



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR**  
**CURSO DE GRADUAÇÃO EM BIOTECNOLOGIA**

**TAYNÁ DA SILVA FIÚZA**

**MAPEAMENTO DE EPÍTOPOS E DESIMUNIZAÇÃO DE UMA L-ASPARAGINASE  
DE *Bacillus subtilis* POR MEIO DE ALGORITMO GENÉTICO**

**FORTALEZA - CE**

**2017**

TAYNÁ DA SILVA FIÚZA

MAPEAMENTO DE EPÍTOPOS E DESIMUNIZAÇÃO DE UMA L-ASPARAGINASE DE  
*Bacillus subtilis* POR MEIO DE ALGORITMO GENÉTICO

Monografia apresentada ao Curso de Graduação em Biotecnologia do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de Bacharel em Biotecnologia.

Orientador: Dr. Raphael Trevizani

FORTALEZA - CE

2017

TAYNÁ DA SILVA FIÚZA

MAPEAMENTO DE EPÍTOPOS E DESIMUNIZAÇÃO DE UMA L-ASPARAGINASE DE  
*Bacillus subtilis* POR MEIO DE ALGORITMO GENÉTICO

Monografia apresentada ao Curso de Graduação em Biotecnologia do Centro de Ciências da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de Bacharel em Biotecnologia.

Aprovada em: 01 de Dezembro de 2017

BANCA EXAMINADORA

---

Dr. Raphael Trevizani (Orientador)  
Fundação Oswaldo Cruz (Fiocruz)

---

Profa. Dra. Carla Freire Celedonio Fernandes  
Fundação Oswaldo Cruz (Fiocruz)

---

Prof. Dr. Hermógenes David de Oliveira  
Universidade Federal do Ceará (UFC)

Às mulheres, cientistas ou não, que lutaram pra  
que eu chegasse aqui.

## AGRADECIMENTOS

À **Fundação Oswaldo Cruz** pelo apoio estrutural e recursos cedidos, ao **Centro de Integração Empresa-Escola** pelo fomento a este trabalho.

À **Universidade Federal do Ceará** e ao **Red River College** pelo suporte acadêmico e pelos incríveis profissionais que pude encontrar, com carinho especial aos **professores do curso de biotecnologia** e ao secretário **Gilmar Ferreira da Costa** pela confiança e por sempre me inspirarem a dar o meu melhor.

Ao orientador **Raphael Trevizani** pelas incontáveis discussões, ensinamentos e pela capacidade e disposição necessários para conduzir este projeto. Às queridas **Jossiely Bezerra** e **Nathalya Fedechen**, conversar com vocês sobre ciência e dar os primeiros passos desse trabalho juntas me ajudou mais do que podem imaginar. Aos colegas da Fiocruz Ceará pelos momentos de descontração e apreciação gastronômica, em especial à **Viviane Macedo** e **Thiago Ranieri** pelo companheirismo e cuidado que tiveram comigo. Aos pesquisadores da referida instituição pela paciência e pelas conversas que tem seus frutos também neste trabalho.

Aos amigos de curso e para a vida: **Fabrcia Diniz** pela nossa instantânea e duradoura conexão, **Igor Duarte** pelas risadas e o carinho que sempre tivemos um pelo outro, **Caroline Landim** pelas lembranças e aprendizado que trago - não apenas - das nossas viagens, **Marcus Rafael** por saber e ensinar a ser sério quando o momento exige e o completo oposto disso em todos os outros, **Lia Pinho** pelo carinho que cresceu entre nós e os puxões de orelha que vieram juntos, **Débora Albuquerque** por todo o apoio incansável e pela confiança, **Alison Rebouças** pelas caronas e bergamotas.

Aos amigos de tanto e tão pouco tempo: **Vanessa Carneiro** pela paciência e por ser abrigo todos esses anos, **Lucas Cordeiro** por estar sempre buscando saber mais e por me ensinar tanto nesse caminho, **Stefany Monteiro** porque os astros nos escreveram com a mão esquerda, **Morgana Feijão** por ser a luz mais bonita que já conheci e **Brenda Sombra** pela cumplicidade, o aconchego e as horas gastas em receitas que nunca funcionariam. À **Jéssica Gabrielle** pelo cosmos que partilhamos e às irmãs, **Anna** e **Ingrid**, pelo carinho e apoio.

Aos meus familiares, em especial minha mãe, **Francisca Alves** pelo inesgotável amor e ao meu pai **Antonio Edilson** pela fiel dedicação, à minha irmã **Taynan Fiúza** pela paciência e por ser fonte inspiração todos os dias, à **Aurora** e **Lunar** pelo carinho cotidiano e pela alegria que me proporcionam.

A Ednardo M. Rodrigues e Alan B. Oliveira pelo *template* L<sup>A</sup>T<sub>E</sub>X fornecido.

“A ciência e a vida cotidiana não podem nem  
devem ser separadas.”

(Rosalind Franklin)

## RESUMO

A L-asparaginase é uma proteína terapêutica usada no tratamento de leucemia linfoblástica aguda cuja utilização pode resultar em sua neutralização pelo sistema imune ou mesmo gerar reações de hipersensibilidade. Componentes da imunidade adaptativa como o MHC-II (Complexo Principal de Histocompatibilidade II) são responsáveis por apresentar peptídeos exógenos processados por células apresentadoras de antígenos. Ferramentas de imunoinformática podem ser usadas para prever a afinidade de ligação de um peptídeo-fenda do MHC-II de forma a localizar epítomos, assim como realizar desimunização por deleção de epítomos. Essas tecnologias podem ser empregadas na geração de *biobetters*, contudo as ferramentas de predição disponíveis não permitem a identificação de aminoácidos promíscuos. O objetivo do trabalho foi desenvolver ferramentas computacionais para desimunização de proteínas, melhorando uma ferramenta de predição de imunogenicidade *in silico*, construindo um algoritmo genético (AG) de desimunização de proteínas e posteriormente aplicando estas ferramentas para identificar e remover os epítomos mais imunogênicos de uma sequência de L-asparaginase de *Bacillus subtilis*. O método de predição de afinidade peptídeo-MHC-II TEPITOPE teve seu desempenho melhorado pelo somatório dos escores de imunogenicidades de epítomos sobrepostos e suavização desses dados através de uma *spline*. Tal protocolo foi aplicado para identificar regiões imunogênicas em uma L-asparaginase de *Bacillus subtilis* (BsA) e essas informações foram usadas em um AG de desimunização de proteínas. O AG utiliza um conjunto de 100 sequências de BsA e realiza operações de recombinação e mutação a cada ciclo apenas nas regiões cujos escores de imunogenicidade estão entre os 20% maiores da proteína e considerando apenas mutações neutras ( $|\Delta\Delta G|$  entre proteína selvagem e mutante é menor que 0.5 kcal/mol), substituindo sequências de maior imunogenicidade por outras de menor. Ao fim de ciclos de execução, o AG foi capaz de, com apenas três mutações, gerar uma proteína cuja imunogenicidade é 63.5% menor que a da sequência original, tendo dois dos epítomos mais imunogênicos removidos, mantendo a estrutura e estabilidade intactas (RMSD=0.01 Å e  $\Delta\Delta G = -0.2935$  kcal/mol). O AG de desimunização permite parametrização e a implementação de novas avaliações de imunogenicidade e estabilidade que aumentem sua eficiência. O método de predição por suavização, o código de visualização dos epítomos e o AG podem ser aplicados de maneira automatizada a um grande número de cadeias aminoacídicas auxiliando trabalhos de engenharia de proteínas.

**Palavras-chave:** Desimunização. Algoritmo Genético. L-asparaginase. *Biobetter*.

## ABSTRACT

L-asparaginase is a therapeutic protein used in the treatment of acute lymphoblastic leukemia, and its use may result in neutralization by the immune system or even produce hypersensitivity reactions. Components of adaptive immunity such as MHC-II (Major Histocompatibility Complex II) are responsible for presenting exogenous peptides processed by antigen-presenting cells. Immunoinformatics tools can be used to predict the binding affinity of a peptide with the MHC-II cleft to localize epitopes, as well as perform epitope deletion deimmunization. These technologies may be employed in the generation of biobetters, however, the predictive tools available do not allow the identification of promiscuous amino acids. The aim of this work was to develop computational tools for protein deimmunization, improve an in silico immunogenicity prediction tool, construct a protein deimmunization genetic algorithm (GA) and later apply these tools to identify and remove the most immunogenic epitopes of a *Bacillus subtilis* L-asparaginase sequence. The TEPITOPE peptide-MHC-II affinity prediction method had its performance improved by summation of immunogenicity scores of overlapping epitopes and smoothing of these data through a spline. This protocol was applied to identify immunogenic regions in a *Bacillus subtilis* L-asparaginase (BsA) and this information was used by a protein deimmunization genetic algorithm. The algorithm uses a set of 100 BsA sequences and performs recombination and mutation operations at each cycle only in regions whose immunogenicity scores are among the top 20% of the protein and considering only neutral mutations ( $|\Delta\Delta G|$  between wild-type and mutant proteins is less than 0.5 kcal/mol), substituting sequences of higher immunogenicity for those with lower ones. At the end of execution cycles, the algorithm was able to, with only three mutations, generate a protein whose immunogenicity is 63.5 % lower than that of the original sequence, having two of the most immunogenic epitopes removed, maintaining structure and stability intact (RMSD=0.01Å and  $\Delta\Delta G = -0.2935$  kcal/mol)). The protein deimmunization genetic algorithm allows parameterization and implementation of new immunogenicity and stability assessments that may increase its efficiency. The smoothing prediction method, the epitope visualization script and the genetic algorithm can be applied in an automated manner to a large number of amino acid chains aiding protein engineering projects.

**Keywords:** Deimmunization. Genetic Algorithm. L-asparaginase. Biobetter.

## LISTA DE FIGURAS

Figura 1 – Representação de modelo de superfície de uma molécula de MHC-II com destaque para a fenda . . . . .	17
Figura 2 – Ativação de células B por vias dependente e independente de célula T . . . .	18
Figura 3 – Representação hipotética de epítomos linear e conformacional em um peptídeo antimicrobiano . . . . .	19
Figura 4 – Fatores de risco associados à imunogenicidade de proteínas terapêuticas . .	20
Figura 5 – Janelas de leitura extraídas de uma sequência . . . . .	22
Figura 6 – Estratégias de deleção de epítomos de célula T . . . . .	25
Figura 7 – Linha do tempo da distribuição de L-asparaginase no Brasil . . . . .	30
Figura 8 – Exemplo de um arquivo do IEDB . . . . .	32
Figura 9 – Representação do arquivo de saída do método TEPITOPE . . . . .	35
Figura 10 – Soma dos escores para cada posição . . . . .	36
Figura 11 – Variação da spline de acordo com o número de nós utilizados . . . . .	37
Figura 12 – Método de contagem de erros . . . . .	38
Figura 13 – População em um algoritmo Genético de Desimunização . . . . .	39
Figura 14 – $\Delta G$ em kcal/mol de duas estruturas de acordo com o FoldX . . . . .	41
Figura 15 – Exemplo de mutação considerando o conjunto de mutações estáveis . . . . .	42
Figura 16 – Exemplo de operação e <i>crossover</i> realizada entre duas sequências A e B . .	43
Figura 17 – Seleção proporcional a aptidão . . . . .	44
Figura 18 – Substituição de sequências . . . . .	45
Figura 19 – Fluxograma de funcionamento do AG . . . . .	46
Figura 20 – Gradiente de imunogenicidade utilizado para visualização no VMD . . . . .	47
Figura 21 – Predição do TEPITOPE para a proteína P18884.1 . . . . .	48
Figura 22 – Epítomos identificados pelo TEPITOPE para a L-asparaginase de <i>Bacillus subtilis</i> . . . . .	49
Figura 23 – Distribuição de Erros no Mapeamento de Epítomos . . . . .	50
Figura 24 – Predição do somatório para a proteína P18884.1 . . . . .	50
Figura 25 – Predição da suavização do somatório para a proteína P18884.1 . . . . .	51
Figura 26 – Erros no Mapeamento de Epítomos . . . . .	52
Figura 27 – Predição da suavização somatório para a proteína P18884.1 . . . . .	52
Figura 28 – Posições escolhidas para mutação e mutações mantidas na população . . . .	54

Figura 29 – Evolução do Algoritmo Genético de acordo com a imunogenicidade do melhor indivíduo a cada geração . . . . .	55
Figura 30 – Sobreposição das estruturas de L-asparaginase de <i>B. subtilis</i> selvagem e mutada	56
Figura 31 – Mapeamento de epítomos em L-asparaginase de <i>B. subtilis</i> mutada em duas posições . . . . .	56

## LISTA DE QUADROS

Quadro 1 – Principais Métodos de Predição de Epítomos Lineares - Ligação ao MHC-II	23
Quadro 2 – Alelos do HLA utilizados pelo TEPITOPE . . . . .	33
Quadro 3 – Matriz utilizada pelo TEPITOPE para calcular a afinidade peptídeo-HLA DRB1*0101 . . . . .	34
Quadro 4 – Vinte primeiras sequências da população final . . . . .	57

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>14</b>
<b>1.1</b>	<b>Proteínas terapêuticas</b>	<b>14</b>
<b>1.2</b>	<b>Sistema imune</b>	<b>15</b>
<b>1.2.1</b>	<b><i>Imunogenicidade</i></b>	<b>19</b>
<b>1.3</b>	<b>Métodos de Predição de Imunogenicidade <i>in silico</i></b>	<b>20</b>
<b>1.4</b>	<b>Desimunização de proteínas <i>in silico</i></b>	<b>24</b>
<b>1.5</b>	<b>L-Asparaginase</b>	<b>27</b>
<b>2</b>	<b>OBJETIVO GERAL</b>	<b>31</b>
<b>2.1</b>	<b>Objetivos específicos</b>	<b>31</b>
<b>3</b>	<b>METODOLOGIA</b>	<b>32</b>
<b>3.1</b>	<b>Predição da Imunogenicidade</b>	<b>32</b>
<b>3.1.1</b>	<b><i>Obtenção dos dados</i></b>	<b>32</b>
<b>3.1.2</b>	<b><i>Uso do TEPITOPE para a predição de epítomos lineares</i></b>	<b>33</b>
<b>3.1.3</b>	<b><i>Destaque dos aminoácidos promíscuos</i></b>	<b>35</b>
<b>3.1.4</b>	<b><i>Discernimento de regiões imunogênicas</i></b>	<b>36</b>
<b>3.2</b>	<b>Avaliação da Predição de Epítomos</b>	<b>37</b>
<b>3.3</b>	<b>Algoritmo genético para desimunização de proteínas</b>	<b>38</b>
<b>3.3.1</b>	<b><i>Conceito</i></b>	<b>38</b>
<b>3.3.2</b>	<b><i>Indivíduo e População</i></b>	<b>39</b>
<b>3.3.3</b>	<b><i>Aptidão</i></b>	<b>39</b>
<b>3.3.4</b>	<b><i>Restrições</i></b>	<b>40</b>
<b>3.3.5</b>	<b><i>Operadores Genéticos</i></b>	<b>41</b>
<b>3.3.5.1</b>	<b><i>Mutação</i></b>	<b>42</b>
<b>3.3.5.2</b>	<b><i>Recombinação</i></b>	<b>42</b>
<b>3.3.5.3</b>	<b><i>Seleção parental</i></b>	<b>43</b>
<b>3.3.6</b>	<b><i>Substituição parental</i></b>	<b>44</b>
<b>3.3.7</b>	<b><i>Convergência e População Final</i></b>	<b>45</b>
<b>3.3.8</b>	<b><i>Funcionamento do AG</i></b>	<b>45</b>
<b>3.4</b>	<b>Softwares Utilizados</b>	<b>46</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>48</b>

<b>4.1</b>	<b>Avaliação da Imunogenicidade</b> . . . . .	48
<b>4.1.1</b>	<b><i>Predição do TEPITOPE</i></b> . . . . .	48
<b>4.1.2</b>	<b><i>Avaliação do Somatório</i></b> . . . . .	49
<b>4.1.3</b>	<b><i>Avaliação da Suavização</i></b> . . . . .	51
<b>4.2</b>	<b>Desimunização</b> . . . . .	53
<b>4.3</b>	<b>Ferramentas Desenvolvidas</b> . . . . .	58
<b>5</b>	<b>CONCLUSÕES</b> . . . . .	59
	<b>REFERÊNCIAS</b> . . . . .	60

# 1 INTRODUÇÃO

Proteínas são macromoléculas formadas por uma ou mais cadeias polipeptídicas, em que cada cadeia é um polímero linear de aminoácidos (WALSH, 2002). Essas moléculas desempenham funções estruturais, de transporte, proteção, regulação ou catálise (MÁRQUEZ-CHAMORRO *et al.*, 2015). Devido à versatilidade de suas aplicações, possuem grande importância econômica.

## 1.1 Proteínas terapêuticas

Desde a década de 1980 as proteínas terapêuticas se estabeleceram como uma nova classe de fármacos com cerca de 264 produtos no mercado (WALSH, 2014). As primeiras proteínas terapêuticas foram versões recombinantes de proteínas encontradas na natureza. Depois dessa fase, o passo seguinte foi o de engenharia de proteínas para melhora de seu potencial clínico (CARTER, 2011).

Geralmente, são necessários de dez a quinze anos de pesquisa e desenvolvimento para a comercialização de um novo biofármaco e os custos associados a esse processo chegaram a cerca de US\$ 53 milhões em 2014, nos Estados Unidos (IFPMA, 2017). Esse elevado custo de investimento é, entretanto, fartamente superado pelo lucro que proporciona: atualmente, proteínas terapêuticas são os biológicos mais importantes em termos de uso clínico e, em 2015, o mercado mundial de proteínas chegou a US\$174.7 bilhões (DEWAN; SULLIVAN, 2016).

Proteínas terapêuticas podem ser separadas em quatro grupos de acordo com suas aplicações e funções (LEADER *et al.*, 2008), sendo eles:

- i Proteínas com atividade enzimática ou regulatória que atuam trazendo uma nova atividade, impulsionando uma via metabólica ou substituindo uma proteína deficiente ou anormal. Exemplos: insulina, somatotropina, beta-gluco cerebrosidase, antitrombina III, proteínas relacionadas às desordens endócrinas, hematológicas e trombooses;
- ii Proteínas com alvos específicos que atuam ligando-se à moléculas específicas e interferindo com sua função, sinalizando para a sua eliminação, transportando compostos para regiões específicas. Este grupo pode ser exemplificado pelo rituximab, infliximab, enfuvirtide e ibritumomab tiuxetan;
- iii Proteínas que atuam como vacinas com a função de proteger contra agentes externos, tratar doenças autoimunes e câncer, dentre as quais podem-se citar o antígeno de superfície

da hepatite B, imunoglobulinas G anti-rhesus e proteína capsídeo L1 do papilomavírus humano;

- iv Proteínas com função diagnóstica e que não são utilizadas para tratar enfermidades, como o glucagon, nofetumomab e antígenos da hepatite C.

Atualmente, proteínas são as moléculas mais importantes em termos de aplicação clínica e seu sucesso se deve principalmente à segurança e eficácia (DIMITROV, 2012), mas também a interações pós-administração mais previsíveis, menos efeitos colaterais, tempo de aprovação mais rápido e singularidade que facilita o patenteamento dessas moléculas em relação a fármacos mais simples como pequenas moléculas (BALDO, 2016). Contudo, respostas imunes às proteínas terapêuticas são responsáveis por um número considerável de insucessos de testes clínicos para novas drogas (RAU *et al.*, 2003) ou até mesmo pela retirada de medicamentos do mercado (CASADEVALL *et al.*, 2002).

O desenvolvimento da resposta imune à proteína é um processo complexo afetado pelas interações entre todos esses fatores, sendo importante entender como funcionam os atores dessa sequência e como eles atuam entre si a fim de minimizar reações indesejadas.

