antígenos menores, devido a menor quantidade de *patches* para analisar (KRAWCZYK et al., 2014). De forma geral, o EpiPred apresentou uma sensibilidade melhor e um nível de precisão similar ao método aleatório, obtendo: 44% de sensibilidade e 14% de precisão, frente aos 23% de sensibilidade e 14% de precisão do método aleatório. Neste contexto, embora o preditor DiscoTope 2.0 não tenha conseguido prever dezoito epítopos, a precisão na predição de oito dos doze epítopos restantes obteve valores superiores em relação ao EpiPred (KRAWCZYK et al., 2014).

Outro teste foi realizado com 15 complexos Ag-Ac modelados. Os resultados obtidos apresentaram desfecho parecido com o teste anterior, sem diferença estatística, indicando que a utilização de estruturas modeladas não atrapalha na predição do EpiPred (KRAWCZYK et al., 2014). Portanto, observa-se que a inclusão de informações advindas do complexo Ag-Ac pode trazer benefícios à precisão das

predições de epítopos, embora ainda apresentem baixa precisão e sensibilidade no caso do EpiPred.

Uma das principais limitações dos algoritmos de predição de epítopos conformacionais está na necessidade da estrutura do antígeno. No entanto, para a maior parte dos antígenos há a disponibilidade somente da sequência de aminoácidos. Em decorrência disso, foram desenvolvidos os preditores SePre e o SEPIa que possibilitam a utilização tanto das sequências primárias quanto as estruturas do antígeno como dados de entrada (*input*).

O preditor SePre emprega o método de aprendizado heterogêneo (*staged heterogeneity learning method*), possuindo duas etapas de aprendizado. A primeira etapa utiliza um algoritmo RF para realizar o treinamento dos padrões de cada modelo de propensão, gerando um subclassificador para cada modelo que, por sua vez, são responsáveis por indicar se o resíduo faz parte ou não do epítopo. A segunda etapa utiliza um algoritmo de árvore de decisão que avalia as informações provenientes da primeira etapa (REN et al., 2017). Um algoritmo foi desenvolvido para unir os aminoácidos antigênicos previstos pelo método de aprendizado de heterogêneo e formar um epítopo descontínuo (REN et al., 2017).

A ferramenta SePre foi comparada ao CBTOPE, que também utiliza a sequência do antígeno como dados de entrada (*input*), assim como as demais ferramentas baseadas em estrutura (DiscoTope 2.0, ElliPro e SEPPA 2.0). O teste de desempenho realizado para este preditor utilizou dois conjuntos de dados heterogêneos formados por estruturas experimentais e um conjunto homólogo com estruturas determinadas computacionalmente. O preditor SePre apresentou melhor precisão dentre as ferramentas comparadas nos três conjuntos de dados. No conjunto homólogo, por exemplo, SePre obteve o valor 0,426 de ACC, enquanto SEPPA 2.0, segundo melhor desempenho para este conjunto, obteve 0,164 de ACC, enquanto que o CBTOPE obteve apenas 0,133 de ACC. Resultados semelhantes foram obtidos em dois estudos de caso utilizando as mesmas ferramentas, indicando uma consistência entre as predições feitas pelo SePre (REN et al., 2017).

Por outro lado, o preditor SEPIa utiliza dois algoritmos de aprendizado de máquina, o RF e Gaussian Naïve Bayes, ambos treinados com 13 recursos para

predição distribuídas em composição de aminoácidos, hidrofilicidade, estrutura secundária, flexibilidade, regiões intrinsicamente desordenadas, tipo de energia, acessibilidade ao solvente, solubilidade e informações evolutivas, obtidas de um conjunto de 1.667 resíduos de ECBC e 16.780 NE. Os resultados destes algoritmos são unidos por um algoritmo de votação (DALKAS; ROOMAN, 2017).

O SEPIa utiliza uma janela de até 9 resíduos para análise da sequência, uma vez que os resultados obtidos durante o teste independente com 407 resíduos de ECBC e 5.192 NE, demostraram que a ferramenta apresenta maior AUC-ROC (0,646) quanto utiliza janelas contendo 9 resíduos (DALKAS; ROOMAN, 2017). O desempenho obtido foi comparado com outras ferramentas de predição que utilizam sequência ($Zhang_{ligado}$, $Zhang_{não\ ligado}$ e CBTOPE) e estrutura (EPCES, EPITOPIA, DiscoTope, BPredictor, SEPPA e EPSVR) no mesmo conjunto de dados independente. Dentre as ferramentas comparadas, o CBTOPE e o EPSVR obtiveram os maiores valores de AUC-ROC (0,607 e 0,606, respectivamente). Neste contexto, percebe-se que o SEPIa apresentou um desempenho ligeiramente superior, apresentando um valor de AUC-ROC 0,039 e 0,04 superior ao CBTOPE e EPSVR, respectivamente (DALKAS; ROOMAN, 2017).

O Quadro 6 reúne as ferramentas apresentadas neste tópico com suas respectivas vantagens e desvantagens.

Quadro 6 – Vantagens e desvantagens das ferramentas para predição de epítopos conformacionais em ordem alfabética (continua).

Ferramenta	Vantagens	Desvantagens
Epitope-Analyser (EA) (MONTIEL-GARCIA et al., 2022)	○ Auxilia na identificação de epítopos conservados entre espécies que podem vir a ser alvos de bNAbs. ○ Disponível em servidor *on-line*.	○ Pode ser dispendioso o trabalho para comparar a semelhança de um epítopo frente a um grande conjunto de outros epítopos. ○ Apresentou apenas 1 teste de certificação.

Quadro 6 – Vantagens e desvantagens das ferramentas para predição de epítopos conformacionais em ordem alfabética (continua).

Ferramenta	Vantagens	Desvantagens
EpiPred (KRAWCZYK et al., 2014)	○ Predição de epítopo direcionada para anticorpos por ajuste geométrico e pontuação assimétrica. ○ Permite a utilização de estruturas modeladas como dados de entrada (*input*).	○ A precisão baixa e diminui para antígenos com muitos resíduos, devido ao aumento do número de *patches*. ○ Disponível apenas como *stand-alone*, o que pode dificultar sua utilização para usuários iniciantes.
PEAc (SELA-CULANG et al., 2014)	○ Predição de epítopo direcionada para um anticorpo pela abordagem *patch-per-Ab* com aumento da precisão quando associada a dados de bloqueio cruzado de Ac (*patch-per-group)*.	○ Prediz apenas fragmentos (*patches*) do epítopo. ○ Necessita de informações sobre o padrão de bloqueio cruzado, além da estrutura tridimensional do Ag, para identificação dos resíduos próximos espacialmente. ○ Disponível apenas como *stand-alone*, o que pode dificultar sua utilização para usuários iniciantes.
SePre (REN et al., 2017)	○ Predição através do método de aprendizado heterogêneo que utiliza a sequência primária do antígeno como *input*. ○ Nos testes independentes, apresentou melhor desempenho que outros algoritmos baseados em sequência e estrutura.	○ O algoritmo está disponível apenas mediante solicitação aos autores, o que torna seu uso mais limitado.

Quadro 6 – Vantagens e desvantagens das ferramentas para predição de epítopos conformacionais em ordem alfabética (conclusão).

Ferramenta	Vantagens	Desvantagens
SEPIa (DALKAS; ROOMAN, 2017)	○ Predição através de um algoritmo de votação que une as estimativas do classificador de Gaussian Naïve Bayes e Random Forest treinados em 13 recursos. ○ Assim como o SePre, o SEPIa também apresenta uma discreta melhora no desempenho quando comparado com outros algoritmos baseados em sequência e estrutura.	○ Disponível apenas como *stand-alone*, o que pode dificultar sua utilização para usuários iniciantes. ○ Limitado a janelas de 9 aminoácidos. ○ Possui um desempenho ligeiramente superior ao CBTOPE e EPSVR.

Fonte: A autora (2022).

5.1.3 Algoritmos para Predição de Epítopos não Peptídicos

Até o momento este presente trabalho relatou algoritmos de predição de epítopos peptídicos de células B. No entanto, outras classes químicas de natureza não peptídicas, como carboidratos e lipídios podem vir a ativar a resposte imune do organismo. Ainda assim, este tema foi pouco explorado dentre os algoritmos de predição e durante a coleta de artigos para esta revisão encontrou-se apenas uma ferramenta para este tipo de predição: o preditor *Non-peptidic epitope prediction* (N-PEP).

A ferramenta N-PEP emprega impressão digital de conectividade estendida (ECFPs), também denominada de impressão digital de Morgan, para vetorizar as estruturas de treinamento que são compostas de 2.140 epítopos de células B e 579 epítopos de células T (ECT) extraídos do IEDB, e 42.643 moléculas extraídas do banco de dados ChEBI (*Chemical Entities of Biological Interest*) (ZIEREP et al.,

2022). As diversas classes químicas não peptídicas foram separadas por meio de análise de enriquecimento de ontologia, dividindo-se em oito *clusters*: i) esteroides ou terpenóides (*cluster* 0); ii) derivados de betaína ou glicerolipídeos (*cluster* 1); iii) derivados de ácidos graxos (*cluster* 2); iv) derivados de acil-CoA (*cluster* 3) que representam os não epítopos, pois não foram encontrados relatos de epítopos de células B ou T para estas estruturas; v) derivados de glicosídeo ou oligossacarídeos (*cluster* 4); vi) entidades moleculares contendo nucleobases (*cluster* 5); vii) pequenas moléculas (*cluster* 6); viii) haletos cíclicos ou fenóis (*cluster* 7) (ZIEREP et al., 2022).

