

# DESENVOLVIMENTO DE MÉTODOS DE BIOINFORMÁTICA PARA BUSCA IN-SILICO DE EPÍTOPOS T CD8+ HUMANOS NO PROTEOMA DE LEISHMANIA BRAZILIENSIS

**Rafael Tosta Santos<sup>1</sup>; Angelo Amâncio Duarte<sup>2</sup>**

1. Bolsista PROBIC/CNPq, Graduando em Engenharia de Computação, Universidade Estadual de Feira de Santana, e-mail: [rafa.ecomp@gmail.com](mailto:rafa.ecomp@gmail.com)

2. Orientador, Departamento de Tecnologia, Universidade Estadual de Feira de Santana, e-mail: [angeloduarte@ecomp.uefs.br](mailto:angeloduarte@ecomp.uefs.br)

**PALAVRAS-CHAVE:** *Web Robots*, epítomos, in-silico.

## INTRODUÇÃO

A leishmaniose é um sério problema de saúde pública e o desenvolvimento de vacinas eficientes é prejudicado pela falta de compreensão global dos mecanismos de defesa do organismo e sua patogênese. Uma atividade chave para o desenvolvimento de vacinas é a identificação dos epítomos que disparam a atividade imunológica no hospedeiro. Essa identificação é custosa tanto em tempo quanto em recursos financeiros, dadas às exigências dos processos de identificação in-vitro (testes laboratoriais) e, principalmente, por causa da enorme diversidade de epítomos originados do proteoma dos patógenos.

Com o avanço da tecnologia diversas áreas de conhecimento e pesquisas estão cada vez mais buscando novos métodos e aplicações para a resolução de problemas. O avanço da informática proporciona um grande poder computacional com o surgimento de novas abordagens para a solução de diversos problemas em diferentes áreas de conhecimento que tornam os custos mais reduzidos, não apenas nos recursos financeiros, mas também nas horas gastas com profissionais, materiais, etc. (DOYTCHINOVA; GUAN; FLOWER, 2008).

A bioinformática é uma das diversas ramificações da informática e visa resolver problemas biológicos aproveitando recursos computacionais como ferramenta de apoio em áreas como a biologia e a medicina. Tornando-se uma vertente amplamente estudada e no processo de desenvolvimento, necessitando assim, de ferramentas novas que atendam as suas crescentes necessidades (BALDI; BRUNAK, 2001).

Atualmente existem disponíveis para consultas pela Internet, *sites* que disponibilizam métodos de predição de epítomos, sendo que cada método utiliza seu próprio algoritmo de predição e apresenta seu próprio índice de classificação para os epítomos. Através de consultas automatizadas aos *sites* preditores e com a utilização e comparação dos diferentes resultados facilita a identificação de potenciais epítomos que promovem resposta imunológica no hospedeiro, reduzindo o universo dos epítomos possíveis para buscas in-vitro.

## METODOLOGIA

Para o acesso automático aos *sites* preditores, foram desenvolvidos sistemas capazes de realizar consultas a *sites* na Internet e obter os resultados gerados através de informações disponibilizadas no formato HTML e extraíndo os dados desejados, esses sistemas são denominados de *Web Robots*. Um *web robot*, é um método de busca focado no conteúdo disponível na Internet, através do código HTML do próprio *site* (BAEZA-YATES; CASTILLO, 2002).

Os métodos de predição de epítomos consistem em reconhecer padrões de determinados epítomos e determinar o nível de compatibilidade de uma sequência genética. Estes métodos podem ser implementados utilizando diversas técnicas, podendo ser por: Métodos probabilísticos, Redes Neurais Artificiais, Cadeias de Markov Escondidas, Matriz de Escore por posição Específicas, entre outros. Porém, nenhuma das metodologias conhecidas de predição de epítomos assumem a existência de eventos evolutivos em seus algoritmos e, portanto, são incapazes de avaliar a seleção natural nas sequências analisadas.