## 1.2 Sistema imune

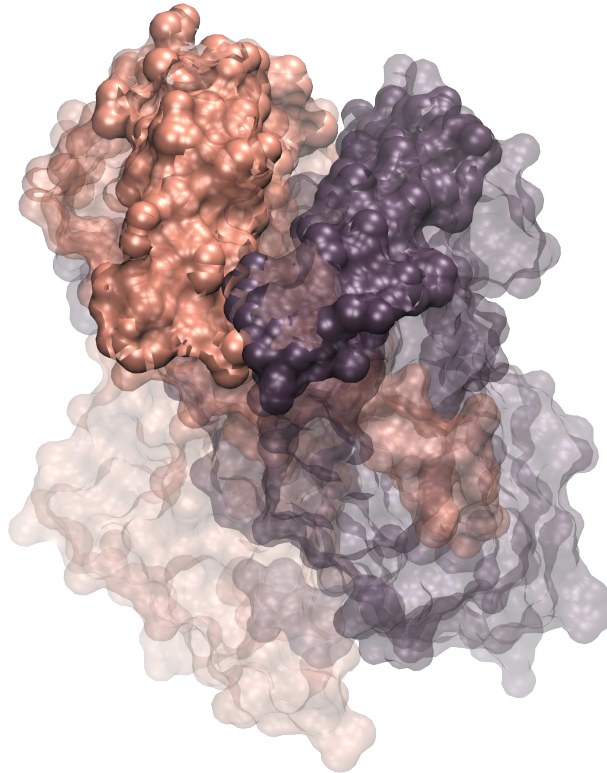
O sistema de defesa de organismos vertebrados, denominado sistema imune, é composto de uma rede complexa de órgãos, células, macromoléculas e citocinas (PARKIN; COHEN, 2001). Todos os animais possuem uma versão ancestral desse sistema, em que moléculas como peptídeos antimicrobianos e células como fagócitos são responsáveis por realizar uma resposta menos específica, porém efetiva contra ameaças. Essa versão é denominada imunidade inata e tem como maior vantagem sua ação rápida, não contando, porém, com uma grande diversidade de respostas ou memórias, isso é, capacidade de reagir com maior rapidez e especificidade a uma ameaça previamente combatida (ABBAS *et al.*, 2014).

A imunidade adquirida é a parte do sistema imune capaz de reconhecer moléculas específicas com as quais o corpo já esteve em contato devido sua característica de memória, produzindo uma proteção imune duradoura e com respostas rápidas. Esta seção da imunidade pode ser dividida em celular e humoral (BACKERT; KOHLBACHER, 2015). As moléculas reconhecidas pelo sistema imune são substâncias capazes de gerar respostas ao se ligarem à moléculas imunologicamente ativas, os antígenos. As porções determinantes desses antígenos são chamadas epítopos (CRUSE *et al.*, 2004).

Na imunidade celular os epítomos são apresentados pelo *Major Histocompatibility Complex* - Complexo de Histocompatibilidade Principal (MHC) e reconhecidos pelas células T (BACKERT; KOHLBACHER, 2015). O MHC de classe II apresenta peptídeos fagocitados do meio extracelular e estimula as imunidades celulares e humorais contra micro-organismos por intermédio de linfócitos T auxiliares. Para que um peptídeo seja capaz de estimular uma resposta celular advinda de linfócitos T auxiliares, o peptídeo deve se ligar ao MHC II em compartimentos da célula apresentadora de antígeno (CASTELLINO *et al.*, 1997), tendo, então, uma certa afinidade de ligação, que é o aspecto mais conhecido da geração de epítomos de célula T (KORBER *et al.*, 2006).

O processamento e apresentação de peptídeos-epítomo são essenciais para a imunidade mediada por células e os MHC são codificados por um complexo genético denominado HLA, localizado no cromossomo seis da espécie humana e que possui 21 genes altamente polimórficos (BACKERT; KOHLBACHER, 2015). Apesar desses polimorfismos, a maioria das diferenças entre moléculas da mesma classe estão em posições da fenda de ligação, que são responsáveis pela restrição das interações MHC-peptídeo (KORBER *et al.*, 2006, Figura 1).

Figura 1 – Representação de modelo de superfície de uma molécula de MHC-II com destaque para a fenda

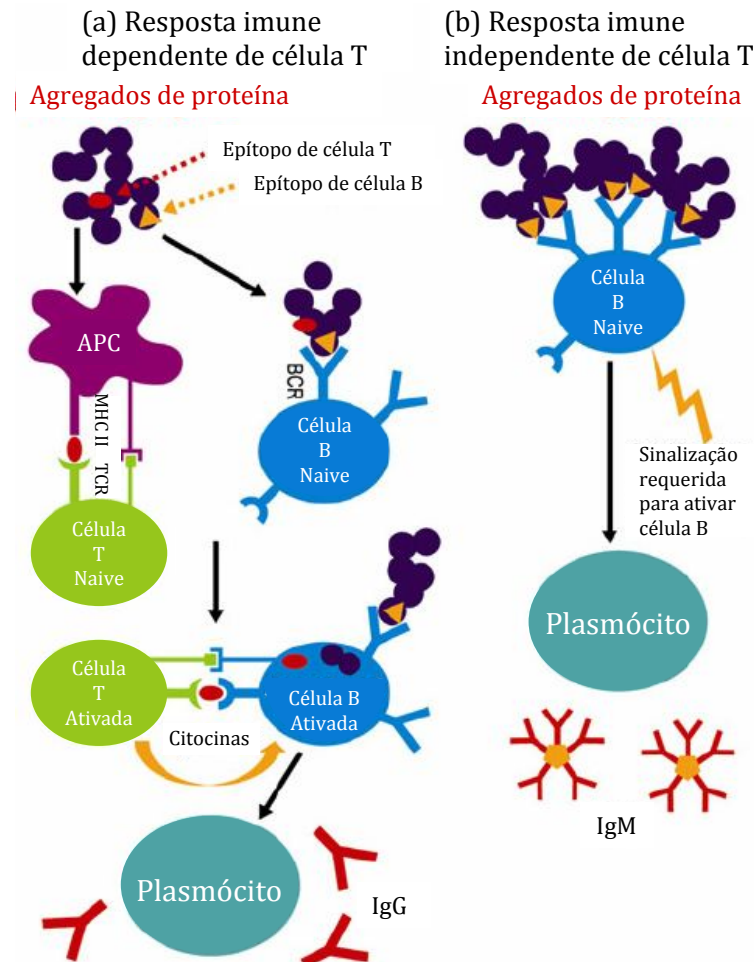


Fonte: Elaborada pela autora (estrutura obtida do PDB: 3PDO). A molécula de MHC de classe II possui duas cadeias,  $\alpha$  (em roxo) e  $\beta$  (em rosa) cada uma com dois domínios ( $\alpha 1$ ,  $\alpha 2$ ,  $\beta 1$ ,  $\beta 2$ ). Os domínios  $\alpha 1$  e  $\beta 1$  contém a fenda do MHC-II, a porção que interage com o peptídeo produzido na célula apresentadora de antígeno e que forma a sinapse imunológica com o receptor de célula T. Nessa figura as porções  $\alpha 1$  e  $\beta 1$  estão em roxo e rosa mais opacos, respectivamente.

O HLA-II possui os loci DR, DP, DQ, DM, DOA e DOB. As fendas de ligação das moléculas codificadas por esses genes são capazes de interagir com motivos de ligação específicos (SETTE *et al.*, 1990) e o conhecimento acerca das fendas codificadas por esses genes polimórficos e suas frequências são de extrema importância para realizar a predição dos peptídeos que serão apresentados pelos MHCs aos linfócitos T (BACKERT; KOHLBACHER, 2015).

Na imunidade humoral, o reconhecimento de antígenos e seus epítomos é, em último estágio, realizado pelas células B (BACKERT; KOHLBACHER, 2015) e a ativação dessa forma de imunidade pode se dar por vias dependentes e independentes de linfócitos T (Figura 2). Epítomos de célula B são definidos como as regiões de antígenos que se ligam a anticorpos livres no plasma ou receptores de membrana e podem ser lineares ou conformacionais.

Figura 2 – Ativação de células B por vias dependente e independente de célula T

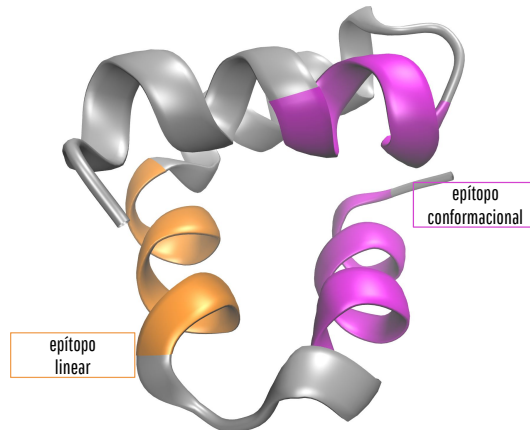


Fonte: Traduzida de (KUMAR *et al.*, 2011)

(a) Na via dependente de célula T, células T auxiliares ativadas pela interação com células apresentadoras de antígeno estimulam a proliferação e posterior diferenciação de células B ligadas ao antígeno. (b) Na via independente, antígenos são capazes de induzir uma resposta rápida, porém, menos específica.

Epítopos lineares (ou contínuos) são quaisquer fragmentos reconhecidos na estrutura primária da proteína capazes de gerar resposta imune (Figura 3, REGENMORTELL 2009). Em contraposição, epítopos conformacionais são porções do arranjo tridimensional da proteína e são formados por segmentos de aminoácidos que estão dispostos de forma descontínua na sequência (ANSARI; RAGHAVA, 2010). Dessa forma, o epítipo conformacional só se manifesta após o enovelamento da proteína nas estruturas terciária ou quaternária. Todos os epítopos apresentados às células T pelo MHC são lineares, e pelo menos 90% dos epítopos de célula B são conformacionais (REGENMORTELL, 1996; BLYTHE; FLOWER, 2005). A presença desses epítopos em proteínas terapêuticas são responsáveis pela imunogenicidade e reações de hipersensibilidade posteriormente geradas.

Figura 3 – Representação hipotética de epítomos linear e conformacional em um peptídeo antimicrobiano



Fonte: Elaborado pela autora (estrutura obtida do PDB: 2N8O).

Aqui estão representados dois epítomos, um linear em laranja e um conformacional cujos dois trechos descontínuos estão representados em rosa.

### 1.2.1 *Imunogenicidade*

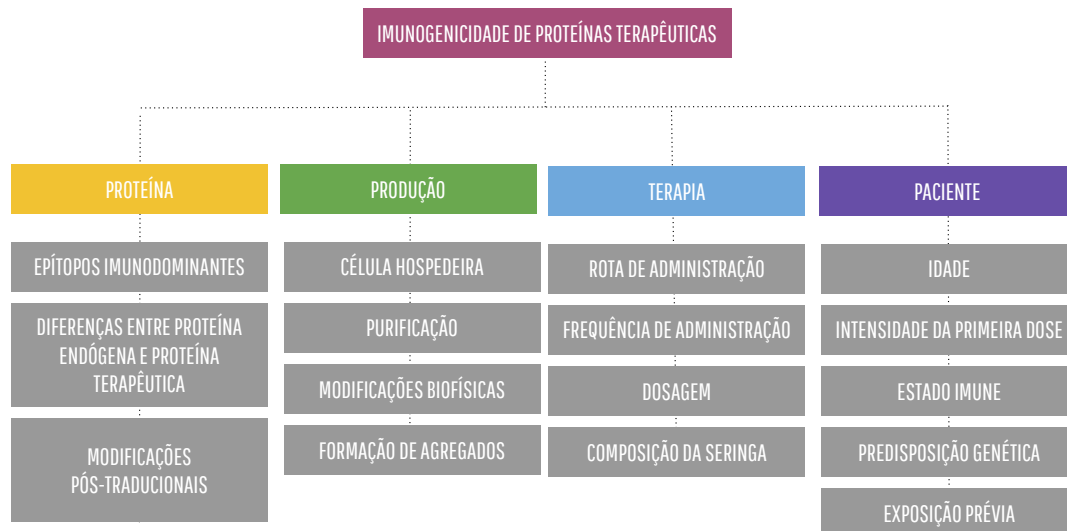
Imunogenicidade é a capacidade que tem uma molécula (proveniente de um patógeno ou não) de induzir uma resposta imune específica quando exposta a componentes do sistema imune (FLOWER, 2007). As respostas imunogênicas podem gerar anticorpos neutralizantes antifármacos que se ligam a sítios específicos da molécula de modo a diminuir ou eliminar sua função biológica (MUKOVOZOV *et al.*, 2008) pela coestimulação de células B por células T-CD4+ ou independentemente de coestimulação (JAWA *et al.*, 2013). Sendo assim, todos os biológicos, inclusive as proteínas terapêuticas, são capazes de induzir respostas imunes indesejadas (SCHELLEKENS, 2002).

Uma proteína terapêutica imunogênica é capaz de comprometer a eficácia do tratamento, podendo levar a efeitos adversos como reações anafiláticas, síndrome da liberação de citocinas, neutralização de proteínas endógenas via reação cruzada e desenvolvimento de mecanismos autoimunes (BAKER *et al.*, 2010), o que já levou ao cancelamento do financiamento de desenvolvimento de biofármacos (FOOD *et al.*, 2013) e à substituição de anticorpos monoclonais murinos por versões humanizadas e humanas (DIMITROV, 2012).

As causas para a indução da imunogenicidade são multifatoriais e algumas não são completamente entendidas (DEEHAN *et al.*, 2015), mas pode-se associar os fatores de imunogenicidade à natureza da proteína, ao paciente ou ao processo de produção. Dentre os elementos de risco associados à natureza da proteína estão a presença de epítomos imunodominantes, as dife-

renças entre a proteína terapêutica utilizada e sua versão endógena (se presente) e seus padrões de glicosilação, oxidação, deamidação, acilação, alquilação, entre outros. Além disso, existem fatores relacionados à terapia, em si, que incluem a rota utilizada e a frequência de administração da molécula, além de sua dosagem e da composição da seringa (seringas confeccionadas com tungstênio tendem a gerar maior agregação de proteínas) (JAWA *et al.*, 2013). Há também fatores ligados ao paciente e que podem comprometer a terapia: idade, predisposição genética, exposição prévia a terapêuticos similares e estado imune do indivíduo (KRISHNA; NADLER, 2016). Por fim, dentro do processo de produção, as causas são: a célula hospedeira utilizada e as diversas etapas de purificação que podem ocasionar modificações biofísicas e a formação de agregados (JAWA *et al.*, 2013, Figura 4).

Figura 4 – Fatores de risco associados à imunogenicidade de proteínas terapêuticas



Fonte: Elaborada pela autora

A imunogenicidade de proteínas terapêuticas pode ser atribuída a muitos fatores que muitas vezes não estão presentes sozinhos. Além dos mencionados na figura, fatores como a predisposição genética do paciente e sua exposição prévia a um terapêutico similar.

### 1.3 Métodos de Predição de Imunogenicidade *in silico*

A identificação de epítopos se torna uma tarefa importante tanto para a pesquisa básica quanto para a clínica, seja pela produção de vacinas, prevenção e diagnóstico de doenças, tratamento de enfermidades ou pela desimunização de proteínas. Até a década de 1990, técnicas dispendiosas e demoradas como radioimunoensaios, imunohistoquímica, ELISA, imunofluorescência, difração de raio-X em cristais e diversas outras eram as únicas alternativas para a avaliação da imunogenicidade de epítopos (YANG; YU, 2009).

O acúmulo de informações referentes ao sequenciamento do genoma humano e de outros organismos, dados de proteômica de diversas espécies, e o crescente número de descobertas epidemiológicas e clínicas relevantes para a pesquisa em imunologia levaram ao surgimento de bancos de dados para organizar essas informações, e as tecnologias *high-throughput* permitiram esse rápido aumento no número de dados disponíveis à comunidade científica. Em outubro de 2017, 203953682 sequências se encontram depositadas no GenBank, em oposição a 424084 epítomos peptídicos depositados no Imune Epitope Database - Banco de Dados de Epítomos Imunes (IEDB), sendo parte desses epítomos procedente de uma mesma proteína. Mesmo se todos esses epítomos correspondessem a sequências diferentes, restariam 99.79% das sequências depositadas sem algum dado imunológico experimental que permitisse a exploração desses dados. Desse modo, se fez necessário o surgimento de ferramentas de análise *in silico* para estimar e prever informações acerca das sequências de imunogenicidade desconhecida e filtrar esses dados a fim reduzir tempo e custo das pesquisas desenvolvidas (DE; TOMAR, 2014).

Há três abordagens gerais para a predição de epítomos: métodos baseados em sequência, métodos baseados na estrutura e métodos de consenso. Esses métodos exploram matrizes buscando sequências-motivo, explorando a relação quantitativa estrutura-atividade, simulações de *docking* molecular, redes neurais artificiais, máquinas de vetores de suporte e outros (DESAI; KULKARNI-KALE, 2014). Especificamente, os principais métodos de predição utilizados para identificar epítomos de MHC-II são TEPITOPE (STURNIOLO *et al.*, 1999), SMM-Align, NN-Align, NetMHCIIpan e Consensus3.

Dentre os métodos baseados em sequência está o TEPITOPE (STURNIOLO *et al.*, 1999), que utiliza as matrizes de afinidade peptídeo-MHC criadas a partir de dados experimentais obtidos da afinidade entre peptídeos e moléculas de MHC-II geradas por diferentes alelos. No algoritmo, todas as janelas de leitura de nove aminoácidos (nonâmeros) são extraídas de uma sequência de proteína (Figura 5). Ao invés de definir um peptídeo como epítomo ou não-epítomo, o uso de matrizes permite que o TEPITOPE atribua aos peptídeos um escore de imunogenicidade, oferecendo ao usuário a escolha na definição de um ponto de corte que define epítomos (acima do ponto de corte) e não-epítomos (abaixo do ponto de corte) (STURNIOLO *et al.*, 1999).

O método SMM-Align realiza suas predições partindo do pressuposto de que aminoácidos em diferentes posições de um peptídeo realizam contribuições independentes para a afinidade de ligação peptídeo-MHC total e considerando também a influência dos resíduos

Figura 5 – Janelas de leitura extraídas de uma sequência

MERIKELRNLMSSQSRTRREILTKTTVDHMAIIKKYTSGRQEKNP

1 MERIKELRN  
 2 ERIKELRNL  
 3 RIKELRNLNLM  
 4 IKELRNLMSS  
 5 KELRNLMSSQ  
 6 ELRNLMSSQS

Fonte: Elaborada pela autora

As seis primeiras janelas de leitura são mostradas na imagem. Epítomos com aminoácidos em comum, como por exemplo o 1 e o 4 podem ser reconhecidos como altamente imunogênicos; desse modo, os aminoácidos compartilhados por esses epítomos (IKELRN) devem receber maior atenção em estratégias de desimunização

flanqueadores de peptídeo (NIELSEN *et al.* 2007). Este método utilizou-se de aprendizagem de máquina para buscar uma matriz de pesos que melhor reproduz os valores de IC50 peptídeo-MHC encontrados para cada peptídeo utilizado em seu conjunto de treinamento (NIELSEN *et al.*, 2007).

O método NN-Align usa informações dos resíduos flanqueadores de peptídeo ao realizar suas previsões e corrige alguns dos problemas identificados nos conjuntos de treinamento de algoritmos de predição baseados em aprendizagem de máquina utilizados em trabalhos anteriores, como o viés gerado por múltiplos epítomos com núcleos de ligação idênticos sendo representados no conjunto de treinamento (NIELSEN; LUND, 2009).

Já como método baseado em estrutura, figura o preditor NetMHCIIpan, que utiliza os resíduos flanqueadores de peptídeo, núcleos de ligação, afinidade de ligação e quaisquer resíduos do MHC-II que tenham sido computados a uma distância que permita interação direta com resíduos do peptídeo analisado (NIELSEN *et al.*, 2008). Ao final, o método utiliza uma rede neural artificial para a realização de previsões generalizáveis para alelos de HLA-DR com poucos ou nenhum dado experimental.

E como método tipo consenso está o Consensus3, que foi desenvolvido após avaliação de métodos predição de MHC de classe II disponíveis e que realizam predição em lote (ARB, TEPITOPE, SVMHC, SYFPEITHI, RANKPEP, SMM-Align, SVRMHC, MHC2PRED, MHCpred). O escore consenso para cada peptídeo é a mediana dos escores de imunogenicidade apontados pelos três melhores preditores para um dado alelo de MHC (WANG *et al.*, 2008; WANG *et al.*, 2010). Esses e outros métodos são listados na ??.