O ECFPs cria padrões circulares em torno de cada átomo de uma molécula, gerando subestruturas que formam a impressão digital desta molécula (ZIEREP et al., 2022). As impressões digitais estão relacionadas a características e foram utilizadas para treinar RF, k-NN e NN. RF e NN obtiveram valores de AUC-ROC semelhantes para todos os *clusters* de ECB e ECT, enquanto k-NN teve desempenho aproximado apenas para o *cluster* 5 de células T (ZIEREP et al., 2022). O algoritmo do tipo RF foi selecionado para um teste com o conjunto independente formado por 47 ECB, 71 ECT e 2.190 estruturas obtidas do ChEBI. RF apresentou bom desempenho novamente com um AUC-ROC superior a 0,69 para todos os *clusters* apesar das poucas estruturas disponíveis de treinamento (ZIEREP et al., 2022).

O N-PEP está disponível em um servidor *on-line* que classifica o tipo de classe química a estrutura de entrada (*input*) integra e prediz a probabilidade desta molécula ser ou não um epítopo de célula B ou T. Ademais, o N-PEP apresenta ao usuário 5 epítopos disponíveis no IEDB que apresentam características de impressão digital próximas a estrutura em análise, permitindo que o usuário analisar epítopos semelhantes (ZIEREP et al., 2022). Por fim, o N-PEP demonstra bom desempenho de previsão de epítopos não peptídicos, alcançando valores de AUC-ROC entre 0,69 a 0,96, mesmo com a baixa quantidade de estruturas disponíveis (ZIEREP et al., 2022).

5.2 ALGORITMOS UTILIZADOS PARA MODELAGEM DE ANTICORPOS

Os artigos selecionados sobre a modelagem de anticorpos podem ser separados a partir de três metodologias: i) modelagem por homologia, utilizando a conformação de anticorpos com estrutura já definida como modelo ("*template*") para a construção da estrutura do anticorpo alvo; ii) modelagem *ab initio*, explorando o espaço conformacional para a modelagem e iii) *redesign* que objetiva modificar a estrutura do anticorpo no intuito de melhorar a afinidade do mesmo com o antígeno original ou direcioná-lo para um novo antígeno (LIANG et al., 2021; MARKS et al., 2017; NIMROD et al., 2018). Vale ressaltar que diversos algoritmos utilizam modelagem por homologia e *ab initio* no mesmo algoritmo e, assim, contornando as limitações intrínsecas de cada metodologia.

5.2.1 Algoritmos de modelagem de anticorpos por homologia e *ab initio*

O pacote Rosetta é amplamente utilizado na imunoinformática por conta de seu amplo leque de ferramentas voltadas à modelagem de anticorpos com diferentes abordagens como, por exemplo, os protocolos RosettaAntibody, RosettaAntibodyDesign (RAbd) e AbDesign (SCHOEDER et al., 2021).

O RosettaAntibody emprega tanto o método de modelagem por homologia, para modelagem das estruturas canônicas, quanto o método *ab initio*, para modelagem do *loop* CDRH3 e simular de forma apropriada sua diversidade conformacional (SCHOEDER et al., 2021; WEITZNER et al., 2017). Como dados de entrada, o RosettaAntibody recebe a sequência do anticorpo, iniciando uma sequência de três etapas que são aplicadas para realizar a modelagem. A primeira é direcionada para a seleção da estrutura modelo para cada *loop* CDR, FR_L e FR_H através de similaridade de sequência utilizando a ferramenta BLAST em um banco de dados, separando cada parte da estrutura de acordo com seu comprimento. Além disso, nesta etapa, é selecionado o modelo de orientação de V_H-V_L, também utilizando BLAST, no intuito de selecionar a orientação mais adequada a partir dos modelos escolhidos anteriormente (1 e 2). A segunda etapa realiza o enxerto dos

modelos e a estrutura é relaxada e na terceira etapa ocorre a modelagem do loop H3 com KIC (*kinematic loop closure*) junto a otimização da orientação da interface V_H-V_L (SCHOEDER et al., 2021; WEITZNER et al., 2017).

O desempenho do RosettaAntibody foi testado no AMA (*Antibody Modeling Assessments*), evento que disponibiliza sequências de estruturas cristalográficas de anticorpos ainda não publicadas com o objetivo de analisar se os modelos previstos computacionalmente se assemelham a estrutura nativa. Na segunda edição do AMA, foram disponibilizados 11 anticorpos para teste, o RosettaAntibody apresentou boa acurácia para modelagem dos *loops* canônicos, além de obter 4 dos melhores modelos de CDRH3 entre os concorrentes (SCHOEDER et al., 2021). Entretanto, as predições realizadas com este protocolo também apresentaram falhas na orientação da interface V_H-V_L e na modelagem dos demais CDRH3 (SCHOEDER et al., 2021).

RosettaAntibody está disponível em formato *stand-alone* e no servidor ROSIE (*Rosetta Online Server that Include Everyone*). Este último apresenta também outras ferramentas baseadas em protocolos do Rosetta como, por exemplo, o SnugDock, ferramenta específica para a realização de *docking* molecular entre antígeno e anticorpo, além de diversas outras ferramentas úteis na área de biologia computacional (MORETTI et al., 2017). Apesar de facilitar o acesso ao uso do protocolo RosettaAntibody, o servidor ROSIE apresenta uma longa fila de espera para a execução dos cálculos, o que pode estar relacionado a robustez do algoritmo e também a alta demanda de trabalhos submetidos.

Uma outra opção de serviço *on-line* para a predição de anticorpos pode ser encontrada no ABodyBuilder, implementado no servidor SabPred. O servidor SabPred, assim como o ROSIE também apresenta diversas ferramentas: ANARCI para numeração de sequência, PEARS para predição de cadeia lateral, SCALOP predição de conformação canônica a partir da sequência de resíduos, TAP para avaliação de desenvolvimento de Acs e Hu-mAb para humanização de Acs (DUNBAR et al., 2016).

De forma similar ao RosettaAntibody, o ABodyBuilder começa selecionando modelos de estruturas. Porém, diferente do RosettaAntibody, o ABodyBuilder procura estruturas com valores de identidade de sequência superior a 80% no banco

de anticorpos SAbDab (*Structural Antibody Database*), com o objetivo de utilizar a mesma estrutura de modelo tanto para o Fv como para a orientação da interface V_H-V_L. Em casos onde não é possível encontrar estruturas com este nível de identidade de sequência, são utilizados dois modelos distintos para a modelagem de V_H e V_L, levando em consideração a estrutura que apresentar a identidade mais alta como modelo da orientação da interface entre estes fragmentos variáveis (LEEM et al., 2016).

O próximo passo é a modelagem dos *loops* CDR utilizando o FREAD, selecionando modelos para cada um dos seis *loops*. Este método apresentou ótimos resultados para a modelagem dos loops canônicos, mas foi menos preciso para o CDRH3 (LEEM et al., 2016). Para os casos onde não é possível encontrar modelos adequados pelo FREAD, é realizada uma busca no banco de dados de F_V. Se mesmo assim não for encontrado é utilizado um *loop* com sequência e comprimento aproximados, obtidos através da pontuação BLOSUM62, com resolução inferior a 2,5Å. Em último caso, é utilizada a ferramenta MODELLER para realizar a modelagem *ab initio* do *loop* (LEEM et al., 2016). Por fim, as cadeias laterais podem ser previstas totalmente ou parcialmente (aproveita algumas cadeias laterais da estrutura modelo) com SCWRL4 e os conflitos estéricos são resolvidos com o relaxamento da estrutura com MODELLER *loop* (LEEM et al., 2016).

O diferencial do ABodyBuilder é a capacidade de modelar nanocorpos e indicar 11 “passivos de sequência” (*sequence liabilities*) na estrutura modelada que representam características que podem atrapalhar o desempenho do anticorpo *in vitro* como, por exemplo, cisteínas não pareadas, oxidação de metionina, entre outros (LEEM et al., 2016). Ademais, esta ferramenta apresenta o limite de desvio (RMSD) que a estrutura alvo foi modelada, dentro de uma determinada probabilidade. Vale ressaltar que este valor encontrado não é o valor real de RMSD, pois dado dentro de uma probabilidade, logo ele apenas estima a acurácia da modelagem (LEEM et al., 2016). Veja um exemplo descrito pelos autores:

> [...] um modelo de domínio V_H com 80% de identidade de sequência para um domínio V_H alvo será modelado em RMSD ≤ 1,0Å com uma probabilidade de 0,75. Isso não significa necessariamente que o domínio VH terá RMSD ≤1,0 Å para o alvo. Em vez disso, afirma que há 75% de chance de que o alvo seja modelado em RMSD ≤1,0Å (LEEM et al., 2016).

O desempenho do preditor ABodyBuilder foi testado para a modelagem de 462 F_V, e 136 estruturas extraídas do PDB, incluindo F_V e nanocorpos (VHH e apenas V_L). Os resultados dos dois conjuntos foram semelhantes aos testes do FREAD anteriormente relatados, onde a menor ACC obtida foi para a modelagem do fragmento CDRH3. Entretanto, nestes testes a ACC melhorou para os *loops* não canônicos com RMSDs entre 0,4 a 0,8Å (LEEM et al., 2016). Além disso, assim como o RosettaAntibody, o ABodyBuilder foi testado com as mesmas 11 estruturas do AMA II e apresentou resultados semelhantes em relação aos demais participantes da competição como, por exemplo, o RosettaAntibody, Kotai Antibody Builder e PIGS, demonstrando assim seu alto desempenho de modelagem (LEEM et al., 2016).

O FREAD também é utilizado em conjunto com um algoritmo *ab initio* do preditor Sphinx para realizar a modelagem apenas dos *loops* do CDR, ao contrário dos demais preditores relatados que também modelaram a região do F_V. Este algoritmo utiliza como dados de entrada (*input*) a estrutura do anticorpo e sequência do *loop*, o que pode representar uma limitação para pesquisas que possuem apenas a sequência de aminoácidos disponível. A partir destes dados, o Sphinx inicia a busca em um banco de dados no intuito de selecionar os fragmentos que apresentem três características semelhantes com a estrutura a ser modelada: i) comprimento menor; ii) resíduos da âncora, localizados nas extremidades inicial e final do fragmento, que devem corresponder ao alvo e iii) similaridade de sequência. O comprimento dos fragmentos selecionados é ajustado pelo algoritmo *ab initio*, fazendo com que cada fragmento seja utilizado na construção de 100 iscas (*decoys*) para a construção do *loop*. A função de energia baseada em RAPDF é utilizada para selecionar as 500 melhores iscas, que por sua vez passam pelo SCWRL4 no âmbito de acrescentar as cadeias laterais, minimizadas pelo protocolo de refinamento de *loop* do Rosetta (MARKS et al., 2017).