Desta forma, é utilizado *Web Robots* para obter as predições de diferentes *sites* preditores e submeter estes resultados ao processo de Calibração e Consolidação, desenvolvido para identificação de potenciais epítomos para buscas in-vitro. Os *sites* preditores utilizados são:

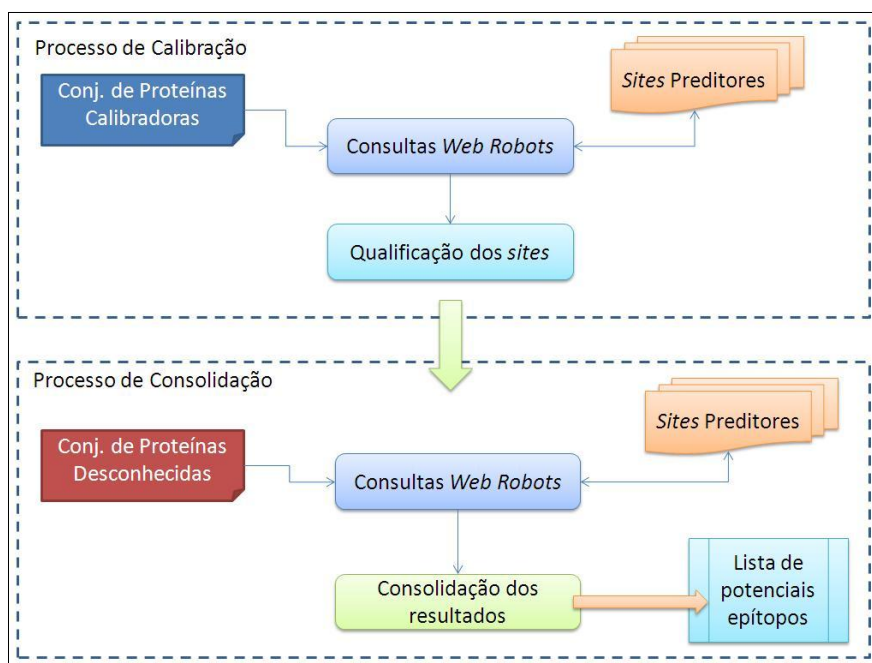
- Bimas (<http://www-bimas.cit.nih.gov/>);
- IEDB Consensus (<http://tools.immuneepitope.org/>);
- NETMHC (<http://www.cbs.dtu.dk/>);
- SYFPEITHI (<http://www.syfpeithi.de>).

O processo de Calibração consiste em determinar a qualidade de predição dos diferentes algoritmos de predição de epítopo a partir de um conjunto de proteínas (conjunto de calibragem de proteína). Cada proteína sobre um conjunto é chamado de proteína de calibração e tem pelo menos um epítopo de calibração (confirmado experimentalmente). Os calibradores é fator essencial para obter resultados que possam ser considerados ótimos, pois a má definição desses calibradores interfere diretamente na qualidade dos algoritmos de previsão de epítopos, por consequência na lista de epítopos candidatos. A nota para cada algoritmo de predição é entre zero e dez. Quanto mais próximo de dez, melhor é o algoritmo e seus resultados possuem mais valor em relação aos outros algoritmos utilizados.

Cada proteína de calibração é individualmente consultada nos diferentes algoritmos de predição de epítopos, e cada algoritmo retorna uma lista ordenada com os melhores epítopos no topo. Esta lista é ordenada de acordo com uma pontuação de peptídeos, sendo que, cada proteína de calibração tem pelo menos um epítopo conhecido e que deverá aparecer no topo da lista. Portanto, a pontuação do epítopo (posição na lista) é utilizada como um indicador de qualidade algoritmo de previsão. Um algoritmo perfeito seria classificar os epítopos calibradores no topo da lista.

O processo de Consolidação cria uma lista de potenciais epítopos para buscas in-vitro. Esta lista é definida a partir de uma pontuação unificada atribuída a cada epítopo baseado nas previsões de todos os *sites* utilizados. Para cada epítopo previsto, a pontuação unificada é definida pela soma da sua pontuação definida a partir da sua posição na lista e a nota do respectivo *site* preditivo.

Na Figura 1 é ilustrado todo o processo realizado na identificação de potenciais epítopos que promovem resposta imunológica no hospedeiro.