Quadro 1 – Principais Métodos de Predição de Epítomos Lineares - Ligação ao MHC-II

Método de Predição	Características	Referência
Sturniolo	Utiliza matrizes geradas pela caracterização topográfica e funcional dos perfis de fendas de alelos do HLA-DR	(STURNIOLO <i>et al.</i> , 1999)
NetMHCIIpan	Rede neural artificial treinada em dados de ligação, cerne de ligação, resíduos flanqueadores do peptídeo e resíduos da molécula de MHC que interagem com o peptídeo ligado	(NIELSEN <i>et al.</i> , 2008)
NN-Align	algoritmo de alinhamento baseado em uma rede neural artificial que elimina cernes de ligação redundantes do seu conjunto de treinamento e considera resíduos flanqueadores do peptídeo	(NIELSEN; LUND, 2009)
SMM-Align	Considera resíduos flanqueadores de peptídeos e foi treinado para nove supertipos de HLA-DR e três alelos H2-IA de ratos	(NIELSEN <i>et al.</i> , 2007)
IEDB Consensus	Baseia-se na obtenção da mediana dos escores dos três melhores métodos de predição para uma dada molécula de MHC-II	(WANG <i>et al.</i> , 2008) (WANG <i>et al.</i> , 2010)
TEPITOPEpan	Extensão do método TEPITOPE para mais de 700 moléculas HLA-DR usando similaridade de sequência entre moléculas com matrizes de ligação definidas e as demais	(ZHANG <i>et al.</i> , 2012)
EpiTOP	Usa o método QSAR (relação quantitativa estrutura-atividade) para estimar imunogenicidade descrevendo volume, hidrofobicidade e polarizabilidade	(DIMITROV <i>et al.</i> , 2010)
ARB	Método da ligação relativa média que prediz valores de IC(50) para ligações MHC-peptídeo utilizando matrizes para 85 alelos MHC do tipo I e 13 alelos de MHC do tipo II construídas através de um banco de dados	(BUI <i>et al.</i> , 2005)
MULTIPRED2	Integra resultados de diversos supertipos de HLA e gera mapas de calor de imunogenicidade e permite visualizar padrões globais de ligação	(ZHANG <i>et al.</i> , 2011)
TEPITOPE	Ferramenta online pra predição de MHC do tipo II que se baseia em matrizes quantitativas	(STURNIOLO <i>et al.</i> , 1999)
PREDIVAC	Predição baseada em resíduos determinantes de especificidade	(OYARZÚN <i>et al.</i> , 2013)
HLAPred	Servidor online permite identificação e predição de peptídeos com sequencia antigênica para alelos de MHC I e II pela identificação de ligantes experimentalmente definidos (advindos do banco de dados MHCBN) e predição de ligantes utilizando-se de matrizes quantitativas de PARKER <i>et al.</i> (1994) e STURNIOLO <i>et al.</i> (1999)	(BHASIN; RAGHAVA, 2007)
FRED	É uma estrutura de software para imunoinformática que propõe uma única interface para a execução de diversos métodos de predição permitindo a realização de modificações customizadas	(FELDHahn <i>et al.</i> , 2009)
MHC2Pred	Método baseado em Máquinas de vetores de suporte que realiza predição de ligantes promíscuos para 42 alelos de MHC II	(LATA <i>et al.</i> , 2007)

Fonte: Elaborado pela autora

O uso de métodos acurados de identificação de epítomos para sequências sem dados experimentais compõe a primeira etapa de qualquer processo de desimunização. O mapeamento produzido com os dados desses métodos e a classificação de epítomos de acordo com sua imunogenicidade possibilita que as estratégias de desimunização mirem nos alvos corretos.

#### 1.4 Desimunização de proteínas *in silico*

Desimunização de proteínas é o processo pelo qual epítomos são eliminados estrategicamente de modo a reduzir a imunogenicidade total de uma proteína terapêutica (JONES *et al.*, 2009; GIOVANNONI, 2003). O objetivo final desse processo é mitigar as respostas de anticorpos antifármacos de modo a melhorar a eficácia do biofármaco (ZHAO *et al.*, 2015).

Dentre as maneiras de diminuir a imunogenicidade de proteínas terapêuticas podem ser citadas: peguilação (TURECEK *et al.*, 2016) e glicosilação (SINCLAIR; ELLIOTT, 2005) para diminuir a taxa de proteólise e diminuir interação com anticorpos e com moléculas de HLA. Apesar de suas aplicações de sucesso, o uso desses métodos pode resultar na produção de anticorpos anti-PEG (ZHANG *et al.*, 2014) e criação de novos epítomos para anticorpos neutralizantes (SINCLAIR; ELLIOTT, 2005), respectivamente.

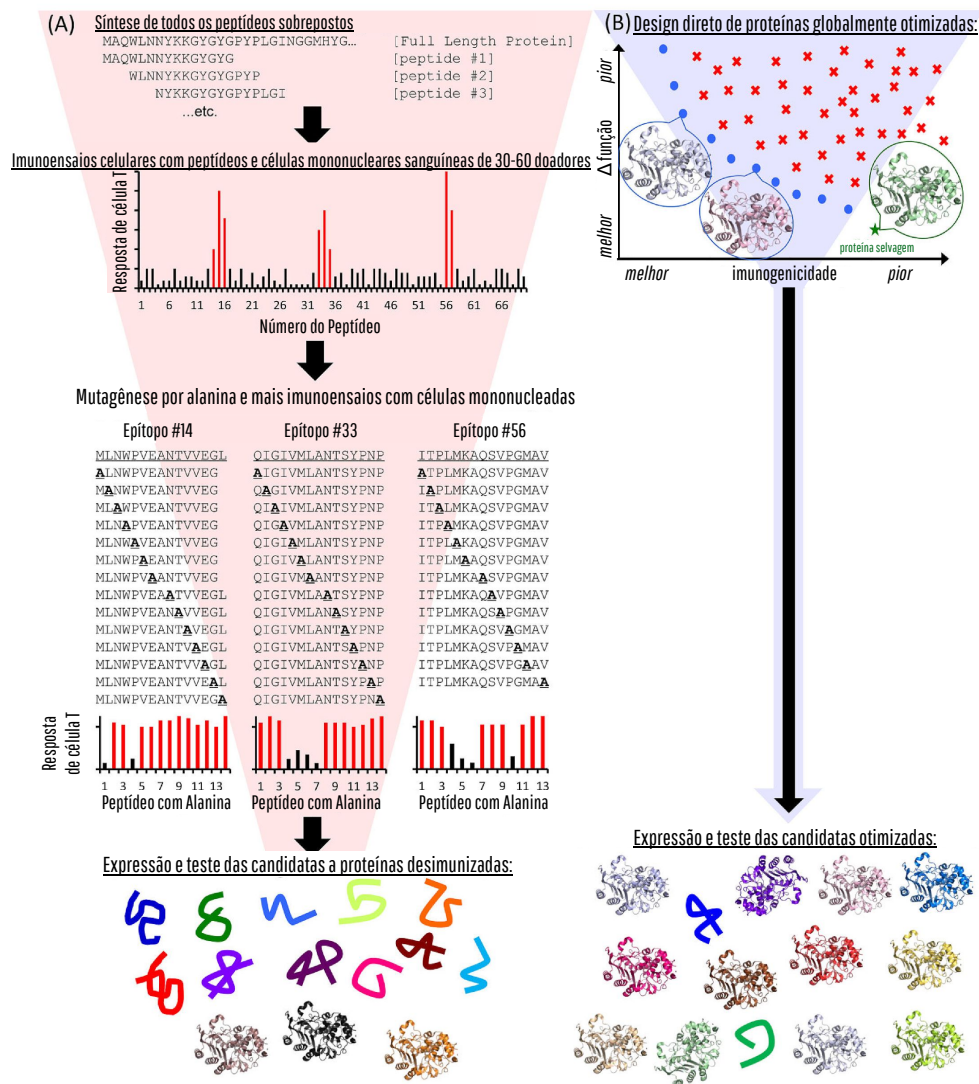
Como alternativa, existe a humanização (LIU *et al.*, 2015), que costuma gerar proteínas com imunogenicidade reduzida em relação à molécula original, embora sem eliminação completa da imunogenicidade (GROOT; SCOTT, 2007). Pode-se ainda realizar fusão com IgGs (ZAMBIDIS; SCOTT, 1996), mas isso onera os custos de produção (SCHMIDT, 2009), diminui a taxa de absorção do complexo e, a depender da proteína, causa um impedimento estérico na ligação proteína efetora-receptor (RATH *et al.*, 2015). Em outra estratégia, podem-se aplicar mudanças na linha de produção para evitar formação de agregados e diminuir a contaminação, mas essa estratégia é menos específica e exige um grau de empirismo que resulta numa solução muitas vezes paleativa (RATANJI *et al.*, 2014).

Por fim, a técnica de modificação da sequência por mutagênese dirigida (TANGRI *et al.*, 2005; GROOT; SCOTT, 2007) permite economizar recursos que seriam despendidos em ensaios e experimentos para avaliar a eficiência de mutantes criados aleatoriamente (Figura 6).

Idealmente, uma estratégia de desimunização por mutagênese dirigida permitiria eliminar totalmente a imunogenicidade de uma proteína sem prejudicar sua estabilidade e afinidade. Contudo, a seleção de mutações que diminuem a imunogenicidade de proteínas mantendo sua função permanece um desafio, pois são  $20^{300} = 2,03 \times 10^{90}$  sequências possíveis para uma

proteína contendo 300 aminoácidos. Adicionalmente, cada mutação realizada no sentido de desimmunizar uma sequência tem potencial para comprometer a estabilidade da conformação nativa. Assim, faz-se necessário surgimento de ferramentas de bioinformática que permitam selecionar mutações desimmunizantes que não perturbem a estrutura da proteína (SALVAT *et al.*, 2015). Dentre as estratégias já desenvolvidas, figuram o software EpiSweep, DP<sup>2</sup>, IP<sup>2</sup> e PEPFR.

Figura 6 – Estratégias de deleção de epítomos de célula T



Fonte: (GRISWOLD; BAILEY-KELLOGG, 2016)

Na estratégia A são muitos os ensaios realizados com *alanine scanning*. Para cada epítopo, toda posição deve ter seu aminoácido substituído por alanina afim de investigar o impacto na imunogenicidade. Contudo, nem todas as substituições que funcionam para reduzir a imunogenicidade do peptídeo permitem à proteína um enovelamento satisfatório. Na estratégia B, tempo e recursos são economizados num protocolo que descarta grande parte dos mutantes inviáveis e seleciona candidatos otimizados tanto para o parâmetro imunogenicidade, quanto para o de estabilidade.

O EpiSweep propõe a desimmunização de proteínas através da criação de bibliotecas

de mutantes com quantidades específicas de mutações (10 sítios e 30 sítios, no trabalho original, SALVAT *et al.* 2017), cujas sequências serão geradas de forma a selecionar mutações que diminuam a imunogenicidade da proteína ou melhorem sua estabilidade. A imunogenicidade é avaliada pelo método TEPITOPE, considerando apenas os 5% epítomos de maior escore e o critério de estabilidade é relacionado à flexibilidade de mutação de um resíduo, segundo a conservação da posição mutada e dos resíduos presentes nessa posição em sequências homólogas.

No DP<sup>2</sup> (*Dynamic Programming for Deimmunizing Proteins* - Programação Dinâmica para Desimunização de Proteínas, em livre tradução), a partir de uma sequência de aminoácidos, o método que identifica conjuntos otimizados de mutações pontuais conservativas que diminuam o número de epítomos de célula T nessas sequências (PARKER *et al.*, 2010). A imunogenicidade das sequências é avaliada pelo método TEPITOPE, considerando apenas os 10% epítomos de maior escore imunogênico identificados para os alelos DRB1\*0101, DRB1\*0301, DRB1\*0401, DRB1\*0701, DRB1\*0801, DRB1\*1101, DRB1\*1301, and DRB1\*1501 do MHC-II. A estabilidade é avaliada por três parâmetros, a diferença entre a proteína selvagem e a mutante utilizando a matriz de substituição BLOSUM-62, a conservação de determinados resíduos de acordo com alinhamentos múltiplos e árvores filogenéticas da sequência e o software FoldX em que o  $\Delta\Delta G$  entre a proteína selvagem e a mutante.

O IP<sup>2</sup> utiliza a mesma avaliação de imunogenicidade que o algoritmo DP<sup>2</sup>, apenas se utilizando de mutações compensadoras em resíduos altamente correlacionados para manter a estabilidade da proteína (PARKER *et al.*, 2011).

O algoritmo PEPFR (*Protein Engineering Pareto FRontier*) divide o espaço de busca de mutantes apropriados utilizando um *trade-off* (compromisso) entre a otimização de parâmetros potencialmente conflitantes, sendo eles imunogenicidade e estabilidade. Posteriormente são aplicados processos de programação dinâmica ou programação inteira para identificar os melhores indivíduos nessas regiões do universo de busca (HE *et al.*, 2012).

Apesar da existência dessas estratégias, existem diversas limitações quanto ao número de mutações possíveis para cada posição, custo computacional associado, número de alelos utilizados para predição de imunogenicidade, número de epítomos considerado para o cálculo de imunogenicidade e a dificuldade de uso dos métodos.

## 1.5 L-Asparaginase

A L-asparaginase é uma proteína terapêutica de natureza enzimática, e portanto, enquadra-se no primeiro grupo da classificação de Leader *et al.* (2008) (Seção 1.1, item i). Lang (1904) foi o primeiro a detectar sua atividade enzimática de degradação utilizando tecidos de carne bovina. Anos depois, Furth e Friedmann (1910) concluíram, após estudar órgãos de porcos e cavalos, que a L-asparaginase ocorria em todos os tecidos animais. Posteriormente, a proteína foi então redefinida como presente no fígado de animais onívoros, em quase todos os tecidos de herbívoros e ausente em órgãos de mamíferos carnívoros, anfíbios e répteis (CLEMENTI, 1922).

O potencial desta proteína como agente anticâncer foi observado pela atividade antitumoral do soro retirado de porcos-da-índia (KIDD, 1953, *Cavia porcellus*). Nesse estudo, injeções intraperitoniais do soro de *Cavia porcellus* geraram uma rápida regressão de dois tipos de linfomas transplantados em camundongos. Outros estudos demonstraram diferenças metabólicas entre células saudáveis e cancerosas na ausência do aminoácido asparagina, sendo as segundas dependentes de fontes exógenas do aminoácido L-asparagina uma vez que possuem deficiência na enzima asparagina sintase, ao contrário de células saudáveis (NEUMAN, 1956). A partir desses estudos, a atividade antitumoral da enzima e a depleção que esta causava no aminoácido asparagina para células cancerígenas foram conectadas (BROOME, 1961).

Posteriormente, foram isoladas duas asparaginases de *Escherichia coli*, uma periplasmática e uma citoplasmática, sendo a última a que exibiu atividade antitumoral (MASHBURN; WRISTON, 1964). As limitações do uso dessa proteína como terapia se devem à sua imunogenicidade e, desde os primeiros relatos (TALLAL *et al.*, 1970; JAFFE *et al.*, 1971; PETERSON *et al.*, 1971), mencionando este entrave diversas abordagens tem sido testadas de forma a reduzir sua capacidade de gerar resposta imune mantendo sua atividade enzimática (ALI *et al.*, 2016).

A L-asparaginase é utilizada no Brasil desde a década de 1970 para o tratamento da leucemia linfoblástica aguda (CFM, 2017). No Brasil, de acordo com o Conselho Federal de Farmácia, entre 2011 e 2014, o gasto do governo federal na compra de medicamentos subiu em 53%, atingindo R\$ 12,9 bilhões - que correspondiam a 13,9% do orçamento disponível para o SUS (CAMBRICOLI, 2015).

De acordo com a Agência Nacional de Vigilância Sanitária (ANVISA) (RDC 55/2010), produto biológico é o medicamento biológico não-novo ou conhecido que contém molécula com atividade biológica conhecida, já registrado no Brasil e que tenha passado por todas as etapas de fabricação. Dentre eles estão:

1. vacinas;
2. soros hiperimunes;
3. hemoderivados;
4. biomedicamentos;
  - a) medicamentos obtidos a partir de fluidos biológicos ou de tecidos de origem animal;
  - b) medicamentos obtidos por procedimentos biotecnológicos.
5. anticorpos monoclonais;
6. medicamentos contendo micro-organismos vivos, atenuados ou mortos;

A L-asparaginase estaria então entre aqueles medicamentos do grupo 4b. Em 2017, uma mudança no fornecedor desse medicamento gerou impactos no sistema de saúde nacional e nos órgãos de representação dos pacientes (Figura 7).

No fim de 2012, a distribuidora Laboratórios Bagó do Brasil S.A informou a suspensão da comercialização do Elspar no Brasil, o que gerou um pedido da SOBOPE (Sociedade Brasileira de Oncologia Pediátrica) para que o Ministério da Saúde adquirisse o medicamento em caráter emergencial até que a situação fosse normalizada. Até 2013, os medicamentos que continham L-asparaginase como princípio ativo eram adquiridos do laboratório Bagó sob o nome de Elspar, fabricados pela Merck Sharp & Dohme (MSD), Lundbeck INC ou OSO Biopharmaceutical; essa aquisição era feita pelos hospitais brasileiros (GADELHA, 2017).

Desse modo, em maio de 2013, o MS realizou sua primeira compra centralizada de Aginasa (da Medac, empresa alemã), medicamento fornecido também pelo laboratório Bagó. Em janeiro do ano seguinte, o laboratório Bagó informou que a importação do medicamento Aginasa seria interrompida devido encerramento das atividades da fabricante do princípio ativo do medicamento no Japão (Kyowa Hakko Kirin Co. Ltda) e ofereceu uma outra asparaginase, recombinante, da Medac (GADELHA, 2017).

O fornecimento de Aginasa pela Bagó estava garantido, segundo o próprio laboratório, para o ano de 2014. A ANVISA pediu então os estudos de estabilidade (pré-clínicos e clínicos) para análise de registro definitivo da Aginasa, ao que a Bagó alegou impossibilidade de fornecimento, devido ao tempo para o cumprimento dos estudos e afirmou que a utilização do medicamento por 30 anos o referendava para o uso. Após reafirmação do posicionamento da Bagó pela SOBOPE, o MS realizou segunda compra centralizada da Aginasa da Medac, via Bagó, para abastecimento até janeiro de 2017 (GADELHA, 2017).

Em maio de 2015, o registro definitivo da Aginasa foi negado pela ANVISA devido

às dezessete pendências, dentre elas a ausência de relatórios dos estudos clínicos. A Bagó recorreu dessa decisão, entretanto o registro da Aginasa permaneceu negado. Em 2016, o MS iniciou novo processo de compra da L-asparaginase antecipando-se ao desabastecimento de 2017 e a consultoria jurídica do MS apontou a necessidade de ocorrência de licitação, uma vez que havia outros fornecedores do medicamento. A Bagó concorreu com seu medicamento, Medac, mas perdeu a licitação para o concorrente de menor preço, do laboratório Beijing (fornecedor Xetley) (GADELHA, 2017).

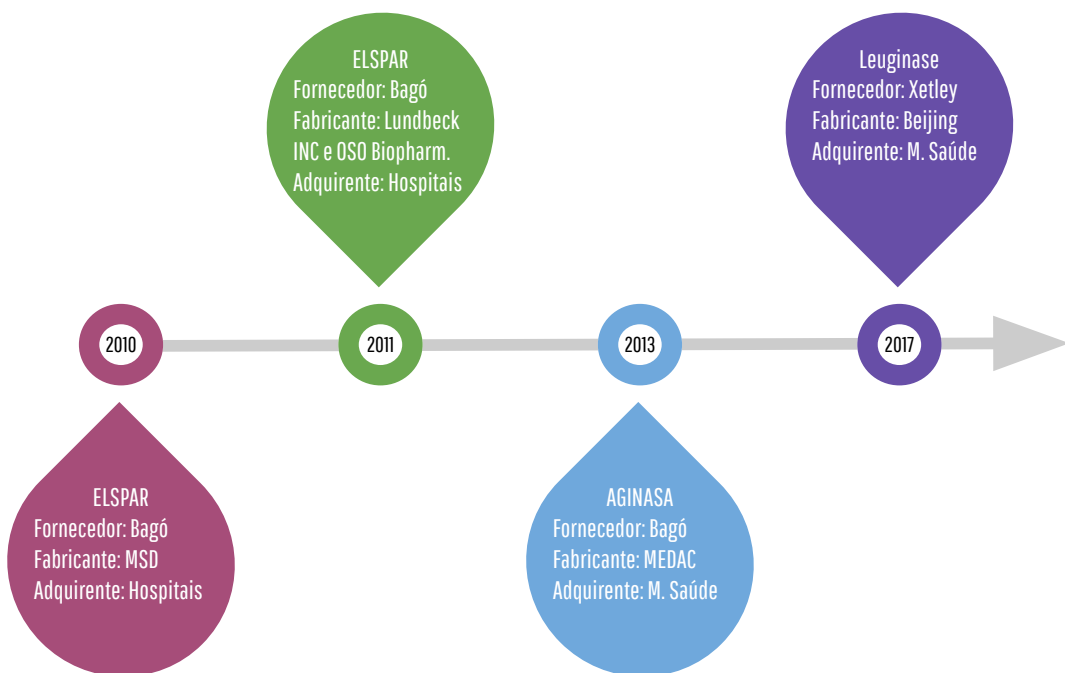
Em janeiro de 2017, o MS então adquiriu o medicamento Leuginase do laboratório Beijing. A partir desse momento, a própria SOBOPE se posicionaram contra a utilização do medicamento Leuginase, alegando: falta de estudos clínicos que comprovem a eficácia do medicamento, efeitos inesperados e incomuns relatados na bula, alta toxicidade e nível impurezas superior ao do medicamento alemão (GADELHA, 2017).