Testes foram realizados no intuito de comparar o desempenho na modelagem de anticorpos entre o preditor FREAD, o algoritmo *ab initio* isolado e o preditor Sphinx, este último com e sem a etapa de minimização. Os resultados mostraram que a ferramenta FREAD conseguiu obter uma média menor para os valores de RMSD. No entanto, o FREAD não conseguiu modelar 9 das 36 estruturas

testadas, enquanto que o Sphinx, em ambas as versões (com e sem minimização), conseguiram modelar todas as amostras com um RMSD ligeiramente maior, sendo a versão do Sphinx com minimização a que apresentou os resultados com menor desvio. Por fim, o algoritmo *ab initio* foi o que apresentou maiores valores de RMSD (MARKS et al., 2017). Além disto, outro teste comparativo foi realizado no âmbito de comparar o Sphinx com as ferramentas *ab initio* MODELLER e RAPPER, onde o Sphinx apresentou em média 0,89Å de RMSD, enquanto que os demais apresentaram valores superiores a 2Å de RMSD (MARKS et al., 2017).

O Sphinx possui uma versão criada especificamente para realizar a modelagem do CDRH3, denominado Sphinx H3. Esta versão difere da versão original pela adição do banco de dados formado apenas por *loops* H3 extraídos dos anticorpos do SAbDab, dados diédricos e quantidade de fragmentos para construção da estrutura. Esta versão foi capaz de produzir bons modelos de H3, inclusive com desempenho comparável ao RosettaAntibody em um tempo menor de execução (MARKS et al., 2017). Por fim, nota-se que as informações advindas de fragmentos com comprimentos diferentes dos *loops* alvo podem também ser utilizados para gerar bons modelos de anticorpos.

O RosettaAntibodyDesign (RAbD) é uma metodologia para planejamento de anticorpos desenvolvida por Adolf-Bryfogle e sua equipe (ADOLF-BRYFOGLE et al., 2018) e foi adicionada ao *framework* Rosetta através da elaboração de mais de 20 mil linhas de código. Esta metodologia pode ser dividida em três etapas.

A primeira etapa consiste em utilizar um banco contendo 3974 fragmentos de regiões CDRs obtidos ao longo de 1.000 estruturas de anticorpos determinadas experimentalmente, selecionados previamente através de uma curadoria realizada pelos próprios autores do RAbD (ADOLF-BRYFOGLE et al., 2018). A junção (*graft*) destes fragmentos é realizada de forma a gerar um novo anticorpo utilizando o algoritmo *Cyclic Coordinate Descent* (Decréscimo de Coordenadas Cíclico), desenvolvido pelos próprios autores do RAbD (ADOLF-BRYFOGLE et al., 2018).

Na segunda etapa, utiliza-se a abordagem perfis de sequência no conjunto de estruturas das regiões CDRs para, desta forma, avaliar o impacto que as substituições de aminoácidos geram nas estruturas dos anticorpos. Por fim, é

realizada uma otimização da estrutura resultante através do método de Monte Carlo já disponível no *framework* Rosetta (ADOLF-BRYFOGLE et al., 2018).

Por fim, na terceira etapa, ferramentas de análise de anticorpos desenvolvidas pelos autores do RAbD são utilizadas para avaliar e selecionar as estruturas de anticorpos mais promissores no intuito de serem sintetizadas e testadas experimentalmente (ADOLF-BRYFOGLE et al., 2018).

Como forma de avaliar a precisão desta metodologia os autores utilizaram a métrica razão risco do *design*, DDR (*Design Risk Ratio*). Esta métrica consiste em pegar o número de vezes em que a estrutura nativa foi encontrada entre as soluções e dividir pelo número de vezes em que a estrutura nativa foi obtida durante os cálculos sem serem listadas entre as soluções (ADOLF-BRYFOGLE et al., 2018). Nesta métrica, valores acima de 1,0 significa que a estrutura nativa foi identificada corretamente entre as soluções em um número maior de vezes do que a mesma foi descartada erroneamente durante os cálculos. Neste sentido, durante o teste realizado com 60 anticorpos obteve-se valores de DRR acima de 2.4 para todos os CDRs sem H3 e DDR acima de 3.5 para o L3 e H1 (ADOLF-BRYFOGLE et al., 2018).

O AbDesign é uma metodologia para modelagem de anticorpos que utiliza protocolos disponíveis no *framework* Rosetta através de rotinas conhecidas como RosettaScripts (LAPIDOTH et al., 2015). Uma das principais vantagens desta abordagem está no fato de que os autores desta metodologia elaboraram dois bancos de dados correlacionados e interligados a partir das estruturas de anticorpos que ocorrem naturalmente obtidas no PDB, são eles: i) Banco de Matriz de posição especifica (PSSM) e ii) Banco de torção da cadeia principal (*backbone*), no intuito de gerar estruturas de anticorpos realistas (LAPIDOTH et al., 2015).

O banco PSSM contém as restrições para o posicionamento dos átomos dos aminoácidos ao longo do chassi do anticorpo, “relaxando” o posicionamento dos átomos nas regiões CDR de forma a permitir diferentes combinações de aminoácidos nestas regiões (LAPIDOTH et al., 2015). O PSSM é utilizado ao longo dos cálculos de modelagem de anticorpos no intuito de restringir a escolha das

sequências de aminoácidos, servindo como ponto de corte para manter a conservação da estrutura do anticorpo (LAPIDOTH et al., 2015).

Durante a modelagem do anticorpo, as conformações pré-computadas no banco PSSM são utilizadas de forma a gerar uma única matriz PSSM para todo o anticorpo ao combinar quatro segmentos PSSMs diferentes (VL, L3, VH e H3). Estes quatro segmentos são substituídos e otimizados de forma a explorar o espaço combinatório, levando em consideração toda a estrutura do anticorpo e não apenas a superfície que interage com o ligante, gerando ao todo 50 estruturas contendo combinações diferentes destes segmentos.

De posse dos modelos otimizados do anticorpo, é realizado o *docking* molecular entre o anticorpo e a molécula alvo através do RosettaDock no intuito de gerar uma orientação inicial de ligação. Após isso, a sequência dos anticorpos é otimizada para interagir com o ligante alvo, levando em consideração tanto as restrições de posicionamento do PSSM, explorando os graus de liberdade através do banco de torções (LAPIDOTH et al., 2015). No intuito de avaliar simultaneamente a estabilidade do anticorpo e a afinidade predita pelo ligante, um algoritmo de lógica *fuzzy* foi elaborado de forma a pontuar as melhores soluções, levando em consideração a correlação entre os valores de energia livre de ligação (k_d) predita e algumas propriedades topológicas como, por exemplo, o tamanho da região de interface do anticorpo (*Buried Surface Area*), a complementariedade estrutural entre o anticorpo e ligante (*Shape Complimentarity*) e a interface do ligante (LAPIDOTH et al., 2015).

O OptMAVEN 2.0 utiliza o banco de dados MAPs (*Modular Antibody Parts*) com o objetivo de formar a estrutura do anticorpo (CHOWDHURY; ALLAN, MARANAS, 2018). O MAPs possui diversas partes de estruturas de anticorpos divididas em regiões variáveis (V), CDR3 e junção (J) para as cadeias pesadas (H) e leves (L): HV, HCDR3, HJ, LV, LCDR3 E LJ (CHOWDHURY; ALLAN; MARANAS, 2018). Essas partes de anticorpos são selecionadas a partir do cálculo da energia de interação com o antígeno alvo em diversas rotações a partir de pesquisa de grade, tendo a soma das energias é minimizada por MILP (*Mixed-Integer Linear Programming*) e selecionando, assim, seis MAPs não conflitantes para cada pose do antígeno. Esses dados são agrupados para formar um arquivo PDB e um FASTA do

conjunto com menor energia (CHOWDHURY; ALLAN, MARANAS, 2018). Cada PDB passa por um relaxamento utilizando o campo de força CHARMM, enquanto que a estabilidade do complexo é avaliada através do método de dinâmica molecular com a ferramenta QwikMD. Por fim, no intuito de diminuir a imunogenicidade do anticorpo, utiliza-se a maturação por afinidade através de IPRO (*IterativeProtein Redesign and Optimization*) (CHOWDHURY; ALLAN; MARANAS, 2018).

Vale ressaltar que OptMAVEN 2.0 realiza o pré-processamento do arquivo de entrada do antígeno, removendo cadeias e átomos não relacionados ao antígeno e adicionando átomos de hidrogênio, processamento ausente na versão 1.0 do programa (CHOWDHURY; ALLAN; MARANAS, 2018). Outra diferença encontrada entre estas versões é que a versão 1.0 necessita de mais espaço de armazenamento e tempo de CPU para realizar a modelagem, algo que foi reduzido em sua versão mais recente, necessitando de 84% menos espaço em disco e utiliza 74% menos recursos computacionais (tempo de CPU), além da automatização de algumas etapas, dispensando a necessidade do comando do usuário (CHOWDHURY; ALLAN; MARANAS, 2018).

No entanto, a ferramenta OptMAVEN apresenta algumas limitações como, por exemplo, a utilização de energia simplificada na etapa MILP, o que limita a estimativa do potencial químico do parátopo, e ao não considerar a estabilidade do anticorpo experimentalmente como, por exemplo, possíveis agregações que levam a perda de funcionalidade (CHOWDHURY; ALLAN; MARANAS, 2018). Entretanto, mesmo com estas limitações, o OptMAVEN continua sendo uma ótima opção para a modelagem de anticorpos, principalmente nos casos onde há apenas a estrutura do antígeno como ponto de partida.