**Figura 1: Processo de identificação de potenciais epítopos**

O processo de calibração precisa ser feito apenas uma única vez para cada tipo de organismo, pois o resultado da Qualificação dos *sites* é baseado em epítopos conhecidos e validados experimentalmente, sendo assim, ao realizar uma busca a partir de proteínas desconhecidas, os epítopos serão selecionados conforme as notas dos *sites* previamente estabelecidos e sua posição na lista de cada *site* utilizado.

## RESULTADOS

Como resultado foi desenvolvido um *software* denominado como EPIBOT que implementa o método computacional proposto e também foi desenvolvido um processo de validação com a finalidade de garantir a eficiência do método computacional de predição de epítomos implementado pelo EPIBOT. Nesta sessão serão descritos todas as etapas envolvidas para a validação dos resultados gerados pelo software EPIBOT. Desta forma, tanto as previsões geradas pelos *sites* quando as geradas pelo EPIBOT serão submetidas a métodos que quantificam suas previsões de forma a serem comparados de maneira quantitativa, ou seja, através a notas atribuídas as suas predições.

Os dados necessários pra o desenvolvimento do processo de validação dos métodos de Calibração e Consolidação foram disponibilizados pelo Dr. Artur Queiroz do Laboratório de Imunologia (LIP) do Centro de Pesquisas Gonçalo Moniz da Fundação Oswaldo Cruz (CPqGM/FIOCRUZ), contendo um conjunto composto por 1661 proteínas e seus respectivos calibradores.

A validação foi feita apenas para o organismo dos Ratos, pois é a atual área de pesquisa do Dr. Artur Queiroz sendo que as proteínas disponibilizadas estão contidas no seu trabalho de pesquisa referente à *Leishmania Braziliensis*, sendo utilizado os alelos H2-Kb, H2-Kd e H2-Kk, cujos tamanhos dos epítomos são 8, 9 e 10 meros. Além disto, para cada alelo dois tipos de proteínas são definidos, sendo do tipo Positivo (composto por 535 proteínas) e Negativo (composto por 1126 proteínas). Cada tipo possui um conjunto de proteínas e outro conjunto contendo seus respectivos epítomos calibradores.

Foi feito a Calibração dos *sites* preditores referente a cada alelo e tamanho, sendo a calibração constituída pelo conjunto de todas as proteínas pertencente ao tipo Positivo, pois esse tipo é composto por epítomos conhecidos e validados experimentalmente e que disparam a atividade imunológica. O conjunto do tipo Negativo também é composto por epítomos conhecidos e validados experimentalmente, porém, possuem características de não dispararem a atividade imunológica no organismo, desta forma esse tipo não é utilizado durante o processo de calibração, apenas na validação dos resultados, pois é verificado se suas predições ocorreram longe das primeiras posições da lista gerada pelo processo de Consolidação.

A Figura 2 mostra os resultados obtidos para cada *site* conforme o tamanho e alelo, a partir dos resultados obtidos, alguns *sites* tem uma nota maior enquanto outros uma nota menor, essa nota baixa é referente à qualidade da predição associada a este *site*.

Alelo H2- Kb			Alelo H2- Kd			Alelo H2- Kk		
Tamanho	Site	Nota	Tamanho	Site	Nota	Tamanho	Site	Nota
8mer	Bimas	5,16	8mer	Bimas	5,53	8mer	Bimas	6,33
	SYFPEITHI	6,52		NETMHC	1,12		SYFPEITHI	6,00
9mer	Bimas	2,23	9mer	Bimas	6,05	9mer	Bimas	5,29
	IEDB_Consensus	0,17		SYFPEITHI	6,11		SYFPEITHI	5,43
10mer	Bimas	2,17		IEDB_Consensus	0,12		IEDB_Consensus	0,11
	IEDB_Consensus	0,13	10mer	Bimas	5,64			
				SYFPEITHI	5,55			

**Figura 2: Qualidade dos *sites* preditores**

A validação dos resultados gerados pelo EPIBOT ocorre com a comparação da lista de potenciais epítomos com as predições geradas por cada *site* predictor individualmente. Desta forma, será utilizado outro conjunto de proteínas, contendo 300 proteínas que possuem epítomos conhecidos, e feito uma análise para verificar se esses epítomos foram previstos nas primeiras posições da lista gerada pelo processo de Consolidação.