Essas alegações foram refutadas pelo MS uma vez que a Aginasa, usada anteriormente, também não possuía estudos clínicos; além disso, os efeitos descritos na bula da Leuginase são os mesmos das demais marcas de L-asparaginase (assim como a toxicidade) e quanto ao nível de impurezas, testes realizados pelo INCQS/Fiocruz, mostraram ausência de contaminantes que pudessem causar danos aos pacientes e atestaram a atividade enzimática da proteína (GADELHA, 2017).

Este relato histórico da comercialização de L-asparaginase no Brasil aponta a crescente necessidade da fabricação nacional deste medicamento. Dentre os benefícios trazidos está a redução dos custos para o Sistema Único de Saúde, o fortalecimento da indústria nacional farmacêutica e do complexo industrial da saúde, geração de empregos diretos e indiretos, ampliação do acesso do usuário e, a depender da natureza do biofármaco (*biobetter* ou biosimilar), a segurança e eficácia do medicamento seria garantida pela obrigatoriedade da realização de extensos estudos clínicos.

A L-asparaginase de *Bacillus subtilis* vem como potencial substituta das formas comercializadas atualmente, de *Escherichia coli* e *Erwinia chrysantemi*. Dentre as vantagens na utilização desse micro-organismo como produtor, estão: o nível de conhecimento acerca dele, sua facilidade de cultivo, ausência de patogenicidade e facilidade de transformação. Contudo, a imunogenicidade da L-asparaginase de *B. subtilis* faz necessários processos de desimunização e validação prévios que justifiquem sua entrada no mercado e este trabalho figura como etapa inicial nesses esforços de desimunização.

Figura 7 – Linha do tempo da distribuição de L-asparaginase no Brasil



Fonte: Elaborada pela autora

Resumo da mudança de fornecedores de L-asparaginase no Brasil desde 2010. Até 2012 a compra de L-asparaginase era feita diretamente pelos hospitais com os Laboratórios Bagó que importava o medicamento de diferentes fornecedores. A partir de 2013 a compra tornou-se centralizada e feita pelo Ministério da Saúde. Entre os anos de 2013 e 2017 mudanças no preço do medicamento e a falta de relatórios de estudos clínicos culminaram na realização de uma licitação internacional para a compra de L-asparaginase. O fornecedor Xetley é o atual fornecedor do medicamento.

## 2 OBJETIVO GERAL

O trabalho desenvolvido teve como objetivo desenvolver ferramentas computacionais para desimunização de proteínas aplicando tais ferramentas a sequência de L-asparaginase de *Bacillus subtilis*.

### 2.1 Objetivos específicos

- Realizar predições de epítomos *in silico* com o método TEPITOPE;
- Avaliar e melhorar as predições realizadas;
- Mapear as predições do melhor método em uma L-asparaginase de *Bacillus subtilis*;
- Construir um algoritmo genético de desimunização de proteínas tendo como base este mapeamento;
- Aplicar esse algoritmo genético em uma sequência de L-asparaginase de *Bacillus subtilis*;
- Obter sequências desimunizadas de L-asparaginase de *Bacillus subtilis*.

### 3 METODOLOGIA

#### 3.1 Predição da Imunogenicidade

A predição da imunogenicidade tem por objetivo encontrar epítomos apenas a partir da sequência de uma proteína e é, portanto, fundamental para um projeto de desimunização de proteínas.

##### 3.1.1 Obtenção dos dados

Para avaliar a acurácia dos métodos de predição *in silico*, são necessárias informações experimentais acerca da imunogenicidade de um número considerável de sequências. Essas informações foram usadas para validar a predição realizada pelos métodos.

Os dados experimentais foram obtidos do banco de dados imunológicos IEDB<sup>1</sup>, onde estão reunidos epítomos de célula T determinados experimentalmente. O número de acesso GI foi utilizado para a obtenção das sequências no formato FASTA de 7183 proteínas através do servidor Batch Entrez<sup>2</sup> do NCBI. Dentre as sequências obtidas no Batch Entrez, 1852 proteínas possuem epítomos padrão de célula T, isto é, são epítomos lineares. Os demais epítomos são carboidratos e pequenas moléculas (carbamazepina, cloranfenicol, cumarina. Figura 8. Seção 3.1).

Figura 8 – Exemplo de um arquivo do IEDB

1	2	3	4
37317	LLGCIITSL	P26662.3	ELISPOT
36795	LKAINWFGQVSTTVAAL	Q07297.1	ELISA
56721	SACDVSVRV	Q96JT2.1	ELISPOT
56722	SACDVSVRVV	Q96JT2.1	51 chromium
70182	VNSFDSSSSSDSLYESI	SRC280373	ELISA
57322	SDYEGRLI	SRC124	in vivo assay
69572	VLKSYVLEGLTAEK	P14013.1	3H-thymidine
74989	YLVTRHADVI	P29846.3	ELISPOT
...	...	...	...

Fonte: Elaborada pela autora.

Exemplo de banco de dados de epítomos de célula T do IEDB. 1. Código do epítomo. 2. Sequência do epítomo. 3. Código identificador da proteína no NCBI, GI. 4. Método de identificação do epítomo.

<sup>1</sup> <http://www.iedb.org/>

<sup>2</sup> <https://www.ncbi.nlm.nih.gov/sites/batchentrez>

### 3.1.2 Uso do TEPITOPE para a predição de epítomos lineares

A avaliação da imunogenicidade foi realizada pelo método de predição TEPITOPE, considerando todos os alelos cujas matrizes de afinidade de ligação estavam disponíveis (Quadro 2).

Quadro 2 – Alelos do HLA utilizados pelo TEPITOPE

DRB1*0101	DRB1*0102	DRB1*0301	DRB1*0305	DRB1*0306
DRB1*0307	DRB1*0308	DRB1*0309	DRB1*0311	DRB1*0401
DRB1*0402	DRB1*0404	DRB1*0405	DRB1*0408	DRB1*0410
DRB1*0423	DRB1*0426	DRB1*0701	DRB1*0703	DRB1*0801
DRB1*0802	DRB1*0804	DRB1*0806	DRB1*0813	DRB1*0817
DRB1*1101	DRB1*1102	DRB1*1104	DRB1*1106	DRB1*1107
DRB1*1114	DRB1*1120	DRB1*1121	DRB1*1128	DRB1*1301
DRB1*1302	DRB1*1304	DRB1*1305	DRB1*1307	DRB1*1311
DRB1*1321	DRB1*1322	DRB1*1323	DRB1*1327	DRB1*1328
DRB1*1501	DRB1*1502	DRB1*1506	DRB5*0101	DRB5*0105

Fonte: Elaborado pela autora

Existem diversos métodos de predição de epítomos lineares, que, embora utilizem-se de métricas diferentes para prever epítomos, possuem acurácia semelhante. O método de predição utilizado neste trabalho foi o TEPITOPE (STURNIOLO *et al.*, 1999), que se baseia em matrizes de predição e foi utilizado anteriormente com relatos de sucesso (PARKER *et al.*, 2010; PARKER *et al.*, 2013). O TEPITOPE considera todas as sequências de 9 resíduos contíguos (nonâmeros) possivelmente criados pela clivagem de uma dada proteína e avalia a imunogenicidade desses nonâmeros, por meio de matrizes de afinidade formuladas de acordo com a contribuição dos aminoácidos em diferentes posições. Essas contribuições foram calculadas com base nos *pocket profiles*, perfis experimentalmente definidos através das variações de afinidade peptídeo-MHC-II causadas pela presença dos diferentes aminoácidos em cada uma das posições do peptídeo (Quadro 3).

O valor de imunogenicidade de um peptídeo é dado pela soma dos valores posição-aminoácido da matriz para um dado alelo de MHC-II. Para esse alelo, o TEPITOPE permite então que apenas cinco resíduos (valores -1.0 e 0) estejam presentes na primeira posição para que um peptídeo seja considerado um epítopo. A ausência de valores na posição 5 e 8 se dá porque as cadeias laterais dos resíduos que poderiam interagir com os aminoácidos nestas posições estão voltadas para a direção oposta à fenda de ligação do MHC-II, não influenciando na afinidade final da ligação.

Cada arquivo de predição gerado pelo TEPITOPE dispõe os nonâmeros em ordem

Quadro 3 – Matriz utilizada pelo TEPITOPE para calcular a afinidade peptídeo-HLA DRB1\*0101

Amino acid/Position	P1	P2	P3	P4	P5	P6	P7	P8	P9
A	-999.00	0.00	0.00	0.00	-	0.00	0.00	-	0.00
D	-999.00	-1.30	-1.30	-2.40	-	-2.70	-2.00	-	-1.90
E	-999.00	0.10	-1.20	-0.40	-	-2.40	-0.60	-	-1.90
F	0.00	0.80	0.80	0.08	-	-2.10	0.30	-	-0.40
G	-999.00	0.50	0.20	-0.70	-	-0.30	-1.10	-	-0.80
H	-999.00	0.80	0.20	-0.70	-	-2.20	0.10	-	-1.10
I	-1.00	1.10	1.50	0.50	-	-1.90	0.60	-	0.70
K	-999.00	1.10	0.00	-2.10	-	-2.00	-0.20	-	-1.70
L	-1.00	1.00	1.00	0.90	-	-2.00	0.30	-	0.50
M	-1.00	1.10	1.40	0.80	-	-1.80	0.09	-	0.08
N	-999.00	0.80	0.50	0.04	-	-1.10	0.10	-	-1.20
P	-999.00	-0.50	0.30	-1.90	-	-0.20	0.07	-	-1.10
Q	-999.00	1.20	0.00	0.10	-	-1.80	0.20	-	-1.60
R	-999.00	2.20	0.70	-2.10	-	-1.80	0.09	-	-1.00
S	-999.00	-0.30	0.20	-0.70	-	-0.60	-0.20	-	-0.30
T	-999.00	0.00	0.00	-1.00	-	-1.20	0.09	-	-0.20
V	-1.00	2.10	0.50	-0.05	-	-1.10	0.70	-	0.30
W	0.00	-0.10	0.00	-1.80	-	-2.40	-0.08	-	-1.40
Y	0.00	0.90	0.80	-1.10	-	-2.00	0.50	-	-0.90

Fonte: (STURNIOLO *et al.*, 1999)

Na primeira posição, quinze aminoácidos realizam uma contribuição muito negativa (-999.00) para o valor total do epítipo. Para as posições 5 e 8 do peptídeo, qualquer resíduo pode se fazer presente.

decrecente de probabilidade de serem epítipos (Figura 9). O mesmo nonúmero pode ser exibido mais de uma vez no arquivo de predição, se tratando então de um *epítipo promíscuo*, isto é, um peptídeo identificado como epítipo para mais de uma molécula de MHC-II. Epítipos promíscuos são de especial interesse para trabalhos de desimunização, devido ao seu potencial de serem reconhecidos por um grande número de indivíduos. Como epítipos diferentes se sobrepõem independentemente em diferentes regiões da sequência, existe também a possibilidade de um mesmo resíduo fazer parte de mais de um epítipo. Esses serão denominados *aminoácidos promíscuos*, e são os alvos de maior interesse para problemas de desimunização por mutagênese sítio-dirigida, por permitirem inativar mais de um epítipo com apenas uma mutação.

Figura 9 – Representação do arquivo de saída do método TEPITOPE

1	2	3	4	5
6.40	38	46	YVQSNGGAI	HLA-DRB1-0701
6.40	38	46	YVQSNGGAI	HLA-DRB1-0703
6.00	99	107	MRTVTPIRM	HLA-DRB1-0701
6.00	99	107	MRTVTPIRM	HLA-DRB1-0703
5.90	191	199	FGISNYCQI	HLA-DRB1-0701
5.90	191	199	FGISNYCQI	HLA-DRB1-0703
5.60	230	238	FRHYDGRTI	HLA-DRB1-1502
5.40	57	65	FKNRFLMSA	HLA-DRB1-0817
5.10	151	159	CNGDTIPRG	HLA-DRB1-0309
5.10	151	159	CNGDTIPRG	HLA-DRB1-0301
5.00	230	238	FRHYDGRTI	HLA-DRB1-0703
5.00	230	238	FRHYDGRTI	HLA-DRB1-0701
5.00	197	205	CQIYPPNAN	HLA-DRB1-0817
4.90	183	191	CRRPNAQRF	HLA-DRB1-0701
4.90	183	191	CRRPNAQRF	HLA-DRB1-0703
4.90	174	182	YRYVAREQS	HLA-DRB1-0309
4.90	255	263	VGYSNAQGV	HLA-DRB1-0703
4.90	255	263	VGYSNAQGV	HLA-DRB1-0701
4.90	96	104	LRQMRTVTP	HLA-DRB1-0410
4.80	57	65	FKNRFLMSA	HLA-DRB1-0813

Fonte: Elaborada pela autora.

1. Escore de imunogenicidade do epítipo. 2. Posição inicial do epítipo. 3. Posição final do epítipo. 4. Sequência do epítipo. 5. Alelo considerado na predição.

### 3.1.3 Destaque dos aminoácidos promíscuos

A importância dos epítopos promíscuos está em sua relevância clínica. Por serem reconhecidos por diversas moléculas do MHC-II, tem potencial para desencadear reações de hipersensibilidade num conjunto diverso de pacientes. Para destacar os aminoácidos promíscuos, isso é, quais regiões possuem alta afinidade com mais de um alelo de MHC, foi realizado um tratamento da saída do TEPITOPE, permitindo então a localização de regiões mais indicadas para mutação. De forma a localizá-los, para cada posição da proteína foi realizado o somatório dos escores de imunogenicidade dos de todos os epítopos preditos para a mesma posição (Figura 10).

Figura 10 – Soma dos escores para cada posição

Q1				Q2		Q3	
1	2	3	4	1	2	1	2
9.7	4	12	SDFGHLW	1	0	1	0
9.0	13	21	KLPTREWC	2	0	2	0
8.8	14	22	LPTREWCS	3	0	3	0
8.4	16	24	TREWCSVA	4	9.7	4	9.7
				5	9.7	5	9.7
				6	9.7	6	9.7
				7	9.7	7	9.7
				8	9.7	8	9.7
				9	9.7	9	9.7
				10	9.7	10	9.7
				11	9.7	11	9.7
				12	9.7	12	9.7
				13	9.0	13	9.0
				14	9.0 + 8.8	14	17.8
				15	9.0 + 8.8	15	17.8
				16	9.0 + 8.8 + 8.4	16	26.2
				17	9.0 + 8.8 + 8.4	17	26.2
				18	9.0 + 8.8 + 8.4	18	26.2
				19	9.0 + 8.8 + 8.4	19	26.2
				20	9.0 + 8.8 + 8.4	20	26.2
				21	9.0 + 8.8 + 8.4	21	26.2
				22	8.8 + 8.4	22	17.2
				23	8.4	23	8.4
				24	8.4	24	8.4

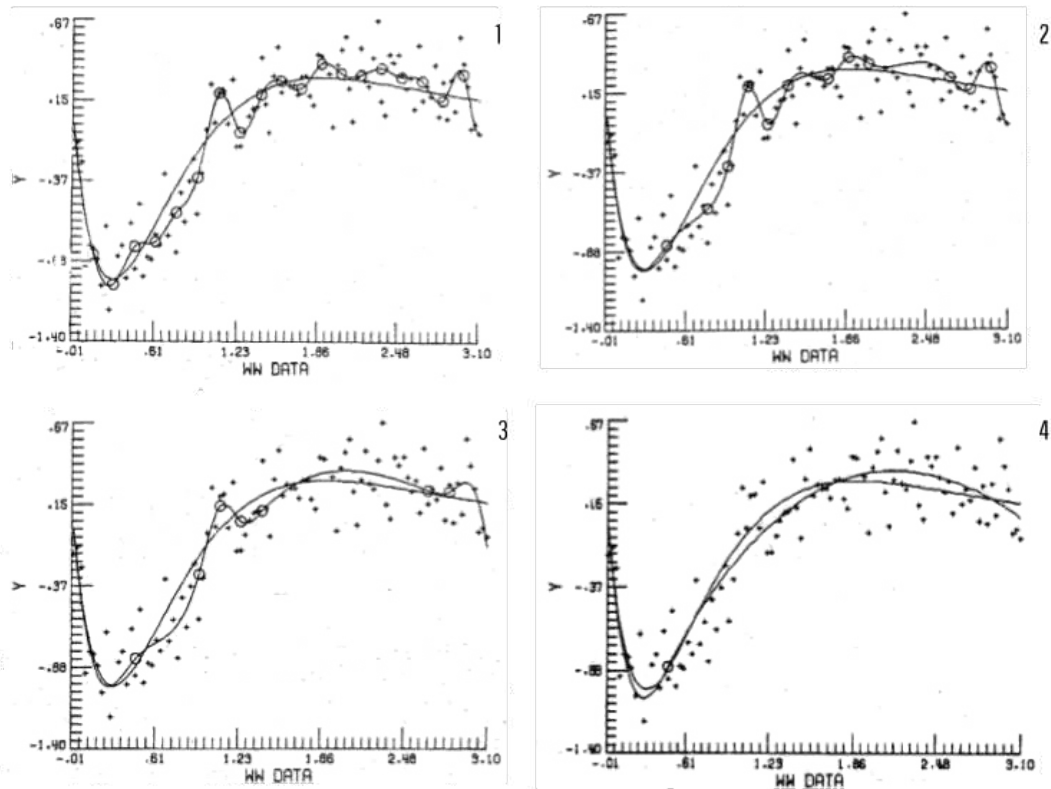
Fonte: Elaborada pela autora

Diferentes cores representam diferentes epítotos. Q1: arquivo de predição do TEPITOPE. 1: Escore imunogênico, 2: Posição inicial do epítoto, 3: Posição final do epítoto, 4: Sequência do epítoto. Q2: operação realizada para cada uma das posições de acordo com a sobreposição. 1: Posição, 2: Escore. O escore de cada epítoto está marcado por uma cor diferente. Os epítotos que se sobrepõem têm seus escores somados. Q3: arquivo em que todas as posições estão representadas com seus escores somados. 1: Posição, 2: Escore de imunogenicidade final.

### 3.1.4 Discernimento de regiões imunogênicas

A *spline* é uma função definida em intervalos contínuos utilizando polinômios, capaz de simplificar funções complexas de maneira acurada que varia de acordo com o número de intervalos (nós) em que são divididos os dados e o grau do polinômio utilizado para representar as seções (Figura 11). Nesse trabalho, os dados somados obtidos na estratégia anterior foram divididos. O número de nós para uma sequência de L aminoácidos é de L/16, ou seja, cada faixa de L/16 aminoácidos a sequência é interpolada por uma *spline* cúbica.

Figura 11 – Variação da spline de acordo com o número de nós utilizados



Fonte: (SMITH, 1982)

Um conjunto de dados complexo pode ter seu comportamento descrito por uma *spline*. 1. e 2. Um número muito grande de nós divide o espaço amostral em intervalos muito pequenos e cria uma função sobreajustada, que não cumpre seu objetivo de suavizar os dados. 3. e 4. Conforme o número de nós é ajustado ao ótimo para aquele conjunto de dados, surge uma curva que permite descrever de maneira simples e eficiente os dados analisados.

### 3.2 Avaliação da Predição de Epítomos

Foram utilizadas três estratégias de predição: (i) TEPITOPE (3.1.2), (ii) somatório das regiões sobrepostas (Somatório, 3.1.3) e (iii) suavização do somatório (Suavização, 3.1.4). Os dados experimentais obtidos anteriormente foram utilizados para analisar quantitativamente como esses comportamentos se traduziam para as proteínas analisadas. Nas saídas dos três métodos de predição os epítomos são dispostos de maneira decrescente, ou seja, os peptídeos mais imunogênicos são os primeiros apresentados no arquivo de saída.

Cada método teve seu erro na predição do epítopo calculada da seguinte forma: o erro atribuído para um epítopo é dado soma de todas as posições apontadas pelo método até a identificação do epítopo correto e o erro total do método é a soma do erro de todos os epítomos (Figura 12).