Por fim, as vantagens e desvantagens destes algoritmos se encontram resumidos no quadro 7.

Quadro 7 – Vantagens e desvantagens das ferramentas para modelagem de anticorpos por homologia em ordem alfabética (continua).

Ferramenta	Vantagens	Desvantagens
AbDesign (LAPIDOTH et al., 2015)	○ Utiliza dois bancos de dados correlacionados e interligados a partir das estruturas de anticorpos que ocorrem naturalmente obtidas no PDB, são eles: i) Banco de Matriz de posição especifica (PSSM) e ii) Banco de torção da cadeia principal (*backbone*), no intuito de gerar estruturas de anticorpos realistas. ○ Utiliza o algoritmo de lógica *fuzzy* para avaliar a estabilidade do anticorpo e a afinidade predita pelo ligante.	○ Disponível apenas como *stand-alone*, o que pode dificultar sua utilização para usuários iniciantes.
ABodyBuilder (LEEM et al., 2016)	○ Modelagem de anticorpos e nanocorpos com ACC sub-Angstrom, identificando 11 “passivos de sequência” que podem interferir no desempenho *in vitro* do anticorpo. ○ Disponível em servidor *on-line* com resultado rápido.	○ Em alguns casos, os “passivos de sequência” podem vir a ser problemáticos somente em certas condições como, por exemplo, pH baixo. Entretanto, o algoritmo ABodyBuilder não realiza esta distinção.

Quadro 7 – Vantagens e desvantagens das ferramentas para modelagem de anticorpos por homologia em ordem alfabética (conclusão).

Ferramenta	Vantagens	Desvantagens
Rosetta Antibody (SCHOEDER et al., 2021; WEITZNER et al., 2017)	○ Modela o anticorpo por homologia, incluindo os *loops* CDR, FR_L, FR_H e orientação V_H-V_L. ○ Explora a diversidade estrutural do CDRH3 com método *de novo.* ○ Disponível em servidor *on-line* e também em versão *stand-alone.*	○ Longa fila de espera no servidor *on-line*.
RAbD (ADOLF-BRYFOGLE et al., 2018)	○ Utiliza a abordagem de perfis de sequência no conjunto de estruturas de CDRs para avaliar o impacto que as substituições de aminoácidos geram nas estruturas dos anticorpos. ○ Otimização da estrutura resultante através do método de Monte Carlo.	○ Disponível apenas como *stand-alone*, o que pode dificultar sua utilização para usuários iniciantes.
Sphix (MARKS et al., 2017)	○ Possui mais estruturas como modelo, já que não limita a busca aos bancos de dados separados por comprimento, como ocorre com as ferramentas ABodyBuilder e RosettaAntibody. ○ Possui uma versão específica para modelagem de H3 (Sphinx H3).	○ É necessário ter disponível a estrutura resolvida do anticorpo, pois modela apenas o CDR. ○ Disponível apenas como *stand-alone*, o que pode dificultar sua utilização para usuários iniciantes.

Fonte: A autora (2022).

5.2.2 Algoritmos de modelagem de anticorpos por *redesign*

O *redesign* de anticorpos, como o próprio nome indica, inclui a modificação da estrutura de um anticorpo com o objetivo de melhorar sua afinidade em relação ao antígeno alvo ou direcionar este anticorpo para um antígeno diferente do seu alvo inicial (LIANG et al., 2021; NIMROD et al., 2018). Nesta seção serão abordados dois protocolos computacionais de *redesign* para auxiliar na engenharia de anticorpos.

Os protocolos computacionais unem diversas ferramentas já descritas na literatura e facilitam a utilização de abordagens de imunoinformática, tanto para pesquisadores mais experientes, ao reunir os diversos algoritmos que possibilitam chegar ao objetivo final da pesquisa e, consequentemente, economizando tempo de pesquisa na literatura, quanto para pesquisadores menos experientes, pois apresenta de forma simples um passo-a-passo para a realização do trabalho.

O protocolo IsAb inicia com as estruturas do anticorpo e do antígeno (Fluxograma 2). Caso a estrutura tridimensional do anticorpo não tenha sido determinada experimentalmente ou não esteja disponível em bancos de dados públicos como, por exemplo, o PDB, o usuário pode utilizar a sequência primária para realizar a modelagem no RosettaAntibody (LIANG et al., 2021), ferramenta descrita detalhadamente na seção 5.2.1 deste presente trabalho e disponível tanto no servidor *on-line* ROSIE, quanto em versão instalável *stand-alone* no pacote Rosetta. De posse da estrutura do antígeno e anticorpo, estes são submetidos ao RosettaRelax com o objetivo de refinar as estruturas e aproximá-las do seu estado nativo, sendo selecionadas as estruturas de pontuação mais baixa dentre os resultados de saída (LIANG et al., 2021). Este recurso de relaxamento (flexibilização) está disponível no RosettaAntibody através da função "*Grafted-Relaxed-Model*". Portanto, se a estrutura do anticorpo tiver sido modelada, só há a necessidade de realizar o procedimento do RosettaRelax apenas para o antígeno (LIANG et al., 2021).

Fluxograma 2 – Procedimento do IsAb.

Fonte: Traduzido de LIANG et al. (2021).

A próxima etapa (Fluxograma 2) tem como objetivo realizar o procedimento de *docking* molecular, onde as melhores conformações de encaixe entre o antígeno e o anticorpo são verificadas. O *docking* global é realizado quando há poucas

informações entre as interações do complexo antígeno-anticorpo (Ag-Ac), sendo indicado o uso do servidor ClusPro na opção avançada “Antibody Mode” (LIANG et al., 2021). Os resultados obtidos são refinados com o procedimento de *docking* local feito pelo SnugDock no modo “*through mode*”, selecionando a solução contendo o complexo Ag-Ac com a menor valor de pontuação de interface (I_sc) (LIANG et al., 2021).

A partir da melhor conformação do complexo Ag-Ac, inicia-se a etapa de *redesign* que inicia com a busca por alaninas (*Alanine scanning*), realizado por um protocolo do framework Rosetta de mesmo nome, onde os aminoácidos presentes na interface Ag-Ac sofrem uma mutação para tornarem-se alanina para, desta forma, calcular a mudança de energia entre o estado inicial (com o aminoácido original) para o estado onde houve a mutação indicando, assim, os principais aminoácidos presentes na interação Ag-Ac, comumente denominados *hotspots* (LIANG et al., 2021). O próximo passo é a maturação de afinidade realizada por diversos *scripts* desenvolvidos no laboratório do Dr. Jens Meiler, os quais estão disponíveis no *website* Rosetta *tutorials* (http://www.meilerlab.org/index.php/rosetta-tutorials). Estes *scripts* são responsáveis por testar os 20 aminoácidos para cada posição do anticorpo (LIANG et al., 2021). Ao final, espera-se que o anticorpo modificado apresente alta afinidade com seu epítopo alvo.

O protocolo IsAb foi aplicado para duas situações: i) tendo a estrutura do anticorpo disponível experimentalmente e ii) estrutura modelada *in silico* com informações sobre o complexo Ag-Ac indisponíveis (LIANG et al., 2021). Nas duas situações, o IsAb se mostrou capaz de produzir anticorpos com boa estabilidade e afinidade predita pelo epítopo alvo (LIANG et al., 2021).

A tarefa de produzir novos anticorpos é um grande desafio, tanto pelo tempo de pesquisa quanto pelo custo experimental envolvidos. Diante deste problema, foi desenvolvido o protocolo de Re-epitoping. Este protocolo inclui ferramentas *in silico* e experimentos *in vitro* para desenvolvimento de novos anticorpos a partir de anticorpos conhecidos e disponíveis em bancos de dados, direcionando-os para os novos antígenos (NIMROD et al., 2018).

A primeira etapa consiste em obter a estrutura tridimensional do antígeno de interesse. Como estudo de caso e demonstração da ferramenta, os autores utilizaram em seu artigo a interleucina-17A (IL-17A), a qual foi modelada no MODELLER (NIMROD et al., 2018). Posteriormente, foram selecionadas diversas estruturas de anticorpos do banco de dados, onde 501 estruturas foram extraídas do PDB com o objetivo de projetar o anticorpo mais indicado para o alvo IL-17A (NIMROD et al., 2018). Os anticorpos foram então submetidos a cálculos de *docking* molecular utilizando os programas ZDOCK e HEX, no intuito de selecionar as soluções presentes entre as 10 melhores soluções para ambos os programas, indicando as estruturas que possam apresentar uma boa complementaridade inicial com o novo antígeno (NIMROD et al., 2018).

Neste sentido, mutações são realizadas no anticorpo escolhido de forma a aumentar a afinidade pelo antígeno na formação do complexo Ag-Ac e, por isso, foram analisados os resíduos dentro de um raio de 5Å do alvo IL-17A para realizar as mutações através do *Discovery Studio protocols* (NIMROD et al., 2018). A avaliação da estabilidade das mutações causadas na estrutura foi realizada a partir de simulações de dinâmica molecular usando o programa GROMACS 4.54, criando, assim, bibliotecas de mutantes estáveis (NIMROD et al., 2018). As interações na formação do complexo foram classificadas por algoritmos RF, o qual foi treinado em um conjunto de 4.635 interações de cadeia pesada e 2.409 interações de cadeia leve, utilizando 8 recursos para analisar o tipo de aminoácido que entra em contato durante a interação Ag-Ac e a sua respectiva pontuação; acessibilidade dos resíduos ao solvente; tipo e comprimento do CDR e a localização do resíduo de contato do anticorpo no CDR (NIMROD et al., 2018). Por fim, foram utilizadas técnicas *in vitro* (*display* de superfície de levedura) para selecionar mutantes e realizar a humanização do anticorpo (NIMROD et al., 2018).