O método de validação constituído pela Análise ROC (*Receiver Operating Characteristic*) classifica os epítomos com finalidade de avaliar a taxa de acerto contida na sua consolidação, e para a comparação desses resultados às previsões de cada *site* também são classificadas pelo mesmo método

de validação, desta forma podemos demonstrar a eficácia do método de predição de epítomos implementado pelo EPIBOT.

Utilizando os testes de *Accuracy* (Avalia a proporção de todas as predições corretas, sobre todos os resultados obtidos) e *Precision* (Indica a capacidade de uma predição detectar corretamente), definidos pela Análise ROC, para os resultados gerados pelo EPIBOT e pelos *sites* preditores é ilustrado na Figura 3 gráficos comparativos que demonstra qualitativamente a eficácia da utilização do método proposto para predição de epítomos implementado pelo software EPIBOT.

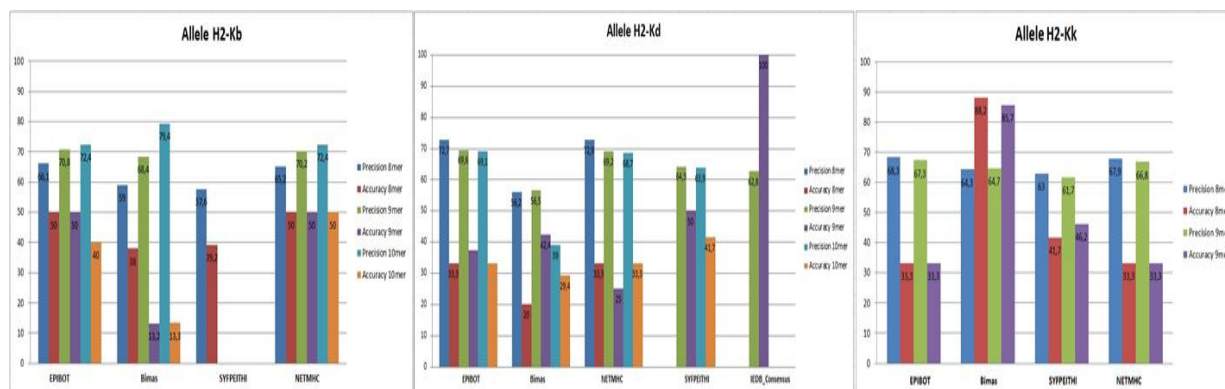


Figura 3: Gráficos dos resultados da validação

## CONSIDERAÇÕES FINAIS

Com os resultados obtidos podemos afirmar que o método de Calibração e Consolidação utilizados pelo EPIBOT não permite um ganho significativo em relação aos métodos existentes na Internet, porém, ao se tratar de uma área de construção de vacinas, sendo que não se há nada preciso para a avaliação de epítomos, qualquer ganho representa um grande passo no processo de pesquisa e desenvolvimento no processo de avaliação in-silico de epítomos. O EPIBOT é uma ferramenta eficaz na obtenção de um conjunto de epítomos capazes de disparar a atividade imunológica para uma determinada vacina é alta, de forma há reduzir o tempo e custo de uma pesquisa.

O sistema computacional desenvolvido, denominado de EPIBOT, está disponível na Internet no endereço <http://sites.ecomp.uefs.br/angeloduarte/epibot>, juntamente com o manual de usuário. Esse artigo tem como finalidade a divulgação na comunidade acadêmica para realização de pesquisas e produção de vacinas a partir de métodos in-silico.

## REFERÊNCIAS

- BAEZA-YATES, R.; CASTILLO, C. Balancing volume, quality and freshness in web crawling. In: SOFT COMPUTING SYSTEMS – Design, Management and Applications, 2002, p. 565–572, Santiago, Chile.
- DOYTCHINOVA, Irini A.; GUAN, Pingping; FLOWER, Darren R. EpiJen: a server for multistep T cell epitope prediction. BMC Bioinformatics. Disponível em: <<http://www.biomedcentral.com/1471-2105/7/131>>. Acesso em: 01 Agosto 2012.
- Parker, K.C. et al. (1994) Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. Journal of Immunology, 152, 163–75.
- Rammensee, H. et al. (1999) SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics, 50, 213–219.