Figura 12 – Método de contagem de erros

Q1			
M1	M2		
1	1		
2	2		
3	3		
4	4		
5	5		
6	6		

Q2	
M1:	$1 + 1 + 2 + 3 + 4 + 5 = 16$
M2:	$1 + 2 + 1 + 2 + 3 + 4 = 13$

Fonte: Elaborada pela autora.

Q1: arquivos de predição de dois métodos, M1 e M2. Q2: somatório de todas as posições apontadas por este método até a identificação do epítipo correto.

Como na Suavização (subseção 3.1.4) apenas alguns dados do conjunto avaliado são considerados na construção da curva, o tamanho da faixa interpolada pela *spline* cúbica foi gradualmente reduzido até que em média, 75% das posições fossem apresentadas no arquivo de predição avaliado. As demais posições foram diretamente copiadas do TEPITOPE e adicionadas a esse arquivo.

### 3.3 Algoritmo genético para desimunização de proteínas

#### 3.3.1 Conceito

O objetivo principal dos algoritmos de otimização consiste em encontrar a melhor solução dentro de um conjunto de soluções possíveis. A qualidade de uma solução é dada pelo valor que ela apresenta diante de uma função de avaliação, e sua viabilidade é dada por um conjunto de restrições.

Algoritmos evolucionários são procedimentos de otimização que buscam as melhores soluções possíveis para o problema proposto partindo de um conjunto de soluções iniciais e gerando, sucessivamente, novas soluções mais adequadas (XINJIE; MITSUO, 2010). Um dos mais utilizados algoritmos de otimização é o Algoritmo Genético (AG) inspirado pela teoria da evolução pela seleção natural proposta por Darwin e Wallace ((HOLLAND, 1975)).

Uma simulação de AG se inicia com um conjunto de representações abstratas

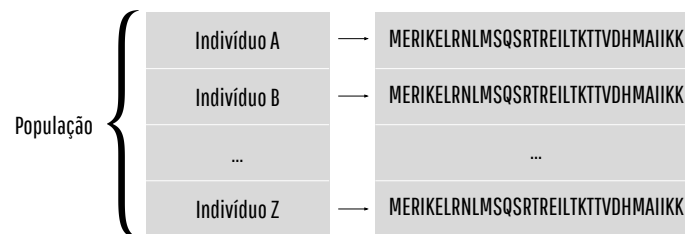
de soluções candidatas ao problema, denominadas *indivíduos*, que serão avaliadas, mutadas, recombinadas, e finalmente selecionadas de forma a gerar um novo grupo de soluções em um processo que será repetido quantas vezes forem necessárias até que se atinja o objetivo final. Nas seções seguintes, serão descritos os conceitos básicos de AG e como foram implementados neste trabalho.

### 3.3.2 *Indivíduo e População*

Cada possível solução para o problema é denominada, isoladamente, de *indivíduo*. Neste trabalho, o indivíduo é a estrutura primária da proteína que se almeja desimunizar. Por analogia, o conjunto de indivíduos é denominado *população* que, por consequência, representa a totalidade de soluções candidatas (Figura 13).

Embora o tamanho máximo da população seja constante, a composição da população muda ao longo das gerações conforme indivíduos mais adaptados são criados. Nesse trabalho utilizou-se uma população  $p = 100$  indivíduos. Para inicializar o algoritmo, necessita-se de um conjunto de soluções prévias, denominadas *população inicial*, que neste trabalho, consiste em  $p_0 = 100$  cópias iguais da sequência-alvo.

Figura 13 – População em um algoritmo Genético de Desimunização



Fonte: Elaborada pela autora  
População de tamanho 26, cada indivíduo é uma sequência de aminoácidos.

### 3.3.3 *Aptidão*

Cada indivíduo possui uma *aptidão* que reflete sua qualidade relativa como potencial solução do problema permitindo que o AG, ao final de sua execução, discrimine os melhores da população. A escolha de uma função de aptidão acurada é de fundamental importância para o sucesso de um AG, pois ela é o critério que determina o sucesso reprodutivo de cada indivíduo: a cada geração os indivíduos com melhor aptidão são os escolhidos para permanecer na população em um processo análogo ao de seleção natural.

Neste trabalho, a aptidão de cada indivíduo consiste na soma dos epítomos de maior escore imunogênico da sequência e, quanto menor esse valor, melhor o indivíduo. O método de predição utilizado foi a Suavização (subseção 3.1.4). Os epítomos que possuem escore entre os 20% maiores eram considerados para o cálculo de imunogenicidade e para definir quais os pontos propícios a mutação. Esse limite foi escolhido pois direciona o algoritmo genético a modificar os epítomos mais influentes na imunogenicidade total da proteína.

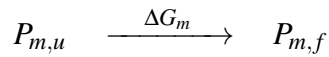
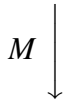
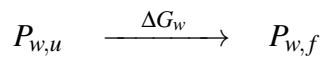
### 3.3.4 Restrições

Enquanto a função de aptidão é o critério que determina a qualidade de um indivíduo frente aos demais da população, permitindo encontrar as melhores soluções, uma *restrição* é uma função que determina a viabilidade de um indivíduo. Uma restrição é, portanto, um critério que deve ser atendido de antemão por um dado indivíduo e delimita, para a população, o espaço de soluções viáveis.

Para este AG, uma solução é ótima se possui imunogenicidade igual a zero. Entretanto, a partir da sequência original, podem-se propor diversos mutantes com imunogenicidade zero cuja conformação tridimensional não é fisicamente possível. Para o problema de desimunização só interessam mutantes de menor imunogenicidade cujas estruturas terciárias sejam livres de colisões estéricas.

Para cada posição imunogênica, são gerados 20 mutantes, cada um com um aminoácido diferente. Como um epítopo possui 9 resíduos, são gerados  $9 \times 20 = 180$  sequências possíveis. A estabilidade desses mutantes é avaliada em função de sua energia livre  $\Delta G$ , pelo FoldX. O FoldX calcula a energia livre utilizando um campo de forças empírico e o valor final de  $\Delta G$  depende das energias relacionadas a: forças de Van der Waals, grupos polares, grupos apolares, ligações de hidrogênio e grupos carregados.

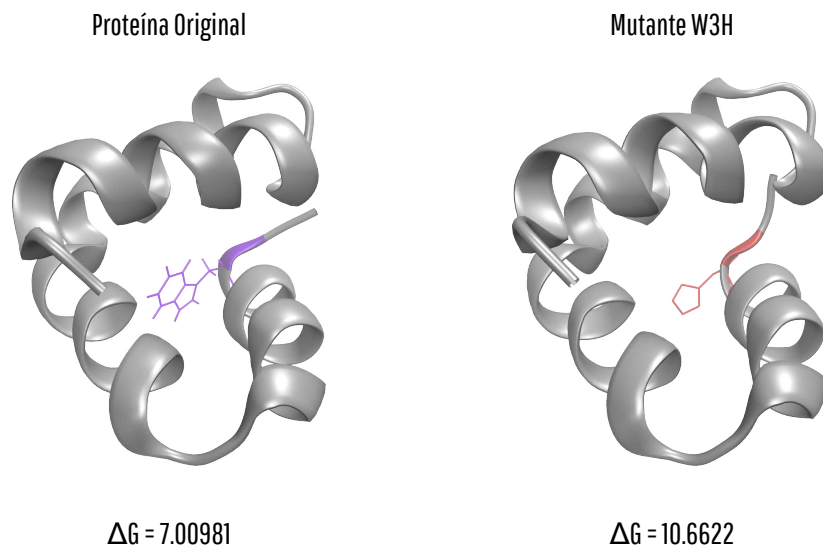
O  $\Delta G$  de cada um dos mutantes é comparado com o  $\Delta G$  da proteína selvagem e aqui se consideram estruturas terciárias possíveis aquelas cujo  $\Delta\Delta G$  mutante-proteína se situa entre  $-0.5$  kcal/mol e  $0.5$  kcal/mol, uma vez que a mutação não deve afetar a estabilidade da proteína selvagem, prejudicando sua viabilidade (Figura 14, Equação 3.1).



$$\Delta\Delta G = \Delta G_w - \Delta G_m \quad (3.1)$$

$P$  = proteína,  $w$  = selvagem,  $m$  = mutado,  $u$  = desenovelado,  $f$  = enovelado,  $\Delta G$  = energia livre de Gibbs,  $\Delta$  = variação

Figura 14 –  $\Delta G$  em kcal/mol de duas estruturas de acordo com o FoldX



Fonte: Elaborada pela autora (PDB: 2N80).

O  $\Delta\Delta G$  entre a proteína original e a mutada impede que a mutação W3H seja considerada uma alternativa viável, uma vez que a diferença ultrapassa o limite de 0.5 kcal/mol estabelecido.

### 3.3.5 Operadores Genéticos

Os operadores genéticos permitem gerar um novo indivíduo a partir indivíduos pré-existent na população sendo, portanto, os responsáveis por guiar a população em direção ao

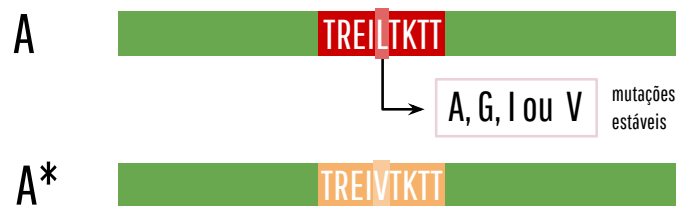
ótimo. São três os operadores genéticos utilizados nesse trabalho: seleção parental, recombinação e mutação.

### 3.3.5.1 Mutação

A *mutação* é o operador genético que serve para inserir e manter variabilidade na população. Análogo à mutação biológica, o operador genético de mutação consiste em uma troca pontual de um dos aminoácidos da sequência (Figura 15).

Neste AG as mutações são feitas sob as seguintes condições: (i) posição da mutação está contida na lista de posições imunogênicas (subseção 3.3.2) e (ii) aminoácido inserido nessa posição não prejudica a estabilidade da proteína, calculada pelo software FoldX (subseção 3.3.4). Dentre as possíveis mutações que atendem a esses critérios, uma delas é escolhida aleatoriamente.

Figura 15 – Exemplo de mutação considerando o conjunto de mutações estáveis



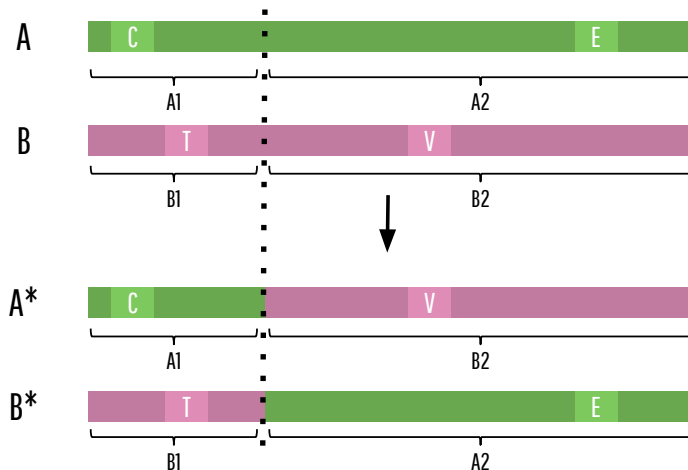
Fonte: Elaborada pela autora

Nesse exemplo, o epítopo TREILKTT tem suas posições contidas na lista de posições imunogênicas e dentre essas seu quinto resíduo (aminoácido L) foi aleatoriamente escolhido para mutação. De acordo com a avaliação de estabilidade de mutação, os mutantes L5A, L5G, L5I e L5V são estáveis, estando assim na lista de mutações permitidas. Desse modo, o operador genético escolheu uma mutação aleatória (V) da lista de mutações permitidas, gerando uma nova sequência.

### 3.3.5.2 Recombinação

Embora a mutação seja o único operador a introduzir novas características, a maior parte da diversidade de uma população advém da recombinação de características pré-existentes em indivíduos. Essa diversidade é originada pelo uso do operador de *recombinação*, na qual partes diferentes de duas soluções existentes são unidas de modo a produzir uma nova sequência que potencialmente possua as melhores características das soluções originais (Figura 16).

Figura 16 – Exemplo de operação e *crossover* realizada entre duas sequências A e B



Fonte: Elaborada pela autora

O operador de recombinação começa com a seleção de dois indivíduos A e B da população, denominados parentais. Uma posição da sequência é escolhida aleatoriamente e os parentais são ‘cortados’ nessa posição, resultando em duas subsequências (A1 e A2, B1 e B2). Dois novos indivíduos, chamados filhos, são então gerados pelas combinações lineares de A1+B2 e B1+A2.

Cada um dos filhos tem sua estabilidade verificada pelo FoldX; caso não atendam ao critério de inclusão ( $-0.05 < \Delta\Delta G < +0.05$ ), são descartados do algoritmo. Cada um dos filhos tem também sua imunogenicidade avaliada. O melhor filho dentre os dois gerados é utilizado para dar prosseguimento ao AG.

### 3.3.5.3 Seleção parental

A seleção parental é o processo pelo qual são escolhidos que indivíduos da população serão mutados e recombinados para gerar os indivíduos da geração seguinte. As três formas mais comuns de se selecionar os parentais são aleatória, roleta ou torneio

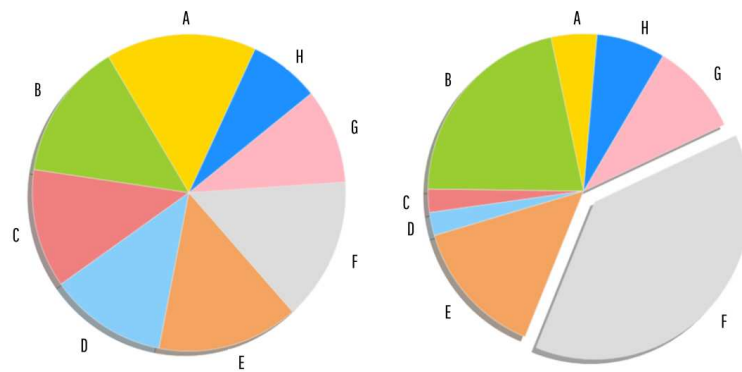
A seleção aleatória consiste em escolher um ou dois indivíduos aleatórios da população, que são selecionados para as operações de mutação e recombinação, respectivamente. Essa metodologia falha ao não discriminar indivíduos mais aptos dos menos aptos, não otimizando o custo computacional do algoritmo e possivelmente aumentando o número de ciclos necessários para atingir o critério de parada estabelecido.

Na roleta, a chance de um indivíduo ser selecionado é proporcional ao seu valor de aptidão. Esse método, se aplicado a uma população com, ao menos, um indivíduo de aptidão consideravelmente superior aos demais, diminui a diversidade populacional ao longo dos ciclos, uma vez que tal indivíduo seria selecionado com frequência, deixando descendentes com suas

características na maior parte das gerações (Figura 17).

Análises preliminares do AG construído indicavam que uso da seleção por roleta traria problemas de diversidade, o que poderia ocorrer devido a presença de indivíduos cuja aptidão superava muito a dos demais. Desse modo, nesse trabalho foi utilizada a seleção por torneio; para este caso, dois indivíduos eram selecionados aleatoriamente da população (100 indivíduos) e o melhor deles era considerado o parental A, e a mesma operação se repetia para seleção do parental B.

Figura 17 – Seleção proporcional a aptidão



Fonte: Elaborada pela autora

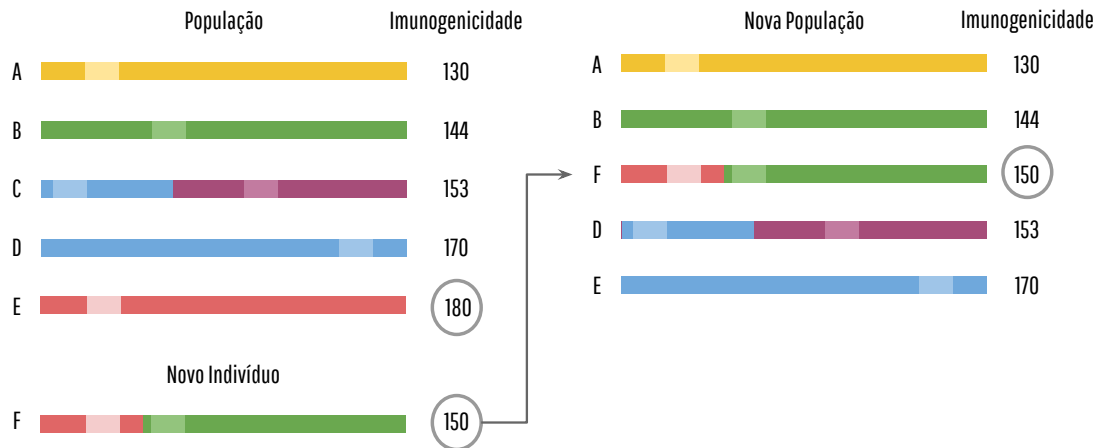
Esquerda: todos os indivíduos A-H tem valores adaptativos semelhantes.

Direita: o indivíduo F tem aptidão muito maior que os demais.

### 3.3.6 Substituição parental

Como consequência das operações de recombinação e mutação, o número de sequências aumenta momentaneamente e substituições devem então ser realizadas para que a população retorne ao seu tamanho original. Caso o valor de imunogenicidade calculado para o filho seja inferior ao de alguma sequência contida na população, esta sequência é substituída e o ciclo recomeça (Figura 18).

Figura 18 – Substituição de sequências



Fonte: Elaborada pela autora

Sequências com escores de imunogenicidade maiores são gradualmente substituídas por aquelas de menor imunogenicidade. No esquema acima o indivíduo F possui imunogenicidade menor que a de alguns indivíduos da população, sendo então incorporado. O indivíduo E de maior imunogenicidade é descartado de forma que a população continue do tamanho original.

### 3.3.7 Convergência e População Final

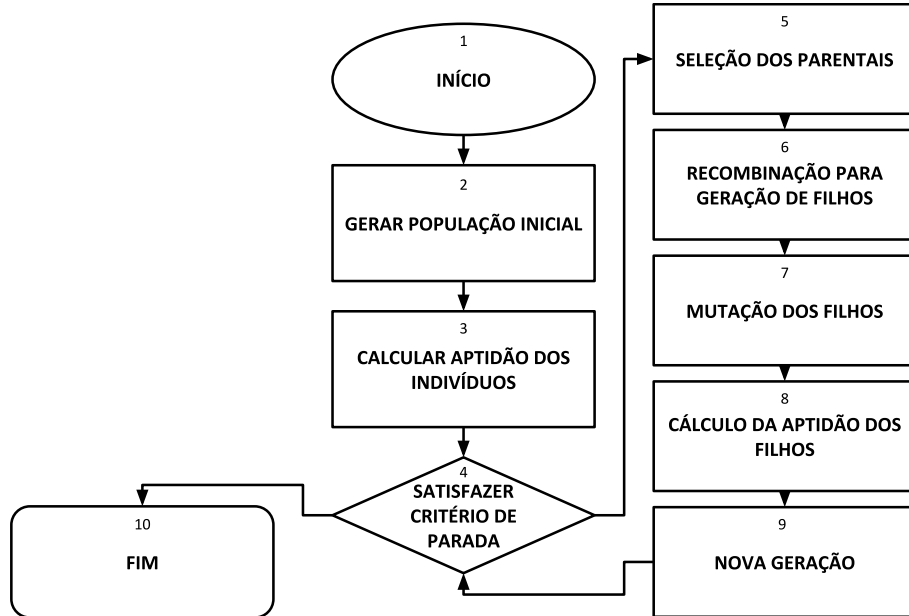
De maneira geral, os valores de aptidão dos indivíduos de um algoritmo genético tendem a convergir após um dado número de gerações. A convergência acontece quando o valor de imunogenicidade não decresce significativamente ao longo de um considerável número de gerações. O número de ciclos necessário para atingir a convergência depende do problema e dos parâmetros utilizados na construção do AG como: variedade da população inicial, taxa de mutação, taxa de recombinação, tamanho da população. No AG construído o critério de parada permaneceu associado ao alcance da convergência e caso um número estabelecido  $n$  de ciclos não fosse suficiente para chegar a essa convergência, o algoritmo era reiniciado partindo da geração  $n$ .