O princípio por trás deste preditor se assemelha bastante ao do OptMAVEN 2.0 (descrito na seção 5.2.1 deste trabalho), uma vez que ambos projetam novos anticorpos a partir da estrutura do antígeno. Entretanto, a metodologia do Re-epitoping seleciona um único anticorpo e realiza modificações estruturais nele de forma a aumentar sua afinidade por um novo antígeno de interesse (NIMROD et al., 2018), enquanto que o OptMAVEN 2.0 utiliza uma combinação de seis partes

obtidas de diversos anticorpos diferentes para formar um novo anticorpo (CHOWDHURY; ALLAN, MARANAS, 2018). Portanto, estas ferramentas apresentam soluções diferentes para o mesmo problema.

O *redesign* se apresenta como uma opção viável para melhoria de afinidade e desenvolvimento de novos anticorpos, mas ainda existem certas dificuldades como falta de automação e possibilidade de reatividade cruzada com o antígeno original para os anticorpos “re-epitopados” (Quadro 8).

Quadro 8 – Vantagens e desvantagens das ferramentas para *redesign* de anticorpos em ordem alfabética.

Ferramenta	Vantagens	Desvantagens
IsAb (LIANG et al., 2021)	○ Protocolo computacional para *redesign* de anticorpos completo que une procedimentos de modelagem de anticorpo, *docking* e maturação de afinidade. ○ Pode ser aplicado para casos que já possuam a estrutura determinada experimentalmente ou apenas a sequência de aminoácidos	○ Não conduz o processo automaticamente, podendo representar um desafio para pesquisadores que são usuários iniciantes em ferramentas de imunoinformática, principalmente na etapa de maturação de afinidade *in silico*.
Re--epitoping (NIMROD et al., 2018)	○ Desenvolvimento de novos anticorpos a partir da estrutura do antígeno ou epítopo utilizando anticorpos já disponíveis em bancos de dados, economizando tempo e custo na pesquisa. ○ O novo anticorpo aproveita as propriedades biofísicas do anticorpo original.	○ O protocolo apresenta etapas *in vitro* para seleção dos mutantes, o que podem aumentar o custo e tempo de pesquisa, mas os pesquisadores podem buscar opções *in silico*. ○ O protocolo não é automatizado. ○ Apresenta possibilidade de reação cruzada do anticorpo com seu antígeno de origem.

Fonte: A autora (2022).

6 CONCLUSÃO

O campo de pesquisa da imunoinformática une os conhecimentos das áreas da imunologia e biologia computacional e, com isso, desenvolve algoritmos que podem auxiliar no desenvolvimento de novas vacinas, encontrar alvos para o diagnóstico de algumas doenças, entre diversos outros campos de aplicação. Neste presente estudo, foi realizada uma revisão bibliográfica sistemática qualitativa de 26 artigos que descrevem ferramentas computacionais voltadas para duas áreas de aplicação: predição *in silico* de epítopos de células B (17 artigos) e modelagem de anticorpos (9 artigos). Estas temáticas foram escolhidas por estarem relacionadas entre si e, com isso, demonstrar o quão amplo pode ser a imunoinformática.

Os epítopos são a porção do antígeno reconhecida pelo sistema imune. Em decorrência disto, foram desenvolvidas ferramentas para predição destas estruturas, tanto para epítopos peptídicos (lineares e conformacionais) quanto para outras classes químicas. Grande parte das ferramentas revisadas neste presente estudo possuem em comum a implementação de algoritmos de aprendizado de máquina em suas metodologias como, por exemplo, RF, SVM, ERT e GB, indicando a importância da aplicação destes tipos de algoritmos em metodologias empregadas na área da imunoinformática. Outras abordagens também podem ser implementadas em ferramentas para a predição de epítopos, tais como os algoritmos do tipo rede neural profunda, que foram implementados nos preditores DRREP e DLBEpitope.

Embora estas ferramentas possuam metodologias semelhantes, diferenças podem ser observadas na constituição do conjunto de dados de treinamento como, por exemplo, o LBtope, que foi a primeira ferramenta a utilizar estruturas confirmadas experimentalmente, e o BepiPred 2.0, que utilizou estruturas cristalográficas para o treinamento. Destaca-se que a inovação também pode vir da escolha dos padrões para identificação dos epítopos, como o iLBE, que utiliza características evolutivas e de sequência para predição. Além disso, as ferramentas podem unir diferentes algoritmos de aprendizado de máquina, como o iBCE-EL, que une ERT e GB.

Outras ferramentas como, por exemplo, BepiTBR e BCEPS, além de realizarem a predição de epítopos de células B, também analisam o modo como os

epítopos interagem com o HLA, sendo este um processo que ocorre naturalmente *in vivo,* e que pode levar à predição de epítopos mais imunogênicos. Além disso, há ferramentas que exploram outros campos de pesquisa, como o NIgPred, que possibilita a predição de epítopos para outras classes de imunoglobulinas menos investigadas, como IgA e IgE, e o N-PEP, que utiliza uma abordagem inovadora de predição de epítopos não peptídicos.

Uma das principais limitações dos algoritmos de predição de epítopos conformacionais está na necessidade da estrutura tridimensional do antígeno. Em decorrência disto, foram desenvolvidos os preditores SePre e o SEPIa, no intuito de possibilitar tanto a utilização da sequência primária, quanto a estrutura tridimensional do antígeno como dados de entrada (*input*). Outra alternativa para solucionar este problema está no preditor EpiPred, o qual permite a utilização de estruturas modeladas como *input*, fornecendo uma opção alternativa quando a estrutura cristalográfica não está disponível.

Salienta-se que estas ferramentas de predição, tanto para epítopos lineares, quanto para os conformacionais, são generalistas, pois são desenvolvidas para a predição de “qualquer” epítopo para “qualquer” anticorpo. Em contrapartida, foram desenvolvidos preditores especializados tais como, por exemplo, o RF-OrgSpec e o PEAc. O preditor RF-OrgSpec utiliza epítopos de *Onchocerca volvulus*, vírus Epstein-Barr e o vírus da hepatite C como conjunto de dados de treinamento, obtendo maior precisão nos resultados de predição para estas espécies, quando comparado com ferramentas mais generalistas que usam epítopos e não epítopos sem restrição de espécie. Enquanto isto, o PEAc identifica epítopos através das informações acerca das interações entre os pares de aminoácidos da interface Ag-Ac, além de outras características de interface, sendo capaz de predizer epítopos específicos para um determinado anticorpo, ao contrário do que ocorre para as ferramentas generalistas, que por utilizarem apenas as características comuns para a seleção de epítopos em potencial, como por exemplo as propriedades físico-químicas e a acessibilidade ao solvente, e podem excluir acidentalmente candidatos à epítopos, caso os mesmos não compartilhem as mesmas características observadas em outros epítopos.

Vale ressaltar que frequentemente as ferramentas apresentam melhor desempenho durante os testes relatados em seus próprios artigos, quando comparadas com as demais ferramentas disponíveis na literatura. Entretanto, observa-se que este desempenho varia de acordo com o conjunto de dados utilizados para o teste. Portanto, durante a escolha da ferramenta que será utilizada em sua pesquisa específica, cada pesquisador deve levar em consideração as vantagens e desvantagens de cada metodologia de acordo com cada aplicação. Por exemplo, o Epitope-Analyser (EA) identifica epítopos conservados entre espécies, e que podem ser alvos de bNAbs. Porém, quando se necessita analisar grandes conjuntos de dados, o EA pode representar um dispendioso trabalho, pois esta ferramenta permite a análise de apenas uma estrutura frente a outra, de cada vez.

Em relação aos algoritmos para modelagem de anticorpos aqui revisados, percebe-se que 75% utilizam a metodologia de modelagem por homologia associada à modelagem *ab initio,* em diferentes etapas do processo. Isto ocorre, devido as estruturas dos anticorpos possuírem regiões com maior diversidade de comprimento e conformações, como a alça CDRH3, e também possuírem regiões com estruturas mais conservadas, como o FR das cadeias leve e pesada e as demais alças do CDR. Diante disso, a união das duas metodologias de modelagem permite explorar as vantagens e mitigar as limitações umas das outras, com o objetivo de modelar anticorpos o mais próximo possível de sua estrutura nativa, apesar do fato de que a precisão da modelagem do CDRH3 ainda se apresente como um desafio para estes *softwares* ou ferramentas.

Algoritmos como, por exemplo, RosettaAntibody e ABodyBuilder, utilizam bancos de fragmentos de regiões conservadas de anticorpos para identificar estruturas de sequência semelhante ao anticorpo a ser modelado, e assim utilizar esta conformação na modelagem. Entretanto, estes fragmentos são separados por comprimento, o que pode limitar as opções de fragmentos modelos. Enquanto isto, a ferramenta Sphinx possui um número maior de estruturas para serem utilizadas como modelo, uma vez que esta ferramenta não limita a busca aos bancos de dados separados por comprimento. Em outro exemplo, o algoritmo AbDesign utiliza dois bancos de dados correlacionados, PSSM e banco de torção da cadeia principal (*backbone*), no intuito de gerar estruturas de anticorpos mais realistas.

O foco das ferramentas de modelagem descritas até aqui está na predição da estrutura tridimensional nativa do anticorpo. Por outro lado, a ferramenta OptMAVEN consegue, a partir da estrutura do antígeno, selecionar porções do anticorpo através de um conjunto de dados MAPs que tenham afinidade com o antígeno, com o intuito de propor um novo anticorpo.

O *redesign* é uma nova abordagem que tem sido desenvolvida recentemente no campo da modelagem de anticorpos, tornando possível a modificação da estrutura de um anticorpo, com o intuito de melhorar sua afinidade pelo antígeno, o que tem sido alcançado através do protocolo IsAb. Ou então, o objetivo pode ser tornar um anticorpo específico para um antígeno que seja diferente do seu alvo inicial, gerando um novo anticorpo com praticamente as mesmas propriedades do anticorpo original, e isto tem sido alcançado com o uso do protocolo de Re-epitoping.