### 3.3.8 Funcionamento do AG

O algoritmo se inicia com uma população de 100 indivíduos<sup>2</sup> cuja aptidão, agora denominada imunogenicidade, é calculada como a soma dos epítopos cujo escore está entre os 20% maiores<sup>3</sup>. Caso o critério de parada não seja atendido<sup>4</sup>, parentais são selecionados por um torneio<sup>5</sup> a fim de serem recombinados e gerarem filhos<sup>6</sup>. O melhor desses filhos é, então, encaminhado para a função de mutação, onde uma posição terá seu aminoácido mutado por outro que permita manter a estabilidade da proteína<sup>7</sup> e caso essa mutação diminua o valor de imunogenicidade do filho, ele passa a incorporar a população do AG, substituindo uma sequência

de imunogenicidade maior<sup>9</sup>. Esse ciclo se repete até que o critério de parada estabelecido seja atendido<sup>4</sup> e então o AG é finalizado<sup>10</sup> (Figura 19).

Figura 19 – Fluxograma de funcionamento do AG



Fonte: Elaborada pela autora

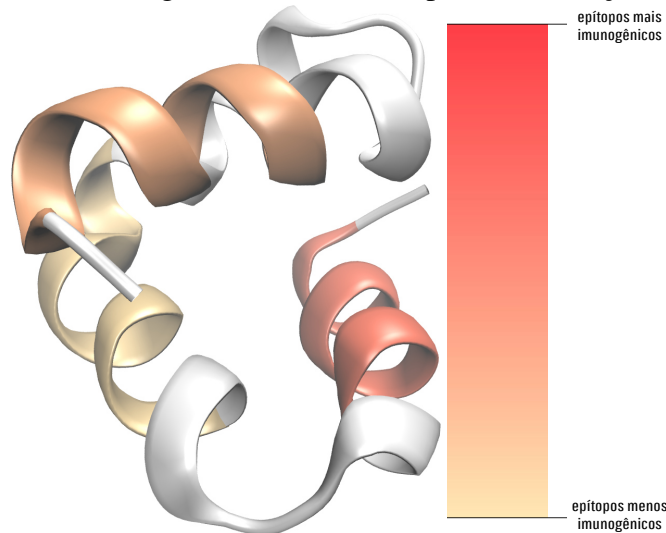
Resumo das operações realizadas pelo algoritmo genético de desimunização desenvolvido em que bordas arredondadas indicam pontos de início ou fim do algoritmo, os retângulos contém processos executados e o losango uma decisão feita com base em uma condição.

### 3.4 Softwares Utilizados

A sequência da L-asparaginase utilizada neste trabalho é uma de *Bacillus subtilis* e foi fornecida por um grupo de pesquisa da Fiocruz Ceará. A L-asparaginase do tipo II de *Bacillus subtilis* foi previamente caracterizada por outros grupos, sendo uma proteína tetramérica com pouca atividade glutaminásica (ONISHI *et al.*, 2011; FENG *et al.*, 2017). Para realização da predição inicial foi utilizada uma implementação do TEPITOPE em Python. Os testes estatísticos T e ANOVA foram automatizados utilizando o módulo Scipy (versão 1.0.0), do Python. Os gráficos foram produzidos utilizando os módulos Matplotlib (versão 2.1.0) e Numpy (versão 1.13.3). O software VMD (versão 1.9.3) foi utilizado para visualização das estruturas e mapeamento dos epítomos. O código utilizado para mutação, escrito em Python, foi adaptado pelo grupo e utiliza o software Modeller (versão 9.17).

O mapeamento de imunogenicidade feito pelo VMD é realizado de forma automatizada por um código em Tcl/Tk desenvolvido pela autora. Este código compartimentaliza os valores de imunogenicidade em 10 faixas de cores linearmente proporcionais (Figura 20).

Figura 20 – Gradiente de imunogenicidade utilizado para visualização no VMD



Fonte: Elaborada pela autora (estrutura obtida do PDB: 2N8O). Os epítomos presentes na estrutura de proteína são identificados de acordo com a escala colorimétrica de imunogenicidade. Os epítomos mais imunogênicos tem cores mais avermelhadas e os menos imunogênicos são aqueles em amarelo. Regiões em branco não tiveram epítomos apontados pelo método de predição.

## 4 RESULTADOS E DISCUSSÃO

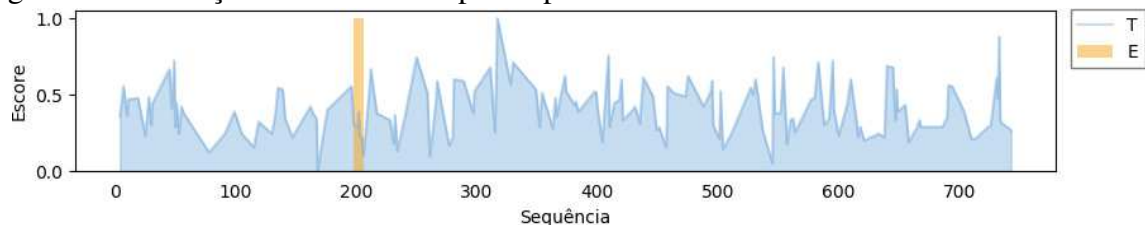
O objetivo desse trabalho foi desenvolver ferramentas computacionais para desimunização de proteínas e aplicá-las a uma sequência de L-asparaginase de *Bacillus subtilis*. Para isso foi realizada (i) a predição *in silico* de epítomos da sequência através de uma modificação do método TEPITOPE, (ii) a identificação dos aminoácidos promíscuos e (iii) a utilização de algoritmo genético para a desimunização da sequência de L-asparaginase.

### 4.1 Avaliação da Imunogenicidade

#### 4.1.1 Predição do TEPITOPE

A predição da imunogenicidade a partir da sequência é utilizada para identificar epítomos que não tenham sido determinados experimentalmente. O método TEPITOPE (subseção 3.1.2), cuja saída aponta nonâmeros e seus escores de imunogenicidade para diferentes alelos do MHC-II, foi escolhido para este trabalho por já ter sido amplamente empregado na identificação de epítomos em outros trabalhos (BRENNAN *et al.*, 2010; DROUIN *et al.*, 2013; GUPTA *et al.*, 2011). Um método de predição de epítomos ideal é capaz de discriminar os diversos nonâmeros avaliados atribuindo-lhes escores de imunogenicidade contrastantes. Existem dois problemas com o modo como os resultados são apresentados por esse método: os escores atribuídos para diferentes epítomos são muitas vezes semelhantes e o método atribui o mesmo escore a todos os aminoácidos de um peptídeo. Os dados obtidos do TEPITOPE não apontam com clareza as regiões mais imunogênicas, pois atribui valores próximos a grande parte dos epítomos (Figura 21).

Figura 21 – Predição do TEPITOPE para a proteína P18884.1



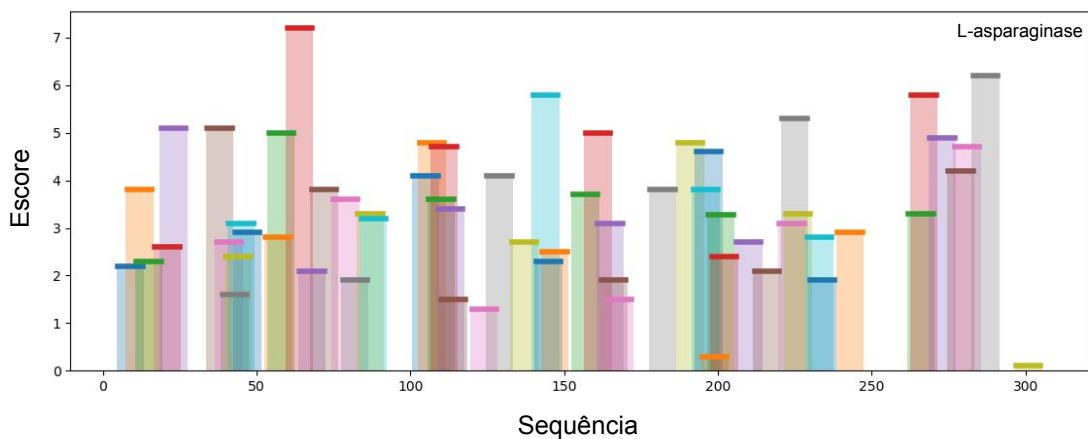
Fonte: Elaborada pela autora

T: escores do TEPITOPE. E: epítomo identificado experimentalmente. Escores de imunogenicidade do TEPITOPE normalizados e atribuídos para cada uma das posições iniciais dos epítomos identificados. Pode-se observar uma baixa distribuição nos escores atribuídos a maioria dos epítomos.

Esse método atribui a todos os resíduos de um peptídeo o mesmo escore e, mesmo

utilizando apenas os nonâmeros de maior escore apontados pelo TEPITOPE, é impossível identificar os aminoácidos promíscuos da sequência. Surge então uma dificuldade em encontrar os resíduos mais propícios a sofrer mutação em uma abordagem de desimunização (Figura 22). A identificação de alguns aminoácidos promíscuos permite trabalhar com um conjunto reduzido de posições-chave para a desimunização e aumenta as chances de sucesso das proteínas propostas, uma vez que um menor número de mutações tem menor potencial de impacto na estabilidade da proteína. Logo, ao invés de utilizar a saída padrão do TEPITOPE, os dados obtidos a partir da saída padrão do TEPITOPE foram adaptados para o evidenciar os aminoácidos promíscuos.

Figura 22 – Epítopos identificados pelo TEPITOPE para a L-asparaginase de *Bacillus subtilis*



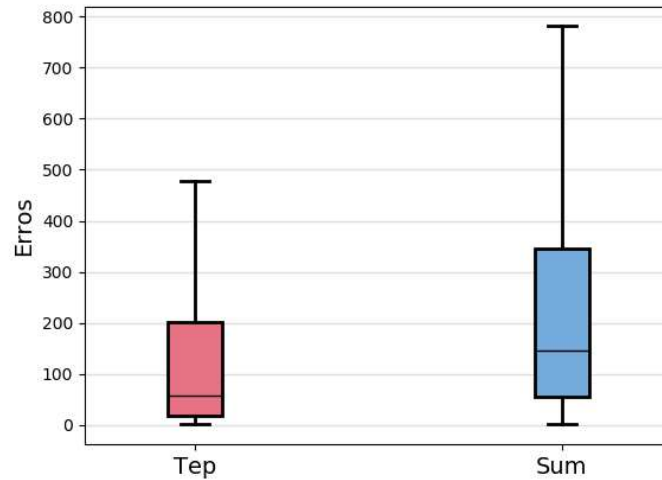
Fonte: Elaborada pela autora.

Diferentes epítopos da sequência de L-asparaginase foram coloridos com diferentes cores. O arquivo de saída do TEPITOPE atribui a todos os aminoácidos de um peptídeo o mesmo escore de imunogenicidade. Existem epítopos que se sobrepõem em muitas posições.

#### 4.1.2 Avaliação do Somatório

A primeira estratégia adotada foi a realização do somatório dos escores de aminoácidos contidos em sobreposições de epítopos, o que permite destacar os epítopos promíscuos (subseção 3.1.3). Entretanto, a distribuição no número de falso-positivos é pior para a estratégia do Somatório que para o método TEPITOPE (Figura 23). São pelo menos 147 erros contabilizados na identificação de 50% dos epítopos identificados pelo TEPITOPE e alguns dos epítopos são identificados após um número de erros muito superior ao do TEPITOPE.

Figura 23 – Distribuição de Erros no Mapeamento de Epítomos

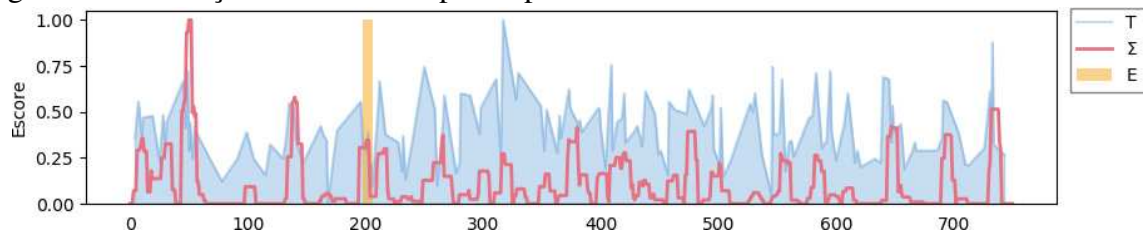


Fonte: Elaborada pela autora

Para o método TEPITOPE, 50% dos epítomos são identificados após pelo menos 58 erros. Para a soma dos escores do TEPITOPE, o número de erros é ainda maior para 50% dos epítomos, pelo menos 147 erros e a variação é a maior dentre as estratégias analisadas. TEP = TEPITOPE, SUM = Somatório.

Possivelmente, isso ocorre porque, mesmo após o Somatório, existem posições de baixa imunogenicidade localizadas entre dois picos obtidos pelo somatório e regiões *indistintas* em que muitos epítomos se sobrepõem uns aos outros em cadeia. Essa diminuição nos escores entre dois valores altos pode dificultar a definição de onde começa e termina um epítopo de interesse, não sendo possível discriminar com clareza regiões de interesse para desimunização (Figura 24).

Figura 24 – Predição do somatório para a proteína P18884.1



Fonte: Elaborada pela autora

T: escores do TEPITOPE.  $\Sigma$ : somatório dos escores do TEPITOPE. E: epítopo identificado experimentalmente. Somatório dos epítomos sobrepostos normalizado em comparação aos dados do TEPITOPE normalizada. Algumas regiões foram destacadas, porém em outras ainda é difícil discriminar aminoácidos de maior importância para desimunização.

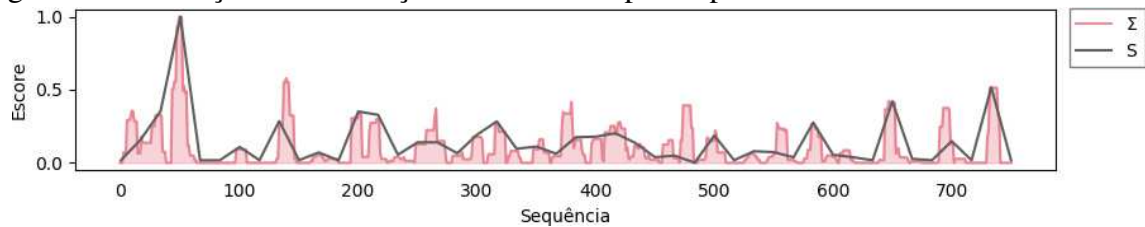
Esses problemas combinados resultam em uma dificuldade na escolha da melhor posição para a desimunização e diminuem a confiabilidade dessa escolha. Além disso, essa variabilidade foi observada nas predições para diversas proteínas, o que apontou espaço para melhorias no método. Tentou-se eliminar essa indiscriminação de epítomos e não-epítomos ainda

presentes com o método Suavização.

#### 4.1.3 Avaliação da Suavização

A decisão pela aplicação de um método de suavização teve como objetivo identificar as regiões de epítomos promíscuos. A escolha da *spline* como forma de suavização dos dados se deu pelo extenso uso dessa abordagem quando se lida com dados complexos cujo comportamento pode ser explicado por funções mais simples. Após essa etapa os dados obtidos mostram variações graduais na pontuação imunogênica das posições ao longo da sequência (Figura 27).

Figura 25 – Predição da suavização do somatório para a proteína P18884.1

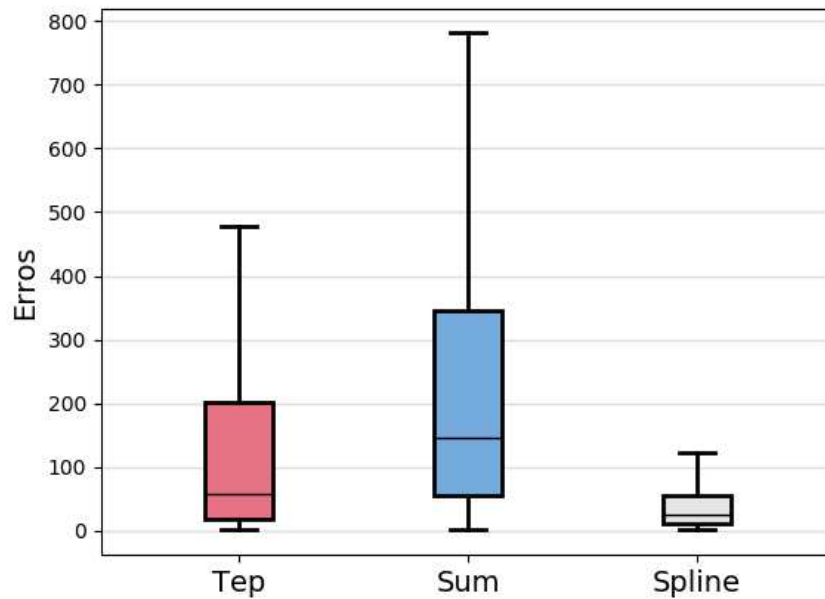


Fonte: Elaborada pela autora

Suavização normalizada do somatório em face da localização do epítipo experimental.  $\Sigma$ : somatório dos escores do TEPITOPE. S: suavização do somatório dos escores do TEPITOPE. E: epítipo identificado experimentalmente.

A distribuição no número de falso-positivos apontados para a Suavização é significativamente menor que os do TEPITOPE e os do Somatório (Figura 26). Dos erros cometidos pela Suavização, 75% estavam na faixa de 0-56.0, enquanto quase 50% dos erros método TEPITOPE e apenas 25% dos erros cometidos pelo Somatório se encontravam nessa faixa. O resultado obtido é estatisticamente significativo e confirma que a estratégia de soma e suavização de escores do TEPITOPE é capaz de identificar aminoácidos promíscuos e seus epítomos.

Figura 26 – Erros no Mapeamento de Epítomos

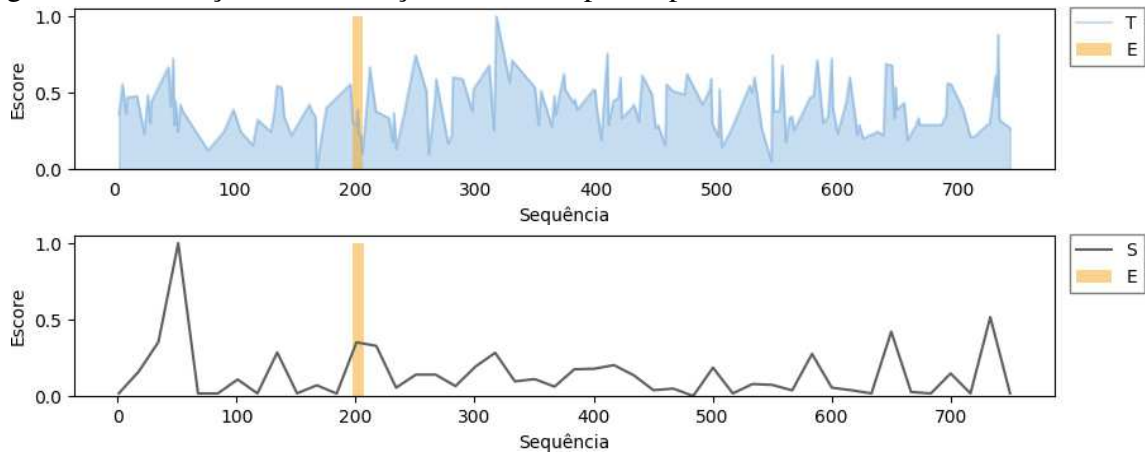


Fonte: Elaborada pela autora  
 Distribuição no número de falso-positivos apontados para todas as proteínas do conjunto teste seção 3.2. TEPITOPE = TEPITOPE, SUM = Somatório, spline = Suavização.  $p < 0.0001$

A suavização permite uma visualização mais clara das regiões imunogênicas em que poucos epítomos ultrapassam um escore de 0.4 e impõe uma hierarquia mais definida entre os aminoácidos promíscuos (Figura 27.2).

Quando esses dados são comparados aos da predição do TEPITOPE torna-se evidente a superioridade da Suavização em discriminar epítomos e apontar aminoácidos promíscuos.

Figura 27 – Predição da suavização somatório para a proteína P18884.1



Fonte: Elaborada pela autora

$\Sigma$ : somatório dos escores do TEPITOPE. S: suavização do somatório dos escores do TEPITOPE. E: epítipo identificado experimentalmente. Acima: predição do TEPITOPE em face da localização do epítipo experimental. Abaixo: suavização normalizada do somatório em face da localização do epítipo experimental. Após a Suavização o epítipo experimental corresponde ao quarto epítipo predito.

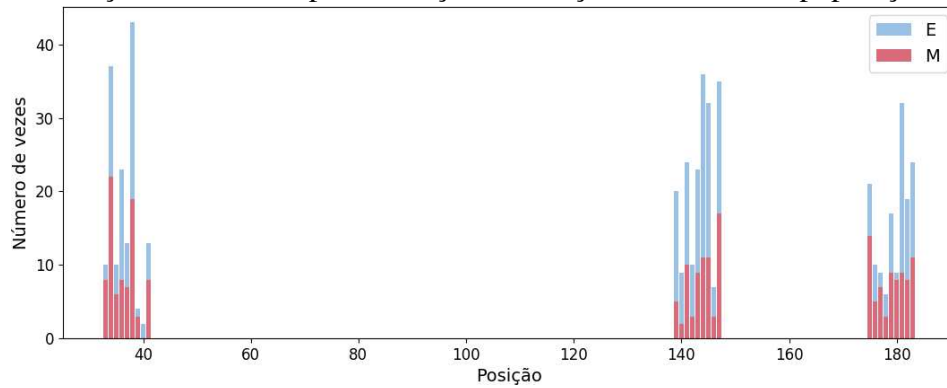
O método de suavização do somatório de epítomos identificados pelo TEPITOPE é uma estratégia de predição de epítomos inovadora em seus princípios. Ao realizar o somatório atribuindo a aminoácidos contidos em diferentes epítomos o valor de imunogenicidade pertencente a aos epítomos em questão, unindo predições feitas para diferentes alelos do MHC-II humano se pôde estabelecer um escore de *imunopromiscuidade* desses aminoácidos. A suavização desses valores de imunopromiscuidade permitiu discriminar resíduos imunogênicos e diminuir o alto número de falsos-positivos que ainda se observava para o método do somatório. Essa redução de falsos positivos é de extrema relevância, uma vez que esforços de desimunização por mutagênese perdem sua eficiência ao realizar mutações em peptídeos que não são apontados como imunogênicos experimentalmente.

De posse da melhor ferramenta de predição de epítomos promíscuos avaliada, a Suavização foi aplicada na avaliação de imunogenicidade de proteínas realizada pelo algoritmo genético de desimunização.

## 4.2 Desimunização

O propósito do trabalho foi construir um algoritmo genético de desimunização de proteínas capaz de apontar mutantes de menor imunogenicidade e com estabilidade comparável a da proteína original. A população era de 100 indivíduos e a imunogenicidade foi calculada como a soma dos 20% piores epítomos encontrados pelo método da Suavização. Devido a restrição de epítomos considerados para o cálculo da imunogenicidade, as mutações realizadas pelo AG só ocorreram em três regiões da sequência da proteína, onde se encontravam os três epítomos mais imunogênicos na sequência de L-asparaginase utilizada, sendo eles Ep30: resíduos 30 a 39, Ep140: 140-149 e Ep175: 175-184 (Figura 28).

Figura 28 – Posições escolhidas para mutação e mutações mantidas na população



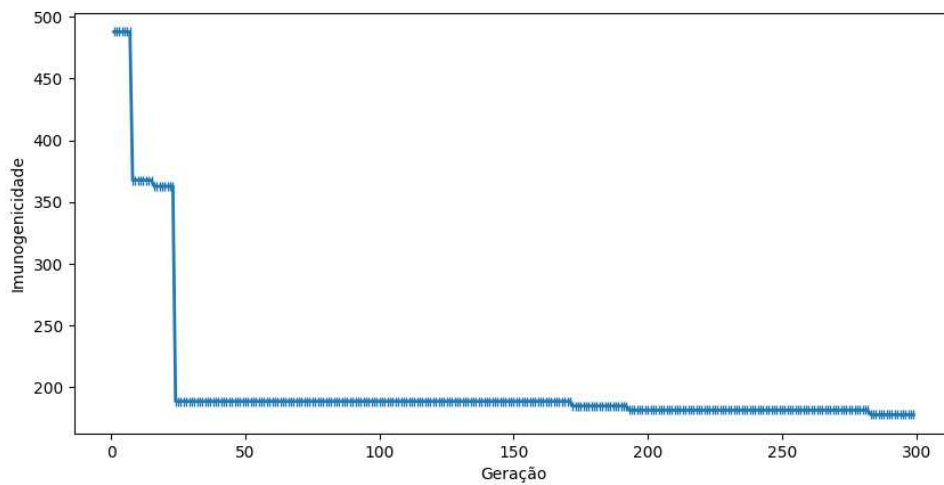
Fonte: Elaborada pela autora

E: mutações escolhidas. M: mutações mantidas. Apenas três regiões se enquadravam entre as 20% mais imunogênicas da proteína, portanto, as operações de mutação realizadas pelo AG aconteceram apenas nessas regiões. Muitas vezes as mutações escolhidas não eram mantidas por não gerar diminuição na imunogenicidade da sequência.

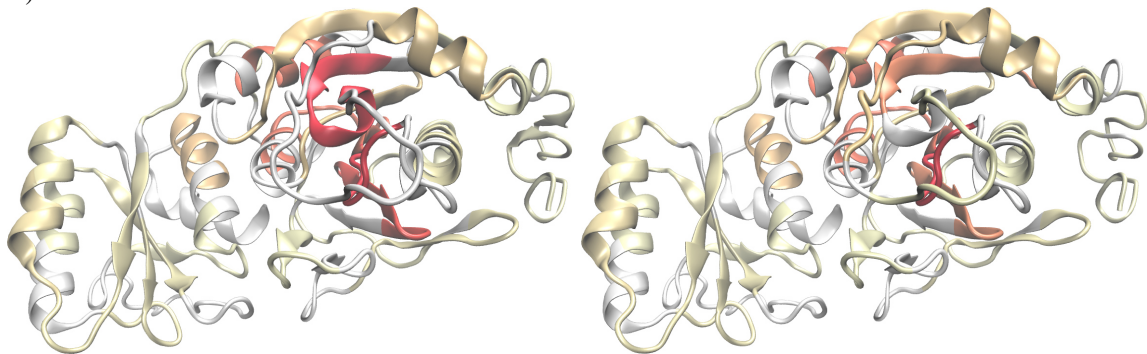
Partindo da sequência original, cuja imunogenicidade somava 487.85. Foram obtidos mutantes pontuais (3), duplos (6), triplos (9) e quádruplos (2) entre os vinte primeiros da população final e a maior imunogenicidade dentre esses foi de 193.18. Esses resultados foram obtidos após trezentas gerações. A evolução do AG e a imunogenicidade calculada para a proteína de menor imunogenicidade é de 177.87 (Figura 29).

Figura 29 – Evolução do Algoritmo Genético de acordo com a imunogenicidade do melhor indivíduo a cada geração

A)



B)



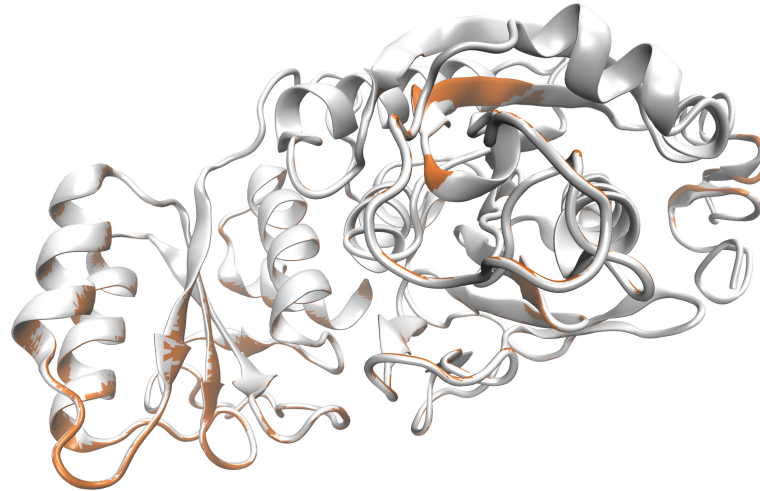
Fonte: Elaborada pela autora

A: O valor de imunogenicidade do melhor indivíduo está representada para cada geração (ciclo) de execução do algoritmo genético e foi reduzida pela metade em menos de cinquenta ciclos. A partir daí as reduções levaram um número maior de ciclos para ocorrer. B: Esquerda: representação estrutural da L-asparaginase selvagem com epítopos mais imunogênicos em vermelho mais forte. Direita: representação estrutural do triplo mutante com redução da imunogenicidade em Ep30 e Ep175.

Ao final dos trezentos ciclos, o melhor indivíduo da população conta com três mutações e tem RMSD de 0.01Å em relação à proteína selvagem. A semelhança entre a estrutura da proteína selvagem e a do triplo mutante é confirmada pelo baixo valor de RMSD entre elas (Figura 30). O  $\Delta\Delta G$  em relação ao da proteína selvagem é de -0.2935 kcal/mol.

Com 300 ciclos, o AG não foi capaz de desimmunizar o Ep140 de maneira a gerar um indivíduo estável e de imunogenicidade menor que os demais indivíduos da população. Possíveis soluções para isso incluem variações no tamanho da população e alterações no número de mutações e recombinações executados por ciclo nos Ep30, Ep140 e Ep175. Por outro lado, caso não sejam encontrados mutantes que conciliem a desimmunização dos três epítopos, isto pode ser um indicativo da possível importância dos resíduos do Ep140 para a manutenção da estabilidade da proteína. Uma estratégia para investigar o papel desse resíduo seria intensificar a

Figura 30 – Sobreposição das estruturas de L-asparaginase de *B. subtilis* selvagem e mutada

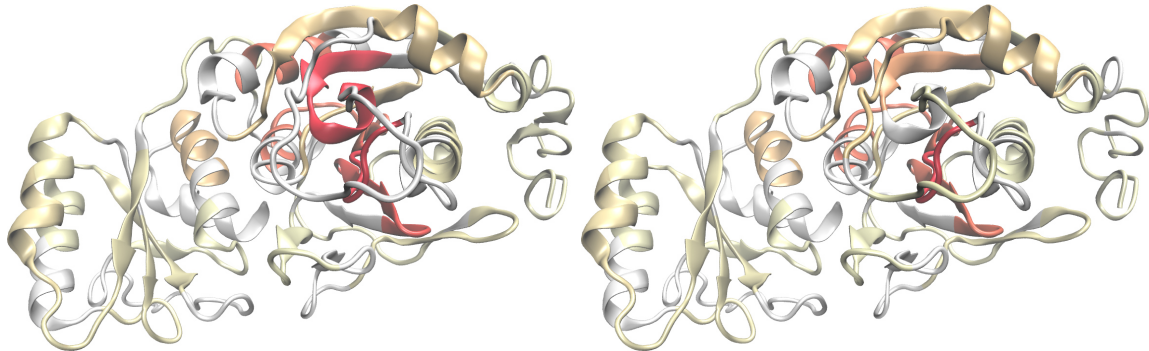


Fonte: Elaborada pela autora  
 Branco: proteína selvagem. Laranja: proteína selvagem. (RMSD: 0.01Å,  
 $\Delta\Delta G = -0.2935$  kcal/mol)

busca local, realizando todas as mutações possíveis no epítipo dessa região e avaliar seu impacto na estabilidade e na imunogenicidade da proteína.

O segundo melhor mutante possui duas mutações e uma diferença de 3.62 no valor de imunogenicidade em relação ao melhor indivíduo. Os mesmos dois epítopos foram eliminados e a diferença de RMSD também é de 0.01Å, e o  $\Delta\Delta G$  é de -0.1924 kcal/mol (Figura 31).

Figura 31 – Mapeamento de epítopos em L-asparaginase de *B. subtilis* mutada em duas posições



Fonte: Elaborada pela autora  
 Esquerda: proteína selvagem. Direita: duplo mutante com redução da imunogenicidade nas regiões dos resíduos Ep30 e Ep175.

Analisando a composição da população é possível observar que a redução no valor de imunogenicidade não é diretamente proporcional ao número de mutações. Os valores de imunogenicidade e o número de mutações dos vinte melhores indivíduos após 300 gerações do algoritmo permitem confirmar esse resultado (Quadro 4). Os indivíduos J e K da população contam com quatro mutações e estão abaixo de indivíduos com uma única mutação como G e H.

Foi ainda avaliada a evolução da população ao longo de 200 ciclos adicionais

Quadro 4 – Vinte primeiras sequências da população final

Indivíduo	Imunogenicidade	Número de Mutações
A	177.87248984	3
B	181.49769884	2
C	183.26250277	3
D	183.66992506	3
E	184.39497231	2
F	184.92853651	3
G	188.65251089	1
H	188.65251443	1
I	189.74914442	3
J	189.74914442	4
K	189.74954005	4
L	190.91199435	2
M	190.91199552	1
N	190.91199552	2
O	192.21781727	2
P	192.21781727	3
Q	192.97533395	2
R	192.97533395	3
S	192.97533395	3
T	193.18369711	3

Fonte: Elaborado pela autora

(totalizando 500 ciclos) para estabilizar e confirmar a composição da população final. Apenas no 438º ciclo foi encontrado um indivíduo de menor imunogenicidade. Embora esse novo mutante contenha quatro mutações, a diferença entre a imunogenicidade do mutante triplo e a deste novo indivíduo é de apenas  $2.249 \times 10^{-6}$ . O maior número de mutações e a diferença desprezível de imunogenicidade podem não justificar a substituição do mutante triplo pelo mutante quádruplo uma vez que cada nova mutação pode estar localizada em um sítio importante para atividade enzimática da proteína. Essa perturbação no sítio ativo só poderia ser verificada experimentalmente.

Os mutantes sugeridos pelo algoritmo genético devem passar por outras etapas de avaliação *in silico*, como avaliações de dinâmica molecular, de forma a estudar a interação do tetrâmetro de L-asparaginase com seu substrato, L-asparagina, e também sua interação com MHC-IIs disponíveis no banco de proteínas PDB. Confirmada a desimunização e mantida a interação proteína-substrato a proteína deve ser então expressa, purificada, caracterizada e submetida a imunoenaios.

Quando comparado com outro trabalho de desimunização *in silico* de L-asparaginases (RAMYA; PULICHERLA, 2015), o protocolo aqui desenvolvido tem como vantagens o emprego de todos os alelos disponíveis para predição no TEPITOPE e a melhoria no método de predição de imunogenicidade utilizado. Adicionalmente, pelo que indicam os dados expostos por

RAMYA; PULICHERLA (2015), apenas a região mais imunogênica de cada L-asparaginase foi considerada para desimunização apesar da presença de muitos outros peptídeos com escore de imunogenicidade próximos. Além disso, ainda no trabalho de RAMYA; PULICHERLA (2015), não são descritos o número de mutações avaliado para uma dada posição e nem quantas posições imunogênicas tiveram seus mutantes avaliados.

### 4.3 Ferramentas Desenvolvidas

As ferramentas aqui desenvolvidas objetivam automatizar a geração de dados a partir de um grande número de sequências, assim como a avaliá-los. Isso permite que esses programas utilizem dados gerados por estratégias de alta vazão (*high throughput*) e auxiliem trabalhos de engenharia de proteínas.

A Suavização é um método capaz de identificar com acurácia epítomos lineares de linfócitos T e que funciona pela busca de aminoácidos promíscuos em uma sequência de proteínas, podendo ser utilizado para quaisquer sequências e modificado para receber informações de outros métodos de predição de epítomos.

O algoritmo genético desenvolvido para este trabalho foi feito para desimunizar qualquer proteína de cadeia única. Ele pode, ainda, ser modificado para atender a outros objetivos como aumento de termoestabilidade, interações proteína-proteína, afinidade enzima-substrato, sendo necessárias apenas a integração de módulos correspondentes a esses objetivos e a alteração da função de aptidão utilizada.

O AG apresenta espaço para melhorias em seus parâmetros e construção, como: congelamento das posições do sítio ativo como imutáveis (caso conhecidas), adição de outros métodos de predição de imunogenicidade, realização de mutações em cada sítio imunogênico a cada geração, inserção de um operador de busca local, incorporação de outras funções de avaliação de imunogenicidade e paralelização do algoritmo.

## 5 CONCLUSÕES

Este trabalho partiu de uma revisão dos métodos de predição de epítomos ligantes ao MHC-II e realizou avaliações do método TEPITOPE com o objetivo de encontrar os epítomos reconhecidos por um grande número de alelos de MHC-II. Para tanto foi necessário introduzir o conceito de aminoácido promíscuo e o método Suavização, que foi desenvolvido para atender a essa necessidade, demonstrou acurácia significativamente maior que os demais métodos analisados.

Os resultados do método Suavização foram utilizados para mapear epítomos presentes em uma L-asparaginase de *Bacillus subtilis*, o que permite ao método ser aplicado para outras proteínas.

O algoritmo genético de desimunização de proteínas aqui desenvolvido foi capaz de gerar um grande número de mutantes de L-asparaginase estáveis e menos imunogênicos, de acordo com os parâmetros *in silico* empregados. Dentre os mutantes propostos ao fim de 300 ciclos, o melhor indivíduo possuía imunogenicidade 63.5% menor que a da sequência original, com mutações em dois dos três epítomos de maior imunogenicidade sem prejuízos à estrutura e estabilidade.

O método da Suavização, o código de mapeamento e visualização de epítomos e o algoritmo genético de desimunização podem ser aplicados para outras sequências imunogênicas. Ademais, algoritmo genético desenvolvido pode ser adaptado para otimização de proteínas de acordo com outros parâmetros. Sendo assim essas ferramentas tem potencial para diminuir o tempo de busca de mutações desimunizantes (ou com outros fins) em proteínas terapêuticas, contribuindo assim com a pesquisa para a produção de *biobetters* nacionais.

## REFERÊNCIAS

- ABBAS, A. K.; LICHTMAN, A. H.; PILLAI, S. **Cellular and Molecular Immunology E-Book**. [S.l.]: Elsevier Health Sciences, 2014.
- ALI, U.; NAVEED, M.; ULLAH, A.; ALI, K.; SHAH, S. A.; FAHAD, S.; MUMTAZ, A. S. L-asparaginase as a critical component to combat acute lymphoblastic leukaemia (all): A novel approach to target all. **European Journal of Pharmacology**, 2016.
- ANSARI, H. R.; RAGHAVA, G. P. Identification of conformational B-cell Epitopes in an antigen from its primary sequence. **Immunome Research**, 2010.
- BACKERT, L.; KOHLBACHER, O. Immunoinformatics and epitope prediction in the age of genomic medicine. **Genome medicine**, BioMed Central, v. 7, n. 1, p. 119, 2015.
- BAKER, M.; REYNOLDS, H. M.; LUMICISI, B.; BRYSON, C. J. Immunogenicity of protein therapeutics: the key causes, consequences and challenges. **Self/nonself**, Taylor & Francis, v. 1, n. 4, p. 314–322, 2010.
- BALDO, B. A. **Safety of Biologics Therapy**. [S.l.]: Springer, 2016.
- BHASIN, M.; RAGHAVA, G. A hybrid approach for predicting promiscuous mhc class i restricted t cell epitopes. **Journal of biosciences**, IAS, v. 32, n. 1, p. 31–42, 2007.
- BLYTHE, M. J.; FLOWER, D. R. Benchmarking b cell epitope prediction: underperformance of existing methods. **Protein Science**, Wiley Online Library, v. 14, n. 1, p. 246–248, 2005.
- BRENNAN, F. R.; MORTON, L. D.; SPINDELDREHER, S.; KIESSLING, A.; ALLENSPACH, R.; HEY, A.; MÜLLER, P.; FRINGS, W.; SIMS, J. Safety and immunotoxicity assessment of immunomodulatory monoclonal antibodies. In: TAYLOR & FRANCIS. **MAbs**. [S.l.], 2010. v. 2, n. 3, p. 233–255.
- BROOME, J. Evidence that the l-asparaginase activity of guinea pig serum is responsible for its antilymphoma effects. **Nature**, Springer, v. 191, n. 4793, p. 1114–1115, 1961.
- BUI, H. H.; SIDNEY, J.; PETERS, B.; SATHIAMURTHY, M.; SINICHI, A.; PURTON, K. A.; MOTHÉ, B. R.; CHISARI, F. V.; WATKINS, D. I.; SETTE, A. Automated generation and evaluation of specific MHC binding predictive tools: ARB matrix applications. **Immunogenetics**, 2005.
- CAMBRICOLI, F. **Gasto federal com remédios sobe 53%**. 2015. Disponível em: <http://www.cff.org.br/noticia.php?id=2935>. Acesso em: 18 setembro 2016.
- CARTER, P. J. Introduction to current and future protein therapeutics: a protein engineering perspective. **Experimental cell research**, Elsevier, v. 317, n. 9, p. 1261–1269, 2011.
- CASADEVALL, N.; NATAF, J.; VIRON, B.; KOLTA, A.; KILADJIAN, J.-J.; MARTIN-DUPONT, P.; MICHAUD, P.; PAPO, T.; UGO, V.; TEYSSANDIER, I. *et al.* Pure red-cell aplasia and antierythropoietin antibodies in patients treated with recombinant erythropoietin. **New England Journal of Medicine**, Mass Medical Soc, v. 346, n. 7, p. 469–475, 2002.
- CASTELLINO, F.; ZHONG, G.; GERMAIN, R. N. Antigen presentation by mhc class ii molecules: invariant chain function, protein trafficking, and the molecular basis of diverse determinant capture. **Human immunology**, Elsevier, v. 54, n. 2, p. 159–169, 1997.