É importante destacar que dentre as 25 ferramentas revisadas, 13 são disponibilizadas através de servidores *on-line*, facilitando o acesso e a utilização por parte dos usuários, tanto os iniciantes na imunoinformática, quanto os que não possuem os recursos computacionais adequados para a execução destas metodologias.

Por fim, particularmente nos campos da predição de epítopos de células B e modelagem de anticorpos, observa-se que há uma ampla gama de ferramentas computacionais voltadas à área da imunoinformática. Entretanto, esta diversidade também dificulta a escolha da ferramenta mais adequada para ser empregada em um determinado projeto de pesquisa. Com isto, ao apresentar as vantagens e desvantagens de cada ferramenta, além do detalhamento de suas respectivas metodologias aqui revisadas, espera-se munir o pesquisador de instrumentos para analisar e escolher as ferramentas computacionais mais adequadas de acordo com as particularidades dos seus respectivos objetivos de pesquisa, em cada projeto individual.

REFERÊNCIAS

ABBOTT, W. Mark; DAMSCHRODER, Melissa M.; LOWE, David C.. Current approaches to fine mapping of antigen-antibody interactions. **Immunology**, [S.L.], v. 142, n. 4, p. 526-535, 11 jul. 2014. Wiley. http://dx.doi.org/10.1111/imm.12284. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4107663/. Acesso em: 19 set. 2022.

ADOLF-BRYFOGLE, Jared; KALYUZHNIY, Oleks; KUBITZ, Michael; WEITZNER, Brian D.; HU, Xiaozhen; ADACHI, Yumiko; SCHIEF, William R.; DUNBRACK, Roland L.. RosettaAntibodyDesign (RAbD): a general framework for computational antibody design. **Plos Computational Biology**, [S.L.], v. 14, n. 4, p. 0-0, 27 abr. 2018. Public Library of Science (PLoS). http://dx.doi.org/10.1371/journal.pcbi.1006112. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5942852/. Acesso em: 14 jun. 2022.

ASHFORD, Jodie; REIS-CUNHA, João; LOBO, Igor; LOBO, Francisco; CAMPELO, Felipe. Organism-specific training improves performance of linear B-cell epitope prediction. **Bioinformatics**, [S.L.], v. 37, n. 24, p. 4826-4834, 21 jul. 2021. Oxford University Press (OUP). http://dx.doi.org/10.1093/bioinformatics/btab536. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8665745/. Acesso em: 9 jun. 2022.

BRASIL. Ministério da Saúde. Secretaria de Ciência, Tecnologia e Insumos Estratégicos. Departamento de Ciência e Tecnologia. **Diretrizes metodológicas: elaboração de revisão sistemática e metanálise de ensaios clínicos randomizados**. Brasília, DF: Ministério da Saúde, 2012.

CARVALHO, Gardenia B. F. et al. Selecting targets for the diagnosis of Schistosoma mansoni infection: An integrative approach using multi-omic and immunoinformatics data. **Plos One**, [S. l.], v. 8, n. 8, p. 0-0, dez. 2017. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5560627/. Acesso em: 06 dez. 2021.

CHIANG, Ming-Hsien; SUNG, Wang-Chou; LIEN, Shu-Pei; CHEN, Ying-Zih; LO, Annie Fei-Yun; HUANG, Jui-Hsin; KUO, Shu-Chen; CHONG, Pele. Identification of novel vaccine candidates againstAcinetobacter baumanniiusing reverse vaccinology. **Human Vaccines & Immunotherapeutics**, [S.L.], v. 11, n. 4, p. 1065-1073, 9 mar. 2015. Informa UK Limited. http://dx.doi.org/10.1080/21645515.2015.1010910. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4514290/. Acesso em: 23 nov. 2021.

CHIU, Mark L. et al. Antibody Structure and Function: The Basis for Engineering Therapeutics. **Antibodies**, v. 8, n. 4, p. 55, 2019. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/31816964/>. Acesso em: 22 nov. 2021.

CHOWDHURY, Ratul; ALLAN, Matthew F.; MARANAS, Costas D.. OptMAVEn-2.0: de novo design of variable antibody regions against targeted antigen epitopes. **Antibodies**, [S.L.], v. 7, n. 3, p. 0-0, 30 jun. 2018. MDPI AG.

http://dx.doi.org/10.3390/antib7030023. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6640672/. Acesso em: 18 jun. 2022.

CRUVINEL, Wilson de Melo; MESQUITA JÚNIOR, Danilo; ARAÚJO, Júlio Antônio Pereira; CATELAN, Tânia Tieko Takao; SOUZA, Alexandre Wagner Silva de; SILVA, Neusa Pereira da; ANDRADE, Luís Eduardo Coelho. Sistema imunitário: parte i. fundamentos da imunidade inata com ênfase nos mecanismos moleculares e celulares da resposta inflamatória. **Revista Brasileira de Reumatologia**, [S.L.], v. 50, n. 4, p. 434-447, ago. 2010. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/s0482-50042010000400008. Disponível em: https://doi.org/10.1590/S0482-50042010000400008. Acesso em: 21 nov. 2021.

DALKAS, Georgios A.; ROOMAN, Marianne. SEPIa, a knowledge-driven algorithm for predicting conformational B-cell epitopes from the amino acid sequence. **Bmc Bioinformatics**, [S.L.], v. 18, n. 1, p. 0-0, 10 fev. 2017. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/s12859-017-1528-9. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5301386/. Acesso em: 7 jun. 2022.

DUNBAR, James; KRAWCZYK, Konrad; LEEM, Jinwoo; MARKS, Claire; NOWAK, Jaroslaw; REGEP, Cristian; GEORGES, Guy; KELM, Sebastian; POPOVIC, Bojana; DEANE, Charlotte M.. SAbPred: a structure-based antibody prediction server. **Nucleic Acids Research**, [S.L.], v. 44, n. 1, p. 474-478, 29 abr. 2016. Oxford University Press (OUP). http://dx.doi.org/10.1093/nar/gkw361. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4987913/. Acesso em: 16 jun. 2022.

HASAN, Md. Mehedi; KHATUN, Mst. Shamima; KURATA, Hiroyuki. ILBE for Computational Identification of Linear B-cell Epitopes by Integrating Sequence and Evolutionary Features. **Genomics, Proteomics & Bioinformatics**, [S.L.], v. 18, n. 5, p. 593-600, out. 2020. Elsevier BV. http://dx.doi.org/10.1016/j.gpb.2019.04.004. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8377379/. Acesso em: 9 jun. 2022.

JESPERSEN, Martin Closter; PETERS, Bjoern; NIELSEN, Morten; MARCATILI, Paolo. BepiPred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes. **Nucleic Acids Research**, [S.L.], v. 45, n. 1, p. 24-29, 2 maio 2017. Oxford University Press (OUP). http://dx.doi.org/10.1093/nar/gkx346. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5570230/. Acesso em: 7 jun. 2022.

KRAWCZYK, Konrad; LIU, Xiaofeng; BAKER, Terry; SHI, Jiye; DEANE, Charlotte M.. Improving B-cell epitope prediction and its application to global antibody-antigen docking. **Bioinformatics**, [S.L.], v. 30, n. 16, p. 2288-2294, 21 abr. 2014. Oxford University Press (OUP). http://dx.doi.org/10.1093/bioinformatics/btu190. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4207425/. Acesso em: 9 jun. 2022.

LAPIDOTH, Gideon D.; BARAN, Dror; PSZOLLA, Gabriele M.; NORN, Christoffer; ALON, Assaf; TYKA, Michael D.; FLEISHMAN, Sarel J.. AbDesign: an algorithm for combinatorial backbone design guided by natural conformations and sequences. **Proteins**: Structure, Function, and Bioinformatics, [S.L.], v. 83, n. 8, p.

1385-1406, 6 jun. 2015. Wiley. http://dx.doi.org/10.1002/prot.24779. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4881815/. Acesso em: 4 jul. 2022.

LEEM, Jinwoo; DUNBAR, James; GEORGES, Guy; SHI, Jiye; DEANE, Charlotte M.. ABodyBuilder: automated antibody structure prediction with data‐driven accuracy estimation. **Mabs**, [S.L.], v. 8, n. 7, p. 1259-1268, 30 ago. 2016. Informa UK Limited. http://dx.doi.org/10.1080/19420862.2016.1205773. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5058620/. Acesso em: 20 jun. 2022.

LEFRANC, Marie-Paule. Immunoglobulin and T Cell Receptor Genes: imgt® and the birth and rise of immunoinformatics. **Frontiers In Immunology**, [S.L.], v. 5, p. 0-0, 2014. Frontiers Media SA. http://dx.doi.org/10.3389/fimmu.2014.00022. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3913909/. Acesso em: 31 ago. 2022.

LIANG, Tianjian; CHEN, Hui; YUAN, Jiayi; JIANG, Chen; HAO, Yixuan; WANG, Yuanqiang; FENG, Zhiwei; XIE, Xiang-Qun. IsAb: a computational protocol for antibody design. **Briefings In Bioinformatics**, [S.L.], v. 22, n. 5, p. 0-0, 20 abr. 2021. Oxford University Press (OUP). http://dx.doi.org/10.1093/bib/bbab143. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8579189/. Acesso em: 20 jun. 2022.

LIU, Tao; SHI, Kaiwen; LI, Wuju. Deep learning methods improve linear B-cell epitope prediction. **Biodata Mining**, [S.L.], v. 13, n. 1, p. 0-0, 17 abr. 2020. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/s13040-020-00211-0. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7371472/. Acesso em: 7 jun. 2022.

MA, Lifei; LI, Huiyang; LAN, Jinping; HAO, Xiuqing; LIU, Huiying; WANG, Xiaoman; HUANG, Yong. Comprehensive analyses of bioinformatics applications in the fight against COVID-19 pandemic. **Computational Biology And Chemistry**, [S.L.], v. 95, n. 0, p. 107599, dez. 2021. Elsevier BV. http://dx.doi.org/10.1016/j.compbiolchem.2021.107599. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8560182/#sec0050title. Acesso em: 17 set. 2022.

MANAVALAN, Balachandran; GOVINDARAJ, Rajiv Gandhi; SHIN, Tae Hwan; KIM, Myeong Ok; LEE, Gwang. IBCE-EL: a new ensemble learning framework for improved linear b-cell epitope prediction. **Frontiers In Immunology**, [S.L.], v. 9, p. 0-0, 27 jul. 2018. Frontiers Media SA. http://dx.doi.org/10.3389/fimmu.2018.01695. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6072840/. Acesso em: 7 jun. 2022.

MARIANO, Diego. Métricas de avaliação em machine learning: acurácia, sensibilidade, precisão, especificidade e f-score. **Bioinfo - Revista Brasileira de Bioinformática e Biologia Computacional**, [S.L.], p. 0-0, jul. 2021. Alfahelix. http://dx.doi.org/10.51780/978-6-599-275326-15. Disponível em: https://bioinfo.com.br/metricas-de-avaliacao-em-machine-learning-acuracia-sensibilidade-precisao-especificidade-e-f-score/. Acesso em: 3 nov. 2022.

MARKS, C.; DEANE, C.M.. Antibody H3 Structure Prediction. **Computational And Structural Biotechnology Journal**, [S.L.], v. 15, p. 222-231, 2017. Elsevier BV.

http://dx.doi.org/10.1016/j.csbj.2017.01.010. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5312500/. Acesso em: 24 nov. 2021.

MARKS, Claire; NOWAK, Jaroslaw; KLOSTERMANN, Stefan; GEORGES, Guy; DUNBAR, James; SHI, Jiye; KELM, Sebastian; DEANE, Charlotte M.. Sphinx: merging knowledge-based andab initioapproaches to improve protein loop prediction. **Bioinformatics**, [S.L.], p. 0-0, 16 jan. 2017. Oxford University Press (OUP). http://dx.doi.org/10.1093/bioinformatics/btw823. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5408792/. Acesso em: 14 jun. 2022.

MARTÍ, Didac; MARTÍN-MARTÍNEZ, Eduard; TORRAS, Juan; BERTRAN, Oscar; TURON, Pau; ALEMÁN, Carlos. In silico antibody engineering for SARS-CoV-2 detection. **Computational And Structural Biotechnology Journal**, [S.L.], v. 19, n. 0, p. 5525-5534, 2021. Elsevier BV. http://dx.doi.org/10.1016/j.csbj.2021.10.010. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8496930/. Acesso em: 17 set. 2022.

MESQUITA JÚNIOR, Danilo; ARAÚJO, Júlio Antônio Pereira; CATELAN, Tânia Tieko Takao; SOUZA, Alexandre Wagner Silva de; CRUVINEL, Wilson de Melo; ANDRADE, Luís Eduardo Coelho; SILVA, Neusa Pereira da. Sistema imunitário - parte II: fundamentos da resposta imunológica mediada por linfócitos t e b. **Revista Brasileira de Reumatologia**, [S.L.], v. 50, n. 5, p. 552-580, out. 2010. Springer Science and Business Media LLC. http://dx.doi.org/10.1590/s0482-50042010000500008. Disponível em: https://doi.org/10.1590/S0482-50042010000500008. Acesso em: 21 nov. 2021.

MONTIEL-GARCIA, Daniel; ROJAS-LABRA, Oscar; SANTOYO-RIVERA, Nelly; REDDY, Vijay S.. Epitope-Analyzer: a structure-based webtool to analyze broadly neutralizing epitopes. **Journal Of Structural Biology**, [S.L.], v. 214, n. 1, p. 0-0, mar. 2022. Elsevier BV. http://dx.doi.org/10.1016/j.jsb.2022.107839. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8829422/. Acesso em: 7 jun. 2022.

MORETTI, Rocco; LYSKOV, Sergey; DAS, Rhiju; MEILER, Jens; GRAY, Jeffrey J.. Web-accessible molecular modeling with Rosetta: the rosetta online server that includes everyone (ROSIE). **Protein Science**, [S.L.], v. 27, n. 1, p. 259-268, 27 out. 2017. Wiley. http://dx.doi.org/10.1002/pro.3313. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5734271/. Acesso em: 23 jun. 2022.

NIMROD, Guy; FISCHMAN, Sharon; AUSTIN, Mark; HERMAN, Asael; KEYES, Feenagh; LEIDERMAN, Olga; HARGREAVES, David; STRAJBL, Marek; BREED, Jason; KLOMPUS, Shelley. Computational Design of Epitope-Specific Functional Antibodies. **Cell Reports**, [S.L.], v. 25, n. 8, p. 2121-2131, nov. 2018. Elsevier BV. http://dx.doi.org/10.1016/j.celrep.2018.10.081. Disponível em: https://bityli.com/TfFyxif. Acesso em: 7 jun. 2022.

PEDOTTI, Mattia; SIMONELLI, Luca; LIVOTI, Elsa; VARANI, Luca. Computational Docking of Antibody-Antigen Complexes, Opportunities and Pitfalls Illustrated by Influenza Hemagglutinin. **International Journal Of Molecular Sciences**, [S.L.], v. 12, n. 1, p. 226-251, 5 jan. 2011. MDPI AG. http://dx.doi.org/10.3390/ijms12010226. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3039950/. Acesso em: 22 ago. 2022.

RAOUFI, Ehsan; HEMMATI, Maryam; EFTEKHARI, Samane; KHAKSARAN, Kamal; MAHMODI, Zahra; FARAJOLLAHI, Mohammad M.; MOHSENZADEGAN, Monireh. Epitope Prediction by Novel Immunoinformatics Approach: a state-of-the-art review. **International Journal Of Peptide Research And Therapeutics**, [S.L.], v. 26, n. 2, p. 1155-1163, 20 ago. 2019. Springer Science and Business Media LLC. http://dx.doi.org/10.1007/s10989-019-09918-z. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7224030/. Acesso em: 22 nov. 2021.

RAS-CARMONA, Alvaro; PELAEZ-PRESTEL, Hector F.; LAFUENTE, Esther M.; RECHE, Pedro A.. BCEPS: a web server to predict linear b cell epitopes with enhanced immunogenicity and cross-reactivity. **Cells**, [S.L.], v. 10, n. 10, p. 2744-2750, 14 out. 2021. MDPI AG. http://dx.doi.org/10.3390/cells10102744. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8534968/. Acesso em: 9 jun. 2022.

REN, Jing; SONG, Jiangning; ELLIS, John; LI, Jinyan. Staged heterogeneity learning to identify conformational B-cell epitopes from antigen sequences. **Bmc Genomics**, [S.L.], v. 18, n. 2, p. 0-0, mar. 2017. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/s12864-017-3493-0. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5374683/#ack-1title. Acesso em: 7 ago. 2022.

SAMPAIO, Rf; MANCINI, Mc. Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. **Revista Brasileira de Fisioterapia**, [S.L.], v. 11, n. 1, p. 83-89, fev. 2007. FapUNIFESP (SciELO). http://dx.doi.org/10.1590/s1413-35552007000100013. Disponível em: https://www.scielo.br/j/rbfis/a/79nG9Vk3syHhnSgY7VsB6jG/?lang=pt&format=html. Acesso em: 29 ago. 2022.

SCHOEDER, Clara T.; SCHMITZ, Samuel; ADOLF-BRYFOGLE, Jared; SEVY, Alexander M.; FINN, Jessica A.; SAUER, Marion F.; BOZHANOVA, Nina G.; MUELLER, Benjamin K.; SANGHA, Amandeep K.; BONET, Jaume. Modeling Immunity with Rosetta: methods for antibody and antigen design. **Biochemistry**, [S.L.], v. 60, n. 11, p. 825-846, 11 mar. 2021. American Chemical Society (ACS). http://dx.doi.org/10.1021/acs.biochem.0c00912. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7992133/. Acesso em: 19 jun. 2022.

SELA-CULANG, Inbal; BENHNIA, Mohammed Rafii-El-Idrissi; MATHO, Michael H.; KAEVER, Thomas; MAYBENO, Matt; SCHLOSSMAN, Andrew; NIMROD, Guy; LI, Sheng; XIANG, Yan; ZAJONC, Dirk. Using a Combined Computational-Experimental Approach to Predict Antibody-Specific B Cell Epitopes. **Structure**, [S.L.], v. 22, n. 4, p. 646-657, abr. 2014. Elsevier BV. http://dx.doi.org/10.1016/j.str.2014.02.003. Disponível em: https://bit.ly/3TqetVU. Acesso em: 10 jun. 2022.

SHER, Gene; ZHI, Degui; ZHANG, Shaojie. DRREP: deep ridge regressed epitope predictor. **Bmc Genomics**, [S.L.], v. 18, n. 6, p. 0-0, out. 2017. Springer Science and Business Media LLC. http://dx.doi.org/10.1186/s12864-017-4024-8. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5629616/. Acesso em: 7 jun. 2022.

SILVA, Cristiane Nonato da. **Avaliação da Variabilidade de uma Biblioteca de Anticorpos construída a partir de sangue de pacientes com Hepatite C Crônica**

com Diferentes Graus de Fibrose. 2019. 70 f. Dissertação (Mestrado) - Curso de Medicina, Universidade Estadual Paulista, Botucatu, 2019. Disponível em: https://repositorio.unesp.br/bitstream/handle/11449/183345/silva_cn_me_bot_int.pdf?sequence=4&isAllowed=y. Acesso em: 20 set. 2022.