CFM. **Justiça Federal proíbe uso do medicamento Leuginase no tratamento de câncer infantil**. 2017. Disponível em: <http://portal.cfm.org.br/index.php?option=comcontentview=articleid=27194:2017-09-26-19-12-02catid=3:portal>. Acesso em: 11 dezembro 2017.

CLEMENTI, A. La désamidation enzymatique de l'asparagine chez les différentes espèces animales et la signification physio logique de sa presence dans l'organisme. **Archives Internationales de Physiologie**, Taylor & Francis, v. 19, n. 4, p. 369–398, 1922.

CRUSE, J.; LEWIS, R.; WANG, H. **Immunology guidebook**. [S.l.]: Academic Press, 2004.

DE, R. K.; TOMAR, N. **Immunoinformatics**. [S.l.]: Springer, 2014. 23–58 p.

DEEHAN, M.; GARCÊS, S.; KRAMER, D.; BAKER, M. P.; RAT, D.; ROETTGER, Y.; KROMMINGA, A. Managing unwanted immunogenicity of biologicals. **Autoimmunity reviews**, Elsevier, v. 14, n. 7, p. 569–574, 2015.

DESAI, D. V.; KULKARNI-KALE, U. T-cell epitope prediction methods: an overview. **Immunoinformatics**, Springer, p. 333–364, 2014.

DEWAN, S. S.; SULLIVAN, L. L. **PROTEIN THERAPEUTICS MARKET - Technology Advances Spur Market Growth of Protein Therapies**. 2016. Disponível em: <http://www.drug-dev.com/Main/Back-Issues/PROTEIN-THERAPEUTICS-MARKET-Technology-Advances-Sp-1211.aspx>. Acesso em: 8 dezembro 2017.

DIMITROV, D. S. Therapeutic proteins. **Therapeutic Proteins: Methods and Protocols**, Springer, p. 1–26, 2012.

DIMITROV, I.; GARNEV, P.; FLOWER, D. R.; DOYTCHINOVA, I. EpiTOP-a proteochemometric tool for MHC class II binding prediction. **Bioinformatics**, 2010.

DROUIN, E. E.; SEWARD, R. J.; STRLE, K.; MCHUGH, G.; KATCHAR, K.; LONDONO, D.; YAO, C.; COSTELLO, C. E.; STEERE, A. C. A novel human autoantigen, endothelial cell growth factor, is a target of t and b cell responses in patients with lyme disease. **Arthritis & Rheumatology**, Wiley Online Library, v. 65, n. 1, p. 186–196, 2013.

FELDHahn, M.; DÖNNES, P.; THIEL, P.; KOHLBACHER, O. Fred—a framework for t-cell epitope detection. **Bioinformatics**, Oxford University Press, v. 25, n. 20, p. 2758–2759, 2009.

FENG, Y.; LIU, S.; JIAO, Y.; GAO, H.; WANG, M.; DU, G.; CHEN, J. Enhanced extracellular production of l-asparaginase from bacillus subtilis 168 by b. subtilis wb600 through a combined strategy. **Applied microbiology and biotechnology**, Springer, v. 101, n. 4, p. 1509–1520, 2017.

FLOWER, D. R. **Immunoinformatics: Predicting immunogenicity in silico**. [S.l.]: Springer Science & Business Media, 2007.

FOOD; ADMINISTRATION, D. *et al.* Immunogenicity assessment for therapeutic protein products. **FDA: Rockville, MD (<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM338856.pdf>)**, 2013.

FURTH, O. v.; FRIEDMANN, M. Über die verbreitung asparaginspaltender organfermente. **Biochem Z**, v. 26, p. 435–440, 1910.

- GADELHA, M. I. **Esclarecimentos L-Asparaginase**. 2017. Disponível em: <[http://portal.arquivos.saude.gov.br/images/pdf/2017/junho/22/2.%20b%20-%20L-ASPARAGINASE\\_CIT\\_22\\_06\\_2017.pdf](http://portal.arquivos.saude.gov.br/images/pdf/2017/junho/22/2.%20b%20-%20L-ASPARAGINASE_CIT_22_06_2017.pdf)>. Acesso em 08 out. 2017.
- GIOVANNONI, G. Strategies to treat and prevent the development of neutralizing anti-interferon- $\beta$  antibodies. **Neurology**, AAN Enterprises, v. 61, n. 9 suppl 5, p. S13–S17, 2003.
- GRISWOLD, K. E.; BAILEY-KELLOGG, C. Design and engineering of deimmunized biotherapeutics. **Current opinion in structural biology**, Elsevier, v. 39, p. 79–88, 2016.
- GROOT, A. S. D.; SCOTT, D. W. Immunogenicity of protein therapeutics. **Trends in immunology**, Elsevier, v. 28, n. 11, p. 482–490, 2007.
- GUPTA, S. K.; SRIVASTAVA, M.; AKHOON, B. A.; SMITA, S.; SCHMITZ, U.; WOLKENHAUER, O.; VERA, J.; GUPTA, S. K. Identification of immunogenic consensus t-cell epitopes in globally distributed influenza-a h1n1 neuraminidase. **Infection, Genetics and Evolution**, Elsevier, v. 11, n. 2, p. 308–319, 2011.
- HE, L.; FRIEDMAN, A. M.; BAILEY-KELLOGG, C. A divide-and-conquer approach to determine the pareto frontier for optimization of protein engineering experiments. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 80, n. 3, p. 790–806, 2012.
- HOLLAND, J. H. **Adaptation in Natural and Artificial Systems: An introductory Analysis with Applications to Biology, Control and Artificial Intelligence**. [s.n.], 1975. 183 p. ISSN 10834419. ISBN 0262581116. Disponível em: <<http://www.citeulike.org/group/664/article/400721>>.
- IFPMA. **THE PHARMACEUTICAL INDUSTRY AND GLOBAL HEALTH**. [S.l.], 2017.
- JAFFE, N.; TRAGGIS, D.; DAS, L.; MOLONEY, W.; HANN, H.; KIM, B.; NAIR, R. L-asparaginase in the treatment of neoplastic diseases in children. **Cancer research**, AACR, v. 31, n. 7, p. 942–949, 1971.
- JAWA, V.; COUSENS, L. P.; AWWAD, M.; WAKSHULL, E.; KROPSHOFER, H.; GROOT, A. S. . D. T-cell dependent immunogenicity of protein therapeutics: Preclinical assessment and mitigation. **Clinical Immunology**, 2013.
- JONES, T. D.; CROMPTON, L. J.; CARR, F. J.; BAKER, M. P. Deimmunization of monoclonal antibodies. **Therapeutic Antibodies: Methods and Protocols**, Springer, p. 405–423, 2009.
- KIDD, J. G. Regression of transplanted lymphomas induced in vivo by means of normal guinea pig serum. **Journal of Experimental Medicine**, Rockefeller University Press, v. 98, n. 6, p. 583–606, 1953.
- KORBER, B.; LABUTE, M.; YUSIM, K. Immunoinformatics comes of age. **PLoS Computational Biology**, Public Library of Science, v. 2, n. 6, p. e71, 2006.
- KRISHNA, M.; NADLER, S. G. Immunogenicity to biotherapeutics—the role of anti-drug immune complexes. **Frontiers in immunology**, Frontiers Media SA, v. 7, 2016.
- KUMAR, S.; SINGH, S. K.; WANG, X.; RUP, B.; GILL, D. Coupling of aggregation and immunogenicity in biotherapeutics: T-and b-cell immune epitopes may contain aggregation-prone regions. **Pharmaceutical research**, Springer, v. 28, n. 5, p. 949–961, 2011.

- LANG, S. Uber desamidierung im tierkorper. **Beitr chem Physiol Pathol**, v. 5, p. 321–345, 1904.
- LATA, S.; BHASIN, M.; RAGHAVA, G. P. S. Application of machine learning techniques in predicting MHC binders. **Methods in molecular biology (Clifton, N.J.)**, 2007.
- LEADER, B.; BACA, Q. J.; GOLAN, D. E. Protein therapeutics: a summary and pharmacological classification. **Nature Reviews Drug Discovery**, 2008.
- LIU, J.; WANG, L.; ZHAO, F.; TSENG, S.; NARAYANAN, C.; SHURA, L.; WILLINGHAM, S.; HOWARD, M.; PROHASKA, S.; VOLKMER, J. *et al.* Pre-clinical development of a humanized anti-cd47 antibody with anti-cancer therapeutic potential. **PloS one**, Public Library of Science, v. 10, n. 9, p. e0137345, 2015.
- MÁRQUEZ-CHAMORRO, A. E.; ASECIO-CORTÉS, G.; SANTIESTEBAN-TOCA, C. E.; AGUILAR-RUIZ, J. S. Soft computing methods for the prediction of protein tertiary structures: A survey. **Applied Soft Computing**, Elsevier, v. 35, p. 398–410, 2015.
- MASHBURN, L. T.; WRISTON, J. C. Tumor inhibitory effect of l-asparaginase from escherichia coli. **Archives of Biochemistry and Biophysics**, Academic Press, v. 105, n. 2, p. 450–453, 1964.
- MUKOVOZOV, I.; SABLJIC, T.; HORTELANO, G.; OFOSU, F. A. *et al.* Factors that contribute to the immunogenicity of therapeutic recombinant human proteins. **THROMBOSIS AND HAEMOSTASIS-STUTTGART-**, FK SCHATTAUER VERLAGSGESELLSCHAFT MBH, v. 99, n. 5, p. 874, 2008.
- NEUMAN, R. E. Dual requirement of walker carcinosarcoma 256 in vitro for asparagine and glutamine. **Science**, v. 124, p. 124–125, 1956.
- NIELSEN, M.; LUND, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. **BMC Bioinformatics**, 2009.
- NIELSEN, M.; LUNDEGAARD, C.; BLICHER, T.; PETERS, B.; SETTE, A.; JUSTESEN, S.; BUUS, S.; LUND, O. Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. **PLoS Computational Biology**, 2008.
- NIELSEN, M.; LUNDEGAARD, C.; LUND, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. **BMC Bioinformatics**, 2007.
- ONISHI, Y.; YANO, S.; THONGSANIT, J.; TAKAGI, K.; YOSHIMUNE, K.; WAKAYAMA, M. Expression in escherichia coli of a gene encoding type ii l-asparaginase from bacillus subtilis, and characterization of its unique properties. **Annals of microbiology**, Springer, v. 61, n. 3, p. 517–524, 2011.
- OYARZÚN, P.; ELLIS, J. J.; BODÉN, M.; KOBE, B. PREDIVAC: CD4+ T-cell epitope prediction for vaccine design that covers 95% of HLA class II DR protein diversity. **BMC bioinformatics**, 2013.
- PARKER, A. S.; CHOI, Y.; GRISWOLD, K. E.; BAILEY-KELLOGG, C. Structure-guided deimmunization of therapeutic proteins. **Journal of Computational Biology**, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 20, n. 2, p. 152–165, 2013.

PARKER, A. S.; GRISWOLD, K. E.; BAILEY-KELLOGG, C. Optimization of therapeutic proteins to delete t-cell epitopes while maintaining beneficial residue interactions. **Journal of bioinformatics and computational biology**, World Scientific, v. 9, n. 02, p. 207–229, 2011.

PARKER, A. S.; ZHENG, W.; GRISWOLD, K. E.; BAILEY-KELLOGG, C. Optimization algorithms for functional deimmunization of therapeutic proteins. **BMC bioinformatics**, BioMed Central, v. 11, n. 1, p. 180, 2010.

PARKER, K. C.; BEDNAREK, M. A.; COLIGAN, J. E. Scheme for ranking potential hla-a2 binding peptides based on independent binding of individual peptide side-chains. **The Journal of Immunology**, Am Assoc Immunol, v. 152, n. 1, p. 163–175, 1994.

PARKIN, J.; COHEN, B. An overview of the immune system. **The Lancet**, Elsevier, v. 357, n. 9270, p. 1777–1789, 2001.

PETERSON, R. G.; HANDSCHUMACHER, R. E.; MITCHELL, M. S. Immunological responses to l-asparaginase. **Journal of Clinical Investigation**, American Society for Clinical Investigation, v. 50, n. 5, p. 1080, 1971.

RAMYA, L.; PULICHERLA, K. K. Studies on deimmunization of antileukaemic l-asparaginase to have reduced clinical immunogenicity-an in silico approach. **Pathology & Oncology Research**, Springer, v. 21, n. 4, p. 909–920, 2015.

RATANJI, K. D.; DERRICK, J. P.; DEARMAN, R. J.; KIMBER, I. Immunogenicity of therapeutic proteins: influence of aggregation. **Journal of immunotoxicology**, Taylor & Francis, v. 11, n. 2, p. 99–109, 2014.

RATH, T.; BAKER, K.; DUMONT, J. A.; PETERS, R. T.; JIANG, H.; QIAO, S.-W.; LENCER, W. I.; PIERCE, G. F.; BLUMBERG, R. S. Fc-fusion proteins and fcγn: structural insights for longer-lasting and more effective therapeutics. **Critical reviews in biotechnology**, Taylor & Francis, v. 35, n. 2, p. 235–254, 2015.

RAU, R.; SANDER, O.; RIEL, P. van; PUTTE, L. van de; HASLER, F.; ZAUG, M.; KNEER, J.; AUWERA, P. van der; STEVENS, R. M. *et al.* Intravenous human recombinant tumor necrosis factor receptor p55-fc igg1 fusion protein ro 45-2081 (lenercept): a double blind, placebo controlled dose-finding study in rheumatoid arthritis. **The Journal of rheumatology**, The Journal of Rheumatology, v. 30, n. 4, p. 680–690, 2003.

REGENMORTEL, M. H. V. Mapping epitope structure and activity: from one-dimensional prediction to four-dimensional description of antigenic specificity. **Methods**, Elsevier, v. 9, n. 3, p. 465–472, 1996.

REGENMORTEL, M. H. V. What is a b-cell epitope? In: \_\_\_\_\_. **Epitope Mapping Protocols: Second Edition**. Totowa, NJ: Humana Press, 2009. p. 3–20. ISBN 978-1-59745-450-6. Disponível em: <[http://dx.doi.org/10.1007/978-1-59745-450-6\\_1](http://dx.doi.org/10.1007/978-1-59745-450-6_1)>.

SALVAT, R. S.; PARKER, A. S.; CHOI, Y.; BAILEY-KELLOGG, C.; GRISWOLD, K. E. Mapping the pareto optimal design space for a functionally deimmunized biotherapeutic candidate. **PLoS computational biology**, Public Library of Science, v. 11, n. 1, p. e1003988, 2015.

- SALVAT, R. S.; VERMA, D.; PARKER, A. S.; KIRSCH, J. R.; BROOKS, S. A.; BAILEY-KELLOGG, C.; GRISWOLD, K. E. Computationally optimized deimmunization libraries yield highly mutated enzymes with low immunogenicity and enhanced activity. **Proceedings of the National Academy of Sciences**, National Acad Sciences, p. 201621233, 2017.
- SCHELLEKENS, H. Immunogenicity of therapeutic proteins: clinical implications and future prospects. **Clinical therapeutics**, Elsevier, v. 24, n. 11, p. 1720–1740, 2002.
- SCHMIDT, S. R. Fusion-proteins as biopharmaceuticals—applications and challenges. **Curr Opin Drug Discov Devel**, v. 12, n. 2, p. 284–295, 2009.
- SETTE, A.; BUUS, S.; APPELLA, E.; ADORINI, L.; GREY, H. M. Structural requirements for the interaction between class ii mhc molecules and peptide antigens. **Immunologic research**, Springer, v. 9, n. 1, p. 2–7, 1990.
- SINCLAIR, A. M.; ELLIOTT, S. Glycoengineering: the effect of glycosylation on the properties of therapeutic proteins. **Journal of pharmaceutical sciences**, Wiley Online Library, v. 94, n. 8, p. 1626–1635, 2005.
- SMITH, P. L. Curve fitting and modeling with splines using statistical variable selection techniques. 1982.
- STURNIOLO, T.; BONO, E.; DING, J.; RADDRIZZANI, L.; TUERECI, O.; SAHIN, U.; BRAXENTHALER, M.; GALLAZZI, F.; PROTTI, M. P.; SINIGAGLIA, F. *et al.* Generation of tissue-specific and promiscuous hla ligand databases using dna microarrays and virtual hla class ii matrices. **Nature biotechnology**, Nature Publishing Group, v. 17, n. 6, p. 555–561, 1999.
- TALLAL, L.; TAN, C.; OETTGEN, H.; WOLLNER, N.; MCCARTHY, M.; HELSON, L.; BURCHENAL, J.; KARNOFSKY, D.; MURPHY, M. L. E. coli l-asparaginase in the treatment of leukemia and solid tumors in 131 children. **Cancer**, Wiley Online Library, v. 25, n. 2, p. 306–320, 1970.
- TANGRI, S.; MOTHÉ, B. R.; EISENBRAUN, J.; SIDNEY, J.; SOUTHWOOD, S.; BRIGGS, K.; ZINCKGRAF, J.; BILSEL, P.; NEWMAN, M.; CHESNUT, R. *et al.* Rationally engineered therapeutic proteins with reduced immunogenicity. **The Journal of Immunology**, Am Assoc Immunol, v. 174, n. 6, p. 3187–3196, 2005.
- TURECEK, P. L.; BOSSARD, M. J.; SCHOETENS, F.; IVENS, I. A. Pegylation of biopharmaceuticals: a review of chemistry and nonclinical safety information of approved drugs. **Journal of pharmaceutical sciences**, Elsevier, v. 105, n. 2, p. 460–475, 2016.
- WALSH, G. **Proteins: biochemistry and biotechnology**. [S.l.: s.n.], 2002.
- WALSH, G. Biopharmaceutical benchmarks 2014. **Nature biotechnology**, Nature Research, v. 32, n. 10, p. 992–1000, 2014.
- WANG, P.; SIDNEY, J.; DOW, C.; MOTHÉ, B.; SETTE, A.; PETERS, B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. **PLoS Computational Biology**, 2008.
- WANG, P.; SIDNEY, J.; KIM, Y.; SETTE, A.; LUND, O.; NIELSEN, M.; PETERS, B. Peptide binding predictions for hla dr, dp and dq molecules. **BMC bioinformatics**, 2010.

XINJIE, Y.; MITSUO, G. **Introduction to Evolutionary Algorithms**. [S.l.: s.n.], 2010. 422 p. ISSN 1098-6596. ISBN 9788578110796.

YANG, X.; YU, X. An introduction to epitope prediction methods and software. **Reviews in medical virology**, Wiley Online Library, v. 19, n. 2, p. 77–96, 2009.

ZAMBIDIS, E. T.; SCOTT, D. W. Epitope-specific tolerance induction with an engineered immunoglobulin. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 93, n. 10, p. 5019–5024, 1996.

ZHANG, F.; LIU, M.-r.; WAN, H.-t. Discussion about several potential drawbacks of pegylated therapeutic proteins. **Biological and Pharmaceutical Bulletin**, The Pharmaceutical Society of Japan, v. 37, n. 3, p. 335–339, 2014.

ZHANG, G. L.; DELUCA, D. S.; KESKIN, D. B.; CHITKUSHEV, L.; ZLATEVA, T.; LUND, O.; REINHERZ, E. L.; BRUSIC, V. MULTIPRED2: A computational system for large-scale identification of peptides predicted to bind to HLA supertypes and alleles. **Journal of Immunological Methods**, 2011.

ZHANG, L.; CHEN, Y.; WONG, H. S.; ZHOU, S.; MAMITSUKA, H.; ZHU, S. Tepitopepan: Extending tepitope for peptide binding prediction covering over 700 hla-dr molecules. **PLoS ONE**, 2012.

ZHAO, H.; VERMA, D.; LI, W.; CHOI, Y.; NDONG, C.; FIERING, S. N.; BAILEY-KELLOGG, C.; GRISWOLD, K. E. Depletion of t cell epitopes in lysostaphin mitigates anti-drug antibody response and enhances antibacterial efficacy in vivo. **Chemistry & biology**, Elsevier, v. 22, n. 5, p. 629–639, 2015.