SINGH, Harinder; ANSARI, Hifzur Rahman; RAGHAVA, Gajendra P. S.. Improved Method for Linear B-Cell Epitope Prediction Using Antigen's Primary Sequence. **Plos One**, [S.L.], v. 8, n. 5, p. 0-0, 7 maio 2013. Public Library of Science (PLoS). http://dx.doi.org/10.1371/journal.pone.0062216. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3646881/. Acesso em: 7 jun. 2022.

SOUSA, Mayara de. **DESCRITORES MORFOLÓGICOS PARA IDENTIFICAÇÃO AUTOMÁTICA DE PADRÕES EPILEPTIFORMES EM SINAIS DE EEG**. 2016. 120 f. Tese (Doutorado) - Curso de Engenharia Elétrica, Universidade Federal de Santa Catarina, Florianópolis, 2016. Disponível em: https://repositorio.ufsc.br/bitstream/handle/123456789/168262/342002.pdf?sequence=1&isAllowed=y. Acesso em: 3 nov. 2022.

TILLER, Kathryn E.; TESSIER, Peter M.. Advances in Antibody Design. **Annual Review Of Biomedical Engineering**, [S.L.], v. 17, n. 1, p. 191-216, 7 dez. 2015. Annual Reviews. http://dx.doi.org/10.1146/annurev-bioeng-071114-040733. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5289076/. Acesso em: 22 nov. 2021.

TOMAR, Namrata; DE, Rajat K. Immunoinformatics: an integrated scenario. Immunology, [S. I.], v. 2, n. 2, p. 153-168, out. 2010. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2967261/. Acesso em: 06 dez. 2021.

TUNG, Chi-Hua; CHANG, Yi-Sheng; CHANG, Kai-Po; CHU, Yen-Wei. NIgPred: class-specific antibody prediction for linear b-cell epitopes based on heterogeneous features and machine-learning approaches. **Viruses**, [S.L.], v. 13, n. 8, p. 1531-1548, 3 ago. 2021. MDPI AG. http://dx.doi.org/10.3390/v13081531. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8402714/. Acesso em: 7 jun. 2022.

VIDARSSON, Gestur; DEKKERS, Gillian; RISPENS, Theo. IgG subclasses and allotypes: from structure to effector functions. **Frontiers in immunology**, v. 5, p. 520, 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4202688/>. Acesso em: 22 nov. 2021.

VITA, Randi; MAHAJAN, Swapnil; A OVERTON, James; DHANDA, Sandeep Kumar; MARTINI, Sheridan; CANTRELL, Jason R; WHEELER, Daniel K; SETTE, Alessandro; PETERS, Bjoern. The Immune Epitope Database (IEDB): 2018 update. **Nucleic Acids Research**, [S.L.], v. 47, n. 1, p. 339-343, 24 out. 2018. Oxford University Press (OUP). http://dx.doi.org/10.1093/nar/gky1006. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6324067/. Acesso em: 1 nov. 2022.

WEITZNER, Brian D; JELIAZKOV, Jeliazko R; LYSKOV, Sergey; MARZE, Nicholas; KURODA, Daisuke; FRICK, Rahel; ADOLF-BRYFOGLE, Jared; BISWAS, Naireeta; DUNBRACK, Roland L; GRAY, Jeffrey J. Modeling and docking of antibody structures with Rosetta. **Nature Protocols**, [S.L.], v. 12, n. 2, p. 401-416, 26 jan. 2017. Springer Science and Business Media LLC.

http://dx.doi.org/10.1038/nprot.2016.180. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5739521/. Acesso em: 9 ago. 2022.

YANG, Guocheng; VELGOS, Stefanie N.; BODDAPATI, Shanta P.; SIERKS, Michael R.. Probing Antibody-Antigen Interactions. **Microbiology Spectrum**, [S.L.], v. 2, n. 1, 17 jan. 2014. American Society for Microbiology. http://dx.doi.org/10.1128/microbiolspec.aid-0010-2013. Disponível em: https://journals.asm.org/doi/10.1128/microbiolspec.AID-0010-2013. Acesso em: 21 nov. 2021.

ZHU, James; GOURU, Anagha; WU, Fangjiang; BERZOFSKY, Jay A.; XIE, Yang; WANG, Tao. BepiTBR: t-b reciprocity enhances b cell epitope prediction. **Iscience**, [S.L.], v. 25, n. 2, p. 0-0, fev. 2022. Elsevier BV. http://dx.doi.org/10.1016/j.isci.2022.103764. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8803616/. Acesso em: 9 jun. 2022.

ZIEREP, Paul F.; VITA, Randi; BLAZESKA, Nina; MOUMBOCK, Aurélien F. A.; GREENBAUM, Jason A.; PETERS, Bjoern; GÜNTHER, Stefan. Towards the prediction of non-peptidic epitopes. **Plos Computational Biology**, [S.L.], v. 18, n. 2, p. 0-0, 18 fev. 2022. Public Library of Science (PLoS). http://dx.doi.org/10.1371/journal.pcbi.1009151. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8893639/. Acesso em: 9 jun. 2022.

APÊNDICE A - ARTIGOS EXCLUÍDOS DA REVISÃO EM ORDEM DE ANO DE PUBLICAÇÃO (continua)

Autores e ano de publicação	Título do artigo e DOI	Nome do algoritmo	Tema relacionado ao algoritmo	Motivo da exclusão
LO et al., 2013	Prediction of conformational epitopes with the use of a knowledge-based energy function and geometrically related neighboring residue characteristics (DOI: 10.1186/1471-2105-14-S4-S3)	CE-KEG	Predição de Epítopos	Servidor *on-line* indisponível (http://cekeg.cs.ntou.edu.tw/)
LI; PANTAZES; MARANAS, 2014	OptMAVEn – A New Framework for the de novo Design of Antibody Variable Region Models Targeting Specific Antigen Epitopes (DOI: 10.1371/journal.pone.0105954)	OptMAVEn	Modelagem de Anticorpo	Possui uma versão mais recente que foi incluída na revisão
LIAN; GE; PAN, 2014	EPMLR: sequence-based linear B-cell epitope prediction method using multiple linear regression (DOI: 10.1186/s12859-014-0414-y)	EPMLR	Predição de Epítopos	Servidor *on-line* indisponível (http://www.bioinfo.tsinghua.edu.cn/epitope/EPMLR/)
MARCATILI et al., 2014	Antibody modeling using the Prediction of ImmunoGlobulin Structure (PIGS) web server (DOI: 10.1038/nprot.2014.189)	Prediction of ImmunoGlobulin Structure (PIGS)	Modelagem de Anticorpo	Servidor *on-line* indisponível (http://www.biocomputing.it/pigs)

APÊNDICE A - ARTIGOS EXCLUÍDOS DA REVISÃO EM ORDEM DE ANO DE PUBLICAÇÃO (continua)

Autores e ano de publicação	Título do artigo e DOI	Nome do algoritmo	Tema relacionado ao algoritmo	Motivo da exclusão
ZHANG et al., 2014	Conformational B-Cell Epitopes Prediction from Sequences Using Cost-Sensitive Ensemble Classifiers and Spatial Clustering (DOI: 10.1155/2014/689219)	CBEP	Predição de Epítopos	Servidor *on-line* indisponível (http://59.73.198.144:8088/CBEP/)
BUJOTZEK et al., 2015	MoFvAb: Modeling the Fv region of antibodies (DOI: 10.1080/19420862.2015.1068492)	MoFvAb	Modelagem de Anticorpo	Uso comercial
SHEN et al., 2015	Predicting linear B-cell epitopes using amino acid anchoring pair composition (DOI: 10.1186/s13040-015-0047-3)	APCpred	Predição de Epítopos	Servidor *on-line* indisponível (http:/ccb.bmi.ac.cn/APCpred/)
KEMMISH; FASNACHT; YAN, 2017	Fully automated antibody structure prediction using BIOVIA tools: Validation study (DOI: 10.1371/journal.pone.0177923)	BIOVIA tools	Modelagem de Anticorpo	Uso comercial
LEPORE et al., 2017	PIGSPro: prediction of immunoGlobulin structures v2 (DOI: 10.1093/nar/gkx334)	PIGSPro	Modelagem de Anticorpo	Servidor *on-line* indisponível (https://www.biocomputing.it/pigspro)

APÊNDICE A - ARTIGOS EXCLUÍDOS DA REVISÃO EM ORDEM DE ANO DE PUBLICAÇÃO (conclusão)

Autores e ano de publicação	Título do artigo e DOI	Nome do algoritmo	Tema relacionado ao algoritmo	Motivo da exclusão
SWINDELLS et al., 2017	abYsis: Integrated Antibody Sequence and Structure—Management, Analysis, and Prediction (DOI: 10.1016/j.jmb.2016.08.019)	abYsis	Modelagem de Anticorpo	Se destina a outra área de imunoinformática (http://www.abysis.org/abysis/)
LAPIDOTH et al., 2018	AbPredict 2: A server for accurate and unstrained structure prediction of antibody variable domains (DOI: 10.1093/bioinformatics/bty822)	AbPredict 2	Modelagem de Anticorpo	Servidor não devolve os resultados (http://abpredict.weizmann.ac.il/bin/steps)
SAETHANG et al., 2018	AbDesigner3D: a structure-guided tool for peptide-based antibody production (DOI: 10.1093/bioinformatics/bty055)	AbDesigner3D	Modelagem de Anticorpo	Servidor *on-line* indisponível (http://sysbio.chula.ac.th/AbDesigner3D) ou (https://hpcwebapps.cit.nih.gov/AbDesigner3D/)
PRIHODA et al. 2022	BioPhi: A platform for antibody design, humanization, and humanness evaluation based on natural antibody repertoires and deep learning (DOI: 10.1080/19420862.2021.2020203)	BioPhi	Modelagem de Anticorpo	Se destina a outra área de imunoinformática (https://biophi.dichlab.org/)

Fonte: A autora (2022).